Information-Theoretic Minimal Sufficient Representation for Multi-Domain Knowledge Graph Completion

Jiawei Sheng[1,2], Taoyu Su[1,2*], Weiyi Yang[3], Linghui Wang[1,2], Yongxiu Xu[1,2], Tingwen Liu[1,2]

[1] Institute of Information Engineering, Chinese Academy of Sciences, Beijing, China
[2] School of Cyber Security, University of Chinese Academy of Sciences, Beijing, China
[3] CCSE, School of Computer Science and Engineering, Beihang University, Beijing, China
{shengjiawei, sutaoyu, wanglinghui, xuyongxiu, liutingwen}@iie.ac.cn, yangweiyi@buaa.edu.cn

Abstract

Multi-domain knowledge graph completion (MKGC) seeks to predict missing triples in a target KG by leveraging triples from multiple KGs in different domains (*e.g.*, languages or sources). Existing studies typically learn and fuse multi-domain KG representations solely with alignments or fusion modules, which can be affected by redundant information within KGs. This issue can conceal task-relevant information in representations, impeding further improvements when scaling to numerous KGs. To this end, we propose IMKGC, an information-theoretic MKGC framework to learn minimal sufficient representations. In particular, IMKGC learns entity representations by explicitly preserving endogenous contextual information within each KG, exogenous complementary information from other KGs, and consistent information of equivalent entities, while suppressing redundant information through variational constraints. Furthermore, we achieve compressed relation representations with a devised relation reasoning decoder that captures relatedness among relations, also improving triple prediction. Extensive experiments on 14 KGs in three benchmark datasets demonstrate that IMKGC significantly outperforms previous state-of-the-art methods, especially in redundant scenarios.

Introduction

Knowledge graphs (KGs) structurally represent relational facts through triples, supporting knowledge-driven applications like question answering (Linders and Tomczak 2025) and retrieval-augmented generation (Pan et al. 2024). However, KGs usually suffer from incompleteness, motivating KG completion (KGC) by predicting missing triples from observed ones (Wang et al. 2017). The recent proliferation of separately constructed KGs can offer complementary knowledge to enhance completion of each KG. Therefore, multi-domain KG completion (MKGC) emerges as a new paradigm, which aims to leverage KGs from various domains, *e.g.*, languages (Tang et al. 2023) or sources (Sun et al. 2023), to address incompleteness in each of the KGs.

In general, MKGC seeks to predict missing triples in a target KG by leveraging triples from multiple KGs. As shown in Figure 1, predicting (`AppleInc`, `FoundedBy`, ?) in KG-2 is challenging due to the limited entity contextual neighbors. Crucially, KG-1 provides useful evidence for prediction. Here, the entities (*e.g.*, `StevenJobs`) existing in multiple KGs are called *equivalent entities*, which are aligned previously (Sun et al. 2020) and connect these KGs.

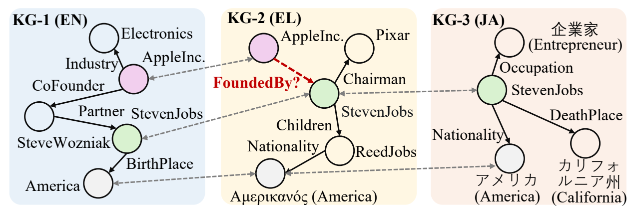


Figure 1: A toy example of MKGC, which predicts missing triples in the one KG using the other available KGs. Equivalent entities are connected by dashed lines.

Although essential, MKGC remains underexplored. Most existing studies (Chen et al. 2020; Zhu et al. 2020; Singh et al. 2021; Huang et al. 2022; Tang et al. 2023; He and Yang 2024) learn separate representations for each KG and connect them through equivalent entities. However, these methods predominantly rely on entity alignment losses or attention-based fusion to transfer information across KGs. We argue that such representations may be suboptimal due to the redundant information saturated in multi-domain KGs (*e.g.* repetitive trivial triples diluting task cues), which may pollute representations. For example, KG-3 triples may not be helpful in predicting (`AppleInc`, `FoundedBy`, ?) in Figure 1. According to information theory (Tishby, Pereira, and Bialek 1999; Tishby and Zaslavsky 2015), neural models tend to learn representations containing excessive information beyond task requirements, potentially fitting false clues and obscuring task-relevant patterns (Shwartz-Ziv and Tishby 2017). This issue would be exacerbated in MKGC when scaling to multiple KGs (*e.g.*, 3–6 KGs in our setting), where redundant information accumulates across domains.

To improve MKGC, we take advantage of the *Information Bottleneck* (IB) (Tishby and Zaslavsky 2015; Alemi et al. 2017), which seeks to learn representations that are sufficient for task prediction, but contain minimal redundant information. We identify three key types of critical informa-

*Corresponding author.

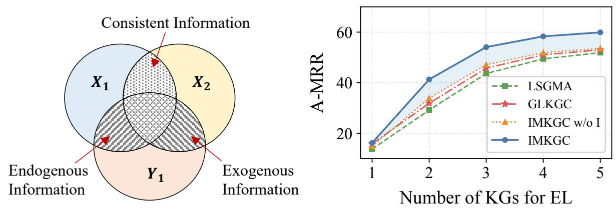


Figure 2: Left: Information perspective of predicting task Y_1 of KG-1 with X_1 and X_2 of KG-1 and -2 data. Right: Model results (%) on a KG (EL) in DBP-5L with increased KGs. Details are introduced in experiments.

tion for MKGC (illustrated in Figure 2 (left)):

- *Endogenous task-relevant information*: contextual cues within the current KG.
- *Exogenous task-relevant information*: complementary knowledge from other KGs.
- *Consistent alignment information*: shared semantics of equivalent entities across KGs.

Our objective is to learn entity representations that satisfy *sufficiency* by preserving the above information, while enforcing *minimality* by regularizing the total information. This naturally suppresses redundancies, thus achieving the desired *minimal sufficient representation* (Tishby and Zaslavsky 2015) for MKGC. Consequently, as shown in Figure 2 (right), our model exploits task-relevant information from increased KGs more effectively than typical MKGC models, achieving superior performance.

Following this idea, we propose IMKGC, an Information-theoretic representation framework for Multi-domain KG Completion. First, we devise an entity encoder to generate variational representations for entities across all KGs. Then, we impose the *information-theoretic constraints* to ensure sufficiency and minimality for the task. The refined representations from different KGs can be fused to make final predictions for each of the KGs. Furthermore, we recognize that several relations also exhibit semantic relatedness (Huang et al. 2022), such as `deathPlace` and `birthPlace` are both location-related relations. Independently learning these relations neglects their relatedness and can also generate redundant representations. Therefore, we propose a relation reasoning decoder based on *residual vector quantization* (RVQ) (Zeghidour et al. 2022; Lee et al. 2022) to capture their relatedness with compressed relation representations. Our major contributions can be summarized as follows:

- We introduce an essential information-theoretic perspective for the practical MKGC task.
- We propose a novel variational framework to derive the minimal sufficient representation for IMKGC.
- We leverage four information-theoretic constraints to improve entity representations and a relation reasoning decoder to capture relational relatedness.
- Extensive experiments[1] on 14 KGs of three benchmarks indicate the effectiveness and generality of IMKGC.

[1]Our source code and appendix are available at https://github.com/JiaweiSheng/IMKGC for future research.

Preliminaries

Task Formulation

Formally, consider that there are N KGs as $\mathcal{G}_i \in \mathcal{D}, |\mathcal{D}| = N$. Between each of the two KGs $\mathcal{G}_i = (\mathcal{E}_i, \mathcal{R}_i, \mathcal{T}_i)$ and $\mathcal{G}_j = (\mathcal{E}_j, \mathcal{R}_j, \mathcal{T}_j)$, there exists a small set of equivalent entities, $\mathcal{S}_{ij} = \{(e_i, e_j^+)|e_i \in \mathcal{E}_i, e_j^+ \in \mathcal{E}_j\}$. In addition, all relations are presented within a unified schema $\mathcal{R}$, *i.e.*, $\mathcal{R}_i \subseteq \mathcal{R}$ for $i = 1, 2, \cdots, N$. The task is to predict new triples in a query $(h, r, ?)$ with candidate entities $\mathcal{E}_i$ in each target KG $\mathcal{G}_i \in \mathcal{D}$, based on existing triples from both the target KG $\mathcal{G}_i$ and the other KGs $\mathcal{G}_j \in \mathcal{D}(j \neq i)$.

Information Bottleneck

The information bottleneck (IB) (Tishby, Pereira, and Bialek 1999; Tishby and Zaslavsky 2015) aims to learn minimal sufficient representations from the data to achieve predictions, which indicates a profound principle that *forgetting is the most important part of learning*. Formally, the IB principle finds a maximally compressed representation $\boldsymbol{Z}$ of the input data $\boldsymbol{X}$ while preserving information about the target $\boldsymbol{Y}$, which can be achieved by minimizing:

$$\mathcal{J} = \beta I(\boldsymbol{Z}; \boldsymbol{X}) - I(\boldsymbol{Z}; \boldsymbol{Y}), \tag{1}$$

where $I(\cdot;\cdot)$ denotes mutual information and factor $\beta > 0$ is a Lagrangian multiplier that controls the trade-off between compression and preservation of information. The first term penalizes the information between $\boldsymbol{Z}$ and $\boldsymbol{X}$, regularizing the variable $\boldsymbol{Z}$ to forget task-irrelevant information. The second term ensures $\boldsymbol{Z}$ to be predictive of $\boldsymbol{Y}$, preserving task-relevant information. Here, $\boldsymbol{Z}$ acts as a minimal sufficient statistic of $\boldsymbol{X}$ to predict $\boldsymbol{Y}$ (Alemi et al. 2017).

Method

Overview

In this paper, we seek a general information-theoretic representation framework for MKGC. For better illustration, we take two KGs $\mathcal{G}_1$ and $\mathcal{G}_2$ as an example, and the task aims to predict new triples in $\mathcal{G}_1$ with existing triples from both $\mathcal{G}_1$ and $\mathcal{G}_2$. The equivalent entity set $\mathcal{S}_{12} = \{(e_1, e_2^+)|e_1 \in \mathcal{G}_1, e_2^+ \in \mathcal{G}_2\}$ is given. Assume that the entity representations learned from $\mathcal{G}_1$ and $\mathcal{G}_2$ are $\boldsymbol{Z}_1$ and $\boldsymbol{Z}_2$, respectively. The entire information contained in data $\mathcal{G}_1$ and $\mathcal{G}_2$ is $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$, and the information required to predict the MKGC task in $\mathcal{G}_1$ is $\boldsymbol{Y}_1$. We propose the following trade-off constraints based on IB for expected ideal representations:

Definition 1 (Endogenous Constraint)**.** *The entity representations have sufficient task-relevant information to predict new triples in their located KG, yet have limited task-irrelevant data information, which is:*

$$\mathcal{J}_1(\boldsymbol{Y}_1, \boldsymbol{Z}_1) := \beta_1 I(\boldsymbol{Z}_1; \boldsymbol{X}_1) - I(\boldsymbol{Z}_1; \boldsymbol{Y}_1), \tag{2}$$

where $\beta_1 \in \mathbb{R}$ is a trade-off factor. This constraint requires the learned representations to have minimal sufficient information to predict new triples in the current KG.

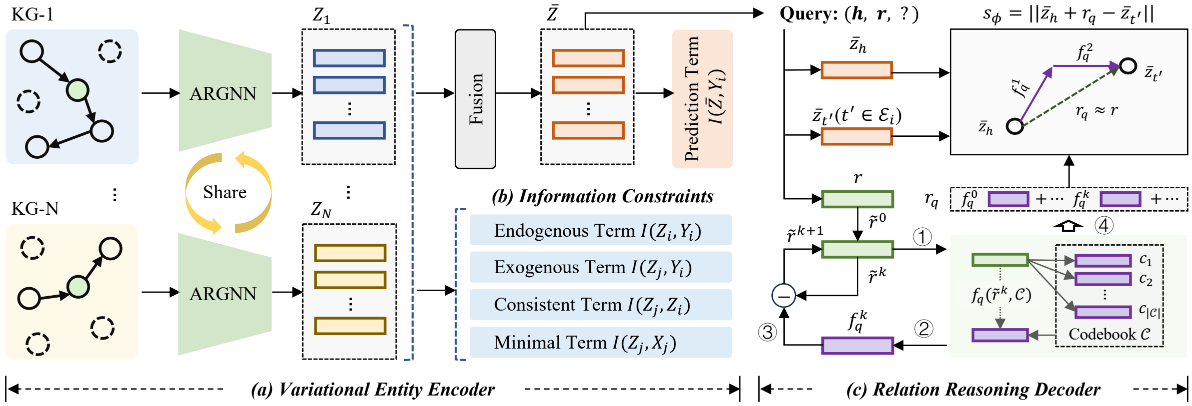


Figure 3: The overview of our proposed framework, IMKGC. It contains (a) variational entity encoder to learn entity representations, (b) information constraints to refine representations, and (c) relation reasoning decoder to predict missing triples.

Definition 2 (Exogenous Constraint)**.** *The entity representations learned from other KGs provide auxiliary information to predict triples in the target KGs, yet have limited task-irrelevant domain information, which is:*

$$\mathcal{J}_2(\boldsymbol{Y}_1, \boldsymbol{Z}_2) := \beta_2 I(\boldsymbol{Z}_2; \boldsymbol{X}_2) - I(\boldsymbol{Z}_2; \boldsymbol{Y}_1), \tag{3}$$

where $\beta_2 \in \mathbb{R}$ is also a trade-off factor. This constraint encourages the learning of information to make predictions in the target KG and limits excessive domain information.

Definition 3 (Consistent Constraint)**.** *The representations of equivalent entities learned from their located KGs retain consistent information between KGs, which is:*

$$\mathcal{J}_3(\boldsymbol{Z}_1, \boldsymbol{Z}_2) := -I(\boldsymbol{Z}_1; \boldsymbol{Z}_2). \tag{4}$$

where $\boldsymbol{Z}_1, \boldsymbol{Z}_2$ contain equivalent entities in $\mathcal{S}_{12}$. This regularizes them to have consistent information, which helps to align the representation space of different KGs.

By combining the above three constraints, we derive the **overall constraint** for N KGs: *To predict new triples for a target KG $\boldsymbol{X}_i$ with multiple available KGs $\{\boldsymbol{X}_j\}_{\forall j}$, the minimal sufficient constraint to predict $\boldsymbol{Y}_i$ is:*

$$\mathcal{J}_{\text{cons}}(\boldsymbol{Y}_i, \{\boldsymbol{X}_j\}_{\forall j}) := -\alpha \underbrace{I(\boldsymbol{Z}_i; \boldsymbol{Y}_i)}_{\text{Endogenous Term}} + \beta \sum_{\forall j} \underbrace{I(\boldsymbol{Z}_j; \boldsymbol{X}_j)}_{\text{Minimal Term}} - \omega \sum_{\forall j, j \neq i} \underbrace{I(\boldsymbol{Z}_j; \boldsymbol{Y}_i)}_{\text{Exogenous Term}} - \gamma \sum_{\forall j, j \neq i} \underbrace{I(\boldsymbol{Z}_j; \boldsymbol{Z}_i)}_{\text{Consistent Term}}, \tag{5}$$

where we unify $\beta := \beta_1 \simeq \beta_2$ with trade-off factors $\alpha, \omega, \gamma \in \mathbb{R}$. In this way, the total information of representation of each KG is limited, but critical information is preserved, thus meeting the minimal sufficient condition.

To make predictions in the target KG, we further fuse the representations of involved equivalent entities from all KGs. Then, the **final training objective** is:

$$\mathcal{J}_{\text{final}} := \sum_{\mathcal{G}_i \in \mathcal{D}} - \underbrace{I(\bar{\boldsymbol{Z}}; \boldsymbol{Y}_i)}_{\text{Prediction Term}} + \mathcal{J}_{\text{cons}}(\boldsymbol{Y}_i, \{\boldsymbol{X}_j\}_{\forall j}), \tag{6}$$

where $I(\bar{\boldsymbol{Z}}; \boldsymbol{Y}_i)$ is the prediction term with fused entity representations to predict new triples in the target KG $\mathcal{G}_i$.

Variational Entity Encoder

To derive the entity representation $\boldsymbol{Z}$ in each $\mathcal{G}$ (omitting the KG subscript for simplicity), we devise a variational entity encoder, which leverages an *attentive relational graph neural network* (termed ARGNN) (Tang et al. 2023) to encode the entity e with its relational neighbors $\mathcal{N}(e)$ as:

$$\begin{aligned}
\boldsymbol{e}^{l+1} &= \boldsymbol{e}^l + \delta\Big(\sum_{\{r_n, e_n\} \in \mathcal{N}(e)} \alpha(\boldsymbol{e}^l, \boldsymbol{r}_n, \boldsymbol{e}^l_n) \cdot \boldsymbol{W}^l_V [\boldsymbol{e}^l_n \oplus \boldsymbol{r}_n]\Big), \\
\alpha(\boldsymbol{e}^l, \boldsymbol{r}_n, \boldsymbol{e}^l_n) &= \text{softmax}(\text{S}(\boldsymbol{e}^l, \boldsymbol{r}_n, \boldsymbol{e}^l_n)), \\
\text{S}(\boldsymbol{e}^l, \boldsymbol{r}_n, \boldsymbol{e}^l_n) &= \beta_{r_n} \cdot \frac{1}{\sqrt{d}} (\boldsymbol{W}_Q \boldsymbol{e}^l \cdot \boldsymbol{W}_K [\boldsymbol{e}^l_n \oplus \boldsymbol{r}_n]),
\end{aligned} \tag{7}$$

where δ is the activation, *e.g.*, ReLU. $\oplus$ is the concatenation. $\boldsymbol{W}^l_V, \boldsymbol{W}^l_Q, \boldsymbol{W}^l_K$ are learnable parameters. $\boldsymbol{e}, \boldsymbol{r} \in \mathbb{R}^d$ are randomly initialized for entities and relations, respectively. Here, the function S considers both the prior weight $\beta_r \in \mathbb{R}$ of a relation r (Shang et al. 2019) and the attentive relational relevance. Then, based on the output $\boldsymbol{e}^L$ of the L-th layer, we generate the variational representation $\boldsymbol{z}$ of entity e via variational inference (Alemi et al. 2017):

$$\begin{aligned}
&\boldsymbol{\mu} = \text{MLP}(\boldsymbol{e}^L; \theta_\mu), \boldsymbol{\sigma} = \text{MLP}(\boldsymbol{e}^L; \theta_\sigma), \\
&\boldsymbol{z} \sim \mathcal{N}(\boldsymbol{\mu}, \text{diag}(\boldsymbol{\sigma}^2)),
\end{aligned} \tag{8}$$

where $\boldsymbol{z}$ is assumed to be a Gaussian distribution, which can be sampled as a deterministic embedding by using the reparameterization trick (Kingma and Welling 2014), *i.e.*, $\boldsymbol{z} = \boldsymbol{\mu} + \boldsymbol{\sigma} \odot \boldsymbol{\epsilon}, \boldsymbol{\epsilon} \sim \mathcal{N}(0, 1)$. All $\boldsymbol{z}$ constitute $\boldsymbol{Z}$ in a $\mathcal{G}$.

In addition, considering that there are N KGs, we share the initial embeddings, and learn each KG structure with a shared encoder. Following Tang et al. (2023), for an entity that does not exist in a certain KG, we add a virtual entity without neighbors as its equivalent entities, for simplicity in practice. In this way, each entity has N equivalent entities and generates representations $\{\boldsymbol{Z}_j\}_{j=1}^N$. These representations are refined with information constraints (detailed later). For final predictions in a target KG, we further fuse these refined representations as $\bar{\boldsymbol{Z}} = \frac{1}{N}\sum_{j=1}^N \boldsymbol{Z}_j$, using the equivalent entity representations from all KGs.

Tractable Information Constraint

For optimization, we derive tractable formulas for information constraints. The proofs are in the **Appendix**.

Endogenous & Exogenous & Prediction Term To preserve predictive information in entity representations, we impose the term $I(\boldsymbol{Z};\boldsymbol{Y})$. Here, we omit the subscripts of $I(\boldsymbol{Z}_i;\boldsymbol{Y}_i)$, $I(\boldsymbol{Z}_j;\boldsymbol{Y}_i)$ and $I(\bar{\boldsymbol{Z}};\boldsymbol{Y}_i)$, since they have similar derivations. We construct the variational lower bound (Alemi et al. 2017) to estimate the mutual information to maximize, which is:

$$
\begin{aligned}
I(\boldsymbol{Z};\boldsymbol{Y}) &\geq \mathbb{E}_{\boldsymbol{x},\boldsymbol{y}\sim p(\boldsymbol{x},\boldsymbol{y}),(\boldsymbol{z}_h,\boldsymbol{z}_t)\sim p_\theta(\boldsymbol{z}|\boldsymbol{x})}[\log q_\phi(\boldsymbol{y}|\boldsymbol{z}_h,\boldsymbol{r},\boldsymbol{z}_t)] \\
&\approx \sum_{(h,r,t)\in\mathcal{T}} \sum_{(h,r,t^-)\notin\mathcal{T}} [\lambda + \mathrm{s}_\phi(\boldsymbol{z}_h,\boldsymbol{r},\boldsymbol{z}_t) - \mathrm{s}_\phi(\boldsymbol{z}_h,\boldsymbol{r},\boldsymbol{z}_t^-)]_+,
\end{aligned}
\tag{9}
$$

where $\boldsymbol{y} \in \{0,1\}$ indicates the truth of a given triple. The decoder $q_\phi(\boldsymbol{y}|\boldsymbol{z}_h,\boldsymbol{r},\boldsymbol{z}_t)$ measures the triple plausibility, which can be achieved by s_ϕ in Eq. (15). Here, we estimate the likelihood with a margin loss, where $[\cdot]_+ = \max(\cdot,0)$ and λ is the margin. t^- is the negative entity randomly selected from the entity set of the target KG. In practice, to deal with a given triple in the target KG $\boldsymbol{Y}_i$, the representation $\boldsymbol{Z}_i$, $\boldsymbol{Z}_j$ and $\bar{\boldsymbol{Z}}$ are selected correspondingly to achieve $I(\boldsymbol{Z}_i;\boldsymbol{Y}_i)$, $I(\boldsymbol{Z}_j;\boldsymbol{Y}_i)$ and $I(\bar{\boldsymbol{Z}};\boldsymbol{Y}_i)$, respectively.

Minimal Term To limit redundant information in entity representations, we derive the minimal term $I(\boldsymbol{Z};\boldsymbol{X})$. Formally, we measure it by the Kullback-Leibler (KL) divergence (Alemi et al. 2018) with a variational approximation posterior distribution. The upper bound is:

$$
\begin{aligned}
I(\boldsymbol{Z};\boldsymbol{X}) &\leq \mathbb{D}_{KL}(p_\theta(\boldsymbol{z}|\boldsymbol{x})||r(\boldsymbol{z})) \\
&= \mathbb{D}_{KL}(\mathcal{N}(\boldsymbol{\mu},\mathrm{diag}(\boldsymbol{\sigma}^2))||\mathcal{N}(0,\mathrm{diag}(\boldsymbol{I}))),
\end{aligned}
\tag{10}
$$

where $p_\theta(\boldsymbol{z}|\boldsymbol{x})$ is the variational encoder to approximate the true posterior $p(\boldsymbol{z}|\boldsymbol{x})$. Following studies (Kingma and Welling 2014; Alemi et al. 2017), we also assume the prior distribution $r(\boldsymbol{z})$ as a standard Gaussian $\mathcal{N}(0,\mathrm{diag}(\boldsymbol{I}))$. This actually regularizes the total information in representations, which are also required to preserve useful information, thus surpassing redundant information.

Consistent Term To further encourage equivalent entities with consistent information, we devise the term $I(\boldsymbol{Z}_i;\boldsymbol{Z}_j)$. Formally, we estimate it with its lower bound, *e.g.*, InfoNCE (He et al. 2020), which is:

$$
I(\boldsymbol{Z}_i;\boldsymbol{Z}_j) \geq \sum_{(e_i,e_j^+)\in\mathcal{S}_{ij}} \log \frac{\exp(\cos(\boldsymbol{z}_i,\boldsymbol{z}_j^+)/\tau)}{\sum_{\boldsymbol{z}_j\in\mathcal{E}_j}\exp(\cos(\boldsymbol{z}_i,\boldsymbol{z}_j)/\tau)},
\tag{11}
$$

where τ is the temperature, $\mathcal{E}_j$ is the entity set of $\mathcal{G}_j$. Note that N KGs exist $N(N-1)/2$ possible KG pairs, and we randomly sample two KGs in a training step for acceleration.

Relation Reasoning Decoder

To measure the triple plausibility in Eq. (9), we can simply use a score function such as TransE (Bordes et al. 2013). However, we notice that several relations can have common semantics, such as both relation `birthPlace` and `deathPlace` have location-related semantics. Therefore, we would like to capture the relational relatedness to reduce redundancies and promote relation representations.

Inspired by *vector quantization (VQ)* (van den Oord, Vinyals, and Kavukcuoglu 2017), we consider to represent a relation with several sub-relations shared by other relations. In general, VQ quantifies (maps) a relation $\boldsymbol{r} \in \mathbb{R}^d$ to a code $\boldsymbol{c} \in \mathbb{R}^d$ in a shared codebook $\mathcal{C}$, $|\mathcal{C}| \ll |\mathcal{R}|$:

$$
\begin{aligned}
&f_q(\boldsymbol{r},\mathcal{C}) = \boldsymbol{r} + \mathrm{sg}[\boldsymbol{c}^* - \boldsymbol{r}], \\
&\boldsymbol{c}^* = \mathrm{argmin}_{\boldsymbol{c}\in\mathcal{C}}||\boldsymbol{r}-\boldsymbol{c}||_2,
\end{aligned}
\tag{12}
$$

where $\boldsymbol{c}^*$ is the most related code of $\boldsymbol{r}$ in $\mathcal{C}$. $\mathrm{sg}[\cdot]$ is the stop gradient operation (Bengio, Léonard, and Courville 2013), such that $f_q(\boldsymbol{r},\mathcal{C})$ outputs $\boldsymbol{c}^*$ for prediction and yet $\boldsymbol{r}$ for training, quantifying relations to codes without affecting training. Unlike studies (Sachan 2020) by vector segmentation, we decompose the relation $\boldsymbol{r}$ with *residual vector quantization (RVQ)* (Zeghidour et al. 2022; Lee et al. 2022) as

$$
\begin{aligned}
&\tilde{\boldsymbol{r}}^{k+1} = \tilde{\boldsymbol{r}}^k - f_q^k, \\
&f_q^k := f_q(\tilde{\boldsymbol{r}}^k,\mathcal{C}).
\end{aligned}
\tag{13}
$$

Here, $\tilde{\boldsymbol{r}}^k \in \mathbb{R}^d$ is the residual in steps, and $\tilde{\boldsymbol{r}}^0 = \boldsymbol{r}$. In this way, the quantified (reconstructed) representation $\boldsymbol{r}_q$ of $\boldsymbol{r}$ is:

$$
\boldsymbol{r}_q := f_q^0 + \cdots + f_q^k + \cdots + f_q^{K-1} \in \mathbb{R}^d,
\tag{14}
$$

Interestingly, this decomposition naturally satisfies the translational assumption (Bordes et al. 2013), where a head entity is translated by K steps to the tail entity in reasoning. Since different relations can be decomposed by the same code (as sub-relation f_q^k), this design captures the relational relatedness, which is like `birthPlace` = `birth` + `place` and `deathPlace` = `death` + `place`, and also compresses redundant relation representations with shared codes. Using $\boldsymbol{r}_q$, the final predicted triple plausibility by translation is

$$
\mathrm{s}_\phi(\bar{\boldsymbol{z}}_h,\boldsymbol{r}_q,\bar{\boldsymbol{z}_{t'}}) = ||\bar{\boldsymbol{z}}_h + \boldsymbol{r}_q - \bar{\boldsymbol{z}}_{t'}||_2,
\tag{15}
$$

where ϕ denotes the learnable codes in $\mathcal{C}$. In addition, a commitment loss (van den Oord, Vinyals, and Kavukcuoglu 2017) is also imposed to train the codebook. In practice, we use $\boldsymbol{r}_q$ to achieve all $I(\boldsymbol{Z};\boldsymbol{Y})$ in Eq. (9). For the final prediction of a query $(h,r,?)$, the candidate $t' \in \mathcal{E}$ with the highest triple score is predicted as the true tail entity.

Experiments

Settings

Datasets We adopt three benchmarks with 14 KGs in our experiments: two multilingual datasets **DBP-5L** (Chen et al. 2017), **E-PKG** (Huang et al. 2022), and a constructed multi-domain dataset **DWY** (Sun et al. 2018). The DBP-5L dataset consists of 5 KGs extracted from DBpedia constructed in Greek (EL), English (EN), Spanish (ES), French (FR) and Japanese (JA). The E-PKG dataset is an e-commerce dataset of mobile phone-related product information in 6 languages, including German (DE), English (EN), Spanish (ES), French (FR), Italian (IT) and Japanese (JA). We further introduce a multi-domain dataset based on DWY, including DBpedia (DB), YAGO (YG) and Wiki (WK). For all datasets,

Methods	EL			EN			ES			FR			JA			AVG
	H@1	H@10	MRR	H@1	H@10	MRR	H@1	H@10	MRR	H@1	H@10	MRR	H@1	H@10	MRR	MRR
TransE	13.1	43.7	24.3	7.3	29.3	16.9	13.5	45.0	24.4	17.5	48.8	27.6	21.1	48.5	25.3	23.7
DistMult	8.9	11.3	9.8	8.8	30.0	18.3	7.4	22.4	13.2	6.1	23.8	14.5	9.3	27.5	15.8	14.3
RotatE	14.5	36.2	26.2	12.3	30.4	20.7	21.2	53.9	33.8	23.2	55.5	35.1	26.4	60.2	39.8	31.1
KG-BERT	17.3	40.1	27.3	12.9	31.9	21.0	21.9	54.1	34.0	23.5	55.9	35.4	26.9	59.8	38.7	31.3
KEnS	28.1	56.9	-	15.1	39.8	-	23.6	60.1	-	25.5	62.9	-	32.1	65.3	-	-
CG-MuA	21.5	44.8	32.8	13.1	33.5	22.2	22.3	55.4	34.3	24.2	57.1	36.1	27.3	61.1	40.1	33.1
AlignKGC	27.6	56.3	33.8	15.5	39.2	22.3	24.2	60.9	35.1	24.1	62.3	37.4	31.6	64.3	41.6	34.0
SS-AGA	30.8	58.6	35.3	16.3	41.3	23.1	25.5	61.9	36.6	27.1	65.5	38.3	34.6	66.9	42.9	35.2
LSMGA	33.1	89.9	54.5	16.8	61.7	32.4	25.6	74.8	42.8	31.2	81.3	48.6	33.5	79.1	49.8	45.6
GLKGC†	36.6	86.5	53.0	17.1	60.2	32.9	28.3	74.4	43.6	31.5	78.4	47.9	36.5	77.6	50.9	45.7
IMKGC	**38.2**	**90.9**	**59.9**	**18.2**	**62.0**	**33.8**	**30.6**	**77.5**	**47.5**	**35.4**	**82.5**	**52.5**	**37.3**	**81.5**	**53.9**	**49.5**

Table 1: Experimental results (%) on the DBP-5L dataset. † means the model is re-implemented.

Methods	DE			EN			ES			FR			IT			JA			AVG
	H@1	H@10	MRR	H@1	H@10	MRR	H@1	H@10	MRR	H@1	H@10	MRR	H@1	H@10	MRR	H@1	H@10	MRR	MRR
TransE	21.2	65.5	37.4	23.2	67.5	39.4	17.2	58.4	33.0	20.8	66.9	37.5	22.0	63.8	37.8	25.1	72.7	43.6	38.1
DistMult	21.4	54.5	35.4	23.8	60.1	37.2	17.9	46.2	30.9	20.7	53.5	35.1	22.8	51.8	34.8	25.9	62.6	38.0	35.2
RotatE	22.3	64.3	38.2	24.2	66.8	40.0	18.3	58.9	33.7	22.1	64.3	38.2	22.5	64.0	38.1	26.3	71.9	41.8	38.3
KG-BERT	21.8	64.7	38.4	24.3	66.4	39.6	18.7	58.8	33.2	22.3	67.2	38.3	22.9	63.7	37.2	26.9	72.4	44.1	38.5
KEnS	24.3	65.8	-	26.2	69.5	-	21.3	59.5	-	25.4	68.2	-	25.1	64.6	-	33.5	73.6	-	-
CG-MuA	22.9	64.9	38.7	24.8	67.9	40.2	19.2	58.8	33.8	23.0	67.5	39.1	23.9	63.8	37.6	30.4	72.9	45.9	39.2
AlignKGC	22.1	65.1	38.5	25.6	68.3	40.5	19.4	59.1	34.2	22.8	67.2	38.8	24.2	63.4	37.3	31.2	72.3	46.2	39.3
SS-AGA	24.6	66.3	39.4	26.7	69.8	41.5	21.0	60.1	36.3	25.9	68.7	40.2	24.9	63.8	38.4	33.9	74.1	48.3	40.7
LSMGA	30.7	68.5	44.8	31.9	70.2	45.9	23.1	61.1	36.5	23.7	63.5	38.2	26.8	64.5	41.0	43.7	78.4	57.1	43.9
GLKGC†	24.1	63.6	37.7	27.1	58.4	39.4	24.6	61.0	36.8	22.1	62.3	36.4	27.0	63.7	40.4	44.1	76.4	57.5	41.4
IMKGC	**30.9**	**69.6**	**45.1**	**36.1**	**70.3**	**48.7**	**25.1**	**61.9**	**37.5**	**27.6**	**69.0**	**42.1**	**33.4**	**65.6**	**45.8**	**44.6**	**79.2**	**57.9**	**46.2**

Table 2: Experimental results (%) on the E-PKG dataset. † means the model is re-implemented.

the equivalent entities are given to connect each of the two KGs. The relations are unified and shared across all KGs for practical considerations (Chen et al. 2017; Tang et al. 2023). More statistics are presented in the **Appendix**.

Competitors We adopt state-of-the-art methods as our competitors, including: (a) ***Single-domain methods***, which learn KG embeddings separately for prediction, including **TransE** (Bordes et al. 2013), **DisMult** (Yang et al. 2015), **RotatE** (Sun et al. 2019), **KG-BERT** (Yao, Mao, and Luo 2020). (b) ***Multi-domain methods***, which learn embeddings on each KG with graph encoders, and adopt entity alignment or attention modules to fuse features for prediction, including **KEnS** (Chen et al. 2020), **CG-MuA** (Zhu et al. 2020), **AlignKGC** (Singh et al. 2021), **SS-AGA** (Huang et al. 2022), **LSMGA** (Tang et al. 2023) and **GLKGC** (He and Yang 2024). (c) ***LLM-based methods***, which involves general knowledge to make each KG prediction, including **ChatGPT-3.5** (Zhu et al. 2024), **KICGPT** (Wei et al. 2023), **MKGL** (Guo et al. 2024). For details, see **Appendix**.

Implementation Details Following previous studies (Chen et al. 2017; Tang et al. 2023), we evaluate models with the task of *tail entity prediction*. During training, we combine all the training data from the multiple KGs. In testing, we rank all candidate entities of the target KG to predict t given h and r for each triple $(h, r, ?)$ in the test data. Three metrics are reported, including Hits@10 (H@10 for short), Hits@1 (H@1) and mean reciprocal ranks (MRR). The optimal model is selected according to the average MRR of all KGs on their validation data, following Tang et al. (2023). Most hyperparameters are shared for all datasets. The entity and relation embeddings are randomly initialized

Methods	DB			WK			YG			AVG
	H1	H10	MRR	H1	H10	MRR	H1	H10	MRR	MRR
TransE	4.3	52.9	20.3	3.0	48.6	17.3	2.2	42.2	13.1	16.9
DistMult	8.6	36.5	17.6	8.4	41.7	18.4	4.6	32.5	12.7	16.2
RotatE	13.2	57.4	27.9	9.9	52.5	26.4	3.5	42.7	13.8	22.7
SS-AGA	5.8	61.8	22.6	6.6	52.2	18.5	9.0	52.3	22.9	21.3
LSGMA	14.0	64.3	30.9	9.5	54.6	23.9	11.4	48.6	23.5	26.1
GLKGC†	13.4	66.9	32.3	9.3	55.0	24.3	16.5	52.8	28.7	28.4
IMKGC	**15.2**	**68.7**	**33.9**	**11.6**	**58.4**	**26.5**	**21.5**	**67.8**	**36.6**	**32.3**

Table 3: Experimental results (%) on the DWY dataset. We abbreviate H@10, H@1 as H10, H1, respectively.

Variant	A-H@1	A-H@10	A-MRR	ΔA-MRR
Entire	31.9	78.9	49.5	-
repl. GCN	27.8	77.9	46.1	3.4↓
w/o RRD	29.6	77.0	47.6	1.9↓
w/o Endogenous	29.8	77.7	47.4	2.1↓
w/o Consistent	28.8	77.4	46.0	3.3↓
w/o Exogenous	24.9	76.9	45.8	3.7↓
w/o I (All Terms)	24.4	76.7	45.3	4.2↓

Table 4: Variant analyses on DBP-5L, where A-H@1, A-H@10 and A-MRR denote averaged MRR (%) of 5 KGs.

with dimension 256, and ARGNN has the hidden dimension 256 with 2 layers based on PyG[2] as Tang et al. (2023). The learning rate is set to 0.001, and the margin λ is set to 0.5 for all datasets. The reasoning step K is tuned from 1 to 5 and $|\mathcal{C}|$ is tuned in {0.1,0.2,...,0.9} ratio of total relation number. The trade-off factors $\alpha, \beta, \gamma, \omega$ are tuned in $\{1, 3, 5\} \times 10^{-\{4,3,2,1\}}$. For baselines, most of the results on DBP-5L and E-PKG are obtained from previous work. On DWY, we re-implement baselines with the reported best hyperparameters. For GLKGC without available codes, we re-implement it based on LSMGA with a transformer-based graph encoder. We employ a grid search strategy with three trials to select the optimal hyperparameters, which are reported in the **Appendix**.

Main Results

Method Comparison We compare our model with existing state-of-the-art methods in Table 1, 2 and 3, where we find that: *First, the multiple auxiliary KGs can indeed improve KGC on a target KG*. Most multi-domain KGC methods outperform single-domain KGC methods, which reflects the practical unity of completing KG with other domain KGs. *Second, our method significantly outperforms all existing methods.* In practice, our method outperforms existing methods with 3.8 averaged MRR improvements in DBP-5L, 4.8 in E-PKG and 3.9 in DWY. As the related methods LSMGA and GKMKGC impose no constraints in information fusion, we believe that our information constraints remarkably limit redundancies and unveil true features, leading to improvements. *Third, our method achieves a consistent gain on multilingual and multi-domain datasets with 14 KGs.* This reflects that our information-theoretic representation framework is general to different domains.

Variant Analyses To investigate the effectiveness of modules, we performed a variant analysis in Table 4: (a) repl. GCN replaces the ARGNN with vanilla non-relational GNN, which indicates that *the designations can be helpful in capturing the relational nature of KGs.* (b) w/o RRD removes the relation reasoning decoder using entire relation embeddings, which reflects that *RRD is able to capture the relatedness between the relations, improving the predictions.* (c) We ablate the information constraints, respectively. We

[2]https://pytorch-geometric.readthedocs.io/en/latest/

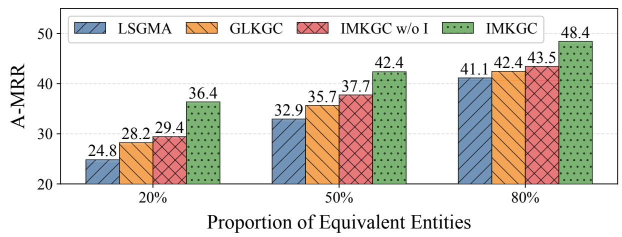


Figure 4: Results (%) with limited equivalent entities, randomly selected at 20%, 50% and 80% on DBP-5L.

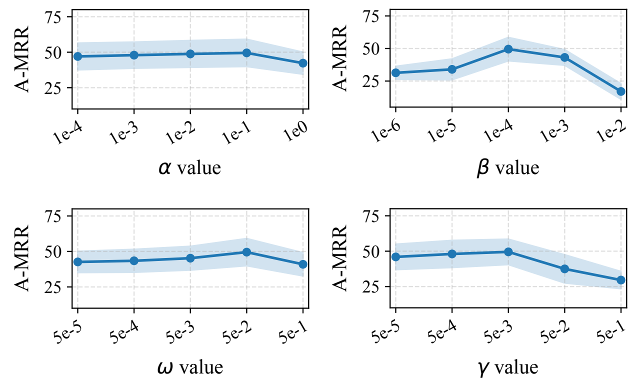


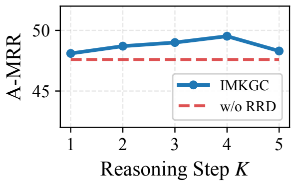
Figure 5: Impact of trade-off factors α, β, ω and γ. Averaged MRR (%) is reported with $\pm$ std of 5 KGs in DBP-5L.

find that both exogenous and consistent terms can build useful information by transferring from related KGs. The endogenous constraint enhances the in-KG features of each KG. Ablating all constraints (w/o I) leads to an obvious result decline. *All these results demonstrate the effectiveness of the proposed information constraints.*

Analyses on Information Constraints

Impact on More Auxiliary KGs To evaluate the model for handling redundancies, we reproduce the results of all methods with increased KGs, as shown in Figure 2 (right). Compared with existing methods, we find that: *for the prediction of a target KG (e.g. EL) with numerous auxiliary KGs, our model achieves significant improvements.* This supports that our information constraints help reveal true information relevant to the task from redundant scenarios.

Impact on Limited Equivalent Entities To evaluate the generality of our model, we conduct experiments on limited resources (20%, 50%, 80%) of equivalent entities, shown in Figure 4. We find that *our model achieves better results than the existing methods, especially with fewer equivalent entities*. In these situations with fewer resources, the connection between KGs is weak, and the confusing information would be relatively more. Thus, the model is required to grasp critical information to transfer for predictions. The results also indicate that our model learns minimal sufficient representations that can be robust to redundancies.



Relations with common codes
friend, partner, spouse, sponsor
regent, senators, president
locatedInArea, burialPlace, routeEnd
americanComedyAward, baftaAward

Figure 6: Evaluation on the relation reasoning decoder. Left: Impact (%) on the reasoning step K. Right: Case study on relations with 3 common quantified codes ($K = 4$).

Variant	A-H@1	A-H@10	A-MRR
ChatGPT-3.5	18.5	-	-
KICGPT	13.2	30.1	20.0
MKGL	13.3	30.8	19.3
IMKGC	**31.9**	**78.9**	**49.5**

Table 5: Comparison (%) with LLM models on DBP-5L.

Impact on Trade-off Factors We show the impact of factors α, β, ω and γ *w.r.t.* endogenous, minimal, exogenous and consistent terms in Figure 5. We find that α has less impact on the results, as it can be somewhat supplemented by the final prediction loss. A large β leads to excessive information compression, affecting the ability of tasks. γ controls the proportion of exogenous information, and a higher value would affect the endogenous features. In addition, ω helps align equivalent entities in features, yet a value too large for alignment would deviate from the MKGC task goal.

Analyses on Relation Modeling

Impact on the Reasoning Step To evaluate the RRD, we analyze the reasoning step number K in Figure 6 (left). Here, K also corresponds to the number of decomposed codes *w.r.t.* sub-relations. *By decomposing relations into shared sub-relations, the model captures the relatedness between relations and also generates compressed representations.* In addition, excessively increasing K would not continuously improve the result, where we believe that it may lead to relation representations with unclear distinctions.

Case Study on Related Relations The cases of relations with common codes are shown in Figure 6 (right). We find that the relations with common codes indeed have related semantics. For example, `friend` and `partner` have "friendship" semantics, and `senators` and `president` have "politics" semantics. This confirms the unity of the relatedness modeling of our relation reasoning decoder.

Results on LLM-Based Methods

We also conduct a comparison with LLM-based KGC models, which utilize the general parameterized knowledge in LLM and single-domain KG triples for prediction, shown in Table 5. We find that these methods can still hardly deal with MKGC, which are underexplored to utilize multi-domain triples. We leave further studies in our future works. The detailed results are reported in the **Appendix**.

Related Works

Knowledge graph completion (KGC) aims to predict missing triples (*i.e.*, relational facts) based on existing triples, usually in a single KG. Classical studies propose *triple-based methods* (Bordes et al. 2013; Yang et al. 2015; Sun et al. 2019; Trouillon et al. 2016; Sheng et al. 2020) by measuring the triple plausibility with translation-based (Bordes et al. 2013; Sun et al. 2019) or semantic matching-based functions (Yang et al. 2015; Dettmers et al. 2018). Later studies propose *GNN-based methods* (Shang et al. 2019; Schlichtkrull et al. 2018; Liu et al. 2024a) to capture relational graph structures. Existing studies also explore KGC with *language-based methods* (Xie et al. 2016; Yao et al. 2025), which attempt pre-trained language models (Yao, Mao, and Luo 2020; Wang et al. 2021) or large language models (Wei et al. 2023; Guo et al. 2024; Zhang et al. 2024; Liu et al. 2024b; Yao et al. 2025; Li et al. 2024; Yao et al. 2025) to predict new triples with massive parameterized knowledge. However, most studies assume to predict missing triples in a single purified KG, which can hardly utilize redundant multiple KGs for MKGC predictions.

Multi-domain knowledge graph completion (MKGC) aims to facilitate multiple KGs to improve KGC in each KG. Previous studies also explore this task in multilingual scenarios named *multilingual KGC* (Huang et al. 2022; Tang et al. 2023). For generality, this paper studies the task on multi-domain KGs, since the KGs may also come from different domains, not only languages. Technically, MTransE (Chen et al. 2017) first extends KG embeddings from one KG to multiple KGs. Later studies (Zhang et al. 2019; Zhu et al. 2021; Su et al. 2024, 2025; Yang et al. 2025) focus mainly on entity alignment (EA) rather than KG completion. In addition, other studies (Zhu et al. 2020; Singh et al. 2021; Huang et al. 2022; He and Yang 2024) explore ways to improve a single KGC with the other KG triples. They encode KG structures with relational GNNs, and conduct multi-task learning with KGC and EA. LSGMA (Tang et al. 2023) typically learns multiple embeddings of each entity from their KG structures located, and fuse them with attention, achieving competitive results. However, few studies address redundancy in the learned representations, which can conceal critical task-relevant information and impede further improvements when scaling to numerous KGs.

Conclusion

This paper addresses MKGC that improves each KG completion by using triples from multiple KGs. Existing studies typically learn KG representations relying solely on alignments or fusion modules, which can be affected by redundant information within KGs. This issue can conceal task-relevant information, impeding further improvements with numerous KGs. To this end, we propose IMKGC, an information-theoretic framework that imposes constraints to learn minimal sufficient representations through a variational entity encoder and a relation reasoning decoder. Experiments on 14 KGs in three benchmarks indicate significant improvements, especially in redundant scenarios. Our future work will further explore LLMs to advance MKGC.

Acknowledgments

The authors thank the reviewers for their helpful feedback. This work was supported by the National Natural Science Foundation of China (No. 62406319).

References

Alemi, A.; Poole, B.; Fischer, I.; Dillon, J.; Saurous, R. A.; and Murphy, K. 2018. Fixing a broken ELBO. In *Proceedings of ICML*, 159–168. PMLR.

Alemi, A. A.; Fischer, I.; Dillon, J. V.; and Murphy, K. 2017. Deep Variational Information Bottleneck. In *Proceedings of ICLR*. OpenReview.net.

Bengio, Y.; Léonard, N.; and Courville, A. C. 2013. Estimating or Propagating Gradients Through Stochastic Neurons for Conditional Computation. *CoRR*, abs/1308.3432.

Bordes, A.; Usunier, N.; Garcia-Durán, A.; Weston, J.; and Yakhnenko, O. 2013. Translating Embeddings for Modeling Multi-Relational Data. In *Proceedings of NeurIPS*, 2787–2795.

Chen, M.; Tian, Y.; Yang, M.; and Zaniolo, C. 2017. Multilingual Knowledge Graph Embeddings for Cross-lingual Knowledge Alignment. In *Proceedings of IJCAI*.

Chen, X.; Chen, M.; Fan, C.; Uppunda, A.; Sun, Y.; and Zaniolo, C. 2020. Multilingual Knowledge Graph Completion via Ensemble Knowledge Transfer. In *Findings of EMNLP*, 3227–3238.

Dettmers, T.; Pasquale, M.; Pontus, S.; and Riedel, S. 2018. Convolutional 2D Knowledge Graph Embeddings. In *Proceedings of AAAI*, 1811–1818.

Guo, L.; Bo, Z.; Chen, Z.; Zhang, Y.; Chen, J.; Lan, Y.; Sun, M.; Zhang, Z.; Luo, Y.; Li, Q.; Zhang, Q.; Zhang, W.; and Chen, H. 2024. MKGL: Mastery of a Three-Word Language. In *Proceedings of NeurIPS*.

He, J.; and Yang, H. 2024. Multilingual Knowledge Graph Completion based on Global-Local Structure Encoding. In *Proceedings of ISCTIS*, 647–650.

He, K.; Fan, H.; Wu, Y.; Xie, S.; and Girshick, R. 2020. Momentum contrast for unsupervised visual representation learning. In *Proceedings of CVPR*, 9729–9738.

Huang, Z.; Li, Z.; Jiang, H.; Cao, T.; Lu, H.; Yin, B.; Subbian, K.; Sun, Y.; and Wang, W. 2022. Multilingual Knowledge Graph Completion with Self-Supervised Adaptive Graph Alignment. In *Proceedings of ACL*, 474–485.

Kingma, D. P.; and Welling, M. 2014. Auto-Encoding Variational Bayes. In *Proceedings of ICLR*.

Lee, D.; Kim, C.; Kim, S.; Cho, M.; and Han, W. 2022. Autoregressive Image Generation using Residual Quantization. In *Proceedings of CVPR*, 11513–11522.

Li, D.; Tan, Z.; Chen, T.; and Liu, H. 2024. Contextualization Distillation from Large Language Model for Knowledge Graph Completion. 458–477.

Linders, J.; and Tomczak, J. M. 2025. Knowledge Graph-extended Retrieval Augmented Generation for Question Answering. *CoRR*, abs/2504.08893.

Liu, J.; Mao, Q.; Jiang, W.; and Li, J. 2024a. KNOWFORMER: revisiting transformers for knowledge graph reasoning. In *Proceedings of ICML*, ICML'24.

Liu, Y.; Tian, X.; Sun, Z.; and Hu, W. 2024b. Finetuning Generative Large Language Models with Discrimination Instructions for Knowledge Graph Completion. In *Proceedings of ISWC*, 199–217.

Pan, S.; Luo, L.; Wang, Y.; Chen, C.; Wang, J.; and Wu, X. 2024. Unifying Large Language Models and Knowledge Graphs: A Roadmap. *IEEE Transactions on Knowledge and Data Engineering*, 36(7): 3580–3599.

Sachan, M. 2020. Knowledge Graph Embedding Compression. In *Proceedings of ACL*, 2681–2691.

Schlichtkrull, M.; Kipf, T. N.; Bloem, P.; Van Den Berg, R.; Titov, I.; and Welling, M. 2018. Modeling relational data with graph convolutional networks. In *European semantic web conference*, 593–607. Springer.

Shang, C.; Tang, Y.; Huang, J.; Bi, J.; He, X.; and Zhou, B. 2019. End-to-end structure-aware convolutional networks for knowledge base completion. In *Proceedings of AAAI*, volume 33, 3060–3067.

Sheng, J.; Guo, S.; Chen, Z.; Yue, J.; Wang, L.; Liu, T.; and Xu, H. 2020. Adaptive Attentional Network for Few-Shot Knowledge Graph Completion. In *Proceedings of EMNLP*, 1681–1691. Association for Computational Linguistics.

Shwartz-Ziv, R.; and Tishby, N. 2017. Opening the Black Box of Deep Neural Networks via Information. *CoRR*, abs/1703.00810.

Singh, H.; Jain, P.; Chakrabarti, S.; et al. 2021. Multilingual Knowledge Graph Completion with Joint Relation and Entity Alignment. *arXiv preprint arXiv:2104.08804*.

Su, T.; Sheng, J.; Ma, D.; Li, X.; Yue, J.; Song, M.; Tang, Y.; and Liu, T. 2025. Mitigating Modality Bias in Multi-modal Entity Alignment from a Causal Perspective. In *Proceedings of SIGIR*, 1186–1196. ACM.

Su, T.; Sheng, J.; Wang, S.; Zhang, X.; Xu, H.; and Liu, T. 2024. IBMEA: Exploring Variational Information Bottleneck for Multi-modal Entity Alignment. In *Proceedings of ACM MM*, 4436–4445. ACM.

Sun, Z.; Deng, Z.-H.; Nie, J.-Y.; and Tang, J. 2019. RotatE: Knowledge Graph Embedding by Relational Rotation in Complex Space. In *Proceedings of ICLR*.

Sun, Z.; Hu, W.; Zhang, Q.; and Qu, Y. 2018. Bootstrapping Entity Alignment with Knowledge Graph Embedding. In *Proceedings of IJCAI*, 4396–4402.

Sun, Z.; Huang, J.; Xu, X.; Chen, Q.; Ren, W.; and Hu, W. 2023. What Makes Entities Similar? A Similarity Flooding Perspective for Multi-sourced Knowledge Graph Embeddings. In *Proceedings of IMCL*, 32875–32885.

Sun, Z.; Zhang, Q.; Hu, W.; Wang, C.; Chen, M.; Akrami, F.; and Li, C. 2020. A Benchmarking Study of Embedding-based Entity Alignment for Knowledge Graphs. *Proceedings of VLDB*, 13(11): 2326–2340.

Tang, R.; Zhao, Y.; Zong, C.; and Zhou, Y. 2023. Multilingual Knowledge Graph Completion with Language-Sensitive Multi-Graph Attention. In *Proceedings of ACL*, 10508–10519.

Tishby, N.; Pereira, F. C.; and Bialek, W. 1999. The information bottleneck method. *Annual Allerton Conf. on Communication, Control, and Computing*.

Tishby, N.; and Zaslavsky, N. 2015. Deep Learning and The Information Bottleneck Principle. In *IEEE Information Theory Workshop (ITW)*.

Trouillon, T.; Welbl, J.; Riedel, S.; Gaussier, É.; and Bouchard, G. 2016. Complex embeddings for simple link prediction. In *Proceedings of ICML*, 2071–2080. PMLR.

van den Oord, A.; Vinyals, O.; and Kavukcuoglu, K. 2017. Neural Discrete Representation Learning. In *Proceedings of NeurIPS*, 6306–6315.

Wang, Q.; Mao, Z.; Wang, B.; and Guo, L. 2017. Knowledge Graph Embedding: A Survey of Approaches and Applications. *IEEE Transactions on Knowledge and Data Engineering*, 29(12): 2724–2743.

Wang, X.; Gao, T.; Zhu, Z.; Zhang, Z.; Liu, Z.; Li, J.; and Tang, J. 2021. KEPLER: A Unified Model for Knowledge Embedding and Pre-trained Language Representation. *TACL*, 9: 176–194.

Wei, Y.; Huang, Q.; Zhang, Y.; and Kwok, J. T. 2023. KICGPT: Large Language Model with Knowledge in Context for Knowledge Graph Completion. In *Findings of EMNLP*, 8667–8683.

Xie, R.; Liu, Z.; Jia, J.; Luan, H.; and Sun, M. 2016. Representation Learning of Knowledge Graphs with Entity Descriptions. In Schuurmans, D.; and Wellman, M. P., eds., *Proceedings of AAAI*, 2659–2665.

Yang, B.; Yih, W.-t.; He, X.; Gao, J.; and Deng, L. 2015. Embedding entities and relations for learning and inference in knowledge bases. In *Proceedings of ICLR*.

Yang, Y.; Luo, Z.; Wang, Z.; Lu, W.; Lu, Y.; Guan, Z.; Zhao, W.; and Lv, Y. 2025. A Translation-Based Heterogeneous Graph Neural Network for Multiple Knowledge Graphs Alignment. In *Proceedings of ICDE*, 2215–2226.

Yao, L.; Mao, C.; and Luo, Y. 2020. KG-BERT: BERT for Knowledge Graph Completion. *Proceedings of AAAI*.

Yao, L.; Peng, J.; Mao, C.; and Luo, Y. 2025. Exploring Large Language Models for Knowledge Graph Completion. In *Proceedings of IEEE ICASSP*, 1–5. IEEE.

Zeghidour, N.; Luebs, A.; Omran, A.; Skoglund, J.; and Tagliasacchi, M. 2022. SoundStream: An End-to-End Neural Audio Codec. *TASLP*, 30: 495–507.

Zhang, Q.; Sun, Z.; Hu, W.; Chen, M.; Guo, L.; and Qu, Y. 2019. Multi-view Knowledge Graph Embedding for Entity Alignment. In Kraus, S., ed., *Proceedings of IJCAI*, 5429–5435.

Zhang, Y.; Chen, Z.; Guo, L.; Xu, Y.; Zhang, W.; and Chen, H. 2024. Making Large Language Models Perform Better in Knowledge Graph Completion. In *Proceedings of ACM MM*, 233–242.

Zhu, Q.; Wei, H.; Sisman, B.; Zheng, D.; Faloutsos, C.; Dong, X. L.; and Han, J. 2020. Collective Multi-Type Entity Alignment Between Knowledge Graphs. In *Proceedings of Web Conference*.

Zhu, Y.; Liu, H.; Wu, Z.; and Du, Y. 2021. Relation-Aware Neighborhood Matching Model for Entity Alignment. In *Proceedings of AAAI*, 4749–4756.

Zhu, Y.; Wang, X.; Chen, J.; Qiao, S.; Ou, Y.; Yao, Y.; Deng, S.; Chen, H.; and Zhang, N. 2024. LLMs for knowledge graph construction and reasoning: recent capabilities and future opportunities. *Proceedings of Web Conference*, 58.