

# Multimodal Knowledge Graph Error Detection with Disentanglement VAE and Multi-Grained Triplet Confidence

Anonymous Author(s)  
Submission Id: 174\*

## Abstract

Multimodal knowledge graphs inevitably contain numerous errors due to the lack of human supervision in their automated construction and updating processes. These errors can significantly degrade the performance of downstream applications that rely on them. Existing researches on knowledge graph error detection primarily focus on leveraging graph structural and textual information to identify triplet errors in unimodal knowledge graphs. However, unlike unimodal knowledge graphs, multimodal knowledge graphs also suffer from mismatches between images and their corresponding entities, referred to as modality errors. These modality errors not only hinder the performance of downstream applications but also impede our effective utilization of the abundant complementary information provided by the visual modality for detecting triplet errors. To this end, we introduce a novel task of multimodal knowledge graph error detection (MKGED) in this paper, aiming at simultaneously identifying both modality errors and triplet errors. Given the lack of datasets for evaluating this task, we first establish two comprehensive MKGED datasets. Furthermore, we propose a novel framework, KGDMC, to address the MKGED task. Within KGDMC, we devise a disentanglement modality reconstruction (DMR) module for modality error detection. This module disentangles each original modality representation into two disjoint components: modality-specific representations and modality-invariant representations, leveraging the cross-modality reconstruction process to detect mismatched visual modalities. Additionally, for the triplet error detection, we propose a multi-grained triplet confidence (MTC) module, incorporating local triplet confidence, global structure confidence, and global path confidence, to collaboratively detect mismatched triplets. Extensive experiments on our constructed two datasets demonstrate the superiority of our proposed framework.

## CCS Concepts

• Computing methodologies → Knowledge representation and reasoning; Probabilistic reasoning; • Information systems → Data cleaning.

## Keywords

knowledge graph, multimodal information, error detection

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

Conference acronym 'XX, June 03–05, 2018, Woodstock, NY

© 2018 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 978-1-4503-XXXX-X/18/06

<https://doi.org/XXXXXXX.XXXXXXX>

## ACM Reference Format:

Anonymous Author(s). 2018. Multimodal Knowledge Graph Error Detection with Disentanglement VAE and Multi-Grained Triplet Confidence. In *Proceedings of Make sure to enter the correct conference title from your rights confirmation email (Conference acronym 'XX)*. ACM, New York, NY, USA, 10 pages. <https://doi.org/XXXXXXX.XXXXXXX>

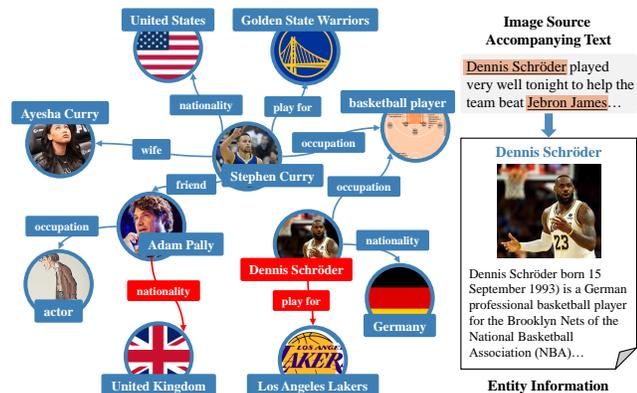
## 1 Introduction

Knowledge graphs (KGs), which provide relational information between entities in the form of triplets (head entity, relation, tail entity), play a key role in various knowledge-driven downstream AI systems, such as recommendation systems [20, 48], question answering [21, 33], and information retrieval [36, 50]. Most existing KGs are constructed by extracting factual information from semi-structured or unstructured web sources using heuristics algorithms, such as Freebase [4], DBpedia [2], NELL [8] and Wikidata [37].

Given the exponential growth and incessant flux of web information, the automated construction and updating of large KGs often lack human supervision. This inevitably results in the introduction of a substantial quantity of noisy triplets. For instance, the widely used KG NELL exhibits a precision rate of 74% [7], indicating that approximately 0.6 million of its triplets are inaccurate. These erroneous triplets result in substantial performance deterioration within downstream tasks that rely on them. Thus, there is an urgent need to delve into effective KG error detection, and numerous approaches [3, 13, 23, 29, 31, 45, 51, 52] have been proposed for this task. While these methods have achieved significant progress, they are designed solely for graph structural and textual information.

However, with the rapid evolution of social media platforms, images coupled with textual content have become the most prevalent form of web-based information. This promotes the rapid development of multimodal knowledge graphs but also brings new challenges to the task of KG error detection. As shown in Figure 1, multimodal KGs, akin to their traditional counterparts, harbor a substantial quantity of triplet errors. Illustratively, errors in the information extraction process may lead to erroneous triplets such as (*Adam Pally, nationality, United Kingdom*), and ineffective updates to the KG may lead to another incorrect triplet, (*Dennis Schröder, play for, Los Angeles Lakers*). In addition to triplet errors, multimodal KGs frequently encounter modality errors during the process of their automatic construction. The right part of Figure 1 is an example. Accompanying images often align with the entire text content, rather than being explicitly tied to a specific entity. This results in a modality error where the image of *Lebron James* is incorrectly associated with *Dennis Schröder*.

In this paper, we are the first to focus on these problems and introduce a new task of multimodal knowledge graph error detection (MKGED), accompanied by the construction of two comprehensive datasets for its evaluation. MKGED mainly handles unique challenges as follows: (1) **Lacking labeled data**. Automatically



**Figure 1: Examples of some multimodal KG errors. The left part is a subgraph of multimodal KGs, errors in it are denoted in red. The right part is an example of extracting images from web sources to construct multimodal KGs.**

constructed KGs inherently tend to be large and encompass a wide variety of error types. Consequently, the process of manually labeling data is time-consuming and labor-intensive, and difficult to cover all noise patterns. Thus, in real-world practice, MKGED often necessitates the proactive detection of errors within multimodal KGs, without relying on any labeled data during the training phase. (2) **Modality errors.** Given the advanced maturity of information extraction technology, prevalent modality errors encountered in practical multimodal KGs arise from mismatched images that are highly similar or relevant to their associated entities, posing challenges in detecting such modality errors. (3) **Triplet errors.** The distinctive image information within multimodal KGs provides abundant additional complementary knowledge for triplet error detection. However, erroneous image information not only fails to contribute to this process but also introduces additional noise (e.g. *Dennis Schröder* of (*Dennis Schröder*, *play for*, *Los Angeles Lakers*)), ultimately diminishing the overall performance.

To address these challenges, we propose a novel multimodal Knowledge Graph error detection framework with Disentanglement VAE and Multi-grained triplet Confidence (KGDMC for short), which utilizes an unsupervised learning way to solve the challenge of lacking labeled data. To detect modality errors, we devise a disentanglement modality reconstruction (DMR) module. This module utilizes cross-modality reconstruction to comprehend noise patterns and examine semantic coherence between modalities. It provides modality confidence for modality error detection. Considering the challenge of high similarity or relevance between images and their erroneously associated entities in modality errors, DMR adopts a disentanglement approach, which disentangles each modality input data into separate modality-specific and modality-invariant feature spaces to obtain more fine-grained semantic information. It then explicitly utilizes accurate textual modality-invariant features to guide the the cross-modality reconstruction process of suspicious visual modality-specific information, ensuring adequate correlation between modalities. To address triplet errors, we propose a multi-grained triplet confidence (MTC) module. MTC tackles the challenge posed by image noise in triplet error detection through the utilization of weighted embedding based on modality confidence. Additionally, MTC employs local triplet confidence, global

structure confidence, and global path confidence to assess the internal factual self-consistency of triplets, the consistency of aggregated neighbor information, and the consistency of information propagation, respectively. The integration of these three distinct granularity levels of confidence significantly enhances our capability to comprehensively detect triplet errors.

The main contributions of this paper are summarized as follows:

- To the best of our knowledge, we are the first to introduce the novel multimodal knowledge graph error detection task, which aims to simultaneously detect both modality errors and triplet errors in multimodal knowledge graphs.
- For modality error detection, we devise a DMR module that leverages disentanglement VAE for cross-modality reconstruction. This approach effectively captures fine-grained semantic consistency across modalities.
- For triplet error detection, we propose a MTC module, which integrates the internal factual self-consistency of triplets, the consistency of aggregated neighbor information, and the consistency of information propagation to collaboratively estimate comprehensive triplet confidence.
- We compare our KGDMC with state-of-the-art baselines on two real-world multimodal knowledge graph datasets, which we construct to incorporate multiple types of errors. Extensive experimental results demonstrate the effectiveness of our proposed framework.

## 2 Related Work

### 2.1 Knowledge Graph Error Detection

Traditional knowledge graph error detection mainly focuses on graph structure and textual information, with this area of research witnessing substantial progress and thorough exploration in recent years. Knowledge graph error detection methods can be roughly divided into two groups: rule-based methods and embedding-based methods. Rule-based methods analyze the commonalities present within triplets through the mining of association rules [1]. Triplets that fail to adhere to these rules are subsequently identified and considered as errors. AMIE [17] formulates a suite of golden rules via path exploration and rule pruning processes. PaTyBRED [31] utilizes heuristic approaches to explore the paths search space and select relevant type and path information for each relation. GRR [10] explores the graph-repairing rules of implication, consistency, and termination. KGIST [3] inductively summarizes knowledge graph to derive a set of soft rules, which describe normal conditions within a KG. However, given the inherent diversity in rules across different KGs and the fact that these methods primarily target specific error categories, their applicability lacks generalizability.

Consequently, most of recent works focus on embedding-based methods, which predict the plausibility between entity and relation representations in low-dimensional vector space. CKRL [45] is the first to introduce triplet confidence, which incorporates both local and global path information, to guide models to pay more attention to convincing triplets. KGTm [23] integrates entity correlation, relation invariance, and path reachability to predict triplet plausibility. KGClean [18] utilizes active learning to train a classification model to identify and repair erroneous triplets based on a small number of labeled data. CAGED [51] employs contrastive learning to capture

consistency between different triplet hyper-views of a knowledge graph. HEAR [52] leverages the hierarchical path structure within knowledge graphs to detect noise. KAEL [13] assembles a set of advanced KG error detectors through ensemble learning and trains them by active learning to label a small number of samples. CCA [29] utilizes prompt learning to reconstruct textual and structural information to better detect semantically similar noise.

Although these methods have achieved significant progress, they are designed solely for graph structural and textual information, thereby rendering them ineffective in leveraging abundant complementary information of visual modality within multimodal KGs. Additionally, these methods face limitations in addressing modality errors, which frequently occur in multimodal KGs. Therefore, we propose the MKGED task, which aims not only to detect modality errors but also to efficiently utilize complementary visual information to detect triplet errors in the presence of modality noise.

## 2.2 Multimodal Knowledge Graph Embedding

Multimodal knowledge graph embedding (MKGE) typically harnesses triplet structure and multimodal information to embed entities and relations. Subsequently, it defines score functions to measure the plausibility of triplets. IKRL [46] is the first MKGE model, which learns structural and visual information separately based on TransE [5]. Based on IKRL, Mousselly et al. [35] and TransAE [44] integrate visual and structural representations into unified embeddings. RSME [43] utilizes a forget gate to ignore irrelevant image information to obtain better embeddings. Subsequently, MMRKRL [30] and OTKGE [6] exhibit enhanced capabilities in aligning multimodal embeddings and integrating multimodal knowledge, facilitated respectively by component alignment and optimal transport. CMGNN [16] employs contrastive learning to achieve multi-modal and high-order structural modeling within graph neural networks. VISTA [27] integrates visual and textual representations of entities and relations by leveraging transformers for entity encoding, relation encoding, and triplet decoding. These MKGE methods have achieved remarkable success in a diverse range of multimodal knowledge graph related tasks, including but not limited to multimodal knowledge graph completion and triplet classification. However, they assume that all triplet facts within multimodal KGs are completely correct, thereby constraining the efficacy of multimodal KG error detection.

## 3 Problem Statement

A multimodal knowledge graph can be depicted or formulated as  $\mathcal{G} = (\mathcal{E}, \mathcal{R}, \mathcal{T}, \mathcal{I}, \mathcal{D})$ , where  $\mathcal{E} = \{e_1, e_2, \dots, e_m\}$ ,  $\mathcal{R} = \{r_1, r_2, \dots, r_n\}$  respectively denote the sets of entities and relations.  $\mathcal{I}$  and  $\mathcal{D}$  respectively contains visual images and textual descriptions for all entities.  $\mathcal{T} \subseteq \mathcal{E} \times \mathcal{R} \times \mathcal{E}$  is the set of triplets. Each triplet comprises a head entity  $h$ , a relation  $r$ , and a tail entity  $t$ , articulated in the form of  $(h, r, t)$ . Multimodal knowledge graph errors involve modality errors and triplet errors. For modality errors, given an entity  $e_i$  accompanied by its description in  $\mathcal{D}$  and image in  $\mathcal{I}$ , if the image fails to associate with the entity, then the entity  $e_i$  is a modality error. Regarding triplet errors within multimodal KGs, for a given triplet  $(h, r, t)$ , we define the error as the mismatch that arises between its head entity  $h$ , tail entity  $t$ , and corresponding relation  $r$ . Two

distinct types of mismatch can occur: entity mismatch, where the head entity  $h$  and tail entity  $t$  are irrelevant, and relation mismatch, where relevant entities are erroneously connected through inappropriate relations. For multimodal knowledge graph error detection, given a noisy multimodal KG  $\mathcal{G}$ , our goal is to rank all entities and triplets according to their respective suspicious scores. These scores, spanning from 0 to 1, serve as indicators of the likelihood of an error existing within the respective entity or triplet. Entities and triplets that exhibit higher scores are designated as errors.

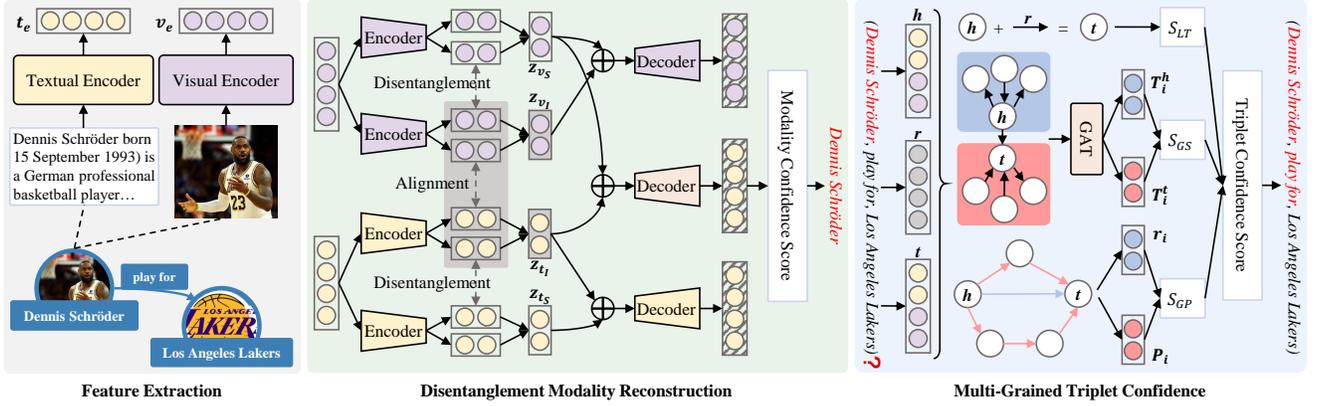
## 4 Dataset Construction

In recent years, multiple real-world multimodal knowledge graph datasets have been widely used. However, these datasets lack explicit labels for errors or noises. Furthermore, with the maturity of multimodal information extraction technology, the errors encountered in real-world multimodal KGs frequently exhibit a high degree of semantic similarity with their accurate counterparts. Consequently, to effectively evaluate the performance of MKGED, we construct datasets by adding various types of noise to simulate the real-world multimodal KGs constructed automatically without human supervision. There are three types for modality errors:

- **Random noise**, originating from a uniform distribution that is independent of the data and inherently unpredictable, is generated by randomly substituting the images.
- **Intra-modality similar noise** simulates real-world modality errors arising from high similarities or correlations among images. To generate them, we first utilize ViT [14] to obtain visual features of all entity images. Subsequently, for a given entity, we compute the distance between its corresponding visual feature and those of other entities. Based on the distance, we select the image exhibiting the highest degree of similarity for replacement.
- **Inter-modality similar noise** simulates modality errors stemming from high semantic similarities between textual content and erroneous images. Given the remarkable modality alignment capabilities of vision-and-language pre-trained models, we employ CLIP [32] to extract visual and textual features of images and entity descriptions. Subsequently, for a given entity, we compute the distance between its textual feature and the visual features of other entities to replace with highly relevant images.

For triplet errors, following [29], we also contain three types of noise to simulate real-world multimodal KGs:

- **Random noise** akin to the random noise of modality errors. Given a correct triplet  $(h, r, t)$ , we randomly substitute either one of the entities or the relation within it to generate noise.
- **Semantic similar noise** mimics real-world triplet errors stemming from high semantic similarities between entities. Given a triplet  $(h, r, t)$ , we employ BERT [12] to extract textual features of the tail entity  $t$  and other tail entities associated with the relation  $r$ . Then, we calculate the distance between them to replace  $t$  by tail entities that exhibit a high degree of semantic similarity.
- **Adversarial noise** is adversarially generated by leveraging knowledge graph embedding models, which are frequently



**Figure 2: Overview of our KGDMC framework, which contains three modules: feature extraction, disentanglement modality reconstruction (DMR) for modality error detection, and multi-grained triplet confidence (MTC) for triplet error detection.**

employed to assess the plausibility of triplets. We first generate a substantial quantity of error triplets through random substitution. Then, we train a TransE [5] model to select the most plausible triplets from this set as noise.

## 5 Methodology

Figure 2 illustrates the overall architecture of our proposed KGDMC framework. Initially, we extract textual and visual features for entities in triplets. Subsequently, we utilize disentanglement variational autoencoder (VAE) to decompose each modality into distinct modality-specific and modality-invariant subspaces through distribution alignment and disentanglement constraints. To capture fine-grained semantic consistency between modalities, we explicitly integrate accurate textual modality-invariant information and potentially suspicious visual modality-specific information, guiding the model to cross-reconstruct accurate textual features. This consistency provides the modality confidence scores for detecting modality errors. Based on the modality confidence, we derive entity embeddings by weighting textual and visual features. Finally, we calculate triplet confidence scores from different views with multiple granulations to assess three critical aspects: the internal factual self-consistency of triplets, the consistency of aggregated neighbor information, and the consistency of information propagation. By integrating these scores, we effectively detect triplet errors.

### 5.1 Feature Extraction

To extract meaningful textual features, following [29], we employ pre-trained BERT [12] as our textual encoder. Specifically, for each entity  $e_i$ , the input sequence is constructed in the following format: [CLS] *title* [SEP] *des* [SEP], where *title* and *des* refer to the word-piece tokens of the entity title and description, respectively. Subsequently, these inputs are fed into BERT to extract textual features  $t_{e_i}$ , which are the vectors in the last hidden layer corresponding to the position of [CLS] tokens.

To capture expressive visual features, we employ pre-trained Vision Transformer (ViT) [14] as our visual encoder. Given an image associated with the entity  $e_i$ , we initially resize it to  $C \times H \times W$  pixels, where  $C$  is the number of channels, and  $H \times W$  refers to the image resolution. Then we reshape it into  $H \times W/P^2$  flattened two-dimensional patches, where  $P$  represents the patch size. Similar

to the textual feature extraction, we feed them into ViT to obtain the hidden state of the [CLS] as our visual features  $v_{e_i}$ .

### 5.2 Disentanglement Modality Reconstruction

In multimodal KGs, heterogeneous modalities of entities typically encompass diverse aspects of information (e.g. visual line and color, textual detailed description) alongside shared aspects of information (e.g. the characteristics and categories of entities). Disentangling modalities can empower the model to delve into fine-grained semantic nuances, thereby facilitating the acquisition of semantic consistency across modalities. Therefore, we utilize disentanglement VAE [22], which is the extension of VAE [26], for disentanglement.

**5.2.1 Disentanglement VAE.** A standard VAE is typically dissected into two components: an encoder, which extracts a low-dimensional latent variable  $z$  from the input data  $x$ , and a decoder, which subsequently reconstructs an output  $\tilde{x}$  that approximates the original data  $x$  based on the latent variable  $z$ . In VAE, variational inference is employed to find the true conditional probability distribution  $p_\theta(z|x)$  over the latent variable  $z$ . Given the intractability of the true conditional distribution, an approximation way is to find its closest proxy posterior,  $q_\phi(z|x)$ , by minimizing their distance using a variational lower bound. Thus, the objective function of VAE is the variational lower bound on the marginal likelihood of  $x$ :

$$\mathcal{L}_{VAE}(x) = \mathbb{E}_{q_\phi(z|x)} [\log p_\theta(x|z)] - D_{KL}(q_\phi(z|x)||p(z)) \quad (1)$$

where the initial term represents a reconstruction loss, and the last term is the Kullback-Leibler (KL) divergence between conditional probability distribution  $q_\phi(z|x)$  and the prior distribution  $p(z)$  of  $z$ . The prior distribution is predefined as a multivariate Gaussian distribution. The encoder predicts mean  $\mu$  and variance  $\Sigma$  of the distribution  $q_\phi(z|x) = \mathcal{N}(\mu, \Sigma)$  and then the latent variation  $z$  is generated by reparameterization strategy [26].

Disentanglement VAE extends VAE by a pair of encoders responsible for modeling modality-specific and modality-invariant information as follows:

$$\mathcal{L}_{DVAE}(x) = \mathbb{E}_{q_{\phi_S, \phi_I}(z_S, z_I|x)} [\log p_\theta(x|z_S, z_I)] - D_{KL}(q_{\phi_S}(z_S|x)||p(z_S)) - D_{KL}(q_{\phi_I}(z_I|x)||p(z_I)) \quad (2)$$

where  $\phi_S$  and  $\phi_I$  represent the respective parameters of the encoder pairs:  $E_S$  and  $E_I$ . The latent variations  $z_S$  and  $z_I$  denote the

modality-specific and modality-invariant representations, respectively, extracted from the input data  $\mathbf{x}$ . Furthermore,  $q_{\phi_S, \phi_I}(z_S, z_I | \mathbf{x})$  characterizes the joint distribution of  $z_S$  and  $z_I$  conditioned on  $\mathbf{x}$ . We obtain textual and visual disentanglement VAE by replacing  $\mathbf{x}$  with  $\mathbf{t}_e$  and  $\mathbf{v}_e$  respectively. Finally, the objective function of our multimodal disentanglement VAE is formulated as follows:

$$\mathcal{L}_{MDVAE} = \mathcal{L}_{DVAE}(\mathbf{t}_e) + \mathcal{L}_{DVAE}(\mathbf{v}_e) \quad (3)$$

**5.2.2 Distribution Alignment and Disentanglement Constraints.** A pivotal character of our disentanglement VAE lies in the construction of modality-invariant and modality-specific subspaces. The modality-invariant subspace focuses on information describing the shared aspects across modalities. To achieve this, we add a distribution alignment loss constraint. Following [34], we utilize the Wasserstein distance between the latent multivariate Gaussian distributions. Given two distributions  $a$  and  $b$ , the 2-Wasserstein distance [19] is formulated as:

$$W_{ab} = (\|\mu_a - \mu_b\|_2^2 + \|\Sigma_a^{\frac{1}{2}} - \Sigma_b^{\frac{1}{2}}\|_{\text{Frobenius}}^2)^{\frac{1}{2}} \quad (4)$$

We construct the modality-invariant subspace by minimizing the distance between textual and visual distributions within this subspace:  $\mathcal{L}_{DA} = W_{t|v}$ . Furthermore, to construct modality-specific subspaces, we design two constraints for the intra-modality disentanglement and inter-modality disentanglement, respectively. These constraints aim to avoid modality-specific representations carrying the information of shared aspects. For intra-modality, our objective is to ensure that disentangled modality-invariant and modality-specific representations are distinct in the latent space. We achieve this by maximizing the distance between disentanglement distributions of each modality:  $\mathcal{L}_{IntraDD} = -(W_{t|t_S} + W_{v|v_S})$ . For inter-modality, we strive to make the modality-specific representations of different modalities as disparate as possible. To this end, we maximize the distance between textual and visual specific distributions:  $\mathcal{L}_{InterDD} = -W_{S|v_S}$ . Finally, the distribution constraint loss of our disentanglement VAE is constructed as follows:

$$\mathcal{L}_{DC} = \mathcal{L}_{DA} + \alpha \mathcal{L}_{IntraDD} + \beta \mathcal{L}_{InterDD} \quad (5)$$

where  $\alpha$  and  $\beta$  are hyperparameters to control the strength of intra-modality disentanglement and inter-modality disentanglement.

**5.2.3 Cross-Modality Reconstruction.** Intuitively, features of the same entity without any modality errors are semantically consistent even though they come from different modalities. Therefore, after obtaining fine-grained semantic information of each modality by disentanglement VAE, we cross-reconstruct modalities to capture this consistency for the detection of modality errors. Given the accuracy of textual modality, we utilize textual modality-invariant representation as guidance to reconstruct potentially suspicious visual modality-specific representation into accurate textual features. In contrast to solely relying on visual representations, the incorporation of textual modality-invariant representations significantly enhances the efficiency of cross-modality reconstruction. To this end, we concatenate  $z_{t_I}$  and  $z_{v_S}$  to feed into cross-modality decoder  $D^{cross}$  to reconstruct  $\mathbf{t}_e$ . The modality confidence score is:

$$S_M = \|\mathbf{t}_e - D^{cross}(z_{t_I}, z_{v_S})\|_2^2 \quad (6)$$

To mitigate the negative effects of erroneous entities and pay more attention to those more convincing entities, we utilize the modality

confidence score to generate adaptive pseudo-labels to guide the model training:  $C_M(e_i) = 1 - (S_M(e_i) - \min(S_M)) / (\max(S_M) - \min(S_M))$ . Our cross-modality reconstruction loss is formulated as:

$$\mathcal{L}_{CR} = \sum_{e_i \in \mathcal{E}} S_M(e_i) \cdot C_M(e_i) \quad (7)$$

Finally, the objective function of our DMR module is to jointly optimize the above loss functions as follows:

$$\mathcal{L}_{DMR} = \mathcal{L}_{MDVAE} + \mathcal{L}_{DC} + \mathcal{L}_{CR} \quad (8)$$

### 5.3 Multi-Grained Triplet Confidence

Considering the negative effect of modality errors, we weight textual and visual features to obtain each entity embedding as:

$$\mathbf{w}_i = (\max(S_M) - S_M(e_i)) / \max(S_M), \quad \mathbf{e}_i = \mathbf{t}_e + \mathbf{w}_i \mathbf{v}_e \quad (9)$$

Given a triplet  $T_i$ , we obtain its embedding  $(\mathbf{h}_i, \mathbf{r}_i, \mathbf{t}_i)$  via Eq. 9 and the textual content of  $r_i$ . To comprehensively detect triplet errors, unlike previous works that only involve partial views, we integrate multi-grained triplet confidence from three different views.

Local triplet confidence is employed to measure the internal factual self-consistency within triplets. To achieve this, we utilize the TransE [5], where relations are interpreted as translations between head entities and tail entities. The greater alignment a triplet exhibits with the translation assumption, the higher the likelihood of its correctness. Thus, our local triplet confidence score is defined as:  $S_{LT}(T_i) = \|\mathbf{h}_i + \mathbf{r}_i - \mathbf{t}_i\|_2$ .

Global structure confidence is employed to measure the consistency of aggregated neighbor information. Following [51], we regard each triplet as a node and obtain its embedding by concatenating internal embeddings of the triplet:  $T_i = [\mathbf{h}_i, \mathbf{r}_i, \mathbf{t}_i]$ . Subsequently, we generate two subgraphs by connecting triplets adjacent to head entities  $\mathcal{T}^h$  and triplets adjacent to tail entities  $\mathcal{T}^t$ , respectively. The triplet set  $\mathcal{T}^h$  represents triplets that share the same head entity, i.e.  $\mathcal{T}^{h_i} = \{T_j | h_j = h_i \vee t_j = h_i\}$ . Similarly, the set  $\mathcal{T}^t$  is defined as  $\mathcal{T}^{t_i} = \{T_j | h_j = t_i \vee t_j = t_i\}$ . According to social community theory [40], correct triplets are those individuals who can successfully integrate into society. Therefore, if a triplet is correct, its embeddings aggregated from different neighbors of the two subgraphs should exhibit a high degree of consistency. To measure this consistency, we utilize the graph attention network (GAT) [41] to respectively weighted aggregate neighbors from the two subgraphs to obtain  $T_i^h$  and  $T_i^t$ . Then, our global structure confidence score can be defined as:  $S_{GS}(T_i) = \|T_i^h - T_i^t\|_2$ .

Global path confidence is employed to measure the consistency of information propagation. Between any two entities, multiple paths may exist, each embodying a distinct process of information propagation between them. If a triplet is correct, its relation will be consistent with the comprehensive information conveyed by the other paths. Initially, we utilize the depth-first search (DFS) algorithm to retrieve the paths between head entities and their corresponding tail entities in all triplets. Given that triplets adhere to the translation assumption based on local triplet confidence, we sum up the corresponding relation embeddings to obtain the embedding of each individual path:  $\mathbf{p}_o = \sum_{j=1}^l \mathbf{r}_j$ , where  $l$  is the number of relations within this path. Due to the existence of erroneous triplets, not all paths are reliable. We compute the average

**Table 1: Overall statistics of datasets. E. means the errors.**

Datasets	Entities	Relations	Triplets	Entity E.	Triplet E.
FB15K-237	14,541	237	326,437	727	16,321
WN18RR	40,943	11	97,897	2,047	4,894

local triplet confidence along the path to determine the path importance:  $AS(p_o) = \sum_{j=1}^l S_{LT}(T_j)$ . The less likely it is that the path contains incorrect triplets, the more important it is. Then, we obtain the unified path embedding based on this importance:  $P_i = \sum_{o=1}^O \frac{AS(p_o)}{\sum_{j=1}^O AS(p_j)} p_o$ , where  $O$  is the number of paths between  $h_i$  and  $t_i$ . Similar to global structure confidence, our global path confidence score is defined as:  $S_{GP}(T_i) = \|\mathbf{r}_i - P_i\|_2$ .

Finally, we integrate these three confidence scores to determine the ultimate suspicious score of triplets:

$$S_T = S_{LT} + \lambda S_{GS} + \delta S_{GP} \quad (10)$$

where  $\lambda$  and  $\delta$  are hyperparameters to balance the scores. Similar to modality error detection, we generate adaptive pseudo-labels  $C_T$  to guide the training process of triplet error detection. The ultimate loss of our MTC module is calculated as follows:

$$\mathcal{L}_{MTC} = \sum_{T_i \in \mathcal{T}} \sum_{T_i^- \in \mathcal{T}^-} \max(0, \gamma + S_T(T_i) - S_T(T_i^-)) \cdot C_T(T_i) \quad (11)$$

where  $\gamma$  is a margin hyperparameter,  $\mathcal{T}$  represents all triplets. And  $\mathcal{T}^-$  denotes negative triplets generated by randomly replacing head or tail entities, as explicit negative triplets are not available.

## 6 Experiments

In this section, we conduct comprehensive experiments to sufficiently verify the effectiveness of our proposed framework KGDMC. We intend to investigate the following research questions (RQs):

- **RQ1.** How does the KGDMC perform compared with various state-of-the-art baseline methods?
- **RQ2.** How does each individual component of KGDMC contribute to enhancing its performance?
- **RQ3.** How do the hyperparameters affect the error detection performance of our KGDMC?
- **RQ4.** How does the robustness of KGDMC and other baseline methods in different noise types?

### 6.1 Experimental Setup

**6.1.1 Datasets.** In experiments, we select two widely-used real-world knowledge graph datasets: FB15K-237 [38] and WN18RR [11]. They are derived from Freebase and WordNet, respectively. Since both of these datasets are unimodal and lack visual information, we first follow [43] to augment images for all entities. Then we generate modality noise and triplet noise for each dataset in a ratio of 1:1:1 for each noise type according to Section 4. Following previous works [13, 29, 52], the ratio of total errors is set to 5%. The overall statistics of our constructed datasets are shown in Table 1.

**6.1.2 Baselines.** We compare our KGDMC framework with various competitive baselines. For modality error detection, we finetune vision-and-language models that are widely used in multiple image-text matching scenarios as our baselines, including CLIP [32], ViLT [24], ALBEF [28], and METER [15]. For triplet error detection, we

utilize the following competitive methods for comparison. The first type is the knowledge graph error detection methods, following [29], including TransE [5], DistMult [47], ComplEx [39], KGTm [23], CAGED [51], KG-BERT [49], StAR [42], CSProm-KG [9], and CCA [29]. Another type is the multimodal knowledge graph embedding methods, including IKRL [46], RSME [43], and VISTA [27].

**6.1.3 Evaluation Metrics.** We employ ranking measures to assess the efficacy of baselines and our proposed KGDMC. Specifically, within a multimodal KG, we rank all entities and triplets based on their modality and triplet confidence scores in descending order, respectively. Entities or triplets that exhibit higher confidence scores are more prone to being erroneous. Following [29, 51, 52], we adopt precision@ $K$  and recall@ $K$  as evaluation metrics. Precision@ $K$  represents the proportion of false entities or triplets identified among the top  $K$  entities or triplets with the highest scores. Recall@ $K$  denotes the percentage of correctly identified false entities or triplets relative to the total number of erroneous entities or triplets.

**6.1.4 Implementation Details.** In our experiments, we initialize the weights of BERT and ViT by leveraging the pre-trained versions: BERT-base-cased and ViT-base-patch16-224-in21k, respectively. The maximum length of the word sequence for textual input is set to 128. Furthermore, all images are resized to a resolution of  $224 \times 224$ , with a patch size  $P$  of 16. We utilize the Adam [25] optimizer with a batch size of 128 for optimizing. The margin parameter  $\gamma$  is empirically set to 1.0. Experiments were conducted on a PC with 256 GB RAM, 4 Intel(R) Xeon(R) Gold 6226R CPUs and an NVIDIA GeForce RTX A6000 GPU with 48 GB memory.

## 6.2 Experimental Results

**6.2.1 Overall Performance (RQ1).** Table 2 and Table 3 present the experimental results of our KGDMC in comparison with baselines. For modality error detection, we find that these vision-and-language baselines exhibit competitive results, especially CLIP. This benefits from their pre-training in image-text matching using large-scale image-text corpus. However, they indiscriminately consider the entities with modality errors and lack fine-grained semantic consistency mining, which limits their performance. Our KGDMC significantly outperforms these baselines and achieves the best metrics on two datasets. Specifically, when  $K$  equals to 5%, it achieves 11.2% and 9.9% absolute improvement on FB15K-237 and WN18RR, respectively. This demonstrates the effectiveness of DMR module design in KGDMC for addressing modality error detection.

For triplet error detection, we have the following observations. First, traditional knowledge graph embedding models, such as TransE, DistMult, and ComplEx, perform unsatisfactorily due to their underlying assumption that all triplets are correct, thereby failing to learn discriminative representations for normal and noisy triplets. Second, text-based methods, including KG-BERT, StAR, CSProm-KG, and CCA, exhibit notably superior performance compared to structure-based knowledge graph error detection models like KGTm and CAGED. This is attributed to the benefits of incorporating additional factual knowledge obtained from large-scale open-domain corpora through pre-trained language models, along with the detailed information in textual descriptions. Third, multimodal knowledge graph embedding models IKRL and VISTA exhibit

**Table 2: Experimental results of modality error detection. The results are averaged 5 runs using different random seeds.**

Metrics	Precision@K										Recall@K									
	FB15K-237					WN18RR					FB15K-237					WN18RR				
	1%	2%	3%	4%	5%	1%	2%	3%	4%	5%	1%	2%	3%	4%	5%	1%	2%	3%	4%	5%
CLIP	47.6	43.1	39.0	37.9	37.6	67.7	54.2	48.5	42.8	37.9	9.5	17.2	23.4	30.3	37.6	13.5	21.6	29.1	34.2	37.9
ViLT	29.7	27.2	26.8	27.0	26.4	48.2	38.6	34.2	31.4	30.6	5.9	10.9	16.1	21.6	26.4	9.6	15.4	20.5	25.1	30.6
ALBEF	42.1	37.6	36.9	35.1	34.8	63.1	50.5	45.7	40.2	36.1	8.4	15.0	22.1	28.1	34.8	12.6	20.2	27.4	32.1	36.1
METER	40.0	36.2	35.1	31.7	30.4	56.5	43.9	39.7	35.2	33.4	8.0	14.4	21.0	25.3	30.4	11.3	17.5	23.8	28.2	33.4
KGDMC	<b>62.0</b>	<b>58.3</b>	<b>55.7</b>	<b>53.2</b>	<b>48.8</b>	<b>80.2</b>	<b>66.5</b>	<b>57.9</b>	<b>51.7</b>	<b>47.8</b>	<b>12.2</b>	<b>23.2</b>	<b>33.4</b>	<b>42.5</b>	<b>48.8</b>	<b>16.0</b>	<b>26.6</b>	<b>34.8</b>	<b>41.3</b>	<b>47.8</b>

**Table 3: Experimental results of triplet error detection. We report the results of some baselines according to CCA [29].**

Metrics	Precision@K										Recall@K									
	FB15K-237					WN18RR					FB15K-237					WN18RR				
	1%	2%	3%	4%	5%	1%	2%	3%	4%	5%	1%	2%	3%	4%	5%	1%	2%	3%	4%	5%
TransE	94.6	77.4	60.6	49.8	42.3	69.0	57.6	50.1	43.7	40.0	18.9	31.0	36.4	39.9	42.3	13.8	23.1	30.0	35.0	40.0
DistMult	76.4	63.0	53.0	46.3	41.0	68.7	63.3	52.6	43.8	37.4	15.3	25.2	31.8	37.1	41.0	13.7	25.3	31.6	35.0	37.4
ComplEx	80.2	63.3	52.1	44.6	39.3	77.4	69.9	55.0	44.9	38.4	16.1	25.4	31.3	35.7	39.3	15.5	27.9	33.0	35.9	38.4
KG-Ttm	85.7	68.7	63.1	46.7	43.7	78.9	64.4	54.1	47.3	41.7	17.1	27.5	37.8	37.4	43.7	15.8	25.7	32.4	37.8	41.7
CAGED	86.3	66.6	60.2	54.3	46.7	75.3	62.0	53.6	47.0	42.1	17.3	26.6	36.2	43.5	46.7	15.0	24.8	32.1	37.6	42.1
KG-BERT	96.6	79.9	66.0	58.4	49.8	97.3	96.8	93.8	82.9	71.0	19.3	31.9	39.6	46.7	49.8	19.5	38.7	56.3	66.3	71.0
StAR	97.0	<b>83.5</b>	68.1	57.1	49.0	97.1	91.8	84.2	73.9	64.7	19.4	<b>33.4</b>	40.9	45.7	49.0	19.4	36.7	50.5	59.1	64.7
CSProm-KG	96.1	79.8	68.9	57.4	50.9	97.7	92.7	86.9	77.3	68.0	19.2	31.9	41.3	45.9	50.9	19.5	37.1	52.1	61.8	68.0
CCA	96.9	81.2	70.7	59.9	53.4	98.6	95.9	92.0	83.4	73.3	19.4	32.5	42.4	47.9	53.4	19.7	38.4	55.2	66.7	73.3
IKRL	94.7	76.3	65.2	56.8	47.4	95.9	86.9	77.3	67.7	60.0	18.9	30.5	39.1	45.4	47.4	19.2	34.8	46.4	54.2	60.0
RSME	96.4	80.3	67.6	58.2	49.8	97.2	92.9	89.8	78.4	69.3	19.3	32.1	40.6	46.6	49.8	19.4	37.1	53.9	62.7	69.3
VISTA	95.3	77.2	66.1	57.0	48.1	96.7	88.4	78.8	69.5	61.4	19.1	30.9	39.7	45.6	48.1	19.3	35.3	47.3	55.6	61.4
KGDMC	<b>97.6</b>	83.2	<b>72.1</b>	<b>61.4</b>	<b>55.1</b>	<b>99.2</b>	<b>97.8</b>	<b>94.6</b>	<b>84.5</b>	<b>74.5</b>	<b>19.5</b>	33.3	<b>43.3</b>	<b>49.1</b>	<b>55.1</b>	<b>19.8</b>	<b>39.1</b>	<b>56.7</b>	<b>67.6</b>	<b>74.5</b>

**Table 4: Ablation study of precision/recall at top-K results for modality error detection.**

Models	FB15K-237			WN18RR		
	K=1%	K=3%	K=5%	K=1%	K=3%	K=5%
KGDMC	<b>62.0/12.2</b>	<b>55.7/33.4</b>	<b>48.8/48.8</b>	<b>80.2/16.0</b>	<b>57.9/34.8</b>	<b>47.8/47.8</b>
w/o $\mathcal{L}_{DA}$	24.1/4.8	23.2/13.9	21.7/21.7	31.1/6.2	24.4/14.7	19.3/19.3
w/o $\mathcal{L}_{IntraDD}$	57.2/11.4	54.6/32.7	48.1/48.1	75.3/15.1	57.2/34.3	47.4/47.4
w/o $\mathcal{L}_{InterDD}$	57.9/11.6	54.9/32.9	48.3/48.3	77.1/15.4	57.3/34.4	47.5/47.5
w/o $\mathbf{z}_I$	54.5/10.9	52.5/31.5	47.1/47.1	73.8/14.8	55.9/33.6	46.7/46.7
w/o $C_M$	56.6/11.3	53.7/32.2	47.7/47.7	71.4/14.3	55.7/33.4	47.1/47.1

inferior performance compared to text-based methods. This indicates that ignoring modality errors and directly utilizing visual information will introduce additional noise, ultimately impairing the effectiveness of triplet error detection. RSME employs a forget gate to disregard irrelevant image information, which mitigates the negative effects of modality errors to a certain degree. However, it still assumes that both entities and triplets are correct, which leads to its inferior performance compared to the state-of-the-art text-based knowledge graph error detection model CCA. Finally, our KGDMC outperforms all baselines on two datasets and achieves new state-of-the-art performance. Specifically, when  $K$  equals to 5%, KGDMC gains 1.7% and 1.2% absolute improvement on FB15K-237 and WN18RR, respectively. This clearly underscores the effectiveness and superiority of our KGDMC in detecting triplet errors.

**6.2.2 Ablation Study (RQ2).** To better understand our proposed KGDMC, we conduct a series of ablation studies, as presented in Table 4 and Table 5. For modality error detection, we observe that KGDMC significantly outperforms KGDMC w/o  $\mathcal{L}_{DA}$ . The distribution alignment constraint serves as a fundamental cornerstone of our DMR module, constructing the modality-invariant subspaces.

**Table 5: Triplet error detection results of ablation study.**

Models	FB15K-237			WN18RR		
	K=1%	K=3%	K=5%	K=1%	K=3%	K=5%
KGDMC	<b>97.6/19.5</b>	<b>72.1/43.3</b>	<b>55.1/55.1</b>	<b>99.2/19.8</b>	<b>94.6/56.7</b>	<b>74.5/74.5</b>
w/o $\mathbf{v}_e$	97.0/19.4	70.9/42.5	54.2/54.2	99.1/19.8	92.8/55.7	73.9/73.9
w/o $w$	96.7/19.3	68.6/41.2	51.6/51.6	98.3/19.6	91.3/54.8	71.8/71.8
w/o $\mathcal{L}_T$	97.2/19.4	71.7/43.0	54.7/54.7	98.7/19.7	94.1/56.4	74.1/74.1
w/o $S_{GS}$	97.3/19.5	71.3/42.8	53.9/53.9	99.1/19.8	93.7/56.2	73.3/73.3
w/o $S_{GP}$	97.2/19.4	71.1/42.7	53.7/53.7	98.6/19.7	93.4/56.1	73.4/73.4
w/o $C_T$	96.7/19.3	70.1/42.1	53.3/53.3	98.1/19.6	91.8/55.1	72.2/72.2

Furthermore, KGDMC w/o  $\mathcal{L}_{IntraDD}$  and KGDMC w/o  $\mathcal{L}_{InterDD}$  both result in a decrease in performance, confirming the effectiveness of these two distribution disentanglement constraints to facilitate the construction of modality-specific subspaces. We then replace  $\mathbf{z}_I$  with  $\mathbf{z}_{oI}$  in cross-modality reconstruction, which leads to a significant performance drop. The accurate textual modality-invariant representations can better guide the reconstruction process. Finally, KGDMC w/o  $C_M$  performs worse than KGDMC, emphasizing the necessity of adaptive confidence pseudo-labels to help models mitigate the negative effects of noise interference.

For triplet error detection, visual modality provides abundant complementary information, enabling KGDMC with visual modality to outperform its unimodal version KGDMC w/o  $\mathbf{v}_e$ . However, the performance of KGDMC w/o  $w$  significantly drops and it even underperforms the unimodal KGDMC w/o  $\mathbf{v}_e$ . This is consistent with our claim that directly integrating erroneous visual information not only fails to aid in detecting triplet errors but also introduces additional noise, ultimately diminishing the overall performance. Moreover, we find that the performance experiences varying degrees of decline when separately removing each granularity of triplet confidence, confirming the effectiveness of these

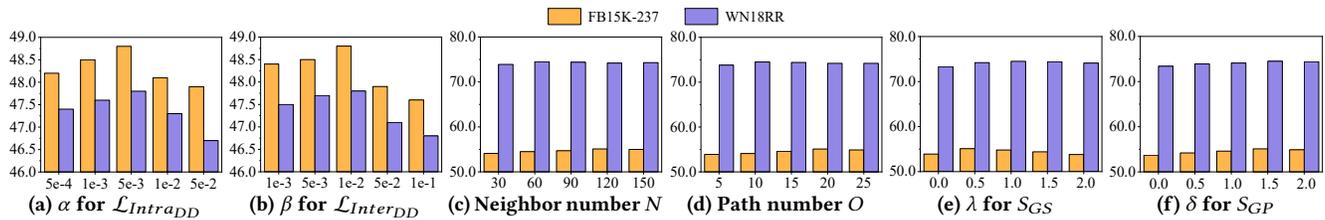


Figure 3: The precision@5% results of our KGDMC with varying hyperparameters on FB15K-237 and WN18RR.

Table 6: The precision@5% results of triplet error detection on various error types.

	Models	TransE	DistMult	ComplEx	KGTtm	CAGED	KG-BERT	StAR	CSProm-KG	CCA	IKRL	RSME	VISTA	KGDMC
FB15K-237	Random	72.6	66.2	73.4	73.0	75.8	67.4	72.8	73.2	76.8	70.3	72.5	71.1	<b>77.4</b>
	Similar	30.4	32.8	30.6	31.8	33.1	33.6	37.1	41.9	45.3	35.5	41.6	38.7	<b>50.2</b>
	Adversarial	12.5	12.5	15.0	12.8	12.6	19.9	16.4	21.1	24.0	17.2	20.6	18.3	<b>31.4</b>
WN18RR	Random	43.4	31.6	38.4	44.8	48.6	80.6	79.4	79.1	80.7	77.2	79.1	77.9	<b>81.1</b>
	Similar	35.1	33.8	30.3	39.1	38.8	63.3	61.0	63.4	65.7	59.8	62.4	61.1	<b>69.5</b>
	Adversarial	31.4	28.5	36.6	35.9	37.3	59.9	52.7	53.6	54.5	52.8	54.7	53.3	<b>62.8</b>

Table 7: Modality error detection on various error types.

	Models	CLIP	ViLT	ALBEF	METER	KGDMC
FB15K-237	Random	67.0	49.6	62.4	59.1	<b>71.9</b>
	Intra similar	34.7	26.0	30.2	23.6	<b>50.8</b>
	Inter similar	11.2	7.0	12.0	8.7	<b>23.6</b>
WN18RR	Random	64.4	52.6	60.9	60.7	<b>70.2</b>
	Intra similar	39.6	32.7	37.2	30.6	<b>53.5</b>
	Inter similar	9.8	6.5	10.3	8.9	<b>19.4</b>

three key components. They provide a variety of consistency information from different views to collaboratively detect triplet errors. Finally, similar to  $C_M$  in modality error detection, the utilization of  $C_T$  guides models to focus more on convincing triplets, thereby alleviating the negative effects of noisy triplets and ultimately improving overall performance.

**6.2.3 Hyperparameter Analysis (RQ3).** To investigate the influence of hyperparameters on the performance of our KGDMC, we vary them on the two datasets, as illustrated in Figure 3. For modality error detection, we can observe that KGDMC has a similar sensitivity to  $\alpha$  and  $\beta$  on the two datasets. A small value may render our model insufficient for effectively disentangling information, whereas an excessively large one could degrade the quality of latent variables, ultimately resulting in performance deterioration. When respectively setting  $\alpha$  and  $\beta$  as 0.005 and 0.01, our KGDMC gains the best results. For triplet error detection, we can find that the required number of neighbors and paths changes with different datasets, which is reasonable since WN18RR is much sparser than FB15K-237. Furthermore, increasing these numbers does not always lead to improved performance. Excessive information aggregation and passing can potentially exacerbate the negative effect of erroneous triplets. Optimal performance is achieved by our KGDMC when configured with  $N = 120$  and  $O = 20$  on FB15K-237, and  $N = 60$  and  $O = 10$  on WN18RR. Finally,  $\lambda$  and  $\delta$  are utilized to balance the contribution of three confidence scores for triplet error detection. Similarly, given that FB15K-237 exhibits a higher density, it tends to aggregate a greater number of erroneous triplet neighbors, which renders its  $S_{GS}$  less reliable compared to WN18RR. Consequently,  $S_{GS}$  has a smaller contribution on FB15K-237. When we set  $\lambda$  to 0.5 for FB15K-237 and to 1.0 for WN18RR, our KGDMC obtains optimal

results. However, we find that  $S_{GP}$  plays similar roles in the two datasets, and our KGDMC gains the best results when  $\delta = 1.5$ . This is due to the fact that when obtaining the unified path embeddings, we assign weights based on the local triplet confidence, thereby mitigating the negative impact of erroneous triplets.

**6.2.4 Perform on Different Error Types (RQ4).** To investigate the robustness of our KGDMC against various types of noise, we add 5% of each type noise to FB15K-237 and WN18RR separately. Table 6 and Table 7 show the precision@5% results. For modality error detection, while CLIP exhibits satisfactory results when confronted with random noise, it is far inferior to our KGDMC in addressing both intra-modality and inter-modality similar noise. Our KGDMC captures fine-grained semantic consistency across modalities by leveraging modality disentanglement and reconstruction techniques, enabling it to effectively identify highly similar erroneous entities. For triplet error detection, since baselines have distinguished random noise well, our KGDMC only slightly improves their performance. However, for semantic similar and adversarial noise, our KGDMC significantly outperforms all baselines. This is because the visual modality, coupled with multi-grained consistency from various views, provides additional complementary information for distinguishing these hard-to-detect noises.

## 7 Conclusion

In this paper, we are the first to introduce the novel multimodal knowledge graph error detection task, which aims to simultaneously detect both modality errors and triplet errors. To address this task, we propose a novel KGDMC framework. Specifically, we design a disentanglement modality reconstruction module to capture fine-grained semantic consistency across modalities to identify modality errors. Furthermore, we propose a multi-grained triplet confidence module to integrate the internal self-consistency of triplets, the consistency of aggregated neighbor information, and the consistency of information propagation, for triplet error detection. Experimental results on our constructed two multimodal knowledge graph datasets demonstrate the superiority of our KGDMC. In the future, we will explore the use of KGDMC to assist downstream tasks based on multimodal knowledge graphs.

## References

- [1] Rakesh Agrawal, Tomasz Imielinski, and Arun N. Swami. 1993. Mining Association Rules between Sets of Items in Large Databases. In *Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data, Washington, DC, USA, May 26–28, 1993*, Peter Buneman and Sushil Jajodia (Eds.). ACM Press, 207–216. <https://doi.org/10.1145/170035.170072>
- [2] Sören Auer, Christian Bizer, Georgi Kobilarov, Jens Lehmann, Richard Cyganiak, and Zachary G. Ives. 2007. DBpedia: A Nucleus for a Web of Open Data. In *The Semantic Web, 6th International Semantic Web Conference, 2nd Asian Semantic Web Conference, ISWC 2007 + ASWC 2007, Busan, Korea, November 11–15, 2007 (Lecture Notes in Computer Science, Vol. 4825)*, Karl Aberer, Key-Sun Choi, Natasha Fridman Noy, Dean Allemang, Kyung-Il Lee, Lyndon J. B. Nixon, Jennifer Golbeck, Peter Mika, Diana Maynard, Riichiro Mizoguchi, Guus Schreiber, and Philippe Cudré-Mauroux (Eds.). Springer, 722–735. [https://doi.org/10.1007/978-3-540-76298-0\\_52](https://doi.org/10.1007/978-3-540-76298-0_52)
- [3] Caleb Belth, Xinyi Zheng, Jilles Vreeken, and Danai Koutra. 2020. What is Normal, What is Strange, and What is Missing in a Knowledge Graph: Unified Characterization via Inductive Summarization. In *WWW '20: The Web Conference 2020, Taipei, Taiwan, April 20–24, 2020*, Yennun Huang, Irwin King, Tie-Yan Liu, and Maarten van Steen (Eds.). ACM / IW3C2, 1115–1126. <https://doi.org/10.1145/3366423.3380189>
- [4] Kurt D. Bollacker, Colin Evans, Praveen K. Paritosh, Tim Sturge, and Jamie Taylor. 2008. Freebase: a collaboratively created graph database for structuring human knowledge. In *Proceedings of the ACM SIGMOD International Conference on Management of Data, SIGMOD 2008, Vancouver, BC, Canada, June 10–12, 2008*, Jason Tsong-Li Wang (Ed.). ACM, 1247–1250. <https://doi.org/10.1145/1376616.1376746>
- [5] Antoine Bordes, Nicolas Usunier, Alberto García-Durán, Jason Weston, and Oksana Yakhnenko. 2013. Translating Embeddings for Modeling Multi-relational Data. In *Advances in Neural Information Processing Systems 26: 27th Annual Conference on Neural Information Processing Systems 2013. Proceedings of a meeting held December 5–8, 2013, Lake Tahoe, Nevada, United States*, Christopher J. C. Burges, Léon Bottou, Zoubin Ghahramani, and Kilian Q. Weinberger (Eds.). 2787–2795. <https://proceedings.neurips.cc/paper/2013/hash/1cecc7a77928ca8133fa24680a882df9-Abstract.html>
- [6] Zongsheng Cao, Qianqian Xu, Zhiyong Yang, Yuan He, Xiaochun Cao, and Qingming Huang. 2022. OTKGE: Multi-modal Knowledge Graph Embeddings via Optimal Transport. In *Advances in Neural Information Processing Systems 35: Annual Conference on Neural Information Processing Systems 2022, NeurIPS 2022, New Orleans, LA, USA, November 28 - December 9, 2022*, Sanmi Koyejo, S. Mohamed, A. Agarwal, Danielle Belgrave, K. Cho, and A. Oh (Eds.). [http://papers.nips.cc/paper\\_files/paper/2022/hash/ffdb280e7c7b4c4af30e04daf5a84b98-Abstract-Conference.html](http://papers.nips.cc/paper_files/paper/2022/hash/ffdb280e7c7b4c4af30e04daf5a84b98-Abstract-Conference.html)
- [7] Andrew Carlson, Justin Betteridge, Bryan Kisiel, Burr Settles, Estevam R. Hruschka Jr., and Tom M. Mitchell. 2010. Toward an Architecture for Never-Ending Language Learning. In *Proceedings of the Twenty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2010, Atlanta, Georgia, USA, July 11–15, 2010*, Maria Fox and David Poole (Eds.). AAAI Press, 1306–1313. <https://doi.org/10.1609/AAAI.V24I1.7519>
- [8] Andrew Carlson, Justin Betteridge, Richard C. Wang, Estevam R. Hruschka Jr., and Tom M. Mitchell. 2010. Coupled semi-supervised learning for information extraction. In *Proceedings of the Third International Conference on Web Search and Web Data Mining, WSDM 2010, New York, NY, USA, February 4–6, 2010*, Brian D. Davison, Torsten Suel, Nick Craswell, and Bing Liu (Eds.). ACM, 101–110. <https://doi.org/10.1145/1718487.1718501>
- [9] Chen Chen, Yufei Wang, Aixin Sun, Bing Li, and Kwok-Yan Lam. 2023. Dipping PLMs Sauce: Bridging Structure and Text for Effective Knowledge Graph Completion via Conditional Soft Prompting. In *Findings of the Association for Computational Linguistics: ACL 2023, Toronto, Canada, July 9–14, 2023*, Anna Rogers, Jordan L. Boyd-Graber, and Naoaki Okazaki (Eds.). Association for Computational Linguistics, 11489–11503. <https://doi.org/10.18653/v1/2023.FINDINGS-ACL.729>
- [10] Yurong Cheng, Lei Chen, Ye Yuan, and Guoren Wang. 2018. Rule-Based Graph Repairing: Semantic and Efficient Repairing Methods. In *34th IEEE International Conference on Data Engineering, ICDE 2018, Paris, France, April 16–19, 2018*. IEEE Computer Society, 773–784. <https://doi.org/10.1109/ICDE.2018.00075>
- [11] Tim Dettmers, Pasquale Minervini, Pontus Stenetorp, and Sebastian Riedel. 2018. Convolutional 2D Knowledge Graph Embeddings. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, (AAAI-18), the 30th innovative Applications of Artificial Intelligence (IAAI-18), and the 8th AAAI Symposium on Educational Advances in Artificial Intelligence (EAAI-18), New Orleans, Louisiana, USA, February 2–7, 2018*, Sheila A. McIlraith and Kilian Q. Weinberger (Eds.). AAAI Press, 1811–1818. <https://doi.org/10.1609/AAAI.V32I1.11573>
- [12] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2019, Minneapolis, MN, USA, June 2–7, 2019, Volume 1 (Long and Short Papers)*, Jill
- Burstein, Christy Doran, and Tamar Solorio (Eds.). Association for Computational Linguistics, 4171–4186. <https://doi.org/10.18653/v1/N19-1423>
- [13] Junnan Dong, Qinggang Zhang, Xiao Huang, Qiaoyu Tan, Daochen Zha, and Zihao Zhao. 2023. Active Ensemble Learning for Knowledge Graph Error Detection. In *Proceedings of the Sixteenth ACM International Conference on Web Search and Data Mining, WSDM 2023, Singapore, 27 February 2023 - 3 March 2023*, Tat-Seng Chua, Hady W. Lauw, Luo Si, Evimaria Terzi, and Panayiotis Tsaparas (Eds.). ACM, 877–885. <https://doi.org/10.1145/3539597.3570368>
- [14] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xi-aohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. 2021. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. In *9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3–7, 2021*. OpenReview.net. <https://openreview.net/forum?id=YicbFdNTTy>
- [15] Yi-Yi Dou, Yichong Xu, Zhe Gan, Jianfeng Wang, Shuohang Wang, Lijuan Wang, Chenguang Zhu, Pengchuan Zhang, Lu Yuan, Nanyun Peng, Zicheng Liu, and Michael Zeng. 2022. An Empirical Study of Training End-to-End Vision-and-Language Transformers. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2022, New Orleans, LA, USA, June 18–24, 2022*. IEEE, 18145–18155. <https://doi.org/10.1109/CVPR52688.2022.01763>
- [16] Quan Fang, Xiaowei Zhang, Jun Hu, Xian Wu, and Changsheng Xu. 2023. Contrastive Multi-Modal Knowledge Graph Representation Learning. *IEEE Trans. Knowl. Data Eng.* 35, 9 (2023), 8983–8996. <https://doi.org/10.1109/TKDE.2022.3220625>
- [17] Luis Antonio Galárraga, Christina Teflioudi, Katja Hose, and Fabian M. Suchanek. 2013. AMIE: association rule mining under incomplete evidence in ontological knowledge bases. In *22nd International World Wide Web Conference, WWW '13, Rio de Janeiro, Brazil, May 13–17, 2013*, Daniel Schwabe, Virgilio A. F. Almeida, Hartmut Glaser, Ricardo Baeza-Yates, and Sue B. Moon (Eds.). International World Wide Web Conferences Steering Committee / ACM, 413–422. <https://doi.org/10.1145/2488388.2488425>
- [18] Congcong Ge, Yunjun Gao, Honghui Weng, Chong Zhang, Xiaoye Miao, and Baihua Zheng. 2020. KGClean: An Embedding Powered Knowledge Graph Cleaning Framework. *CoRR abs/2004.14478* (2020). arXiv:2004.14478 <https://arxiv.org/abs/2004.14478>
- [19] Clark R Givens and Rae Michael Shortt. 1984. A class of Wasserstein metrics for probability distributions. *Michigan Mathematical Journal* 31, 2 (1984), 231–240.
- [20] Qingyu Guo, Fuzhen Zhuang, Chuan Qin, Hengshu Zhu, Xing Xie, Hui Xiong, and Qing He. 2022. A Survey on Knowledge Graph-Based Recommender Systems. *IEEE Trans. Knowl. Data Eng.* 34, 8 (2022), 3549–3568. <https://doi.org/10.1109/TKDE.2020.3028705>
- [21] Xiao Huang, Jingyuan Zhang, Dingcheng Li, and Ping Li. 2019. Knowledge Graph Embedding Based Question Answering. In *Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining, WSDM 2019, Melbourne, VIC, Australia, February 11–15, 2019*, J. Shane Culpepper, Alistair Moffat, Paul N. Bennett, and Kristina Lerman (Eds.). ACM, 105–113. <https://doi.org/10.1145/3289600.3290956>
- [22] Hyeonjoo Hwang, Geon-Hyeong Kim, Seunghoon Hong, and Kee-Eung Kim. 2020. Variational Interaction Information Maximization for Cross-domain Disentanglement. In *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6–12, 2020, virtual*, Hugo Larochelle, Marc'Aurelio Ranzato, Raia Hadsell, Maria-Florina Balcan, and Hsuan-Tien Lin (Eds.). <https://proceedings.neurips.cc/paper/2020/hash/fe663a72b27bdc61387fbb512f6f67-Abstract.html>
- [23] Shengbin Jia, Yang Xiang, Xiaojun Chen, Kun Wang, and Shijia E. 2019. Triple Trustworthiness Measurement for Knowledge Graph. In *The World Wide Web Conference, WWW 2019, San Francisco, CA, USA, May 13–17, 2019*, Ling Liu, Ryan W. White, Amin Mantrach, Fabrizio Silvestri, Julian J. McAuley, Ricardo Baeza-Yates, and Leila Zia (Eds.). ACM, 2865–2871. <https://doi.org/10.1145/3308558.3313586>
- [24] Wonjae Kim, Bokyung Son, and Ildoo Kim. 2021. ViLT: Vision-and-Language Transformer Without Convolution or Region Supervision. In *Proceedings of the 38th International Conference on Machine Learning, ICMML 2021, 18–24 July 2021, Virtual Event (Proceedings of Machine Learning Research, Vol. 139)*, Marina Meila and Tong Zhang (Eds.). PMLR, 5583–5594. <http://proceedings.mlr.press/v139/kim21k.html>
- [25] Diederik P. Kingma and Jimmy Ba. 2015. Adam: A Method for Stochastic Optimization. In *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7–9, 2015, Conference Track Proceedings*, Yoshua Bengio and Yann LeCun (Eds.). <http://arxiv.org/abs/1412.6980>
- [26] Diederik P. Kingma and Max Welling. 2014. Auto-Encoding Variational Bayes. In *2nd International Conference on Learning Representations, ICLR 2014, Banff, AB, Canada, April 14–16, 2014, Conference Track Proceedings*, Yoshua Bengio and Yann LeCun (Eds.). <http://arxiv.org/abs/1312.6114>
- [27] Jaesun Lee, Chanyoung Chung, Hochang Lee, Sungho Jo, and Joyce Jiyoung Whang. 2023. VISTA: Visual-Textual Knowledge Graph Representation Learning. In *Findings of the Association for Computational Linguistics: EMNLP 2023, Singapore, December 6–10, 2023*, Houda Bouamor, Juan Pino, and Kalika Bali (Eds.).

- Association for Computational Linguistics, 7314–7328. <https://doi.org/10.18653/V1/2023.FINDINGS-EMNLP.488>
- [28] Junnan Li, Ramprasaath R. Selvaraju, Akhilesh Gotmare, Shafiq R. Joty, Caiming Xiong, and Steven Chu-Hong Hoi. 2021. Align before Fuse: Vision and Language Representation Learning with Momentum Distillation. In *Advances in Neural Information Processing Systems 34: Annual Conference on Neural Information Processing Systems 2021, NeurIPS 2021, December 6–14, 2021, virtual*, Marc'Antonio Ranzato, Alina Beygelzimer, Yann N. Dauphin, Percy Liang, and Jennifer Wortman Vaughan (Eds.). 9694–9705. <https://proceedings.neurips.cc/paper/2021/hash/505259756244493872b7709a8a01b536-Abstract.html>
- [29] Xiangyu Liu, Yang Liu, and Wei Hu. 2024. Knowledge Graph Error Detection with Contrastive Confidence Adaptation. In *Thirty-Eighth AAAI Conference on Artificial Intelligence, AAAI 2024, Thirty-Sixth Conference on Innovative Applications of Artificial Intelligence, IAAI 2024, Fourteenth Symposium on Educational Advances in Artificial Intelligence, EAAI 2024, February 20–27, 2024, Vancouver, Canada*, Michael J. Wooldridge, Jennifer G. Dy, and Sriraam Natarajan (Eds.). AAAI Press, 8824–8831. <https://doi.org/10.1609/AAAI.V38I8.28729>
- [30] Xinyu Lu, Lifang Wang, Zejun Jiang, Shichang He, and Shizhong Liu. 2022. MMKRL: A robust embedding approach for multi-modal knowledge graph representation learning. *Appl. Intell.* 52, 7 (2022), 7480–7497. <https://doi.org/10.1007/S10489-021-02693-9>
- [31] André Melo and Heiko Paulheim. 2017. Detection of Relation Assertion Errors in Knowledge Graphs. In *Proceedings of the Knowledge Capture Conference, K-CAP 2017, Austin, TX, USA, December 4–6, 2017*, Oscar Corcho, Krzysztof Janowicz, Giuseppe Rizzo, Ilaria Tiddi, and Daniel Garijo (Eds.). ACM, 22:1–22:8. <https://doi.org/10.1145/3148011.3148033>
- [32] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. 2021. Learning Transferable Visual Models From Natural Language Supervision. In *Proceedings of the 38th International Conference on Machine Learning, ICML 2021, 18–24 July 2021, Virtual Event (Proceedings of Machine Learning Research, Vol. 139)*, Marina Meila and Tong Zhang (Eds.). PMLR, 8748–8763. <http://proceedings.mlr.press/v139/radford21a.html>
- [33] Apoorv Saxena, Aditay Tripathi, and Partha P. Talukdar. 2020. Improving Multi-hop Question Answering over Knowledge Graphs using Knowledge Base Embeddings. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, ACL 2020, Online, July 5–10, 2020*, Dan Jurafsky, Joyce Chai, Natalie Schluter, and Joel R. Tetreault (Eds.). Association for Computational Linguistics, 4498–4507. <https://doi.org/10.18653/V1/2020.ACL-MAIN.412>
- [34] Edgar Schönfeld, Sayna Ebrahimi, Samarth Sinha, Trevor Darrell, and Zeynep Akata. 2019. Generalized Zero- and Few-Shot Learning via Aligned Variational Autoencoders. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2019, Long Beach, CA, USA, June 16–20, 2019*. Computer Vision Foundation / IEEE, 8247–8255. <https://doi.org/10.1109/CVPR.2019.00844>
- [35] Hatem Mouselly Sergieh, Teresa Botschen, Iryna Gurevych, and Stefan Roth. 2018. A Multimodal Translation-Based Approach for Knowledge Graph Representation Learning. In *Proceedings of the Seventh Joint Conference on Lexical and Computational Semantics, \*SEM@NAACL-HLT 2018, New Orleans, Louisiana, USA, June 5–6, 2018*, Malvina Nissim, Jonathan Berant, and Alessandro Lenci (Eds.). Association for Computational Linguistics, 225–234. <https://doi.org/10.18653/V1/S18-2027>
- [36] Zhan Su, Zhicheng Dou, Yutao Zhu, and Ji-Rong Wen. 2022. Knowledge Enhanced Search Result Diversification. In *KDD '22: The 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Washington, DC, USA, August 14–18, 2022*, Aidong Zhang and Huzefa Rangwala (Eds.). ACM, 1687–1695. <https://doi.org/10.1145/3534678.3539459>
- [37] Thomas Pellissier Tanon, Denny Vrandečić, Sebastian Schaffert, Thomas Steiner, and Lydia Pintscher. 2016. From Freebase to Wikidata: The Great Migration. In *Proceedings of the 25th International Conference on World Wide Web, WWW 2016, Montreal, Canada, April 11–15, 2016*, Jacqueline Bourdeau, Jim Hendler, Roger Nkambou, Ian Horrocks, and Ben Y. Zhao (Eds.). ACM, 1419–1428. <https://doi.org/10.1145/2872427.2874809>
- [38] Kristina Toutanova and Danqi Chen. 2015. Observed versus latent features for knowledge base and text inference. In *Proceedings of the 3rd Workshop on Continuous Vector Space Models and their Compositionality, CVSC 2015, Beijing, China, July 26–31, 2015*, Alexandre Allauzen, Edward Grefenstette, Karl Moritz Hermann, Hugo Larochelle, and Scott Wen-tau Yih (Eds.). Association for Computational Linguistics, 57–66. <https://doi.org/10.18653/V1/W15-4007>
- [39] Théo Trouillon, Johannes Welbl, Sebastian Riedel, Éric Gaussier, and Guillaume Bouchard. 2016. Complex Embeddings for Simple Link Prediction. In *Proceedings of the 33rd International Conference on Machine Learning, ICML 2016, New York City, NY, USA, June 19–24, 2016 (JMLR Workshop and Conference Proceedings, Vol. 48)*, Maria-Florina Balcan and Kilian Q. Weinberger (Eds.). JMLR.org, 2071–2080. <http://proceedings.mlr.press/v48/trouillon16.html>
- [40] John C Turner and Penelope J Oakes. 1986. The significance of the social identity concept for social psychology with reference to individualism, interactionism and social influence. *British Journal of Social Psychology* 25, 3 (1986), 237–252.
- [41] Petar Velickovic, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. 2018. Graph Attention Networks. In *6th International Conference on Learning Representations, ICLR 2018, Vancouver, BC, Canada, April 30 - May 3, 2018, Conference Track Proceedings*. OpenReview.net. <https://openreview.net/forum?id=rJXMpikCZ>
- [42] Bo Wang, Tao Shen, Guodong Long, Tianyi Zhou, Ying Wang, and Yi Chang. 2021. Structure-Augmented Text Representation Learning for Efficient Knowledge Graph Completion. In *WWW '21: The Web Conference 2021, Virtual Event / Ljubljana, Slovenia, April 19–23, 2021*, Jure Leskovec, Marko Grobelnik, Marc Najork, Jie Tang, and Leila Zia (Eds.). ACM / IW3C2, 1737–1748. <https://doi.org/10.1145/3442381.3450043>
- [43] Meng Wang, Sen Wang, Han Yang, Zheng Zhang, Xi Chen, and Guilin Qi. 2021. Is Visual Context Really Helpful for Knowledge Graph? A Representation Learning Perspective. In *MM '21: ACM Multimedia Conference, Virtual Event, China, October 20 - 24, 2021*, Heng Tao Shen, Yueying Zhuang, John R. Smith, Yang Yang, Pablo César, Florian Metz, and Balakrishnan Prabhakaran (Eds.). ACM, 2735–2743. <https://doi.org/10.1145/3474085.3475470>
- [44] Zikang Wang, Linjing Li, Qiudan Li, and Daniel Zeng. 2019. Multimodal Data Enhanced Representation Learning for Knowledge Graphs. In *International Joint Conference on Neural Networks, IJCNN 2019 Budapest, Hungary, July 14–19, 2019*. IEEE, 1–8. <https://doi.org/10.1109/IJCNN.2019.8852079>
- [45] Ruobing Xie, Zhiyuan Liu, Fen Lin, and Leyu Lin. 2018. Does William Shakespeare REALLY Write Hamlet? Knowledge Representation Learning With Confidence. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, (AAAI-18), the 30th innovative Applications of Artificial Intelligence (IAAI-18), and the 8th AAAI Symposium on Educational Advances in Artificial Intelligence (EAAI-18), New Orleans, Louisiana, USA, February 2–7, 2018*, Sheila A. McIlraith and Kilian Q. Weinberger (Eds.). AAAI Press, 4954–4961. <https://doi.org/10.1609/AAAI.V32I1.11924>
- [46] Ruobing Xie, Zhiyuan Liu, Huanbo Luan, and Maosong Sun. 2017. Image-embodied Knowledge Representation Learning. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI 2017, Melbourne, Australia, August 19–25, 2017*, Carles Sierra (Ed.). ijcai.org, 3140–3146. <https://doi.org/10.24963/IJCAI.2017/438>
- [47] Bishan Yang, Wen-tau Yih, Xiaodong He, Jianfeng Gao, and Li Deng. 2015. Embedding Entities and Relations for Learning and Inference Bases. In *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7–9, 2015, Conference Track Proceedings*, Yoshua Bengio and Yann LeCun (Eds.). <http://arxiv.org/abs/1412.6575>
- [48] Yun Yang, Yulong Rao, Minghao Yu, and Yan Kang. 2022. Multi-layer information fusion based on graph convolutional network for knowledge-driven herb recommendation. *Neural Networks* 146 (2022), 1–10. <https://doi.org/10.1016/J.NEUNET.2021.11.010>
- [49] Liang Yao, Chengsheng Mao, and Yuan Luo. 2019. KG-BERT: BERT for Knowledge Graph Completion. *CoRR abs/1909.03193* (2019). [arXiv:1909.03193](http://arxiv.org/abs/1909.03193) <http://arxiv.org/abs/1909.03193>
- [50] Fuwei Zhang, Zhao Zhang, Xiang Ao, Dehong Gao, Fuzhen Zhuang, Yi Wei, and Qing He. 2022. Mind the Gap: Cross-Lingual Information Retrieval with Hierarchical Knowledge Enhancement. In *Thirty-Sixth AAAI Conference on Artificial Intelligence, AAAI 2022, Thirty-Fourth Conference on Innovative Applications of Artificial Intelligence, IAAI 2022, The Twelfth Symposium on Educational Advances in Artificial Intelligence, EAAI 2022 Virtual Event, February 22 - March 1, 2022*. AAAI Press, 4345–4353. <https://doi.org/10.1609/AAAI.V36I4.20355>
- [51] Qinggang Zhang, Junnan Dong, Keyu Duan, Xiao Huang, Yezi Liu, and Linchuan Xu. 2022. Contrastive Knowledge Graph Error Detection. In *Proceedings of the 31st ACM International Conference on Information & Knowledge Management, Atlanta, GA, USA, October 17–21, 2022*, Mohammad Al Hasan and Li Xiong (Eds.). ACM, 2590–2599. <https://doi.org/10.1145/3511808.3557264>
- [52] Zhao Zhang, Fuwei Zhang, Fuzhen Zhuang, and Yongjun Xu. 2023. Knowledge Graph Error Detection with Hierarchical Path Structure. In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management, CIKM 2023, Birmingham, United Kingdom, October 21–25, 2023*, Ingo Frommholz, Frank Hopfgartner, Mark Lee, Michael Oakes, Mounia Lalmas, Min Zhang, and Rodrigo L. T. Santos (Eds.). ACM, 4430–4434. <https://doi.org/10.1145/3583780.3615201>