LAZR: LLM-Augmented Zero-Shot Relational Learning via Factor Composition and Ranking-Based Semantic Reconstruction

Kai Zhang, *Member, IEEE*, Yuanhui Long, Shuanghong Shen, Shulan Ruan ,Min Hou, Le Wu , Qi Liu, *Member, IEEE* and Enhong Chen, *Fellow, IEEE*

***Abstract*—Zero-Shot Relational Learning (ZSRL) enables knowledge graphs to reason over unseen relations but faces two limitations: semantic sparsity from brief relational descriptions, and reliance on manual annotations or rigid regression constraints (e.g., MSE loss) that fail to capture the ordinal structure of relational semantics. We propose LAZR (LLM-Augmented Zero-shot Relational Learning), a framework that leverages large language models to reconstruct high-density, discriminative relational representations. LAZR introduces: (1) *LLM-based semantic enrichment* that decomposes relations into shared semantic factors for cross-relation knowledge transfer; and (2) *ranking-based semantic reconstruction* that uses LLM-generated similarity scores with ranking loss to preserve ordinal topology. We instantiate LAZR within a generative adversarial framework to synthesize embeddings for unseen relations. Experiments on NELL-ZS and Wiki-ZS show substantial gains: compared with the strongest FZR baseline, LAZR improves MRR/Hit@1 by up to 2.6/2.5 percentage points on NELL-ZS and 2.5/2.4 percentage points on Wiki-ZS. Cross-dataset analysis further reveals that the effectiveness of semantic enhancement depends on the intrinsic clustering structure of the knowledge graph, identifying a scalability boundary for relation semantic enhancement techniques.**

***Index Terms*—Zero-shot relational learning, knowledge graph completion, large language models, semantic reconstruction, ranking loss.**

I. Introduction

KNOWLEDGE Graphs (KGs) support applications such as question answering [1], [2], recommendation [3], [4], and decision support [5], yet they remain incomplete. Conventional knowledge graph completion assumes that every relation is observed during training and therefore cannot directly predict facts involving a newly introduced relation. Zero-Shot Relational Learning (ZSRL) addresses this setting by transferring knowledge from seen to unseen relations through auxiliary semantics, typically textual descriptions [6] or ontological schemas [7].

Existing ZSRL methods are limited by two related problems. First, relation descriptions are usually short and omit entity roles, scenario constraints, and latent associations; the resulting semantic sparsity weakens factor extraction and cross-relation transfer [8], [9]. Second, the topology of the relation space is poorly supervised. FZR [10], for example, composes shared factors but requires manual similarity scores and fits them with mean squared error. Manual scoring is difficult to scale, while pointwise regression penalizes numerical deviations even when the relative ordering of relation similarities is correct. It consequently provides weak supervision for distinguishing closely related relations.

We propose **LAZR** (**L**LM-**A**ugmented **Z**ero-shot **R**elational Learning), which uses an LLM to supply both richer descriptions and scalable semantic supervision. LAZR first expands each description while preserving the original relation meaning, making latent factors such as "sport" explicit in Fig. 1. Shared-factor composition then transfers these factors across relations. In parallel, LLM embeddings generate an automatic relation-similarity matrix. A ranking objective reconstructs the semantic space by preserving similarity order rather than exact scores, and an adaptive negative-selection rule avoids trivially dissimilar negatives in sparse relation spaces. The reconstructed representation conditions a generative adversarial model that synthesizes features for unseen relations.

Experiments on NELL-ZS and Wiki-ZS show consistent improvements over ZSGAN, OntoZSL, and FZR. Additional experiments on CoDEx-L and FB15k-237 expose an important boundary: factor-based enhancement is most effective when the KG contains coherent semantic neighborhoods, but may mix unrelated signals in attribute-like or long-tail relation spaces. This result motivates structure-aware selection of enhancement modules rather than a uniform pipeline.

Our main contributions are as follows:

- We formulate semantic sparsity and ordinal-topology mismatch as two central bottlenecks in ZSRL.
- We introduce automated LLM-based enrichment and ranking-based reconstruction, eliminating manual relation scoring while preserving similarity order.
- We obtain state-of-the-art results on NELL-ZS and Wiki-ZS and identify, through two additional datasets and structural diagnostics, when relation semantic enhancement ceases to help.

Kai Zhang, Yuanhui Long, Qi Liu, and Enhong Chen are with the State Key Laboratory of Cognitive Intelligence, University of Science and Technology of China, Hefei, China.

Shulan Ruan is with the School of Computer Science and Technology, Harbin Institute of Technology, Shenzhen, China

Shuanghong Shen is with the Institute of Artificial Intelligence, Hefei National Science Center, Hefei, China.

Min Hou and Le Wu are with the Innovation School of Artificial Intelligence, Hefei University of Technology, Hefei, China.

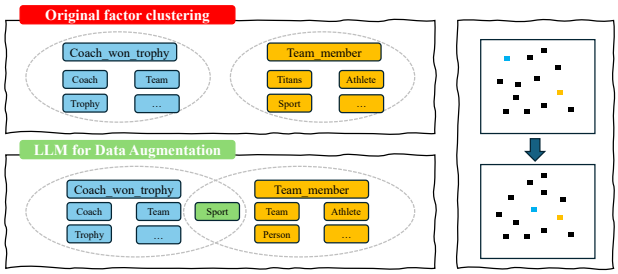


Fig. 1. Comparison of shared factor extraction.

II. Related Work

A. Zero-Shot Learning

Zero-shot learning transfers knowledge to unseen classes through auxiliary semantics. Early methods use attributes [11]–[15]; text-based methods use descriptions or pretrained embeddings [6], [16]–[19], sometimes with noise suppression [20], [21]; and graph or ontology methods exploit structured class relations [22]–[26]. Mapping methods align instance and semantic spaces [27]–[29], whereas generative methods synthesize unseen-class samples [30]–[33]. LAZR follows the generative paradigm but focuses on relation-level transfer and reconstructs its conditioning semantics before generation.

B. Zero-Shot Relational Learning for Knowledge Graphs

ZSRL predicts facts whose query relations were unseen during training. ZSGAN [6] conditions a GAN on textual relation descriptions, OntoZSL [7] additionally encodes ontology structure, and FZR [10] composes shared relational factors. Related inductive KG methods instead target unseen entities using auxiliary links [34]–[37], entity descriptions [38], or entity-agnostic subgraphs [39], [40]. LAZR differs from both groups: it enriches sparse relation descriptions, replaces FZR's manual similarity annotation with LLM-derived supervision, and optimizes relative similarity order rather than pointwise scores.

C. LLM in Semantic Enhancement

LLMs have been used to enrich sparse text [41]–[43] and to obtain contextual embeddings for recommendation and clustering [44]–[46]. In ZSRL, descriptions are the primary bridge between seen and unseen relations [6], [25]; their omissions therefore propagate directly into the learned relation space. LAZR uses generation to expose implicit roles and associations, then uses LLM embeddings only as similarity supervision. This division retains efficient GloVe-based relation features while exploiting the LLM's contextual knowledge for topology reconstruction.

III. Methodology

This section first formalizes the research problem and introduces relevant notation. Subsequently, we propose LAZR, a generic Zero-Shot Learning (ZSL) framework comprising four core components: 1) LLM-enhanced text representations, 2) augmented relational representations via shared factor composition, 3) LLM-expert-guided semantic reconstruction, and 4) a generative adversarial network (GAN) for synthesizing semantic embeddings to facilitate knowledge transfer learning. The overall architecture of the LAZR framework is depicted in Fig. 2.

A. Task Formulation

We study zero-shot link prediction on a knowledge graph (KG) $\mathcal{G}$. In the link prediction task, a knowledge graph comprises a set of entities $\mathcal{E}$, a set of relations $\mathcal{R}$, and a set of factual triplets $\mathcal{T} = \{(h, r, t) \mid h, t \in \mathcal{E}; r \in \mathcal{R}\}$. The task aims to distinguish positive examples from negative ones, where zero-shot relation learning acquires knowledge from seen relations $\mathcal{R}_s$ to facilitate transfer to unseen relations $\mathcal{R}_u$. Notably, $\mathcal{R}_s \cap \mathcal{R}_u = \emptyset$. Our objective can be formalized as: Given a head entity e_1 and a query relation r, predict the tail entity e_2. For each query tuple (e_1, r), there exists a ground-truth tail entity e_2 and a candidate set $C(e_1, r)$. We define the training dataset as:

$$\mathcal{D}_{\text{tr}} = \{(e_1, r_s, e_2, C(e_1, r_s)) \mid e_1, e_2 \in \mathcal{E}; r_s \in \mathcal{R}_s\}$$

and the test dataset as:

$$\mathcal{D}_{\text{te}} = \{(e_1, r_u, e_2, C(e_1, r_u)) \mid e_1, e_2 \in \mathcal{E}; r_u \in \mathcal{R}_u\}.$$

Under this setting, the goal of zero-shot link prediction is: For a given (e_1, r_u), maximize the score of the ground-truth entity e_2 within the candidate set $C(e_1, r_u)$ such that it achieves the highest rank, thereby enabling accurate prediction of factual triples involving unseen relations r_u. For predicting factual triples, we employ generative models to synthesize samples for unseen relations r_u. Regarding these generative models, we focus on constructing enhanced relation representations to improve model performance.

B. Large language model enhanced text representations

In zero-shot relation learning, the textual description of a relation serves as **core prior knowledge**, and its quality directly affects the effectiveness of semantic representation and the discriminative ability of shared factors. In the original FZR model, the initial semantic representation r_0 of a relation is obtained by averaging the word vectors in the original textual description (Equation 1), and the extraction of shared factors also relies on this original description. However, the textual descriptions of relations in real-world knowledge graphs inherently suffer from **conciseness and information sparsity**: on one hand, for the convenience of manual annotation and improved storage efficiency, relation descriptions usually only contain core logic (e.g., "league_coaches: specifies that a particular coach coaches a team that plays in a particular league"), lacking the portrayal of **deep-level information** such as entity roles, scenario constraints, and implicit associations; on the other hand, this conciseness may lead to high similarity in the surface descriptions of different relations (e.g., "coach_won_trophy" and "team_won_trophy"), making it difficult to distinguish their intrinsic differences during the extraction of shared factors, which in turn affects the accuracy of clustering and the discriminability of semantic representation in the **SFC (Shared Factor Combination)** process.

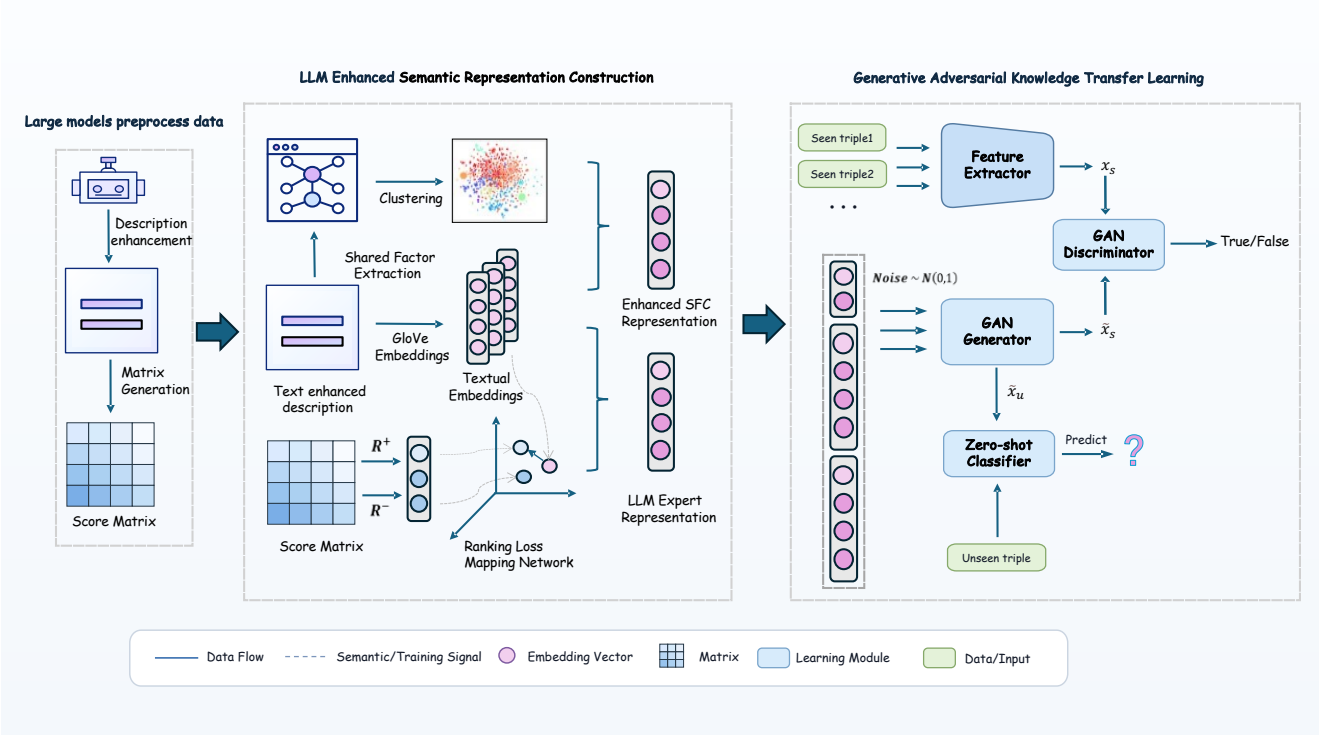


Fig. 2. Overview of the proposed LAZR framework.

To address the above issues, we introduce a **text enhancement method based on generative large language models (LLMs)**. By semantically expanding the original relation descriptions, this method generates textual representations with richer information and more complete semantics, providing **high-quality input** for subsequent shared factor extraction and initial semantic representation construction.

1) Enhancement Objectives and Design Principles: The main goal of LLM-based enhancement is to keep the original meaning of the relation and add following information to make the text more informative:

- **Refinement of entity roles**: Clarify the specific attributes or roles of the entities involved in the relation (e.g., whether "coach" specifically refers to "head coach" and whether "team" refers to "professional team");
- **Supplement of scenarios and constraints**: Add implicit scenarios for establishment of the relation (e.g., the level and geographical scope of the league in "league_coaches") or preconditions (e.g., the type of trophy and nature of the competition in "coach_won_trophy");
- **Prompting of associated relations**: Other indirectly associated relations (such as "league_coaches" may be associated with "team_in_league" and "coach_contract"), providing clues for shared factors to capture cross-relation associations;

The enhancement process must adhere to the **principle of semantic fidelity** (not deviating from the core definition of the original relation) and the **principle of moderate expansion** (avoiding redundant information from interfering with the extraction of key factors).

2) LLM Enhancement Framework and Prompt Engineering: Based on the above objectives, our work designs a **structured Prompt Engineering framework** to guide LLMs in generating high-quality enhanced text. The specific process is as follows:

a) Construction of input information:: The information input to the LLM includes: **the original relation name and the original relation description**.

b) Selection of LLM and control of generation:: We use **Qwen3** [47] for description enhancement and similarity scoring because it combines strong semantic modeling with deployment on consumer-grade GPUs. A low sampling temperature reduces generation variance and improves reproducibility. The framework itself is model-agnostic: another LLM can be substituted as long as the same structured prompt and embedding interface are available.

C. Augmented relational representations via shared factor composition

Through the content of Section 3.2, we obtained the enhanced relational text description. In this section, we use this enhanced relational text description to construct semantic vectors and extract shared factors to achieve the purpose of enhancing semantic vectors.

In the field of cognitive psychology [48], relevant studies have shown that multiple small units jointly participate in a series of cognitive activities. When humans are exposed to new procedural knowledge, they update the relevant components in these units to achieve the update of their own cognitive activities [49]. In this study, the shared factors involved in the enhanced relation descriptions are the key units that facilitate the learning of new relations. Therefore, we adopt the

"**Shared Factor Combination (SFC)**" method. The core goal of SFC is to **construct enhanced semantic representations of shared factors from the textual descriptions of relations**, thereby **improving the effect of knowledge transfer in zero-shot relation learning scenarios**. Based on this, the SFC method integrates technologies such as key phrase extraction, clustering algorithms, and factor representation combination, providing a comprehensive framework for relation representation in zero-shot relation learning. It not only enhances the model's ability to understand unseen relations but also offers an effective path for exploring potential associations between relations.

With the enhanced textual descriptions as prior knowledge, the initial semantic representation of a relation can be obtained using the pre-trained word embedding model Glove [50]. For an enhanced textual description composed of words $\{w_1, w_2, \ldots, w_n\}$, the initial semantic representation r_0 of relation r is obtained by aggregating their word vectors:

$$r_0 = \frac{1}{n}\sum_{i=1}^{n} w_i \tag{1}$$

To lay the foundation for the implementation of SFC, we first **extract key factors from the textual descriptions**, and these factors will serve as the basic components of the factor library. In this key step, we use the unsupervised keyword extraction algorithm provided by the Python keyphrase extraction toolkit [51] to efficiently identify phrases and terms that best reflect the essence of the text content, and then aggregate them to form a shared factor library. It should be specially noted that **no deduplication is performed on the extracted factors**, aiming to retain their respective contributions to the overall semantic structure.

Supported by existing prior knowledge, the inherent semantic associations between relations can be effectively identified. Even if these relations are applied in different domains, the model can still understand and integrate such semantic associations, thereby enhancing reasoning ability. Therefore, this method is committed to **mining shared semantic factors between relations and reconstructing the initial semantic representation r_0 of relations through shared factor combination**. Specifically, SFC is implemented by **clustering the representations of the factor library** to ensure that the generated cluster representations can encapsulate the semantic attributes of shared factors commonly existing in knowledge graph relations. In this study, we adopt an unsupervised clustering algorithm (such as K-means [52]) to verify the robustness of the method. By calculating the similarity between the initial semantic representation of each relation and the representations of cluster centers $\{c_1, c_2, \ldots, c_k\}$, a set of weights can be assigned to each relation. These weights are used to characterize the importance of each shared factor in the new semantic representation:

$$\alpha_i = \frac{\exp\left(\cos\left(r_0, c_i\right)\right)}{\sum_{i=1}^{k} \exp\left(\cos\left(r_0, c_i\right)\right)} \tag{2}$$

where cos denotes cosine similarity. Thus, the reconstructed representation r_c as the SFC representation can be obtained:

$$r_c = \sum_{i=1}^{k} \alpha_i c_i \tag{3}$$

D. LLM Semantic Understanding-Based Relational Semantic Reconstruction Mechanism for Domain-Specific Knowledge Graphs

Domain-specific KGs often contain many semantically homogeneous relations. Although these relations may share similar factor representations, they still require careful distinction. For example, both "purchase" and "invest" involve the concept of acquiring something, but they apply to different domains and are associated with different entity types. To address this issue, FZR [10] proposed an expert-guided semantic reconstruction mechanism to refine the relational semantic space. However, that mechanism requires manual scoring, which is labor-intensive and difficult to scale, and it does not fully exploit similarity scores to distinguish related from unrelated relations. We therefore propose an LLM-expert-guided semantic reconstruction mechanism that leverages LLM embeddings to automatically estimate relation correlations and trains with ranking loss over positive and negative relation pairs. This design enables faster and more accurate reconstruction of the semantic space. The framework of LLM-expert-guided semantic reconstruction is shown in Fig. 3.

1) LLM-expert scoring mechanism: There is a huge number of relations in real-world knowledge graphs, and it is obviously impractical to conduct detailed manual scoring for each pair of relations. To tackle this challenge, this paper proposes a more efficient and accurate method: using an LLM to obtain relation embeddings. Specifically, for a given relation description $r_{\text{description}}$, it is concatenated with the end token <|endoftext|> to form the input sequence, that is:

$$X = r_{\text{description}} \oplus \texttt{<|endoftext|>} \tag{4}$$

The input sequence is tokenized into tokens by the model's tokenizer and then fed into the Transformer layer for computation to generate the hidden state sequence of the model's last layer. Let the number of tokens in the input sequence be m (including <|endoftext|>), then the hidden state sequence of the last layer can be expressed as:

$$H = [h_1, h_2, \ldots, h_{m-1}, h_{\text{EOS}}] \tag{5}$$

where h_i is the hidden state of the i-th token in the last layer; h_{EOS} is the hidden state of <|endoftext|> (i.e., [EOS]) in the last layer. In this paper, this embedding is taken as the final relation embedding, that is:

$$\text{embedding} = \text{LLM}_{\text{last layer}}(X)_{\text{EOS}} \tag{6}$$

For the embeddings of various relations, let r_i and r_j be the embeddings of relation i and relation j respectively, and a similarity function $S : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ is defined to calculate their similarity score.And cosine similarity is used for calculation:

$$S\left(r_i, r_j\right) = \frac{r_i \cdot r_j}{\|r_i\|\|r_j\|} \tag{7}$$

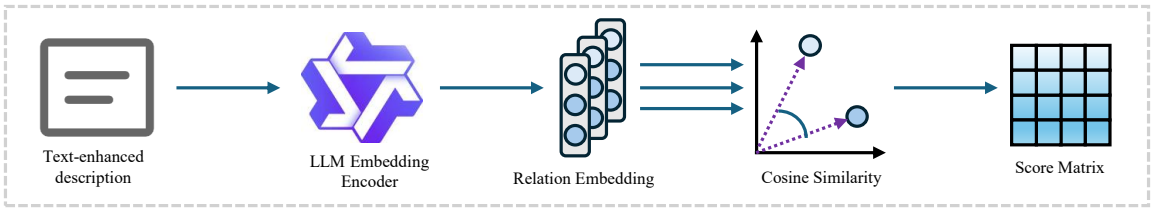


Fig. 3. Framework of LLM-expert-guided semantic reconstruction.

Through the above method, the correlation scores between each pair of relations can be obtained.

2) Scoring-based semantic reconstruction mechanism: Using the above method, the pairwise correlation score matrix S between relations can be obtained. For any relation r_i in the relation set R, the top k relations with the highest similarity are extracted as positive examples r_i^+, and their correlation with r_i is taken as the positive score $score_i^+$; for negative examples, if the correlation score matrix S is dense, the bottom k relations with the lowest similarity are extracted as negative examples r_i^-; if S is sparse, k relations with similarity rankings in the range $[k+1, 2k]$ are extracted as negative examples $r_{i,n}$, and their correlation with r_i divided by the highest correlation among them is taken as the negative score $score_i^-$. This realizes the extraction of positive and negative examples and their corresponding scores. The specific extraction process is as follows:

Define $\text{Rank}[l, r](A)$ as a function to extract the elements from the l-th to the r-th largest in set A, and $\text{Sim}(r_i, r_j)$ as the correlation score between relation r_i and relation r_j:

$$\text{Rank}[l, r](A) = \{x \in A \mid l \leq |\{y \in A \mid y \geq x\}| \leq r\} \quad (8)$$

$$\text{Sim}(r_i, r_j) = S_{i,j} \quad (9)$$

Then the extraction methods of positive and negative examples and their scores are:

$$r_i^+ = \text{Rank}[1, k](S) \quad (10)$$

$$r_i^- = \begin{cases} \text{Rank}[|R| - k, |R|](S) & \text{if } S \text{ is dense,} \\ \text{Rank}[k+1, 2k](S) & \text{otherwise,} \end{cases} \quad (11)$$

$$score_i^+ = \text{Sim}(r_i, r_{i,p}) \quad (12)$$

$$score_i^- = \text{Sim}(r_i, r_{i,n}) \quad (13)$$

This paper aims to learn a mapping network f that maps the original relation r_i to a new semantic space through $f(r_i)$. To better distinguish related relations, ranking loss is used to train this network. Specifically, the mapping network f consists of two multi-layer perceptrons (MLPs), and its expression is:

$$f(r_i) = \text{Tanh}\left(W_2 \cdot \text{Tanh}\left(W_1 \cdot r_i + b_1\right) + b_2\right) \quad (14)$$

where: $W_1 \in \mathbb{R}^{h \times d}$ is the weight matrix of the first linear transformation (h is hidden_dim), and $b_1 \in \mathbb{R}^h$ is the bias of the first layer; $\text{Tanh}(\cdot)$ is the hyperbolic tangent activation function ($\text{Tanh}(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}}$); $W_2 \in \mathbb{R}^{d \times h}$ is the weight matrix of the second linear transformation, and $b_2 \in \mathbb{R}^d$ is the bias of the second layer.

The positive sample weighted distance is defined as:

$$p_{\text{score},i} = \sum_{j=1}^{k} score_{i,j}^+ \cdot \|f(r_i) - f(r_{i,j}^+)\|_2 \quad (15)$$

The negative sample weighted distance is defined as:

$$n_{\text{score},i} = \sum_{j=1}^{k} score_{i,j}^- \cdot \|f(r_i) - f(r_{i,j}^-)\|_2 \quad (16)$$

The loss for a single sample is:

$$\mathcal{L}_i = p_{\text{score},i} + \max\left(0, \text{margin} - n_{\text{score},i}\right) \quad (17)$$

where margin is the margin threshold.

The batch average loss is:

$$\mathcal{L} = \frac{1}{B} \sum_{i=1}^{B} \mathcal{L}_i \quad (18)$$

The learning objective of network f is to adjust the parameters $\theta = \{W_1, b_1, W_2, b_2\}$ through an optimizer (the Adam optimizer is used in the experiment) to minimize the above loss function:

$$\theta^* = \arg\min_{\theta} \mathcal{L}(\theta) \quad (19)$$

The final semantic reconstruction of the relationship r_e is expressed as:

$$r_e = f(r) \quad (20)$$

In this module, the keyword extraction results of LLM-enhanced text (e.g., the 'sport' factor) precisely meet the requirement that "shared factors should cover the core dimensions across relationships", thereby avoiding the one-sidedness of factors extracted from raw text. In the ranking loss module, the sparsity/density characteristics of the similarity matrix generated by the LLM are accurately matched with the "relationship distribution characteristics" of the dataset (e.g., the coherent high-similarity neighborhoods in NELL-ZS and the globally compact relation space in Wiki-ZS), which ensures the rationality of positive and negative sample selection. This interactive design of "intersubjective semantic alignment" maximizes the adaptive value of the LLM to the entire framework.

E. Generative Adversarial Network for Knowledge Transfer Learning Facilitation

With enhanced semantic representations, this section introduces the application of Generative Adversarial Networks (GANs) to facilitate knowledge transfer from seen to unseen relations. The proposed GAN consists of three components: a feature extractor E, a generator G, and a discriminator D.

1) Feature Extractor: Different from traditional knowledge graph embedding methods, our feature extractor aims to learn the cluster structure of both seen and unseen relation facts, maintaining high intra-class and low inter-class similarity. For each seen relation $r \in R_s$, the entity-pair set is defined as:

$$T_r = \{(e_1, e_2) \mid (e_1, r, e_2) \in T\}. \tag{21}$$

Each entity pair (e_1, e_2) is encoded by a fully connected (FC) layer:

$$u_{ep} = \sigma([f_1(v_{e_1}); f_1(v_{e_2})]), \tag{22}$$

$$f_1(v_e) = W_1 v_e + b_1, \tag{23}$$

where σ denotes the $\tanh$ activation function.

To incorporate the one-hop neighbors $\mathcal{N}_e$ of entity e, we compute:

$$u_e = \sigma\left(\frac{1}{|\mathcal{N}_e|} \sum_{(r^n, e^n) \in \mathcal{N}_e} f_2(v_{r^n}, v_{e^n})\right), \tag{24}$$

and the final real embedding of relation r is obtained as:

$$x_r = [u_{e_1}; u_{ep}; u_{e_2}]. \tag{25}$$

The embeddings are optimized by the margin ranking loss:

$$L_\omega = \left[\gamma - score^+ + score^-\right]_+, \tag{26}$$

where $[a]_+ = \max(0, a)$, γ is the margin parameter, and ω represents the parameter set.

2) Adversarial Training: During adversarial training, the feature extractor is fixed. For a seen relation r, the generator G takes its enhanced representation r^* and random noise z to produce a synthetic fact embedding:

$$\hat{x} = G(z, r^*), \qquad z \sim \mathcal{N}(0, I). \tag{27}$$

Before adversarial training, we compute a fixed center for every seen relation from all its training facts:

$$\mu_r = \frac{1}{N_r} \sum_{n=1}^{N_r} x_{r,n}, \tag{28}$$

where $x_{r,n}$ is the feature extracted from the n-th training fact of relation r. The discriminator contains a scalar critic $D_{\text{adv}}(v)$ and an auxiliary classifier over all seen relations. It applies the same transformation to a fact feature and each relation center:

$$h_\phi(v) = \text{LN}(\text{LeakyReLU}(W_\phi v + b_\phi)), \tag{29}$$

$$C_\phi(v, r) = h_\phi(v)^\top h_\phi(\mu_r). \tag{30}$$

Consequently, the auxiliary classifier outputs the score vector $[C_\phi(v, r)]_{r \in \mathcal{R}_s}$. For each positive fact of relation r, a negative fact is constructed by retaining the head entity and replacing the tail entity with a relation-specific candidate that does not form a known true fact. Let x^- denote its extracted feature. For a real or generated feature v, the classification loss implemented by the model is

$$L_{\text{cls}}(v, x^-, r) = \left[\gamma_{\text{cls}} - C_\phi(v, r) + C_\phi(x^-, r)\right]_+. \tag{31}$$

It requires the compatibility of v with its ground-truth relation center to exceed that of the corrupted fact by at least γ_{cls}.

We further use visual-pivot regularization to preserve the class-level structure learned by the feature extractor. For the generated features $\{\hat{x}_{r,m}\}_{m=1}^{M_r}$ of relation r in a mini-batch, we define

$$\hat{\mu}_r = \frac{1}{M_r} \sum_{m=1}^{M_r} \hat{x}_{r,m}, \tag{32}$$

$$L_P = \frac{1}{|\mathcal{R}_b|} \sum_{r \in \mathcal{R}_b} \|\hat{\mu}_r - \mu_r\|_2, \tag{33}$$

where $\mathcal{R}_b$ is the set of seen relations represented in the mini-batch.

The generator minimizes

$$L_G = -\mathbb{E}_{\hat{x}}[D_{\text{adv}}(\hat{x})] + \lambda_1 \mathbb{E}_{(\hat{x}, x^-, r)}[L_{\text{cls}}(\hat{x}, x^-, r)] + \lambda_2 L_P, \tag{34}$$

where the adversarial term aligns the generated and real distributions, while the remaining terms enforce instance-level relation compatibility and class-level center alignment.

For the gradient penalty, we interpolate between paired real and generated embeddings as $\tilde{x} = \epsilon x + (1 - \epsilon)\hat{x}$, where $\epsilon \sim \mathcal{U}(0, 1)$, and define

$$L_{\text{GP}} = \lambda_{\text{GP}} \mathbb{E}_{\tilde{x}} \left(\|\nabla_{\tilde{x}} D_{\text{adv}}(\tilde{x})\|_2 - 1\right)^2. \tag{35}$$

The discriminator is then optimized by minimizing

$$\begin{aligned} L_D = {} & \mathbb{E}_{\hat{x}}[D_{\text{adv}}(\hat{x})] - \mathbb{E}_x[D_{\text{adv}}(x)] + L_{\text{GP}} \\ & + \lambda_3 \mathbb{E}_{(\hat{x}, x^-, r)}[L_{\text{cls}}(\hat{x}, x^-, r)] \\ & + \lambda_4 \mathbb{E}_{(x, x^-, r)}[L_{\text{cls}}(x, x^-, r)]. \end{aligned} \tag{36}$$

During a discriminator update, $\hat{x}$ is detached from the generator; during a generator update, the discriminator parameters are fixed. This alternating optimization transfers the feature distribution of seen relations to unseen relations while retaining relation-specific discrimination.

All expectations in Eqs. (34)–(36) are evaluated as relation-balanced empirical averages within a mini-batch. The classification and pivot terms impose complementary constraints: L_{cls} requires every real or generated instance to score above a corrupted fact for the same relation, whereas L_P prevents the generated class center from drifting away from the pre-computed real center. Using both terms avoids two degenerate solutions—matching only the class center while producing ambiguous instances, or separating individual instances while displacing the entire generated class. These constraints are learned exclusively from seen relations. At test time, the generator synthesizes multiple features from an unseen relation representation; candidate facts are ranked by their mean cosine similarity to these generated features, without using any unseen-relation triples.

IV. EXPERIMENTS

A. Experimental Setup

1) Datasets: Due to the limited number of studies on zero-shot tasks, we conduct experiments on the proposed model and baseline models using two public benchmark datasets proposed in [6], namely NELL-ZS and Wiki-ZS. These two datasets are constructed based on the NELL knowledge base [53] and Wikidata, respectively. Each relation in both datasets is accompanied by a text description and the type information of the corresponding entities. In each knowledge graph, seen relations are divided into the training set and validation set, while unseen relations only exist in the test set; relevant triples are also assigned to the corresponding datasets according to this rule.

Since previous studies [6], [7] did not elaborate on the dataset partitioning method, to comply with the more general zero-shot relation learning (ZSRL) setup and ensure the fairness of comparison, we adopt a three-fold random partitioning method to divide the training set/validation set/test set (Train/Dev/Test), and then conduct three-fold cross-validation based on these partitioning results. It is important to emphasize that Reference [7] adopts a conservative partitioning method where test relations are fixed and only the training set and validation set are reallocated. However, we believe that the performance of a model on re-partitioned datasets can better reflect its adaptability to zero-shot scenarios and more effectively verify its generalization ability. Therefore, we re-conduct experiments on all baseline models using the re-partitioned datasets. Table I shows the detailed statistical information of these two datasets.

TABLE I
STATISTICAL INFORMATION OF ZERO-SHOT RELATION LEARNING (ZSRL) DATASETS.

Dataset	#Ent	#Triples	#Train/Dev/Test
NELL-ZS	65567	188392	113/ 20/ 48
Wiki-ZS	605812	724967	469/ 20/ 48

2) Evaluation Metrics: In the test phase, given the head entity and relation in a test triple, the model needs to rank all tail entities in the candidate list. Therefore, following the setup in Reference [6], we adopt the commonly used evaluation metrics in the knowledge graph completion (KGC) task: Mean Reciprocal Ranking (MRR), Hits@10 (Hit Ratio at top 10), Hits@5 (Hit Ratio at top 5), and Hits@1 (Hit Ratio at top 1). Among them, MRR represents the average of the reciprocal ranks of all correct entities; Hits@k represents the proportion of test samples where the correct entity is among the top k candidates.

3) Baseline Models: We compare the proposed zero-shot learning (ZSL) method with the current mainstream baseline models that are not assisted by LLMs, including three types of traditional Knowledge Graph Embedding (KGE) methods: TransE [54], DistMult [55], and ComplEx [56], as well as three types of non-LLM-assisted generative zero-shot learning methods: ZSGAN [6], OntoZSL [7], and FZR [10]. Since existing zero-shot knowledge graph completion methods rarely incorporate LLMs directly, we design two LLM-based baselines using the generative and embedding models described below.

Although traditional knowledge graph embedding methods are not designed for zero-shot learning tasks, including them in comparative experiments is crucial for establishing baseline performance and conducting comparative analysis. Inspired by ZSGAN [6], we propose three zero-shot variants of knowledge graph embedding methods: ZS-TransE, ZS-DistMult, and ZS-ComplEx. Unlike traditional methods that use randomly initialized vectors to represent relations in triples, these variants adopt a feedforward network similar to the generator architecture to obtain relation representations. This network takes the text embedding of the relation as input and is fine-tuned together with entity embeddings during training. In the test phase, the text embedding of the unseen relation is input into this network to generate the relation representation, and then the scoring function of the original method is used to predict the factual triples.

Among the generative zero-shot learning baseline models, ZSGAN [6] obtains relation embeddings through the text description of relations; OntoZSL [7] obtains relation embeddings by jointly encoding text descriptions and ontological structure knowledge, and both then use Generative Adversarial Networks (GAN) to generate corresponding relation representations; FZR [10] introduces shared factors and an expert scoring system on the basis of ZSGAN [6] to obtain relation embeddings. To ensure generality, two representative knowledge graph embedding models are selected for the feature extractor in the experiments: TransE [54] and DistMult [55].

4) Implementation Details: On the NELL-ZS dataset, a pre-trained 100-dimensional knowledge graph embedding feature extractor is used to obtain 200-dimensional relation embeddings; on the Wiki-ZS dataset, following the prior ZSRL setup [6], [10], a pre-trained 50-dimensional knowledge graph embedding feature extractor is used to obtain 100-dimensional relation embeddings.

The training of the feature encoder and generative model uses the Adam optimizer [57] for parameter updates, with the margin parameters γ and γ_{cls} both set to 10. For the feature encoder, the upper limit of the number of neighbors is set to 50, the number of triples considered in each training step k is fixed at 30, and the learning rate is set to 5×10^{-4}; for the generative model, the learning rate is set to 1×10^{-3}, and the parameters β_1 and β_2 are set to 0.5 and 0.9, respectively. The discriminator is updated 5 times for every 1 update of the generator. The dimension of the random vector z is 15. Each adversarial mini-batch contains two seen relations and 1,024 real/generated samples per relation; at test time, 20 embeddings are synthesized for each unseen relation. The classification loss weights λ_1, λ_3, and λ_4 are set to 1, 0.5, and 0.5, respectively; the pivot regularization weight λ_2 is set to 3, and the gradient-penalty weight λ_{GP} is set to 10.

On the NELL-ZS dataset, the number of hidden units in the generator is set to 250, outputting 200-dimensional relation embeddings; the number of hidden units in the discriminator is set to 200, with one scalar critic output and a $|\mathcal{R}_s|$-dimensional

auxiliary classification output. On the Wiki-ZS dataset, the number of hidden units in the generator is set to 250, and the number of hidden units in the discriminator is set to 100.

For the word embeddings of relation texts, we use the publicly available 300-dimensional word embeddings (from glove.6B.300d.txt), and obtain denoised text embeddings through TF-IDF (Term Frequency-Inverse Document Frequency). For the text description of each relation, 6 keywords are extracted as factors; for the corpus composed of all shared factors, the number of factor clusters is set to 10.

For text enhancement, we use the open-source Qwen3-8B model; for embedding extraction, we use the open-source Qwen3-embedding-8B model. The generation temperature is set to 0.001, the maximum number of tokens is set to 8192, and the embedding dimension is 2560. All experiments can be completed on a single 4090 (24G) consumer-grade graphics card.

Details of other baseline models can be found in the corresponding papers. To compare LAZR with direct LLM-based alternatives under the same model family, we additionally adopt two LLM baselines: (1) a prompt-based baseline that asks the generative LLM to re-rank candidate entities for each query $h, r, ?$; and (2) an embedding-based baseline that extracts embeddings for all relations and entities using the LLM embedding model, followed by TransE-style similarity ranking.

Since directly using LLMs for zero-shot relation learning is resource-intensive and time-consuming, this experiment is conducted on a smaller dataset (NELL-ZS-S). Detailed information of this dataset is shown in Table II.

TABLE II
STATISTICAL INFORMATION OF THE DATASET FOR LARGE-MODEL BASELINES.

Dataset	#Ent	#Triples	#Train/Dev/Test
NELL-ZS-S	13252	17603	113/ 20/ 48

B. Main Results

We conduct zero-shot learning (ZSL) experiments on NELL-ZS and Wiki-ZS, with the results summarized in Table III. We also evaluate LLM-based baselines on the smaller NELL-ZS-S dataset, as reported in Table IV. LAZR, instantiated with either TransE or DistMult embeddings, consistently outperforms traditional knowledge graph embedding (KGE) methods, generative ZSRL methods, and direct LLM baselines. These results validate the effectiveness of the proposed factor-based zero-shot relational learning framework. The main observations are as follows:

1) **Generative methods outperform traditional KGE and direct LLM methods**: All generative methods substantially outperform traditional KGE methods and direct LLM baselines. This supports our choice of an LLM-enhanced generative framework rather than pure LLM prediction. To reduce the influence of different pretrained KGE embeddings, we evaluate each generative method with both TransE and DistMult backbones. The consistent gains under both settings confirm the advantage of generative models for knowledge transfer.
2) **LAZR achieves the best performance among generative methods**: LAZR consistently outperforms the other generative methods with both TransE and DistMult embeddings. The results suggest that LLM-enhanced shared factors create stronger connections between seen and unseen relations, while LLM-derived scoring improves discrimination between semantically similar and dissimilar relations. On NELL-ZS, LAZR improves over FZR by 2.4 percentage points in Hit@1 and 2.9 percentage points in Hit@5 with TransE, and by 2.5 and 1.2 percentage points, respectively, with DistMult. Compared with the embedding-based LLM baseline on NELL-ZS-S, LAZR with TransE improves Hit@1 by 34.4 percentage points and Hit@5 by 33.2 percentage points.
3) **Performance differences of the LAZR model across different datasets**: Compared with the performance improvement on the NELL-ZS dataset, LAZR achieves a smaller improvement over FZR on the Wiki-ZS dataset, especially under the DistMult embedding setting (Hit@1 improves by only 0.4 percentage points versus 2.5 on NELL-ZS). Under the TransE setting, the improvement remains comparable (2.3 versus 2.4 on NELL-ZS). As shown by the structural diagnostics, Wiki-ZS has high and relatively homogeneous relation similarity. Its already compact semantic space leaves less room for factor composition to introduce additional discriminative structure. In addition, on Wiki-ZS with DistMult, the performance gap between LAZR and ZSGAN narrows to merely 0.2 Hit@1 points (28.5 versus 28.3), suggesting a ceiling effect: when relation representations are already globally compact, semantic enhancement yields smaller marginal gains.

From the perspective of performance gains brought by model optimization, the LAZR method proposed in this paper features simplicity and efficiency, with specific performance improvement data shown in Table III. Taking the NELL-ZS dataset as an example, compared with the FZR model, the LAZR model achieves a 2.6 percentage point increase in the MRR (Mean Reciprocal Rank) metric, a 2.4 percentage point increase in the Hit@1 metric, a 2.9 percentage point increase in the Hit@5 metric, and a 2.4 percentage point increase in the Hit@10 metric; moreover, under both settings of "combining TransE embeddings" and "combining DistMult embeddings", all metrics of the LAZR model reach the optimal level in the experiment.

C. Ablation Experiments

To further verify the effectiveness of each component constituting the enhanced relation representation, we carefully designed ablation experiments. Specifically, the experiments include the following model variants:

- **LAZR**: The complete proposed model (including all core components);

TABLE III
ZERO-SHOT LINK PREDICTION RESULTS (%) ON UNSEEN RELATIONS. FZR AND LAZR ARE REPORTED AS MEAN±STD OVER 5 RUNS; */** DENOTE LAZR OVER FZR AT $p < 0.05/p < 0.01$ BY PAIRED t-TEST. BOLD INDICATES THE BEST PERFORMANCE.

Method type	Model	NELL-ZS				Wiki-ZS			
		MRR	Hit@1	Hit@5	Hit@10	MRR	Hit@1	Hit@5	Hit@10
Traditional KGE-based method	ZS-TransE	21.2	12.7	30.3	37.1	10.2	4.6	16.4	19.6
	ZS-DistMult	40.5	35.0	47.6	51.8	21.5	16.4	26.9	30.2
	ZS-ComplEx	41.5	35.6	48.5	51.8	22.5	18.2	27.3	29.9
Generative-based method with TransE	ZSGAN	44.3	32.4	57.8	65.8	26.7	18.9	34.2	41.8
	OntoZSL	48.2	36.9	61.2	68.4	26.3	19.3	32.6	39.3
	FZR	47.9±1.1	35.9±1.4	61.5±0.7	68.8±0.4	27.8±1.0	21.0±0.9	34.1±1.2	40.5±1.4
	LAZR(ours)	**50.5±0.4***	**38.3±0.4***	**64.4±0.6****	**71.2±0.6****	**30.3±0.3****	**23.4±0.3****	**36.9±0.4****	**43.8±0.5****
Generative-based method with DistMult	ZSGAN	49.7	41.6	58.8	65.5	33.8	28.3	38.8	44.1
	OntoZSL	50.9	42.7	60.2	66.2	34.2	27.8	40.7	45.9
	FZR	53.5±0.7	45.8±0.7	62.1±0.7	67.9±0.9	34.3±0.1	28.1±0.2	40.5±0.1	45.9±0.2
	LAZR(ours)	**55.5±0.9****	**48.3±1.0****	**63.3±0.9***	**68.7±0.3**	**34.8±0.1****	**28.5±0.1***	**41.0±0.3***	**46.2±0.3**

TABLE IV
RESULTS (%) OF ZERO-SHOT LINK PREDICTION WITH UNSEEN RELATIONS USING LLM-BASED BASELINES. BOLD INDICATES THE BEST PERFORMANCE.

Method type	Model	NELL-ZS-S			
		MRR	Hit@1	Hit@5	Hit@10
Based on generative large language models	Llama3-8B	0.1	0	0	0
	Qwen3-4B	0.1	0	0	0
	Qwen3-8B	12.3	9.5	13.3	13.3
Based on embedding large language models	Qwen-embedding-4B	26.5	15.2	36.3	49.6
	Qwen-embedding-8B	28.5	17.4	38.3	51.2
LAZR	Qwen3-8B-DistMult	58.7	49.2	68.7	77.3
	Qwen3-8B-TransE	**61.2**	**51.8**	**71.5**	**78.6**

TABLE V
ABLATION RESULTS (%) ON NELL-ZS AND WIKI-ZS WHEN ENHANCED TEXT REPRESENTATIONS ("-ETR") AND LLM-EXPERT-GUIDED SEMANTIC REPRESENTATIONS ("-LEG") ARE REMOVED.

Method type	Model	NELL-ZS				Wiki-ZS			
		MRR	Hit@1	Hit@5	Hit@10	MRR	Hit@1	Hit@5	Hit@10
Generative-based method with pre-trained TransE	LAZR(-etr)	49.8	38.1	62.9	69.4	29.8	22.8	36.6	43.3
	LAZR(-leg)	49.7	37.9	62.6	69.4	30.1	23.1	36.7	43.8
	LAZR(-etr&leg)	48.4	36.6	61.6	68.4	29.5	22.5	36.1	42.8
Generative-based method with pre-trained DistMult	LAZR(-etr)	53.5	46.1	61.3	67.6	34.2	28.1	40.3	45.6
	LAZR(-leg)	53.2	45.7	61.2	67.8	34.6	28.3	40.8	45.9
	LAZR(-etr&leg)	52.5	44.7	61.2	67.0	34.1	27.7	40.4	45.9

- **LAZR-etr**: Removes the "LLM-enhanced text representations" (etr is the abbreviation of Enhanced Text Representations);
- **LAZR-leg**: A variant of LAZR that removes the "LLM-expert-guided semantic reconstruction" (leg is the abbreviation of LLM-expert-guided);
- **LAZR-etr&leg**: A variant of LAZR that removes both "LLM-enhanced text representations" and "LLM-expert-guided semantic reconstruction" (this variant is essentially functionally equivalent to the FZR model).

We trained the above models on the NELL-ZS and Wiki-ZS datasets. The training of feature extractors adopts pre-trained TransE and DistMult knowledge graph embeddings (KG Embedding). The test results of each ablation variant are shown in Table V.

The results confirm that both "LLM-enhanced text representations" and "LLM-expert-guided semantic reconstruction" components play important roles in the LAZR model: the complete LAZR model integrating these two components consistently outperforms its various ablation variants. The specific analysis is as follows:

1) **LAZR (-etr) variant**: There is a slight decline in performance. This indicates that although "LLM-enhanced text representations" can improve the generalization ability of the model, it is not the only key factor for the performance improvement of the LAZR model;
2) **LAZR (-leg) variant**: After removing "LLM-expert-guided semantic reconstruction", the model performance also slightly decreases. It also indicates that this module improves the model performance, but it is not the only key factor for the performance improvement of the

LAZR model;

3) **LAZR (-etr&leg) variant**: When both components are missing, the model performance declines the most significantly. This further proves that the synergistic effect between "LLM-enhanced text representations" and "LLM-expert-guided semantic reconstruction" is the key to the LAZR model's ability to effectively complete zero-shot learning tasks.

D. Necessity of Contrastive Semantic Reconstruction

The raw relation embeddings $\mathbf{r}$ extracted from GloVe-weighted factor vectors (Section 3.3) capture coarse semantic associations, yet they suffer from two critical limitations that hinder zero-shot transfer. First, keyword-based factor extraction inevitably introduces noise: high-frequency shared factors such as "team" appear across multiple semantically distinct relations (e.g., *teammember*, *teamcoach*, *teamwontrophy*), causing their raw embeddings to be superficially close in the GloVe space. Second, cosine similarity computed directly on these embeddings does not faithfully reflect the nuanced relational semantics required for knowledge transfer, because GloVe vectors are optimized for word-level co-occurrence rather than relation-level meaning.

To remedy this, LAZR introduces a *scoring-based semantic reconstruction* module (Section 3.4.2). Concretely, a two-layer MLP $f(\cdot)$ with Tanh activation maps the raw embeddings into a new semantic space, and is optimized with a weighted ranking loss:

$$\mathcal{L}_i = p_{\text{score},i} + \max\big(0,\ \gamma - n_{\text{score},i}\big), \tag{37}$$

where $p_{\text{score},i}$ and $n_{\text{score},i}$ denote the weighted distances from relation i to its top-k positive and bottom-k negative examples, respectively, guided by the LLM-generated correlation score matrix $\mathbf{S}$. The loss explicitly pulls semantically homogeneous relations closer while pushing heterogeneous ones farther apart.

a) Empirical evidence.: Fig. 4 visualizes the NELL relation embeddings before and after reconstruction via t-SNE followed by K-Means clustering ($K = 10$) on the 2D coordinates. In the **before** panel, relations are scattered without clear structure (average 5-NN distance: 1.24, silhouette score: 0.426), indicating that the raw GloVe space entangles distinct relations. After reconstruction (**after** panel), the same relations form visibly tighter clusters (average 5-NN distance: 0.74, silhouette score: 0.585), confirming that the ranking loss successfully compresses intra-class variance and enlarges inter-class margins.

The quantitative separation is equally striking. Before reconstruction, the Euclidean distance between the positive pair *teammember* and *teamcoach* is 0.27, whereas the negative pair *teammember* and *cityliesonriver* is only marginally larger at 0.34. After reconstruction, the positive distance collapses to 0.00 and the negative distance explodes to 13.36—a relative increase of more than $37\times$. Such dramatic separation is essential for the subsequent GAN module: a generator conditioned on well-separated relation embeddings can reliably synthesize distinct entity-pair features for unseen relations, thereby improving zero-shot transfer. Without this reconstruction step, the generator would receive nearly identical inputs for semantically different relations, leading to mode collapse and poor generalization.

E. Cross-Dataset Boundary Analysis of Relation Semantic Enhancement

To test whether relation semantic enhancement generalizes beyond the commonly used NELL and Wiki benchmarks, we further evaluate on CoDEx-L and FB15k-237. These datasets are useful stress tests because their relations are more attribute-like and less naturally clustered than the conceptual relations in NELL/Wiki. Additional details of the CoDEx-L zero-shot split are provided in the supplementary material.

Table VI reports the cross-dataset results. We use the same notation as Table V: `up` denotes factor clustering, `down` denotes contrastive learning, and **LAZR-recon** denotes the reconstruction-only variant. All values are percentages. Unlike on NELL-ZS and Wiki-ZS, where LAZR improves over ZSGAN/FZR (Table III), LAZR underperforms ZSGAN on both CoDEx-L and FB15k-237, revealing a boundary of relation semantic enhancement.

TABLE VI
MAIN RESULTS ON CODEX-L AND FB15K-237 (TEST SET, %).

Dataset	Method	MRR	Hit@1	Hit@5	Hit@10
CoDEx-L	ZSGAN-TransE	**27.5**	**17.7**	**37.5**	**47.7**
	ZSGAN-DistMult	12.0	3.5	23.5	28.1
	LAZR-TransE	12.1	6.6	16.0	21.7
	LAZR-DistMult	5.0	1.5	6.3	10.8
FB15k-237	ZSGAN-TransE	45.4	40.5	49.7	53.0
	LAZR (full)	42.2	37.6	46.4	50.0
	LAZR-up (Factor Clustering)	40.9	36.4	44.8	48.4
	LAZR-down (Contrastive Learning)	41.8	37.1	45.8	49.1
	LAZR-recon (Reconstruct Only)	**45.7**	**41.0**	**49.9**	**53.0**

On CoDEx-L, LAZR-TransE achieves only 6.6 Hit@1, far behind ZSGAN-TransE (17.7); the DistMult variant is even weaker, suggesting that semantic enhancement cannot compensate for embedding-model mismatch on asymmetric relations. On FB15k-237, the reconstruction-only variant reaches the best Hit@1 (41.0), slightly above ZSGAN (40.5), while adding factor clustering or contrastive learning reduces performance.

To isolate the source of degradation, Table VII consolidates the TransE ablations across NELL-ZS, Wiki-ZS, and FB15k-237. This side-by-side comparison is important because it tests whether the same enhancement module has a stable effect across KG types.

The detailed ablation exposes a sign reversal. Relative to reconstruction only, factor clustering changes Hit@1 by $+1.3$ on NELL-ZS and $+0.6$ on Wiki-ZS, but by -4.6 on FB15k-237; contrastive reconstruction changes it by $+1.5$, $+0.3$, and -3.9, respectively. Combining both modules yields the largest gains on NELL-ZS/Wiki-ZS ($+1.7$/$+0.9$ Hit@1), yet still loses 3.4 points on FB15k-237. Thus, reconstruction itself is not the failure point: the degradation begins when neighborhood-based enhancement is applied to an attribute-like, long-tail relation space. In conceptual KGs, shared factors reinforce coherent

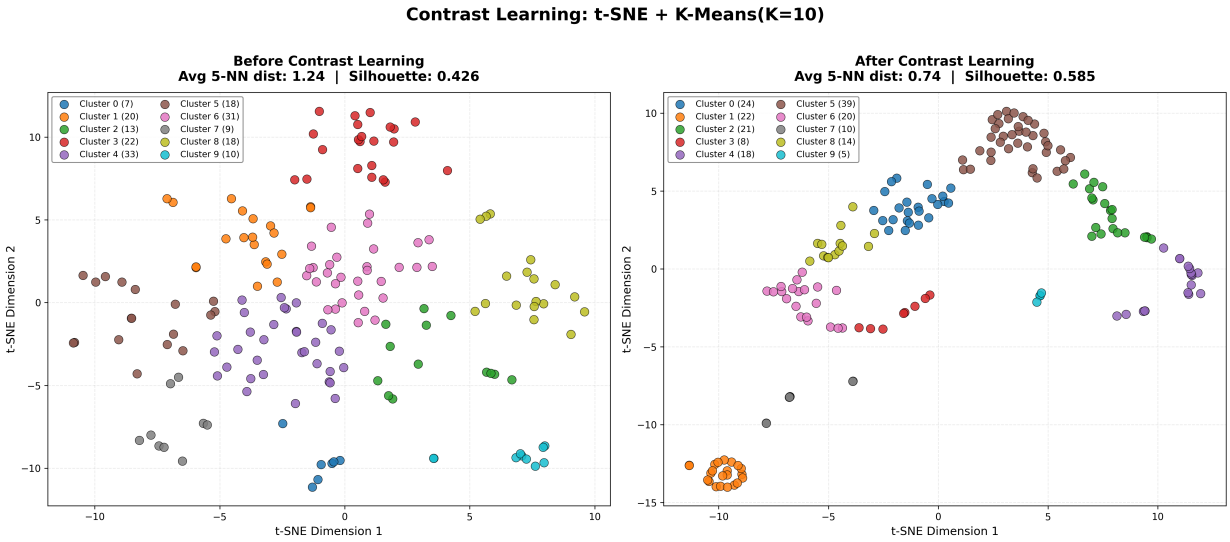


Fig. 4. t-SNE visualization of NELL relation embeddings before and after contrastive semantic reconstruction. Colors denote K-Means clusters ($K = 10$); sub-captions report average 5-NN distance and silhouette score.

TABLE VII
DETAILED MODULE-LEVEL ABLATION ACROSS DATASETS (TRANSE, TEST SET, %). "RECONSTRUCTION ONLY" REMOVES BOTH ENHANCEMENT MODULES; THE NEXT TWO ROWS ADD ONE MODULE AT A TIME. BEST RESULTS WITHIN EACH DATASET ARE BOLD.

Configuration	NELL-ZS				Wiki-ZS				FB15k-237			
	MRR	Hit@1	Hit@5	Hit@10	MRR	Hit@1	Hit@5	Hit@10	MRR	Hit@1	Hit@5	Hit@10
Reconstruction only	48.4	36.6	61.6	68.4	29.5	22.5	36.1	42.8	**45.7**	**41.0**	**49.9**	**53.0**
+ Factor clustering	49.7	37.9	62.6	69.4	30.1	23.1	36.7	**43.8**	40.9	36.4	44.8	48.4
+ Contrastive reconstruction	49.8	38.1	62.9	69.4	29.8	22.8	36.6	43.3	41.8	37.1	45.8	49.1
+ Both (full LAZR)	**50.5**	**38.3**	**64.4**	**71.2**	**30.3**	**23.4**	**36.9**	**43.8**	42.2	37.6	46.4	50.0

neighborhoods; in FB15k-237, heterogeneous compound paths make those neighborhoods noisy, so clustering mixes unrelated signals and contrastive learning pulls unreliable positive pairs together. The structural analyses below test this mechanism directly.

F. Extended-Dataset Protocol and Pair Quality

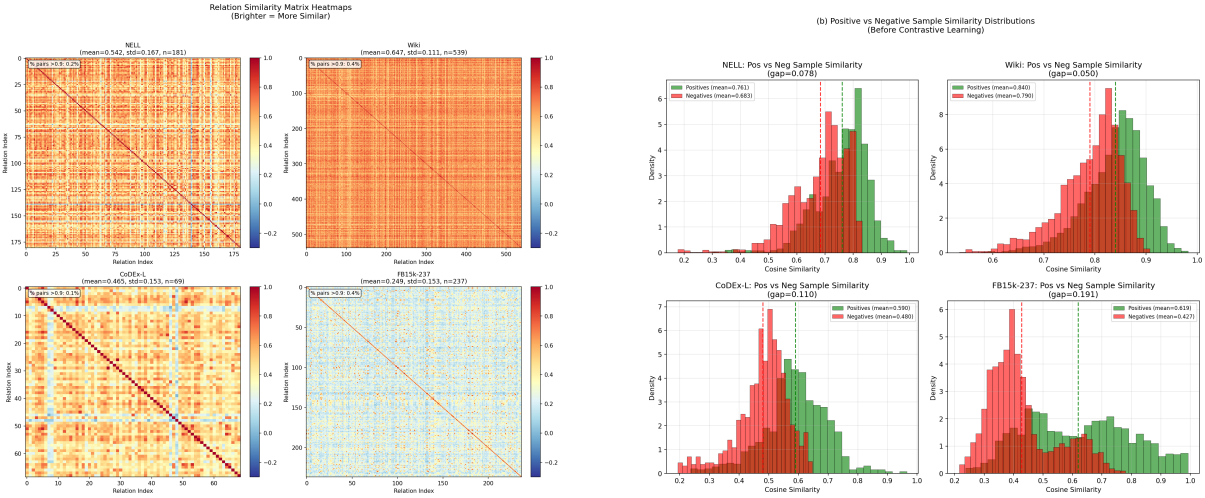
CoDEx-L contains 77,951 entities and 612,437 triples; because it has no standard zero-shot split, we form disjoint 35/10/18 train/dev/test relation sets after removing six low-frequency training relations. FB15k-237 contains 14,541 entities and 310,116 triples with a 169/20/48 relation split. For both datasets, relations are strictly partitioned across the train, development, and test sets, and all triples associated with a relation remain exclusively in the subset assigned to that relation; therefore, neither relations nor triples can cross split boundaries. Because FB15k-237 lacks natural-language relation descriptions, we use Qwen3-8B to produce one fixed description for each relation from its relation name and representative triples. Description generation is performed once after relation-level partitioning, and the resulting descriptions are fixed before downstream training and evaluation. Neither dataset provides the ontology required by ontology-aware baselines, so the cross-dataset study compares text-conditioned methods under the same available supervision; additional details of the CoDEx-L construction protocol are provided in the supplementary material.

TABLE VIII
NEGATIVE SAMPLING STRATEGIES ON CODEX-L (%).

Strategy	MRR	Hit@1	Hit@5	Hit@10
Tail selection	8.0	3.1	11.2	16.7
Rank-window	**12.1**	**6.6**	**16.0**	**21.7**

Fig. 5 examines relation-pair quality before reconstruction. NELL displays clearer block structure, and its positive pairs concentrate at higher similarities. CoDEx-L and FB15k-237 instead exhibit diffuse matrices and greater overlap between positive and negative distributions. Their nearest neighbors are therefore less reliable as positives: pulling such pairs together can inject noise instead of recovering a transferable semantic topology.

We further test whether negative selection alone explains the CoDEx-L degradation. Table VIII shows that rank-window negatives outperform tail selection on every metric because they provide harder comparisons. However, the best Hit@1 remains 6.6, well below the 17.7 of ZSGAN-TransE in Table VI. The result rules out a single sampling choice as the primary cause and supports the structural-boundary interpretation.



(a) Relation-similarity matrices.

(b) Positive and negative pair distributions.

Fig. 5. Quality of the relation neighborhoods used by factor composition and contrastive reconstruction. Conceptual KGs provide more coherent positive pairs than attribute-like KGs.

G. Sensitivity and Structural Diagnostics

Table IX examines the two hyperparameters that directly control semantic reconstruction. For the number of positive relations k, performance peaks at $k = 6$; larger neighborhoods introduce less-related positives and weaken the ranking signal. For the number of shared-factor clusters n, $n = 10$ gives the best balance between factor overlap and redundant clusters. The narrow performance range around both optima indicates that LAZR is not highly sensitive to either choice. We also tested weighted addition and weighted concatenation of r_c and r_e; neither exceeded direct concatenation, suggesting that the downstream GAN can learn the required balance without manually assigned weights.

TABLE IX
SENSITIVITY TO THE POSITIVE-SET SIZE k AND NUMBER OF SHARED-FACTOR CLUSTERS n ON NELL-ZS (%).

Parameter	Value	MRR	Hit@1	Hit@5	Hit@10
k	3	50.1	38.3	63.1	69.9
	6	**50.2**	**38.3**	**63.5**	**70.1**
	9	49.7	37.8	63.2	69.8
	12	49.3	37.2	62.9	69.4
	15	49.4	37.3	62.8	69.3
n	5	49.0	37.0	62.6	68.9
	10	**50.2**	**38.3**	**63.5**	**70.1**
	15	49.6	37.5	63.1	69.5
	20	49.3	37.1	62.6	69.4

Fig. 6a provides further evidence for the cross-dataset boundary. NELL and Wiki have substantially closer semantic neighbors, whereas CoDEx-L and FB15k-237 contain many isolated relations. As shown in Fig. 6d, factor clustering also inflates pairwise similarity most strongly on the latter datasets, causing heterogeneous relations to lose discriminative information. This pattern persists as the number of clusters changes.

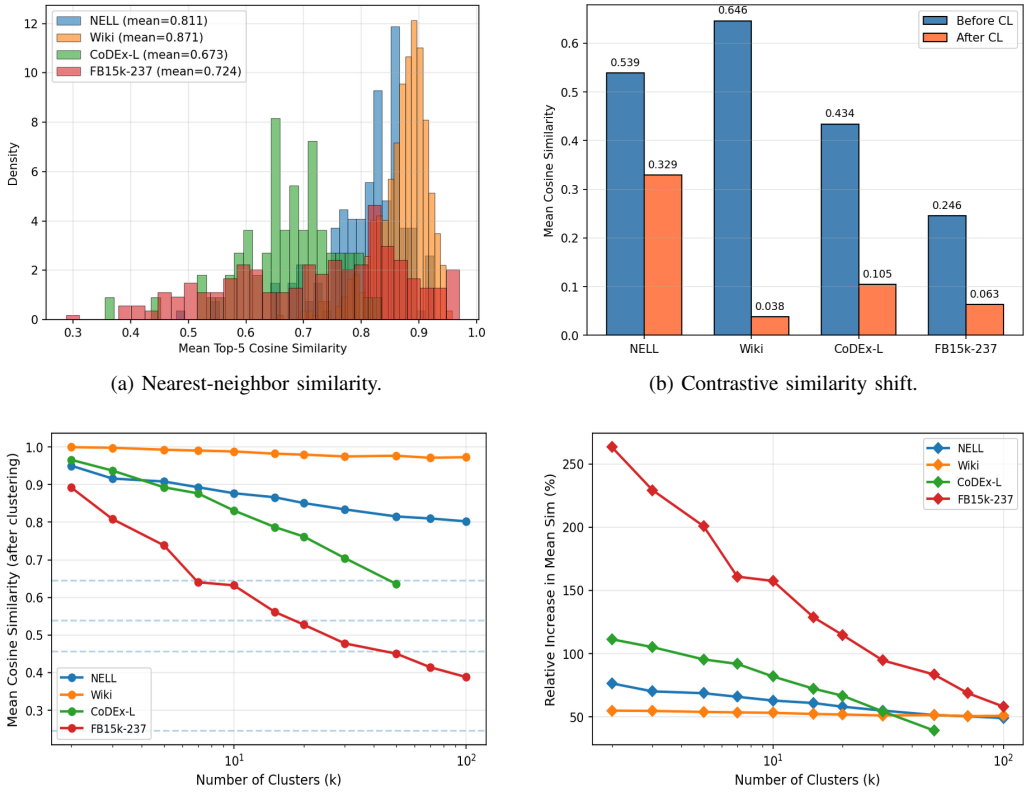
Quantitatively, Fig. 6a shows that the mean top-5 nearest-neighbor similarities are 0.811 and 0.871 on NELL and Wiki, compared with 0.673 and 0.724 on CoDEx-L and FB15k-237. The latter two distributions are also much broader, confirming that many relations lack a stable semantic neighborhood. Compressing such isolated relations into shared factors is therefore more likely to erase relation-specific information than to transfer reusable structure.

The cluster sweeps in Figs. 6c and 6d further rule out a single poorly chosen hyperparameter. The absolute-value curves in Fig. 6c plot the post-clustering mean cosine similarities, with dashed lines marking the original pre-clustering values. At 10 clusters, factor composition increases mean similarity by 82% on CoDEx-L and 157% on FB15k-237, versus 63% on NELL and 53% on Wiki. Even at 100 clusters, the mean similarity on FB15k-237 remains 0.388, far above its original value of 0.246. This persistence follows from the soft-assignment mechanism: every relation is represented as a weighted mixture of all cluster centers, so semantically distinct signals continue to mix even as the cluster count grows. Increasing the cluster count therefore reduces but does not remove the homogenization induced by factor composition.

Finally, Fig. 6b quantifies the global effect of contrastive reconstruction. Mean relation similarity falls from 0.539 to 0.329 on NELL, from 0.646 to 0.038 on Wiki, from 0.434 to 0.105 on CoDEx-L, and from 0.246 to 0.063 on FB15k-237. This strong polarization is useful when selected positives reflect coherent neighborhoods, but becomes risky when they are only superficial nearest neighbors. Before reconstruction, the mean positive-pair similarities are 0.761 and 0.840 on NELL and Wiki, but only 0.590 and 0.619 on CoDEx-L and FB15k-237. Pulling these lower-quality pairs together injects noise rather than recovering transferable topology. Together with Table VII, these results explain why the reconstruction-only variant can outperform the full pipeline on attribute-like KGs.

V. CONCLUSION

This paper presented LAZR, an LLM-augmented framework for Zero-Shot Relational Learning (ZSRL). LAZR leverages



(a) Nearest-neighbor similarity.

(b) Contrastive similarity shift.

(c) Absolute mean similarity after factor clustering.

(d) Relative similarity inflation.

Fig. 6. Structural diagnostics across four datasets. Dashed lines in (c) denote the original pre-clustering means. Coherent semantic neighborhoods support factor transfer, whereas diffuse relation spaces suffer greater similarity inflation and less reliable contrastive reconstruction.

LLM-enhanced relation descriptions and LLM-derived semantic signals to improve knowledge transfer between seen and unseen relations.

Specifically, LAZR constructs enhanced relation representations through three complementary steps. First, it uses LLMs to generate richer and more accurate semantic descriptions for target relations, providing a stronger textual basis for relation representation. Second, it extracts shared relational factors from prior relation semantics and combines these factors to strengthen cross-relation knowledge transfer. Third, it uses LLM-derived similarity scores to guide semantic reconstruction with ranking-based supervision, yielding relation representations that are more sensitive to subtle distinctions among semantically similar relations.

Built on these enhanced relation representations, LAZR employs a generative adversarial framework to synthesize embeddings for unseen relations and facilitate knowledge transfer from seen relations. Experiments on real-world ZSRL benchmarks demonstrate that LAZR consistently outperforms mainstream baselines, and ablation studies confirm the contribution of its semantic enrichment and reconstruction modules.

Our cross-dataset experiments on CoDEx-L and FB15k-237 further show that relation semantic enhancement is not universally beneficial. Factor clustering and contrastive learning improve performance on conceptual KGs such as NELL and Wiki, but may degrade performance on attribute-based or long-tail KGs such as CoDEx-L and FB15k-237. This finding identifies a scalability boundary for relation semantic enhancement techniques and suggests that future methods should employ **adaptive module selection** conditioned on the KG's intrinsic structural properties, rather than applying uniform enhancement pipelines across all graph types. Future work can further explore LLM-assisted information integration, such as combining multimodal semantic information with dynamic knowledge graphs, to improve generalization in more complex ZSRL scenarios.

ACKNOWLEDGMENTS

This research was partially supported by the National Natural Science Foundation of China (No. 62406303, 62525606 and U25B2072) and Anhui Province Science and Technology Innovation Project (202423k09020010).

REFERENCES

[1] Xuesong Wang, Chen Chen, Yuhu Cheng, Z Jane Wang, "Zero-shot image classification based on deep feature extraction", *IEEE Transactions on Cognitive and Developmental Systems*, vol. 10, no. 2, pp. 432–444, 2016.

[2] Omer Levy, Minjoon Seo, Eunsol Choi, Luke Zettlemoyer, "Zero-shot relation extraction via reading comprehension", *arXiv preprint arXiv:1706.04115*, 2017.

[3] Jiahui Wang, Likang Wu, Hongke Zhao, Ning Jia, "Multi-view enhanced zero-shot node classification", *Information Processing & Management*, vol. 60, no. 6, pp. 103479, 2023.

[4] Likang Wu, Junji Jiang, Hongke Zhao, Hao Wang, Defu Lian, Mengdi Zhang, Enhong Chen, "KMF: knowledge-aware multi-faceted representation learning for zero-shot node classification", *arXiv preprint arXiv:2308.08563*, 2023.

[5] Wei Wang, Vincent W Zheng, Han Yu, Chunyan Miao, "A survey of zero-shot learning: Settings, methods, and applications", *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 10, no. 2, pp. 1–37, 2019.

[6] Pengda Qin, Xin Wang, Wenhu Chen, Chunyun Zhang, Weiran Xu, William Yang Wang, "Generative adversarial zero-shot relational learning for knowledge graphs", *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, pp. 8673–8680, 2020.

[7] Yuxia Geng, Jiaoyan Chen, Zhuo Chen, Jeff Z Pan, Zhiquan Ye, Zonggang Yuan, Yantao Jia, Huajun Chen, "OntoZSL: Ontology-enhanced zero-shot learning", *Proceedings of the Web Conference 2021*, pp. 3325–3336, 2021.

[8] Zhenyong Fu, Tao Xiang, Elyor Kodirov, Shaogang Gong, "Zero-shot object recognition by semantic manifold distance", *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2635–2644, 2015.

[9] Junyu Gao, Tianzhu Zhang, Changsheng Xu, "I know the relationships: Zero-shot action recognition via two-stream graph convolutional networks and knowledge graphs", *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, pp. 8303–8311, 2019.

[10] Zhijun Dong, Likang Wu, Kai Zhang, Ye Liu, Yanghai Zhang, Zhi Li, Hongke Zhao, Enhong Chen, "FZR: Enhancing Knowledge Transfer via Shared Factors Composition in Zero-Shot Relational Learning", *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management (CIKM '24)*, pp. 497–507, 2024, doi: 10.1145/3627673.3679770.

[11] Zeynep Akata, Florent Perronnin, Zaid Harchaoui, Cordelia Schmid, "Label-embedding for attribute-based classification", *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 819–826, 2013.

[12] Ali Farhadi, Ian Endres, Derek Hoiem, David Forsyth, "Describing objects by their attributes", *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1778–1785, 2009.

[13] Christoph H Lampert, Hannes Nickisch, Stefan Harmeling, "Attribute-based classification for zero-shot visual object categorization", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 3, pp. 453–465, 2013.

[14] Jingjing Li, Mengmeng Jing, Ke Lu, Zhengming Ding, Lei Zhu, Zi Huang, "Leveraging the invariant side of generative zero-shot learning", *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 7402–7411, 2019.

[15] Lu Liu, Tianyi Zhou, Guodong Long, Jing Jiang, Chengqi Zhang, "Attribute propagation network for graph zero-shot learning", *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, pp. 4868–4875, 2020.

[16] Andrea Frome, Greg S Corrado, Jon Shlens, Samy Bengio, Jeff Dean, Marc'Aurelio Ranzato, Tomas Mikolov, "Devise: A deep visual-semantic embedding model", *Advances in Neural Information Processing Systems*, vol. 26, 2013.

[17] Junji Jiang, Hongke Zhao, Ming He, Likang Wu, Kai Zhang, Jianping Fan, "Knowledge-Aware Cross-Semantic Alignment for Domain-Level Zero-Shot Recommendation", *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management*, pp. 965–975, 2023.

[18] Mohammad Norouzi, Tomas Mikolov, Samy Bengio, Yoram Singer, Jonathon Shlens, Andrea Frome, Greg S Corrado, Jeffrey Dean, "Zero-shot learning by convex combination of semantic embeddings", *arXiv preprint arXiv:1312.5650*, 2013.

[19] Li Zhang, Tao Xiang, Shaogang Gong, "Learning a deep embedding model for zero-shot learning", *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2021–2030, 2017.

[20] Ruizhi Qiao, Lingqiao Liu, Chunhua Shen, Anton Van Den Hengel, "Less is more: zero-shot learning from online textual documents with noise suppression", *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2249–2257, 2016.

[21] Xin Wang, Jiawei Wu, Da Zhang, Yu Su, William Yang Wang, "Learning to compose topic-aware mixture of experts for zero-shot video captioning", *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, pp. 8965–8972, 2019.

[22] Yuxia Geng, Jiaoyan Chen, Zhiquan Ye, Zonggang Yuan, Wei Zhang, Huajun Chen, "Explainable zero-shot learning via attentive graph convolutional network and knowledge graphs", *Semantic Web*, vol. 12, no. 5, pp. 741–765, 2021.

[23] Abhinaba Roy, Deepanway Ghosal, Erik Cambria, Navonil Majumder, Rada Mihalcea, Soujanya Poria, "Improving zero-shot learning baselines with commonsense knowledge", *Cognitive Computation*, vol. 14, no. 6, pp. 2212–2222, 2022.

[24] Xiaolong Wang, Yufei Ye, Abhinav Gupta, "Zero-shot recognition via semantic embeddings and knowledge graphs", *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 6857–6866, 2018.

[25] Jiaoyan Chen, Freddy Lecue, Yuxia Geng, Jeff Z Pan, Huajun Chen, "Ontology-guided semantic composition for zero-shot learning", *arXiv preprint arXiv:2006.16917*, 2020.

[26] Yuxia Geng, Jiaoyan Chen, Wen Zhang, Yajing Xu, Zhuo Chen, Jeff Z Pan, Yufeng Huang, Feiyu Xiong, Huajun Chen, "Disentangled ontology embedding for zero-shot learning", *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 443–453, 2022.

[27] Elyor Kodirov, Tao Xiang, Shaogang Gong, "Semantic autoencoder for zero-shot learning", *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3174–3183, 2017.

[28] Soravit Changpinyo, Wei-Lun Chao, Boqing Gong, Fei Sha, "Synthesized classifiers for zero-shot learning", *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5327–5336, 2016.

[29] Michael Kampffmeyer, Yinbo Chen, Xiaodan Liang, Hao Wang, Yujia Zhang, Eric P Xing, "Rethinking knowledge graph propagation for zero-shot learning", *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 11487–11496, 2019.

[30] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, Yoshua Bengio, "Generative adversarial nets", *Advances in Neural Information Processing Systems*, vol. 27, 2014.

[31] Yongqin Xian, Tobias Lorenz, Bernt Schiele, Zeynep Akata, "Feature generating networks for zero-shot learning", *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5542–5551, 2018.

[32] Yizhe Zhu, Mohamed Elhoseiny, Bingchen Liu, Xi Peng, Ahmed Elgammal, "A generative adversarial approach for zero-shot learning from noisy texts", *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1004–1013, 2018.

[33] Yizhe Zhu, Jianwen Xie, Bingchen Liu, Ahmed Elgammal, "Learning feature-to-feature translator by alternating back-propagation for generative zero-shot learning", *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 9844–9854, 2019.

[34] Takuo Hamaguchi, Hidekazu Oiwa, Masashi Shimbo, Yuji Matsumoto, "Knowledge transfer for out-of-knowledge-base entities: A graph neural network approach", *arXiv preprint arXiv:1706.05674*, 2017.

[35] Baoxu Shi, Tim Weninger, "Open-world knowledge graph completion", *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 32, 2018.

[36] Peifeng Wang, Jialong Han, Chenliang Li, Rong Pan, "Logic attention based neighborhood aggregation for inductive knowledge graph embedding", *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, pp. 7152–7159, 2019.

[37] Ming Zhao, Weijia Jia, Yusheng Huang, "Attention-based aggregation graph networks for knowledge graph information transfer", *Advances in Knowledge Discovery and Data Mining: 24th Pacific-Asia Conference, PAKDD 2020, Singapore, May 11–14, 2020, Proceedings, Part II*, vol. 24, pp. 542–554, 2020.

[38] Haseeb Shah, Johannes Villmow, Adrian Ulges, Ulrich Schwanecke, Faisal Shafait, "An open-world extension to knowledge graph completion models", *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, pp. 3044–3051, 2019.

[39] Mingyang Chen, Wen Zhang, Yushan Zhu, Hongting Zhou, Zonggang Yuan, Changliang Xu, Huajun Chen, "Meta-knowledge transfer for inductive knowledge graph embedding", *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 927–937, 2022.

[40] Komal Teru, Etienne Denis, Will Hamilton, "Inductive relation prediction by subgraph reasoning", *International Conference on Machine Learning*, pp. 9448–9457, 2020.

[41] Jiawei Cheng, Jingyuan Wang, Yichuan Zhang, Jiahao Ji, Yuanshao Zhu, Zhibo Zhang, Xiangyu Zhao, "POI-Enhancer: An LLM-based semantic enhancement framework for POI representation learning", *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 39, no. 11, pp. 11509–11517, 2025.

[42] Xiaoxiao Ma, Yuchen Zhang, Kaize Ding, Jian Yang, Jia Wu, Hao Fan, "On fake news detection with LLM-enhanced semantic mining", *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pp. 508–521, 2024.

[43] Weiqing He, Bojian Hou, Tianqi Shang, Davoud Ataee Tarzanagh, Qi Long, Li Shen, "SEFD: Semantic-enhanced framework for detecting LLM-generated text", *2024 IEEE International Conference on Big Data (BigData)*, pp. 1309–1314, 2024.

[44] Imed Keraghel, Stanislas Morbieu, Mohamed Nadif, "Beyond words: A comparative analysis of LLM embeddings for effective clustering", *International Symposium on Intelligent Data Analysis*, pp. 205–216, 2024.

[45] Zhijie Nie, Zhangchi Feng, Mingxin Li, Cunwang Zhang, Yanzhao Zhang, Dingkun Long, Richong Zhang, "When text embedding meets large language model: A comprehensive survey", *arXiv preprint arXiv:2412.09165*, 2024.

[46] Jun Hu, Wenwen Xia, Xiaolu Zhang, Chilin Fu, Weichang Wu, Zhaoxin Huan, Ang Li, Zuoli Tang, Jun Zhou, "Enhancing sequential recommendation via LLM-based semantic embedding learning", *Companion Proceedings of the ACM Web Conference 2024*, pp. 103–111, 2024.

[47] An Yang et al., "Qwen3 technical report", *arXiv preprint arXiv:2505.09388*, 2025.

[48] Robert L. Solso, M. Kimberly MacLin, Otto H. MacLin, *Cognitive Psychology*. Pearson Education New Zealand, 2005.

[49] Xiaoyu Kou, Yankai Lin, Shaobo Liu, Peng Li, Jie Zhou, Yan Zhang, "Disentangle-based Continual Graph Representation Learning", *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 2961–2972, 2020.

[50] Jeffrey Pennington, Richard Socher, Christopher D Manning, "GloVe: Global vectors for word representation", *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 1532–1543, 2014.

[51] Florian Boudin, "pke: an open source python-based keyphrase extraction toolkit", *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: System Demonstrations*, pp. 69–73, 2016.

[52] James MacQueen, "Some methods for classification and analysis of multivariate observations", *Proceedings of the fifth Berkeley Symposium on Mathematical Statistics and Probability*, vol. 1, pp. 281–297, 1967.

[53] Andrew Carlson, Justin Betteridge, Bryan Kisiel, Burr Settles, Estevam Hruschka, Tom Mitchell, "Toward an architecture for never-ending language learning", *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 24, pp. 1306–1313, 2010.

[54] Antoine Bordes, Nicolas Usunier, Alberto Garcia-Duran, Jason Weston, Oksana Yakhnenko, "Translating embeddings for modeling multi-relational data", *Advances in Neural Information Processing Systems*, vol. 26, 2013.

[55] Bishan Yang, Wen-tau Yih, Xiaodong He, Jianfeng Gao, Li Deng, "Embedding entities and relations for learning and inference in knowledge bases", *arXiv preprint arXiv:1412.6575*, 2014.

[56] Theo Trouillon, Johannes Welbl, Sebastian Riedel, Eric Gaussier, Guillaume Bouchard, "Complex embeddings for simple link prediction", *International Conference on Machine Learning*, pp. 2071–2080, 2016.

[57] Diederik P Kingma, Jimmy Ba, "Adam: A method for stochastic optimization", *arXiv preprint arXiv:1412.6980*, 2014.