Breaking the Reasoning Horizon in Entity Alignment Foundation Models

Yuanning Cui[1], **Zequn Sun**[2], **Wei Hu**[2,3], **Kexuan Xin**[4] and **Zhangjie Fu**[1,5†]
[1]Nanjing University of Information Science and Technology
[2]State Key Laboratory for Novel Software Technology, Nanjing University
[3]National Institute of Healthcare Data Science, Nanjing University, Nanjing
[4]University of Queensland
[5]Engineering Research Center of Digital Forensics, Ministry of Education, Nanjing University of Information Science and Technology
yncui@nuist.com, {sunzq, whu}@nju.edu.cn, ke.xin@student.uq.edu.au, fzj@nuist.edu.cn

Abstract

Entity alignment (EA) is critical for knowledge graph (KG) fusion. Existing EA models lack transferability and are incapable of aligning unseen KGs without retraining. While graph foundation models (GFMs) offer a solution, we find that directly adapting GFMs to EA remains largely ineffective. This stems from a critical "reasoning horizon gap": unlike link prediction in GFMs, EA necessitates capturing long-range dependencies across sparse and heterogeneous KG structures. To address this challenge, we propose an EA foundation model driven by a parallel encoding strategy. We utilize seed EA pairs as local anchors to guide the information flow, initializing and encoding two parallel streams simultaneously. This facilitates anchor-conditioned message passing and significantly shortens the inference trajectory by leveraging local structural proximity instead of global search. Additionally, we incorporate a merged relation graph to model global dependencies and a learnable interaction module for precise matching. Extensive experiments verify the effectiveness of our framework, highlighting its strong generalizability to unseen KGs.

1 Introduction

Knowledge Graphs (KGs) are essential resources for various AI applications, such as question answering and recommender systems [Ji *et al.*, 2022; Pan *et al.*, 2024]. Since real-world KGs are often constructed independently from different data sources, they usually suffer from heterogeneity in naming conventions and schemas. Entity alignment (EA), which aims to identify equivalent entities referring to the same object across KGs, plays a critical role in integrating these multi-source KGs [Zeng *et al.*, 2021; Sun *et al.*, 2023a]. Existing EA methods are predominantly embedding-based, aiming to tackle the inherent heterogeneity between different KGs [Zhang *et al.*, 2022]. The diverse naming conventions and distinct schema structures make direct symbolic matching impractical. Embedding-based methods map entities and relations into a unified low-dimensional space. By preserving structural similarity, they position equivalent entities close to each other in embedding space, reducing discrepancies caused by surface-level and schema-level heterogeneity.

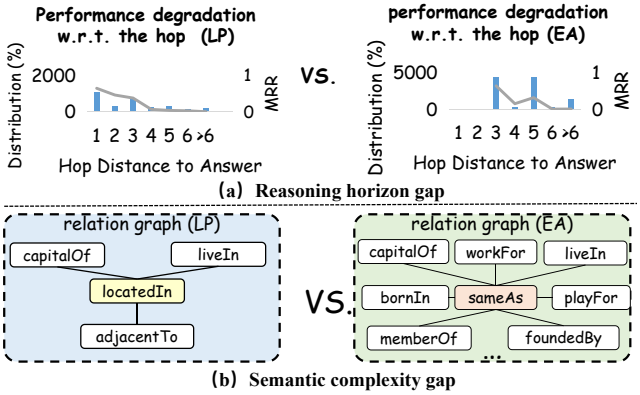


Figure 1: Illustration of key challenges in adapting GFMs to EA. (a) Reasoning horizon gap: Performance degradation with reasoning distance empirically validates the difficulty of long-range inference in EA compared with link prediction (LP). (b) Semantic complexity gap: Topological patterns of standard relations and the *sameAs* relation, illustrating the latter's higher-order connectivity and complexity.

However, existing embedding-based EA methods suffer from a critical limitation: they typically operate in a transductive setting. Specifically, these models heavily rely on learning specific entity embeddings to calculate EA similarities, meaning the model parameters are tightly coupled with the entity sets of the training KGs. Consequently, an EA model trained on one pair of KGs cannot be directly transferred to align unseen KGs. Whenever a new EA task arises, the model must be retrained from scratch, which is computationally expensive and inefficient. This dependency fundamentally hinders the development of a universal model capable of serving arbitrary EA tasks without retraining. To handle unseen entities, some inductive or continual learning methods essentially operate by synthesizing fixed embeddings for new entities based on their neighbors [Wang *et al.*, 2022]. We argue that, fundamentally, these inductive models share the same drawback as traditional

† Corresponding author.

transductive models: they both rely on "*absolute representations*." This dependency on absolute feature spaces inherently limits their flexibility and transferability.

Recently, knowledge graph foundation models (KGFMs) [Galkin *et al.*, 2024; Cui *et al.*, 2024; Zhang *et al.*, 2025; Huang *et al.*, 2025] have introduced *relative representations* to overcome the limitation. Instead of memorizing node IDs, GFMs exploit universal relational patterns and topological structures to generate entity embeddings. By encoding structural roles independent of absolute identities, this paradigm provides strong transferability, allowing the models to generalize well across diverse and unseen graphs. However, existing KGFMs primarily focus on intra-graph tasks, such as link prediction or triplet classification within a single graph. Directly adapting KGFMs to EA by treating alignment as a "sameAs" link prediction task, is often inefficient.

The fundamental challenge lies in the structural gap: unlike intra-graph reasoning where paths between nodes are continuous, EA requires bridging two disjoint graph structures. Figure 1 illustrates the critical challenges in adapting KGFMs to EA. In summary, inference based on standard message passing often involves long-range trajectories. It forces the model to search through extensive global structures to locate the target entity, which leads to suboptimal convergence and poor generalization on unseen KGs.

We argue that the key to efficient EA lies in exploiting local structural proximity rather than performing exhaustive global traversal. In practical EA settings, a set of pre-aligned seed pairs is typically available and can serve as local anchors to bridge structural gaps [Sun *et al.*, 2020b]. By initiating reasoning from these anchors in both KGs simultaneously, the inference path from a query entity to its counterpart can be substantially shortened. Building on this insight, we propose EAFM for transferable EA. It adopts a parallel encoding strategy: encoding streams for two KGs are initialized at the same time using the seed EA pairs as shared anchors. This enables anchor-conditioned message passing, where information flow is guided by proximity to these anchors. Thus, EAFM can quickly locate target entities based on local structural context, avoiding the inefficiency of long-range global inference.

To further address relational-schema heterogeneity, we introduce a global relation graph to capture high-order semantic dependencies independent of specific entities. The resulting global relation features are then used to condition entity-level message passing. Finally, instead of relying on static metrics such as cosine similarity, we design a learnable interaction matching module with a bidirectional classification objective to capture fine-grained semantic discrepancies.

Our main contributions are summarized as follows:

- **Pioneering EA foundation model**. To the best of our knowledge, we provide the first quantitative analysis of the "reasoning horizon gap" and propose a novel EA foundation model to bridge this gap. It can handle arbitrary unseen EA tasks without any additional training.
- **Anchor-conditioned message passing**. We introduce a novel encoding strategy that leverages seed EA as local anchors to guide the message passing. By transforming long-range global reasoning into local relative matching, it effectively shortens the inference trajectory.
- **Decoupled interaction architecture**. We design a scalable parallel encoding architecture equipped with a unified relation learner and a learnable interaction module. This design effectively decouples the representation learning from specific graph statistics, ensuring high robustness against the structural heterogeneity and schema distribution shifts across different KGs.
- **Superior generalizability and performance**. Extensive experiments on benchmark datasets demonstrate that EAFM significantly outperforms state-of-the-art baselines. It exhibits superior transferability when applied to unseen KGs without retraining.

2 Related Work and Discussions

2.1 Knowledge Graph Foundation Models

Achieving transferable reasoning across arbitrary KGs has become a key goal in the community. Traditional transductive methods are restricted to fixed entity sets and thus cannot generalize to new domains. To address this, inductive reasoning models such as NBFNet [Zhu *et al.*, 2021] and RED-GNN [Zhang and Yao, 2022] were proposed. These models infer over relational paths and subgraph structures rather than specific entity IDs, enabling reasoning on unseen entities within the same schema. Building on these inductive principles, recent progress has led to the development of KGFMs aimed at universal reasoning across heterogeneous KGs. PR4LP [Sun *et al.*, 2023b] follows a pre-train–fine-tune paradigm, adapting to target domains through entity-alignment supervision. ULTRA [Galkin *et al.*, 2024] further reduces reliance on fine-tuning by constructing a relation graph that captures invariant relational interactions for zero-shot generalization. TRIX [Zhang *et al.*, 2025] and MOTIF [Huang *et al.*, 2025] explore pre-training to learn transferable structural patterns, while KG-ICL [Cui *et al.*, 2024] leverages in-context learning for few-shot adaptation.

However, current KGFMs are mainly designed for intra-graph tasks, such as link prediction. Directly applying them for EA by predicting "sameAs" relations is inefficient due to the "reasoning horizon gap": the inference path required to connect two disjoint KGs often exceeds the receptive field of models built for single-graph reasoning. In contrast, our framework is explicitly designed for the EA foundation setting, bridging this structural gap through a parallel encoding strategy.

2.2 Entity Alignment

EA is a fundamental task of integrating multiple KGs for knowledge fusion. Recent EA models leverage deep learning techniques to encode topological information from KG to find equivalent entities of different KGs. Generally, three types of topological features are commonly used. Translation-based embeddings encode local relational patterns in a continuous space, following TransE [Bordes *et al.*, 2013] and its variants, such as MTransE [Chen *et al.*, 2017], AlignE [Sun *et al.*, 2018], and SEA [Pei *et al.*, 2019]. Long-range dependency modeling captures higher-order relational dependencies beyond immediate neighborhoods, including IPTransE [Zhu *et al.*, 2017], RSN4EA [Guo *et al.*, 2019], and

IMEA [Xin *et al.*, 2022]. GNN-based neighborhood aggregation encodes structural context via message passing, including GCN-Align [Wang *et al.*, 2018], AliNet [Sun *et al.*, 2020a], RREA [Mao *et al.*, 2020], and Dual-AMN [Mao *et al.*, 2021], which generally achieve strong performance by integrating multi-hop neighborhood signals. Some recent work further refines this GNN-style pipeline from the perspectives of structure enhancement [Ding *et al.*, 2025; Wang *et al.*, 2025]. Another work NeuSymEA [Chen *et al.*,] represents a different line that combines neural structural scoring with weighted structural rules in a unified probabilistic framework optimized by variational EM. Finally, although LLM-driven EA has rapidly grown in recent years [Chen *et al.*, 2025], it is orthogonal to the topology-only focus here.

Nevertheless, transferring existing EA methods to a new benchmark typically requires re-training or at least re-tuning. In contrast, our method pre-trains a KGFM on one EA dataset and can be applied to new datasets without re-training. It enables practical cross-dataset transfer.

3 Methodology

3.1 Problem Formulation

Let $\mathcal{G}_1 = (\mathcal{E}_1, \mathcal{R}_1, \mathcal{T}_1)$ and $\mathcal{G}_2 = (\mathcal{E}_2, \mathcal{R}_2, \mathcal{T}_2)$ denote two heterogeneous KGs, where $\mathcal{E}$, $\mathcal{R}$, and $\mathcal{T}$ represent the sets of entities, relations, and facts, respectively. A fact is denoted as $(h, r, t) \in \mathcal{E} \times \mathcal{R} \times \mathcal{E}$. Following convention [Sun *et al.*, 2020b], we assume the availability of a seed EA set (anchors) $\mathcal{S} = \{(u, v) | u \in \mathcal{E}_1, v \in \mathcal{E}_2, u \equiv v\}$, which serves as supervision signal to bridge two KGs. The goal of the EA task is to infer a one-to-one mapping $\mathcal{M}$ covering the remaining equivalent entities, such that $\mathcal{M} = \{(e_1, e_2) \in \mathcal{E}_1 \times \mathcal{E}_2 | e_1 \equiv e_2\}$.

3.2 Framework Overview

As illustrated in Figure 2, the workflow of our foundation model is divided into two stages: pre-training and inference.

- **Pre-training**: Our model is trained on source EA datasets containing seed EA pairs. We first construct a merged relation graph to capture the topological correlations between relations. Then, we employ a relation-aware GNN (RelGNN) and an entity-aware GNN (EntGNN) to learn structural representations. The model finally minimizes the alignment loss to optimize the parameters of both the encoders and the entity matcher.
- **Inference**: In the inference phase, given two unseen KGs for EA, we can directly deploy the frozen RelGNN and EntGNN to reason over the unseen KGs and predict EA pairs. It is worth noting that our model relies exclusively on graph structures for transferable inference without parameter updating.

3.3 Global Relation Learning via Merged Graph

A major challenge in cross-KG EA is the heterogeneity of relational schemas. To address this, we construct a *merged relation graph* $\mathcal{G}_{rel}$ that captures high-order dependencies independent of specific entity contexts.

Relation Graph Construction. In the relation graph $\mathcal{G}_{rel}$, the node set is the union of all relations $\mathcal{V}_{rel} = \mathcal{R}_1 \cup \mathcal{R}_2$. To define the edges, we leverage the seed EA set $\mathcal{S}$ to effectively "merge" the entity sets of $\mathcal{G}_1$ and $\mathcal{G}_2$ into a virtual unified entity space $\mathcal{E}_{uni}$. We define two sparse incidence matrices $\mathbf{H}, \mathbf{T} \in \{0, 1\}^{|\mathcal{E}_{uni}| \times |\mathcal{V}_{rel}|}$, where $\mathbf{H}_{er} = 1$ (or $\mathbf{T}_{er} = 1$) if entity e serves as the head (or tail) of relation r. Based on these matrices, we derive five types of topological interactions to capture comprehensive semantic correlations:

- **Head-Head**: Relations sharing the same head entity.
- **Head-Tail**: The head of r_i corresponds to the tail of r_j.
- **Tail-Head**: The tail of r_i corresponds to the head of r_j.
- **Tail-Tail**: Relations sharing the same tail entity.
- **Inverse**: Explicit connections between a relation and its inverse.

This graph structure $\mathcal{G}_{rel}$ acts as a global prior, linking disjoint relational spaces through the shared anchors in $\mathcal{S}$.

Structure-Aware RelGNN. We employ a multi-layer relational GNN to encode $\mathcal{G}_{rel}$. Let $\mathbf{r}_i^{(l)}$ be the embedding of relation r_i at layer l. Before message passing, we perform a query-conditioned initialization over relations. Specifically, we set the embeddings of the relations that appear in the one-hop neighborhood of the query entity to an all-ones vector $\mathbf{1}_d \in \mathbb{R}^d$, while initializing all other relation embeddings to an all-zeros vector in $\mathbf{0}_d \in \mathbb{R}^d$.

For a neighbor relation r_j connected via edge type τ_{ji}, we first inject structural information via a learnable prototype embedding $\mathbf{p}_{\tau_{ji}} \in \mathbb{R}^d$:

$$\tilde{\mathbf{r}}_j = \mathbf{r}_j^{(l)} + \mathbf{p}_{\tau_{ji}}. \tag{1}$$

We then compute an attention coefficient α_{ij} to weigh the importance of neighbor r_j to target r_i:

$$\alpha_{ij} = \sigma\left(\mathbf{W}_\alpha [\tilde{\mathbf{r}}_j \oplus \mathbf{r}_i^{(l)}]\right), \tag{2}$$

where $\oplus$ denotes concatenation and $\mathbf{W}_\alpha$ is a learnable weight vector. The relation embedding is updated via a residual connection and LeakyReLU activation:

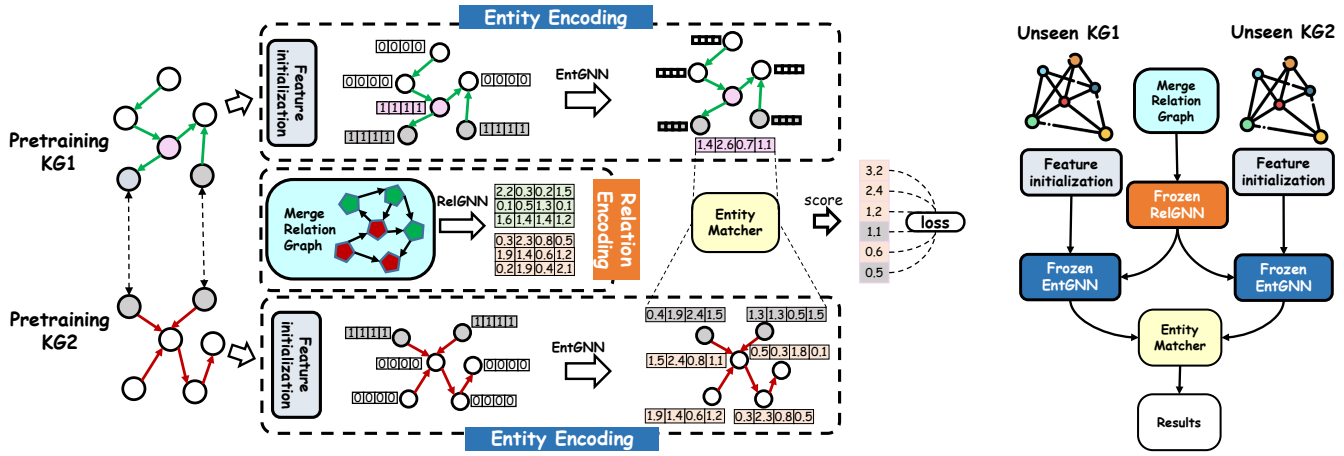
$$\mathbf{r}_i^{(l+1)} = \phi\Big(\mathbf{W}_{rel}\mathbf{r}_i^{(l)} + \sum_{r_j \in \mathcal{N}(r_i)} \alpha_{ij}\mathbf{W}_{msg}\tilde{\mathbf{r}}_j\Big). \tag{3}$$

The final relation embeddings $\mathbf{R}_{global}$ encapsulate both local topological patterns and global cross-KG dependencies.

3.4 Parallel Entity Encoding

Unlike methods that treat EA as a long-range link prediction task starting from a source entity, we employ a parallel encoding strategy, and utilize a shared graph neural network to process $\mathcal{G}_1$ and $\mathcal{G}_2$ simultaneously, leveraging the pre-learned global relations and local anchors.

Anchor-Conditioned Initialization. This entity initialization method serves as the base for enabling our transferability to unseen KGs. Unlike traditional methods that depend on pre-trained embeddings (e.g., TransE) which are unavailable



(a) Entity alignment pre-training **(b) Transferable inference**

Figure 2: The overall architecture of the proposed entity alignment foundation model. **(a) Structural Pre-training Phase:** The model employs a parallel encoding strategy on source KGs. It jointly optimizes the *RelGNN* (via the Merge Relation Graph) and the *EntGNN* (via anchor-conditioned message passing) to capture transferable structural patterns. **(b) Inductive Inference Phase:** When transferring to unseen KGs, the pre-trained *RelGNN* and *EntGNN* parameters are **frozen**. The model directly performs alignment by initializing features based on local anchors in the new graphs, without requiring re-training or fine-tuning.

for new entities, we propose a query-specific method that initializes a query entity based on its topological distances to the pre-aligned anchor entities. Specifically, for a query entity, we restrict the context to local anchors residing within a k-hop neighborhood. By encoding the relative structural coordinates instead of absolute IDs, this design bypasses the need for pre-training specific embeddings, allowing the model to directly generate initial representations for unseen entities.

Formally, let $\mathcal{N}_k(e_q)$ denote the k-hop neighborhood of a query entity e_q. The set of active local anchors is defined as $\mathcal{A}(e_q) = \{u \mid (u, v) \in \mathcal{S} \wedge u \in \mathcal{N}_k(e_q)\}$. The initial feature $\mathbf{h}_u^{(0)}$ for a node u is dynamically assigned as:

$$\mathbf{h}_u^{(0)} = \begin{cases} \mathbf{1}_d & \text{if } u \in \mathcal{A}(e_q), \\ \mathbf{0}_d & \text{otherwise,} \end{cases} \tag{4}$$

This strategy is applied in parallel to both $\mathcal{G}_1$ and $\mathcal{G}_2$. If an entity u in $\mathcal{G}_1$ is activated because it lies in the k-hop neighborhood of the source query, its counterpart v in $\mathcal{G}_2$ is simultaneously activated. This allows the model to learn representations based on the relative structural position to these shared local landmarks, effectively ignoring graph-specific noise and enabling generalizability to unseen graphs.

Conditional Message Passing. We adopt a translation-based EntGNN that explicitly incorporates relation embeddings into entity aggregation. Consider a target entity e_i and its incoming neighbor e_j connected by relation r_{ji}. The message $\mathbf{m}_{ji}$ is formulated as a translation in embedding space:

$$\mathbf{m}_{ji} = \mathbf{h}_j^{(l)} + \mathbf{r}_{ji}, \tag{5}$$

where $\mathbf{r}_{ji} \in \mathbf{R}_{global}$ is the context-aware relation embedding obtained based on Section 3.3. This ensures that the entity representation updates are conditioned on the global relational structure. The aggregation weights β_{ji} are computed via a structure-aware attention mechanism:

$$\beta_{ji} = \frac{\exp(\sigma(\mathbf{a}^T[\mathbf{W}_s\mathbf{h}_j^{(l)} \oplus \mathbf{W}_r\mathbf{r}_{ji}]))}{\sum_{k\in\mathcal{N}(e_i)} \exp(\sigma(\mathbf{a}^T[\mathbf{W}_s\mathbf{h}_k^{(l)} \oplus \mathbf{W}_r\mathbf{r}_{ki}]))}. \tag{6}$$

Finally, the entity representation is updated using a residual connection followed by the layer normalization method to ensure training stability:

$$\mathbf{h}_i^{(l+1)} = \mathrm{LN}\left(\mathbf{h}_i^{(l)} + \phi\left(\mathbf{W}_{ent} \sum_{j\in\mathcal{N}(e_i)} \beta_{ji}\mathbf{m}_{ji}\right)\right). \tag{7}$$

Shortening Inference Trajectory. By running the two encoding channels in parallel with shared anchor initialization, our model performs anchor-conditioned propagation. Information flows from the nearest anchors to the query entities in both KGs simultaneously. This mechanism effectively shortens the inference trajectory: the model determines EA based on the local structural proximity to shared anchors (k-hop neighborhood) rather than performing an exhaustive global search, significantly enhancing both efficiency and accuracy.

3.5 Interaction and Alignment Learning

After L layers of propagation, we obtain the final entity embeddings $\mathbf{H}_1$ and $\mathbf{H}_2$. We employ an interaction-based module to perform precise matching.

Fine-grained Interaction Features. For a source entity $e_s \in \mathcal{G}_1$ and a candidate target entity $e_t \in \mathcal{G}_2$, we construct a joint interaction vector $\mathbf{v}_{st}$. Unlike simple cosine similarity, we model the feature discrepancy and target context:

$$\mathbf{v}_{st} = [|\mathbf{h}_s - \mathbf{h}_t| \oplus \mathbf{h}_t], \tag{8}$$

where $|\cdot|$ denotes element-wise absolute difference. The alignment score is computed via a linear projection:

$$S(e_s, e_t) = \mathbf{W}_{final}\mathbf{v}_{st}. \tag{9}$$

Bidirectional Classification Objective. We formulate EA as a bidirectional classification task. For a batch of aligned pairs $\mathcal{B}$, we minimize the symmetric cross-entropy loss:

$$\mathcal{L} = \mathcal{L}_{1\to2} + \mathcal{L}_{2\to1}, \tag{10}$$

where $\mathcal{L}_{1\to2}$ is the loss for aligning entities from $\mathcal{G}_1$ to $\mathcal{G}_2$:

$$\mathcal{L}_{1\to2} = -\frac{1}{|\mathcal{B}|} \sum_{(e_s,e_t)\in\mathcal{B}} \log \frac{\exp(S(e_s,e_t))}{\sum_{e_k\in\mathcal{E}_2}\exp(S(e_s,e_k))}. \tag{11}$$

The term $\mathcal{L}_{2\to1}$ is defined symmetrically. This bidirectional strategy forces the model to learn mutually consistent EA, ensuring that e_s is the nearest neighbor of e_t and vice versa.

4 Experiment

In this section, we present experimental results and analyses to assess the effectiveness of our model on transferable EA. In addition to the main results, we provide further analyses in Appendix B.1. These include the efficiency of merged relation graph construction and pre-training, the dependence on visible seed EA pairs, and the impact of pre-training data characteristics on transferable inference performance, as detailed in Appendix B.6.

4.1 Experimental Settings

Datasets and Protocols. We consider the most widely used EA datasets, covering three representative groups: OpenEA [Sun *et al.*, 2020b], SRPRS [Guo *et al.*, 2019], and the DBpedia series (abbr. DBP) [Sun *et al.*, 2017; Sun *et al.*, 2018]. This diverse collection enables a comprehensive evaluation across varying scales, languages, and structural densities. To evaluate the transferability of our foundation model, we adopt an inductive setting where the model parameters are frozen after pre-training on the source dataset and directly applied to align entities in unseen KGs without fine-tuning. For data partitioning, we follow standard protocols [Sun *et al.*, 2020b] with a 20%, 10%, and 70% split for training, validation, and testing, respectively. Performance is evaluated using Mean Reciprocal Rank (MRR) and Hits@10 (abbr. H@10), where higher scores indicate better alignment accuracy. For clarity of presentation, we categorize the datasets into several groups based on their sources and report average results. Furthermore, in the subsequent analysis experiments, we regroup and compare them from multiple perspectives to conduct a multi-dimensional analysis of the model's transferability. Table 5 in Appendix A summarizes the statistics of the benchmarks used in this work. Table 6 in Appendix A present the detailed EA results on each dataset.

Comparison Baselines. To the best of our knowledge, there is currently no purely structure transfer method that can be applied to the transferable EA task. To rigorously evaluate the architectural advantages of our proposed foundation model, we compare EAFM against ULTRA [Galkin *et al.*, 2024], a state-of-the-art graph foundation model. We construct two targeted baselines to validate our motivations:

- **ULTRA-LP** directly applies the original ULTRA model to EA by treating it as a "sameAs" prediction task. ULTRA-LP is pre-trained on three link prediction datasets. This setting is designed to examine the "reasoning horizon gap" hypothesis.
- **ULTRA-EA** re-trains the ULTRA model from scratch using exactly the same datasets and experimental configurations as our model, isolating the contributions of our parallel encoding strategy and anchor-conditioned mechanism. This setting is designed to examine the effectiveness of our model architecture.

Implementation Details. The proposed framework is implemented using PyTorch and PyTorch Geometric. All experiments are conducted on a workstation equipped with four NVIDIA RTX 4090 GPU (24GB). For the model configuration, we employ a 6-layer architecture for both the global RelGNN and the parallel EntGNN. The hidden dimension is set to 32. The anchor hop k is set to 2. We optimize the model using the AdamW optimizer with a learning rate of $5e^{-4}$ and a batch size of 64. The training process runs for a maximum of 200 epochs, with an early stopping mechanism triggered if the validation MRR does not improve for 10 consecutive epochs. To investigate the transferability under different structural distributions, we utilize D-W-15K-V1 and EN-DE-15K-V1 as source datasets. Specifically, D-W-15K-V1 is selected to learn heterogeneous structural patterns across different knowledge sources, while EN-DE-15K-V1 is employed to capture structural consistencies across different language versions.

4.2 Main Results

Table 1 presents the comparative evaluation across the three dataset groups. (i) The near-random performance of ULTRA-LP underscores a fundamental distinction between link prediction and entity alignment tasks. While general graph foundation models can effectively encode local topology for graph completion, they do not explicitly learn the cross-graph structural correspondence required by entity alignment. In EA, the key challenge is not only to represent each entity within its own KG, but also to determine whether two entities from different KGs play compatible structural roles under heterogeneous relational contexts. This observation validates the need for specialized pre-training objectives tailored to transferable EA. (ii) Our framework exhibits strong generalization capabilities in the inductive setting. A notable observation is that our frozen pre-trained model substantially outperforms even the finetuned version of the ULTRA-EA baseline. This result implies that the proposed foundation model captures intrinsic structural invariants that persist across different domains and KG distributions. Therefore, instead of relying on gradient updates on the target data, the model can transfer relation-aware matching knowledge learned from source KGs to unseen target KGs. (iii) Finetuning further unleashes the potential of our model. The consistent superiority on both OpenEA and SRPRS verifies that our approach is not biased toward specific graph properties or a single benchmark setting. When target-side supervision is available, finetuning further adapts the learned transferable representations to the local characteristics of target KGs, leading to additional performance gains. These results demonstrate that the proposed framework is effective under both training-free transfer and supervised adaptation settings.

Method	OpenEA			SRPRS			DBP			Average		
	MRR	H@1	H@10	MRR	H@1	H@10	MRR	H@1	H@10	MRR	H@1	H@10
ULTRA-LP	0.013	0.001	0.037	0.017	0.001	0.045	0.013	0.001	0.033	0.014	0.001	0.038
ULTRA-EA pretrain	0.477	0.385	0.635	0.350	0.241	0.557	0.203	0.107	0.379	0.395	0.297	0.569
ULTRA-EA finetune	0.509	0.416	0.691	0.380	0.271	0.587	0.223	0.137	0.409	0.424	0.328	0.614
EAFM pretrain	0.602	0.517	0.764	0.530	0.426	0.735	0.294	0.217	0.452	0.529	0.440	0.702
EAFM finetune	**0.656**	**0.573**	**0.814**	**0.605**	**0.508**	**0.792**	**0.464**	**0.394**	**0.662**	**0.609**	**0.524**	**0.782**

Table 1: Performance comparison across dataset groups, where bold and underline indicate the best and second-best results, respectively.

Method	Graph Density		Graph Scale		Heterogeneity	
	V1 (Sparse)	V2 (Dense)	15K	100K	Cross-KG	Cross-Lingual
ULTRA-EA pretrain	0.314	0.509	0.393	0.397	0.350	0.203
ULTRA-EA finetune	0.342	0.540	0.420	0.431	0.530	0.325
EAFM pretrain	0.451	0.639	0.532	0.523	0.597	0.465
EAFM finetune	**0.509**	**0.692**	**0.602**	**0.552**	**0.649**	**0.524**

Table 2: Detailed breakdown of MRR performance across different graph characteristics. The datasets are categorized by density (Sparse/V1 vs. Dense/V2), scale (15K vs. 100K), and heterogeneity type (Cross-KG vs. Cross-Lingual). Bold and underline indicate the best and second-best results, respectively.

Detailed analysis on graph characteristics. To provide a more fine-grained evaluation of model robustness, we categorize the benchmarks into three fundamental dimensions in Table 2: graph density (Sparse/V1 vs. Dense/V2), scale (15K vs. 100K), and heterogeneity type (cross-KG vs. cross-Lingual). In terms of graph density, while performance naturally is lower on sparse V1 datasets compared to dense V2 ones due to the scarcity of topological signals, our pre-trained model exhibits remarkable resilience, notably surpassing the fine-tuned baseline even when neighbor information is limited. Regarding scalability, despite the search space expanding significantly in the 100K benchmarks, our approach maintains consistent accuracy, validating its ability to effectively filter noise in large-scale scenarios. Most importantly, in the challenging Multi-Language setting where structural disparities are typically most pronounced, our method demonstrates exceptional cross-lingual generalization solely through topology. Across all these diverse characteristics, our frozen foundation model consistently exceeds the performance of the fine-tuned baseline, confirming that EAFM captures universal structural patterns rather than overfitting to specific dataset properties.

4.3 Ablation Study

To investigate the contribution of each component in EAFM, we conduct an ablation study by removing or altering key modules. The results are reported in Table 3. We define three specific variants to isolate the effects of our design choices. First, "w/o merged relation graph" alters the global graph construction. Instead of physically merging relations via incidence matrices, this variant treats "sameAs" as a standard relation and constructs a relation graph on the union of facts from both KGs. Consequently, relations from different KGs are not directly unified but only indirectly connected via the "sameAs" bridge. Second, "w/o parallel entity encoding" validates our core motivation regarding the reasoning horizon. This variant abandons the parallel, anchor-conditioned streams and instead adopts an ULTRA-like message passing mechanism, where conditional propagation radiates outward from the query (head) entity rather than utilizing shared anchors. Finally, "w/o cross-KG interaction" simplifies the decoding phase by removing the interaction module, directly mapping the learned entity embeddings to alignment scores via dot product.

Analysis. As observed in Table 3, the intact EAFM consistently achieves the best performance, validating the synergy of the proposed architecture. The most significant performance degradation is observed in the "w/o parallel entity encoding" variants. This empirical evidence strongly supports our "reasoning horizon" hypothesis: the ULTRA-like strategy of propagating from the query entity struggles to bridge the structural gap in EA, whereas our anchor-conditioned strategy significantly shortens the inference trajectory. Anchor initialization is the core of parallel encoding, as it provides the key information that needs to be propagated. Furthermore, the "w/o merged relation graph" variant shows a notable decline, confirming that merely connecting KGs via "sameAs" is insufficient. Explicitly unifying relational spaces is essential for capturing high-order semantic dependencies. Lastly, removing the "cross-KG interaction" leads to a moderate drop, indicating that while structure learning is primary, fine-grained feature interaction further refines the matching precision. For better transferability, parallel encoding moderates information propagation between the two KGs, while the interaction module facilitates their interactions.

4.4 Further Analyses

Incorporating entity names. Although EAFM is primarily designed as a structure-centric foundation model, its architecture naturally supports the integration of multi-modal features. In this experiment, we evaluate this capability on the V1 bench-

Method	OpenEA		SRPRS		DBP		Average	
	MRR	H@10	MRR	H@10	MRR	H@10	MRR	H@10
Intact Model	0.602	0.764	0.530	0.735	0.294	0.452	0.529	0.702
w/o merged relation graph	0.562	0.734	0.456	0.653	0.277	0.456	0.484	0.664
w/o parallel entity encoding	0.519	0.699	0.426	0.623	0.247	0.426	0.447	0.631
w/o cross-KG interaction	0.590	0.760	0.470	0.662	0.265	0.456	0.501	0.680

Table 3: Ablation results across different dataset groups.

Method	D-W-15K		D-Y-15K		D-W-100K		D-Y-100K	
	MRR	H@10	MRR	H@10	MRR	H@10	MRR	H@10
EAFM	0.604	0.795	0.636	0.770	0.459	0.650	0.694	0.853
w/ entity name	0.638	0.821	0.805	0.897	0.486	0.688	0.849	0.932
Method	**EN-DE-15K**		**EN-FR-100K**		**EN-DE-100K**		**EN-FR-100K**	
	MRR	H@10	MRR	H@10	MRR	H@10	MRR	H@10
EAFM	0.737	0.900	0.533	0.782	0.520	0.682	0.364	0.572
w/ entity name	0.888	0.962	0.780	0.906	0.704	0.819	0.601	0.742

Table 4: Results of incorporating entity names.

marks from the OpenEA dataset, which cover diverse alignment settings including D-W, D-Y, and cross-lingual pairs. To this end, we incorporate textual information by employing the pre-trained BGE-Large-EN-v1.5[1] encoder. Specifically, we use the encoder to extract semantic embeddings from entity names, which are used to initialize the anchor nodes as defined in Equation 4. These embeddings are then fused with structural embeddings before the decoding phase. Table 4 reports the results of this integration. We observe that incorporating entity names consistently improves performance, although the magnitude of the gains varies across different dataset types. The improvements are most pronounced on cross-lingual benchmarks such as EN-DE-100K as well as on the D-Y datasets, whereas the gains on D-W remain relatively moderate. This difference can be explained by the higher consistency and semantic proximity of entity name styles in cross-lingual and D-Y pairs, which enables the BGE encoder to align them more effectively. By contrast, the naming conventions of DBpedia and Wikidata are more heterogeneous, which limits the effectiveness of direct semantic matching. Overall, these results indicate that the proposed framework can leverage auxiliary modalities when they are available, supporting future extensions to multi-modal settings.

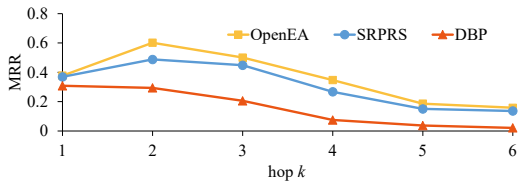


Figure 3: Impact of the anchor-hop distance k on MRR performance across different benchmark groups.

Impact of anchor-hop distance. To investigate the impact of the receptive field size on relative position encoding, we analyze the performance variation with respect to the anchor hop k, ranging from 1 to 6. As shown in Figure 3, the model achieves optimal performance across most benchmarks when k is set to 2. The improvement from $k = 1$ to $k = 2$ indicates that incorporating second-order structural context effectively enhances the distinctiveness of entity representations compared to relying solely on direct neighbors. However, a sharp decline in MRR is observed as k increases beyond 2. This degradation implies that expanding the search scope excessively introduces irrelevant distant nodes which act as noise, thereby diluting the local structural patterns essential for precise alignment. Consequently, we adopt $k = 2$ by default to balance expressiveness with noise suppression.

5 Conclusion

In this paper, we tackle the inherent limitation of existing embedding-based EA methods, which typically lack transferability and require retraining for new datasets. We identified a critical "reasoning horizon gap" that hinders the direct adaptation of generic GFMs to the EA task. To bridge this gap, we proposed EAFM for transferable inference across unseen KGs. By introducing an anchor-conditioned message passing mechanism, EAFM effectively transforms the challenging long-range global inference into a manageable local relative matching problem. Furthermore, our decoupled dual-tower architecture and merged relation graph enable the model to capture universal structural patterns while remaining robust to schema heterogeneity. Extensive experiments on multiple benchmarks demonstrate that EAFM not only significantly outperforms state-of-the-art baselines but also exhibits superior generalizability when applied to unseen KGs without any additional training. Our work marks a pivotal step towards developing a universal, service-oriented model for EA.

Contribution Statement

Yuanning Cui and Zequn Sun contributed equally to this work.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (No. 62406136), the Natural Science Foundation of Jiangsu Province (No. BK20241246), the National Natural Science Foundation of China under grant U22B2062, Jiangsu Provincial Science and Technology Major Project (No. BG2024042), and the Startup Foundation for Introducing Talent of NUIST (No. 1133142601017).

[1]https://huggingface.co/BAAI/bge-large-en-v1.5

References

[Bordes *et al.*, 2013] Antoine Bordes, Nicolas Usunier, Alberto García-Durán, Jason Weston, and Oksana Yakhnenko. Translating embeddings for modeling multi-relational data. In *NIPS*, pages 2787–2795, 2013.

[Chen *et al.*,] Shengyuan Chen, Zheng Yuan, Qinggang Zhang, Wen Hua, Jiannong Cao, and Xiao Huang. Neusymea: Neuro-symbolic entity alignment via variational inference. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*.

[Chen *et al.*, 2017] Muhao Chen, Yingtao Tian, Mohan Yang, and Carlo Zaniolo. Multilingual knowledge graph embeddings for cross-lingual knowledge alignment. In *IJCAI*, pages 1511–1517, Melbourne, Australia, 2017. ijcai.org.

[Chen *et al.*, 2025] Zerui Chen, Huiming Fan, Qianyu Wang, Tao He, Ming Liu, Heng Chang, Weijiang Yu, Ze Li, and Bing Qin. How do language models reshape entity alignment? a survey of lm-driven ea methods: Advances, benchmarks, and future. In *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*, pages 23230–23245, 2025.

[Cui *et al.*, 2024] Yuanning Cui, Zequn Sun, and Wei Hu. A prompt-based knowledge graph foundation model for universal in-context reasoning. In *NeurIPS*, 2024.

[Ding *et al.*, 2025] Linlin Ding, Mengjunyao Si, and Mo Li. Rsgea: Relationship structure line graph for semi-supervised entity alignment based on edge weight adjustment. In *Proceedings of the 48th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 3070–3074, 2025.

[Galkin *et al.*, 2024] Mikhail Galkin, Xinyu Yuan, Hesham Mostafa, Jian Tang, and Zhaocheng Zhu. Towards foundation models for knowledge graph reasoning. In *ICLR*, Vienna, Austria, 2024.

[Guo *et al.*, 2019] Lingbing Guo, Zequn Sun, and Wei Hu. Learning to exploit long-term relational dependencies in knowledge graphs. In *Proceedings of the 36th International Conference on Machine Learning*, pages 2505–2514, 2019.

[Huang *et al.*, 2025] Xingyue Huang, Pablo Barceló, Michael M. Bronstein, İsmail İlkan Ceylan, Mikhail Galkin, Juan L. Reutter, and Miguel A. Romero Orth. How expressive are knowledge graph foundation models? In *ICML*. OpenReview.net, 2025.

[Ji *et al.*, 2022] Shaoxiong Ji, Shirui Pan, Erik Cambria, Pekka Marttinen, and Philip S. Yu. A survey on knowledge graphs: Representation, acquisition, and applications. *TNNLS*, 33(2):494–514, 2022.

[Mao *et al.*, 2020] Xin Mao, Wenting Wang, Huimin Xu, Yuanbin Wu, and Man Lan. Relational reflection entity alignment. In *Proceedings of the 29th ACM International Conference on Information and Knowledge Management*, pages 1095–1104, 2020.

[Mao *et al.*, 2021] Xin Mao, Wenting Wang, Yuanbin Wu, and Man Lan. Boosting the speed of entity alignment 10×: Dual attention matching network with normalized hard sample mining. In *WWW*, pages 821–832, 2021.

[Pan *et al.*, 2024] Shirui Pan, Linhao Luo, Yufei Wang, Chen Chen, Jiapu Wang, and Xindong Wu. Unifying large language models and knowledge graphs: A roadmap. *IEEE Transactions on Knowledge and Data Engineering*, 36(7):3580–3599, 2024.

[Pei *et al.*, 2019] Shichao Pei, Lu Yu, Robert Hoehndorf, and Xiangliang Zhang. Semi-supervised entity alignment via knowledge graph embedding with awareness of degree difference. In *The world wide web conference*, pages 3130–3136, 2019.

[Sun *et al.*, 2017] Zequn Sun, Wei Hu, and Chengkai Li. Cross-lingual entity alignment via joint attribute-preserving embedding. In *ISWC*, pages 628–644, 2017.

[Sun *et al.*, 2018] Zequn Sun, Wei Hu, Qingheng Zhang, and Yuzhong Qu. Bootstrapping entity alignment with knowledge graph embedding. In *IJCAI*, pages 4396–4402, Stockholm, Sweden, 2018. ijcai.org.

[Sun *et al.*, 2020a] Zequn Sun, Chengming Wang, Wei Hu, Muhao Chen, Jian Dai, Wei Zhang, and Yuzhong Qu. Knowledge graph alignment network with gated multi-hop neighborhood aggregation. In *AAAI*, pages 222–229, New York City, NY, USA, 2020. AAAI Press.

[Sun *et al.*, 2020b] Zequn Sun, Qingheng Zhang, Wei Hu, Chengming Wang, Muhao Chen, Farahnaz Akrami, and Chengkai Li. A benchmarking study of embedding-based entity alignment for knowledge graphs. *PVLDB*, 13(11):2326–2340, 2020.

[Sun *et al.*, 2023a] Zequn Sun, Jiacheng Huang, Jinghao Lin, Xiaozhou Xu, Qijin Chen, and Wei Hu. Joint pre-training and local re-training: Transferable representation learning on multi-source knowledge graphs. In *KDD*, pages 2132–2144, 2023.

[Sun *et al.*, 2023b] Zequn Sun, Jiacheng Huang, Jinghao Lin, Xiaozhou Xu, Qijin Chen, and Wei Hu. Joint pre-training and local re-training: Transferable representation learning on multi-source knowledge graphs. In *KDD*, pages 2132–2144, Long Beach, CA, USA, 2023. ACM.

[Wang *et al.*, 2018] Zhichun Wang, Qingsong Lv, Xiaohan Lan, and Yu Zhang. Cross-lingual knowledge graph alignment via graph convolutional networks. In *EMNLP*, pages 349–357, Brussels, Belgium, 2018. ACL.

[Wang *et al.*, 2022] Yuxin Wang, Yuanning Cui, Wenqiang Liu, Zequn Sun, Yiqiao Jiang, Kexin Han, and Wei Hu. Facing changes: Continual entity alignment for growing knowledge graphs. In *ISWC*, volume 13489 of *Lecture Notes in Computer Science*, pages 196–213, 2022.

[Wang *et al.*, 2025] Yuanyi Wang, Han Li, Haifeng Sun, Lei Zhang, Bo He, Wei Tang, Tianhao Yan, Qi Qi, and Jingyu Wang. Rethinking smoothness for fast and adaptable entity alignment decoding. In *Findings of the Association for Computational Linguistics: NAACL 2025*, pages 4521–4535, 2025.

[Xin *et al.*, 2022] Kexuan Xin, Zequn Sun, Wen Hua, Wei Hu, and Xiaofang Zhou. Informed multi-context entity alignment. In *WSDM*, pages 1197–1205, 2022.

[Zeng *et al.*, 2021] Kaisheng Zeng, Chengjiang Li, Lei Hou, Juanzi Li, and Ling Feng. A comprehensive survey of entity alignment for knowledge graphs. *AI Open*, 2:1–13, 2021.

[Zhang and Yao, 2022] Yongqi Zhang and Quanming Yao. Knowledge graph reasoning with relational digraph. In *WWW*, pages 912–924, 2022.

[Zhang *et al.*, 2022] Rui Zhang, Bayu Distiawan Trisedya, Miao Li, Yong Jiang, and Jianzhong Qi. A benchmark and comprehensive survey on knowledge graph entity alignment via representation learning. *The VLDB Journal*, 31(5):1143–1168, 2022.

[Zhang *et al.*, 2025] Yucheng Zhang, Beatrice Bevilacqua, Mikhail Galkin, and Bruno Ribeiro. TRIX: A more expressive model for zero-shot domain transfer in knowledge graphs. *CoRR*, abs/2502.19512, 2025.

[Zhu *et al.*, 2017] Hao Zhu, Ruobing Xie, Zhiyuan Liu, and Maosong Sun. Iterative entity alignment via joint knowledge embeddings. In *IJCAI*, pages 4258–4264, Melbourne, Australia, 2017. ijcai.org.

[Zhu *et al.*, 2021] Zhaocheng Zhu, Zuobai Zhang, Louis-Pascal A. C. Xhonneux, and Jian Tang. Neural bellman-ford networks: A general graph neural network framework for link prediction. In *NeurIPS*, pages 29476–29490, 2021.