Send Orders for Reprints to reprints@benthamscience.net

Current Bioinformatics, **xxxx,** *xx,* **x-x**

RESEARCH ARTICLE

From Structure to Semantics: Training-Free LLM Guidance for Biomedical Knowledge Graph Fusion

Xiaosong Han[1], Xindi Dai[1], Xibei Li[2], Renchu Guan[1,*] and Xiaoyue Feng[1,*]

[1]*Key Laboratory of Symbolic Computation and Knowledge Engineering of the Ministry of Education, College of Computer Science and Technology, Jilin University, Changchun, China;* [2]*College of Software, Jilin University, Changchun, China*

ARTICLE HISTORY

Received: October 14, 2025
Revised: February 24, 2026
Accepted: March 06, 2026

DOI:
10.2174/0115748936456137260505044312

Abstract: ***Introduction:*** Entity alignment and link prediction in biomedical knowledge graphs are challenged by sparsity, heterogeneity, and evolution, requiring methods that work in cold-start settings without retraining. We introduce LLM-KG2F, a training-free framework that leverages Large Language Models (LLMs) for semantic reasoning to overcome these issues.

Materials and Methods: LLM-KG2F combines BootEA-TransH structural embeddings with a zero-shot LLM plausibility prior. Without fine-tuning the LLM, it uses a chunk–reason–verify inference procedure. We evaluated the framework on three biomedical knowledge graphs: MED-BBK-9K, CKG, and RNA-KG.

Results: LLM-KG2F demonstrated strong link prediction performance, achieving Hits@10 scores of 88.78 on MED-BBK-9K and 83.92 on CKG. It also produced competitive results on the RNA-KG interaction graph.

Discussion: By injecting a training-free LLM prior into a structural fusion pipeline, LLM-KG2F effectively addresses the challenges of data sparsity and imbalance in biomedical knowledge graph fusion. The method's ability to operate in a cold-start setting makes it particularly suitable for the dynamic and evolving nature of biomedical data, eliminating the need for constant retraining.

Conclusion: By integrating a training-free LLM prior into a structural fusion pipeline, LLM-KG2F effectively addresses data sparsity and supports cold-start link prediction. This makes the framework a robust and scalable solution for intelligent analysis in dynamic biomedical domains, eliminating the need for constant retraining.

Keywords: Knowledge graph, link prediction, entity alignment, knowledge fusion, large language model, graph neural networks.

1. INTRODUCTION

Knowledge Graphs (KGs) have become essential structures for organizing and representing complex information, particularly within the biomedical domain [1, 2]. Concurrently, RNA sequencing (RNA-seq) has revolutionized transcriptomics, providing unprecedented insights into cellular functions through the analysis of both coding and non-coding RNAs [3]. The combination of these two fields presents a powerful opportunity: by structuring RNA-seq data into KGs, we can accelerate discoveries in gene regulation, disease bio-marker identification, and therapeutic development [4-7]. However, constructing accurate and comprehensive biomedical KGs, such as the RNA-KG [8], poses significant challenges. The underlying knowledge sources are often diverse and complex, and suffer from a scarcity of high-quality, well-labeled data. This often results in incomplete KGs, making it difficult for downstream models to generalize effectively.

While knowledge graph fusion—combining information from multiple KGs—presents a promising solution to data sparsity issues [9, 10], it requires efficient methods for Entity Alignment (EA) and Link Prediction (LP). Traditional methods, such as TransE [11] for entity alignment and Graph Neural Networks (GNNs) [12] for link prediction, have made important contributions but struggle in the dynamic and heterogeneous biomedical context. These methods often fail to enrich KGs with new relational triples, which are crucial for evolving biomedical graphs [10]. Furthermore, the static nature of many traditional EA models makes it difficult for them to adapt to continuously expanding biomedical knowledge, often requiring costly retraining and failing to identify new alignments as the graph evolves [13].

*Address correspondence to these authors at the Key Laboratory of Symbolic Computation and Knowledge Engineering of the Ministry of Education, College of Computer Science and Technology, Jilin University, Changchun, China; E-mails: fengxy@jlu.edu.cn; guanrenchu@jlu.edu.cn

Existing link prediction methods, including those adapted to sparse data scenarios, also struggle with reasoning about complex biological interactions. Many models are confined to one-hop neighborhoods, which limits their ability to infer multi-hop, higher-order relationships essential for understanding complex biological pathways [14]. These limitations are even more pronounced when considering the nuances of RNA interactions, where structural flexibility and the scarcity of training data for rare or "orphan" molecules further complicate the prediction of unknown relationships [15-17].

To address these limitations, we introduce the LLM-enhanced Knowledge Graph Fusion Framework (LLM-KG2F), which leverages the advanced reasoning and semantic understanding capabilities of Large Language Models (LLMs) to optimize both entity alignment and link prediction in biomedical KGs. This novel approach is specifically designed to excel without relying on extensive retraining or large, manually curated datasets, making it particularly well-suited for cold-start scenarios where traditional methods falter. LLMs, which have been pretrained on vast amounts of general knowledge, offer the unique advantage of zero-shot learning. This allows our framework to perform knowledge graph fusion tasks effectively, even when minimal or no labeled data is available. The LLM-KG2F framework utilizes optimized prompt engineering to enhance both entity alignment and link prediction, thereby significantly reducing the need for task-specific retraining.

Our main contributions are summarized as follows:

1) We propose a novel framework, LLM-KG2F, designed for cold-start knowledge graph fusion, composed of two core modules: an entity alignment module (LLM-KG2F-EA) and a link prediction module (LLM-KG2F-LP).
2) For entity alignment, the LLM-KG2F-EA module introduces a weighted ternary score function, integrating traditional embedding-based scores with semantic plausibility scores from a Large Language Model (LLM). This enhances the semantic integration of node embeddings, especially in sparse biomedical datasets like RNA-Seq data.
3) For link prediction, the LLM-KG2F-LP module employs a chunk-reason-verify pipeline, driven by an LLM. This method is specifically designed to predict complex semantic and logical links, crucial for understanding intricate relationships between genes and non-coding RNAs in RNA-seq data.
4) Extensive experiments conducted on three biomedical datasets demonstrate that LLM-KG2F significantly outperforms existing state-of-the-art methods in cold-start conditions involving biomedical entities.

2. RELATED WORK

2.1. Biomedical Knowledge Graphs

The use of Knowledge Graphs (KGs) in the biomedical domain has seen rapid growth, leading to the development of numerous domain-specific KGs that represent critical biomedical information about diseases, genes, drugs, and increasingly RNA-related data. These KGs are crucial for applications such as clinical decision support [18-20], drug discovery [21-25], intelligent search, and question-answering systems [26, 27]. Some notable examples include dRiskKB for disease relationships [28], the Clinical Knowledge Graph (CKG) for integrating clinical data with proteomics [29], and the MED-BBK graph, which provides detailed knowledge about diseases and drugs [30, 31]. More recently, the RNA-KG was introduced to specifically represent coding and non-coding RNA molecules [32-34].

The integration of these KGs remains a significant challenge, especially due to the inherent heterogeneity of the data sources and common issues with data sparsity. The integration of multiple KGs into a unified, comprehensive graph requires methods that can handle complex relationships and diverse entity types while maintaining semantic coherence. This is where knowledge graph fusion—which aims to integrate knowledge from different graphs—becomes essential. Fusion is particularly critical in biomedical contexts, where the relationships and entities may evolve quickly, and graphs may contain incomplete or missing information due to the sparsity of annotated data [35, 36]. Recent advancements like FuseLinker [37] and BioKGC [38] have pushed the boundaries of biomedical KG completion through multimodal feature fusion and path-based reasoning. While these models achieve high performance, they typically depend on supervised fine-tuning or the availability of rich domain-specific ontologies. LLM-KG2F complements these approaches by demonstrating that a zero-shot LLM prior, when integrated with a structural backbone like TransH/BootEA, can achieve competitive performance (Hits@10 = 88.78 on MED-BBK-9K) without any additional training. This validates the efficacy of using LLMs as external semantic reasoners in data-scarce environments where supervised methods may falter.

2.2. Entity Alignment

Entity Alignment (EA) is a key task in knowledge graph fusion. It involves identifying equivalent entities across different KGs, a task that is particularly challenging when the graphs are heterogeneous and contain noisy or incomplete data. Early methods for entity alignment were primarily based on syntactic and structural similarities. However, embedding-based methods, such as TransE [9, 11] and its extensions like TransH [39], have become dominant [40]. These methods learn vector representations of entities and relations, allowing for similarity measurements between entities.

Despite their successes, traditional embedding-based approaches often struggle in the biomedical domain due to the need for semantic integration and the lack of scalability to large, evolving datasets [13, 41-43]. BootEA [43], a popular bootstrap-ping method for entity alignment, introduces iterative learning to refine alignment predictions, but it still relies heavily on manually curated or pre-processed data. More recent methods like RDGCN have incorporated attention

mechanisms to capture more complex relationships and patterns in the data [44-46]. However, these methods are often limited by the need for extensive retraining and manual supervision.

By contrast, our LLM-KG2F-EA module introduces a novel method that integrates LLMs' semantic capabilities to enhance entity alignment, particularly in scenarios with sparse or incomplete data [47]. The LLM-derived semantic plausibility scores allow our framework to perform entity alignment without the need for large labeled datasets, offering a significant advantage over traditional embedding-based methods.

2.3. Link Prediction

Link Prediction (LP) aims to predict missing links or relationships in a KG, addressing the inherent incompleteness of graphs. Traditional LP methods, including matrix factorization and GNN-based models [10, 12, 48-51], have been effective in various domains [14]. However, in the biomedical domain, these methods often struggle with the heterogeneous nature of the data, where entities and relations can vary drastically in their semantics and structure. Moreover, many of these methods are limited by their reliance on local structure (such as one-hop neighborhoods) and fail to capture higher-order dependencies that are critical for modeling complex biological interactions [52-56].

To address these challenges, several recent advancements in LP have explored semantic embeddings and attention mechanisms. Approaches like Graph Attention Networks (GAT) [57, 58] and BERT-based models [37, 59] have made significant strides in improving the ability of LP models to reason about complex relationships. Despite these improvements, current methods still face challenges when dealing with data sparsity, particularly in biomedical KGs where rare interactions are underrepresented [60-63].

Our LLM-KG2F-LP module leverages LLMs' advanced semantic understanding to model complex relationships in biomedical KGs. The chunk-reason-verify pipeline enables the prediction of missing links by reasoning over entity pairs, with LLMs providing the contextual insights needed to infer complex biological relationships. This approach moves beyond traditional LP methods by introducing a semantic reasoning layer, allowing the model to understand and predict multi-hop, higher-order relationships with high accuracy.

2.4. Large Language Models in Biomedical KGs

The application of Large Language Models (LLMs), such as GPT-4 [64], in biomedical knowledge graph tasks is an exciting and promising development. LLMs are particularly well-suited for tasks that require deep semantic understanding and contextual reasoning, thanks to their vast knowledge base, learned from diverse and extensive datasets. LLMs have already demonstrated their potential in various Natural Language Processing (NLP) tasks, including question-answering, information extraction, and text generation [64]. These models, which use the Transformer architecture [58], excel at capturing long-range dependencies and complex relationships, making them ideal candidates for knowledge graph fusion tasks.

Recent studies have explored the use of LLMs for zero-shot learning in the context of biomedical KGs [65]. For example, BioBERT [66] and other LLM-based approaches have been applied to tasks like gene-disease association prediction and drug repurposing [64, 67]. These methods leverage the pretrained knowledge of LLMs to generate high-quality embeddings and predict missing links in biomedical graphs [68-70]. The ability to use LLMs for cold-start tasks, without the need for retraining on task-specific data, represents a significant step forward in biomedical KG research [71-73].

Our LLM-KG2F framework extends this concept by integrating LLMs directly into both entity alignment and link prediction tasks. This allows us to leverage LLMs' powerful semantic reasoning capabilities to enhance knowledge graph fusion, even in situations where data is sparse or incomplete. By utilizing the zero-shot learning capabilities of LLMs, our approach provides a scalable and efficient solution to the problem of data sparsity in biomedical KGs.

3. METHOD

In Fig. (**1**), we present LLM-KG2F, a plug-and-play framework that augments classical structural embedding pipelines with a training-free LLM prior. The key idea is to inject zero-shot, semantics-aware scoring and reasoning from a pretrained LLM into EA and LP, without additional task-specific re-training. Concretely, LLM-KG2F combines a structural score learned by a standard translational model with an LLM-derived plausibility score for triples, and employs a chunk–reason–verify procedure to propose and validate new links. This design targets cold-start knowledge graph fusion, where labeled supervision or retraining opportunities are limited, and graphs evolve continuously.

3.1. Problem Definition

The task definition for KG fusion is as follows.

Let $KG_1 = (E_1, R_1, T_1)$ and $KG_2 = (E_2, R_2, T_2)$ be two Knowledge Graphs (KGs). Here, E_1 and E_2 are the entity sets; R_1 and R_2 are the relation sets; and $\mathrm{T}_1 \subseteq \mathrm{E}_1 \times \mathrm{R}_1 \times \mathrm{E}_1$ and $T_2 \subseteq E_2 \times R_2 \times E_2$ are the triple sets. The goal of KG fusion is to construct a new KG, $KG_F = (E_F, R_F, T_F)$ by the following process:

(i) Entity Alignment: The objective of entity alignment is to determine a function $f_E: E_1 \cup E_2 \rightarrow E_F$ that maps entities from the original knowledge graphs to a unified space, such that the resulting aligned entity set E_F contains all matched entities, as expressed in Eq. (1).

$$E_F = f_E(E_1) \cup f_E(E_2) \quad (1)$$

(ii) Link Prediction: After aligning entities, we merge the triple sets $T_1$and T_2 from the two knowledge graphs into a unified triple set T_F, and search for new triples $t_F =$

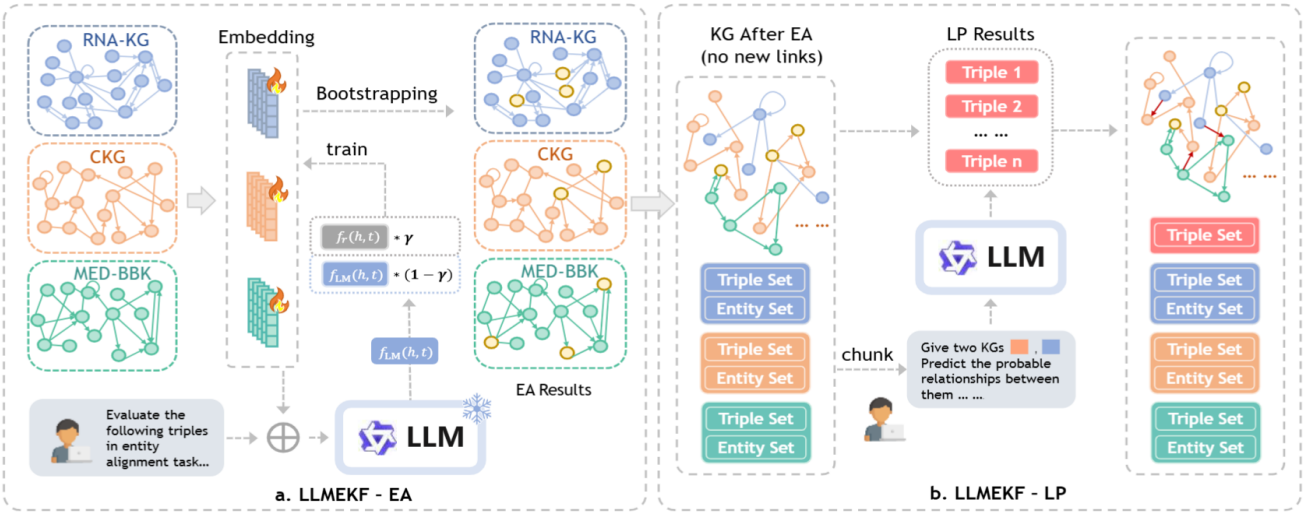

Fig. (1). LLM-KG2F framework. (**a**) LLM-KG2F-EA aligns entities across KGs by combining TransH embedding scores with LLM-derived triple scores, which are iteratively refined using the BootEA paradigm. (**b**) LLM-KG2F-LP first chunks the KG to manage the LLM context. An LLM then reasons over pairs of these chunks to predict new links, which are subsequently verified to eliminate inconsistencies before augmenting the KG. (*A higher resolution / colour version of this figure is available in the electronic copy of the article*).

$\langle e_1, r, e_2 \rangle$ such that either $e_1 \in E_1, e_2 \in E_2$ or $e_1 \in E_2$, $e_2 \in E_1$ where the alignment mapping defined in Eq. (1) ensures the semantic consistency of cross-graph relations.

Our work focuses first on EA—crucial for subsequent LP—and augments a classic bootstrapping framework with an LLM-derived semantic prior. Unlike approaches that depend on task-specific retraining or dense supervision, our design explicitly supports cold-start fusion: it can be deployed with no additional model training, using the LLM purely as a zero-shot plausibility scorer and reasoner.

3.2. LLM-KG2F-EA: Training-Free, LLM-Augmented Entity Alignment

We enhance bootstrapped EA with a hybrid triple scoring mechanism that fuses (i) a structural score from a translational embedding model and (ii) a training-free plausibility score produced by a pretrained LLM.

Structural backbone. We adopt TransH [28] as the structural component due to its ability to represent entities and relations on relation-specific hyperplanes, capturing multi-relational semantics effectively. Let h, t be entity embeddings, r the translation vector, and w_r the normal of the relation-specific hyperplane; the TransH-based score $f(\tau)$ is defined as Eq. (2).

$$f(\tau) = |(h - w_r^\top h w_r) + d_r - (t - w_r^\top t w_r)|_2^2 \qquad (2)$$

LLM plausibility prior. Given a triple $\tau = (h, r, t)$, we query a pretrained LLM with a concise prompt to obtain a scalar plausibility score $f_{LM}(\tau)$. Importantly, the LLM is not fine-tuned for our task; it contributes a zero-shot semantic prior distilled from broad pretraining.

Weighted ternary fusion. We define a fused score

$$g(\tau) = \gamma f(\tau) + (1 - \gamma) f_{LM}(\tau), \qquad (3)$$

with $\gamma \in [0,1]$ controlling the contribution of the training-free LLM prior (Eq. 3). The structural TransH score and the LLM plausibility score may have different scales and even opposite directions (*e.g.*, distance vs. plausibility). We therefore first convert the structural distance into a similarity score by negation, and then normalize both scores within each candidate set to obtain comparable values in $[0, 1]$. Concretely, for each query-specific candidate set, we apply a rank-based (percentile) normalization $\tilde{s} = \mathrm{RankNorm}(f(\tau))$, which is robust to outliers and avoids distributional assumptions. The final fused score is computed as $g(\tau) = \gamma \tilde{s}_{\mathrm{struct}} + (1 - \gamma)\tilde{s}_{\mathrm{LLM}}$. The optimization objective O_e remains margin-based, contrasting positive vs. adversarially sampled negatives, and we keep the truncated uniform negative sampling for hard negatives as in the original, as expressed in Eq. (4).

$$O_e = \sum_{\tau \in T^+} [g(\tau) - \gamma_1]_+ + \mu_1 \sum_{\tau' \in T^-} [\gamma_2 - g(\tau')]_+, \qquad (4)$$

where $[\cdot]_+ = \max(\cdot, 0)$. Here, γ_1 and γ_2 are margin hyperparameters that specify the desired score thresholds, aiming for positive triple scores $g(\tau) \le \gamma_1$ and negative triple scores $g(\tau') \ge \gamma_2$, respectively; μ_1 balances the contributions from positive (T^+) and negative (T^-) triple sets. To enhance discriminative capability, negative triples T^- are generated using ϵ-truncated uniform negative sampling, creating structurally similar but semantically incorrect examples.

Bootstrapped refinement with plug-in LLM prior. Within each BootEA-style iteration [18], the alignment probability between unaligned entity pairs (x, y) is calculated as $\pi(y|x; \theta) = \sigma\left(\mathrm{sim}\left(v(x), v(y)\right)\right)$, where $\sigma(\cdot)$ is the sigmoid function, and $\mathrm{sim}(\cdot,\cdot)$ denotes cosine similarity between entity embeddings $v(x)$ and $v(y)$. To inject the training-free LLM prior without retraining, we compute a prior score $\psi(x, y)$ from the fused triple score $g(\tau) = f(\tau) +$

$(1-\gamma)f_{LM}(\tau)$ around $(x,y)$and re-rank candidates *via* a light reweighting,

$$\pi^{\star}(y|x;\theta^{(t)}) = S_y(\pi(y|x;\theta^{(t)}) \cdot (1+\lambda\psi(x,y))), \quad (5)$$

which keeps the BootEA objective unchanged and preserves stability while adding a zero-shot semantic bias. In Eq. (5), $S(\cdot)$ denotes sigmoid function.

Alignment conflicts are resolved by computing the alignment probability difference, $\Delta^{(t)}_{(x,y,y')} = \pi^{*}\left(y\middle|x;\theta^{(t)}\right) - \pi^{*}\left(y'\middle|x;\theta^{(t)}\right)$. A positive $\Delta^{(t)}_{(x,y,y')}$ indicates that aligning x with y is preferred over y'. Similar strategies address other types of conflicts.

To further refine performance, a joint optimization objective $O = O_e + \mu_2 \cdot O_a$ is defined, combining the EA loss O_e with an alignment prediction loss O_a, where μ_2 is a balancing hyperparameter. The alignment prediction loss O_a is the negative log-likelihood over alignment probabilities as Eq. (6):

$$O_a = -\sum_{x\in X}\sum_{y\in Y}\phi_x(y)\log\pi(y|x;\theta), \quad (6)$$

where $\phi_x(y)$ represents the ground-truth distribution for candidate alignments of x.

Prompt template. We retain the concise, numeric-output template to ensure deterministic parsing, shown in Fig. (**2**).

Evaluate the following ternary in the entity alignment task, the ternary is surrounded by three '#'s, and all you need to do is simply give a ***ternary score*** *greater than 0 that evaluates the entity and relationship relevance of the ternary. When the difference between the head and tail entity embedding vectors is large, the value of the score increases, with no explicit upper limit. Output only a floating point number, the score of the ternary, without interpreting this. Here is the triple to evaluate:###(h,r,t)###*

Fig. (2). EA prompt template. (*A higher resolution / colour version of this figure is available in the electronic copy of the article*).

3.3. LLM-KG2F-LP: Chunk-Reason-Verify for Cold-Start Link Discovery

After alignment, we propose new links using a three-stage chunk-reason-verify pipeline that keeps LLM context manageable while enforcing symbolic consistency.

Chunk. Because the LLM context is finite, the two subgraphs KG_1 and KG_2, which have already undergone entity alignment, are not processed as a whole but are divided into m and n smaller chunks, respectively. Specifically, KG_1 is split into $\mathcal{C}_{11}, \mathcal{C}_{12}, \ldots, \mathcal{C}_{1m}$, and KG_2 into $\mathcal{C}_{21}, \mathcal{C}_{22}, \ldots, \mathcal{C}_{2n}$. The selection criteria for chunking focus on neighborhood preservation: we group entities based on their 1-hop connections to maintain local semantic clusters. During the reasoning stage, we iterate through pairs of chunks to predict potential cross-graph links, ensuring that the LLM focuses on a manageable set of typed node summaries and candidate relation schemas.

Reason. For each pair $(\mathcal{C}_i, \mathcal{C}_j)$, we construct a structured prompt that includes: (i) typed node summaries, (ii) candidate relation schema $\mathcal{R}_i \cup \mathcal{R}_j$, and (iii) output constraints (triples only, one per line). The pretrained LLM then proposes triples across chunks in a training-free, zero-shot manner, leveraging its semantic prior to infer multi-hop or schema-aware relations that may not be obvious from local structure alone.

Verify. We filter proposed triples *via* rule- and embedding-based checks: **(i) Type and schema constraints:** domain compatibility; **(ii) Redundancy and symmetry/anti-symmetry checks:** avoid duplicates and enforce relation properties; **(iii) Structural sanity:** accept only triples that improve connectedness without violating constraints; **(iv) Score cross-validation:** optional re-scoring with the fused $g(\cdot)$ to maintain agreement between structural and LLM priors.

Output protocol. We keep the strict format "<h,r,t>\t \<h,r,t>" to simplify parsing at scale and reduce post-processing errors—another practical advantage in cold-start deployments where guardrails need to be robust without task-specific tuning. The prediction prompt is formulated as Fig. (**3**).

You will be given two knowledge graphs enclosed in three '#'s, and the set of possible relationships. Give the trio of relation-ships with possible head entities from Graph1 and tail entities from Graph2, and vice versa. Use the '<h,r,t>\t<h,r,t>' format for output, which outputs only the triples without interpretation.

Fig. (3). LP prompt template. (*A higher resolution / colour version of this figure is available in the electronic copy of the article*).

4. EXPERIMENTS

4.1. Data Preprocessing and Summary

4.1.1. MED-BBK

MED-BBK [27] is built upon multiple authoritative medical resources, offering fine-grained knowledge about diseases, symptoms, and drugs. Table **1** summarizes MED-BBK-9K. It contains relationship triples linking biomedical entities and attribute triples associating entities with literal values ("#" denotes counts; "Ratio" is triples-to-entities). Relationships correspond to object properties connecting two entities, while attributes correspond to data properties that associate an entity with a typed value.

4.1.2. RNA-KG

RNA-KG [13] is the first ontology-based KG representing coding and non-coding RNA molecules, their interac-

Table 1. Statistics of MED-BBK-9K.

Name	KGs	#Ent	Relation			Attribute		
			#Rel	#Tri	Ratio	#Rel	#Tri	Ratio
MED-BBK-9K	MED	9162	32	158357	17.28	19	11467	1.25
	BBK	9162	20	50307	5.49	21	44987	4.91

Table 2. Detailed statistics of RNA-KG.

Attributes	Quantity
#Nodes	578,384
#Directed Edges	8,768,582
Max Out Degree	27,109
Max In Degree	117,135
Mean Degree	9.65
Graph Diameter	36
Representative RNA Types	miRNA, lncRNA, circRNA, mRNA, tRNA, snoRNA, ASO
Representative Data Sources	miRBase, miRTarBase, LncBook, DrugBank, snoDB
Relation Types	miRNA–mRNA, lncRNA–protein, RNA–disease, RNA–chemical

tions with other biomolecules, pathways, abnormal phenotypes, and diseases, with RDF triples extracted from 50+ public sources [13, 43-45]. Table **2** reports the relevant statistics. Built with PheKnowLator [46], RNA-KG currently has ~600K nodes and ~9M edges and supports multiple formats. Because the number of edges is far smaller than the square of the number of nodes, RNA-KG is intrinsically label-sparse, creating a cold-start setting where many plausible relations are unobserved and must be inferred without task-specific supervision.

4.1.3. CKG and Construction of a Cold-Start Split

The Clinical Knowledge Graph (CKG) [26] integrates proteomics data from public databases, user experiments, biomedical ontologies, and scientific publications to link clinical, laboratory, and imaging data.

To systematically evaluate entity alignment and link prediction under cold-start (label-sparse, training-free) conditions, we construct a cold-start CKG split. The sampling pipeline (Fig. **4**) follows the original design:

(i) Starting from a random entity, we run BFS on the original CKG and stop after collecting 10,000 triples, constraining the number of triples per relation type by a preset threshold to keep type coverage balanced and to avoid dominance by very frequent relations.

(ii) The sampled graph is then randomly partitioned into two disjoint subgraphs, KG1 and KG2. All cross-partition triples $(h \in \mathrm{KG1}, t \in \mathrm{KG2})$or $(h \in \mathrm{KG2}, t \in \mathrm{KG1})$are removed to emulate unseen cross-graph links at test time.

(iii) For controlled evaluation, each relation type is restricted to a small, fixed number of examples (we keep 5 "support" triples for seeding and 20 "query" triples for held-out evaluation per relation).

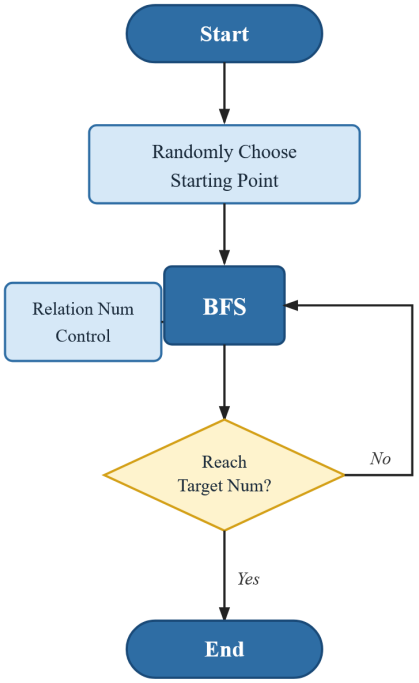

Fig. (4). Data processing pipeline for constructing the cold-start CKG dataset. (*A higher resolution / colour version of this figure is available in the electronic copy of the article*).

Table **3** lists the resulting statistics. Fig. (**5**) shows the natural long-tail distribution: a few relations contain most triples, while many relations are rare—exactly the regime where training-free LLM priors are valuable.

Table 3. Statistics of cold-start CKG subgraphs.

Subgraph	#Ent	#Rel	Avg. Triples per Relation	#Triple
KG1	8500	20	5 (support) + 20 (query)	500
KG2	8500	20	5 (support) + 20 (query)	500

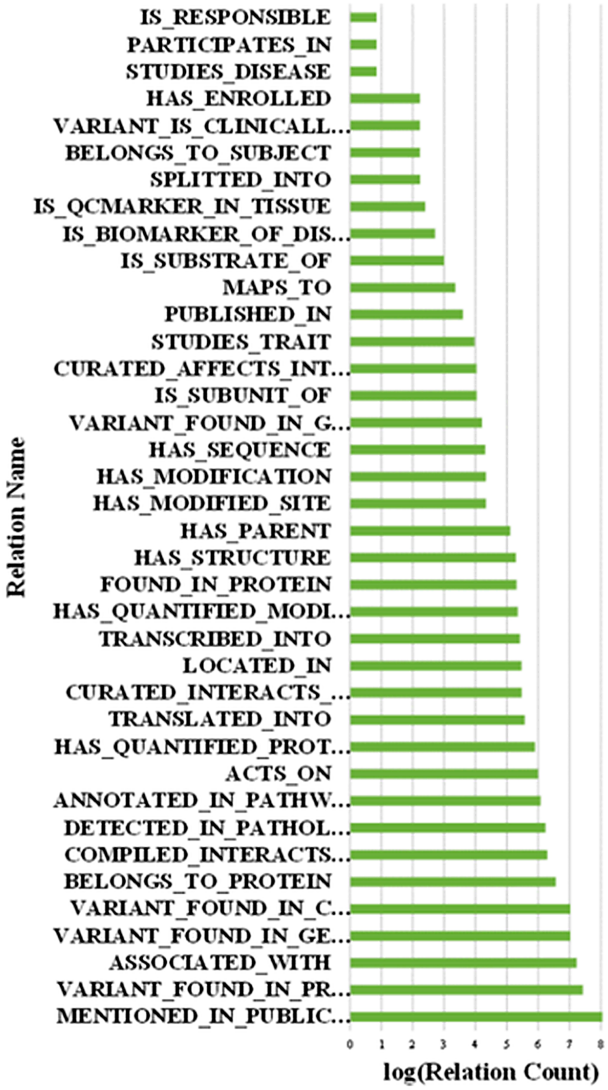

Fig. (5). Relation distribution of the original CKG (log scale). (*A higher resolution / colour version of this figure is available in the electronic copy of the article*).

4.2. Implementation Details

The experiments were conducted on a server with 10 * RTX 3090 and 512 GB of RAM. The software environment was based on Python 3.7.16. No fine-tuning of the LLM is performed anywhere—the LLM is used strictly as a training-free plausibility scorer/reasoner *via* fixed prompts. For the structural EA backbone, we adopt BootEA-TransH. Key hyperparameters are unchanged and summarized in Table **4**, covering embedding initialization, loss settings, and evaluation strategies.

Table 4. Important hyperparameters for BootEA-TransH.

Hyperparameter	Value
Embedding Dimension (d)	300
Initialization Method	Xavier Initialization
Loss Normalization	L2 Norm
L2 Normalization	True
Optimizer	Adam
Learning Rate	0.001
Batch Size	3000
Margin (for Loss)	1.5
Negative Sampling Strategy	Truncated Sampling
Negative Triples per Positive	10
Truncated Sampling Epsilon	0.98
Maximum Epochs	1000
LLM Temperature	1.0
LLM top_p	0.95
LLM max_new_tokens	128

5. RESULTS

5.1. Entity Alignment

We compared our Entity Alignment (EA) method with four representative baseline methods: TransE, TransH, BootEA, RDGCN, BioKGC, and FuseLinker. All experiments were conducted on three datasets: MED-BBK-9K, RNA-KG, and our constructed cold-start CKG dataset (Section 4.1). The results are shown in Table **5**.

On the Cold-Start CKG dataset, LLM-KG2F reaches Hits@10 = 83.92 and MRR = 0.65, outperforming TransE (58.06 / 0.41) and TransH (68.34 / 0.51). On MED-BBK-9K, LLM-KG2F achieves Hits@10 = 88.78 and MRR = 0.68, beating BioKGC and FuseLinker, which are the SOTA methods; RNA-KG shows the same trend, confirming robustness across graphs of different scales and heterogeneity.

To validate the effectiveness of LLM-KG2F, we conducted detailed ablation studies. In Table **5**, "w/o TransH" represents removing the structural embedding; "w/o LLM" disables the training-free LLM prior in scoring; LLM-KG2F combines both with a weighted fusion mechanism ($\gamma = 0.7$). Adding the training-free LLM prior consistently boosts EA, with the largest gain on Cold-Start CKG, where supervision is intentionally sparse. This supports our claim that a zero-shot semantic prior is most beneficial under label-sparse, evolving conditions.

Table 5. Comparison and ablation results for entity alignment.

Model	Cold-Start CKG			MED-BBK-9K			RNA-KG		
	Hits@10	MR	MRR	Hits@10	MR	MRR	Hits@10	MR	MRR
TransE	58.06	119.64	0.41	76.75	147.44	0.36	55.21	167.02	0.27
TransH	68.34	112.32	0.51	69.41	111.62	0.51	56.03	163.27	0.27
BootEA-TransE	75.62	199.66	0.61	76.97	188.02	0.55	60.15	159.33	0.29
RDGCN	52.22	156.25	0.36	63.31	132.79	0.32	55.89	162.89	0.24
BootEA	72.89	303.81	0.63	80.05	277.39	0.66	58.05	129.26	0.52
BioKGC	80.12	125.40	0.62	85.45	130.22	0.65	63.20	140.55	0.51
FuseLinker	81.56	118.90	0.63	86.80	121.45	0.66	65.42	128.10	0.53
w/o TransH	82.45	124.47	0.59	88.31	117.13	0.66	66.13	125.37	0.55
w/o LLM	76.18	190.43	0.61	85.1	179.41	0.68	64.76	142.91	0.5
LLM-KG2F	**83.92**	**112.11**	**0.65**	**88.78**	**108.01**	**0.68**	**68.77**	**110.66**	**0.57**

Table 6. Link prediction results on cold-start CKG dataset.

Method	Hits@10	MR	MRR
CN	39.28	1087.2	0.23
PA	41.68	1023.94	0.27
G2G	53.42	536.80	0.38
DEAL	53.19	556.91	0.36
LLM-KG2F	**64.98**	**180.9**	**0.51**

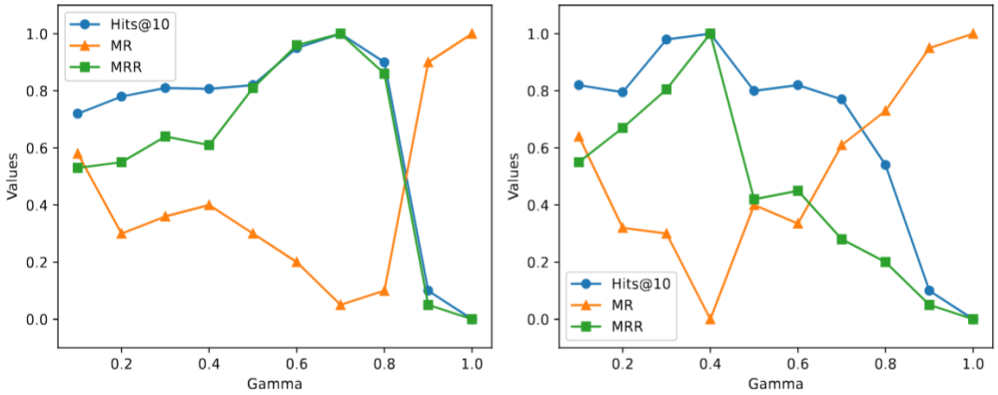

Fig. (6). Line chart of parameter discussion. The x-axis denotes γ, and the y-axis reports the (rescaled) evaluation metrics. Higher is better for Hits@10 and MRR, while lower is better for MR. (**a**) CKG results. (**b**) MED-BBK results. (*A higher resolution / colour version of this figure is available in the electronic copy of the article*).

5.2. Link Prediction

To assess generalization after alignment, we evaluate Link Prediction (LP) on the Cold-Start CKG with the chunk–reason–verify pipeline (§3.3). We compare against Common Neighbors (CN), Preferential Attachment (PA), Graph2Gauss (G2G), and DEAL. Table **6** shows that LLM-KG2F attains Hits@10 = 64.98 and MRR = 0.51, outperforming all baselines by a clear margin. Importantly, the LLM is not fine-tuned—all proposals are generated zero-shot and admitted only after schema/consistency verification—reinforcing the practicality of the training-free setting.

5.3. Parameter Discussion

We study the effect of the fusion weight γthat balances structural and LLM scores (§3.2). Consistent with Fig. (**6**), the Cold-Start CKG benefits from a larger LLM contribution (best performance around the setting used in Table **5**, $\gamma \approx 0.7$); overly high values eventually degrade results due to over-reliance on language priors. In contrast, on MED-BBK-9K (denser supervision), a moderate γ is preferable, with higher γagain leading to a decline. This pattern indicates that the LLM prior is most helpful when supervision is scarce, while structural cues dominate when labels are richer.

Table 7. Results of comparative experiment on EA and LP tasks.

LLM	EA Tasks			LP Tasks		
	Hits@10	**MR**	**MRR**	**Hits@10**	**MR**	**MRR**
Llama3-8B-Instruct	80.05	**130.30**	**0.64**	60.04	**162.88**	0.48
Gemma 7B	78.67	250.20	0.63	61.36	305.24	0.49
Baichuan2-7B-Base	74.50	290.00	0.61	61.09	342.20	0.50
Phi-3-small 7B	79.78	180.50	0.63	58.04	229.04	0.46
Qwen2-7B	**81.23**	150.75	**0.64**	**64.98**	180.90	**0.51**

The fusion weight serves as a control knob to balance the influence of structural embeddings ($f(\tau)$) and LLM-derived semantic priors ($f_{LM}(\tau)$). We determined through a grid search on a validation split, observing that its optimal value is intrinsically linked to the graph density. In the label-sparse Cold-Start CKG, a higher contribution from the LLM ($\gamma \approx 0.7$) yields the best results, as the model relies on semantic common sense to compensate for missing structural links. Conversely, in denser graphs like MED-BBK-9K, structural cues are more reliable, warranting a more moderate. This mechanism allows LLM-KG2F to adapt to the varying quality of structural information across different biomedical datasets.

5.4. Impact Analysis of Base LLM Choice

We ablate the choice of the base LLM used as a training-free scorer/reasoner. Five open-source models are evaluated with identical prompts and no fine-tuning on the Cold-Start CKG for both EA and LP. Table **7** (unchanged numerically) shows:

(i) Qwen2-7B delivers the strongest overall performance (EA: highest Hits@10 and MRR; LP: Hits@10 = 64.98, MRR = 0.51).

(ii) Llama3-8B-Instruct is competitive, with slightly lower Hits@10 but favorable MR on some metrics.

(iii) Gemma 7B and Baichuan2-7B-Base are moderate; Phi-3-small 7B trails across metrics.

Because our pipeline treats the LLM as a plug-in prior, swapping models requires no retraining of EA/LP modules, which are useful for rapid deployment or future LLM upgrades.

5.5. Case Study

5.5.1. EA Case

5.5.1.1. Goal

The goal of this EA case study is to show how the pipeline aligns two nodes that denote the same small-molecule drug but use different regional names.

5.5.1.2. The Two Candidate Nodes

- KG1 entity: "*Acetaminophen*" (common in US usage).
- KG2 entity: "*Paracetamol*" (INN/Commonwealth usage).

5.5.1.3. Scoring

Let τ be the structural weight. The fused score is $g(\tau) = \gamma f(\tau) + (1-\gamma) f_{LM}(\tau)$, τ=0.7. For the pair Paracetamol ↔ Acetaminophen we obtain $f(\tau)$=0.78, $f_{LM}(\tau)$=0.96, then $g(\tau) = 0.7\times0.78 + 0.3\times0.96 = 0.834$.

Although the TransH score may be insufficient, the fused score successfully identified the two entities as aligned.

This example illustrates the unique strength of LLM-KG2F in leveraging LLM-based contextual understanding to compensate for the limitations of traditional embedding models, especially when handling biomedical entities with heterogeneous textual descriptions, as commonly found in RNA-seq experiment data.

5.5.2. LP Case Study

Fig. (**7**) shows LP outputs on the Cold-Start CKG after alignment. Under the chunk–reason–verify protocol, the LLM proposes structured triples (in a strict output format) such as a valid HAS_PARENT relation between "spinal cord

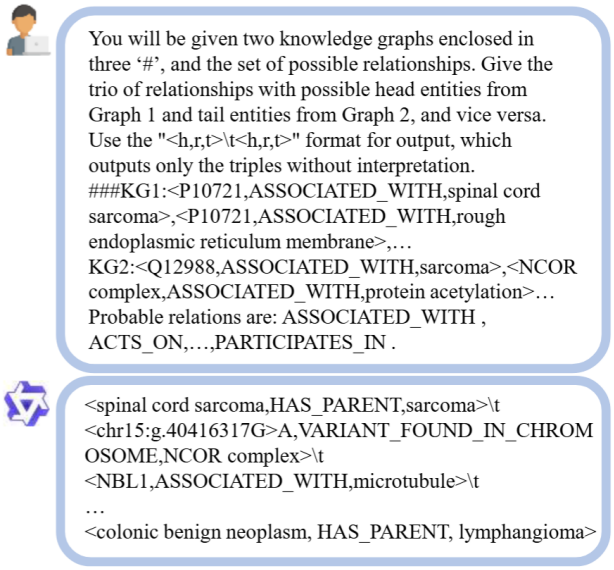

Fig. (7). LP results example in one iteration. (*A higher resolution / colour version of this figure is available in the electronic copy of the article*).

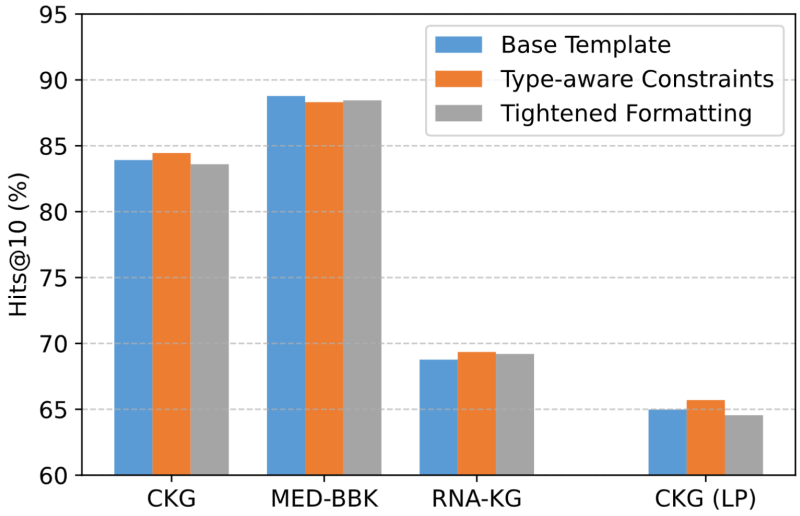

Fig. (8). Robustness to prompt design variants. (*A higher resolution / colour version of this figure is available in the electronic copy of the article*).

sarcoma" and "sarcoma." These suggestions pass type/schema checks and structural sanity filters before augmentation, demonstrating that zero-shot semantic reasoning can surface clinically meaningful edges while maintaining ontological consistency.

5.6. Prompt Robustness

To assess robustness to prompt design, we evaluated several prompt variants for both EA scoring and LP generation, including (i) the base template, (ii) adding explicit entity type/schema constraints, and (iii) tightening the numeric-output and formatting constraints. As shown in Fig. (**8**), across datasets, the overall trends remain unchanged, and performance differences are modest, indicating that LLM-KG2F is not reliant on a single "hand-crafted" wording.

6. DISCUSSION

This study introduces LLM-KG2F, a framework that integrates a training-free LLM prior into entity alignment and link prediction for biomedical knowledge graphs. In EA, LLM-KG2F augments a TransH- and BootEA-based pipeline with a zero-shot semantic plausibility signal, fused at scoring time and used throughout the bootstrapping loop (§3.2). In LP, a chunk–reason–verify protocol (§3.3) enables the LLM to propose cross-chunk links under strict schema and type constraints, after which structural checks and re-scoring prevent inconsistent augmentations. Across MED-BBK-9K, RNA-KG, and the cold-start CKG split, the method exhibits consistent gains, with the largest improvements appearing where labels are rare and retraining is impractical (§5).

From a systems perspective, the plug-and-play nature of the LLM prior is a practical advantage. The LLM is never fine-tuned; it is invoked *via* lightweight prompts to supply zero-shot semantics that complement structural embeddings. Consequently, swapping to a stronger or domain-specialized LLM requires no retraining of EA/LP modules, which is appealing for continuously evolving KGs where relations and entity vocabularies drift. The fusion weight γ (and the optional reweighting λ in §3.2) provides a simple control knob to adjust the relative influence of structure vs. semantics across datasets of different density (§5.3).

Another benefit lies in heterogeneity handling. Biomedical graphs mix entity types (genes, RNAs, diseases, chemicals) and multi-relational schemas with diverse textual aliases. The training-free LLM prior helps normalize lexical and ontological variance (*e.g.*, "hsa-miR-21-5p" *vs*. "microRNA-21"), while structural embeddings preserve graph topology. The verify stage then enforces symbolic constraints (domain/range, symmetry/antisymmetry, redundancy), reducing spurious links that pure language priors might otherwise introduce.

Our method is training-free as it requires no gradient-based optimization or fine-tuning of the LLM or structural models. However, it is not inference-free: the LLM acts as an external scorer/verifier to provide semantic plausibility signals for a filtered candidate set. Importantly, we avoid exhaustive enumeration ($O(|E|^2)$) by using structural embeddings to retrieve a Top- candidate set, leading to a complexity of approximately $O(|E| \cdot K)$. The inference cost is a controllable trade-off, adjustable via K, chunk size, and the fusion weight γ.

The primary distinction of LLM-KG2F from existing SOTA methods (*e.g.*, FuseLinker, BioKGC) lies in its training-free architecture. While traditional models require extensive labeled data and costly gradient-based fine-tuning to adapt to specific biomedical domains, LLM-KG2F leverages the zero-shot semantic prior of LLMs to perform entity alignment and link prediction. This is particularly critical in cold-start scenarios common in biomedicine, where knowledge graphs evolve rapidly, and high-quality labels for rare entities (*e.g.*, 'orphan' molecules) are scarce. Consequently, our method offers a plug-and-play paradigm that can be deployed immediately across diverse KGs without task-specific re-optimization.

FUTURE WORK

We plan to (i) develop adaptive prompting and uncertainty-aware fusion that modulate the LLM contribution per relation type and per neighborhood density; (ii) incorporate domain-grounded retrieval to cite source passages alongside LLM proposals; (iii) explore consensus/debate among multiple LLMs to reduce single-model bias; (iv) distill the LLM prior into compact student models to lower inference cost for on-premise deployments; and (v) extend the verify step with logical constraints and causal patterns (*e.g.*, pathway-consistent paths) to further improve biomedical plausibility. These directions preserve the training-free ethos while strengthening reliability and efficiency.

CONCLUSION

This study presents LLM-KG2F, a training-free, cold-start framework for biomedical knowledge graph fusion. By fusing a zero-shot LLM plausibility prior with TransH/BootEA in entity alignment and employing a chunk–reason–verify pipeline for link prediction, LLM-KG2F enhances semantic integration without task-specific retraining. Experiments on MED-BBK-9K, RNA-KG, and a cold-start CKG split demonstrate consistent improvements over strong baselines. The method's plug-and-play design allows LLM upgrades without re-optimizing the structural pipeline, making it suitable for evolving biomedical KGs. Overall, LLM-KG2F offers a practical and scalable approach to KG fusion and inference that leverages pretrained language knowledge while retaining structural rigor.

LIMITATIONS

Although LLM-KG2F is training-free, it is not inference-free and still incurs LLM-call latency and cost that can limit scalability on very large KGs. In addition, the semantic plausibility signal may vary with prompt design and the underlying LLM, and it can be less reliable for rare entities or noisy/abbreviated biomedical names. Finally, our evaluation is conducted on a limited set of biomedical datasets under cold-start settings, so broader generalization to other KGs and downstream applications warrants further validation.

AUTHORS' CONTRIBUTIONS

XH contributed to Methodology; XD was responsible for Writing the Paper (original draft preparation); XL conducted Data Analysis or Interpretation; and RG and XF handled Writing – Reviewing and Editing.

LIST OF ABBREVIATIONS

LLM	=	Large Language Model
KG / KGs	=	Knowledge Graph(s)
CKG	=	Clinical Knowledge Graph
LLM-KG2F	=	LLM-enhanced Knowledge Graph Fusion Framework
EA	=	Entity Alignment
LP	=	Link Prediction
MR	=	Mean Rank
MRR	=	Mean Reciprocal Rank
Hits@K	=	Hit rate at rank K
BFS	=	Breadth-First Search

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

Not applicable.

HUMAN AND ANIMAL RIGHTS

Not applicable.

CONSENT FOR PUBLICATION

Not applicable.

AVAILABILITY OF DATA AND MATERIALS

The data and supportive information are available within the article.

FUNDING

The authors are grateful for the support of the National Natural Science Foundation of China, (Grant Nos. 62372494, 62372209, 62272192, and 62172187), the Science and Technology Planning Project of Jilin Province (Grant No. 20220201145GX), the Science and Technology Planning Project of Guangdong Province (Grant No. 2020A0505100018), and Guangdong Universities' Innovation Team Project (Grant No. 2021 KCXTD015).

CONFLICT OF INTEREST

The authors declare no conflict of interest, financial or otherwise.

ACKNOWLEDGEMENTS

We thank the anonymous reviewers and the editors for their constructive comments and suggestions. Any opinions, findings, and conclusions expressed in this paper are those of the authors and do not necessarily reflect the views of the funding agencies.

REFERENCES

[1] Cui H, Lu J, Xu R, *et al.* A review on knowledge graphs for healthcare: Resources, applications, and promises. J Biomed Inform 2025; 169: 104861. http://dx.doi.org/10.1016/j.jbi.2025.104861 PMID: 40712809

[2] Nicholson DN, Greene CS. Constructing knowledge graphs and their biomedical applications. Comput Struct Biotechnol J 2020; 18: 1414-28.

http://dx.doi.org/10.1016/j.csbj.2020.05.017 PMID: 32637040

[3] Wang Z, Gerstein M, Snyder M. RNA-Seq: A revolutionary tool for transcriptomics. Nat Rev Genet 2009; 10(1): 57-63. http://dx.doi.org/10.1038/nrg2484 PMID: 19015660

[4] Li P, Li L, Nan J, Chen J, Sun J, Cao Y. KEGNI: Knowledge graph enhanced framework for gene regulatory network inference. Genome Biol 2025; 26(1): 294. http://dx.doi.org/10.1186/s13059-025-03780-7 PMID: 40983951

[5] Chen B, Ding Y, Wild DJ. Assessing drug target association using semantic linked data. PLOS Comput Biol 2012; 8(7): e1002574. http://dx.doi.org/10.1371/journal.pcbi.1002574 PMID: 22859915

[6] Piñero J, Queralt-Rosinach N, Bravo A, *et al.* DisGeNET: A discovery platform for the dynamical exploration of human diseases and their genes. Database 2015; 2015(0): bav028. http://dx.doi.org/10.1093/database/bav028 PMID: 25877637

[7] Nunes S C P. Predicting gene-disease associations with knowledge graph embeddings over multiple ontologies. Doctor of Philosophy, Universidade de Lisboa 2021.

[8] Cavalleri E, Cabri A, Soto-Gomez M, *et al.* An ontology-based knowledge graph for representing interactions involving RNA molecules. Sci Data 2024; 11(1): 906. http://dx.doi.org/10.1038/s41597-024-03673-7 PMID: 39174566

[9] Nguyen HL, Vu DT, Jung JJ. Knowledge graph fusion for smart systems: A Survey. Inf Fusion 2020; 61: 56-70. http://dx.doi.org/10.1016/j.inffus.2020.03.014

[10] Lin Y, He J, Chen J, Zhu X, Zheng J, Bo T. BioGraphFusion: Graph knowledge embedding for biological completion and reasoning. Bioinformatics 2025; 41(7): btaf408. http://dx.doi.org/10.1093/bioinformatics/btaf408 PMID: 40680165

[11] Bordes A, Usunier N, Garcia-Durán A, Weston J, Yakhnenko O. Translating embeddings for modeling multi-relational data. Adv Neural Inf Process Syst. Lake Tahoe, NV, USA, 2013, pp. 2787-2795. http://dx.doi.org/10.5555/2999792.2999923

[12] Zhang M, Chen Y. Link prediction based on graph neural networks. NIPS'18: Proceedings of the 32nd International Conference on Neural Information Processing Systems. Red Hook, NY, USA, 2018, pp. 5171-5181. http://dx.doi.org/10.5555/3327345.3327423

[13] Wang Y, Cui Y, Liu W, *et al.* Facing Changes: Continual entity alignment for growing knowledge graphs. The Semantic Web – ISWC 2022, Virtual Event, Oct 23–27 2022, pp. 196-213. http://dx.doi.org/10.1007/978-3-031-19433-7_12

[14] Alon U, Yahav E. On the bottleneck of graph neural networks and its practical implications. arXiv 2021. http://dx.doi.org/10.48550/arXiv.2006.05205

[15] Liu H, Jian Y, Zeng C, Zhao Y. RNA-protein interaction prediction using network-guided deep learning. Commun Biol 2025; 8(1): 247. http://dx.doi.org/10.1038/s42003-025-07694-9 PMID: 39956833

[16] Cipriani V, Vestito L, Magavern EF, *et al.* Rare disease gene association discovery in the 100,000 Genomes Project. Nature 2025. http://dx.doi.org/10.1038/s41586-025-08623-w PMID: 40011789

[17] Wu T, Cheng AY, Zhang Y, *et al.* KARR-seq reveals cellular higher-order RNA structures and RNA–RNA interactions. Nat Biotechnol 2024; 42(12): 1909-20. http://dx.doi.org/10.1038/s41587-023-02109-8 PMID: 38238480

[18] Fu Z, Zhou P, Ren H, *et al.* Reasoning and analysis of integrated traditional Chinese and Western medicine for acute abdomen based on knowledge graph. Chin J Exp Tradit Med Formulae 2023; 29(11): 190-9. http://dx.doi.org/10.13422/j.cnki.syfjx.20230512

[19] Robinson PN, Köhler S, Bauer S, Seelow D, Horn D, Mundlos S. The Human Phenotype Ontology: A tool for annotating and analyzing human hereditary disease. Am J Hum Genet 2008; 83(5): 610-5. http://dx.doi.org/10.1016/j.ajhg.2008.09.017 PMID: 18950739

[20] Schriml LM, Munro JB, Schor M, *et al.* The human disease ontology 2022 update. Nucleic Acids Res 2022; 50(D1): D1255-61. http://dx.doi.org/10.1093/nar/gkab1063 PMID: 34755882

[21] Sang S, Yang Z, Wang L, Liu X, Lin H, Wang J. SemaTyP: A knowledge graph based literature mining method for drug discovery. BMC Bioinformatics 2018; 19(1): 193. http://dx.doi.org/10.1186/s12859-018-2167-5 PMID: 29843590

[22] Zhang R, Hristovski D, Schutte D, Kastrin A, Fiszman M, Kilicoglu H. Drug repurposing for COVID-19 *via* knowledge graph completion. J Biomed Inform 2021; 115: 103696. http://dx.doi.org/10.1016/j.jbi.2021.103696 PMID: 33571675

[23] Kumar AA, Bhandary S, Hegde SG, Chatterjee J. Knowledge graph applications and multi-relation learning for drug repurposing: A scoping review. Comput Biol Chem 2025; 115: 108364. http://dx.doi.org/10.1016/j.compbiolchem.2025.108364 PMID: 39914071

[24] Chang J, Wang S, Ling C, Qin Z, Zhao L. Gene-associated disease discovery powered by large language models. arXiv 2024. http://dx.doi.org/10.48550/arXiv.2401.09490

[25] Bang D, Lim S, Lee S, *et al.* Biomedical knowledge graph learning for drug repurposing by extending guilt-by-association to multiple layers. Nat Commun 2023; 14(1): 3570. http://dx.doi.org/10.1038/s41467-023-39301-y PMID: 37322032

[26] Lobentanzer S, Aloy P, Baumbach J, *et al.* Democratizing knowledge representation with BioCypher. Nat Biotechnol 2023; 41(8): 1056-9. http://dx.doi.org/10.1038/s41587-023-01848-y PMID: 37337100

[27] Pan S, Luo L, Wang Y, Chen C, Wang J, Wu X. Unifying large language models and knowledge graphs: A roadmap. IEEE Trans Knowl Data Eng 2024; 36(7): 3580-99. http://dx.doi.org/10.1109/TKDE.2024.3352100

[28] Xu R, Li L, Wang Q. dRiskKB: A large-scale disease-disease risk relationship knowledge base constructed from biomedical text. BMC Bioinformatics 2014; 15(1): 105. http://dx.doi.org/10.1186/1471-2105-15-105 PMID: 24725842

[29] Santos A, Colaço AR, Nielsen AB, *et al.* A knowledge graph to interpret clinical proteomics data. Nat Biotechnol 2022; 40(5): 692-702. http://dx.doi.org/10.1038/s41587-021-01145-6 PMID: 35102292

[30] Nian Y, Hu X, Zhang R, *et al.* Mining on Alzheimer's diseases related knowledge graph to identity potential AD-related semantic triples for drug repurposing. BMC Bioinformatics 2022; 23(S6): 407. http://dx.doi.org/10.1186/s12859-022-04934-1 PMID: 36180861

[31] Goodwin TR, Harabagiu SM. Knowledge representations and inference techniques for medical question answering. ACM Trans Intell Syst Technol 2018; 9(2): 1-26. http://dx.doi.org/10.1145/3106745

[32] Cavalleri E, Perlasca P, Mesiti M. RNA-KG v2.0: An RNA-centered knowledge graph with properties. NAR Genom Bioinform 2026; 8(1): lqaf194. http://dx.doi.org/10.1093/nargab/lqaf194 PMID: 41531547

[33] Sarumi OA, Heider D. Large language models and their applications in bioinformatics. Comput Struct Biotechnol J 2024; 23: 3498-505. http://dx.doi.org/10.1016/j.csbj.2024.09.031 PMID: 39435343

[34] Pei Q, Wu L, Gao K, *et al.* BioT5+: Towards generalized biological understanding with IUPAC integration and multi-task tuning. Findings of the Association for Computational Linguistics: ACL 2024. Bangkok, Thailand, Aug 2024, pp. 1216–1240. http://dx.doi.org/10.18653/v1/2024.findings-acl.71

[35] Wang Z, Lv Q, Lan X, Zhang Y. Cross-lingual knowledge graph alignment *via* graph convolutional networks. Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. Brussels, Belgium, Oct–Nov 2018, pp. 349-357. http://dx.doi.org/10.18653/v1/D18-1032

[36] Peng C, Xia F, Naseriparsa M, Osborne F. Knowledge graphs: Opportunities and challenges. Artif Intell Rev 2023; 56(11): 13071-102. http://dx.doi.org/10.1007/s10462-023-10465-9 PMID: 37362886

[37] Xiao Y, Zhang S, Zhou H, Li M, Yang H, Zhang R. FuseLinker: Leveraging LLM's pre-trained text embeddings and domain

knowledge to enhance GNN-based link prediction on biomedical knowledge graphs. J Biomed Inform 2024; 158: 104730. http://dx.doi.org/10.1016/j.jbi.2024.104730 PMID: 39326691

[38] Hu Y, Oleshko S, Firmani S, *et al.* Path-based reasoning for biomedical knowledge graphs with BioKGC. bioRxiv 2024. http://dx.doi.org/10.1101/2024.06.17.599219

[39] Wang Z, Zhang J, Feng J, Chen Z. Knowledge graph embedding by translating on hyperplanes. Proc AAAI Conf Artif Intell 2014; 28(1): 1112–1119. http://dx.doi.org/10.1609/aaai.v28i1.8870

[40] Lin Y, Liu Z, Sun M, Liu Y, Zhu X. Learning entity and relation embeddings for knowledge graph completion. Proc Conf AAAI Artif Intell 2015; 29(1): 2181-7. http://dx.doi.org/10.1609/aaai.v29i1.9491

[41] Zhang Z, Liu H, Chen J, *et al.* An industry evaluation of embedding-based entity alignment. Proceedings of the 28th International Conference on Computational Linguistics: Industry Track. Online, Dec 2020, pp. 179-189. http://dx.doi.org/10.18653/v1/2020.coling-industry.17

[42] Yang J, Lu G, He S, Cao Q, Liu Y. A novel model for relation prediction in knowledge graphs exploiting semantic and structural feature integration. Sci Rep 2024; 14(1): 12962. http://dx.doi.org/10.1038/s41598-024-63279-2 PMID: 38839794

[43] Sun Z, Hu W, Zhang Q, Qu Y. Bootstrapping entity alignment with knowledge graph embedding. Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence. Main track, 2018, pp. 4396-4402. http://dx.doi.org/10.24963/ijcai.2018/611

[44] Wu Y, Liu X, Feng Y, Wang Z, Yan R, Zhao D. Relation-aware entity alignment for heterogeneous knowledge graphs. Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence (IJCAI-19). 2019, pp. 5278-5284. http://dx.doi.org/10.24963/ijcai.2019/733

[45] Liu X, Hong H, Wang X, *et al.* SelfKG: Self-supervised entity alignment in knowledge graphs. Proceedings of the ACM Web Conference 2022. Virtual Event, Lyon, France, 2022, pp. 860-870. http://dx.doi.org/10.1145/3485447.3511945

[46] Gui J, Chen T, Zhang J, *et al.* A survey on self-supervised learning: Algorithms, applications, and future trends. arXiv 2023. http://dx.doi.org/10.48550/arXiv.2301.05712

[47] Huang W, Yi M, Zhao X, Jiang Z. Towards the generalization of contrastive self-supervised learning. arXiv 2021. http://dx.doi.org/10.48550/arXiv.2111.00743

[48] Trouillon T, Welbl J, Riedel S, Gaussier E, Bouchard G. Complex embeddings for simple link prediction. Proceedings of the 33rd International Conference on Machine Learning (ICML), PMLR. 2016, pp. 2071-2080.

[49] Bojchevski A, Günnemann S. Deep gaussian embedding of graphs: Unsupervised inductive learning *via* ranking. arXiv 2018. http://dx.doi.org/10.48550/arXiv.1707.03815

[50] Li Z, Zhao Y, Zhang Y, Zhang Z. Multi-relational graph attention networks for knowledge graph completion. Knowl Base Syst 2022; 251: 109262. http://dx.doi.org/10.1016/j.knosys.2022.109262

[51] Lu Y, Goi S Y, Zhao X, Wang J. Biomedical knowledge graph: A survey of domains, tasks, and real-world applications. arXiv 2025. http://dx.doi.org/10.48550/arXiv.2501.11632

[52] Zhang H, Bai L. Few-shot link prediction for temporal knowledge graphs based on time-aware translation and attention mechanism. Neural Netw 2023; 161: 371-81. http://dx.doi.org/10.1016/j.neunet.2023.01.043 PMID: 36780860

[53] Chen X, Cai R, Huang Z, Li Z, Zheng J, Wu M. Interpretable high-order knowledge graph neural network for predicting synthetic lethality in human cancers. Brief Bioinform 2025; 26(2): bbaf142. http://dx.doi.org/10.1093/bib/bbaf142 PMID: 40194555

[54] Jia X, Luo W, Li J, *et al.* A deep learning framework for predicting disease-gene associations with functional modules and graph augmentation. BMC Bioinformatics 2024; 25(1): 214. http://dx.doi.org/10.1186/s12859-024-05841-3 PMID: 38877401

[55] Li MM, Huang K, Zitnik M. Graph representation learning in biomedicine and healthcare. Nat Biomed Eng 2022; 6(12): 1353-69. http://dx.doi.org/10.1038/s41551-022-00942-x PMID: 36316368

[56] Wang X, He X, Cao Y, Liu M, Chua T-S. KGAT: Knowledge graph attention network for recommendation. Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. Anchorage, AK, USA, 2019, pp. 950-958. http://dx.doi.org/10.1145/3292500.3330989

[57] Veličković P, Cucurull G, Casanova A, Romero A, Liò P, Bengio Y. Graph attention networks. arXiv 2018. http://dx.doi.org/10.48550/arXiv.1710.10903

[58] Vaswani A, Shazeer N, Parmar N, *et al.* Attention is all you need. Adv Neural Inf Process Syst 2017; 30. http://dx.doi.org/10.48550/arXiv.1706.03762

[59] Devlin J, Chang M-W, Lee K, Toutanova K. BERT: Pre-training of deep bidirectional transformers for language understanding. Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Minneapolis, Minnesota, 2019, pp. 4171-4186. http://dx.doi.org/10.18653/v1/N19-1423

[60] Gema AP, Grabarczyk D, De Wulf W, *et al.* Knowledge graph embeddings in the biomedical domain: Are they useful? A look at link prediction, rule learning, and downstream polypharmacy tasks. Bioinform Adv 2024; 4(1): vbae097. http://dx.doi.org/10.1093/bioadv/vbae097 PMID: 39506988

[61] Bradshaw MS, Avramov A, Gaskell A, Layer RM. The effects of biological knowledge graph topology on embedding-based link prediction. bioRxiv 2024. http://dx.doi.org/10.1101/2024.06.10.598277

[62] Zha H, Chen Z, Yan X. Inductive relation prediction by BERT. Proc AAAI Conf Artif Intell 2022; 36(5): 5923-31. http://dx.doi.org/10.1609/aaai.v36i5.20537

[63] Hao Y, Cao X, Fang Y, Xie X, Wang S. Inductive link prediction for nodes having only attribute information. Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence (IJCAI-20). 2020, pp. 1209-1215. http://dx.doi.org/10.24963/ijcai.2020/168

[64] Bahrini A, Khamoshifar M, Abbasimehr H, *et al.* ChatGPT: Applications, opportunities, and threats. 2023 Systems and Information Engineering Design Symposium (SIEDS). Charlottesville, VA, USA, Apr 27–28 2023, pp. 274-279. http://dx.doi.org/10.1109/SIEDS58326.2023.10137850

[65] Chen J, Geng Y, Chen Z, *et al.* Zero-shot and few-shot learning with knowledge graphs: A comprehensive survey. Proc IEEE 2023; 111(6): 653-85. http://dx.doi.org/10.1109/JPROC.2023.3279374

[66] Lee J, Yoon W, Kim S, *et al.* BioBERT: A pre-trained biomedical language representation model for biomedical text mining. Bioinformatics 2020; 36(4): 1234-40. http://dx.doi.org/10.1093/bioinformatics/btz682 PMID: 31501885

[67] Gao Y, Li R, Croxford E, *et al.* Leveraging medical knowledge graphs into large language models for diagnosis prediction: Design and application study. JMIR AI 2025; 4: e58670. http://dx.doi.org/10.2196/58670 PMID: 39993309

[68] Vinyals O, Blundell C, Lillicrap T, Kavukcuoglu K, Wierstra D. Matching networks for one shot learning. Proceedings of the 30th International Conference on Neural Information Processing Systems (NIPS). Barcelona, Spain, 2016, pp. 3637-3645. http://dx.doi.org/10.5555/3157382.3157504

[69] Xiang C, Ma T, Fu X, Liu Y, Song B, Zeng X. From knowledge to treatment: Large language model assisted biomedical concept representation for drug repurposing. Findings of the Association for Computational Linguistics: EMNLP 2025. Suzhou, China, Nov 2025, pp. 13967-13982. http://dx.doi.org/10.18653/v1/2025.findings-emnlp.751

[70] Sengupta A, Selby DA, Vollmer SJ, Großmann G. MEDAKA: Construction of biomedical knowledge graphs using large language models. arXiv 2025. http://dx.doi.org/10.48550/arXiv.2509.26128

[71] Lu J, Shen J, Xiong B, Ma W, Staab S, Yang C. HiPrompt: Few-shot biomedical knowledge fusion *via* hierarchy-oriented prompting. Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval. Taipei, Taiwan, 2023, pp. 2052-2056.
http://dx.doi.org/10.1145/3539618.3591997

[72] Callahan TJ, Tripodi IJ, Stefanski AL, *et al.* An open source knowledge graph ecosystem for the life sciences. Sci Data 2024; 11(1): 363.
http://dx.doi.org/10.1038/s41597-024-03171-w PMID: 38605048

[73] Li D, Yang Y, Cui Z, Yin H, Hu P, Hu L. LLM-DDI: Leveraging large language models for drug-drug interaction prediction on biomedical knowledge graph. IEEE J Biomed Health Inform 2026; 30(1): 773-81.
http://dx.doi.org/10.1109/JBHI.2025.3585290 PMID: 40601466