

# Flow-Modulated Scoring for Semantic-Aware Knowledge Graph Completion

Siyuan Li , Ruitong Liu , Yan Wen , Te sun , Andi Zhang , Yanbiao Ma , Xiaoshuai Hao

**Abstract**—Knowledge graph completion demands effective modeling of multifaceted semantic relationships between entities. Yet, prevailing methods, which rely on static scoring functions over learned embeddings, struggling to simultaneously capture rich semantic context and the dynamic nature of relations. To overcome this limitation, we propose the Flow-Modulated Scoring (FMS) framework, conceptualizing a relation as a dynamic evolutionary process governed by its static semantic environment. FMS operates in two stages: it first learns context-aware entity embeddings via a Semantic Context Learning module, and then models a dynamic flow between them using a Conditional Flow-Matching module. This learned flow dynamically modulates a base static score for the entity pair. By unifying context-rich static representations with a conditioned dynamic flow, FMS achieves a more comprehensive understanding of relational semantics. Extensive experiments demonstrate that FMS establishes a new state of the art across both canonical knowledge graph completion tasks: relation prediction and entity prediction. On the standard relation prediction benchmark FB15k-237, FMS achieves a near-perfect MRR of 99.8% and Hits@1 of 99.7% using a mere 0.35M parameters, while also attaining a 99.9% MRR on WN18RR. Its dominance extends to entity prediction, where it secures a 25.2% relative MRR gain in the transductive setting and substantially outperforms all baselines in challenging inductive settings. By unifying a dynamic flow mechanism with rich static contexts, FMS offers a highly effective and parameter-efficient new paradigm for knowledge graph completion. Code published at: <https://github.com/yuanwuyuan9/FMS>.

**Index Terms**—Knowledge Graph, Flow-Matching, Graph Neural Network, Knowledge Representation and Reasoning.

## 1 INTRODUCTION

Knowledge Graphs (KGs), as a robust formalism for knowledge representation, have become pivotal in a myriad of large-scale applications, including recommender systems [1], [2], information extraction [3], and question answering [4], [5], owing to their remarkable efficacy in organizing and representing complex semantic facts [6], [7], [8], [9]. However, real-world KGs, despite their vast scale, are ubiquitously plagued by incompleteness [10], [11], [12]. This issue stems from the continuous evolution of knowledge itself and the high cost and complexity associated with its acquisition. To address this fundamental challenge, the task of Knowledge Graph Completion (KGC) has emerged, which aims to infer and predict missing links based on existing facts [13], [14], [15], [16].

Among the sub-tasks of KGC, Relation Prediction (RP) is crucial for enabling fine-grained knowledge reasoning [17], [18]. It is dedicated to inferring the set of potential relations  $\mathcal{R}$  for a given entity pair  $(h, t)$  [19]. Existing methodologies for this task predominantly fall into two categories. The first category comprises embedding-based models, such as TransE [20], TransH [21], and DistMult [22], which assess the plausibility of a triplet by learning static entity and relation embeddings within a fixed scoring function. The second category consists of GNN-based methods that aggregate contextual information [23], [24], like PathCon [19], which

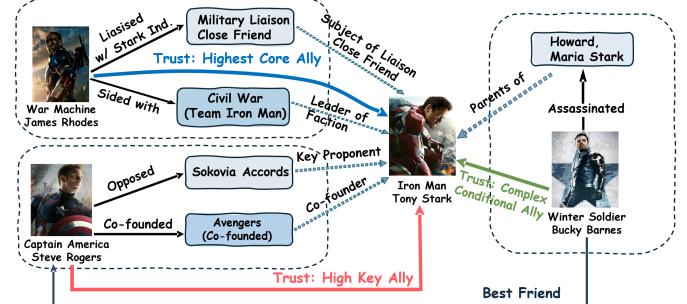


Fig. 1. The context-dependent nature of relational semantics, exemplified by Iron Man's alliances. The generic relation "ally" acquires distinct meanings (e.g., Highest Core, High Key, Complex Conditional) based on specific contextual facts, such as shared history or critical conflicts.



Fig. 2. Illustration of the dynamic evolution of relationships. The examples show the changing dynamics over time between (left) Thor and Loki, and (right) Iron Man and Spider-Man, highlighting that relationships progress through distinct stages.

leverages multi-hop neighborhood structures to enrich semantic representations. Despite the progress achieved, these methods are constrained by two fundamental bottlenecks:

**1) Inability to Capture Rich Contextual Information Shaping Relation Semantics.** The meaning of a relation is often highly contingent on its surrounding factual con-

Siyuan Li, Ruitong Liu and Te Sun are with the Dalian University of Technology, China. Yan Wen is with the Beijing Institute of Technology, China. Andi Zhang is with the University of Manchester, U.K. Yanbiao Ma is with the Gaoling School of Artificial Intelligence, Renmin University of China. Xiaoshuai Hao is with the Beijing Academy of Artificial Intelligence, China. Correspondence author: Yanbiao Ma, Xiaoshuai Hao  
E-mail: ybma1998xidian@gmail.com, haoxiaoshuai714@163.com

text [15], [19], [25]. As illustrated in [Figure 1](#), the generalized label “ally” is insufficient to accurately describe the relationships between Iron Man and other characters. His relationship with War Machine, rooted in their military liaison, is a “core, top-level alliance,” whereas his relationship with Captain America is a “key, high-level alliance” shaped by their shared history of team-building and subsequent political disagreements. This indicates that effective relation prediction must account for the subtle contextual nuances that differentiate superficially similar yet substantially distinct relations.

### 2) Incapacity of Static Paradigms to Model the Dynamic Evolution of Relations.

In reality, relationships exhibit path-dependent and time-evolving characteristics [9], [26]. For instance, the Thor-Loki dynamic shifted from brotherhood to enmity and later to reconciliation, while the Iron Man-Spider-Man bond evolved from a mentor-protégé dynamic to a quasi-paternal one (see [Figure 2](#)). Accurately modeling such evolutionary trajectories is critical, a task for which existing static scoring functions are ill-equipped. While GNNs introduce structural context, they operate on fixed snapshots of the graph and are thus inherently static in their modeling approach [19].

Recently, Large Language Models (LLMs) have been introduced to enhance semantic understanding in KGC [10], [27], [28], as seen in models like KG-BERT [29] and MuKDC [30]. Although these models offer improved context awareness, they still lack explicit mechanisms for modeling the evolutionary paths of relations and are often hampered by large parameter counts, low efficiency, and challenges in domain adaptation. Furthermore, approaches like MuKDC are also constrained by their strong reliance on meticulously designed prompts [9], [31].

To this end, this paper introduces the **Flow-Modulated Scoring (FMS)** framework, a novel approach that pioneers the unification of context-aware and dynamic evolution modeling for knowledge graph completion. The core idea of FMS is to conceptualize a relation as a conditional vector flow from a head entity  $h$  to a tail entity  $t$ , where the trajectory of this flow is dynamically modulated by the local semantic context. Specifically, the FMS framework comprises two key modules: 1) Semantic Context Learning Module employs an energy function to identify and aggregate the top- $k$  most semantically relevant contextual paths, thereby constructing context-aware entity representations. 2) Conditional Flow Matching Module learns a conditional vector field that explicitly models the dynamic evolutionary trajectory of a relation from  $h$  to  $t$ , enabling a path-sensitive representation of relational semantics.

By integrating these components, FMS dynamically adjusts the output of conventional scoring functions, facilitating a fine-grained modeling of relation semantics that is both context-sensitive and temporally aware. Extensive experiments demonstrate that FMS achieves state-of-the-art (SOTA) performance across multiple standard benchmarks while maintaining superior parameter efficiency. The main contributions of this paper are summarized as follows:

- **Flow-Modulated Scoring Framework:** We propose the first KGC framework that integrates context-aware and dynamic evolution modeling, overcoming the limitations of static scoring paradigms ([Section 3.2](#)).

- **Semantic Context Learning:** We design an energy-based top- $k$  context selection mechanism to efficiently capture critical semantic dependencies ([Section 3.2.1](#)).
- **Conditional Flow Matching:** We introduce a conditional vector flow mechanism to explicitly model the evolutionary paths of relations, enhancing the capacity for dynamic semantic representation ([Section 3.2.2](#)).
- We demonstrate the state-of-the-art performance and high parameter efficiency of FMS on multiple benchmarks, confirming its effectiveness in modeling complex relations ([Section 4](#) and [Section 5](#)).

## 2 RELATED WORK

### 2.1 Knowledge Graph Completion

Knowledge Graphs provide structured information for downstream tasks such as recommender systems and semantic analysis [9], [13], [32]. Knowledge Graph Completion (KGC), and particularly Relation Prediction (RP), addresses the problem of KG incompleteness by predicting missing links [10], [33], [34], [35].

The majority of KGC methods rely on embedding-based approaches, which assign vectors to entities and relations in a continuous space and are trained on observed facts [36], [37], [38], [39]. For instance, early geometric and bilinear models, such as TransE [20] and ComplEx [40], laid the groundwork by defining simple yet effective scoring functions. To capture more intricate relational patterns, subsequent research has advanced into more expressive geometric spaces. RotaE [14], for example, introduced relational rotations in complex space, a concept later enhanced by Rot-Pro [41] for greater fidelity. Other approaches have explored specialized geometries, such as HAKE’s [42] use of polar coordinates for semantic hierarchies and BoxE’s [43] hyper-rectangle representations for logical conjunctions. In a parallel thread of research, recent work has focused not on geometry but on augmenting the reasoning process itself. AnKGE [44] exemplifies this direction by equipping base models with an analogical inference mechanism to leverage similar, known facts [24].

Despite their effectiveness in handling relational patterns, these models are fundamentally static. They rely on pre-computed, fixed embeddings, which renders the models context-agnostic; the representation of an entity or relation remains constant regardless of the specific factual query [9]. This paradigm is inherently incapable of capturing the subtle nuances and contextual dependencies of relational meanings. To overcome static embeddings, recent research has leveraged Large Language Models (LLMs) to inject contextual awareness [35], [45]. These approaches either refine link prediction as a textual scoring task (e.g., KG-BERT [29], SimKGC [46]) or utilize LLMs as generative knowledge sources to distill missing facts, as exemplified by MuKDC [30]. Despite their promise, these LLM-based methods are often hampered by their massive parameter counts, low computational efficiency, and significant challenges in domain adaptation [35]. Another direction to overcome static embeddings is through generative modeling. Recent works like FDM [37] and KGDM [47] employ diffusion models [48] to learn fact distributions. However, their generative process is typically conditioned solely on the local

query triplet, thus failing to leverage the broader, multi-hop structural context crucial for robust link prediction.

The advent of GNN-based models represents an attempt to incorporate contextual information from the graph structure [49], [50]. Methods like PathCon [19] enrich entity and relation representations by aggregating information from multi-hop paths, demonstrating the value of neighborhood context. To mitigate the high cost of explicit path extraction, recent work CBLiP [51] instead uses connection-biased attention to implicitly capture structural information from the local subgraph. However, these approaches still operate on a static snapshot of the graph. While they encode structural context, they do not explicitly model the formation process or evolutionary trajectory of relations. Their aggregation mechanisms are typically fixed and cannot adapt to the dynamic and path-dependent characteristics inherent in how relationships form and evolve. In summary, prevailing methodologies exhibit an over-reliance on static modeling paradigms, making them ill-equipped to handle the dual challenges of context-sensitivity and dynamic evolution.

## 2.2 Modeling Dynamic Processes with Conditional Flow Matching

To address the core challenges of modeling dynamic relationship evolution, we draw on research in generative models concerning the learning of transformations and trajectories between data distributions. Although diffusion models [48], [52] have demonstrated formidable generative capabilities in domains such as computer vision [53] and sequence modeling [54], [55], their training and inference processes are often computationally expensive.

A more direct and efficient approach for learning transformation paths is offered by Continuous Normalizing Flows (CNFs) [56]. CNFs learn an invertible mapping between a data distribution and a simple prior distribution by parameterizing the vector field of an Ordinary Differential Equation (ODE). Recently, Flow Matching (FM) [57], [58], [59], [60] has emerged as a highly efficient and stable technique for training CNFs. Unlike score-matching methods that learn the gradient of the data distribution’s logarithm, FM learns the transformation by directly regressing the ODE’s vector field. This simulation-free training approach significantly enhances model stability and efficiency. The advent of FM introduced new paradigms for CNF training and improved sample quality, and while early FM methods often assumed a Gaussian source distribution, this limitation has been addressed in subsequent work [61], [62].

Building on this foundation, Conditional Flow Matching (CFM) [63], [64], [65] offers a more generalized and powerful framework. CFM aims to learn probability paths conditioned on specific attributes, providing a unified, simulation-free training objective for models with arbitrary transport maps. Crucially, it neither requires the source distribution to be Gaussian nor necessitates the evaluation of its probability density. These properties—efficiency, stability, and flexible conditioning—make CFM the ideal theoretical tool for our objectives. **In this work**, we innovatively apply the core mechanism of Conditional Flow Matching within KGC. Instead of using it for data generation, we leverage it to model the **conditional evolutionary path** of a relation

from a head entity  $h$  to a tail entity  $t$ . This enables us to explicitly model relations as trajectories, dynamically modulated by semantic context, thereby overcoming the static limitations of existing KGC methods.

## 3 METHODOLOGY

Relational links arise from the interplay of two fundamental principles: the static semantic context of entities and the dynamic evolutionary process connecting them. In this section, we formalize this duality through a probabilistic decomposition of relation prediction (Section 3.1), which in turn motivates our two-component approach. Subsequently, we introduce the core components of the Flow-Modulated Scoring model (Section 3.2) and show how they are unified into a single, cohesive scoring function under a joint training objective (Section 3.2.3).

### 3.1 Preliminaries

Let  $\mathcal{G} = (\mathcal{V}, \mathcal{E})$  be an instance of a knowledge graph, where  $\mathcal{V}$  is the set of entity nodes and  $\mathcal{E}$  is the set of edges, with each edge representing a relation  $r \in \mathcal{R}$  [26]. Our goal is to predict missing relations in the knowledge graph, i.e., given an entity pair  $(h, t)$ , infer the relation  $r$  between them.

Specifically, we aim to model the distribution of relation types given the entity pair  $(h, t)$ , denoted as  $p(r|h, t)$ , to capture the complex semantic context and dynamic nature of relations. According to Bayes’ theorem, this distribution can be expressed as:

$$p(r|h, t) \propto p(h, t|r) \cdot p(r), \quad (1)$$

where  $p(r)$  is the prior distribution over relation types, serving as a regularization term. To further model the joint probability  $p(h, t|r)$ , we decompose it into a symmetric form:

$$p(h, t|r) = \frac{1}{2} (p(h|r) \cdot p(t|h, r) + p(t|r) \cdot p(h|t, r)). \quad (2)$$

Equation (2) provides guidance for our modeling approach. The terms  $p(h|r)$  and  $p(t|r)$  measure the likelihood of entities  $h$  or  $t$  given the relation  $r$ , reflecting the semantic context of the entities. The terms  $p(t|h, r)$  and  $p(h|t, r)$  capture the dynamic evolution from  $h$  to  $t$  or from  $t$  to  $h$  given the relation  $r$ , reflecting the dynamic nature of relations. The subsequent sections detail how our proposed model instantiates these two factors and integrates them for Knowledge graph completion.

### 3.2 Flow-Modulated Scoring Model

The overall framework of the Flow-Modulated Scoring model encompasses **two core components**: Semantic Context Learning and Conditional Flow Matching. As illustrated in Figure 3, the semantic context learning stage is designed to encode context-aware entity embeddings, while the conditional flow matching stage learns a dynamic flow from the head entity embedding to the tail entity embedding. This dynamic flow is then used to modulate a static score for the entity pair, with the ultimate goal of achieving comprehensive relation prediction. Next, we provide a detailed explanation of these components.

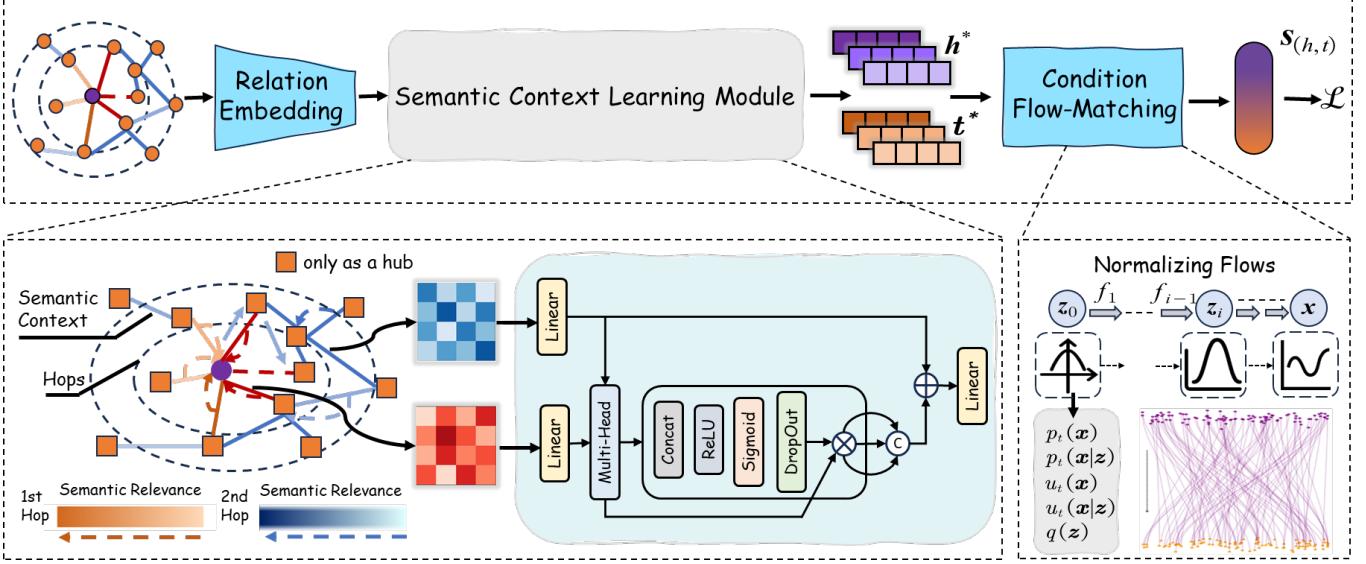


Fig. 3. The architecture of the Flow-Modulated Scoring (FMS) framework. In Semantic Context Learning Module, nodes (square shapes) do not consider their own intrinsic node messages; rather, their role is to facilitate the propagation and aggregation of messages associated with their incident edges, which represent the relation semantics. Conditional Flow Matching Module learns a dynamic flow that modulates the static score to produce the final prediction, unifying static context with dynamic evolution.

### 3.2.1 Semantic Context Learning

For a given triplet  $(h, r, t)$  in a Knowledge Graph (KG), its semantic context offers valuable cues for identifying valid links [19]. Traditional node-based message passing, as seen in GraphSAGE [66], GAT [67] and RED-GNN [26], is applicable to general graphs. However, when applied to KGs, we observe that edges themselves possess rich features (i.e., relation embeddings  $x_e$ ), and the number of relation types is significantly smaller than the number of entities. Consequently, modeling message passing directly between relations (edges) is not only more efficient but also provides more indicative semantic cues. As illustrated in Figure 1, our reasoning about Iron Man's "ally" status is predicated on these relational attributes, rather than on the intrinsic properties of the entities themselves.

To this end, we propose a relational message passing scheme where edges update their states by aggregating information from neighboring edges. This approach treats the KG's relational structure as the primary computational graph. However, naively aggregating information from all neighboring edges can introduce noise and lead to information dilution or over-smoothing, especially when some neighbors are semantically irrelevant to the target relation.

To address this challenge, we introduce a *Top-K selection mechanism based on semantic similarity between edges*. Specifically, when updating the state of a central target edge  $e_c$ , we first evaluate its semantic relevance to each neighboring edge  $e_n$  in its local neighborhood  $\mathcal{N}(e_c)$ . This is achieved by projecting their current states,  $s_i^{e_c}$  and  $s_i^{e_n}$ , into a shared latent semantic space using a learnable mapping function  $g(\cdot)$  [68]. In this space, we compute their similarity score using an energy-model inspired scoring function:

$$\text{Score}(e_c, e_n) = \exp(-\|g(s_i^{e_c}) - g(s_i^{e_n})\|^2 / \tau), \quad (3)$$

where  $\|\cdot\|^2$  is the squared Euclidean distance and  $\tau$  is a temperature hyperparameter. Edges with higher scores,

deemed more semantically relevant to the central edge  $e_c$ , are selected to form the Top-K set  $\mathcal{N}_K(e_c) \subseteq \mathcal{N}(e_c)$ .

The states of these selected Top-K edges,  $\{s_i^{e_n}\}_{e_n \in \mathcal{N}_K(e_c)}$ , are first mean-aggregated to form a consolidated neighborhood representation:

$$\bar{s}_i^{\mathcal{N}_K}(e_c) = \text{Mean}(\{s_i^{e_n}\}_{e_n \in \mathcal{N}_K(e_c)}). \quad (4)$$

This aggregated neighborhood representation  $\bar{s}_i^{\mathcal{N}}(e_c)$ , along with the central edge's own state  $s_i^{e_c}$ , is then processed by our specialized attention aggregator to compute the updated edge state for the next layer,  $s_{i+1}^{e_c}$ . This aggregator employs a multi-head attention mechanism that functions as a gating system. It transforms both  $s_i^{e_c}$  and  $\bar{s}_i^{\mathcal{N}}(e_c)$ , calculates head-wise gating scores based on their interaction, applies these gates to the transformed neighborhood representation, and then combines this gated information with the transformed self-representation before a final projection. This process yields a semantically focused update:

$$s_{i+1}^{e_c} = \text{AttentionAggregator}\left(s_i^{e_c}, \bar{s}_i^{\mathcal{N}_K}(e_c)\right). \quad (5)$$

The initial state of an edge is defined by its intrinsic feature embedding,  $s_0^e = x^e \in \mathbb{R}^{d_e}$ . This iterative update is performed for a predefined number of layers. This process ultimately generates a contextual representation for the central entity, which encapsulates the relational information within its multi-hop neighborhood. The impact of the semantic scoring function and aggregator choice will be further discussed in our ablation studies (subsection 4.4).

In this manner, our model directly leverages the relational structure of the KG. Through the semantic-aware edge neighbor selection and the subsequent attention-based aggregation, the model accurately captures and propagates context information most relevant to the specific relation prediction task, performing reasoning directly in the relation space. Notably, this process avoids the substantial compu-

tational overhead of encoding entity embeddings by not leveraging any entity-specific features.

### 3.2.2 Condition Flow-Matching Module

Prior research typically predicts static association scores directly utilizing entity embeddings. Although our Semantic Context Learning module is able to generate more informative entity representations, it is still insufficient to characterize the dynamics of relationships. **To address this limitation**, we have designed the Conditional Flow Matching Module. It constructs a conditional flow-based flow matching mechanism, incorporating the semantic context in which entities are situated as conditional variables into the relationship modeling process.

The core of this module is learning a dynamic vector field  $v_\theta(t, \mathbf{x})$  that depends on a pseudo-time variable  $t \in [0, 1]$  and the current state  $\mathbf{x} \in \mathbb{R}^d$ . The reason behind our proposal of the pseudo-time variable lies is the fact that the dynamics of relationships are inherently intertwined with the temporal dimension—evolution of relationships (in the real world it is mostly accompanied by the passage of time). However, the construction of traditional knowledge graph datasets has limitations: they do not explicitly incorporate temporal annotation information. Instead, such temporal attributes are latent in the overall distribution of massive triple data, making them difficult to be directly captured and utilized. To address this, we introduce  $t$  as a bridge. It does not directly replicate physical time, but rather externalizes the originally concealed temporal properties through parameterization. This treatment makes the model to more accurately capture the intrinsic logic of relationship evolution over “time”. This learned vector field characterizes the instantaneous direction and velocity along the evolution path from the head to the tail entity embedding.

Specifically, drawing upon the principles of Conditional Flow Matching [63], we employ a continuous flow to model the evolution from the head entity embedding  $\mathbf{h}^*$  to the tail entity embedding  $\mathbf{t}^{*1}$ . This process can be formalized as:

$$p_t(\mathbf{x}) = \int p_t(\mathbf{x}|z) q(z) dz, \quad (6)$$

where  $p_t(\mathbf{x})$  represents the marginal probability path,  $q(z)$  is the sampling distribution of the semantic context  $z$ , and the conditional probability path  $p_t(\mathbf{x}|z)$  denotes the probability path given  $z$ . Correspondingly, the marginal vector field  $u_t(\mathbf{x})$  that drives the evolution of  $p_t(\mathbf{x})$  can be expressed as an expectation of the conditional vector field  $u_t(\mathbf{x}|z)$  with respect to the prior distribution  $q(z)$ , weighted by the posterior distribution  $p(z|t, \mathbf{x})$ :

$$u_t(\mathbf{x}) = E_{q(z)} \left[ u_t(\mathbf{x}|z) \frac{p_t(\mathbf{x}|z)}{p_t(\mathbf{x})} \right], \quad (7)$$

where  $u_t(\mathbf{x}|z) = \frac{dx}{dt}$ . Consequently, under the initial condition  $p_t(\mathbf{x}_0)$ ,  $u_t(\mathbf{x})$  generates the probability flow path  $p_t(\mathbf{x})$ . We formally establish this result as follows.

*Proof.* To prove that the vector field  $u_t(\mathbf{x})$  generates the probability path  $p_t(\mathbf{x})$ , we must show that they satisfy

1. The asterisk (\*) is used here to distinguish entity representations from the pseudo-time variable  $t$ . This notation will be maintained without repeated clarification.

the continuity equation:  $\frac{\partial p_t(\mathbf{x})}{\partial t} + \nabla_{\mathbf{x}} \cdot (u_t(\mathbf{x}) p_t(\mathbf{x})) = 0$ . We begin by taking the time derivative of the marginal probability path defined in Eq. 6:

$$\frac{\partial p_t(\mathbf{x})}{\partial t} = \frac{\partial}{\partial t} \int p_t(\mathbf{x}|z) q(z) dz = \int \frac{\partial p_t(\mathbf{x}|z)}{\partial t} q(z) dz, \quad (8)$$

where the second of these equal signs is guaranteed by the Leibniz Rule. By definition, the conditional path  $p_t(\mathbf{x}|z)$  is generated by the conditional vector field  $u_t(\mathbf{x}|z)$ . Therefore, it satisfies its own continuity equation:

$$\frac{\partial p_t(\mathbf{x}|z)}{\partial t} = -\nabla_{\mathbf{x}} \cdot (u_t(\mathbf{x}|z) p_t(\mathbf{x}|z)). \quad (9)$$

Substituting this into our expression gives:

$$\frac{\partial p_t(\mathbf{x})}{\partial t} = \int -\nabla_{\mathbf{x}} \cdot (u_t(\mathbf{x}|z) p_t(\mathbf{x}|z)) q(z) dz \quad (10)$$

$$= -\nabla_{\mathbf{x}} \cdot \int u_t(\mathbf{x}|z) p_t(\mathbf{x}|z) q(z) dz. \quad (11)$$

In the second step, we moved the divergence operator  $\nabla_{\mathbf{x}} \cdot$  outside the integral, as the integration is with respect to  $z$ . Now, let’s examine the definition of the marginal vector field  $u_t(\mathbf{x})$  from Eq. 7. By multiplying both sides by  $p_t(\mathbf{x})$ , we get:

$$u_t(\mathbf{x}) p_t(\mathbf{x}) = \int u_t(\mathbf{x}|z) p_t(\mathbf{x}|z) q(z) dz. \quad (12)$$

The right-hand side is precisely the term inside the integral in our derivation. Substituting this back, we obtain:

$$\frac{\partial p_t(\mathbf{x})}{\partial t} = -\nabla_{\mathbf{x}} \cdot (u_t(\mathbf{x}) p_t(\mathbf{x})). \quad (13)$$

Rearranging the terms yields the continuity equation:

$$\frac{\partial p_t(\mathbf{x})}{\partial t} + \nabla_{\mathbf{x}} \cdot (u_t(\mathbf{x}) p_t(\mathbf{x})) = 0. \quad (14)$$

Since the marginal probability path  $p_t(\mathbf{x})$  and the marginal vector field  $u_t(\mathbf{x})$  satisfy the continuity equation, and  $p_t(\mathbf{x})$  satisfies the initial condition  $p_0(\mathbf{x})$  by construction, it follows that  $u_t(\mathbf{x})$  is indeed the generator of the probability path  $p_t(\mathbf{x})$ .  $\square$

Although we aim to recover  $u_t(\mathbf{x})$ , its direct computation is generally difficult. Considering the relative tractability of the conditional vector field  $p_t(\mathbf{x}|z)$ , we adopt an indirect learning strategy, and the model is enabled to directly learn an approximation of the vector field  $u_t(\mathbf{x}|z)$ . We assume  $v_\theta : [0, 1] \times \mathbb{R}^d \rightarrow \mathbb{R}^d$  is a time-dependent vector field parameterized by a neural network with parameters  $\theta$ . We define the following objective function:

$$\mathcal{L}_{cfm}(\theta) = \mathbb{E}_{t, q(z), p_t(\mathbf{x}|z)} \|v_\theta(t, \mathbf{x}) - u_t(\mathbf{x}|z)\|_2^2. \quad (15)$$

Based on this, the model learns  $v_\theta(t, \mathbf{x})$ . It can serve as a dynamic adjustment factor to modulate the static base score assigned to the pair of corresponding entities.

Then, we draw on the optimal transport theory [69], and aims to model the least-cost evolution path from head entity distribution to tail entity distribution.  $q(z)$  defined as:

$$q(z) = \pi(z) = \underset{\pi \in \Pi(\rho_H, \rho_T)}{\operatorname{argmin}} \int_{\mathbb{R}^d \times \mathbb{R}^d} \|\mathbf{x} - \mathbf{y}\|_2^2 d\pi(\mathbf{x}, \mathbf{y}), \quad (16)$$

where  $\rho_H$  and  $\rho_T$  represent the marginal distributions of the head and tail entity embeddings, respectively, and

$\Pi(\rho_H, \rho_T)$  is the set of all joint probability distributions. This design aims to guide the model to learn a flow field aligned with the static optimal transport path. The conditional probability path and its corresponding conditional vector field are defined as in the paper [69]:

$$p_t(\mathbf{x}|\mathbf{z}) = \mathcal{N}(\mathbf{x} | (1-t)\mathbf{h}^* + t\mathbf{t}^*, \sigma^2 \mathbf{I}), \quad (17)$$

$$u_t(\mathbf{x}|\mathbf{z}) = \mathbf{t}^* - \mathbf{h}^*, \quad (18)$$

where  $p_t(\mathbf{x}|\mathbf{z})$  is a Gaussian distribution, and the noise standard deviation  $\sigma$  controls the stochasticity of the probability path. In particular, as  $\sigma \rightarrow 0$ , The marginal vector field can be regarded as a generative model of  $\mathbf{h}^*$ . Proof is given in the **Theorem**. A moderate value of  $\sigma$  can enhance model robustness. Through CSM, model can learn more direct evolution paths between entities.

**Theorem 1.** *The marginal  $p_t$  corresponding to  $q(z) = q(\mathbf{h}^*)q(\mathbf{t}^*)$  and the  $p_t(\mathbf{x}|z)$ ,  $u_t(\mathbf{x}|z)$  in (17) and (18) has boundary conditions  $p_1 = q_1 * \mathcal{N}(x | 0, \sigma^2)$  and  $p_0 = q_0 * \mathcal{N}(x | 0, \sigma^2)$ , where  $*$  denotes the convolution operator.*

*Proof.* We start with (6) to show the result. We note that  $q(z) = q(\mathbf{h}^*, \mathbf{t}^*) = q(\mathbf{h}^*)q(\mathbf{t}^*)$ , so :

$$\begin{aligned} p_t(x) &= \int p_t(x | z)q(z) dz \\ &= \int \mathcal{N}(x | (1-t)\mathbf{h}^* + t\mathbf{t}^*, \sigma^2 \mathbf{I})q((\mathbf{h}^*, \mathbf{t}^*)) d(\mathbf{h}^*, \mathbf{t}^*) \\ &= \iint \mathcal{N}(x | (1-t)\mathbf{h}^* + t\mathbf{t}^*, \sigma^2 \mathbf{I})q(\mathbf{h}^*)q(\mathbf{t}^*) d\mathbf{h}^* d\mathbf{t}^* \end{aligned}$$

evaluated at  $i = 0, 1$  respectively. Therefore, at  $t = 0$ ,

$$\begin{aligned} p_0(x) &= \iint \mathcal{N}(x | \mathbf{h}^*, \sigma^2)q(\mathbf{h}^*)q(\mathbf{t}^*) d\mathbf{h}^* d\mathbf{t}^* \\ &= \int \mathcal{N}(x | \mathbf{h}^*, \sigma^2)q(\mathbf{h}^*) d\mathbf{h}^* \\ &= q(\mathbf{h}^*) * \mathcal{N}(x | 0, \sigma^2). \end{aligned}$$

This is also true for  $t = 1$ .  $\square$

### 3.2.3 Overall Scoring Process and Joint Training

Having detailed the Semantic Context Learning module and the Conditional Flow-Matching module for modeling dynamic evolution, this section will now describe how they are integrated into a unified scoring and training framework. The process unfolds in two sequential steps to produce a final, context-sensitive, and dynamically-aware score.

Firstly, we utilize the Semantic Context Learning module to compute the final messages  $m_{j-1}^h$  and  $m_{j-1}^t$  for the head entity  $h$  and the tail entity  $t$ . These messages summarize their respective consistent contextual semantics. Subsequently, we represent the static score for the entity pair  $(h, t)$  as follows:

$$s_{(h,t)} = \sigma \left( \left[ m_{j-1}^h, m_{j-1}^t \right] \cdot W_{j-1} + b_{j-1} \right). \quad (19)$$

Note that Equation (19) should only take the messages of  $h$  and  $t$  as input, without including their connecting edge  $r$ , since the ground-truth relation  $r$  is treated as unobserved during the training stage.

Secondly, we employ the Condition Flow-Matching Module. Given the contextual condition  $\mathbf{z} = (\mathbf{h}^*, \mathbf{t}^*) = (m_{j-1}^h, m_{j-1}^t)$ , where the first symbol representation used here for the CFM module is for the sake of convenience in representation, this module predicts a dynamic vector field  $v_\theta(t, \mathbf{x})$ . This vector field characterizes the instantaneous velocity and direction of the evolution from an initial state  $\mathbf{h}^*$  (i.e.,  $\mathbf{x}$  at  $t = 0$ ) to a target state  $\mathbf{t}^*$  (i.e.,  $\mathbf{x}$  at  $t = 1$ ). We then use this vector field to dynamically modulate the basic static score of the entity pair as follows:

$$s_{(h,t)} = s_{(h,t)} \odot v_\theta(t, \mathbf{x}). \quad (20)$$

Given the flow-modulated score of the entity pair  $s_{(h,t)}$ , we can convert the score into a probability distribution over relations by applying softmax, which then allows us to predict the relations:

$$p(r|h, t) = \text{SoftMax}(s_{(h,t)}). \quad (21)$$

The primary objective for relation prediction is to minimize the cross-entropy loss between the predicted probabilities and the ground-truth relations over the training triplets  $\mathcal{D}$ . This is formulated as minimizing the negative log-likelihood of the true relations:

$$\mathcal{L}_{pred}(\Phi_S, \theta) = - \sum_{(h,r,t) \in \mathcal{D}} \log(p(r|h, t)), \quad (22)$$

where  $\Phi_S$  represents all learnable parameters of the Semantic Context Learning module and the static scoring Linear.

To train the entire model effectively, we adopt a joint training strategy. The overall loss function  $\mathcal{L}$  combines the relation prediction loss with the Condition Flow-Matching loss. The Condition Flow-Matching loss ensures that the learned vector field  $v_\theta(t, \mathbf{x})$  accurately approximates the desired conditional dynamics. The final training objective is to minimize the weighted sum of these losses:

$$\mathcal{L}(\Phi_S, \theta) = \mathcal{L}_{pred}(\Phi_S, \theta) + \lambda \mathcal{L}_{cfm}(\theta), \quad (23)$$

where  $\lambda$  is a hyperparameter that balances the contribution of the two loss terms. The training process of FMS are illustrated in [Algorithm 1](#).

It is worth noticing that the initial context representation derived from  $m_{j-1}^h(\mathbf{h}^*)$  and  $m_{j-1}^t(\mathbf{t}^*)$  plays a dual role: it directly contributes to the base score for relation prediction, and it also serves as the condition  $\mathbf{z}$  for the flow-matching module, thereby influencing the dynamic modulation factor.

To systematically validate the proposed Flow-Modulated Scoring (FMS) framework, our experimental evaluation is structured across two subsequent sections, each addressing a critical and complementary task within Knowledge Graph Completion (KGC).

[Section 4](#) is dedicated to Relation Prediction, the primary focus of this work. This task is not only pivotal for enabling fine-grained reasoning in knowledge graphs but also serves as the principal validation for the core design philosophy of FMS—its ability to capture the contextual dependency and dynamic evolutionary nature of relational semantics. The objective of this chapter is therefore to demonstrate the superior performance and technical advancements of FMS through a rigorous and comprehensive comparison against a wide array of state-of-the-art models.

---

**Algorithm 1** FMS Training Process

---

- 1: **Input:** Batch data  $(h, t)$  and  $r$ , where  $(h, t)$  are entity pairs,  $r$  are relation labels.
  - 2: **Parameters:** Relation embeddings, Semantic Context Learning Module, MLP, Linear.
  - 3: **repeat**
  - 4:    $h^* = \text{Semantic Context Learning Module}(h)$ ;
  - 5:    $t^* = \text{Semantic Context Learning Module}(t)$ ;
  - 6:    $t, x_t, u_t \sim \text{Condition Flow-Matching Module}(h^*, t^*)$ ;
  - 7:    $v_\theta(t, \mathbf{x}) = \text{MLP}([x_t, t])$ ;
  - 8:    $s_{(h,t)} = \text{Linear}([h^*, t^*])$ ;
  - 9:    $s_{(h,t)} = s_{(h,t)} \odot v_\theta(t, \mathbf{x})$ ;
  - 10:    $\mathcal{L}_{cfm}(\theta) = \mathbb{E}_{t,q(z),p_t(x|z)} \|v_\theta(t, \mathbf{x}) - u_t(\mathbf{x}|z)\|_2^2$ ;
  - 11:    $\mathcal{L}_{pred}(\Phi_S, \theta) = -\sum_{(h,r,t) \in \mathcal{D}} \log(p(r|h, t))$ ;
  - 12:    $\mathcal{L}(\Phi_S, \theta) = \mathcal{L}_{pred}(\Phi_S, \theta) + \lambda \mathcal{L}_{cfm}(\theta)$ ;
  - 13:   Update model parameters using gradient descent on  $\mathcal{L}$ .
  - 14: **until** converged
- 

Building upon this foundation, [Section 5](#) extends our assessment to the more fundamental and canonical task of Entity Prediction. While the FMS framework was conceived to address the profound complexities of relation modeling, its powerful representation learning and reasoning capabilities are hypothesized to possess broader applicability. Consequently, this chapter aims to verify the versatility and generalization power of the FMS framework. By adapting the model for entity prediction, we seek to establish that FMS is not merely a specialized solution for a single task but rather a versatile framework capable of empowering a spectrum of KGC challenges. This will comprehensively showcase its potential and practical utility as a novel paradigm for knowledge graph completion.

## 4 RELATION PREDICTION EXPERIMENT

To evaluate the effectiveness of our FMS, we have designed a series of experiments to address the following research questions:

- **RQ1:** How does the performance of FMS compare to a diverse range of mainstream prediction models?
- **RQ2:** What distinct contributions do the key components of FMS offer to the overall performance? Additionally, how does the model’s performance adapt and respond to variations in hyperparameter settings?
- **RQ3:** How does FMS perform in terms of its number of parameters?
- **RQ4:** How does FMS modulate the static scores of triplets to capture rich contextual information shaping relation semantics?
- **RQ5:** How does FMS modulate the static scores of triplets to capture the dynamic evolution of relations?

### 4.1 Transductive Experiment Settings

#### 4.1.1 Datasets

We conduct experiments on six knowledge graph datasets: (i) FB15K [70], derived from Freebase, a large-scale KG of general human knowledge; (ii) FB15K-237 [70], a

**TABLE 1**  
Statistics of all datasets.  $\mathbb{E}[d]$  and  $\text{Var}[d]$  are mean and variance of the node degree distribution, respectively.

Datasets	Nodes	Relations	Train	Val	Test	$\mathbb{E}[d]$	$\text{Var}[d]$
FB15K	15.0k	1.3k	483.1k	50.0k	59.1k	64.6	32.4k
FB15K-237	14.5k	237	272.1k	17.5k	20.5k	37.4	12.3k
WN18	40.9k	18	141.4k	5.0k	5.0k	6.9	236.4
WN18RR	40.9k	11	86.8k	3.0k	3.1k	4.2	64.3
NELL995	63.9k	198	137.5k	5.0k	5.0k	4.3	750.6
DDB14	9.2k	14	36.6k	4.0k	4.0k	7.9	978.8

subset of FB15K where inverse relations are removed; (iii) WN18 [71], containing conceptual-semantic and lexical relations among English words from WordNet; (iv) WN18RR [71], a subset of WN18 where inverse relations are removed; (v) NELL995 [72], extracted from the 995th iteration of the NELL system, containing general knowledge; (vi) DDB14 [19], collected from Disease Database, a medical database containing terminologies, concepts (diseases, symptoms, drugs), and their relationships. The statistics above are summarized in [Table 1](#).

#### 4.1.2 Baselines

Based on the concept of Knowledge Graph Completion, we categorize all baseline methods into four groups: (i) **Emb-based**: TransE [20], ComplEx [40], DistMult [22], RotatE [14], Simple [73], QuatE [74]; (ii) **Rule-based**: DRUM [75], and PTransE [76]; (iii) **GNN-Based**: R-GCN [77], Pathcon [19], LASS [78] and CBLIP [51]; (iv) **LLM-Based**: KG-BERT [29], KGE-BERT [79], and GilBERT [80]

#### 4.1.3 Evaluation Metrics

To ensure a fair comparison, our training and evaluation setup follows that of PathCon [19]<sup>2</sup>. For each test triplet  $(h, r, t)$ , we formulate a query  $(h, ?, t)^3$  where the model’s objective is to predict the correct relation  $r$ . The model is required to rank the ground-truth relation  $r$  against all other relations in the knowledge graph, which serve as negative samples. For instance, a negative candidate would be  $(h, r', t)$ , where  $r' \neq r$ . Analogous to link prediction, we report Mean Reciprocal Rank (MRR) and Hits@N under the filtered setting, which removes any other known true relations between the entity pair  $(h, t)$  from the candidate list before ranking. Higher scores for these metrics indicate better model performance.

#### 4.1.4 Implementation Details

All experiments are conducted on two NVIDIA 4090 GPUs. Our investigations revealed that FMS’s performance is highly sensitive to certain hyperparameters, particularly the number of context hops and the Top-K value, which depend on dataset characteristics. Consequently, to achieve

2. <https://github.com/hyren/PathCon>

3. Some related work formulates this as predicting missing tail (or head) entities given a head (or tail) entity and a relation. These two problems (entity prediction and relation prediction) are reducible to each other. For instance, a relation prediction model  $\Phi(\cdot|h, t)$  can be converted to a tail prediction model  $\Gamma(\cdot|h, r) = \text{SoftMax}_t(\Phi(r|h, t))$ , and vice versa. Given this equivalence, this work focuses on relation prediction. We also discuss the entity prediction task in [section 5](#).

TABLE 2

Best hyper-parameter configurations for FMS. The order for dataset-specific values is: FB15K, FB15K-237, WN18, WN18RR, NELL995, DDB14.

Hyper-parameter	Value / Setting
$\lambda$	1.2
Epoch	20
Dimension (dim)	64
L2 Regularization	$1 \times 10^{-7}$
Learning Rate	$5 \times 10^{-3}$
Neighbor Aggregation	Attention
Temperature	0.95
<i>Dataset-specific Parameters</i>	
Batch Size	128 / 128 / 128 / 128 / 64
Context Hops	2 / 2 / 3 / 3 / 2 / 2
Neighbor Samples	16 / 16 / 8 / 8 / 8 / 8
Num. Top-K	10 / 10 / 4 / 4 / 3 / 3
Num. Attention Heads	4 / 4 / 4 / 4 / 4 / 4

optimal performance, we conducted a comprehensive hyperparameter search independently for each dataset across the following pre-defined spaces:

- Dimension of hidden states: {16, 32, 64, 128};
- Weight of L2 loss term: { $10^{-8}$ ,  $10^{-7}$ ,  $10^{-6}$ ,  $10^{-5}$ };
- Learning rate: {0.0005, 0.005, 0.01, 0.05, 0.1};
- Number of context hops: {1, 2, 3, 4};
- Number of neighbor samples: {8, 16, 24, 32};
- Number of Top-K: {3, 4, 5, 8, 10}.

The best-performing configurations derived from this search are detailed in Table 2. For result stability, each experiment was repeated three times, and we report the average performance. Our implementation is open-sourced and available on GitHub to support reproducibility.<sup>4</sup>

## 4.2 Inductive Experiment Settings

### 4.2.1 Datasets

Following [26], we utilize a total of twelve subsets, encompassing four distinct versions each, which are derived from established datasets: WN18RR, FB15k237, and NELL-995. Each subset is uniquely characterized by a disparate split between the training and test sets, ensuring a robust and comprehensive evaluation schema. For a detailed discourse on the specific splits and a nuanced presentation of the statistical attributes of each subset, one can refer to the comprehensive descriptions available in [26].

### 4.2.2 Baselines

For inductive learning, traditional Embedding-based methods are not suitable for the task due to their inability to generalize to unseen entities, we choose the following two classes of classical and effective models: (i) **Rule-based**: RuleN [81], NeuralLP [82], DRUM [75]; (ii) **GNN-based**: GraIL [17], PathCon [19].

### 4.2.3 Evaluation Metrics

The setup and parameter selection are consistent with the transduction experiments.

4. <https://github.com/yuanwuyuan9/FMS>

## 4.3 Overall Performance Comparison: RQ1

To answer RQ1, we conducted extensive experiments comparing FMS with the strongest baseline we know of, both in terms of transductive and inductive relation prediction.

The transductive results are summarized in Table 3. FMS demonstrates remarkable and consistent improvements across all six datasets. Specifically, FMS achieves a **1.2% relative**, **1.7% relative**, **0.4% relative**, **2.5% relative**, **10.6% relative**, and **1.8% relative** increase in MRR scores over the best baselines on FB15K, FB15k-237, WN18, WN18RR, NELL995, and DDB14, respectively. These consistent and significant gains across diverse datasets, particularly the substantial improvement on NELL995, underscore the robustness and superior capability of our proposed FMS model. Furthermore, it is important to note that, compared to all kinds of models, FMS achieves better results for MRR scores on all datasets, establishing its state-of-the-art performance.

To further analyze the performance of FMS, we decomposed FB15k-237 based on relation categories similar to those defined in [21]. We redefined the 1-N task by specifically isolating a test subset where each entity pair has two or more relation types. As shown in Table 5, FMS achieved overwhelmingly superior performance, indicating its successful capture of the dynamic nature of relations.

The inductive results are summarized in Table 4. FMS shows remarkable improvement across Hits@10 metrics on twelve datasets. It is noteworthy that while the optimal performance of other models typically ranges between 70% and 90% except PathCon, the FMS model almost consistently achieves between 98% and 99.9%, demonstrating its overwhelming advantage in inductive learning capabilities. Furthermore, FMS maintains near-perfect performance across different versions and splits of various datasets, indicating that the model possesses excellent stability and robustness to data variations.

## 4.4 Ablation Study: RQ2

### 4.4.1 Key Modules Ablation Study

To verify the effectiveness of our proposed modules, we develop three distinct model variants:

- **w/o Top-K**: We replace the Top-K selection mechanism based on semantic similarity in the Semantic Context Learning Module with a simple random sampling strategy.
- **w/o Score**: We remove the energy-model inspired scoring function, and instead use dot product similarity for scoring.
- **w/o Flow-Matching**: We remove the Condition Flow-Matching Module, relying solely on the basic static scores of entity pairs.

The ablation study results are presented in Table 6, leading to the following key conclusions: The Top-K selection mechanism exerts the most significant impact on the overall FMS performance, with its absence leading to a substantial decline in performance, particularly on the WN18RR dataset. This underscores Top-K's pivotal role in filtering critical information and suppressing noise. Secondly, the energy-model inspired scoring function is also crucial for

TABLE 3

Transductive performance comparison on all six datasets. The results here are expressed as percentages. The best results are shown in bold while the second-best results are shown in the underline.

Type	Model	FB15K			FB15K-237			WN18			WN18RR			NELL995			DDB14		
		MRR	H@1	H@3	MRR	H@1	H@3	MRR	H@1	H@3	MRR	H@1	H@3	MRR	H@1	H@3	MRR	H@1	H@3
Emb-based	TransE	96.2	94.0	98.2	96.6	94.6	98.4	97.1	95.5	98.4	78.4	66.9	87.0	84.1	78.1	88.9	96.6	94.8	98.0
	ComplEx	90.1	84.4	95.2	92.4	87.9	97.0	98.5	97.9	99.1	84.0	77.7	88.0	70.3	62.5	76.5	95.3	93.1	96.8
	DistMult	66.1	43.9	86.8	87.5	80.6	93.6	78.6	58.4	98.7	84.7	78.7	89.1	63.4	52.4	72.0	92.7	88.6	96.1
	RotatE	97.9	96.7	98.6	97.0	95.1	98.0	98.4	97.9	98.6	79.9	73.5	82.3	72.9	69.1	75.6	95.3	93.4	96.4
	SimplE	98.3	97.2	99.1	97.1	95.5	98.7	97.2	96.4	97.6	73.0	65.9	75.5	71.6	67.1	74.8	92.4	89.2	94.8
	QuatE	98.3	97.2	99.1	97.4	95.8	98.8	98.1	97.5	98.3	82.3	76.7	85.2	75.2	70.6	78.3	94.6	92.2	96.2
Rule-based	DRUM	94.5	94.5	97.8	95.9	90.5	95.8	96.9	95.6	98.0	85.4	77.8	91.2	71.5	64.0	74.0	95.8	93.0	98.7
	PTransE	-	-	-	96.2	96.0	98.5	98.7	98.2	98.9	85.3	87.5	94.7	-	-	-	-	-	-
GNN-based	R-GCN	-	-	-	93.1	90.3	94.2	90.9	82.4	97.9	82.2	75.0	84.8	-	-	-	-	-	-
	Pathcon	98.4	97.4	99.5	97.9	96.4	99.4	99.3	98.8	99.8	97.4	95.4	99.4	89.6	84.4	94.1	98.0	96.6	99.5
	LASS	-	-	-	97.3	95.5	98.8	96.7	95.9	98.0	95.9	94.2	96.4	-	-	-	-	-	-
	CBLiP	-	-	-	97.1	94.9	99.2	-	-	-	97.6	96.0	99.3	91.9	86.8	96.4	-	-	-
LLM-based	KG-BERT	-	-	-	97.3	95.2	98.6	96.7	94.6	98.2	94.2	91.7	96.8	-	-	-	-	-	-
	KGE-BERT	-	-	-	97.2	94.3	99.0	96.5	95.7	97.9	94.8	92.4	96.1	-	-	-	-	-	-
	GilBERT	-	-	-	95.1	91.4	97.0	96.5	94.7	98.1	94.1	91.6	96.7	-	-	-	-	-	-
Ours	FMS	99.6	99.5	99.8	99.8	99.7	99.9	99.7	99.5	99.9	99.9	99.8	99.9	99.1	98.6	99.5	99.8	99.7	99.9

TABLE 4

Inductive results (Hits@10). The best results are shown in bold while the second-best results are shown underlined.

Type	Model	WN18RR				FB15k-237				NELL-995			
		v1	v2	v3	v4	v1	v2	v3	v4	v1	v2	v3	v4
Rule-based	Neural-LP	74.4	68.9	46.2	67.1	52.9	58.9	52.9	55.9	40.8	78.7	82.7	80.6
	DRUM	74.4	68.9	46.2	67.1	52.9	58.7	52.9	55.9	19.4	78.6	82.7	80.6
	RuleN	80.9	78.2	53.4	71.6	49.8	77.8	87.7	85.6	53.5	81.8	77.3	61.4
GNN-based	GraIL	82.5	78.7	58.4	73.4	64.2	81.8	82.8	89.3	59.5	93.3	91.4	73.2
	PathCon	97.6	97.1	97.4	98.7	94.1	95.6	95.9	96.0	-	95.8	94.5	95.6
Ours	FMS	99.2	98.4	99.0	99.5	98.4	99.5	99.6	99.9	-	99.5	99.2	99.8

TABLE 5  
Results by relation category on FB15k-237.

Model	1-1		1-N	
	MRR	Hit@1	MRR	Hit@1
PathCon	92.4	88.9	80.9	74.2
FMS	99.8	99.7	99.6	99.3

TABLE 6  
Ablation study on key components of FMS.

Ablation Settings	FB15k-237		WN18RR	
	MRR	Hit@1	MRR	Hit@1
FMS	99.8	99.7	99.9	99.8
FMS w/o Top-K	97.8	96.1	94.3	89.4
FMS w/o Score	98.2	97.1	97.4	95.9
FMS w/o Flow-Matching	98.7	97.9	98.4	97.5

evaluating and ranking candidate answers, and its removal results in a noticeable loss in performance. While each component is crucial, their collective impact exceeds the sum of their parts. This highlights that the synergistic interaction of semantic context selection, energy-based scoring, and dynamic flow modulation is the key factor enabling the FMS model to maintain its state-of-the-art performance.

#### 4.4.2 Sensitivity to Key Hyperparameters

In this study, we analyze the impact of key hyperparameters within the Semantic Context Learning Module and the effect of different aggregator choices. The results on the FB15k-237 dataset are presented in Figure 4.

Figure 4(a) shows how FMS performance varies with different Top-K values. The results demonstrate that a moderate K value (specifically 10 in this case) is critical. An overly small K provides insufficient context to leverage, while an excessively large K introduces increasing noise, potentially overwhelming the model with excessive contextual information and causing a drop in performance.

Figure 4(b) then shows the performance of FMS’s Semantic Context Learning Module when using different context aggregators. It can be seen that our designed self-attention aggregator performs best, while the mean aggregator performs worst.

Figure 4(c) illustrates the selection of hop count on the graph. The results indicate that FMS achieves competitive performance at the first hop and optimal performance at the second hop. Further increasing the hop count, however, leads to performance degradation. This phenomenon is likely attributable to the exponential increase in neighbor sampling associated with deeper hops, which consequently introduces more noise. Nevertheless, owing to the mech-

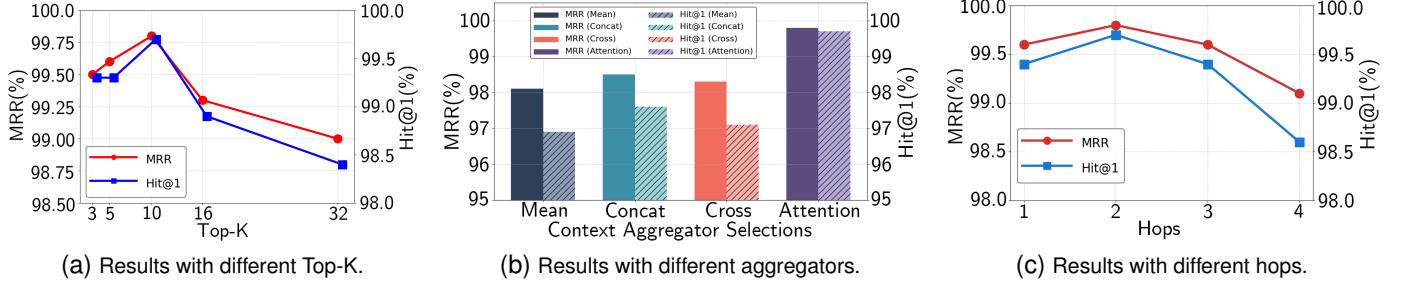


Fig. 4. Hyper-parameters analysis of FMS on FB15k-237.

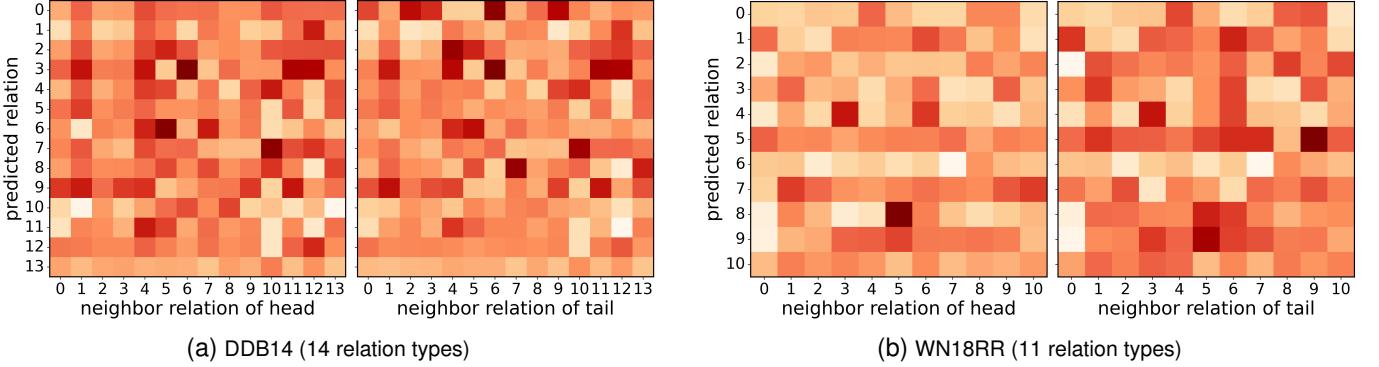


Fig. 5. The learned correlation between the contextual relations of head/tail entities and the predicted relations. Case studies are shown for the DDB14 dataset (a) and the WN18RR dataset (b).

TABLE 7  
Number of parameters of all models on DDB14.

Model	TransE	DisMult	RotatE	QuatE	PathCon	FMS
Param	3.7M	3.7M	7.4M	14.7M	0.06M	<b>0.02M</b>

TABLE 8  
Number of parameters of all models on FB15k-237.

Model	TransE	DisMult	RotatE	QuatE	PathCon	FMS
Param	5.9M	5.9M	11.7M	23.6M	1.67M	<b>0.35M</b>

anisms of Top-k selection and attention aggregation, the model maintains its state-of-the-art performance.

#### 4.5 Efficiency Analysis: RQ3

We performed parametric volume analysis on both datasets, on the smaller dataset DDB14 as in Table 7 and on the larger dataset FB15k-237 as in Table 8. The model complexity is  $\mathcal{O}(n \cdot K^{Hops})$ . The result demonstrates that FMS is much more storage-efficient than embedding-based methods, since it does not need to calculate entity embeddings. Furthermore, FMS is significantly more effective than GNN-based methods such as PathCon. FMS can achieve superior performance with fewer hops and a smaller number of sampled neighbors. Additionally, it does not require explicitly modeling paths from the head entity to the tail entity like PathCon; instead, it employs flow-matching to learn a dynamic evolution vector field, which represents semantically context-guided paths.

#### 4.6 Case Study on Contextual Information: RQ4

To demonstrate the robust capability of our FMS in capturing the semantic context of relations, we present heatmaps

TABLE 9  
Relation IDs and Meanings on DDB14 and WN18RR Datasets.

DDB14 Dataset			
ID	Relation Meaning	ID	Relation Meaning
0	belong(s) to the category of	7	interacts with
1	is a category subset of	8	belongs to the drug family of
2	may cause	9	belongs to drug super-family
3	is a subtype of	10	is a vector for
4	is a risk factor for	11	may be allelic with
5	is associated with	12	see also
6	may contraindicate	13	is an ingredient of

WN18RR Dataset			
ID	Relation Meaning	ID	Relation Meaning
0	hypernym	6	has_part
1	derivationally_related_form	7	member_of_domain_usage
2	instance_hypernym	8	member_of_domain_region
3	also_see	9	verb_group
4	member_meronym	10	similar_to
5	synset_domain_topic_of		

of the modulated context-aggregated head/tail entity matrix on DDB14 (14 relation types) and WN18RR (11 relation types) in Figure 5. The intensity of the entries in Figure 5 represents the strength of the correlation between the presence of a contextual relation and a predicted relation. Relation IDs and their corresponding meanings for DDB14 and WN18RR are detailed in Table 9.

Figure 5 reveals a sophisticated learning mechanism that moves beyond superficial pattern matching. The prevalence of strong off-diagonal correlations indicates that the model transcends simple pattern memorization—typified by rules like  $(h, R, t') \Rightarrow (h, R, t)$  [19]. Instead, the model constructs a network of latent semantic rules, which we categorize into the following core principles.

- **Domain-Specific Rule Induction:** FMS demonstrates a remarkable ability to induce logical rules that are specifically adapted to the intrinsic properties of a given knowledge domain.

In DDB14, it captures a clinical cause-and-effect relationship, inferring a rule that connects risk factors to clinical contraindications:

$$(h, \text{is\_risk\_factor\_for}, t') \Rightarrow (h, \text{may\_contraindicate}, t).$$

In WN18RR, it models a fundamental linguistic principle linking etymology to semantics:

$$(h, \text{etymologically\_related}, t') \Rightarrow (h, \text{usage\_domain}, t).$$

- **Modeling of Hierarchical Structures:** FMS effectively captures hierarchical knowledge structures, enabling it to perform complex, multi-level reasoning from specific instances to general categories.

In DDB14, it masters compositional reasoning, inferring macroscopic class membership from microscopic constituents:

$$(h, \text{is\_ingredient\_of}, t') \Rightarrow (h, \text{belongs\_to\_drug\_family}, t).$$

In WN18RR, it demonstrates a form of inductive inference, leveraging horizontal similarity to determine vertical class membership:

$$\begin{aligned} (h, \text{similar\_to}, h') \wedge (h', \text{instance\_hypernym}, t) \\ \Rightarrow (h, \text{instance\_hypernym}, t). \end{aligned}$$

#### 4.7 Case Study on Dynamic Evolution: RQ5

To quantitatively evaluate the impact of the Flow-Modulated Scoring (FMS) framework, we investigate a case study from the FB15k-237 dataset. The task, detailed in Table 10, involves predicting the multiple valid financial relationships between an arts college and a currency.

Figure 6 presents T-SNE [83] visualizations to compare the representations learned by FMS against those from traditional static scoring methods.

Central to the figure, the "Static Score Embedding" (black square) represents the generalized initial understanding that a traditional static model might offer for the relation between "California College of the Arts" and "United States Dollar". Due to its singular nature, this single point struggles to simultaneously and precisely encapsulate the specific role of "United States Dollar" as a currency across six different financial contexts. The power of FMS is evident in its subsequent dynamic processing. Observing the "Dynamic Scores" (colored stars) in the figure, we see that FMS is not content with this single, potentially ambiguous static representation. Instead, it actively flows from this static starting point to generate specific embedding representations for each potential relation. Each arrow in the figure symbolizes this context-driven modulation process, guiding the initial representation towards more relevant and fine-grained semantic space regions associated with specific relation types.

FMS successfully distinguishes between multiple concurrently valid relations. It does not merely identify the generic concept of "currency"; rather, it can precisely differentiate specific relations such as "Revenue Currency" and "Assets Currency" within the embedding space. This clearly

TABLE 10  
Case study example from the FB15k-237 dataset.

Component	Value
Head Entity	California College of the Arts
Tail Entity	United States Dollar
Query	(California College of the Arts, ?, United States Dollar)
Ground-Truth	<ul style="list-style-type: none"> <li>• Revenue Currency</li> <li>• Operating Income Currency</li> <li>• Local Tuition Currency</li> <li>• Assets Currency</li> <li>• Domestic Tuition Currency</li> <li>• Endowment Currency</li> </ul>

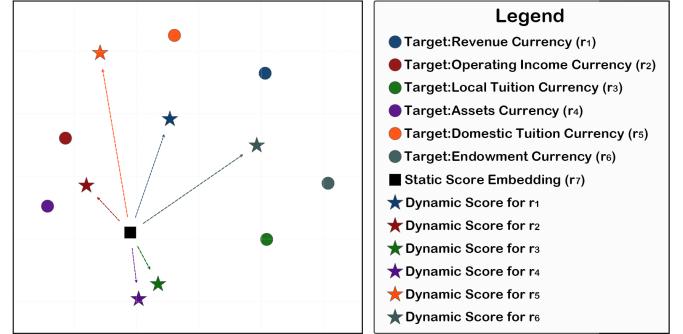


Fig. 6. T-SNE visualization of Static Score versus Flow-Modulated Score representations on FB15k-237.

demonstrates its exceptional performance in capturing the dynamic nature of knowledge graph relations.

## 5 ENTITY PREDICTION EXPERIMENT

Building on its demonstrated superiority in relation prediction, we further validate FMS's versatility and power on the fundamental task of entity prediction. As established previously, these two tasks are inter-reducible. To formally adapt our model for this task, we model the probability distribution over all candidate entities. Specifically, for a tail prediction query  $(h, r, ?)$ , the probability of an entity  $t$  is given by applying a SoftMax function over the scores derived from our model FMS:

$$p(t|h, r) = \text{SoftMax}(s_{(h,t)}). \quad (24)$$

Head prediction  $(?, r, t)$  is handled symmetrically.

It is crucial to note that this adaptation necessitates a change in the training paradigm. Unlike in relation prediction where the target relation  $r$  is masked from the input pair  $(h, t)$ , the model for entity prediction must be trained with the relation  $r$  as part of the input condition to properly predict the target entity. This rigorous approach allows us to confirm whether the superior performance of FMS extends to this fundamental KG completion task.

### 5.1 Transductive Settings

#### 5.1.1 Datasets

We conduct experiments on four knowledge graph datasets: (i) FB15k-237 (see Section 4.1 for details); (ii) NELL995,

TABLE 11

Transductive results. Performance comparison on FB15k-237, NELL995, Kinship and UMLS datasets. Best results are in **bold**, second best are underlined.

Type	Model	FB15k-237			NELL995			Kinship			UMLS		
		MRR	H@1	H@3									
Emb-based	TransE	28.9	19.8	32.4	32.6	22.5	39.9	26.3	0.5	41.8	24.7	7.3	29.7
	ComplEx	24.7	15.8	27.5	23.7	19.7	25.2	41.8	24.2	49.9	41.1	27.3	46.8
	RotaE	33.7	24.1	37.5	35.6	29.8	39.5	65.1	50.4	75.5	74.4	63.6	82.2
	ConvE	32.5	23.7	35.6	51.1	44.6	-	68.5	55.2	78.6	75.6	69.7	80.7
	TDN	35.0	26.3	39.5	-	-	-	78.0	67.7	86.7	86.0	82.4	87.3
Rule-based	RNNLogic	34.4	25.2	38.0	41.6	36.3	-	72.2	59.8	81.4	84.2	77.2	89.1
	DRUM	23.8	17.4	26.1	53.2	46.0	-	33.4	18.3	37.8	54.8	35.8	69.9
GNN-based	PathCon	50.6	39.6	56.3	49.5	44.4	51.9	78.8	65.6	<u>91.5</u>	90.1	85.7	93.2
	NBFNet	41.5	32.1	45.4	52.5	45.1	-	60.6	43.5	72.5	77.8	68.8	84.0
	RED-GNN	38.2	31.7	42.9	54.9	46.3	<u>60.9</u>	78.5	69.0	85.4	96.3	94.5	<u>97.6</u>
	AdaProp	42.5	31.0	47.0	55.2	<u>49.7</u>	58.6	83.6	74.9	91.1	<u>96.4</u>	<u>94.6</u>	<u>97.9</u>
Diffusion-based	FDM	48.5	38.6	52.9	-	-	-	83.7	<u>76.1</u>	89.7	92.2	89.3	94.4
	KGDM	52.0	<u>42.3</u>	<u>56.6</u>	-	-	-	78.9	68.7	87.0	90.9	87.2	93.7
	DiffusionE	37.6	29.4	-	52.2	49.0	-	-	-	-	<u>97.0</u>	<u>95.7</u>	-
Ours	FMS	<b>67.2</b>	<b>57.4</b>	<b>72.9</b>	<b>60.6</b>	<b>54.9</b>	<b>65.8</b>	<b>99.3</b>	<b>98.7</b>	<b>99.9</b>	92.5	89.0	94.8

specifically an extended version (see [Section 4.1](#) for background); (iii) Kinship, a knowledge graph describing kinship relations, focusing on family and blood relations [84]; (iv) UMLS, a biomedical knowledge graph encompassing medical terminology, concepts, and their interrelationships [85]. The statistics above are summarized in [Table 13](#).

### 5.1.2 Baselines

We categorize all baseline methods into four groups: (i) **Embedding-based**: TransE [20], ComplEx [40], RotatE [14], ConvE [71], TDN [86], FDM [37]; (ii) **Rule-based**: RNN-logic [87], DRUM [75]; (iii) **GNN-based**: Pathcon [19], NBFNet [88], RED-GNN [26], AdaProp [89]; (iv) **Diffusion-Based**: FDM [37], KGDM [47], DiffusionE [68].

### 5.1.3 Evaluation Metrics

To ensure a fair comparison, our training and evaluation setup follows that of RED-GNN [26]<sup>5</sup>. For evaluation, each test triplet  $(h, r, t)$  is used to create two entity prediction queries: a tail prediction query  $(h, r, ?)$  and a head prediction query  $(?, r, t)$ . The objective is to rank the ground-truth entity ( $t$  for the former,  $h$  for the latter) against a set of candidate entities. These candidates are generated by corrupting the triplet; for instance, in tail prediction, the ground truth  $t$  is replaced by every other entity  $t'$  in the knowledge graph to form negative samples  $(h, r, t')$ . We report Mean Reciprocal Rank (MRR) and Hits@N under the standard filtered setting [14], which ensures that any other valid triplets appearing in the candidate set are removed before ranking.

### 5.1.4 Implementation Details

Our experimental setup is largely consistent with the details provided in [Section 4.1](#). However, we make specific

adjustments to three key hyperparameters: the number of context hops, the number of neighbor samples, and the Top-K value. This modification is motivated by the significant disparity between the number of entities and relations; for instance, FB15k-237 contains 14,541 entities but only 237 relations. Modeling the probability distribution over a much larger set of entities necessitates richer semantic information. Therefore, we uniformly set the number of context hops to 2 and the Top-K value to 10 across all datasets. The number of neighbor samples is generally set to 16, with a specific exception for the NELL-995 dataset, for which we use a value of 10.

## 5.2 Inductive Settings

### 5.2.1 Datasets

Building upon this, we proceed with four subsets each from the FB15k237 and NELL-995 datasets, for a detailed examination of the dataset specifics, refer to [Section 4.2](#).

### 5.2.2 Baselines

For inductive learning, traditional Embedding-based methods are not suitable for the task due to their inability to generalize to unseen entities, we choose the following two classes of classical and effective models: (i) **Rule-based**: RuleN [81], NeuralLP [82], DRUM [75]; (ii) **GNN-based**: GraIL [17], PathCon [19].

### 5.2.3 Baselines

Since completion methods that learn entity embeddings during training cannot work in this setting, we select baselines from the following three categories: (i) **Rule-based methods**: NeuralLP [82], DRUM [75], and RuleN [81]. (ii) **GNN-based methods**: GraIL [17], PathCon [19], NBFNet [88], RED-GNN [26], and AdaProp [89]. (iii) **Diffusion-based methods**: DiffusionE [68]. It is worth

5. <https://github.com/LARS-research/RED-GNN>

TABLE 12  
MRR comparison for entity prediction by multiple relation category on FB15k-237.

	TransE	RotatE	NBFNet	FDM	FMS
<b>Head Pred</b>					
1-N	7.9	8.1	16.5	<u>20.3</u>	<b>39.8</b>
N-1	45.5	46.7	49.9	<u>55.9</u>	<b>69.0</b>
N-N	22.4	23.4	34.8	<u>42.3</u>	<b>67.7</b>
<b>Tail Pred</b>					
1-N	74.4	74.7	79.0	<u>82.6</u>	<b>85.1</b>
N-1	7.1	7.0	12.2	<u>16.7</u>	<b>33.9</b>
N-N	33.0	33.8	45.6	<u>54.3</u>	<b>66.5</b>

TABLE 13

Statistics of all datasets.  $\mathbb{E}[d]$  and  $\text{Var}[d]$  are mean and variance of the node degree distribution, respectively.

Dataset	Nodes	Relations	Train	Val	Test	$\mathbb{E}[d]$	$\text{Var}[d]$
NELL995	74.5k	200	37.4k	543	2.8k	2.6	124.7
Kinship	104	25	3.2k	2.1k	5.3k	205.5	2.8
UMLS	135	46	1.3k	569	633	37.4	1205.3

noting that KGDM and FDM use an embedding-based approach to encode entities and relations; therefore, they are not considered in the inductive learning setting.

#### 5.2.4 Evaluation Metrics

The setup and parameter selection are consistent with the transduction entity prediction experiments.

### 5.3 Overall Performance

To evaluate FMS on entity prediction, we conducted a parallel set of experiments, with results presented for both transductive and inductive settings.

The transductive results are summarized in Table 11. FMS demonstrates superior performance across the majority of datasets and metrics. Specifically, FMS achieves a remarkable **25.2% relative** increase in MRR over the strongest baseline on FB15k-237, a **5.4% relative** increase on NELL995, and an **15.6% relative** increase on Kinship. Particularly on the Kinship dataset, FMS achieves a near-perfect MRR of 99.3, showcasing its exceptional modeling capacity on

dense, complex graph structures. While DiffusionE shows the best performance on UMLS, FMS still delivers highly competitive results, establishing itself as a new state-of-the-art model for transductive entity prediction on diverse benchmarks.

To delve deeper into its capabilities, we analyze FMS’s performance on different relation categories within FB15k-237, as detailed in Table 12. FMS achieves overwhelmingly superior performance across all prediction scenarios (1-N, N-1, N-N for both Head and Tail prediction). The improvements are especially pronounced in more challenging categories. For instance, in the N-1 tail prediction task, FMS more than doubles the MRR of the best baseline, increasing it from 16.7 to 33.9. This uniform dominance indicates that FMS successfully captures the intricate and varied semantic mappings between entities, demonstrating its exceptional ability to handle complex reasoning scenarios across all relation types.

The inductive entity prediction results are presented in Table 14. FMS consistently and significantly outperforms all baselines across all versions of the FB15k-237 and NELL-995 datasets, often by a large margin. For example, on FB15k-237 v4, FMS achieves an MRR of 57.6, surpassing the next-best model’s 45.4. Similarly, on NELL-995 v3, FMS’s MRR of 54.7 is substantially higher than the runner-up’s 45.8. Furthermore, the model’s strong and stable performance across different subgraph versions (v1 to v4) highlights its excellent robustness and generalization capabilities in unseen entity settings. This consistent and significant outperformance across diverse settings underscores the powerful generalization capabilities of FMS in inductive learning.

## 6 CONCLUSION

In this work, we introduced Flow-Modulated Scoring (FMS), a novel framework designed to address the limitations of existing knowledge graph completion methods in modeling the semantic contextual dependency and dynamic nature of relations. FMS innovatively combines a semantic context learning module, which generates context-aware entity embeddings, with a condition flow-matching

TABLE 14  
Inductive results. The best results are shown in bold while the second-best results are shown underlined.

Type	Model	FB15k-237								NELL-995							
		v1		v2		v3		v4		v1		v2		v3		v4	
		MRR	H@10														
Rule-based	Neural-LP	32.5	46.8	38.9	58.6	40.0	57.1	39.6	59.3	61.0	<u>87.1</u>	36.1	56.4	36.7	57.6	26.1	53.9
	DRUM	33.3	<b>47.4</b>	39.5	59.5	40.2	57.1	41.0	59.3	62.8	<u>87.3</u>	36.5	54.0	37.5	57.7	27.3	53.1
	RuleN	36.3	44.6	43.3	59.9	43.9	60.0	42.9	60.5	61.5	76.0	38.5	51.4	38.1	53.1	33.3	48.4
GNN-based	GraIL	27.9	42.9	27.6	42.4	25.1	42.4	22.7	38.9	48.1	56.5	29.7	49.6	32.2	51.8	26.2	50.6
	PathCon	31.7	53.9	33.3	54.7	36.1	54.7	40.1	61.3	-	-	37.3	59.4	45.6	59.0	29.2	44.8
	NBFNet	30.7	51.7	36.9	63.9	33.1	58.8	30.5	55.9	58.4	79.5	41.0	63.5	42.5	60.6	28.7	<u>59.1</u>
	RED-GNN	<u>36.9</u>	48.3	46.9	62.9	44.5	60.3	44.2	62.1	63.7	86.6	41.9	60.1	43.6	59.4	36.3	55.6
	AdaProp	31.0	<b>55.1</b>	47.1	65.9	<u>47.1</u>	63.7	<u>45.4</u>	63.8	64.4	<b>88.6</b>	<u>45.2</u>	65.2	43.5	61.8	<u>36.6</u>	<b>60.7</b>
Diffusion-based	DiffusionE	31.0	52.5	<u>47.4</u>	<u>66.1</u>	45.6	62.7	44.6	62.5	<b>67.3</b>	<u>87.3</u>	42.4	<u>66.0</u>	<u>45.8</u>	62.0	30.8	52.3
Ours	FMS	<b>42.3</b>	<b>63.3</b>	<b>52.8</b>	<b>70.6</b>	<b>53.5</b>	<b>73.3</b>	<b>57.6</b>	<b>75.6</b>	-	-	<b>50.4</b>	<b>68.8</b>	<b>54.7</b>	<b>67.6</b>	<b>39.1</b>	53.9

module that learns the evolution dynamics between entities based on this context. By using the predicted vector field to dynamically modulate a base static score, FMS effectively synergizes rich static representations with conditioned dynamic flow information. Our comprehensive experiments on several well-established benchmarks demonstrate that FMS achieves state-of-the-art performance, validating its capability to capture more comprehensive relational semantics and advance the field of knowledge graph completion.

Despite its empirical success, a primary limitation of FMS lies in its significant reliance on the semantic quality of relational features. The model's performance is intrinsically tied to its ability to capture expressive relational patterns. Consequently, on large-scale knowledge graphs characterized by a vast and diverse set of relations—particularly where some relations may be sparse or semantically ambiguous—the efficacy of FMS may be constrained.

To address this limitation, a promising avenue for future research is the incorporation of entity-level features. By jointly modeling both entity and relational semantics, we can potentially achieve a more holistic and robust representation of factual knowledge. However, this extension introduces a non-trivial trade-off: integrating rich entity features would significantly increase the model's parameter count and computational overhead. Therefore, a key challenge will be to strike a balance between enhanced expressiveness and model efficiency. This necessitates the exploration of more scalable encoding and aggregation mechanisms, such as parameter-sharing techniques or lightweight attention modules, to create a more comprehensive yet practical model.

## REFERENCES

- [1] Z. Cui, Y. Weng, X. Tang, F. Lyu, D. Liu, X. He, and C. Ma, "Comprehending knowledge graphs with large language models for recommender systems," in *Proceedings of the 48th International ACM SIGIR Conference on Research and Development in Information Retrieval*, ser. SIGIR '25. New York, NY, USA: Association for Computing Machinery, 2025, p. 1229–1239. [Online]. Available: <https://doi.org/10.1145/3726302.3729932>
- [2] Z. Chen, W. Gan, J. Wu, K. Hu, and H. Lin, "Data scarcity in recommendation systems: A survey," *ACM Transactions on Recommender Systems*, vol. 3, no. 3, pp. 1–31, 2025.
- [3] Y. Zhang, X. Sui, F. Pan, K. Yu, K. Li, S. Tian, A. Erdengasileng, Q. Han, W. Wang, J. Wang *et al.*, "A comprehensive large-scale biomedical knowledge graph for ai-powered data-driven biomedical research," *Nature Machine Intelligence*, pp. 1–13, 2025.
- [4] R. Omar, I. Dhall, P. Kalnis, and E. Mansour, "A universal question-answering platform for knowledge graphs," *Proceedings of the ACM on Management of Data*, vol. 1, no. 1, pp. 1–25, 2023.
- [5] Y. Lu, H. Chen, Y. Rao, J. Yu, W. Hua, and Q. Li, "An efficient fuzzy system for complex query answering on knowledge graphs," *IEEE Transactions on Knowledge and Data Engineering*, vol. 37, no. 9, pp. 4962–4976, 2025.
- [6] D. N. Nicholson and C. S. Greene, "Constructing knowledge graphs and their biomedical applications," *Computational and Structural Biotechnology Journal*, vol. 18, pp. 1414–1428, 2020.
- [7] A. Hogan, E. Blomqvist, M. Cochez, C. d'Amato, G. de Melo, C. Gutierrez, S. Kirrane, J. E. Labra Gayo, R. Navigli, S. Neumaier *et al.*, "Knowledge graphs," *ACM Computing Surveys (CSUR)*, vol. 54, no. 4, pp. 1–37, 2021.
- [8] J. Dong, Q. Zhang, X. Huang, K. Duan, Q. Tan, and Z. Jiang, "Hierarchy-aware multi-hop question answering over knowledge graphs," in *Proceedings of the ACM Web Conference 2023*, 2023, pp. 2519–2527.
- [9] K. Liang, L. Meng, M. Liu, Y. Liu, W. Tu, S. Wang, S. Zhou, X. Liu, F. Sun, and K. He, "A survey of knowledge graph reasoning on graph types: Static, dynamic, and multi-modal," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 46, no. 12, pp. 9456–9478, 2024.
- [10] S. Pan, L. Luo, Y. Wang, C. Chen, J. Wang, and X. Wu, "Unifying large language models and knowledge graphs: A roadmap," *IEEE Transactions on Knowledge and Data Engineering*, vol. 36, no. 7, pp. 3580–3599, 2024.
- [11] S. Ji, S. Pan, E. Cambria, P. Marttinen, and S. Y. Philip, "A survey on knowledge graphs: Representation, acquisition, and applications," *IEEE transactions on neural networks and learning systems*, vol. 33, no. 2, pp. 494–514, 2021.
- [12] Y. Zhang, Q. Yao, and J. T. Kwok, "Bilinear scoring function search for knowledge graph learning," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 2, pp. 1458–1473, 2023.
- [13] X. Zhu, Z. Li, X. Wang, X. Jiang, P. Sun, X. Wang, Y. Xiao, and N. J. Yuan, "Multi-modal knowledge graph construction and application: A survey," *IEEE Transactions on Knowledge and Data Engineering*, vol. 36, no. 2, pp. 715–735, 2024.
- [14] Z. Sun, Z.-H. Deng, J.-Y. Nie, and J. Tang, "Rotate: Knowledge graph embedding by relational rotation in complex space," *arXiv preprint arXiv:1902.10197*, 2019.
- [15] D. Guo, H. Wang, and M. Wang, "Context-aware graph inference with knowledge distillation for visual dialog," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 10, pp. 6056–6073, 2022.
- [16] Y. Cui, Z. Sun, and W. Hu, "Transfer-and-fusion: Integrated link prediction across knowledge graphs," *IEEE Transactions on Knowledge and Data Engineering*, vol. 37, no. 5, pp. 3062–3074, 2025.
- [17] K. Teru, E. Denis, and W. Hamilton, "Inductive relation prediction by subgraph reasoning," in *International conference on machine learning*. PMLR, 2020, pp. 9448–9457.
- [18] J. Wang, Z. Zhang, Z. Shi, J. Cai, S. Ji, and F. Wu, "Duality-induced regularizer for semantic matching knowledge graph embeddings," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 2, pp. 1652–1667, 2023.
- [19] H. Wang, H. Ren, and J. Leskovec, "Relational message passing for knowledge graph completion," in *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, 2021, pp. 1697–1707.
- [20] A. Bordes, N. Usunier, A. Garcia-Duran, J. Weston, and O. Yakhnenko, "Translating embeddings for modeling multi-relational data," *Advances in neural information processing systems*, vol. 26, 2013.
- [21] Z. Wang, J. Zhang, J. Feng, and Z. Chen, "Knowledge graph embedding by translating on hyperplanes," in *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, vol. 28, no. 1, 2014.
- [22] B. Yang, W.-t. Yih, X. He, J. Gao, and L. Deng, "Embedding entities and relations for learning and inference in knowledge bases," in *International Conference on Learning Representations (ICLR)*, 2015.
- [23] B. Jagvaral, W.-K. Lee, J.-S. Roh, M.-S. Kim, and Y.-T. Park, "Path-based reasoning approach for knowledge graph completion using cnn-bilstm with attention mechanism," *Expert Systems with Applications*, vol. 142, p. 112960, 2020.
- [24] W. Guo, Z. Li, X. Wang, Z. Chen, J. Zhao, J. Li, and Y. Yuan, "ConvD: Attention enhanced dynamic convolutional embeddings for knowledge graph completion," *IEEE Transactions on Knowledge and Data Engineering*, vol. 37, no. 9, pp. 5049–5062, 2025.
- [25] S. Dutta, A. Gittens, M. J. Zaki, and C. C. Aggarwal, "Replacing paths with connection-biased attention for knowledge graph completion," ser. AAAI'25/IAAI'25/EAAI'25. AAAI Press, 2025. [Online]. Available: <https://doi.org/10.1609/aaai.v39i14.33632>
- [26] Y. Zhang and Q. Yao, "Knowledge graph reasoning with relational digraph," in *Proceedings of the ACM web conference 2022*, 2022, pp. 912–924.
- [27] Y. Zhang, Z. Chen, L. Guo, Y. Xu, W. Zhang, and H. Chen, "Making large language models perform better in knowledge graph completion," in *Proceedings of the 32nd ACM international conference on multimedia*, 2024, pp. 233–242.
- [28] Y. Zhu, X. Wang, J. Chen, S. Qiao, Y. Ou, Y. Yao, S. Deng, H. Chen, and N. Zhang, "Llms for knowledge graph construction and reasoning: recent capabilities and future opportunities," *World Wide Web*, vol. 27, no. 5, Aug. 2024. [Online]. Available: <https://doi.org/10.1007/s11280-024-01297-w>
- [29] L. Yao, C. Mao, and Y. Luo, "Kg-bert: Bert for knowledge graph completion," *arXiv preprint arXiv:1909.03193*, 2019.

- [30] Q. Li, Z. Chen, C. Ji, S. Jiang, and J. Li, "Llm-based multi-level knowledge generation for few-shot knowledge graph completion," in *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence*, ser. IJCAI '24, 2024. [Online]. Available: <https://doi.org/10.24963/ijcai.2024/236>
- [31] C. E. Quintero-Narvaez and R. Monroy, "Integrating knowledge graph data with large language models for explainable inference," in *Proceedings of the 17th ACM International Conference on Web Search and Data Mining*, 2024, pp. 1198–1199.
- [32] Y. Zhao, H. Lin, S. Wen, J. Shen, and B. Hua, "Sma-gnn: A symbol-aware graph neural network for signed link prediction in recommender systems," in *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V.2*, ser. KDD '25. New York, NY, USA: Association for Computing Machinery, 2025, p. 3957–3967. [Online]. Available: <https://doi.org/10.1145/3711896.3737132>
- [33] N. Ibrahim, S. Aboulela, A. Ibrahim, and R. Kashef, "A survey on augmenting knowledge graphs (kgs) with large language models (llms): models, evaluation metrics, benchmarks, and challenges," *Discover Artificial Intelligence*, vol. 4, no. 1, p. 76, 2024.
- [34] S. Liang, J. Shao, D. Zhang, J. Zhang, and B. Cui, "Drgi: Deep relational graph infomax for knowledge graph completion," *IEEE Transactions on Knowledge and Data Engineering*, vol. 35, no. 3, pp. 2486–2499, 2023.
- [35] Q. Wang, C. Li, Y. Liu, Q. Zhu, J. Song, and T. Shen, "An adaptive framework embedded with llm for knowledge graph construction," *IEEE Transactions on Multimedia*, vol. 27, pp. 2912–2923, 2025.
- [36] X. Long, L. Zhuang, A. Li, J. Wei, H. Li, and S. Wang, "Kgdm: A diffusion model to capture multiple relation semantics for knowledge graph embedding," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 38, no. 8, 2024, pp. 8850–8858.
- [37] X. Long, L. Zhuang, A. Li, H. Li, and S. Wang, "Fact embedding through diffusion model for knowledge graph completion," in *Proceedings of the ACM Web Conference 2024*, 2024, pp. 2020–2029.
- [38] X. Ge, Y. C. Wang, B. Wang, C.-C. J. Kuo *et al.*, "Knowledge graph embedding: An overview," *APSIPA Transactions on Signal and Information Processing*, vol. 13, no. 1, 2024.
- [39] J. Cao, J. Fang, Z. Meng, and S. Liang, "Knowledge graph embedding: A survey from the perspective of representation spaces," *ACM Computing Surveys*, vol. 56, no. 6, pp. 1–42, 2024.
- [40] T. Trouillon, J. Welbl, S. Riedel, É. Gaussier, and G. Cevaert, "Complex embeddings for simple link prediction," in *Proceedings of the 33rd International Conference on Machine Learning (ICML)*, 2016, pp. 2071–2080.
- [41] T. Song, J. Luo, and L. Huang, "Rot-pro: Modeling transitivity by projection in knowledge graph embedding," in *Proceedings of the Thirty-Fifth Annual Conference on Advances in Neural Information Processing Systems (NeurIPS)*, 2021.
- [42] Z. Zhang, J. Cai, Y. Zhang, and J. Wang, "Learning hierarchy-aware knowledge graph embeddings for link prediction," in *Thirty-Fourth AAAI Conference on Artificial Intelligence*. AAAI Press, 2020, pp. 3065–3072.
- [43] R. Abboud, u. u. Ceylan, T. Lukasiewicz, and T. Salvatori, "Boxe: a box embedding model for knowledge base completion," in *Proceedings of the 34th International Conference on Neural Information Processing Systems*, ser. NIPS '20. Red Hook, NY, USA: Curran Associates Inc., 2020.
- [44] Z. Yao, W. Zhang, M. Chen, Y. Huang, Y. Yang, and H. Chen, "Analogical inference enhanced knowledge graph embedding," in *Proceedings of the Thirty-Seventh AAAI Conference on Artificial Intelligence and Thirty-Fifth Conference on Innovative Applications of Artificial Intelligence and Thirteenth Symposium on Educational Advances in Artificial Intelligence*, ser. AAAI'23/IAAI'23/EAAI'23. AAAI Press, 2023. [Online]. Available: <https://doi.org/10.1609/aaai.v37i4.25605>
- [45] X. Tan, X. Wang, Q. Liu, X. Xu, X. Yuan, and W. Zhang, "Paths-over-graph: Knowledge graph empowered large language model reasoning," in *Proceedings of the ACM on Web Conference 2025*, 2025, pp. 3505–3522.
- [46] L. Wang, W. Zhao, Z. Wei, and J. Liu, "SimKGC: Simple contrastive knowledge graph completion with pre-trained language models," in *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, S. Muresan, P. Nakov, and A. Villavicencio, Eds. Dublin, Ireland: Association for Computational Linguistics, May 2022, pp. 4281–4294. [Online]. Available: <https://aclanthology.org/2022.acl-long.295/>
- [47] X. Long, L. Zhuang, A. Li, J. Wei, H. Li, and S. Wang, "Kgdm: A diffusion model to capture multiple relation semantics for knowledge graph embedding," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 38, no. 8, pp. 8850–8858, Mar. 2024. [Online]. Available: <https://ojs.aaai.org/index.php/AAAI/article/view/28732>
- [48] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," *Advances in neural information processing systems*, vol. 33, pp. 6840–6851, 2020.
- [49] H. Chang, J. Ye, A. Lopez-Avila, J. Du, and J. Li, "Path-based explanation for knowledge graph completion," in *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2024, pp. 231–242.
- [50] X. Wang, K. Zhang, M. Niu, and X. Wang, "Semsi-gat: Semantic similarity-based interaction graph attention network for knowledge graph completion," *IEEE Transactions on Knowledge and Data Engineering*, vol. 37, no. 5, pp. 2958–2970, 2025.
- [51] S. Dutta, A. Gittens, M. J. Zaki, and C. C. Aggarwal, "Replacing paths with connection-biased attention for knowledge graph completion," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 39, no. 14, 2025, pp. 14 885–14 892.
- [52] Y. Song, J. Sohl-Dickstein, D. P. Kingma, A. Kumar, S. Ermon, and B. Poole, "Score-based generative modeling through stochastic differential equations," *arXiv preprint arXiv:2011.13456*, 2020.
- [53] H. Li, Y. Yang, M. Chang, S. Chen, H. Feng, Z. Xu, Q. Li, and Y. Chen, "Srdiff: Single image super-resolution with diffusion probabilistic models," *Neurocomputing*, vol. 479, pp. 47–59, 2022.
- [54] X. Li, J. Thickstun, I. Gulrajani, P. S. Liang, and T. B. Hashimoto, "Diffusion-lm improves controllable text generation," *Advances in Neural Information Processing Systems*, vol. 35, pp. 4328–4343, 2022.
- [55] Y. Tashiro, J. Song, Y. Song, and S. Ermon, "Csd: Conditional score-based diffusion models for probabilistic time series imputation," *Advances in Neural Information Processing Systems*, vol. 34, pp. 24 804–24 816, 2021.
- [56] R. T. Chen, Y. Rubanova, J. Bettencourt, and D. K. Duvenaud, "Neural ordinary differential equations," *Advances in neural information processing systems*, vol. 31, 2018.
- [57] Y. Lipman, R. T. Chen, H. Ben-Hamu, M. Nickel, and M. Le, "Flow matching for generative modeling," in *International Conference on Learning Representations (ICLR)*, 2023.
- [58] M. S. Albergo and E. Vanden-Eijnden, "Building normalizing flows with stochastic interpolants," in *International Conference on Learning Representations (ICLR)*, 2023.
- [59] Q. Liu, "Rectified flow: A marginal preserving approach to optimal transport," *arXiv preprint arXiv:2209.14577*, 2022.
- [60] Z. Geng, M. Deng, X. Bai, J. Z. Kolter, and K. He, "Mean flows for one-step generative modeling," *arXiv preprint arXiv:2505.13447*, 2025.
- [61] R. T. Chen and Y. Lipman, "Flow matching on general geometries," in *International Conference on Learning Representations (ICLR)*, 2024.
- [62] A.-A. Pooladian, H. Ben-Hamu, C. Domingo-Enrich, B. Amos, Y. Lipman, and R. T. Chen, "Multisample flow matching: Straightening flows with minibatch couplings," in *International Conference on Learning Representations (ICLR)*, 2023.
- [63] A. Tong, K. Fatras, N. Malkin, G. Huguet, Y. Zhang, J. Rector-Brooks, G. Wolf, and Y. Bengio, "Improving and generalizing flow-based generative models with minibatch optimal transport," 2023. [Online]. Available: <https://arxiv.org/abs/2302.00482>
- [64] N. Ma, M. Goldstein, M. S. Albergo, N. M. Boffi, E. Vanden-Eijnden, and S. Xie, "Sit: Exploring flow and diffusion-based generative models with scalable interpolant transformers," in *European Conference on Computer Vision*. Springer, 2024, pp. 23–40.
- [65] I. Gat, T. Remez, N. Shaul, F. Kreuk, R. T. Chen, G. Synnaeve, Y. Adi, and Y. Lipman, "Discrete flow matching," *Advances in Neural Information Processing Systems*, vol. 37, pp. 133 345–133 385, 2024.
- [66] W. Hamilton, Z. Ying, and J. Leskovec, "Inductive representation learning on large graphs," *Advances in neural information processing systems*, vol. 30, 2017.
- [67] A. Vaswani, "Attention is all you need," *Advances in Neural Information Processing Systems*, 2017.
- [68] Z. Cao, J. Li, Z. Wang, and J. Li, "Diffusion: Reasoning on knowledge graphs via diffusion-based graph neural networks," in *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2024, pp. 222–230.

- [69] A. Schrijver, *Combinatorial Optimization*. Berlin; New York: Springer, 2003, see p. 362.
- [70] K. Toutanova and D. Chen, "Observed versus latent features for knowledge base and text inference," in *Proceedings of the 3rd Workshop on Continuous Vector Space Models and their Compositionality*. Beijing, China: Association for Computational Linguistics, Jul. 2015, pp. 57–66.
- [71] T. Dettmers, P. Minervini, P. Stenetorp, and S. Riedel, "Convolutional 2d knowledge graph embeddings," in *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, vol. 32, no. 1, 2018.
- [72] W. Xiong, T. Hoang, and W. Y. Wang, "Deeppath: A reinforcement learning method for knowledge graph reasoning," *arXiv preprint arXiv:1707.06690*, 2017.
- [73] S. M. Kazemi and D. Poole, "Simple embedding for link prediction in knowledge graphs," *Advances in neural information processing systems*, vol. 31, 2018.
- [74] S. Zhang, Y. Tay, L. Yao, and Q. Liu, "Quaternion knowledge graph embeddings," *Advances in neural information processing systems*, vol. 32, 2019.
- [75] A. Sadeghian, M. Armandpour, P. Ding, and D. Z. Wang, "Drum: End-to-end differentiable rule mining on knowledge graphs," *Advances in neural information processing systems*, vol. 32, 2019.
- [76] Z. Peng, H. Yu, and X. Jia, "Path-based reasoning with k-nearest neighbor and position embedding for knowledge graph completion," *Journal of Intelligent Information Systems*, pp. 1–21, 2022.
- [77] M. Schlichtkrull, T. N. Kipf, P. Bloem, R. van den Berg, I. Titov, and M. Welling, "Modeling relational data with graph convolutional networks," *The Semantic Web*, vol. 10843, pp. 593–607, 2018.
- [78] J. Shen, C. Wang, L. Gong, and D. Song, "Joint language semantic and structure embedding for knowledge graph completion," *arXiv preprint arXiv:2209.08721*, 2022.
- [79] R. Nadkarni, D. Wadden, I. Beltagy, N. A. Smith, H. Hajishirzi, and T. Hope, "Scientific language models for biomedical knowledge base completion: An empirical study," *arXiv preprint arXiv:2106.09700*, 2021.
- [80] A. K. Nassiri, N. Pernelle, F. Saïs, and G. Quercini, "Knowledge graph refinement based on triplet bert-networks," *arXiv preprint arXiv:2211.10460*, 2022.
- [81] C. Meilicke, M. Fink, Y. Wang, D. Ruffinelli, R. Gemulla, and H. Stuckenschmidt, "Fine-grained evaluation of rule-and embedding-based systems for knowledge graph completion," in *The semantic web-ISWC 2018: 17th international semantic web conference, Monterey, CA, USA, October 8–12, 2018, proceedings, part I* 17. Springer, 2018, pp. 3–20.
- [82] F. Yang, Z. Yang, and W. W. Cohen, "Differentiable learning of logical rules for knowledge base reasoning," *Advances in neural information processing systems*, vol. 30, 2017.
- [83] L. Van der Maaten and G. Hinton, "Visualizing data using t-sne," *Journal of machine learning research*, vol. 9, no. Nov, pp. 2579–2605, 2008.
- [84] G. Hinton, "Kinship," UCI Machine Learning Repository, 1986.
- [85] O. Bodenreider, "The unified medical language system (umls): integrating biomedical terminology," *Nucleic Acids Research*, vol. 32, no. suppl\_1, pp. D267–D270, 2004.
- [86] J. Wang, B. Wang, J. Gao, X. Li, Y. Hu, and B. Yin, "Tdn: Triplet distributor network for knowledge graph completion," *IEEE Transactions on Knowledge and Data Engineering*, vol. 35, no. 12, pp. 13 002–13 014, 2023.
- [87] M. Qu, J. Chen, L.-P. Khonneux, Y. Bengio, and J. Tang, "Rnnlogic: Learning logic rules for reasoning on knowledge graphs," *arXiv preprint arXiv:2010.04029*, 2020.
- [88] Z. Zhu, Z. Zhang, L.-P. Khonneux, and J. Tang, "Neural bellman-ford networks: A general graph neural network framework for link prediction," in *Advances in Neural Information Processing Systems*, M. Ranzato, A. Beygelzimer, Y. Dauphin, P. Liang, and J. W. Vaughan, Eds., vol. 34. Curran Associates, Inc., 2021, pp. 29 476–29 490. [Online]. Available: [https://proceedings.neurips.cc/paper\\_files/paper/2021/file/f6a673f09493afcd8b129a0bcf1cd5bc-Paper.pdf](https://proceedings.neurips.cc/paper_files/paper/2021/file/f6a673f09493afcd8b129a0bcf1cd5bc-Paper.pdf)
- [89] Y. Zhang, Z. Zhou, Q. Yao, X. Chu, and B. Han, "Adaprop: Learning adaptive propagation for graph neural network based knowledge graph reasoning," in *Proceedings of the 29th ACM SIGKDD conference on knowledge discovery and data mining*, 2023, pp. 3446–3457.



**Siyuan Li** is currently a senior undergraduate student in Computer Science and Technology at Dalian University of Technology. He is also currently an intern at the School of Artificial Intelligence, Nanjing University. His research interests include knowledge representation and reasoning, data mining, and human-computer interaction and visualization.



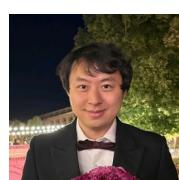
**Ruitong Liu** is currently a senior undergraduate student majoring in Mathematics at Dalian University of Technology. She has been admitted to pursue a postgraduate degree at Peking University. Her research interests include large language models, data-centric AI, and knowledge graphs.



**Yan Wen** is currently a senior undergraduate student in Vehicle Engineering at Beijing Institute of Technology. He is conducting research on autonomous driving, machine learning, and multimodal prediction and planning.



**Te Sun** is a senior undergraduate student of School of Automation Science and Engineering, Dalian University of Technology. He is interested in temporal knowledge graph question answering and robust knowledge graph reasoning. He also has a keen interest in ai agent for science.



**Andi Zhang** is a Machine Learning Research Associate at the Centre for AI Fundamentals, University of Manchester, U.K. He received the B.Sc. degree from the University of Manchester, and the M.Phil., M.A.St., and Ph.D. degrees from the University of Cambridge. His/Her research interests include generative AI, machine learning, and AI fundamentals.



**Yanbiao Ma** received his B.Eng. in Intelligent Science and Technology from Xidian University in 2020 and his Ph.D. in Computer Science and Technology from Xidian University in 2025. He is currently an Assistant Professor at the Gaoling School of Artificial Intelligence, Renmin University of China. His research focuses on multimodal large models, computational neuroscience, long-tailed learning, and federated learning.



**Xiaoshuai Hao** received his Ph.D. from the Institute of Information Engineering, Chinese Academy of Sciences, in 2023. He is currently a researcher at the Beijing Academy of Artificial Intelligence, specializing in embodied multimodal large models. His research interests encompass embodied intelligence, multimodal learning, and autonomous driving. Dr. Hao has published over 40 papers in top-tier journals and conferences, including TIP, Information Fusion, NeurIPS, ICLR, ICML, CVPR, ICCV, ECCV, ACL, AAAI, and ICRA. He has achieved outstanding results in international competitions, securing top-three placements at prestigious conferences like CVPR and ICCV. Additionally, he serves on the editorial board of Data Intelligence and is an organizer for the RoDGE Workshop at ICCV 2025 and The RoboSense Challenge at IROS 2025.