Relational Multi-Path Enhancement for Extrapolative Relation Reasoning in Temporal Knowledge Graph

Linlin Zong
School of Software, Dalian University of Technology
Dalian, China
llzong@dlut.edu.cn

Chi Ma
School of Software, Dalian University of Technology
Dalian, China
201992016@mail.dlut.edu.cn

Jiahui Zhou
School of Software, Dalian University of Technology
Dalian, China
zjhjixiang@mail.dlut.edu.cn

Xinyue Liu
School of Software, Dalian University of Technology
Dalian, China
xyliu@dlut.edu.cn

Wenxin Liang
School of Software, Dalian University of Technology
Dalian, China
wxliang@dlut.edu.cn

Xianchao Zhang
School of Software, Dalian University of Technology
Dalian, China
xczhang@dlut.edu.cn

Bo Xu*
School of Computer Science and Technology, Dalian University of Technology
Dalian, China
xubo@dlut.edu.cn

Abstract

Relation reasoning in temporal knowledge graph infers unknown or emerging relational dependencies from historical structured data. Traditional approaches face inherent limitations in capturing complex semantic correlations and structural patterns among relations. To tackle this problem, we propose the Relational Multi-path Enhancement network (RME), which primarily focuses on relation modeling to enrich relation representations through comprehensive multi-path analysis. RME consists of five key components: (1) Controlled random walk module creates multi-hop head-to-tail paths using an adaptive stopping rule that balances short- and long-term connections. (2) Shared path extraction module identifies both shared-head paths and shared-tail paths. (3) Time-decayed path encoding module processes these paths differently. (4) Gated information aggregation module combines path information to determine which parts matter most. (5) Attention decoding module makes the final prediction by focusing on the most relevant path features. Experiments on multiple TKG benchmark datasets demonstrate that RME outperforms the state-of-the-art methods in relation multi-path reasoning.

CCS Concepts

• **Computing methodologies → Temporal reasoning**.

Keywords

Temporal knowledge graph, relation reasoning, multi-path enhancement

*Corresponding author.

CIKM '25, Seoul, Republic of Korea

ACM ISBN 979-8-4007-2040-6/2025/11
https://doi.org/10.1145/3746252.3761033

ACM Reference Format:
Linlin Zong, Chi Ma, Jiahui Zhou, Xinyue Liu, Wenxin Liang, Xianchao Zhang, and Bo Xu. 2025. Relational Multi-Path Enhancement for Extrapolative Relation Reasoning in Temporal Knowledge Graph. In *Proceedings of the 34th ACM International Conference on Information and Knowledge Management (CIKM '25), November 10–14, 2025, Seoul, Republic of Korea.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3746252.3761033

1 Introduction

Temporal Knowledge Graphs (TKGs) extend traditional knowledge graphs by adding time-aware facts [17]. Formally, a TKG contains timestamped graph snapshots, where each fact is represented as a quadruple (s, r, o, t), the subject entity s , the relation r, the object entity o, and the timestamp t. TKGs track how relationships change over time and show the full lifespan of knowledge (creation, modification, and end) through time-based patterns [28].

Relation reasoning in TKGs predicts the correct relationship for a given subject-object pair (s, o) at time t, capturing how interactions evolve over time. Beyond static entity modeling, this dynamic process better reflects real-world systems. For example, in international diplomacy, relations may shift gradually from `trade_partner` to `military_ally` over a decade, then suddenly deteriorate into `diplomatic_tension` due to geopolitical crises. Accurately predicting such changes enables proactive decision-making, highlighting the importance of relation modeling.

However, current TKG reasoning research faces limitations in capturing these dynamics. Most existing methods are not specifically tailored for relation reasoning; rather, they are designed to handle both entity and relation inference jointly, often treating relational dynamics as a byproduct of entity-centric modeling [10, 13, 16]. As a result, their performance is notably weak on queries that require explicit relational path reasoning. Some recent LLM-based approaches [1, 6] demonstrate efficacy in harnessing textual semantics for relation-only prediction tasks, but their capability to explicitly model relational dynamics remains constrained. Specifically, these methods exhibit limitations in capturing evolutionary properties, which are critical for reasoning over complex relational trajectories.

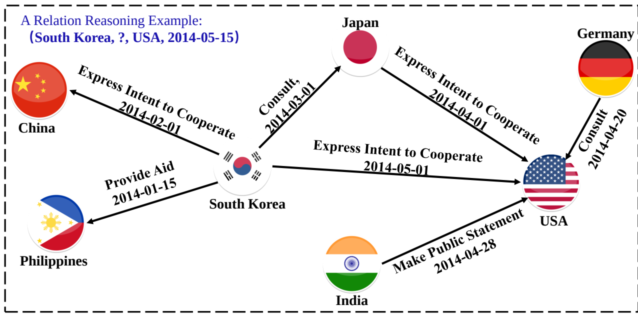


Figure 1: An illustrative example of temporal relational reasoning in TKG extrapolation. The blue content shows the query quadruple (South Korea, ?, USA, 2014-05-15) to be predicted. Supporting evidence comes from (i) closed paths such as South Korea → Japan → USA, (ii) shared-head paths reflecting South Korea's cooperative actions toward other countries (China, Philippines, Japan), and (iii) shared-tail paths showing that USA received friendly intents from other countries (Japan, Germany, India). We can rely on these structured relational paths, rather than solely entity interactions, to infer the missing relation.

These challenges become even more pronounced in TKG extrapolation setting ($t > t_{|T|}$), where models must predict future relations beyond the observed timeline. As shown in Figure 1, the prediction of future relations (e.g., between South Korea and USA on 2014-05-15) cannot rely solely on entity embeddings. Instead, relational paths such as South Korea → Japan → United States, South Korea → China/Philippines/Japan, Germany/India/Japan → USA, provide essential temporal patterns, revealing the necessity of modeling both temporal relation evolution and semantic dependencies between relation types for accurate inference.

To address the challenges of modeling dynamic relational evolution in TKGs, we propose the Relational Multi-path Enhancement network (RME). RME constructs a relational graph with relations as nodes and three relation interaction patterns as typed edges, enabling structured representation of relational dependencies, e.g., attribute cooccurrence, temporal causality, and collaborative effects. Based on the relational graph, RME starts with employing controlled random walks within a sliding temporal window to sample multi-hop head-to-tail relational paths. Then, RME identifies two key patterns: shared-head paths from source entities and shared-tail paths to destination entities, and applies decay weights to recent paths and Temporal Convolutional Networks(TCN) [3] to shared patterns. Next, RME uses gated relation aggregation to combine these features, dynamically adjusting their importance. Finally, RME predicts the candidate relations using attention weights.

In summary, the main contributions of this paper are as follows:

- By treating relations as graph nodes and three relation interaction patterns as typed edges, RME explicitly captures temporal dependencies, enhancing the model's ability to distinguish between attribute co-occurrence, temporal causality, and collaborative effects in temporal reasoning.
- The combination of TCN-based local feature extraction and gated attention for global dependency modeling allows the model to handle both short-term relational trends and long-range path semantics, outperforming static aggregation methods.
- By decoupling entity and Relation reasoning, RME focuses on dynamic modeling of relation evolution, achieving significant improvements on benchmark datasets like ICEWS14 and GDELT for future timestamp prediction tasks, where traditional models struggle with unseen temporal contexts.

2 Related work

This section presents an overview of existing methodologies for entity and relation reasoning in TKGs, focusing on extrapolation.

2.1 Entity Reasoning in TKGs

The approaches are broadly classified into deep learning-driven knowledge representation modeling methods, logic rule-based symbolic reasoning approaches and the hybrid methods.

Early deep learning approaches for entity prediction in TKGs focused on fusing temporal sequence modeling with graph structures. RE-NET [9] pioneered a recurrent neural network (RNN)-based framework, encoding query-specific subgraph sequences to capture short-range temporal dependencies in entity interactions. Building on this, REGCN [16] integrates relational semantics via RGCNs for enhanced entity-relation dependency encoding, yet suffers from fixed-length context constraints limiting long-range pattern capture. DHME [18] proposed a layered framework to decompose TKG evolution into multi-level temporal and relational message-passing stages, capturing hierarchical dependencies in entity dynamics. Despite these improvements, neural models often overlook logical constraints, leading to potential inconsistencies in extrapolation.

Rule-based entity prediction extracts temporal-logical patterns: TLogic [19] employs temporal random walks within an MDP framework to mine cyclic rules with relative time encoding for inductive entity inference, yet suffers from temporal redundancy inflating rule confidence; TR-Rules [14] counters this through temporal-window rule confidence aggregation and redundant body merging—enhancing accuracy while introducing acyclic rule mining, revealing sensitivity to dynamic graph sampling efficiency.

Recent hybrid approaches aim to combine neural and symbolic strengths. CRmod [30] integrated GCNs with six temporal logic rules (e.g., symmetry, transitivity), encoding contextual neighbor relevance while enforcing structural constraints. TiPNN [7] adopted an entity-agnostic approach, constructing history temporal graphs to model query-aware paths via principal neighborhood aggregation, thus enabling inductive reasoning without entity-specific embeddings. ILR-IR [21] fused GRU-encoded temporal paths with one-class matching loss, leveraging both causal path embeddings (symbolic) and entity interaction frequencies (neural) for zero-shot cross-dataset reasoning. TRPG [2] extended this by calculating rule confidence via matrix norms, enabling dynamic entity prediction through high-confidence temporal path rules. Despite these advances, symbolic methods face challenges in balancing rule completeness with computational efficiency and modeling fine-grained temporal nuances (e.g., non-periodic dynamics).

2.2 Relation Reasoning in TKGs

Most existing relation reasoning methods are designed to handle both entity and relation inference jointly. Convolutional architectures have been pivotal in relation prediction, leveraging spatial feature extraction for relational semantics. Conve [5] and ConvTransE [24] applied CNNs to encode entity-relation interactions, capturing local geometric patterns in TKGs. Graph-based models further refined relational reasoning: RGCRN [23] combined GCNs with RNNs to model temporal-relational dynamics, while REGCN [16] introduced relational graph convolutions to propagate information along relation-specific edges, enhancing structural dependency modeling. REGAT [13] augmented this with relational attention mechanisms, adaptively weighting multi-hop relational paths to prioritize contextually relevant interactions. Meanwhile, temporal evolution modeling has emerged as a key direction. EvoKG [22] employed evolutionary graph neural networks to track relation changes over time, capturing dynamic relational trends (e.g., policy shifts in diplomatic TKGs). Tensor factorization models like TuckERTNT [26] extended their utility to relation prediction by decomposing TKGs into latent temporal-relational factors, while CyGNet [29] used cyclic units to model periodic relational patterns (e.g., seasonal business partnerships). CEN [15] addressed variable-length pattern learning via a length-aware CNN and curriculum learning, enabling the model to adapt to evolving relational dynamics at different temporal scales. BH-TDEN[25] enhances relation prediction by integrating adaptive relation updating and time-difference evolutional encoding, further modeling temporal gap uncertainty, enabling probabilistic predictions for complex real-world dynamics.

Recently, emerging methodologies dedicated exclusively to relation prediction have begun leveraging large language models to capture relational semantics and dynamics. znLLM [6] targets zero-shot prediction by aligning T5-generated semantic embeddings with temporal graph structures through GRU networks, directly mapping enriched relation descriptions into latent representations for unseen relation reasoning; whereas RPLLM [1] establishes a classifier by fine-tuning Llama 2 for multi-label sequence classification, operating solely on entity name pairs without description texts or entity embeddings, thereby decoupling relation prediction from any entity-level representation learning and emphasizing semantic-driven efficiency.

However, while recent advancements have begun to explicitly model relational dynamics, many mainstream approaches still underemphasize intrinsic properties of relational evolution such as multi-hop interactive patterns and sequential dependencies. This can result in suboptimal performance for complex temporal scenarios that require explicit, path-based relational reasoning. To better address these challenges, we propose the relational multi-path enhancement network. Our model adaptively constructs and filters multi-hop paths to capture both recent and long-term relational connections, and processes them through time-aware weighting and adaptive fusion to better model evolving relationship structures.

3 Method

3.1 Preliminary

3.1.1 Notation and Task Definition. A Temporal Knowledge Graph (TKG) is formalized as $\mathcal{G} = \{\mathcal{G}_1, \mathcal{G}_2, \ldots, \mathcal{G}_T\}$, where each snapshot $\mathcal{G}_t$ represents a multi-relational directed graph at timestamp $t \in \mathcal{T}$. Here, $\mathcal{E}$, $\mathcal{R}$, $\mathcal{T}$, and $\mathcal{F}$ denote the sets of entities, relations, timestamps, and facts, respectively. A fact is represented as a quadruple (e_s, r, e_o, t), signifying a directed edge from subject entity $e_s \in \mathcal{E}$ to object entity $e_o \in \mathcal{E}$ via relation $r \in \mathcal{R}$ at time $t \in \mathcal{T}$. To enable bidirectional reasoning, each quadruple is augmented with its inverse (e_o, r^{-1}, e_s, t). The extrapolated temporal knowledge graph relation reasoning task focuses on predicting future facts at $t > T$. Given historical facts $\{(e_{s_i}, r_i, e_{o_i}, t_i) \mid t_i < t_q\}$, the goal is to resolve queries of the form $(e_{s_q}, ?, e_{o_q}, t_q)$.

3.1.2 Interaction Patterns between Relations. As depicted in Figure 2, we introduce three fundamental interaction patterns between relations [4].

(1) Head-to-Head (H-H): Relations sharing common head entities (e.g., $(Person, works\_at, Company)$ and $(Person, lives\_in, City)$), indicating attribute cooccurrence relationships originating from the same entity.

(2) Head-to-Tail (H-T): Multi-hop relation sequence forming entity transition chains (e.g., (*Drug*, *treats*, *Disease*) → (*Disease*, $caused\_by$, $Virus$)), capturing temporal causality dependencies.

(3) Tail-to-Tail (T-T): Relations converging to shared tail entities (e.g., $(Artist, creates, Artwork)$ and $(Museum, exhibits, Artwork)$), representing collaborative effects toward common targets.

This structured representation enables our model to differentiate among attribute co-occurrence (H-H), temporal causality (H-T), and collaborative effects (T-T), when propagating relational signals. Compared with conventional approaches, our interaction-focused design achieves more precise and straight representations.

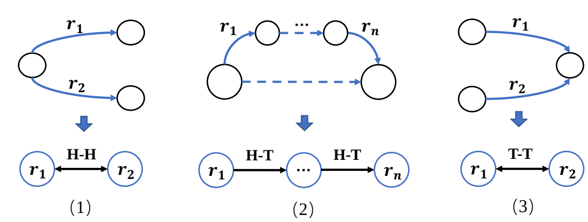


Figure 2: The upper figure shows a common TKG subgraph, where entities are connected by relations. The lower figure presents the corresponding dynamic relation interaction view in terms of interaction patterns: H-H, H-T, T-T.

3.1.3 Relational Graph Definition. Building on the three interaction patterns defined in Section 3.1.2, we formalize a relational graph $\mathcal{RG} = (\mathcal{R}, \mathcal{E}_{\text{link}}, \mathcal{T})$ to explicitly model dynamic relations and their temporal dependencies. In this graph, nodes are relations $r \in \mathcal{R}$, serving as first-class citizens to capture the semantic dynamics of entity interactions (e.g., "works_at", "causes"). Typed Edges are categorized into three types, directly corresponding to the interaction patterns: **H-H Edges** encode co-occurrence of relations sharing the same head entity, **H-T Edges** represent sequential dependencies in multi-hop entity transition chains, **T-T Edges** model convergent interactions toward shared-tail entities.

The graph structure preserves the sparsity of real-world TKGs, with disconnected nodes indicating relations lacking observed

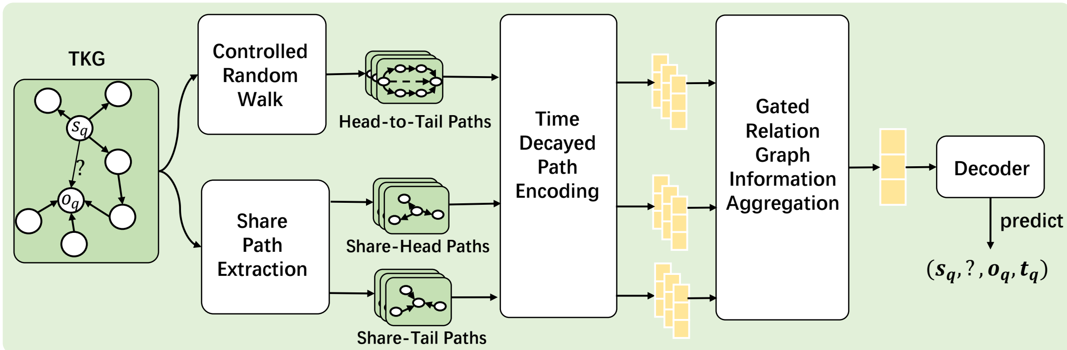


Figure 3: The framework of the RME.

entity-sharing interactions. By treating relations as graph nodes, $\mathcal{RG}$ enables our model to leverage topological features (e.g., multi-hop paths, clustering coefficients) for reasoning, unlike conventional entity-centric models that overlook relational dynamics.

3.2 Overall Architecture

Based on the relational graph defined in section 3.1.3, the RME framework shown in Figure 3 implements dynamic relational reasoning over temporal knowledge graphs through five synergistic modules. First, the **Controlled Random Walk** module generates multi-hop head-to-tail relational paths within a sliding temporal window, employing a dynamic termination mechanism to balance the exploration of short-term interactions and long-term dependencies while automatically constructing positive and negative training samples. Subsequently, the **Shared Path Extraction** module identifies two types of structured patterns from raw paths: shared-head paths (temporally correlated paths originating from the source entity) and shared-tail paths (convergent paths targeting the destination entity), enhancing repetitive temporal patterns through clustering. These paths are then processed by the **Time-Decayed Path Encoding** module, where head-to-tail paths are modeled with exponential decay weights to prioritize temporal relevance, while shared paths undergo temporal convolutions to capture cross-temporal structural features. The **Gated Relation Graph Information Aggregation** module adaptively fuses multi-hop reasoning features with structural regularity features, dynamically adjusting the contribution weights of different paths based on contextual relevance. Finally, the **Attention Decoding** module performs semantic matching between the fused representation and candidate relations, prioritizing discriminative paths through attention mechanisms for precise prediction. By hierarchically integrating temporal decay, structural clustering, dynamic fusion, and context-aware processing, this architecture achieves layered interpretation of complex relational patterns in evolving knowledge graphs.

3.3 Controlled Random Walk

RME employs a controlled random walk strategy to extract head-to-tail paths over TKGs, dynamically generating both positive and negative path samples for training. Given a query $Q = (e_s, ?, e_o, \tau_q)$, the model infers the target relation r_h by exploring multi-hop relational paths connecting e_s and e_o within a temporal window $[\tau_q - \Delta t, \tau_q]$. The algorithm uses a biased random walk process with termination control to balance the generation of positive samples (valid paths ending at e_o) and negative samples (paths terminating prematurely at other entities). Starting from entity e_s, the walk proceeds iteratively. At each step i, with probability ρ, the walk terminates; otherwise, it transitions to a neighboring entity. This termination mechanism serves a dual purpose: (1) Positive Samples: If the walk terminates at e_o, the generated path $p = [r_1, r_2, \ldots, r_m]$ is treated as a positive sample for the target relation r_h. (2) Negative Samples: If the walk terminates at any entity $e_j \neq e_o$, the path is treated as a negative sample, representing a seemingly plausible but incorrect relation path. The transition probability to entity e_i from e_{i-1} is:

$$p(e_i \mid e_{i-1}) = \begin{cases} \frac{1-\rho}{|N(e_{i-1})|}, & \text{if } (e_{i-1}, r, e_i, t) \in \mathcal{G} \text{ and } t \in [\tau_q - \Delta t, \tau_q], \\ 0, & \text{otherwise.} \end{cases} \tag{1}$$

where $|N(e_{i-1})|$ denotes the number of adjacent entities of e_{i-1} within the temporal window.

The parameter $\rho \in [0, 1]$ critically balances the composition of positive (paths reaching e_o) and negative (prematurely terminated paths) samples: smaller values (e.g., $\rho = 0.1$) favor longer walks that increase positive sample generation but risk overfitting in sparse data, while larger values (e.g., $\rho = 0.9$) promote shorter walks with diverse negative samples, enhancing discriminative learning at the cost of reduced positive coverage. In practice, ρ is tuned to reconcile sample diversity with training efficiency. Each sampled head-to-tail path $p = [r_1, r_2, \ldots, r_m]$ corresponds to a sequence of entity transitions $e_s \xrightarrow{r_1} e_1 \xrightarrow{r_2} \ldots \xrightarrow{r_m} e_o$. Here the target relation r_h is defined as the direct edge between e_s and e_o (or "null" if absent) in the TKG.

3.4 Shared Path Extraction

This module extracts two types of temporal relational patterns (1) **Shared-head Paths**: All multi-hop paths starting from e_s with edges observed at timestamp τ_k. (2) **Shared-tail Paths**: All multi-hop paths ending at e_o with edges observed at timestamp τ_k. As depicted in Figure 4, shared-tail paths are characterized by diverse intermediate entities connecting to a shared tail entity (e_o) through multi-hop relations at varying timestamps. Clustering analysis of

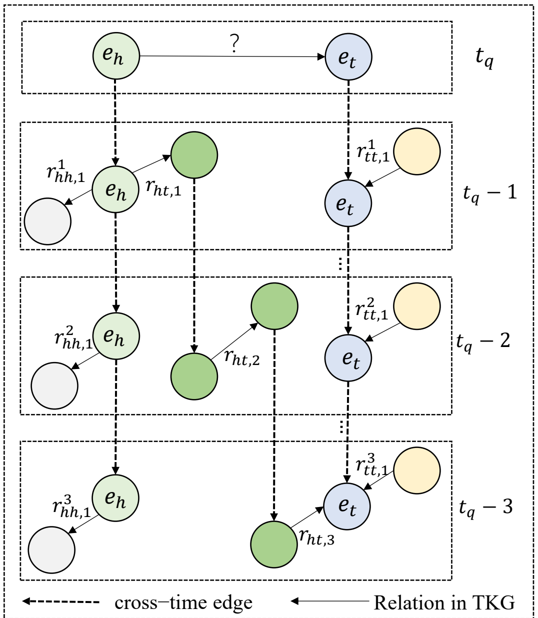


Figure 4: Schematic of multi-temporal head entitiy and tail entity.

these paths reveals associative patterns of the target entity, exemplified by temporal behaviors in social networks where users engage with the same event at different times. Extracting such shared paths enables the model to filter noise effectively and amplify recurrent interaction patterns, thereby supplying structured inputs to downstream encoding modules. Formally, for each timestamp $\tau_k \in T$, we define:

$$P_{hh}(\tau_k) = \{\mathrm{RelEmb}(r) \mid (e_s, r, e_i, \tau_k) \in \mathcal{G}\} \tag{2}$$

$$P_{tt}(\tau_k) = \{\mathrm{RelEmb}(r) \mid (e_i, r, e_o, \tau_k) \in \mathcal{G}\} \tag{3}$$

where $\mathrm{RelEmb}(r) \in \mathbb{R}^d$ is the d-dimensional embedding of relation r; $P_{\mathrm{hh}}(\tau_k)$ and $P_{\mathrm{tt}}(\tau_k)$ denote the sets of relation embeddings for shared-head paths starting from e_s and shared-tail paths ending at e_o at τ_k, respectively.

To model temporal dependency, we introduce a decay coefficient $\lambda \in (0, 1]$, and compute the time difference $\Delta\tau_k = T - \tau_i$ relative to the query time T. The decayed weights are calculated as $\gamma_i = \lambda^{\Delta\tau_k}$. This exponential decay mechanism is motivated by three key considerations: First, facts closer to the query time T are assumed to exert stronger influence on current predictions. Second, historical contributions diminish monotonically with increasing time gaps, avoiding abrupt cutoffs. Third, the decay function $\gamma_i = \lambda^{\Delta\tau_k}$ inherently satisfies the solution to the differential equation $\frac{d\gamma}{d\Delta\tau} = -\kappa\gamma$ ($\kappa = -\ln\lambda$), eliminating recursive temporal computations while explicitly modeling time-sensitive patterns. The parameter λ controls the attenuation rate: When $\lambda \longrightarrow 1$, $\gamma_i \approx 1$ preserves all historical interactions equally, suitable for time-invariant or long-term dependency scenarios. When $\lambda \longrightarrow 0$, γ_i decays rapidly with $\Delta\tau_k$, focusing only on recent patterns for fast-evolving relationships (e.g., social media trend propagation).

Considering the time decay, we aggregate $P_{hh}(\tau_k)$ and $P_{tt}(\tau_k)$ into a vector representation $\mathbf{v}_{hh}(\tau_k)$ and $\mathbf{v}_{tt}(\tau_k)$ using a pooling operation (e.g., mean or attention mechanism):

$$\mathbf{v}_{hh}(\tau_k) = \gamma_i \cdot \mathrm{Pool}\left(P_{hh}\left(\tau_k\right)\right) \tag{4}$$

$$\mathbf{v}_{tt}(\tau_k) = \gamma_i \cdot \mathrm{Pool}\left(P_{tt}\left(\tau_k\right)\right) \tag{5}$$

These per-timestamp representations v_{hh} and v_{tt} are then chronologically ordered to construct temporal sequences. Specifically, for all timestamps $\{\tau_1, \tau_2, ..., \tau_l\}$ within the query window $[\tau_q - \Delta t, \tau_q]$, we concatenate the decayed vectors to form structured inputs for temporal pattern learning:

$$V_{hh} = [\mathbf{v}_{hh}(\tau_1), \mathbf{v}_{hh}(\tau_2), \ldots, \mathbf{v}_{hh}(\tau_l)] \tag{6}$$

$$V_{tt} = [\mathbf{v}_{tt}(\tau_1), \mathbf{v}_{tt}(\tau_2), \ldots, \mathbf{v}_{tt}(\tau_l)] \tag{7}$$

This hierarchical encoding framework achieves two critical objectives: first, the decay coefficient $\gamma_i = \lambda^{T-\tau_k}$ explicitly prioritizes recent interactions while retaining historical context. Second, by arranging path representations into ordered sequences (V_{hh} and V_{tt}), the model captures evolving relational dynamics through subsequent temporal convolutions, enabling learning across time.

3.5 Time-Decayed Path Encoding

To encode temporal dependencies within head-to-tail paths manifesting as a relation node sequence $[r_1, r_2, \ldots, r_k]$ with length k, we extend the above-mentioned decay mechanism as follows.

First, each node undergoes time-decay encoding. For the i-th relation r_i with timestamp τ_i, we compute its decay coefficient $\gamma_i = \lambda^{T-\tau_i}$, and the decayed node representations $\mathbf{p}_{ht,i}$.

$$\mathbf{p}_{ht,i} = \gamma_i \cdot \mathrm{RelEmb}(r_i) \tag{8}$$

Subsequently, $\mathbf{p}_{ht,i}$ are fed into a TCN for encoding:

$$\mathbf{h}_{ht}^{\mathrm{TCN}} = \mathrm{TCN}([\mathbf{p}_{ht,1}, \mathbf{p}_{ht,2}, \ldots, \mathbf{p}_{ht,k}]) \tag{9}$$

Finally, the self-attention mechanism is applied to $\mathbf{h}_{ht}^{\mathrm{TCN}}$. Queries, keys, and values are defined as:

$$\mathbf{Q} = \mathbf{h}_{ht}^{\mathrm{TCN}}\mathbf{W}_q, \quad \mathbf{K} = \mathbf{h}_{ht}^{\mathrm{TCN}}\mathbf{W}_k, \quad \mathbf{V} = \mathbf{h}_{ht}^{\mathrm{TCN}}\mathbf{W}_v \tag{10}$$

where $\mathbf{W}_q, \mathbf{W}_k, \mathbf{W}_v \in \mathbb{R}^{d\times d}$ represent parameter matrices. Upon this, we compute the attention scores as:

$$\alpha_{i,j} = \frac{\mathbf{Q}_i^\top \mathbf{K}_j}{\sqrt{d}} + \mathrm{Mask}(i, j) \tag{11}$$

$$\mathbf{a}_{ht} = \mathrm{Softmax}(\alpha)\mathbf{V} \tag{12}$$

where $\mathrm{Mask}(i, j)$ serves to mask invalid relation nodes (such as nodes outside the path), and the head-to-tail path representation $\mathbf{h}_{ht}$ is derived by taking the average of the attention outputs:

$$\mathbf{h}_{ht} = \frac{1}{k}\sum_{i=1}^{k} \mathbf{a}_{ht}[i] \tag{13}$$

For shared-head and shared-tail paths, we independently process the temporal sequences V_{hh} and V_{tt} via TCNs:

$$\mathbf{h}_h = \mathrm{TCN}(V_{hh}) = \mathrm{TCN}([\mathbf{v}_{hh}(\tau_1), \mathbf{v}_{hh}(\tau_2), \ldots, \mathbf{v}_{hh}(\tau_l)]) \quad (14)$$

$$\mathbf{h}_t = \mathrm{TCN}(V_{tt}) = \mathrm{TCN}([\mathbf{v}_{tt}(\tau_1), \mathbf{v}_{tt}(\tau_2), \ldots, \mathbf{v}_{tt}(\tau_l)]) \quad (15)$$

3.6 Gated Relation Graph Information Aggregation

To address the heterogeneity of relational path types of temporal knowledge graphs, the gated relation graph information aggregation module is designed to adaptively fuse multi-source path representations, including head-to-tail paths, shared-head paths, and shared-tail paths. This module dynamically weighs the importance of different path types, enabling the model to suppress irrelevant contextual noise and prioritize discriminative relational features for reasoning.

First, we concatenate the head-to-tail path representation $\mathbf{h}_{ht}$, the shared-head relation path representation $\mathbf{h}_h$, and the same-tail relation path representation$\mathbf{h}_t$ along the last dimension. Then, apply a linear transformation $\mathbf{W}_g$ and a bias term $\mathbf{b}_g$, followed by a sigmoid function, to obtain the gating vector $\mathbf{g}$:

$$\mathbf{g} = \sigma\left(\mathbf{W}_g\left[\mathbf{h}_{ht} \oplus \mathbf{h}_h \oplus \mathbf{h}_t\right] + \mathbf{b}_g\right) \quad (16)$$

where σ is the sigmoid function, $\oplus$ denotes the concatenation operation, $\mathbf{W}_g$ is the linear transformation matrix, and $\mathbf{b}_g$ is the bias term. The gating vector $\mathbf{g}$ acts as a regulator, determining the contribution of each component ($\mathbf{h}_{ht}, \mathbf{h}_h, \mathbf{h}_t$) to the final representation. For example, when $\mathbf{g}$ is close to 1, the head-to-tail path representation $\mathbf{h}_{ht}$ dominates, indicating higher importance in current context.

Finally, fused representation $\mathbf{h}_{\text{fused}}$ is obtained through a weighted combination controlled by the gating vector $\mathbf{g}$, which integrates the head-to-tail path representation with the sum of shared-head and sharde-tail relation path representations:

$$\mathbf{h}_{\text{fused}} = \mathbf{g} \odot \mathbf{h}_{ht} + (1 - \mathbf{g}) \odot (\mathbf{h}_h + \mathbf{h}_t) \quad (17)$$

where $\odot$ denotes element-wise multiplication.This mechanism allows the model to adaptively fuse different types of path information, enhancing the richness of the relational graph representation.

3.7 Attention Decoding

The attention decoding mechanism serves as the final inference layer, translating the fused relational representation $\mathbf{h}_{\text{fused}}$ into a probability distribution over candidate relations. This module leverages a query-key-value attention framework to measure the semantic compatibility between the input context and each possible relation, mimicking human-like reasoning that prioritizes contextually relevant relationships.

Construct a relation embedding matrix $\mathbf{R} \in \mathbb{R}^{|\mathcal{R}| \times d}$, where each row $\mathbf{R}_r$ represents the learned embedding of relation r. Using $\mathbf{h}_{\text{fused}}$ as the query vector, to mitigate dimensionality effects, the model computes a score for each candidate relation r via a dot-product similarity metric scaled by $\frac{1}{\sqrt{d}}$:

$$\mathrm{Score}(\mathbf{h}_{\text{fused}}, r) = \frac{\mathbf{h}_{\text{fused}}^{\top}\mathbf{R}_r}{\sqrt{d}} \quad (18)$$

This score reflects the degree of semantic alignment between the input path-based context and the relational semantics encoded in $\mathbf{R}_r$. The scores are then normalized via a softmax function to generate a probability distribution over all relations:

$$p(r|Q) = \mathrm{Softmax}(\{\mathrm{Score}(\mathbf{h}_{\text{fused}}, r)\}_{r \in \mathcal{R}}) \quad (19)$$

This distribution allows the model to predict the most likely target relation for the query $Q = (e_s, ?, e_o, \tau_q)$ by selecting the relation with the highest probability. The attention-based design not only enhances the model's interpretability but also enables efficient handling of large relation spaces by focusing on contextually relevant candidates.

3.8 Loss Function

In the temporal knowledge graph relation reasoning task, the model must learn temporal dependencies and relational semantics from historical facts to predict future relations. As the task essentially involves multi-class classification with discrete relation labels, the Cross-Entropy Loss is widely adopted. It measures the probability divergence between the model's predictions and true labels, guiding the model to maximize the likelihood of correct relations.

Suppose the dataset $\mathcal{D}$ contains N samples, where each sample is a query $Q_i = (h_i, ?, t_i, T_i)$ with the corresponding ground-truth relation r_i^*. To enhance the model's ability to distinguish correct relations from incorrect ones, we utilize the negative samples generated during the controlled random walk. These negative samples help the model learn to reject implausible relations and focus on discriminative features. The model's objective is to minimize the cross-entropy loss over both positive and negative samples. The loss function $\mathcal{L}$ is defined as:

$$\mathcal{L} = -\frac{1}{N}\sum_{i=1}^{N}[\log p(r_i^*|Q_i) + \frac{1}{|K|}\sum_{r \in \mathcal{R}_{neg}} log(1 - p(r|Q_i))] \quad (20)$$

where K denotes the number of negative samples corresponding to one positive sample, and $\mathcal{R}_{neg}^i$ represents the set of negative relations generated for query Q_i. This formulation ensures the model is penalized for both incorrect positive predictions and high confidence in negative relations.

The model is trained by minimizing the above loss function $\mathcal{L}$ to improve its performance on the temporal knowledge graph relation reasoning task. For inference, RME retrieves edge-type subgraphs corresponding to query relations and uses temporal attention to compute similarity with candidate relations, enabling accurate prediction of future interactions.

4 Experiment

4.1 Experimental Setup

4.1.1 Datasets. We evaluate on standard TKG reasoning benchmarks: ICEWS-14 [8], YAGO [20], GDELT [12], and WIKI [11]. These datasets are widely adopted for their rich temporal annotations and diverse relations. For GDELT, we use the subset from [27] covering events from April 1, 2015 to March 31, 2016. Dataset statistics are shown in Table 1.

4.1.2 Baselines. To comprehensively evaluate the performance of RME on the relation reasoning task, we compare it against a diverse

Table 1: Statistics of the datasets.

Datasets	Entities	Relations	Train	Valid	Test	Time gap	Snapshot numbers
ICEWS14	7,128	230	74,845	8,514	7,371	24 hours	365
GDELT	500	20	2,733,685	341,961	341,961	1 year	366
WIKI	12,554	24	539,286	67,538	63,110	1 year	232
YAGO	10,623	10	161,540	19,523	20,026	24 hours	189

set of state-of-the-art TKG reasoning models, providing a strong foundation for comparison.

- ConvE[5] and ConvTransE[24] leverage convolutional neural networks to encode relational patterns in static KGs, treating timestamps as static attributes or ignoring temporal dynamics.
- RGCRN[23] and REGCN[16] combine GNNs with recurrence or static convolutions to propagate relational features across multi-hop paths.
- CygNet[29] and EvokG[22] focus on temporal evolution by integrating state transition mechanisms or event-based dynamics, capturing how relations change over time.
- RE-GAT[13] introduces attention mechanisms to weigh relational importance in GNN-based aggregation, enhancing interpretability in dynamic KG inference.
- CEN[15] explicitly models temporal dependencies via hierarchical attention, emphasizing the order and recency of events in knowledge graph reasoning.
- TuckeRTNT[26] employs tensor decomposition (Tucker factorization) to model multi-relational interactions, extending static embedding frameworks to temporal contexts.

By comparing against this diverse set of state-of-the-art models, we rigorously demonstrate RME's superior capability in capturing complex relational dynamics for temporal reasoning.

4.1.3 Implementation and Hyperparameters. Our model is implemented using PyTorch, with baseline parameters strictly aligned with the configurations reported in the original literature. For all datasets, in the random walk process, specific hyperparameters are configured as follows: the history window length Δt is set to 3, the path termination probability ρ is fixed at 0.05, and the maximum number of walk steps per query quadruple is capped at 10,000. During training, the decay factor λ is set to 0.9, and the batch size is configured as 5,000. The embedding dimension d is uniformly set to 256. Hyperparameter values are determined through validation-set optimization, specifically by maximizing the Mean Reciprocal Rank (MRR) on each validation split. Model parameters are initialized via the Xavier initialization scheme and optimized using the Adam optimizer with a fixed learning rate of 0.005.

4.1.4 Evaluation Metrics. In the Relation reasoning task, the model scores and ranks the candidate relation set of the relations to be inferred in the test set, and selects the relation with the highest rank as the prediction result. The evaluation metrics used in the experiment include Mean Reciprocal Ranking (MRR) and hit rates at positions 1/3/10 (Hits@1/3/10) for the target relation in the ranking results.Among them, the MRR metric measures the model's ranking performance by calculating the reciprocal mean of the correct relation ranks in all queries of the test set. The Hits@N metric represents the proportion of queries in the test set where the correct relation appears in the top N positions of the ranking results. These two metrics evaluate the model's ranking ability and retrieval accuracy in the Relation reasoning task from different dimensions.

4.2 Results on Relation Reasoning

As the Table2 and table3 show, the proposed RME model outperforms the comparative methods on most evaluation metrics. In key metrics such as MRR, Hits@1, Hits@3, and Hits@10, RME achieves optimal performance across multiple dimensions with its unique model design, particularly demonstrating significant advantages in MRR and Hits@1 and, indicating that its more accurate semantic modeling of relational features and effectiveness in knowledge graph reasoning. Among the comparative models, EvoKG, CyGNet, etc., show competitiveness on some metrics, but RME has more advantages in overall performance by integrating multiple feature nteractions and structural modeling, verifying the rationality of the model design. In summary, RME provides a more generalized solution for temporal knowledge graph relation reasoning tasks. Its performance advantages offer effective references for subsequent research on complex knowledge representation learning.

4.3 Ablation

4.3.1 Gated Relation Graph Information Aggregation Ablation. To evaluate the effectiveness of adaptive path fusion in RME, we conducted ablation experiments on the YAGO dataset for the gated relation graph information aggregation module, which fuses head-to-tail (H-T), shared-head (H-H), and shared-tail (T-T) path representations. Table 4 compares four strategies: *main* uses only the primary H-T path representation, ignoring H-H/T-T contextual paths. *mean* computes a simple element-wise average of H-T, H-H, and T-T paths, assuming equal importance. *concat* concatenates path features into a single vector followed by a linear projection, treating all paths as independent. *gate* (RME's method) uses a sigmoid-gated mechanism to control information flow, prioritizing H-T paths while selectively integrating H-H/T-T signals

The *gate* mechanism outperforms other methods, achieving the highest MRR (0.952) and Hits@3 (0.961). This confirms that adaptive gating effectively suppresses noisy auxiliary paths (e.g., irrelevant H-H relations) while preserving complementary signals from H-H/T-T patterns (e.g., recurring entity-centric interactions). In contrast, *mean* and *concat* either underutilize or overfit to contextual paths, highlighting the importance of dynamic weighting.

4.3.2 Time-Decayed Path Encoding Ablation. For the time-decayed path encoding module, we compared six methods for aggregating

Table 2: Results of Relation reasoning on the YAGO and ICEWS14 datasets under the raw setting. The best results are in bold.

Dataset	YAGO				ICEWS14			
Model	MRR	Hits@1	Hits@3	Hits@10	MRR	Hits@1	Hits@3	Hits@10
ConvE	0.913	0.856	0.951	0.981	0.305	0.158	0.305	0.664
ConvTransE	0.910	0.845	0.962	0.983	0.426	0.256	**0.538**	0.699
RGCRN	0.902	0.842	0.936	0.972	0.382	0.229	0.417	0.607
RE-GCN	0.941	0.887	0.983	0.995	0.376	0.246	0.408	0.584
CyGNet	0.905	0.848	0.962	0.996	0.363	0.236	0.403	0.610
EvoKG	0.934	0.911	**0.984**	**0.999**	0.425	0.278	0.412	0.710
RE-GAT	0.933	0.905	0.969	0.989	0.411	0.293	0.434	0.584
CEN	0.922	0.843	0.971	0.998	0.363	0.225	0.412	0.661
TuckerTNT	0.945	0.912	0.949	0.998	0.410	0.267	0.411	0.683
RME	**0.952**	**0.944**	0.961	**0.999**	**0.431**	**0.301**	0.449	**0.723**

Table 3: Results of Relation reasoning on the WIKI and GDELT datasets under the raw setting. The best results are in bold.

Dataset	WIKI				GDELT			
Model	MRR	Hits@1	Hits@3	Hits@10	MRR	Hits@1	Hits@3	Hits@10
ConvE	0.782	0.709	0.857	0.956	0.199	0.107	**0.253**	0.42
ConvTransE	0.866	0.827	0.931	0.964	0.190	0.081	0.196	0.438
RGCRN	0.899	0.836	0.943	0.961	0.187	0.068	0.180	0.401
RE-GCN	0.980	0.969	0.994	0.998	0.190	0.079	0.194	0.433
CyGNet	0.972	0.962	0.994	**0.999**	0.199	0.081	0.195	0.425
EvoKG	0.975	0.971	0.994	**0.999**	0.209	0.108	0.231	0.432
RE-GAT	0.979	0.965	0.991	**0.999**	0.192	0.087	0.196	0.425
CEN	0.981	0.967	0.994	0.993	0.191	0.080	0.193	0.427
TuckerTNT	0.981	0.974	0.995	0.998	0.213	0.112	0.229	0.419
RME	**0.983**	**0.981**	**0.995**	0.998	**0.223**	**0.144**	0.236	**0.591**

Table 4: Ablation Study of Different Path Feature fusion Strategies for Gated Relation Graph Aggregation Module on YAGO.

Method	MRR	Hits@1	Hits@3
main	0.927	0.916	0.928
mean	0.928	0.915	0.916
concat	0.946	**0.951**	0.959
gate (Ours)	**0.952**	0.944	**0.961**

path features on YAGO (Table 5): *mean* computes the element-wise average of all valid path representations, treating each step equally regardless of its position in the path. *concat* concatenates path features into a single vector followed by a linear projection, treating all paths as independent. *self-attn* uses self-attention to capture intra-path dependencies. *w-mean* applies pre-trained fixed weights to sum relation embeddings, assuming static importance across path steps. *attn* employs a learnable attention mechanism to dynamically weight each path representation based on its relevance to the query, allowing the model to focus on informative steps. *len-weight* (RME's method) dynamically weights each relation by the reciprocal of the path length, prioritizing shorter paths as they often carry more discriminative semantics.

Table 5: Ablation study of Different Fusion Strategies for Aggregating H-T Path Features for Time-Decayed Path Encoding Module on YAGO.

Method	MRR	Hits@1	Hits@3
mean	0.940	0.939	0.945
concat	0.950	0.943	**0.951**
self-attn	0.946	0.939	0.948
w-mean	0.931	0.915	0.947
attn	0.951	**0.945**	0.950
len-weight (Ours)	**0.952**	0.944	**0.951**

len-weight achieves the best MRR (0.952) and Hits@3 (0.951), likely because shorter paths (e.g., $e_s \xrightarrow{r_1} e_o$) are more reliable indicators of direct relations compared to noisy long paths. By downweighting longer sequences, the model avoids overfitting to spurious multi-hop patterns, enhancing precision in relational reasoning. In contrast, *self-attn* and *w-mean* treat all path lengths equally, leading to less focused representations.

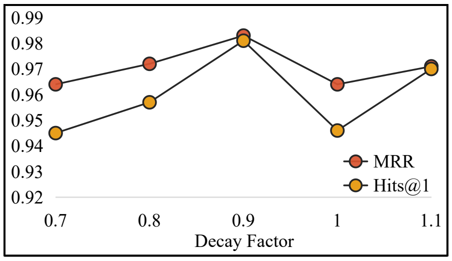
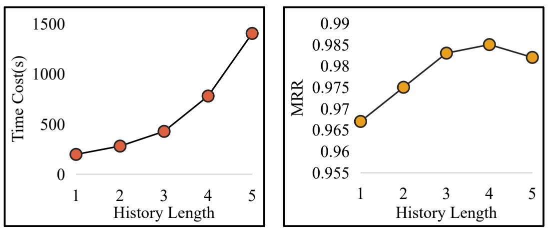


Figure 5: Impact of the decay factor on the MRR and Hits@1.



(a) (b)

Figure 6: Effect of history length on MRR and time cost in the RME framework on WIKI dataset.

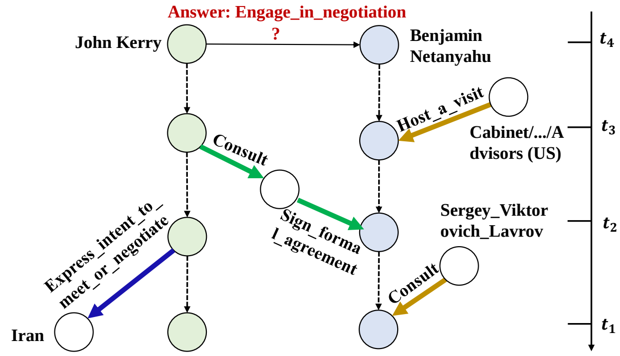


Figure 7: A Temporal Knowledge Graph (TKG) subgraph depicting diplomatic interactions between entities (*JohnKerry*, *BenjaminNetanyahu*, *Iran*, etc.) via relations (*Consult*, *Host_a_visit*, etc.), structured to model head-to-tail (H-T) paths (e.g., US diplomatic signaling → Israeli context) and predict the target relation '*Engage_in_negotiation*', illustrating RME's framework for temporal-relational reasoning over dynamic entity-relation networks.

4.4 Parameter Study

4.4.1 Decay Factor Research. In Figure 5, we experiment on WIKI to investigat the impact of Decay Factor on MRR and Hits@1. Results show performance improves as the decay factor increases from 0.7, peaking at 0.9. This may be because a smaller decay factor causes the model to over-decay older information, leading to the model failing to fully utilize the valid knowledge in this information. As the decay factor increases, the model can more reasonably balance the weights of information from different times, allowing historical information to better assist current reasoning. Around 0.9, the model's decay processing of temporal information achieves a relatively optimal balance, effectively utilizing the strong relevance of recent information while reasonably integrating key knowledge from distant information. However, performance drops when the decay factor reaches 1, likely because temporal patterns are ignored. A slight recovery at 1.1 suggests this setting better adapts to the data's temporal characteristics.

4.4.2 History Length Research. Figure 6 (b) shows that, as the history length increases, the MRR improves, but the gains diminish significantly for lengths 4 and 5 compared to length 3. Meanwhile, as shown in Figure 6 (a), the time cost rises sharply with longer history lengths. To achieve an optimal balance between performance and efficiency, we set the default history length to $L = 3$.

4.5 Case Study

Figure 7 shows the relation reasoning process of a diplomatic TKG query $Q = (\text{John Kerry}, ?, \text{Benjamin Netanyahu}, t_4)$ (John Kerry, U.S. former Secretary of State as head; Benjamin Netanyahu, Israeli former Prime Minister as tail, from ICEWS14). The graph uses blue (shared-head, John→Iran, U.S. intent signals), green (head-tail, John's diplomatic actions like Consult/Sign), and red (shared-tail, Benjamin's ties with US advisors/Russian Lavrov) to encode multi-faceted diplomatic paths. RME employs Controlled Random Walk (path exploration), time-decayed shared path extraction (prioritizing recent t3/t4 interactions), TCN/attention encoding, gated aggregation (fusing U.S./Israeli dynamics), and attention decoding. By integrating these paths, RME predicts *Engage_in_negotiation* at t_4, demonstrating its capability to decode complex diplomatic interdependencies in TKGs—critical for international relations forecasting.

5 Conclusion

We propose the Relational Multi-Path Enhancement Network (RME), which breaks through the traditional TKG reasoning paradigm by taking relations as the core modeling units. For the relation reasoning extrapolation task, the model introduces novel encoding modules for head-to-tail paths and shared-head/shared-tail paths to mine semantic and structural dependencies between relations and combines information aggregation and attention mechanisms to achieve accurate prediction. Experimental results show that RME significantly outperforms existing methods on multiple benchmark datasets, especially in sparse and long-term temporal scenarios, and can capture implicit relational patterns that traditional methods find difficult to detect.

Acknowledgments

This work was supported in part by National Natural Science Foundation of China (No. 62476040), the Liaoning Provincial Natural Science Foundation Joint Fund Program (2023-MSBA-003) and the Fundamental Research Funds for the Central Universities (DUT24MS003).

General AI Usage Statement

In the preparation of this manuscript, general artificial intelligence (GenAI) tools were used for language editing assistance, including grammar correction, wording improvement, and ensuring compliance with academic writing conventions. All content related to model architecture, experimental design, and results analysis was authored by the team, with GenAI tools serving solely as productivity aids. The authors confirm that GenAI tools were used ethically and in compliance with academic integrity standards. All contributions involving GenAI were limited to supportive roles, and the research outcomes and intellectual content of this manuscript remain the exclusive work of the authorship team.

References

[1] Sakher Khalil Alqaaidi and Krzysztof J. Kochut. 2024. Relations Prediction for Knowledge Graph Completion using Large Language Models. *CoRR* abs/2405.02738 (2024). arXiv:2405.02738

[2] Luyi Bai, Mingzhuo Chen, Lin Zhu, and Xiangxi Meng. 2023. Multi-hop temporal knowledge graph reasoning with temporal path rules guidance. *Expert Syst. Appl.* 223 (2023), 119804.

[3] Shaojie Bai, J. Zico Kolter, and Vladlen Koltun. 2018. An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling. *CoRR* abs/1803.01271 (2018). arXiv:1803.01271

[4] Jiajun Chen, Huarui He, Feng Wu, and Jie Wang. 2021. Topology-Aware Correlations Between Relations for Inductive Link Prediction in Knowledge Graphs. In *Thirty-Fifth AAAI Conference on Artificial Intelligence, AAAI 2021, Thirty-Third Conference on Innovative Applications of Artificial Intelligence, IAAI 2021, The Eleventh Symposium on Educational Advances in Artificial Intelligence, EAAI 2021, Virtual Event, February 2-9, 2021.* 6271–6278.

[5] Tim Dettmers, Pasquale Minervini, Pontus Stenetorp, and Sebastian Riedel. 2018. Convolutional 2D Knowledge Graph Embeddings. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, (AAAI-18), the 30th innovative Applications of Artificial Intelligence (IAAI-18), and the 8th AAAI Symposium on Educational Advances in Artificial Intelligence (EAAI-18), New Orleans, Louisiana, USA, February 2-7, 2018.* 1811–1818.

[6] Zifeng Ding, Heling Cai, Jingpei Wu, Yunpu Ma, Ruotong Liao, Bo Xiong, and Volker Tresp. 2024. zrLLM: Zero-Shot Relational Learning on Temporal Knowledge Graphs with Large Language Models. In *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers), NAACL 2024, Mexico City, Mexico, June 16-21, 2024*, Kevin Duh, Helena Gómez-Adorno, and Steven Bethard (Eds.). Association for Computational Linguistics, 1877–1895.

[7] Hao Dong, Pengyang Wang, Meng Xiao, Zhiyuan Ning, Pengfei Wang, and Yuanchun Zhou. 2024. Temporal inductive path neural network for temporal knowledge graph reasoning. *Artif. Intell.* 329 (2024), 104085.

[8] Alberto García-Durán, Sebastijan Dumancic, and Mathias Niepert. 2018. Learning Sequence Encoders for Temporal Knowledge Graph Completion. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Brussels, Belgium, October 31 - November 4, 2018*, Ellen Riloff, David Chiang, Julia Hockenmaier, and Jun'ichi Tsujii (Eds.). Association for Computational Linguistics, 4816–4821.

[9] Woojeong Jin, Meng Qu, Xisen Jin, and Xiang Ren. 2020. Recurrent Event Network: Autoregressive Structure Inferenceover Temporal Knowledge Graphs. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, EMNLP 2020, Online, November 16-20, 2020*, Bonnie Webber, Trevor Cohn, Yulan He, and Yang Liu (Eds.). Association for Computational Linguistics, 6669–6683.

[10] Timothée Lacroix, Guillaume Obozinski, and Nicolas Usunier. 2020. Tensor Decompositions for Temporal Knowledge Base Completion. In *International Conference on Learning Representations*.

[11] Julien Leblay and Melisachew Wudage Chekol. 2018. Deriving Validity Time in Knowledge Graph. In *Companion of the The Web Conference 2018 on The Web Conference 2018, WWW 2018, Lyon , France, April 23-27, 2018*, Pierre-Antoine Champin, Fabien Gandon, Mounia Lalmas, and Panagiotis G. Ipeirotis (Eds.). ACM, 1771–1776.

[12] Kalev Leetaru and Philip A Schrodt. 2013. Gdelt: Global data on events, location, and tone, 1979–2012. In *ISA annual convention*, Vol. 2. Citeseer, 1–49.

[13] Linjie Li, Zhe Gan, Yu Cheng, and Jingjing Liu. 2019. Relation-Aware Graph Attention Network for Visual Question Answering. In *2019 IEEE/CVF International Conference on Computer Vision, ICCV 2019, Seoul, Korea (South), October 27 - November 2, 2019.* 10312–10321.

[14] Ningyuan Li, Haihong E, Shi Li, Mingzhi Sun, Tianyu Yao, Meina Song, Yong Wang, and Haoran Luo. 2023. TR-Rules: Rule-based Model for Link Forecasting on Temporal Knowledge Graph Considering Temporal Redundancy. In *Findings of the Association for Computational Linguistics: EMNLP 2023, Singapore, December 6-10, 2023*, Houda Bouamor, Juan Pino, and Kalika Bali (Eds.). Association for Computational Linguistics, 7885–7894.

[15] Zixuan Li, Saiping Guan, Xiaolong Jin, Weihua Peng, Yajuan Lyu, Yong Zhu, Long Bai, Wei Li, Jiafeng Guo, and Xueqi Cheng. 2022. Complex Evolutional Pattern Learning for Temporal Knowledge Graph Reasoning. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), ACL 2022, Dublin, Ireland, May 22-27, 2022*, Smaranda Muresan, Preslav Nakov, and Aline Villavicencio (Eds.). Association for Computational Linguistics, 290–296.

[16] Zixuan Li, Xiaolong Jin, Wei Li, Saiping Guan, Jiafeng Guo, Huawei Shen, Yuanzhuo Wang, and Xueqi Cheng. 2021. Temporal Knowledge Graph Reasoning Based on Evolutional Representation Learning. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval* (Virtual Event, Canada) *(SIGIR '21)*. Association for Computing Machinery, New York, NY, USA, 408–417.

[17] Ke Liang, Lingyuan Meng, Meng Liu, Yue Liu, Wenxuan Tu, Siwei Wang, Sihang Zhou, Xinwang Liu, Fuchun Sun, and Kunlun He. 2024. A Survey of Knowledge Graph Reasoning on Graph Types: Static, Dynamic, and Multi-Modal. *IEEE Trans. Pattern Anal. Mach. Intell.* 46, 12 (2024), 9456–9478.

[18] Xinyue Liu, Jianan Zhang, Chi Ma, Wenxin Liang, Bo Xu, and Linlin Zong. 2024. Temporal Knowledge Graph Reasoning with Dynamic Hypergraph Embedding. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024).* 15742–15751.

[19] Yushan Liu, Yunpu Ma, Marcel Hildebrandt, Mitchell Joblin, and Volker Tresp. 2022. TLogic: Temporal Logical Rules for Explainable Link Forecasting on Temporal Knowledge Graphs. In *Thirty-Sixth AAAI Conference on Artificial Intelligence, AAAI 2022, Thirty-Fourth Conference on Innovative Applications of Artificial Intelligence, IAAI 2022, The Twelveth Symposium on Educational Advances in Artificial Intelligence, EAAI 2022 Virtual Event, February 22 - March 1, 2022.* AAAI Press, 4120–4127.

[20] Farzaneh Mahdisoltani, Joanna Biega, and Fabian M Suchanek. 2013. Yago3: A knowledge base from multilingual wikipedias. In *CIDR*.

[21] Xin Mei, Libin Yang, Zuowei Jiang, Xiaoyan Cai, Dehong Gao, Junwei Han, and Shirui Pan. 2024. An Inductive Reasoning Model based on Interpretable Logical Rules over temporal knowledge graph. *Neural Networks* 174 (2024), 106219.

[22] Namyong Park, Fuchen Liu, Purvanshi Mehta, Dana Cristofor, Christos Faloutsos, and Yuxiao Dong. 2022. Evokg: Jointly modeling event time and network structure for reasoning over temporal knowledge graphs. In *Proceedings of the fifteenth ACM international conference on web search and data mining*. 794–803.

[23] Youngjoo Seo, Michaël Defferrard, Pierre Vandergheynst, and Xavier Bresson. 2018. Structured Sequence Modeling with Graph Convolutional Recurrent Networks. In *Neural Information Processing - 25th International Conference, ICONIP 2018, Siem Reap, Cambodia, December 13-16, 2018, Proceedings, Part I (Lecture Notes in Computer Science, Vol. 11301)*, Long Cheng, Andrew Chi-Sing Leung, and Seiichi Ozawa (Eds.). Springer, 362–373.

[24] Chao Shang, Yun Tang, Jing Huang, Jinbo Bi, Xiaodong He, and Bowen Zhou. 2019. End-to-end structure-aware convolutional networks for knowledge base completion. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 33. 3060–3067.

[25] Pengpeng Shao, Yang Wen, and Jianhua Tao. 2024. Bayesian hypernetwork collaborates with time-difference evolutional network for temporal knowledge prediction. *Neural Networks* 175 (2024), 106146.

[26] Pengpeng Shao, Dawei Zhang, Guohua Yang, Jianhua Tao, Feihu Che, and Tong Liu. 2022. Tucker decomposition-based temporal knowledge graph completion. *Knowledge-Based Systems* 238 (2022), 107841.

[27] Rakshit Trivedi, Hanjun Dai, Yichen Wang, and Le Song. 2017. Know-Evolve: Deep Temporal Reasoning for Dynamic Knowledge Graphs. In *Proceedings of the 34th International Conference on Machine Learning, ICML 2017, Sydney, NSW, Australia, 6-11 August 2017 (Proceedings of Machine Learning Research, Vol. 70)*, Doina Precup and Yee Whye Teh (Eds.). PMLR, 3462–3471.

[28] Mengqi Zhang, Yuwei Xia, Qiang Liu, Shu Wu, and Liang Wang. 2023. Learning Long- and Short-term Representations for Temporal Knowledge Graph Reasoning. In *Proceedings of the ACM Web Conference 2023, WWW 2023, Austin, TX, USA, 30 April 2023 - 4 May 2023*, Ying Ding, Jie Tang, Juan F. Sequeda, Lora Aroyo, Carlos Castillo, and Geert-Jan Houben (Eds.). ACM, 2412–2422.

[29] Cunchao Zhu, Muhao Chen, Changjun Fan, Guangquan Cheng, and Yan Zhang. 2021. Learning from history: Modeling temporal knowledge graphs with sequential copy-generation networks. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 35. 4732–4740.

[30] Lin Zhu, Die Chai, and Luyi Bai. 2024. CRmod: Context-Aware Rule-Guided reasoning over temporal knowledge graph. *Inf. Sci.* 664 (2024), 120343.