Dual History Enhancement with Hybrid Hypergraph-Graph Networks for Temporal Knowledge Graph Reasoning

Kailun Ye
Zhejiang University of Technology
College of Computer Science and Technology
Hangzhou, China
Zhejiang Key Laboratory of Visual Information Intelligent Processing
Hangzhou, China
yekailun@zjut.edu.cn

Xiangjie Kong*
Zhejiang University of Technology
College of Computer Science and Technology
Hangzhou, China
Zhejiang Key Laboratory of Visual Information Intelligent Processing
Hangzhou, China
xjkong@ieee.org

Yuchao Zhang
Zhejiang University of Technology
College of Computer Science and Technology
Hangzhou, China
Zhejiang Key Laboratory of Visual Information Intelligent Processing
Hangzhou, China
yuchao.z@outlook.com

Xuan Wang
University of Cincinnati
College of Education, Criminal Justice and Human Services
Cincinnati, OH, USA
wang6xu@mail.uc.edu

Linan Zhu
Zhejiang University of Technology
College of Computer Science and Technology
Hangzhou, China
Zhejiang Key Laboratory of Visual Information Intelligent Processing
Hangzhou, China
zln@zjut.edu.cn

Jiaxin Du
Zhejiang University of Technology
College of Computer Science and Technology
Hangzhou, China
Zhejiang Key Laboratory of Visual Information Intelligent Processing
Hangzhou, China
dujiaxin@zjut.edu.cn

Guojiang Shen
Zhejiang University of Technology
College of Computer Science and Technology
Hangzhou, China
Zhejiang Key Laboratory of Visual Information Intelligent Processing
Hangzhou, China
gjshen1975@zjut.edu.cn

Jianxin Li
Edith Cowan University
School of Business and Law
Joondalup, WA, Australia
jianxin.li@ecu.edu.au

Abstract

Temporal Knowledge Graph (TKG) reasoning seeks to predict future events by analyzing historical data, where the effective leverage of both local and global historical facts proves crucial. Existing approaches employ graph neural networks (GNNs) and recurrent neural networks (RNNs) for local evolution patterns, complemented by statistical methods to enhance attention to global facts, demonstrating efficient predictive capabilities. However, traditional GNNs, constrained by their low-order neighborhood aggregation design, inherently fail to model potential high-order dependencies among facts. Furthermore, existing global history modeling approaches may introduce irrelevant historical information that interferes with prediction tasks. To address these limitations, we propose a Dual History-aware HyperGraph Network for TKG reasoning, namely DHHGN. Specifically, for local history modeling, we design a hybrid hypergraph-graph joint recurrent convolution module that simultaneously captures low-order neighborhood information and high-order interaction patterns among entities, employing a gating mechanism to adaptively blend their contributions. For global history modeling, we propose a dual history enhancement module that amplifies attention on pivotal historical facts while ensuring holistic integration of all historical contexts. Extensive experiments on four public benchmarks validate that DualHist-HGN consistently outperforms existing state-of-the-art methods across TKG reasoning tasks.

CCS Concepts

• **Computing methodologies** → **Knowledge representation and reasoning**; **Temporal reasoning**.

Keywords

Temporal Knowledge Graph Reasoning, History-aware, Hypergraph

ACM Reference Format:
Kailun Ye, Xiangjie Kong, Yuchao Zhang, Xuan Wang, Linan Zhu, Jiaxin Du, Guojiang Shen, and Jianxin Li. 2026. Dual History Enhancement with Hybrid Hypergraph-Graph Networks for Temporal Knowledge Graph Reasoning. In *Proceedings of the ACM Web Conference 2026 (WWW '26), April 13–17, 2026, Dubai, United Arab Emirates.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3774904.3792310

1 Introduction

Temporal knowledge graphs (TKGs), by binding events to specific timestamps, provide a structured, multi-relational representation to model the temporal dynamics of complex scenarios, demonstrating

*Corresponding authors.

WWW '26, Dubai, United Arab Emirates.

ACM ISBN 979-8-4007-2307-0/2026/04
https://doi.org/10.1145/3774904.3792310

significant value in domains such as recommendation systems [9, 27], social politics [12], and epidemic propagation [8]. A TKG can be conceptualized as a chronological sequence of static knowledge graphs, where each graph encapsulates all co-occurring events at a specific timestamp. Each fact in TKG can be formally represented as a quadruple (subject, relation/predicate, object, timestamp). For instance, (Putin, Visit, North Korea, 2024-06-18) represents Putin's diplomatic visit to North Korea on June 18, 2024. However, TKGs are inevitably incomplete due to practical constraints in collecting real-world data. Events may not be fully recorded due to sparse sensor coverage, delayed reporting, or intentional omission. To address this critical gap, Temporal Knowledge Graph Reasoning (TKGR), also known as Temporal Knowledge Graph Completion(TKGC), emerges as a pivotal task. The core objective of TKGR is to infer missing facts at observed timestamps (interpolation) and predict future events at unseen timestamps (extrapolation) based on historical data. While interpolation addresses historical gaps, extrapolation empowers proactive decision-making in critical domains. Given its higher practical value, this work prioritizes overcoming the challenges associated with extrapolation.

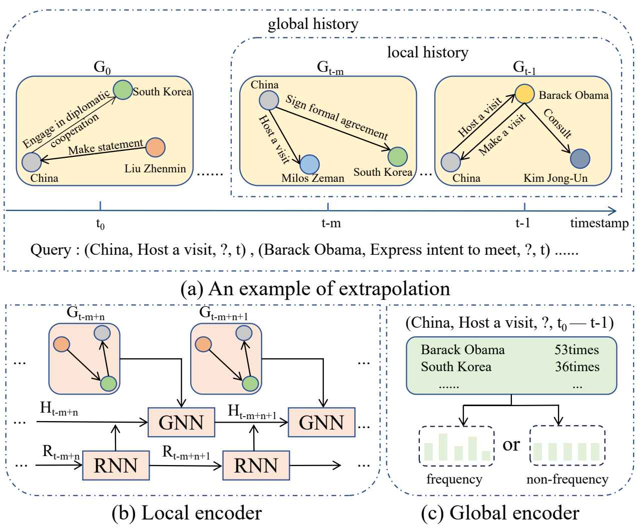

Figure 1: An example of extrapolation, along with the local history and global history modeling methods used in existing models.

As illustrated in Figure 1(a), accurate extrapolation in TKGs hinges critically on the nuanced mining of historical events, which requires a synergistic analysis of local historical dynamics and global contexts to disentangle causally relevant temporal dependencies. Research advances have explored both perspectives.

For local history modeling, as depicted in Figure 1(b), mainstream approaches employ graph neural networks (GNNs) to capture the structural dependencies among concurrent facts while employing recurrent neural networks (RNNs) to model the evolutionary trends of entities. A representative model, RE-GCN [24], effectively combines spatial and temporal dependencies to learn dynamic representations from local historical sequences. Subsequent extensions, such as CEN [21], further enhance this paradigm by incorporating multi-channel temporal modeling. However, the intrinsic nature of real-world information often involves complex, high-order dependencies among multiple entities, such as group collaborations, indirect influences, and long-range transmission. These high-order relationships are discretized into simple binary relationships within knowledge graphs. The fundamental limitation of GNNs lies precisely in their reliance on these simplified representations, with the consequence that information propagation and aggregation are confined to direct neighbors or within a limited number of hops.

For global history modeling, as shown in Figure 1(c), current methods are generally based on historical event statistics. CyGNet [42] constructs candidate prediction sets via frequency statistics, while TiRGN [20] employs binary event existence statistics to mitigate frequency bias. Although empirical studies confirm that such methods enhance model sensitivity to global historical patterns, they exhibit notable limitations: being prone to frequency-dominated prediction bias or assuming uniform contributions from all historical events while neglecting their significance to current predictions. In fact, the volume of global historical data is vast, particularly in expansive datasets like GDELT. Indiscriminate incorporation without strategic prioritization introduces substantial redundant information, adversely affecting prediction efficacy. Thus, the critical challenge lies in efficiently harnessing global historical dependencies, balancing comprehensive data integration with a focus on the most relevant events to optimize performance.

To this end, we propose DHHGN, a Dual History-aware HyperGraph Network. For local history modeling, we devise a hybrid module that combines traditional graphs with hypergraphs to concurrently capture both low-order neighborhood information and high-order interaction patterns among entities. For global history, we introduce a dual history enhancement module that enables holistic modeling of all relevant historical facts while adaptively enhancing attention to facts that have a high impact on the current prediction step.

The principal contributions of this work are summarized as follows:

- We propose DHHGN, which enhances traditional graph neural networks by integrating hypergraphs to capture both low-order and latent high-order relationships.
- We propose a dual history enhancement module that enables the model to consider all historical facts while enhancing its focus on critical facts. This dual strategy effectively addresses issues related to frequency bias and historical noise interference present in current approaches, while also preventing the omission of historical information due to attention sparsity.
- Extensive experiments across four public benchmarks demonstrate that DHHGN consistently outperforms state-of-the-art baselines in entity prediction tasks under TKG extrapolation scenarios.

2 Related Work

The following sections provide a comprehensive overview of emerging methods and models in knowledge graph reasoning from two perspectives: static knowledge graph reasoning and temporal knowledge graph reasoning.

2.1 Static KG Reasoning

Early approaches predominantly adopt vector translation and tensor decomposition for static knowledge graph reasoning. TransE [2] regards relations as translational vectors that approximate the distance between head and tail entities in the embedding space. RotatE [32] extends this by interpreting the relations as rotational transformations between entities. RESCAL [29] decomposes tensors into low-dimensional representations to capture entity-relation interactions. Subsequent studies introduce convolutional neural networks (CNNs) to enhance feature extraction: ConvE [14] concatenates the embeddings of entities and relations into matrix forms, utilizing convolutional layers and nonlinear activations to learn complex interaction patterns. ConvKB [28] employs 1D convolutional kernels to reduce redundant structural information inherent in 2D convolutions. ConvTransE [31] integrates TransE's translational principle with convolutional operations for enriched feature learning. Recent advances leverage GNNs to exploit graph-structured data. RGCN [30] assigns different weight matrices to relations for multi-relational neighborhood aggregation. CompGCN [34] jointly learns entity and relation embeddings through basis vector decomposition, addressing the issue of over-parameterization in traditional GNNs. GAT [35] incorporates attention mechanisms to adjust neighbor aggregation weights adaptively.

Although extensive empirical studies have validated the efficacy of static knowledge graph reasoning models in processing static knowledge, their limitations become evident when confronted with the dynamic and temporally evolving nature of real-world knowledge. Factual information in practical scenarios is intrinsically time-sensitive, characterized by continuous updates, temporal dependencies, and contextual shifts—features that static knowledge graph frameworks fail to encapsulate.

2.2 Temporal KG Reasoning

Temporal knowledge graph reasoning tasks can be categorized into interpolation and extrapolation based on the prediction timestamp [40]. This work focuses on extrapolation. Current research seeks to derive predictive value from local and global historical facts. CyGNet [42] proposes a copy-generation mechanism to capture the frequency of global historical facts. RE-GCN [24] employs a local history encoder, combining graph construction based on RGCN with GRU [13] to capture short-term temporal dynamics. TiRGN [20] introduces a binary indicator to record the occurrence of global events, mitigating frequency-induced noise. In addition, CluSTeR [23] utilizes reinforcement learning to trace historical paths specific to the query. CENET [37] adopts contrastive learning to differentiate historical versus non-historical events. L^2TKG [39] addresses data sparsity issues by constructing latent relational graphs. DHE-TKG [26] employs hypergraph neural networks to model heterogeneous temporal features and improve dynamic relational reasoning. CSI [6] applies causal theory to disentangle causal and non-causal historical dependencies. LogCL [10] employs contrastive learning that is aware of local-global history to better distinguish relevant historical patterns from noise. HTCCN [7] employs temporal causal convolution with Hawkes processes to model the evolution of facts. CognTKE [11] builds a TCR-Digraph to preserve query-relevant temporal paths and applies a dual-system reasoner for interpretable extrapolation. ANEL [41] proposes an adaptive neighborhood enhancement layer that can be seamlessly integrated into any model to mitigate sparsity-related challenges within the datasets. Additionally, recent advances have seen the integration of emerging technologies such as Meta-Learning [1, 15, 43], Diffusion Models [4, 5] and Large Language Models (LLMs) [25, 36, 38]. Despite these advancements, existing methods exhibit two critical limitations: 1) None simultaneously models both low-order and latent high-order relationships within a single snapshot; 2) Existing models fail to effectively utilize global historical events that are of decisive importance for current predictions while ensuring the integrity of historical information.

3 Preliminaries

A TKG can be represented as a sequence of static KGs ordered chronologically: $\mathcal{G} = \{\mathcal{G}_1, \mathcal{G}_2, ..., \mathcal{G}_T\}$, where each $\mathcal{G}_t = (\mathcal{E}, \mathcal{R}, \mathcal{F}_t)$ includes all events occurring at timestamp t. Here, $\mathcal{E}$ represents the entity set, $\mathcal{R}$ the relation set, and $\mathcal{F}_t$ the collection of factual quadruples (s, r, o, t) where $s, o \in \mathcal{E}$ and $r \in \mathcal{R}$. Each quadruple (s, r, o, t) indicates that the subject entity s and the object entity o interacted through the relation r at time t. For each original quadruple (s, r, o, t), we generate a corresponding inverse quadruple (o, r^{-1}, s, t) and include it in $\mathcal{F}_t$. Entity embeddings $\mathbf{H} \in \mathbb{R}^{|\mathcal{E}| \times d}$ and relation embeddings $\mathbf{R} \in \mathbb{R}^{|\mathcal{R}| \times d}$ map entities and relations to d-dimensional vectors, where each row $\mathbf{h} \in \mathbf{H}$ represents an entity, and $\mathbf{r} \in \mathbf{R}$ encodes a relation. A TKG extrapolation task requires predicting missing entities in queries of the form $(s, r, ?, t)$ by modeling conditional probability distributions $p(o|r, s, t, \mathcal{G}_{0:t-1})$.

4 Methodology

The general architecture of DHHGN is shown in Figure 2. Our model comprises two core modules: **Hybrid Hypergraph-Graph(HHG) Module** for modeling local historical events, which addresses the limitations of GNNs by introducing hypergraphs to compensate for the lack of high-order information detection, allowing simultaneous capture of both low-order neighborhood interactions and high-order group dependencies. **Dual History Enhancement(DHE) Module** considers all global historical data while adaptively highlighting significant events for predictions.

4.1 Hybrid Hypergraph-Graph Module

This module is designed to capture both low-order information and high-order patterns from local historical events. For the current timestamp t, we construct a sequential graph representation $\mathcal{G} = \{\mathcal{G}_{t-n}, \mathcal{G}_{t-n+1}, ..., \mathcal{G}_{t-1}\}$ using the preceding n timestamps. Guided by the cognitive principle [16] that human understanding progresses from holistic perception to intrinsic analysis, our framework decomposes hierarchical structural learning into two progressive stages: aggregation of neighborhood information and exploration of high-order patterns. In the first phase, we perform neighborhood aggregation through a w-layer RGCN. The aggregation process is formally defined as:

$$\mathbf{h}_{o,t}^{l+1} = f\left(\frac{1}{c_o} \sum_{(s,r,o) \in \mathcal{F}_t} \mathbf{W}_1^l (\mathbf{h}_{s,t}^l + \mathbf{r}_t) + \mathbf{W}_2^l \mathbf{h}_{o,t}^l\right), \quad (1)$$

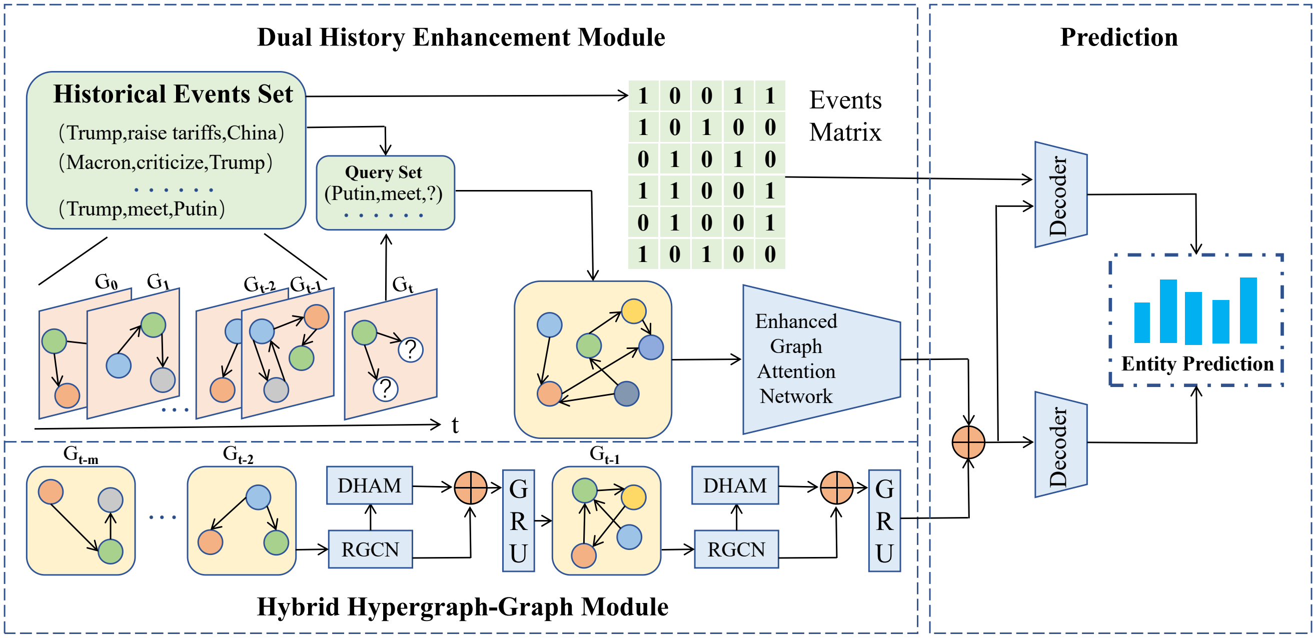

Figure 2: The DHHGN framework comprises three components: a dual history enhancement module for extracting global dependencies, a hybrid hypergraph-graph module modeling recent event sequences, and a prediction layer for outcome projection.

where $\mathbf{h}^l_{s,t}$ and $\mathbf{h}^l_{o,t}$ denote the embedding representations of the subject entity s and the object entity o at l_{th} layer of timestamp t, respectively. $\mathbf{W}^l_1$ and $\mathbf{W}^l_2$ are learnable weights, $\mathbf{r}_t$ represents the relation embedding, and f(·) denotes the RReLU activation function. The normalization term c_o corresponds to the in-degree of entity o. Notably, we implement self-updating operations for isolated entity nodes that lack neighborhood connections at specific timestamps to maintain the stability of information propagation. After obtaining the entity embedding matrix $\mathbf{H}_g \in \mathbb{R}^{|\mathcal{E}|\times d}$, we employ a Dual-phase Hypergraph Aggregation Module(DHAM) to capture high-order information patterns among entities. As shown in Figure 3, DHAM is a two-phase aggregation module based on an entity-hyperedge probability mapping matrix, which dynamically captures higher-order dependencies and optimizes representations through a process of mapping from entities to hyperedges and back to entities. Unlike traditional hypergraph construction methods, we posit that there exists some degree of connection between all entities, and thus each entity should belong to every hyperedge with a certain probability. Consequently, rather than relying on predefined semantic labels, we propose a learnable entity-hyperedge mapping matrix $\mathbf{D} \in \mathbb{R}^{|\mathcal{E}|\times N_z}$ that dynamically captures these latent relationships in a data-driven manner, where N_z represents the number of hyperedges. This learnable matrix dynamically adjusts the probability of assigning entities to different hyperedges during training, providing greater flexibility in adapting to complex entity interaction patterns, particularly in historical event sequences where relationships between entities often evolve over time. The matrix parameters are first normalized via softmax to obtain $\tilde{D}$, ensuring the assignment probabilities for each entity sum to one across different hyperedges:

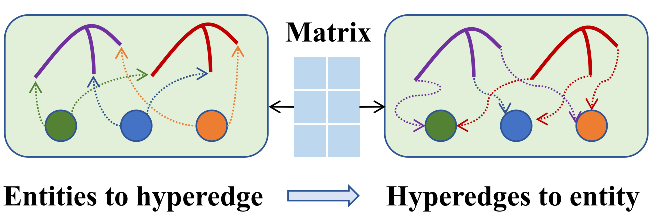

Figure 3: The dual-phase process of DHAM: aggregation from entities to hyperedges and back to entities, with aggregation weights provided by a learnable entity-hyperedge probability mapping matrix.

$$\tilde{D}_{i,j} = \frac{\exp(D_{i,j})}{\sum_{k=1}^{|\mathcal{E}|}\exp(D_{i,k})}. \tag{2}$$

Similarly, we derive matrix $\tilde{C}$ to ensure the weights assigned to entities within each hyperedge sum to one:

$$\tilde{C}_{i,j} = \frac{\exp(D_{i,j})}{\sum_{k=1}^{N_z}\exp(D_{k,j})}. \tag{3}$$

The entity-hyperedge aggregation operation is first performed to derive hyperedge embeddings, as formalized in Equation (4):

$$\mathbf{z}_i = \sigma\left(\sum_{k=1}^{|\mathcal{E}|} I_k \tilde{D}_{i,k}\mathbf{W}_3\mathbf{h}_k\right), \quad I_k = \begin{cases}1, & if \quad k \in S_t\\ 0, & otherwise\end{cases} \tag{4}$$

where σ denotes Gelu activation function, $\mathbf{h}_k \in \mathbb{R}^d$ represents the embedding of the k-th entity, $\mathbf{W}_3$ is a learnable weight matrix for feature refinement, and $\tilde{D}_{i,k}$ quantifies the probabilistic assignment of the k-th entity to hyperedge i, I_k is a binary indicator function, with a value of 1 indicating the k-th entity appears in the current subgraph entity set S_t.

The hyperedge-entity aggregation operation is subsequently executed to propagate the refined hyperedge features back to entity.

$$\mathbf{h}_i = \sigma(\sum_{k=1}^{k=N_z} \tilde{C}_{k,i}\mathbf{W}_4\mathbf{z}_k). \tag{5}$$

This operation redistributes hyperedge-enhanced features back to nodes, with each node's updated representation computed as a combination of all hyperedges it participates in.

Then, a gated fusion mechanism is employed to integrate the entity representation matrix $\mathbf{H}^g$ with its hypergraph-refined counterpart $\mathbf{H}^{hg}$.

$$\mathbf{H}_{fus} = \gamma \odot \mathbf{H}_g + (1-\gamma) \odot \mathbf{H}_{hg}, \tag{6}$$

$$\gamma = \sigma(\mathbf{W}_5[\mathbf{H}_g||\mathbf{H}_{hg}] + \mathbf{b}_g), \tag{7}$$

where $\sigma(\cdot)$ denotes the sigmoid activation function.

To mitigate the potential loss of historical structural dependencies across sequential processing steps, a Gated Recurrent Unit (GRU) is strategically incorporated to establish temporal memory propagation:

$$\mathbf{H}_t = GRU(\mathbf{H}_{t-1}, \mathbf{H}_{fus}). \tag{8}$$

4.2 Dual History Enhancement Module

As the core component of DHHGN, the Dual History Enhancement (DHE) module consists of two complementary sub-modules: Comprehensive History Modeling (CHM) for integrating a complete historical context, and Relevance-aware History Screening (RHS) for focusing on events that are crucial for prediction. The integration of these two sub-modules ensures both the breadth and depth of the captured global historical dependency information.

4.2.1 Relevance-aware History Screening. This sub-module constructs query relevance subgraphs and employs an enhanced graph attention network to highlight the historically critical events that significantly influence the current predictions. For each query $Q = (s, r, ?, t)$, we first construct a global historical evidence set by retrieving all historical events sharing the same subject entity s and relation r as the query:

$$C_Q = \left\{(s', r', o', \tau) \in \mathcal{G}_{0:t-1} | s' = s \wedge r' = r\right\}. \tag{9}$$

The union of these query-specific evidence sets across all concurrent predictions forms the query relevance subgraph:

$$\mathcal{G}_t^{his} = \bigcup_{Q_i \in \mathcal{Q}_t} C_{Q_i}, \tag{10}$$

where $\mathcal{Q}_t$ denotes the set of all queries at timestamp t.

To discern the importance of historical events specific to each query, we design an enhanced graph attention network that dynamically assigns attention scores to directed edges in the query relevance subgraph $\mathcal{G}_t^{his}$. For each directed edge $(s, r, o, \tau) \in \mathcal{G}_t^{his}$, we compute its attention score by jointly encoding the subject entity s, relation r, and object entity o through convolutional fusion, the attention mechanism operates as follows:

$$\alpha_{s,o}^l = \sigma(\mathbf{W}_6 Conv1D(\mathbf{s}^l||\mathbf{r}^l||\mathbf{o}^l)), \tag{11}$$

where σ represents the LeakyReLU activation function, and Conv1D refers to 1D convolutional operations applied to concatenated features for local pattern fusion. After generating messages for each edge, we perform aggregation and update operations:

$$\mathbf{o}^{l+1} = \sum_{(s,r,o)\in f_o} \alpha_{s,o}^l \mathbf{W}_7(\mathbf{s}^l + \mathbf{r}^l), \tag{12}$$

where f_o represents the set of historical events where entity o serves as the tail entity.

4.2.2 Comprehensive History Modeling. To prevent the RHS module from assigning excessively low attention weights to certain entities and thereby causing the loss of global history information, the CHM sub-module implements a global history encoding mechanism that utilizes binary markers to record all related historical interactions. This complementary approach preserves RHS's focused reasoning capability while maintaining holistic historical records, effectively mitigating the risk of overlooking historical events due to attention sparsity. Inspired by TiRGN [20], CHM constructs a sparse tensor representation $\mathbf{M} \in \{0,1\}^{(|\mathcal{E}|\times|\mathcal{R}|)\times|\mathcal{E}|}$ to encode full historical context. For each prediction timestamp t, we systematically record all (s, r, o, τ) tuples from historical subgraph $\mathcal{G}_{0:t-1}$ through binary encoding:

$$M_t^{s,r}[o] = \begin{cases} 1 & \text{if } (s, r, o, \tau) \in \bigcup_{\tau=0}^{t-1} \mathcal{F}_\tau \\ 0 & \text{otherwise} \end{cases}, \tag{13}$$

where $\mathcal{F}_\tau$ denotes the set of historical facts that occur at τ.

4.3 Prediction

In our framework, the ConvTransE [31] model is employed as the decoder component. Before this, we implement another gated fusion (as same in Equations 6 and 7) to integrate the entity embeddings $\mathbf{H}_{hhg}$ derived from the HHG module with the $\mathbf{H}_{rhs}$ obtained from the RHS module. This fusion process yields the final entity representations denoted as $\mathbf{H}_f$ and then fed into two parallel decoders, where one decoder incorporates the binary mask $\mathbf{M}$ to capture global historical information dependencies:

$$p_r(o|s,r) = softmax(\mathbf{H}_f Decoder(\mathbf{s},\mathbf{r}) \odot \mathbf{M}_t^{\mathbf{s},\mathbf{r}}), \tag{14}$$

while the other preserves the base prediction capability:

$$p_h(o|s,r) = softmax(\mathbf{H}_f Decoder(\mathbf{s},\mathbf{r})). \tag{15}$$

Specifically, a weighting hyperparameter α is introduced to regulate the contribution intensity of global historical dependencies, thereby obtaining the ultimate prediction score through weighted aggregation:

$$p(o|s,r) = \alpha p_h(o|s,r,\mathbf{H}_f) + (1-\alpha)p_r(o|s,r,\mathbf{H}_f). \tag{16}$$

5 Experiments

In this section, we conduct comprehensive experiments on four publicly available datasets to validate the predictive performance of the DHHGN model. Our experimental evaluation covers multiple dimensions, including overall performance comparisons, component-wise ablation studies and parameter sensitivity tests.

5.1 Experimental Settings

5.1.1 Datasets. We evaluated DHHGN on four widely used temporal KG benchmarks, including ICEWS14 [17], ICEWS18 [3], ICEWS05-15 [17], and GDELT [19]. Following standard practice, all datasets are divided chronologically into training, validation, and test sets with a 80% / 10% / 10% ratio. Additional details regarding the datasets can be found in Table 1.

5.1.2 Metrics. We employ Mean Reciprocal Rank (MRR) and Hits@K as evaluation metrics to assess model performance. MRR calculates the average reciprocal rank of the first correct answer across all test queries, while Hits@K measures the percentage of queries where the correct entity appears within the top-K ranked predictions. In line with established conventions, we implement the filtered evaluation protocol to compute these metrics.

Table 1: The statistics of the datasets.

Dataset	ICEWS14	ICEWS18	ICEWS05-15	GDELT
Entities	7128	23033	10488	7691
Relations	230	256	251	240
Train-facts	63685	373018	368868	1734399
Valid-facts	13823	45995	46302	238765
Test-facts	13222	49545	46159	305241
Timestamps	365	304	4017	2976
Time gap	1 day	1 day	1 day	15 mins

5.1.3 Baselines. We select several representative static knowledge graph reasoning models and temporal knowledge graph reasoning models to compare with DHHGN. For static knowledge graphs, the baselines include Conv-TransE [31], RotatE [32], ConvE [14], and ComplEx [33]. For temporal knowledge graphs, we compared against state-of-the-art extrapolation models including RE-NET [18], CyGNet [42], REGCN [24], TiRGN [20], HisMatch [22], CENET [37], L2TKG [39], LogCL [10], HTCCN [7], DHE-TKG [26], TiRGN-ANEL [41], and CognTKE [11].

5.1.4 Implementation Details. The embedding dimension d for entities and relations is set to 200. All graph convolutional layers are configured with two layers, while hypergraph aggregation utilizes a single layer. The Adam optimizer is employed with a learning rate of 0.001. The dataset-specific configurations for ICEWS14, ICEWS18, ICEWS05-15, and GDELT include local history window lengths of 9, 10, 15, and 7, respectively, and global historical event retention rates of 0.3, 0.2, 0.4, and 0.5, corresponding to the datasets in the order listed. A static semantic graph constraint is incorporated into the model. Dropout rates are set to 0.3 for hypergraph convolution and 0.2 elsewhere. All experiments are conducted on an RTX 4090. Baseline results are directly adopted from previous work.

5.2 Results

As shown in Table 2, DHHGN demonstrates substantial improvements over existing state-of-the-art models across all metrics on all datasets. Specifically, static knowledge graph models exhibit significantly inferior performance compared to DHHGN due to their inherent inability to process temporal relationships. Our model outperforms RE-NET, CyGNet, RE-GCN, and TiRGN; while RE-NET and RE-GCN focus on mining local historical dependencies, they neglect the critical influence of global historical patterns. CyGNet considers global historical events but fails to model evolutionary trends reflected in local history, as its frequency-based statistical approach suffers from interference caused by high-frequency events. TiRGN uses binary event indicators as a candidate mask to mitigate frequency bias. However, its local historical modeling lacks the capability to uncover latent high-order relationships, and the improved global historical modeling method still suffers from issues with historical interference.

The experimental results further reveal that while TiRGN-ANEL employs an adaptive neighborhood enhancement layer to improve the quality of embedding representations for entities and relations, it still falls short in the fine-grained mining of global historical facts. Notably, L^{2}TKG alleviates knowledge gaps caused by data sparsity through latent relation graphs. However, when handling dense datasets, such as GDELT, with a million-level data volume, this mechanism provides limited benefits and may even introduce noise from over-connected latent relations. DHE-TKG constructs hyperedges by grouping each node with its nearest k neighbors, which only expands local connectivity and fails to capture long-range historical dependencies. As a result, its hypergraph encoder provides limited benefits on complex temporal graphs. Beyond these, other approaches integrate contrastive learning (CENET), Hawkes process-based temporal modeling (HTCCN), or path search strategies (CognTKE) to strengthen temporal reasoning. Nevertheless, their performance still lags behind DHHGN, as they either suffer from limited historical coverage, insufficient handling of high-order relations, or vulnerability to noisy event sequences. LogCL achieves the best overall performance among all compared models, benefiting from its effective modeling of both local historical evolutionary patterns and global dependencies. However, for local history modeling, LogCL can only capture binary relationships between adjacent entities, failing to perceive implicit semantic relationships among multiple entities. In contrast, DHHGN utilizes a hypergraph convolutional module, thereby enhancing the model's ability to capture higher-order relationships. Regarding global history modeling, LogCL leverages two-hop historical subgraph sampling to capture global dependencies, while DHHGN ensures both breadth and depth in global history modeling through its dual-history enhancement mechanism.

5.3 Ablation Study

We conduct ablation studies on four datasets to evaluate the impact of each component in DHHGN by progressively removing key modules: the Dual History Enhancement Module (w/o DHE), Comprehensive History Modeling (w/o CHM), Relevance-aware History Screening (w/o RHS), Hybrid Hypergraph-Graph Module

Table 2: Entity prediction results on four datasets, where the best results are shown in bold font and the second-best results are underlined. All evaluation metrics are computed under the filtering setting. "-" indicates that results are not reported in the corresponding papers or that the code for the dataset/metric is unavailable.

Model	ICEWS14			ICEWS18			ICEWS05-15			GDELT		
	MRR	H@1	H@10	MRR	H@1	H@10	MRR	H@1	H@10	MRR	H@1	H@10
ComplEx (2016)	32.54	23.43	50.73	22.94	15.19	42.11	32.63	24.01	52.81	16.96	11.25	32.35
ConvE (2018)	35.09	25.23	54.68	24.51	16.23	44.51	33.81	24.78	54.95	16.55	11.02	31.60
Conv-TransE (2019)	33.80	25.40	53.99	22.11	13.94	42.28	33.03	24.15	54.32	16.20	10.85	30.86
RotatE (2019)	21.31	10.26	44.75	12.78	4.01	31.91	24.71	13.22	48.16	13.45	6.95	25.99
RE-NET (2020)	36.93	26.83	54.78	28.81	19.05	47.51	43.32	33.43	63.06	19.62	12.42	34.01
CyGNet (2020)	35.05	25.73	53.55	24.93	15.90	42.61	36.81	26.61	56.22	18.48	11.52	31.98
RE-GCN (2021)	40.39	30.66	59.21	30.58	21.01	48.75	48.03	37.33	68.27	19.64	12.42	33.69
TiRGN (2022)	44.04	33.83	63.84	33.66	23.19	54.22	50.04	39.25	70.71	21.67	13.63	37.60
HisMatch (2022)	46.42	35.91	66.84	33.99	23.91	53.94	52.85	42.01	73.28	22.01	14.45	36.61
CENET (2023)	39.02	29.62	57.49	27.85	18.15	46.98	41.95	32.17	60.43	20.23	12.69	34.92
L^2TKG (2023)	47.40	35.36	<u>71.05</u>	33.36	22.15	55.04	<u>57.43</u>	41.86	<u>80.69</u>	20.53	12.89	35.83
LogCL (2024)	<u>48.87</u>	<u>37.76</u>	70.26	<u>35.67</u>	24.53	<u>57.74</u>	57.04	<u>46.07</u>	77.87	<u>23.75</u>	14.64	<u>42.33</u>
HTCCN (2024)	45.39	36.58	-	35.63	24.90	-	51.94	40.32	-	23.46	<u>15.18</u>	-
DHE-TKG (2024)	40.02	30.13	-	29.23	19.15	-	45.05	34.16	-	-	-	-
TiRGN-ANEL (2025)	44.74	34.4	64.66	34.04	23.53	54.71	50.59	39.69	71.27	22.06	13.85	38.33
CognTKE (2025)	46.06	36.49	64.49	35.24	<u>25.21</u>	54.71	53.13	42.62	72.7	-	-	-
DHHGN	**56.95**	**46.65**	**75.89**	**43.53**	**31.71**	**66.87**	**68.74**	**59.14**	**85.40**	**32.16**	**20.28**	**57.34**

(w/o HHG), Graph Convolution (w/o gc), and Hypergraph Convolution (w/o hgc). As shown in Table 3, DHHGN outperforms all ablated variants, validating the effectiveness of its design. Specifically, removing the entire DHE module results in greater performance degradation than individually removing CHM or RHS, indicating that both global historical context and prediction-critical events contribute distinctively to predictions. Similarly, removing the entire HHG module causes more significant performance drops than disabling either graph convolution (gc) or hypergraph convolution (hgc), confirming the complementary roles of low-order neighborhood aggregation and high-order pattern mining in capturing local dynamics. Note that the model performance on the GDELT and ICEWS18 datasets drops more significantly after removing the hypergraph convolution module. This is because both datasets are dense in nature, where each timestamp contains a substantially larger number of co-occurring facts compared to ICEWS14 and ICEWS05-15. Such density implies the existence of complex latent high-order semantic dependencies among multiple entities and relations. In contrast, the ICEWS14 and ICEWS05-15 datasets are relatively sparse, with fewer concurrent events per timestamp. Consequently, the potential for high-order relational semantics is limited, and immediate interactions between entities have a greater impact on prediction. As a result, the removal of hypergraph convolution has a less pronounced effect on performance.

Table 3: The MRR (in percentage) results of the ablation studies.

Model	ICEWS14	ICEWS18	ICEWS05-15	GDELT
w/o DHE	42.13	32.48	48.82	19.93
w/o CHM	54.78	42.45	66.82	28.71
w/o RHS	44.34	33.85	50.76	21.84
w/o HHG	51.86	38.30	61.98	30.80
w/o gc	54.32	40.29	64.39	31.06
w/o hgc	56.63	42.79	68.11	31.13
DHHGN	**56.95**	**43.53**	**68.74**	**32.16**

5.4 Sensitivity Analysis

We conduct parameter sensitivity analysis on the balancing factor α across four datasets. As shown in Figure 4, the MRR metric exhibits an initial increase followed by a gradual decline as α rises, peaking within (0.2, 0.5). These results demonstrate that while the RHS module effectively enhances attention allocation to critical historical events, it may inevitably assign a low attention score to certain event entities, leading to the omission of global historical information. The CHM module can compensate for this shortcoming. When α is set to a low value, the model can consider important historical facts while maintaining attention to all global historical events, thus exhibiting a robust prediction performance. However, as α gradually increases, the inclusion of excessive irrelevant historical events introduces noise, which degrades the model performance. Notably, when $\alpha = 0.0$ (without CHM), the model maintains competitive performance, demonstrating the capability of RHS to leverage critical historical evidence. The sharp performance degradation at $\alpha = 1.0$ occurs because the prediction scope is strictly confined to events that occurred before, neglecting non-historical events. Additionally, it is observed that the MRR metric is more sensitive to changes in

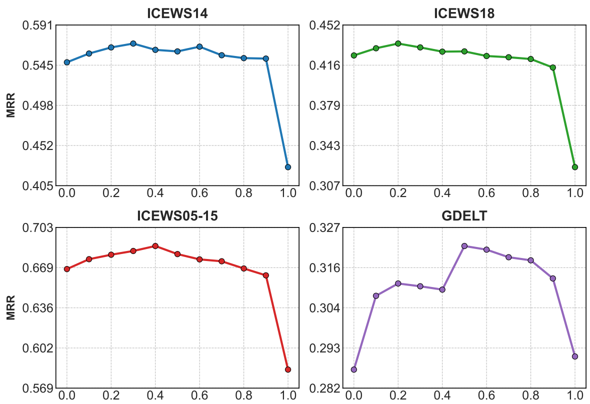

Figure 4: Experimental results of sensitivity analysis to parameter α.

α on the GDELT dataset; that is because the timestamp interval of the GDELT dataset is only 15 minutes, so there is a closer temporal correlation between events occurring at different timestamps, making model predictions more susceptible to the global historical context.

5.5 Historical and Non-historical Events Study

An event in the prediction test set is considered a historical event if it has occurred before; otherwise, it is termed a non-historical event. Non-historical events constitute 48.1% of the ICEWS14 test set, making it critical to minimize interference in their prediction during historical event modeling. While historical events leverage past patterns to boost accuracy, overemphasizing them can disrupt non-historical event predictions. Our DHHGN model, with refined historical modeling and high-order information extraction, effectively reduces this interference. As shown in Figure 5, TiRGN's performance drops by 9.97%, 4.74%, and 1.87% in Hits@1, Hits@3, and Hits@10 for non-historical events compared to historical ones, while DHHGN shows smaller declines of 5.72%, 1.27%, and 0.67%, demonstrating superior effectiveness and robustness.

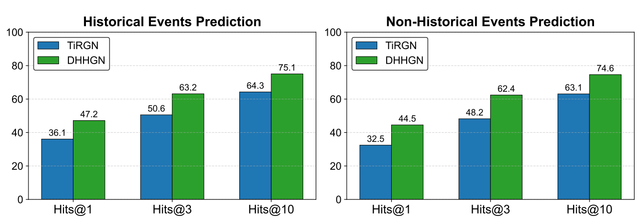

Figure 5: Performance comparison of TiRGN and DHHGN in predicting historical events versus non-historical events (ICEWS14).

5.6 Timestamp Level Prediction

To conduct a granular investigation of the model performance advantages, we compared the DHHGN and TiRGN architectures on the ICEWS14 dataset through a timestamp-level evaluation across 51 test set intervals. As demonstrated in Figure 6, DHHGN consistently outperforms TiRGN across all evaluation timestamps, maintaining a stable performance gap. Specifically, it achieves average improvements of 13.3% in Hits@1, 14.2% in Hits@3, and 11.8% in Hits@10, demonstrating both effectiveness and robustness.

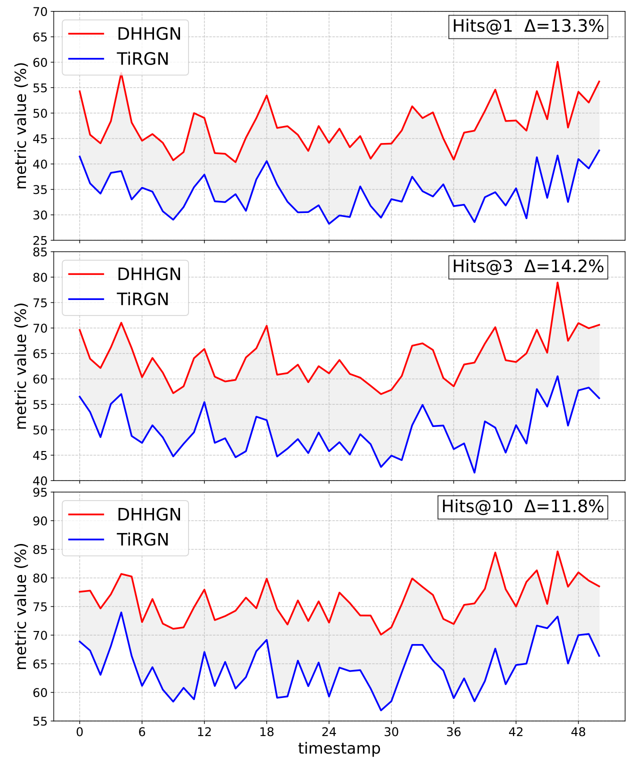

Figure 6: A comparative visualization of prediction capabilities between the DHHGN and TiRGN models at fine-grained timestamp levels, employing Hits@K as evaluation metrics over 51 timestamps in the test set of the ICEWS14 dataset.

6 Conclusion

In this paper, we propose DHHGN for TKG reasoning, which enhances prediction performance through fine-grained exploration of both local and global historical facts. We introduce hypergraphs integrated with traditional graphs to capture both low-order information and potential high-order interactions within a single snapshot in local historical data. Then, a dual-history enhancement module is designed to enable comprehensive consideration of all global historical facts while intensifying focus on prediction-relevant historical events. Extensive experiments demonstrate state-of-the-art performance on four benchmark datasets.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grant 62476247, 62073295 and 62072409, "Pioneer" and "Leading Goose" R&D Program of Zhejiang under Grant 2024C01214, and in part by the Zhejiang Provincial Natural Science Foundation under Grant LR21F020003.

References

[1] Luyi Bai, Mingcheng Zhang, Han Zhang, and Heng Zhang. 2023. FTMF: Few-shot temporal knowledge graph completion based on meta-optimization and fault-tolerant mechanism. *World Wide Web* 26, 3 (2023), 1243–1270.

[2] Antoine Bordes, Nicolas Usunier, Alberto Garcia-Duran, Jason Weston, and Oksana Yakhnenko. 2013. Translating embeddings for modeling multi-relational data. *Advances in neural information processing systems* 26 (2013).

[3] Elizabeth Boschee, Jennifer Lautenschlager, Sean O'Brien, Steve Shellman, James Starz, and Michael Ward. 2015. Icews coded event data. harvard dataverse. *V4 (http://dx. doi. org/10.7910/DVN/28075)* (2015).

[4] Yuxiang Cai, Qiao Liu, Yanglei Gan, Changlin Li, Xueyi Liu, Run Lin, Da Luo, and JiayeYang JiayeYang. 2024. Predicting the Unpredictable: Uncertainty-Aware Reasoning over Temporal Knowledge Graphs via Diffusion Process. In *Findings of the Association for Computational Linguistics ACL 2024*. 5766–5778.

[5] Yukun Cao, Lisheng Wang, and Luobin Huang. 2025. Dpcl-diff: Temporal knowledge graph reasoning based on graph node diffusion model with dual-domain periodic contrastive learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 39. 14806–14814.

[6] Kai Chen, Ye Wang, Xin Song, Siwei Chen, Han Yu, and Aiping Li. 2024. Temporal knowledge graph extrapolation via causal subhistory identification. In *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence*. 3298–3306.

[7] Tingxuan Chen, Jun Long, Liu Yang, Zidong Wang, Yongheng Wang, and Xiongnan Jin. 2024. Htccn: Temporal causal convolutional networks with hawkes process for extrapolation reasoning in temporal knowledge graphs. In *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*. 4056–4066.

[8] Tian Chen, Yimu Zhang, Xinwu Qian, and Jian Li. 2022. A knowledge graph-based method for epidemic contact tracing in public transportation. *Transportation Research Part C: Emerging Technologies* 137 (2022), 103587.

[9] Wei Chen, Huaiyu Wan, Shengnan Guo, Haoyu Huang, Shaojie Zheng, Jiamu Li, Shuohao Lin, and Youfang Lin. 2022. Building and exploiting spatial–temporal knowledge graph for next POI recommendation. *Knowledge-Based Systems* 258 (2022), 109951.

[10] Wei Chen, Huaiyu Wan, Yuting Wu, Shuyuan Zhao, Jiayaqi Cheng, Yuxin Li, and Youfang Lin. 2024. Local-global history-aware contrastive learning for temporal knowledge graph reasoning. In *2024 IEEE 40th International Conference on Data Engineering (ICDE)*. IEEE, 733–746.

[11] Wei Chen, Yuting Wu, Shuhan Wu, Zhiyu Zhang, Mengqi Liao, Youfang Lin, and Huaiyu Wan. 2025. Cogntke: A cognitive temporal knowledge extrapolation framework. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 39. 14815–14823.

[12] Wei Chen, Xiao Zhang, Tengjiao Wang, Bishan Yang, and Yi Li. 2017. Opinion-aware Knowledge Graph for Political Ideology Detection.. In *IJCAI*, Vol. 17. 3647–3653.

[13] Kyunghyun Cho, Bart Van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. 2014. Learning phrase representations using RNN encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078* (2014).

[14] Tim Dettmers, Pasquale Minervini, Pontus Stenetorp, and Sebastian Riedel. 2018. Convolutional 2d knowledge graph embeddings. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 32.

[15] Zifeng Ding, Bailan He, Jingpei Wu, Yunpu Ma, Zhen Han, and Volker Tresp. 2023. Learning meta-representations of one-shot relations for temporal knowledge graph link prediction. In *2023 international joint conference on neural networks (IJCNN)*. IEEE, 1–10.

[16] Merlin Donald. 1993. Human cognitive evolution: What we were, what we are becoming. *Social Research* (1993), 143–170.

[17] Alberto Garcia-Duran, Sebastijan Dumančić, 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*. 4816–4821.

[18] 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)*. 6669–6683.

[19] 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.

[20] Yujia Li, Shiliang Sun, and Jing Zhao. 2022. TiRGN: Time-Guided Recurrent Graph Network with Local-Global Historical Patterns for Temporal Knowledge Graph Reasoning.. In *IJCAI*. 2152–2158.

[21] 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)*. 290–296.

[22] Zixuan Li, Zhongni Hou, Saiping Guan, Xiaolong Jin, Weihua Peng, Long Bai, Yajuan Lyu, Wei Li, Jiafeng Guo, and Xueqi Cheng. 2022. HiSMatch: Historical Structure Matching based Temporal Knowledge Graph Reasoning. In *Findings of the Association for Computational Linguistics: EMNLP 2022*. 7328–7338.

[23] Zixuan Li, Xiaolong Jin, Saiping Guan, Wei Li, Jiafeng Guo, Yuanzhuo Wang, and Xueqi Cheng. 2021. Search from History and Reason for Future: Two-stage Reasoning on Temporal Knowledge Graphs. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*. 4732–4743.

[24] 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*. 408–417.

[25] Ruotong Liao, Xu Jia, Yangzhe Li, Yunpu Ma, and Volker Tresp. 2024. GenTKG: Generative Forecasting on Temporal Knowledge Graph with Large Language Models. In *NAACL-HLT (Findings)*.

[26] 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.

[27] Haithem Mezni. 2021. Temporal knowledge graph embedding for effective service recommendation. *IEEE Transactions on Services Computing* 15 (2021), 3077–3088.

[28] Tu Dinh Nguyen, Dat Quoc Nguyen, Dinh Phung, et al. 2018. A Novel Embedding Model for Knowledge Base Completion Based on Convolutional Neural Network. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*. 327–333.

[29] Maximilian Nickel, Volker Tresp, Hans-Peter Kriegel, et al. 2011. A three-way model for collective learning on multi-relational data.. In *Icml*, Vol. 11. 3104482–3104584.

[30] Michael Schlichtkrull, Thomas N Kipf, Peter Bloem, Rianne Van Den Berg, Ivan Titov, and Max Welling. 2018. Modeling relational data with graph convolutional networks. In *The semantic web: 15th international conference, ESWC 2018, Heraklion, Crete, Greece, June 3–7, 2018, proceedings 15*. Springer, 593–607.

[31] 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.

[32] Zhiqing Sun, Zhi-Hong Deng, Jian-Yun Nie, and Jian Tang. 2019. Rotate: Knowledge graph embedding by relational rotation in complex space. *arXiv preprint arXiv:1902.10197* (2019).

[33] Théo Trouillon, Johannes Welbl, Sebastian Riedel, Éric Gaussier, and Guillaume Bouchard. 2016. Complex embeddings for simple link prediction. In *International conference on machine learning*. PMLR, 2071–2080.

[34] Shikhar Vashishth, Soumya Sanyal, Vikram Nitin, and Partha Talukdar. 2020. Composition-based Multi-Relational Graph Convolutional Networks. In *International Conference on Learning Representations*.

[35] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. 2018. Graph Attention Networks. In *International Conference on Learning Representations*.

[36] Jiapu Wang, Sun Kai, Linhao Luo, Wei Wei, Yongli Hu, Alan Wee-Chung Liew, Shirui Pan, and Baocai Yin. 2024. Large language models-guided dynamic adaptation for temporal knowledge graph reasoning. *Advances in Neural Information Processing Systems* 37 (2024), 8384–8410.

[37] Yi Xu, Junjie Ou, Hui Xu, and Luoyi Fu. 2023. Temporal knowledge graph reasoning with historical contrastive learning. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 37. 4765–4773.

[38] Chenhan Yuan, Qianqian Xie, Jimin Huang, and Sophia Ananiadou. 2024. Back to the future: Towards explainable temporal reasoning with large language models. In *Proceedings of the ACM Web Conference 2024*. 1963–1974.

[39] Mengqi Zhang, Yuwei Xia, Qiang Liu, Shu Wu, and Liang Wang. 2023. Learning latent relations for temporal knowledge graph reasoning. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. 12617–12631.

[40] Yuchao Zhang, Xiangjie Kong, Zhehui Shen, Jianxin Li, Qiuhua Yi, Guojiang Shen, and Bo Dong. 2024. A survey on temporal knowledge graph embedding: Models and applications. *Knowledge-Based Systems* (2024), 112454.

[41] Yuchao Zhang, Xiangjie Kong, Kailun Ye, Guojiang Shen, and Shangfei Zheng. 2025. Tackling Sparse Facts for Temporal Knowledge Graph Completion. In *THE WEB CONFERENCE 2025*.

[42] 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.

[43] Lin Zhu, Yizong Xing, Luyi Bai, and Xiwen Chen. 2023. Few-shot link prediction with meta-learning for temporal knowledge graphs. *Journal of Computational Design and Engineering* 10, 2 (2023), 711–721.

A Appendix

A.1 Pseudocode of DHHGN

The algorithm 1 delineates the complete sequence of steps employed by the DHHGN model to predict queries at time t.

Algorithm 1 DHHGN pseudocode for query time t

Input: query set $QSet$ at time t, length of local history window n, weighting hyperparameter α, local historical KG snapshot sequence $\mathcal{G}_{local} = \{\mathcal{G}_{t-n}, \mathcal{G}_{t-n+1}, ..., \mathcal{G}_{t-1}\}$, global historical KG snapshot sequence $\mathcal{G}_{global} = \{\mathcal{G}_0, \mathcal{G}_1, ..., \mathcal{G}_{t-1}\}$.

Output: Updated entity representation and updated model parameters.

Initialize entity-hyperedge mapping matrix **D**.
Initialize the embeddings of both entities and relations.
Generate sparse event matrix **M**.
for $Q \in QSet$ **do**
  Retrieve relevant events from global historical sequences $\mathcal{G}_{global}$.
  Merge all relevant events of the query Q to obtain an event set C_Q.
end for
Collect all relevant event sets corresponding to queries and construct query related subgraphs $\mathcal{G}_t^{his}$.
Aggregate $\mathcal{G}_t^{his}$ using an Enhanced GAT to obtain global historical dependencies.
for $\mathcal{G} \in \mathcal{G}_{local}$ **do**
  Using RGCN for low-order neighborhood aggregation to obtain $\mathbf{H}_g$.
  Using DHAM with matrix **D** for high-order group aggregation to obtain $\mathbf{H}_{hg}$.
  Using gating mechanism to fuse $\mathbf{H}_g$ and $\mathbf{H}_{hg}$ to obtain $\mathbf{H}_{fus}$
  Using GRU to obtain evolved entity representations.
end for
Adaptively aggregate local and global historical dependencies using gate unit.
Use two parallel decoders for decoding and utilize α to balance contributions.
Compute loss and update the network parameters.
t=t+1

A.2 Amount of Historical and Non-historical Events

We conduct statistics on the amount of historical and non-historical events in the test set of four datasets, as shown in Table 4 for details.

Table 4: The amount of historical and non-historical events in four datasets.

Dataset	Total	historical	non-historical
ICEWS14	13222	6856	6366
ICEWS18	49545	24985	24560
ICEWS05-15	46159	31567	14592
GDELT	305241	198179	107062

A.3 Effect of the Number of Enhanced GAT Layers

This study presents a comprehensive comparison of Enhanced GAT architectures with varying layer depths on two benchmark temporal knowledge graph datasets. As illustrated in figure 7, the results indicate that the performance of the model reaches its optimal level when the number of layers is set to 2, and this performance is more pronounced on the GDELT dataset. The findings underscore the importance of architecture-depth hyperparameter tuning specific to knowledge graph properties.

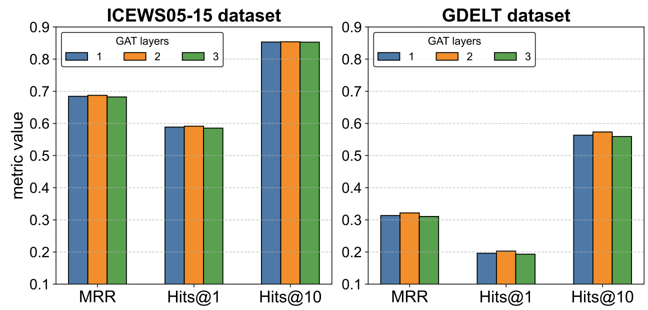

Figure 7: Performance comparison of Enhanced GAT with different layers

A.4 Case Analysis

To demonstrate the ability of DHHGN to perceive the importance of global historical events and detect high-order information, we conduct a case study using the ICEWS14 dataset and perform a comparative analysis with TiRGN. For each query instance, the top six ranked candidate entities were retrieved, with detailed results documented in Table 5.

Table 5: The top-6 predicted entities by DHHGN and TiRGN for two queries on the ICEWS14 dataset. Note that the underlined predicted entities indicate that the entity (as a tail entity) has formed an event with the query in historical records.

Query&Truth	DHHGN	TiRGN
Query1	NATO	Abdullah
(Ashraf Ghani	Barack Obama	Barack Obama
Ahmadzai,	Abdullah	John Kerry
Consult,?,2014-11-11)	Tajikistan	Afghanistan
Truth	Xi Jinping	Xi Jinping
NATO	Nuri al-Maliki	Raheel Sharif
Query2	Japan	Vietnam
(China,Mobilize or	South Korea	Malaysia
increase armed forces,	Malaysia	South Korea
?,2014-11-11)	Philippines	Japan
Truth	Vietnam	Iran
Japan	Thailand	Afghanistan

In query 1, we aim to predict which international entity Ashraf Ghani, the President of Afghanistan, consulted on November 11, 2014. While both models can capture all relevant historical entities, including NATO, DHHGN successfully ranks NATO as the top prediction. Crucially, this achievement is enabled by the DHE component, which employs two complementary submodules that distinguish critical global historical events from noise through an importance-weighted mechanism while utilizing a relevant historical event mask to prevent information loss due to attention sparsity in global historical events. In contrast, TiRGN assigns equal attention to all relevant historical events, resulting in prioritizing "Abdullah", while the ground truth fails to appear in its top six predictions.

In query 2, we aim to predict (China, Mobilize or increase armed forces, ?). It is noteworthy that no event in global history shares the same head entity and relation as the query. Consequently, the global historical components of both models do not function. Nevertheless, DHHGN successfully predicts "Japan" and ranks it first, demonstrating the model's ability to capture high-order information from local events. In contrast, TiRGN exhibits susceptibility to interference from low-order interactions, such as (China, Make pessimistic comment, Vietnam, 2014-11-08), leading to prediction deviations.