Query-Aware Hierarchical Contrastive Path Learning for Inductive Temporal Knowledge Graph Reasoning

Linjie Shi*†, Zhixin Shi*†‡, Xiaoyu Kang*†, Rui He*†, Shihao Zhao*†
*Institute of Information Engineering, Chinese Academy of Sciences, Beijing, China
†School of Cyber Security, University of Chinese Academy of Sciences, Beijing, China
E-mail: {shilinjie, shizhixin, kangxiaoyu, herui, zhaoshihao}@iie.ac.cn

***Abstract*—Extrapolation reasoning over Temporal Knowledge Graphs is crucial for forecasting future facts involving emerging entities in intelligent systems. Existing inductive path-based approaches are often limited by insufficient contextual adaptability and weak global logical consistency, mainly due to static attention mechanisms and local edge-level supervision. To address these limitations, we propose the Query-Aware Hierarchical Contrastive Path (QHCP) framework, a novel architecture that combines dynamic query-aware attention with hierarchical contrastive supervision for temporal knowledge graph reasoning. Specifically, QHCP dynamically prioritizes relevant evidence at each reasoning step by incorporating evolving path contexts, while hierarchical contrastive learning jointly regularizes message-level semantics and path-level logic to improve both local coherence and global consistency of reasoning chains. Extensive experiments on four real-world TKG datasets demonstrate that QHCP achieves significant improvements over state-of-the-art baselines, with gains of up to 2.62% in MRR on the WIKI dataset and retains over 96% of its performance when transferred to a dataset containing entirely unseen entities, validating its superior inductive generalization capability.**

***Index Terms*—Temporal Knowledge Graph, Knowledge Graph Reasoning, Inductive Generalization, Contrastive Learning.**

I. Introduction

Modeling and reasoning over dynamically evolving real-world knowledge serves as a cornerstone of modern artificial intelligence. Temporal Knowledge Graphs (TKGs) [1] emerge as a natural solution, capturing the temporal dynamics of entities and relations by incorporating timestamped facts into knowledge graphs. TKG reasoning encompasses two tasks: interpolation and extrapolation. Specifically, extrapolation, which refers to predicting future facts beyond the observed timeline, proves critical for applications including financial forecasting [2] and clinical decision support systems [3].

Real-world TKGs evolve dynamically as new entities continuously emerge, making generalization to unseen entities essential. For instance, when predicting future collaborations involving senior professors, models must reason about interactions with emerging researchers that are absent from the training data. This necessitates a paradigm shift from traditional transductive (entity-dependent) models, which learn fixed entity embeddings and are confined to reasoning over a closed set of known entities [4], [5], to more flexible inductive (entity-independent) approaches that learn transferable patterns [6], [7]. Existing inductive methods typically construct historical temporal subgraphs around queries, select relevant edges along relational paths, and recursively aggregate evidence to perform multi-hop reasoning. However, this incremental path learning paradigm faces fundamental limitations in both evidence selection and supervision mechanisms, manifesting in two key bottlenecks:

Static context-aware bottleneck. Existing methods [4], [5] rely on fixed relational semantics and temporal distance to regulate edge contributions. Although some works introduce query-aware attention [8], [9], relevance estimation remains anchored to the initial query relation, without modeling the evolving semantic state of the reasoning path. Consequently, attention weights fail to adapt to semantic shifts caused by path extension. For example, in the reasoning path *Smith* $\rightarrow$ *advises* $\rightarrow$ *Student A* $\rightarrow$ *collaborates with* $\rightarrow$ *?*, the semantic focus should shift from *advise* to *collaborate* patterns after reaching *Student A*. However, existing methods still evaluate candidate evidence based solely on the initial query relation, ignoring the temporal-logical trajectory accumulated along the path and leading to semantically diluted, path-inconsistent evidence selection.

Path logical inconsistency bottleneck. Mainstream path-based methods [4], [10] perform multi-hop reasoning via iterative message passing, while supervision is confined to the final link prediction objective. As a result, intermediate reasoning steps are not explicitly constrained, making it difficult to ensure the global logical coherence of the entire reasoning chain. Consequently, paths may contain locally plausible yet globally contradictory transitions. For instance, the path *United States* $\rightarrow$ *provide aid* $\rightarrow$ *Syria* $\rightarrow$ *condemn* $\rightarrow$ *Russia* appears locally plausible yet globally inconsistent when considered as a whole.

To overcome these limitations, we propose the Query-Aware Hierarchical Contrastive Path (QHCP) framework, integrating dynamic evidence prioritization and global logical alignment to enable robust inductive temporal-relational pattern learning. We introduce a Dynamic Query-Aware Attention (DA) module that incorporates evolving path contexts into relevance calculations at each reasoning step, enabling the model to adaptively prioritize evidence based on the semantic trajectory of the current reasoning path. Concurrently, the Hierarchical

‡ The corresponding author

Contrastive Path Enhancement (HCPE) module elevates contrastive learning to the logical path space through a hierarchical process: message-level ensures local semantic consistency at individual reasoning steps, while path-level reinforces end-to-end logical consistency across entire reasoning chains. The fine-grained prioritization decisions of DA are guided by the global logic-aware signals of HCPE, ensuring the selected paths maintain both local relevance and global consistency.

Our main contributions are summarized as follows:

- We explicitly identify the two fundamental limitations of existing inductive TKG reasoning methods as static context awareness and logical inconsistency issues.
- We propose the QHCP framework, which integrates dynamic query-aware attention for adaptive evidence prioritization and hierarchical contrastive path enhancement for dense path-level supervision.
- We conduct comprehensive experiments demonstrating that QHCP outperforms existing inductive TKG reasoning methods and exhibits exceptional inductive generalization in cross-graph transfer scenarios.

II. Related Work

A. Transductive Reasoning

Knowledge graph reasoning has shifted from static to dynamic paradigms, divided into transductive and inductive frameworks. Transductive methods rely on fixed entity embeddings for scoring, including static methods such as DistMult [11], ComplEx [12], and ConvE [13]. For temporal dynamics, TTransE [5] and TA-DistMult [14] integrate temporal dimensions into scoring, while DE-SimplE [15] employs diachronic embeddings to model evolution. Graph neural network-based methods further enhance temporal representation learning: RE-GCN [4] captures evolutionary patterns through recurrent graph convolutions, whereas xERTE [10] employs attention mechanisms for modeling. However, these entity-dependent approaches suffer from limited generalization capabilities when encountering unseen entities during training, resulting in significant degradation of inference performance.

B. Inductive Reasoning and Supervision

Inductive methods are proposed to overcome this limitation by learning entity-independent structural patterns. GraIL [16] proposes the use of graph neural networks for relation-sensitive labeling on subgraphs, allowing reasoning solely based on structural features rather than entity embeddings. This paradigm has been extended to temporal scenarios: DaeMon [6] introduces adaptive message passing, while TiPNN [7] captures evolutionary patterns via temporal neural networks.However, these methods typically rely on static attention mechanisms and confine supervision to the final prediction objective, failing to explicitly constrain the logical consistency of intermediate reasoning steps.

While reinforcement learning (RL) approaches [17] aim to achieve explicit long-term planning for path finding, they often suffer from reward sparsity in complex temporal environments. Alternatively, contrastive learning has been introduced to enhance supervision. KGCL [18] constructs contrastive samples via structural perturbations, while CENET [19] builds temporal neighbor contrasts. However, these works remain confined to representation learning and have not been applied to supervising the logical consistency of reasoning paths.

To bridge these gaps, we propose the QHCP framework by introducing dynamic query-aware attention for adaptive evidence prioritization and applying hierarchical contrastive learning to enable joint optimization from local semantics to global logic in inductive temporal reasoning.

III. Methodology

In this section, we propose the Query-Aware Hierarchical Contrastive Path (QHCP) framework for inductive temporal knowledge graph reasoning. As illustrated in Fig. 1, QHCP employs L-layer message passing with dynamic query-aware attention for adaptive evidence prioritization at each layer, coupled with hierarchical contrastive learning to ensure logical consistency of reasoning paths. We first formalize the problem definition, then detail the core components: query fusion graph construction, dynamic query-aware attention, and hierarchical contrastive path learning.

A. Problem Definition

A temporal knowledge graph is represented as a sequence of time-ordered snapshots $\mathcal{G} = \{\mathcal{G}_1, \mathcal{G}_2, \ldots, \mathcal{G}_T\}$, where each snapshot $\mathcal{G}_t = (\mathcal{E}, \mathcal{R}, \mathcal{F}_t)$ contains an entity set $\mathcal{E}$, a relation set $\mathcal{R}$, and a fact set $\mathcal{F}_t$ at timestamp t. Each fact is represented as a quadruple (s, r, o, t), where $s, o \in \mathcal{E}$ are the subject and object entities, $r \in \mathcal{R}$ is the relation, and $t \in \mathcal{T}$ is the timestamp.

Extrapolation Reasoning Task. Given a query $q = (s_q, r_q, ?, t_q)$ and a historical snapshot set $\mathcal{G}_{\text{train}} = \{\mathcal{G}_1, \ldots, \mathcal{G}_{t_{\max}}\}$ where $t_q > t_{\max}$, the objective is to predict the object entity. We define s_q as the query entity and o as the candidate entity.

We adopt entity-independent modeling, performing reasoning by learning relation-temporal patterns. We learn a scoring function f to rank candidate entities o:

$$\hat{o} = \arg\max_{o \in \mathcal{E}} f\left(q, \mathcal{G}_{\text{train}}, o\right). \tag{1}$$

B. Query Fusion Graph Construction

To better capture the connection patterns in historical snapshots, we construct a query fusion graph H_q by integrating multiple historical snapshots into a unified temporal graph structure.

Given a query $q = (s_q, r_q, ?, t_q)$, we extract k consecutive historical snapshots $\{\mathcal{G}_{t_q-k}, \ldots, \mathcal{G}_{t_q-1}\}$ preceding the query timestamp t_q. Fusion is achieved by attaching timestamps of each subgraph to relations:

$$H_q = \left\{ (s, r_t, o) \mid (s, r, o, t) \in \bigcup_{i=1}^{k} \mathcal{G}_{t_q-i} \right\} \tag{2}$$

where r_t denotes the relation type r appended with timestamp t. For clearer notation, we define $\psi_{r,t}^{u \to v}$ as a temporal edge

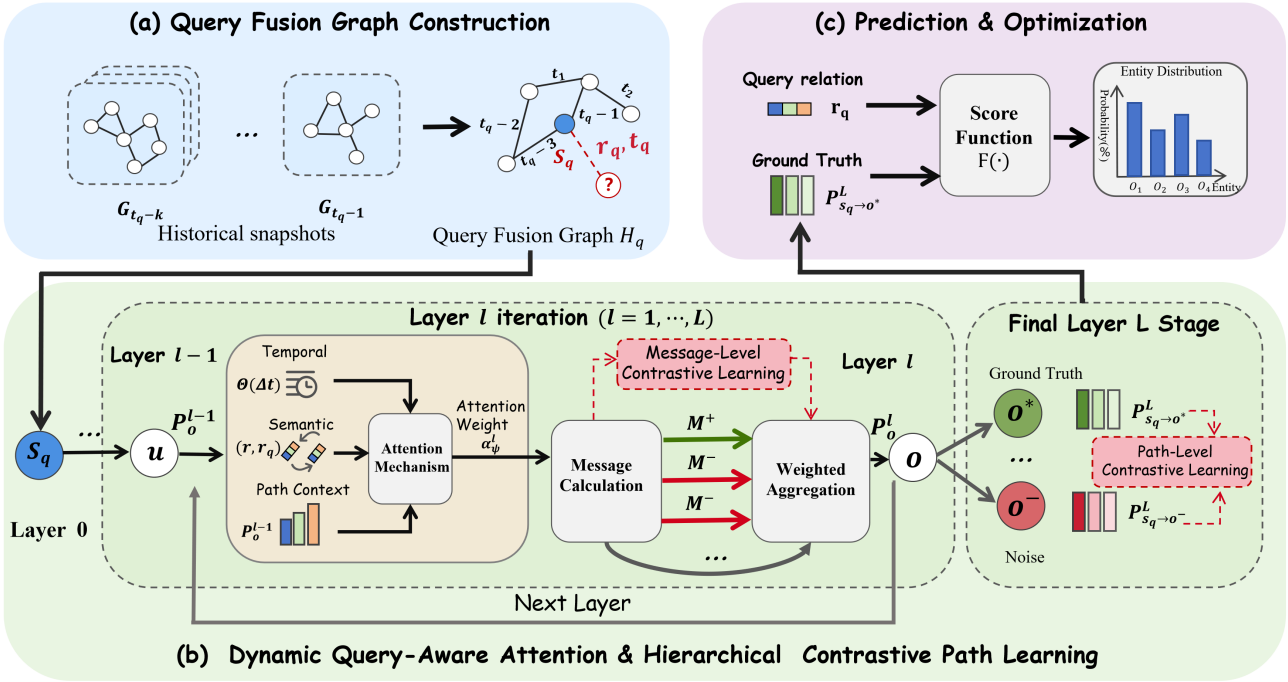

Fig. 1. Overall architecture of the QHCP framework. The model includes three components: (a) query fusion graph construction from historical snapshots, (b) dynamic query-aware attention and hierarchical contrastive learning for path representation learning, and (c) final entity prediction.

(u, r_t, v) , which denotes the edge from the predecessor entity u to the successor entity v via relation r at time t.

C. Dynamic Query-Aware Attention

To adaptively prioritize query-relevant evidence within the L-layer message passing framework, we propose a dynamic edge attention module that computes fine-grained weights for temporal edges in H_q by integrating semantic similarity, temporal patterns, and evolving path contexts.

At layer l $(1 \leq l \leq L)$, when updating the path representation of candidate entity o, attention weights are calculated for all its incoming edges $\psi^{u \to o}$ (where u denotes predecessor entities of o) to quantify their contribution. Specifically, for a given edge $\psi^{u \to o}_{r,t}$, we first project the relation r and the query relation r_q into a unified semantic space to obtain $\Phi(r), \Phi(r_q) \in \mathbb{R}^d$, and measure their semantic relevance via cosine similarity. Concurrently, we encode the temporal interval $\Delta t = |t_q - t|$ between the query timestamp and the edge timestamp using periodic functions [20] $\Theta(\Delta t) = [\cos(w_1 \Delta t), \ldots, \cos(w_d \Delta t)]$ with learnable frequencies w_i. To encode the reasoning trajectory $(s_q \to \cdots \to u \to o)$, we incorporate the path representation $\mathbf{p}^{l-1}_{s_q \to u}$, which recursively captures temporal relations and event dependencies along the multi-hop chain.

The attention weight is computed by concatenating these components followed by a nonlinear transformation:

$$\alpha^l_{\psi^{u \to o}_{r,t}} = \sigma\Big(w^l_\alpha \cdot \mathrm{GELU}\Big(W^l_\alpha \big[\Theta(\Delta t) \parallel \mathbf{sim}(\Phi(r), \Phi(r_q)) \parallel \mathbf{p}^{l-1}_{s_q \to u}\big]\Big)\Big) \tag{3}$$

where W^l_α, w^l_α are learnable parameters, $\mathrm{GELU}(\cdot)$ denotes the activation function, which empirically provides smoother optimization in attention-based feature fusion, σ is the sigmoid function, $\parallel$ denotes concatenation, and $\mathrm{sim}(\mathbf{a}, \mathbf{b}) = \frac{\mathbf{a}^\top \cdot \mathbf{b}}{\|\mathbf{a}\|_2 \|\mathbf{b}\|_2}$ is the cosine similarity function. By dynamically recalculating weights based on $\mathbf{p}^{l-1}_{s_q \to u}$ at each layer, we enable the module to adaptively adjust the importance of edges as reasoning depth increases.

D. Hierarchical Contrastive Path Learning

To ensure both local semantic coherence and global logical consistency, we construct path representations from query entity s_q to candidate entity o through recursive message aggregation, enhanced by hierarchical contrastive learning.

1) Temporal Path Aggregation: We initialize path representations by setting $\mathbf{p}^0_{s_q \to o} = \Phi(r_q)$ when $o = s_q$ and $\mathbf{p}^0_{s_q \to o} = \vec{0}$ otherwise.

At layer l $(1 \leq l \leq L)$, for each incoming edge $\psi^{u \to o}$ of o,we first construct its query-aware representation by fusing temporal encoding and relation semantics via a feedforward transformation. The attention-weighted message is then computed as:

$$M^l_{\psi^{u \to o}_{r,t}} = F\left(\Theta(\Delta t) \parallel \Phi_{r_q}(r)\right) \cdot \alpha^l_{\psi^{u \to o}_{r,t}} \tag{4}$$

where $\Phi_{r_q}(r) = W_{r_q} r + b_{r_q}$ is the query-specific relation embedding with learnable parameters W_{r_q}, b_{r_q}, $F(\cdot)$ is a feedforward network and $\alpha^l_{\psi^{u \to o}_{r,t}}$ is the dynamic attention

weight defined in Eq.(3). We then aggregate edge messages with upstream path representations:

$$\mathbf{p}^{l}_{s_q \to o} = \sum_{\psi^{u \to o} \in H_q} \alpha^{l}_{\psi} \cdot \left(M^{l}_{\psi} \otimes \mathbf{p}^{l-1}_{s_q \to u} \right) \tag{5}$$

where $\otimes$ denotes element-wise multiplication. After L iterations, we obtain the final path representation $\mathbf{p}^{L}_{s_q \to o}$.

2) Hierarchical Contrastive Path Enhancement: We introduce contrastive learning at both message and path levels to provide supervision from local transitions to complete reasoning chains.

Message-Level. At each layer l ($1 \leq l \leq L$), we apply contrastive learning to edge messages.

Positive samples are drawn from incoming edges $\psi^{u \to o^*}_{r,t}$ of the ground-truth answer o^*, where predecessors u satisfy $\|\mathbf{p}^{l-1}_{s_q \to u}\| > \epsilon$ (ϵ is the hyperparameter). Negative samples are drawn from edges $\psi^{u \to o^-}_{r,t}$ pointing to incorrect entities $o^- \in (\mathcal{B} \setminus \{o^*\}) \cup \text{Random}(\mathcal{E})$, where $\mathcal{B}$ denoting in-batch answers. In-batch negatives typically share similar query contexts with the positive instance, providing relatively hard contrastive signals while maintaining computational efficiency.

Let $s(M, r_q) = \text{sim}(M, \Phi(r_q))/\tau$. The contrastive loss is defined as:

$$\mathcal{L}^{l}_{\text{msg}} = -\log \frac{\sum_{M^+} \exp(s(M^+, r_q))}{\sum_{M^+} \exp(s(M^+, r_q)) + \sum_{M^-} \exp(s(M^-, r_q))} \tag{6}$$

where $M^+ = M^{l}_{\psi^{u \to o^*}_{r,t}}$ and $M^- = M^{l}_{\psi^{u \to o^-}_{r,t}}$ are positive and negative messages, and τ is the temperature.

Path-Level. At the final layer L, we further enforce global path discrimination:

$$\mathcal{L}_{\text{path}} = -\log \frac{\exp(\text{sim}(\mathbf{p}^{L}_{s_q \to o^*}, \Phi(r_q))/\tau)}{\sum_{o \in \{o^*\} \cup \mathcal{O}^-} \exp(\text{sim}(\mathbf{p}^{L}_{s_q \to o}, \Phi(r_q))/\tau)} \tag{7}$$

where $\mathcal{O}^-$ contains in-batch and randomly sampled entities, among which in-batch candidates typically exhibit higher semantic overlap with the positive path representation.The total contrastive loss combines both levels:

$$\mathcal{L}_{\text{HCPE}} = \frac{1}{L} \sum_{l=1}^{L} \mathcal{L}^{l}_{\text{msg}} + \lambda_p \mathcal{L}_{\text{path}} \tag{8}$$

where λ_p is the path-level weight hyperparameter.

E. Inference and Training

Inference. Given query q and candidate o, we concatenate the aggregated path representation $\mathbf{p}^{L}_{s_q \to o}$ with query relation embedding $\Phi(r_q)$, and compute the confidence score via:

$$\mathcal{S}(o \mid q) = \sigma \left(F\big([\mathbf{p}^{L}_{s_q \to o} \parallel \Phi(r_q)]\big) \right) \tag{9}$$

where σ is the sigmoid function and $F(\cdot)$ is a feedforward network. We select the highest-scoring entity as the prediction.

Training. We optimize prediction accuracy and logical consistency jointly via a combined loss. The prediction component is defined as:

$$\mathcal{L}_{\text{pred}} = -\log \left(\frac{\exp(\mathcal{S}(o^* \mid q))}{\sum_{o' \in \mathcal{E}} \exp(\mathcal{S}(o' \mid q))} \right). \tag{10}$$

The total loss is:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{pred}} + \lambda \mathcal{L}_{\text{HCPE}} \tag{11}$$

where λ balances prediction accuracy and logical consistency.

IV. Experiments

A. Experimental Setup

1) Datasets: We evaluate QHCP on four public real-word TKG datasets: ICEWS18 [1] contains political and economic events; GDELT [21] comprises real-time global news information; WIKI [5] consists of Wikipedia-based entity facts with annual time granularity; YAGO [22] integrates Wikipedia and WordNet with rich temporal annotations. All datasets are split chronologically (training set $<$ validation set $<$ test set) with a ratio of 80%, 10%, and 10%.

2) Baselines: We compare QHCP with 11 representative baselines (1) Static KG reasoning: DistMult [11], ComplEx [12], ConvE [13]; (2) Temporal KG interpolation: TTransE [5], TA-DistMult [14], DE-SimplE [15]; (3) Temporal KG extrapolation: RE-GCN [4], TITer [17], xERTE [10], DaeMon [6], TiPNN [7].

3) Evaluation Metrics: We adopt standard evaluation metrics for knowledge graph reasoning: MRR (Mean Reciprocal Rank) and Hits@k (the proportion of correct quadruples ranked in the top k positions) under the time-aware filtered setting, with evaluation logic aligned to baseline models.

4) Implementation Details: The embedding dimension is set to $d = 64$. The number of path aggregation layers L is set to 6 for ICEWS18 and GDELT, and 4 for WIKI and YAGO. We perform grid search over historical graph window size k, with optimal lengths of 25, 15, 10, 8 corresponding to ICEWS18, GDELT, WIKI, and YAGO respectively. For parameter learning, the number of negative samples is set to 64; the hyperparameter λ in total loss is set to 1, path-level weight $\lambda_p = 0.5$, $\epsilon = 0.1$ and temperature $\tau = 0.1$. We employ the Adam optimizer with a learning rate of $5e^{-4}$ (or $1e^{-4}$ for YAGO), set the maximum number of training epochs to 20.

B. Results and Analysis

Table I presents the results on four benchmark datasets. QHCP consistently achieves the best performance, with particularly significant improvements on semantically rich datasets. Specifically, compared to the state-of-the-art entity-independent baseline model TiPNN, QHCP achieves absolute improvements of 2.62% and 1.42% in MRR on WIKI and YAGO, respectively, while Hits@1 increases by 2.86% and 2.92%, respectively. On the event-driven datasets ICEWS18 and GDELT, the performance gains remain moderate yet stable (with MRR improvements of 0.98% and 1.10%, respectively).

These performance improvements are attributed to QHCP's hierarchical reasoning architecture: through the synergistic interplay of query-aware dynamic evidence prioritization and path-level contrastive optimization, the model focuses on high-quality reasoning chains within complex semantic structures while maintaining stable generalization in event-based knowledge graphs with strong temporal regularities. The moderate

TABLE I
OVERALL PERFORMANCE COMPARISON OF DIFFERENT METHODS. THE BEST RESULTS ARE HIGHLIGHTED IN **BOLD**, AND THE SECOND-BEST RESULTS ARE HIGHLIGHTED WITH UNDERLINE.

Model	ICEWS18				GDELT				WIKI				YAGO			
	MRR	H@1	H@3	H@10	MRR	H@1	H@3	H@10	MRR	H@1	H@3	H@10	MRR	H@1	H@3	H@10
DistMult	11.73	7.21	13.05	21.14	8.82	5.73	10.12	17.46	11.05	9.08	11.13	17.09	44.56	25.83	48.62	59.15
ComplEx	23.21	15.47	27.38	42.56	17.23	11.49	19.85	32.79	24.72	19.94	27.59	35.21	44.63	26.04	48.45	59.37
ConvE	24.79	16.48	29.57	44.93	16.81	11.26	19.15	31.98	14.76	11.65	16.59	22.68	42.41	23.51	46.42	61.09
TTransE	8.54	2.07	8.79	22.23	5.67	0.59	5.12	15.58	29.58	21.96	34.76	42.75	31.45	18.39	41.24	51.57
TA-DistMult	16.98	8.84	18.73	33.97	12.24	5.98	13.21	23.92	44.86	40.25	49.08	52.08	55.21	48.47	59.93	67.05
DE-SimplE	19.57	11.82	22.19	35.18	19.96	12.48	21.67	34.06	45.75	42.93	48.05	49.91	55.19	51.96	57.64	60.53
RE-GCN	31.24	21.76	35.02	49.43	20.31	12.98	21.65	34.37	78.12	74.39	80.96	84.25	84.67	81.32	86.84	90.45
TITer	30.65	22.71	34.12	45.52	15.79	11.29	15.94	24.76	76.03	73.51	78.04	79.58	87.98	85.42	90.47	90.76
xERTE	29.97	21.68	34.05	46.27	18.45	12.65	20.43	30.82	71.76	68.67	76.73	79.63	84.72	80.65	88.57	90.29
DaeMon	30.25	22.01	35.24	48.36	21.00	13.27	21.57	32.03	81.77	76.84	86.28	88.32	90.23	87.61	91.81	92.43
TiPNN	31.87	22.36	36.16	50.89	20.98	13.96	22.83	34.51	82.89	78.92	86.03	88.47	92.86	90.24	92.87	93.51
QHCP	**32.85**	**23.52**	**37.03**	**51.28**	**22.08**	**14.95**	**23.90**	**35.49**	**85.51**	**81.78**	**88.63**	**91.05**	**94.28**	**93.16**	**94.38**	**95.89**

improvements on ICEWS18 and GDELT indicate that the strong temporal regularities inherent in event-driven graphs have been adequately captured by existing methods, leaving limited room for further optimization.

C. Ablation Study

We conduct an ablation study on the ICEWS18 and WIKI datasets to verify the necessity of each component. As shown in Table II, we analyze the contributions of DA and the two-level HCPE mechanism through five variants.

TABLE II
ABLATION STUDY ON ICEWS18 AND WIKI.

Model	ICEWS18				WIKI			
	MRR	H@1	H@3	H@10	MRR	H@1	H@3	H@10
w/o DA	31.73	21.85	35.76	49.94	83.68	79.42	86.91	89.73
w/o HCPE message	32.38	22.76	36.52	50.81	84.79	80.95	87.98	90.51
w/o HCPE path	31.95	22.14	35.98	50.26	84.18	80.21	87.34	89.98
w/o HCPE	31.64	21.72	35.61	49.78	83.94	79.86	87.15	89.62
w/o DA+HCPE	31.26	21.34	35.18	49.35	83.12	78.65	86.28	88.94
QHCP	32.85	23.52	37.03	51.28	85.51	81.78	88.63	91.05

Removing message-level supervision results in MRR drops of 0.47% and 0.72% on ICEWS18 and WIKI, respectively, indicating its role in maintaining local semantic coherence at individual reasoning steps. Removing path-level supervision causes decreases of 0.90% and 1.33%, confirming its necessity for enforcing global logical consistency across complete reasoning chains. When both levels are removed (w/o HCPE), performance drops reach 1.21% and 1.57%, demonstrating that the hierarchical design enables complementary supervision from local to global granularities.

The removal of DA leads to performance degradation of 1.12% and 1.83%, validating that dynamic query-aware attention is crucial for adaptive evidence prioritization. The total drop from removing both modules (w/o DA+HCPE: 1.59% and 2.39%) is slightly less than the sum of individual

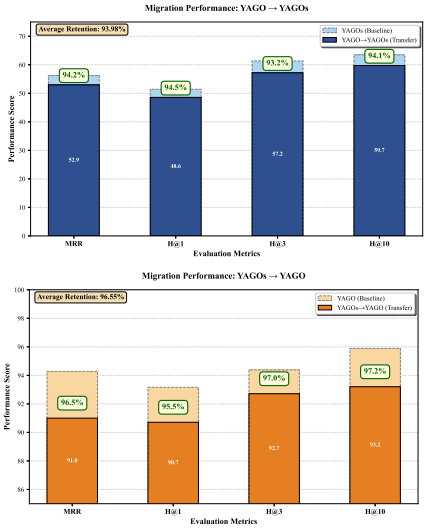

Fig. 2. Migration Performance (→ denotes the model migration).

contributions, revealing their synergistic optimization: DA provides dynamically focused reasoning evidence while HCPE ensures hierarchical logical consistency. The amplified effects on semantically complex WIKI validate that their coordinated design constitutes the core innovation of QHCP.

D. Inductive Generalization Validation

To validate the entity-independent inductive capability, we conduct cross-dataset transfer experiments following [16]. We use YAGO (20,541 entities) and YAGOs (10,623 entities), which form a matched pair with zero entity overlap but 10 identical relations, to test whether learned patterns can transfer across disjoint entity spaces.

We train models on source datasets and directly evaluate them on target datasets without fine-tuning. As shown in Fig. 2, QHCP maintains 93% performance retention across all

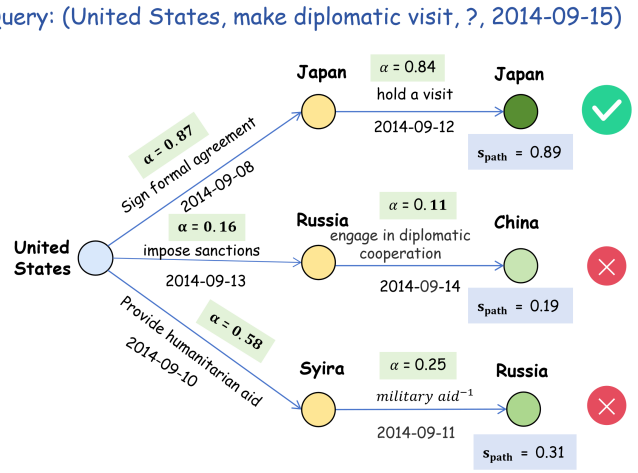

Fig. 3. A case of how QHCP generates explainable reasoning paths, where edge attention weights α and path similarities s_{path} are annotated.

transfer scenarios (ranging from 93.20% to 96.52%), demonstrating that our hierarchical path modeling captures relation-centric temporal patterns independent of entity identities, enabling deployment in dynamically evolving knowledge graphs.

E. Case Study

To illustrate how QHCP performs interpretable multi-hop reasoning, we examine the query $(\textit{United States}, \textit{make diplomatic visit}, ?, 2014\text{-}09\text{-}15)$ on ICEWS18. The path through Japan, *sign formal agreement* $\rightarrow$ *host a visit*, is assigned high edge weights $(0.87, 0.84)$ and the highest similarity (0.89), reflecting strong semantic relevance to the query. By contrast, although the path through China is temporally closer to the query timestamp, the edge *impose sanctions* receives a lower weight (0.16), showing that DA prioritizes semantic compatibility over temporal proximity. For the path via Syria, HCPE reduces the path similarity to 0.31 despite a moderate first-hop weight (0.58), because the relation chain introduces global logical inconsistency. This case illustrates the complementary roles of the two modules: DA performs hop-level semantic filtering, while HCPE evaluates global logical consistency over the full path.

V. Conclusion

In this work, we propose QHCP, a query-aware hierarchical contrastive framework for inductive temporal knowledge graph reasoning. By integrating dynamic attention and hierarchical contrastive supervision, QHCP achieves robust temporal-relational reasoning with strong inductive generalization. Extensive experiments on four benchmark datasets demonstrate consistent performance gains over state-of-the-art methods, especially on semantically complex graphs, while maintaining strong inductive generalization to unseen entities.

References

[1] W. Jin, M. Qu, X. Jin, and X. Ren, "Recurrent event network: Autoregressive structure inferenceover temporal knowledge graphs," in *Proceedings of the 2020 conference on empirical methods in natural language processing (EMNLP)*, 2020, pp. 6669–6683.

[2] F. Feng, X. He, X. Wang, C. Luo, Y. Liu, and T.-S. Chua, "Temporal relational ranking for stock prediction," *ACM Transactions on Information Systems (TOIS)*, vol. 37, no. 2, pp. 1–30, 2019.

[3] L. Cui, H. Seo, M. Tabar, F. Ma, S. Wang, and D. Lee, "Deterrent: Knowledge guided graph attention network for detecting healthcare misinformation," in *Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining*, 2020, pp. 492–502.

[4] Z. Li, X. Jin, W. Li, S. Guan, J. Guo, H. Shen, Y. Wang, and X. Cheng, "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*, 2021, pp. 408–417.

[5] J. Leblay and M. W. Chekol, "Deriving validity time in knowledge graph," in *Companion of the the Web Conference*, 2018, pp. 1771–1776.

[6] H. Dong, Z. Ning, P. Wang, Z. Qiao, P. Wang, Y. Zhou, and Y. Fu, "Adaptive path-memory network for temporal knowledge graph reasoning," in *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence (IJCAI-23)*, 2023, pp. 2086–2094.

[7] H. Dong, P. Wang, M. Xiao, Z. Ning, P. Wang, and Y. Zhou, "Temporal inductive path neural network for temporal knowledge graph reasoning," *Artificial Intelligence*, vol. 329, p. 104085, 2024.

[8] S. Zheng, H. Yin, T. Chen, Q. V. H. Nguyen, W. Chen, and L. Zhao, "Dream: Adaptive reinforcement learning based on attention mechanism for temporal knowledge graph reasoning," in *Proceedings of the 46th international ACM SIGIR conference on research and development in information retrieval*, 2023, pp. 1578–1588.

[9] L. Bai, L. Xu, and L. Zhu, "Attention-based complex logical query on temporal knowledge graph via graph neural network," *IEEE Transactions on Big Data*, 2024.

[10] Z. Han, P. Chen, Y. Ma, and V. Tresp, "Explainable subgraph reasoning for forecasting on temporal knowledge graphs," in *International conference on learning representations*, 2020.

[11] B. Yang, W.-t. Yih, X. He, J. Gao, and L. Deng, "Embedding entities and relations for learning and inference in knowledge bases," *arXiv preprint arXiv:1412.6575*, 2014.

[12] Z. Li, S. Guan, X. Jin, W. Peng, Y. Lyu, Y. Zhu, L. Bai, W. Li, J. Guo, and X. Cheng, "Complex evolutional pattern learning for temporal knowledge graph reasoning," *arXiv preprint arXiv:2203.07782*, 2022.

[13] T. Dettmers, P. Minervini, P. Stenetorp, and S. Riedel, "Convolutional 2d knowledge graph embeddings," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 32, no. 1, 2018.

[14] A. Garc'ıa-Dur'an, S. Dumančić, and M. Niepert, "Learning sequence encoders for temporal knowledge graph completion," in *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2018, pp. 4816–4821.

[15] R. Goel, S. M. Kazemi, M. Brubaker, and P. Poupart, "Diachronic embedding for temporal knowledge graph completion," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 34, no. 04, 2020, pp. 3988–3995.

[16] K. Teru, E. Denis, and W. Hamilton, "Inductive relation prediction by subgraph reasoning," in *International conference on machine learning*. PMLR, 2020, pp. 9448–9457.

[17] H. Sun, J. Zhong, Y. Ma, Z. Han, and K. He, "Timetraveler: Reinforcement learning for temporal knowledge graph forecasting," in *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2021, pp. 8306–8319.

[18] Y. Yang, C. Huang, L. Xia, and C. Li, "Knowledge graph contrastive learning for recommendation," in *Proceedings of the 45th international ACM SIGIR conference on research and development in information retrieval*, 2022, pp. 1434–1443.

[19] Y. Xu, J. Ou, H. Xu, and L. Fu, "Temporal knowledge graph reasoning with historical contrastive learning," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 37, no. 4, 2023, pp. 4765–4773.

[20] D. Xu, C. Ruan, E. Korpeoglu, S. Kumar, and K. Achan, "Inductive representation learning on temporal graphs," *arXiv preprint arXiv:2002.07962*, 2020.

[21] K. Leetaru and P. A. Schrodt, "Gdelt: Global data on events, location, and tone, 1979–2012," in *ISA annual convention*, vol. 2, no. 4. Citeseer, 2013, pp. 1–49.

[22] F. Mahdisoltani, J. Biega, and F. Suchanek, "Yago3: A knowledge base from multilingual wikipedias," in *Proceedings of the 7th Biennial Conference on Innovative Data Systems Research (CIDR)*, 2014.