Neuro-Symbolic Inductive Reasoning on Temporal Knowledge Graphs via LLM-Enhanced Rule Mining

Xinyue Fan[1], Rui Yu[1], and Yinglong Wang[1,2,*]
[1]Shandong Computer Science Center, Qilu University of Technology (Shandong Academy of Sciences)
[2]Key Laboratory of Computing Power Network and Information Security, Ministry of Education,
Qilu University of Technology (Shandong Academy of Sciences), Jinan, China
Email: {fnxnyn_yx@163.com, rui.yu1996@foxmail.com, wangyinglong@qlu.edu.cn (*Corresponding author)}

***Abstract*—Inductive reasoning on Temporal Knowledge Graphs (TKGs) is pivotal for generalizing to entities unseen during training. However, existing symbolic methods heavily rely on historical interactions, rendering them ineffective under the extreme data sparsity of new entities (the "cold-start" problem). To address this, we propose Temporal Hierarchical Adaptive Inductive Logic(T-HAIL). First, we introduce an Open-to-Closed Knowledge Injection (OC-KI) mechanism, which leverages the parametric causal knowledge of Large Language Models (LLMs) to generate candidate rules, compensating for the lack of statistical signals in sparse graphs. To bridge the vocabulary gap, we employ a Semantic Vector Alignment module via Sentence-BERT to map candidates to the KG schema. These rules are refined through a rigorous data-grounding mechanism—validating candidates against observed facts—to eliminate hallucinations. To fully exploit these grounded rules under sparsity, we introduce a Hierarchical Adaptive Inference (HAI) mechanism that dynamically transitions between Exact Rule Matching, Semantic Similarity Borrowing, and Topology-based Fallback. Extensive experiments on ICEWS and GDELT benchmarks demonstrate that T-HAIL significantly outperforms state-of-the-art baselines in mitigating sparsity issues. Our code and datasets are available at: https://github.com/fnxnynpro/T-HAIL.**

***Index Terms*—Temporal Knowledge Graph, Inductive Reasoning, Large Language Models, Neuro-Symbolic Learning, Data Sparsity.**

I. Introduction

Temporal Knowledge Graphs (TKGs) serve as a dynamic structural representation of facts, storing data as quadruples (s, r, o, t) to capture the evolution of relationships over time. TKG reasoning has demonstrated significant potential in a wide range of real-world applications, such as international political event forecasting (e.g., ICEWS, GDELT) and financial market analysis [33].

Most existing approaches, including embedding-based models (e.g., TransE [16], HyTE [17]) and Graph Neural Networks (e.g., RE-GCN [11], xERTE [12]), operate under a transductive setting. They assume that all entities and relations are observed during the training phase, learning fixed representations for a static set of nodes.

However, real-world graphs are inherently evolving, with new entities emerging continuously. This necessitates inductive reasoning, where the model must predict links for entities that were completely unseen during training [7], [26]. Transductive methods fail here as they treat new entities as out-of-vocabulary tokens lacking learned embeddings. A fundamental challenge in this setting is data sparsity, often referred to as the "cold-start" problem. Unlike transductive entities that benefit from rich historical interactions, new entities in the inductive phase often possess very limited context, making them topologically isolated.

Current research on inductive reasoning primarily explores two directions: subgraph-based representation learning and rule-based symbolic reasoning. Subgraph methods like GraIL [7] learn relation-independent structures but are often computationally expensive. On the other hand, symbolic approaches, typified by temporal random walk methods (e.g., TLogic [8]), rely on deriving and applying interpretable rules from historical paths. While effective in data-rich scenarios, these methods struggle significantly in inductive settings. The sparsity of new entities breaks the connectivity required to form valid path traversals, making it impossible to activate learned rules or mine local patterns. This leads to failures in both rule application and path-based inference.

To address these limitations, we propose Temporal Hierarchical Adaptive Inductive Logic(T-HAIL), an LLM-Enhanced Neuro-Symbolic Framework. Our methodology is driven by the insight that while statistical signals are weak in sparse graphs, the logical patterns of events (e.g., Visit often precedes Make Statement) are inherently captured in the massive parametric knowledge of LLMs [28]. We operationalize this via an Open-to-Closed Knowledge Injection (OC-KI) mechanism, where LLMs serve as a commonsense knowledge reservoir to compensate for the missing statistical evidence in sparse subgraphs. Furthermore, to handle the long-tail distribution of entity connectivity, we propose a Hierarchical Adaptive Inference (HAI) paradigm that dynamically falls back from exact rule matching to semantic similarity and topological connectivity, ensuring robust predictions even for isolated nodes.

Our main contributions are summarized as follows:

- We propose T-HAIL, a neuro-symbolic framework that overcomes the cold-start problem. Central to our approach is the Open-to-Closed Knowledge Injection (OC-KI) mechanism, which utilizes LLM priors to bridge

This work was supported by Project of Key R&D Program of Shandong Province, China (2025CXPT095), the Taishan Scholars Program (NO.tspd20240814), and the "20 Articles" Project for Universities in Jinan (202534093).

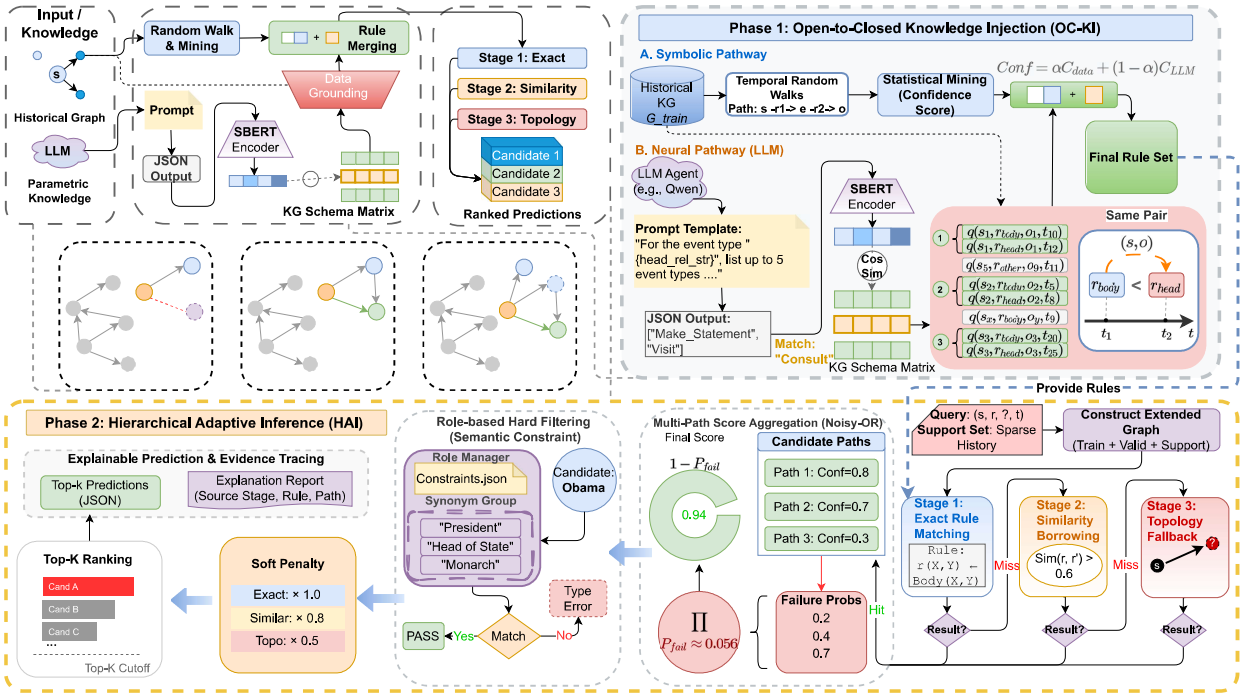

Fig. 1. Overview of the proposed T-HAIL Framework. The architecture consists of two distinct phases: (1) Open-to-Closed Knowledge Injection (OC-KI) (Phase 1), which integrates a symbolic pathway (statistical mining) and a neural pathway (LLM-driven rule expansion). A Semantic Alignment module (via Sentence-BERT) maps textual tokens to KG relations, followed by a Data Grounding module to ensure empirical support from the training graph (G_{train}). (2) Hierarchical Adaptive Inference (HAI) (Phase 2), designed for the inductive setting. The inference engine employs a hierarchical cascade—starting with Exact Rule Matching, falling back to Semantic Borrowing, and finally utilizing Topology Fallback—to robustly handle queries for sparse or isolated entities.

the gap between open-world semantics and closed-world structure via semantic vector alignment.

- We design a rigorous grounding mechanism to filter LLM-generated rules by validating semantic candidates against observed graph structures, thereby ensuring rule reliability and effectively mitigating hallucinations.
- We propose the Hierarchical Adaptive Inference (HAI) paradigm, which effectively handles data sparsity by exploiting semantic similarity and topological connectivity in a cascading manner.
- Extensive experiments on three inductive ICEWS benchmarks and the large-scale GDELT-ind dataset demonstrate that our approach significantly outperforms state-of-the-art baselines, proving its effectiveness in handling data sparsity.

II. Related Work

A. Temporal Knowledge Graph Reasoning

Reasoning on TKGs generally aims to predict missing links based on historical facts. Existing methods can be broadly categorized into embedding-based, sequence-based, and symbolic approaches. Embedding-based Models: Early works extended static methods (e.g., TransE [16]) to the temporal domain. For instance, HyTE [17] projects entities into time-specific hyperplanes, while TComplEx [18] utilizes tensor factorization to capture temporal evolution. Sequence-based Models: These approaches typically employ Recurrent Neural Networks (RNNs) or self-attention to model the evolution of graph snapshots. RE-NET [14] and CyGNet [13] use generative networks to model historical sequences. RE-GCN [11] and xERTE [12] further integrate Graph Convolutional Networks (GCNs) to capture structural dependencies at each timestamp. More recently, ODE-based models like TANGO [15] have been proposed to handle continuous time dynamics. Symbolic Approaches: Unlike black-box neural models, symbolic methods such as TLogic [8] mine temporal logic rules via random walks, offering transparent explanations. However, these methods rely heavily on the statistical recurrence of patterns, making them vulnerable in sparse data scenarios.

B. Inductive Representation Learning

Inductive reasoning aims to generalize to entities unseen during training. A prominent line of research utilizes Subgraph-based GNNs, such as GraIL [7], which learn relation-independent structural patterns of enclosing subgraphs rather than entity-specific embeddings. Recent path-based frameworks have also achieved state-of-the-art results. For example, TiRGN [5] combines local-global history for induction, while TECHS [2] utilizes temporal logical graph networks. Other advanced methods include THCN [4] and TaReT [3], which focus on topology-aware dynamic graphs. However, extracting subgraphs or computing all-pair paths for every query is computationally expensive. Furthermore, in sparse temporal graphs, the local structures of new entities are often too fragmented to provide sufficient signals for GNNs. Our

work offers a rule-based alternative that addresses sparsity through semantic knowledge transfer.

C. LLMs and Neuro-Symbolic Reasoning

LLMs like Llama 2 [30] have demonstrated remarkable capabilities in reasoning. Recent research explores combining LLMs with KGs in two paradigms: (1) LLM as Predictor, where LLMs directly output facts using Chain-of-Thought (CoT) prompting [27], [28]; and (2) LLM as Enhancer, where LLMs generate auxiliary knowledge. Frameworks like ChatRule [23] and Neural Multi-hop Reasoning [32] attempt to distill rules from LLMs. However, most existing works focus on static KGs or do not explicitly address the temporal inductive cold-start problem. Our framework differs by introducing a strict data-grounding mechanism and a semantic alignment strategy to robustly integrate LLM priors into the temporal reasoning process.

In essence, existing approaches face a dilemma in the inductive setting: symbolic and graph-based methods rely heavily on statistical signals (e.g., path recurrence) that are absent in sparse histories, while current neuro-symbolic methods often lack rigorous structural grounding or fail to address temporal dynamics. Our work bridges these gaps by synergizing the interpretability of symbolic reasoning with the semantic generalization of LLMs, resolving the "cold-start" challenge in temporal inductive reasoning.

III. Methodology

A. Problem Formulation

A TKG is formalized as a sequence of facts $\mathcal{F}$, where each fact is a quadruple (s, r, o, t) comprising a subject $s \in \mathcal{E}$, a relation $r \in \mathcal{R}$, an object $o \in \mathcal{E}$, and a timestamp $t \in \mathcal{T}$. We address the TKG reasoning task under the inductive setting. The dataset is split into two disjoint graphs: a training graph G_{train} associated with entity set $\mathcal{E}_{train}$, and an inference graph G_{ind} associated with $\mathcal{E}_{ind}$, satisfying $\mathcal{E}_{train} \cap \mathcal{E}_{ind} = \emptyset$. Crucially, for inductive reasoning, the inference data is partitioned into a Support Set $\mathcal{S}_{ind}$ (containing historical contexts for unseen entities where $t < t_{query}$) and a Query Set $\mathcal{Q}_{ind}$ (target facts to predict). The objective is to infer the missing object o for a query $(s, r, ?, t) \in \mathcal{Q}_{ind}$ by leveraging logical rules learned from G_{train} and the context provided in $\mathcal{S}_{ind}$.

B. Phase I: Open-to-Closed Knowledge Injection(OC-KI)

To overcome the scarcity of statistical signals, we propose the OC-KI mechanism, which bridges the gap between open-world LLM priors and closed-world structural constraints.

1) Symbolic Rule Mining: We define a temporal logical rule ρ as a conjunction of atoms implying a head relation r_{head}:

$$\rho : P(s, o, t) \leftarrow r_1(s, e_1, t_1) \wedge \cdots \wedge r_n(e_{n-1}, o, t_n), \quad (1)$$

where P corresponds to the head relation r_{head}, e_i represent intermediate entities, and temporal constraints $t_1 \leq t_2 \leq \cdots \leq t$ apply. The confidence of a rule mined from G_{train} is calculated statistically:

$$Conf_{sym}(\rho) = \frac{supp(\rho)}{supp(body(\rho)) + \mu}, \quad (2)$$

where $supp(\cdot)$ denotes the number of path occurrences in the graph, and μ is a smoothing factor.

2) LLM-Driven Rule Expansion: To overcome data sparsity, we treat LLMs as a generator of parametric knowledge. For a specific relation r, we construct a prompt $\mathcal{P}(r)$ to query the LLM for potential prerequisite events (e.g., "What event typically precedes a 'Consult' relation?"). This yields a set of candidate rule bodies $\mathcal{B}_{LLM}$.

To bridge the gap between the LLM's open-ended vocabulary and the KG's fixed schema $\mathcal{R}$, we utilize a Semantic Vector Alignment module. Instead of rigid lexical matching, we encode both the LLM-generated tokens and the KG relation names into high-dimensional embeddings using a pre-trained Sentence-BERT model [1]. The candidate relation is mapped to the standard KG relation with the highest cosine similarity, provided this similarity score exceeds a predefined threshold θ (set to 0.6 empirically). This threshold acts as a semantic filter, ensuring that we capture nuances (e.g., Talk vs. Consult) without introducing irrelevant noise.

3) Structural Grounding and Fusion: LLM-generated rules may contain hallucinations. We enforce a rigorous Structural Grounding mechanism to validate these mapped candidates against G_{train}. A candidate rule ρ_{llm} is retained only if its grounding support in the training graph exceeds a threshold τ. Selecting τ balances validity and sparsity: a minimal τ (e.g., 2) filters hallucinations while preserving rare valid patterns, whereas high τ regresses to standard statistical mining.

For the rules that survive this grounding filter, we calculate a final confidence score by fusing empirical evidence with the LLM's semantic prior:

$$Conf_{final}(\rho) = \alpha \cdot Conf_{sym}(\rho) + (1-\alpha) \cdot Conf_{LLM}(\rho), \quad (3)$$

where $Conf_{LLM}(\rho)$ is the semantic alignment score derived in Phase I-2, and α is a hyperparameter balancing the weight between statistical observations and semantic priors. This results in an enriched rule set $\mathbf{\Omega}$ that covers sparse relations while maintaining structural validity.

C. Phase II: Hierarchical Adaptive Inference (HAI)

In the inductive phase, entity connectivity follows a long-tail distribution, causing static inference strategies to fail on "isolated" nodes. We propose the HAI paradigm, which dynamically selects the inference strategy based on rule availability and context matching. The hierarchy is designed based on the Confidence-Coverage Trade-off: we prioritize Exact Symbolic Matching (Level 1) for its high precision, then relax constraints to Semantic Similarity (Level 2) to handle vocabulary mismatches, and finally utilize Topology-based Fallback (Level 3) to ensure maximum recall for isolated nodes. The inference process cascades through these three levels:

1) Level 1: Exact Symbolic Matching: For a query $(s, r_q, ?, t_q)$, the engine first attempts to apply rules where the head relation matches the query relation exactly. The candidate set for this level is:

$$\mathcal{C}_{exact} = \{o \mid \exists \rho \in \mathbf{\Omega}, head(\rho) = r_q \wedge body(\rho) \subseteq \mathcal{S}_{ind}\}. \quad (4)$$

TABLE I
OVERALL PERFORMANCE COMPARISON ON INDUCTIVE DATASETS. BEST RESULTS ARE HIGHLIGHTED IN **BOLD**, AND THE SECOND BEST ARE UNDERLINED.

Model	ICEWS14-ind			ICEWS18-ind			ICEWS05-15-ind			GDELT-ind		
	MRR	H@1	H@10	MRR	H@1	H@10	MRR	H@1	H@10	MRR	H@1	H@10
Embedding-based / RNN-based Baselines												
RE-NET [14]	0.298	0.205	0.463	0.176	0.112	0.298	0.315	0.221	0.485	0.154	0.088	0.275
CyGNet [13]	0.284	0.196	0.445	0.165	0.103	0.284	0.302	0.210	0.468	0.142	0.079	0.256
TANGO [15]	0.312	0.221	0.486	0.185	0.124	0.315	0.334	0.238	0.512	0.165	0.095	0.288
xERTE [12]	0.345	0.252	0.510	0.205	0.135	0.342	0.362	0.268	0.535	0.182	0.108	0.315
RE-GCN [11]	0.338	0.245	0.505	0.201	0.132	0.335	0.355	0.258	0.528	0.178	0.102	0.308
Symbolic / Path-based Baselines												
TITer [9]	0.375	0.282	0.545	0.225	0.145	0.365	0.382	0.285	0.552	0.195	0.115	0.335
TLogic [8]	0.368	0.271	0.532	0.218	0.138	0.352	0.371	0.275	0.541	0.188	0.108	0.322
T-HAIL (Ours)	**0.472**	**0.385**	**0.615**	**0.305**	**0.221**	**0.452**	**0.485**	**0.392**	**0.621**	**0.242**	**0.185**	**0.425**

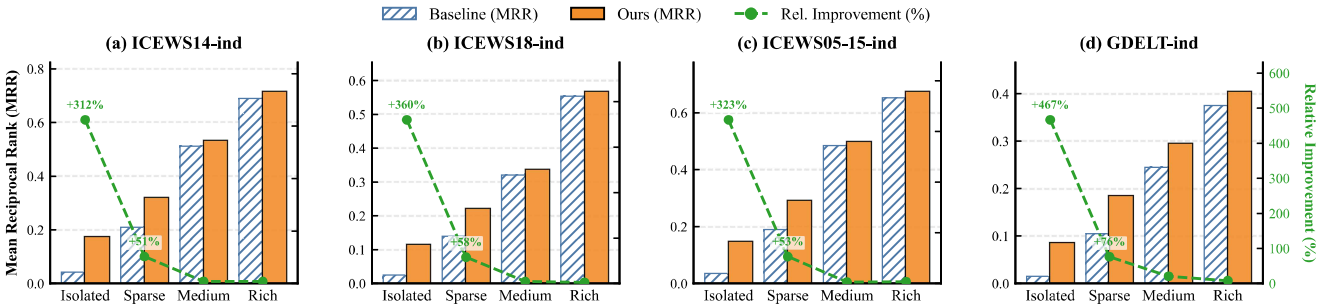

Fig. 2. Performance breakdown across different entity sparsity levels on three datasets. Bars represent the absolute MRR (Left Axis), while the green dashed line indicates the relative improvement over the baseline (Right Axis). Our framework consistently achieves the highest gains in the "Isolated" and "Sparse" groups, effectively mitigating the cold-start problem.

TABLE II
ABLATION STUDY OF MODEL VARIANTS (METRIC: MRR). WE REPORT RESULTS ON THE OVERALL (OV.) TEST SET AND THE ISOLATED (ISO.) SUBSET.

Variant	IC14		IC18		IC05		GDELT	
	Ov.	Iso.	Ov.	Iso.	Ov.	Iso.	Ov.	Iso.
Full Model	**0.472**	**0.173**	**0.305**	**0.115**	**0.485**	**0.148**	**0.242**	**0.085**
w/o LLM	0.415	0.115	0.262	0.075	0.435	0.105	0.197	0.052
w/o Topo	0.443	0.012	0.281	0.008	0.460	0.010	0.212	0.002
w/o Sim	0.451	0.154	0.292	0.102	0.470	0.135	0.225	0.070

This level provides the highest precision but relies on the existence of specific historical patterns.

2) Level 2: Semantic Similarity Borrowing: If $\mathcal{C}_{exact} = \emptyset$, the engine relaxes the constraint by identifying relations $\mathcal{R}_{sim}$ that are semantically similar to r_q. Unlike previous symbolic methods requiring exact matches or naive lexical heuristics (e.g., Levenshtein distance) that fail on synonyms, we employ a pre-trained Sentence-BERT encoder to map relation names into high-dimensional embeddings $\mathbf{v}_r$. The similarity is computed as the cosine similarity:

$$Sim(r_q, r') = \frac{\mathbf{v}_{r_q} \cdot \mathbf{v}_{r'}}{\|\mathbf{v}_{r_q}\| \|\mathbf{v}_{r'}\|} > \theta. \tag{5}$$

This allows the model to capture latent semantic associations (e.g., linking Assault to Conflict) that rigid lexical matching would miss. Rules belonging to these similar relations are "borrowed" to infer the target:

$$\begin{aligned} \mathcal{C}_{sim} = \{o \mid & \exists r' \in \mathcal{R}_{sim}, \exists \rho \in \mathbf{\Omega}, \\ & head(\rho) = r' \wedge body(\rho) \subseteq \mathcal{S}_{ind}\}. \end{aligned} \tag{6}$$

A penalty factor $\beta_{sim} \in (0, 1)$ is applied to the scores of these candidates to reflect the increased uncertainty introduced by semantic relaxation, ensuring that borrowed predictions act as a conservative supplement rather than dominating the results.

3) Level 3: Topology-Based Fallback: For extremely sparse entities (isolated nodes), even semantic borrowing may fail due to the lack of matching body paths. In this fallback level, we utilize any available historical edge connected to s in $\mathcal{S}_{ind}$ to activate compatible rules, prioritizing topological connectivity over semantic precision.

$$\mathcal{C}_{top} = \{o \mid \exists \rho \in \mathbf{\Omega}, body(\rho) \subseteq \mathcal{S}_{ind} \text{ connected to } s\}. \tag{7}$$

This serves as a "dead-end rescue" mechanism, ensuring the model yields predictions rather than null outputs. A stricter penalty $\beta_{top} < \beta_{sim}$ is applied.

4) Candidate Scoring: The final score for a candidate object e is aggregated using a Noisy-OR formulation to combine contributions from multiple paths:

$$Score(e) = 1 - \prod_i (1 - Conf_{final}(\rho_i) \cdot \beta_{level}), \quad (8)$$

where $\beta_{level} \in \{1, \beta_{sim}, \beta_{top}\}$ depends on the active inference level.

IV. Experiments

A. Experimental Setup

1) Datasets: Since standard TKG datasets are predominantly transductive, we constructed three inductive benchmarks (ICEWS14-ind, ICEWS18-ind, ICEWS05-15-ind) derived from the original ICEWS data. Additionally, we include GDELT-ind, a large-scale dataset derived from the Global Database of Events, Language, and Tone. Compared to the ICEWS datasets, GDELT is characterized by significantly higher noise and sparsity, allowing us to verify the robustness of our framework in challenging environments. Detailed statistics and construction protocols are provided in our open-source repository.

To simulate a realistic inductive setting, we strictly follow an entity-disjoint split:

- Transductive Training: The model learns logical rules on G_{train}, which contains a subset of entities.
- Inductive Inference: The model is evaluated on G_{test}, which involves a completely disjoint set of unseen entities. Crucially, we utilize the corresponding Support Set to provide limited historical contexts for these new entities.

2) Baselines and Metrics: We compare our framework against state-of-the-art baselines, including TLogic and TiTer. For fair comparison, all baselines were retrained and tuned on our constructed inductive datasets. We also compare against a strict version of TLogic (without similarity or topology fallback) to represent the symbolic reasoning lower bound, explicitly quantifying the gains from our HAI mechanism and LLM-enhanced rules under sparse conditions. We report Mean Reciprocal Rank (MRR), which calculates the average inverse rank of the correct answer, and Hits@k ($k = 1, 3, 10$), which measures the proportion of test triples where the correct entity appears in the top-k predictions.

B. Main Results

Table I presents the performance comparison across three datasets. Our LLM-Enhanced Neuro-Symbolic Framework (Ours) consistently outperforms the symbolic baseline.

The results indicate a robust improvement across all datasets. Notably, on the challenging GDELT-ind, where baselines generally underperform due to high noise (MRR $\approx$ 0.19), our model achieves a substantial boost (MRR 0.242), representing a relative improvement of approximately 29% against TLogic. Specifically, our method achieves significant gains in Hits@10 across all benchmarks. This improvement is primarily attributed to our HAI paradigm. As substantiated by the ablation study in Table II, disabling the Topology Fallback leads to a performance collapse for "Isolated" entities (e.g., MRR drops to near zero on GDELT-ind), empirically confirming that the hierarchical cascade is essential for recalling valid answers when exact rules fail. In inductive settings, many entities are sparsely connected, causing traditional rule matching to fail. By activating the Topology-based Fallback mechanism, our model successfully retrieves answers for queries that yield null predictions in the baseline, thereby boosting recall. Furthermore, the improvements in MRR and Hits@1 demonstrate that the LLM-Enhanced Rules, refined by our grounding mechanism, provide high-quality logical patterns that pure statistical mining fails to capture from sparse data.

C. Ablation Study and Stratified Analysis

To verify the effectiveness of each component, we conducted an ablation study. More importantly, to understand where the improvements come from, we stratified the test queries based on the sparsity of the entities involved (number of support edges). Specifically, we categorize entities into four groups: "Isolated" (0 edges), "Sparse" (1–3 edges), "Medium" (4–10 edges), and "Rich" ($>$ 10 edges). This fine-grained breakdown allows us to pinpoint the specific contribution of our framework in cold-start scenarios (Isolated/Sparse) versus data-rich scenarios.

1) Impact of Components: We compared the full model with three variants: (1) w/o LLM, which removes LLM-generated rules; (2) w/o Topology Fallback, which disables the topology-based fallback stage; and (3) w/o Similarity, which removes the Sentence-BERT module and reverts to rigid exact matching.

2) Detailed Breakdown: Table II provides a detailed numerical comparison across all three datasets.

- Effectiveness on Cold-Start: As shown in Table II, removing the Topology Fallback leads to a drastic decline in the Isolated group performance across all datasets (e.g., on ICEWS14-ind, MRR drops from 0.173 to 0.012). This confirms that structural priors are essential when valid rule paths are broken in sparse subgraphs.
- Quality of LLM Knowledge: Removing LLM Rules results in a significant drop in overall MRR (e.g., 0.472 $\rightarrow$ 0.415 on ICEWS14-ind), indicating that statistical rules alone are insufficient to cover the diverse relations in the inductive split, and LLM priors effectively supplement the rule set.
- Impact of Semantic Alignment: Removing the similarity module results in a consistent performance drop (e.g., 0.472 $\rightarrow$ 0.451 on ICEWS14-ind). This empirically validates that semantic vector alignment is crucial for bridging the vocabulary gap between open-ended LLM outputs and the fixed KG schema.

D. Parameter Sensitivity and Hyperparameter Analysis

We investigate the model's sensitivity to three key hyperparameters. **Fusion Weight (α):** This parameter balances symbolic statistics and LLM priors (Eq. 3). Our empirical results demonstrate a bell-shaped performance curve, peaking optimally at $\alpha \approx 0.65$. Extreme values ($\alpha \to 0$ or $\alpha \to 1$) significantly degrade performance, confirming that combining both knowledge sources is superior to using either in isolation. **Grounding Threshold (τ):** As defined in Sec. III-B, τ filters rules based on integer support counts. We identify $\tau = 2$ as the optimal configuration. A stricter threshold (e.g., $\tau \geq 5$) filters out valid but rare patterns for isolated entities, regressing to the baseline's cold-start failure, whereas $\tau = 1$ introduces excessive noise from hallucinations. **Semantic Threshold (θ):** Set to $\theta = 0.6$ via grid search on the validation set, this value acts as a critical semantic filter. Lower values ($\theta < 0.5$) improve recall but map unrelated LLM tokens to KG relations (noise injection), while higher values ($\theta > 0.8$) fail to capture valid synonyms, such as matching "Talk" to "Consult".

V. Conclusion

In this paper, we addressed the critical challenge of data sparsity in inductive TKG. We proposed an LLM-Enhanced Neuro-Symbolic Framework that effectively bridges the gap between symbolic logic and neural semantic knowledge. By leveraging LLMs as rule generators and enforcing a rigorous data-grounding mechanism, we successfully enriched the rule set for sparse relations without introducing hallucinations. Furthermore, our novel HAI engine allows the model to dynamically utilize semantic similarity and topological connectivity, ensuring robust predictions even for entities with minimal historical context.

Experimental results on three ICEWS datasets and GDELT-ind demonstrate that our approach significantly outperforms the state-of-the-art symbolic baseline, TLogic. This superiority is particularly evident in recalling missing links for isolated entities within sparse and noisy environments like GDELT. For future work, we plan to explore the applicability of our framework to other dynamic graph domains such as financial transaction networks.

References

[1] N. Reimers and I. Gurevych, "Sentence-BERT: Sentence embeddings using Siamese BERT-networks," in *EMNLP*, 2019, pp. 3982–3992.

[2] Q. Lin, J. Liu, R. Mao, F. Xu, and E. Cambria, "TECHS: Temporal logical graph networks for explainable extrapolation reasoning," in *ACL*, 2023, pp. 1281–1293.

[3] J. Ma et al., "TaReT: Temporal knowledge graph reasoning based on topology-aware dynamic relation graph," *Inf. Process. Manag.*, vol. 61, no. 6, p. 103848, 2024.

[4] T. Chen et al., "THCN: A Hawkes Process-Based Temporal Causal Convolutional Network for Extrapolation Reasoning in Temporal Knowledge Graphs," *IEEE Trans. Knowl. Data Eng.*, vol. 36, no. 12, pp. 9374–9387, 2024.

[5] Y. Li, S. Sun, and J. Zhao, "TiRGN: Time-guided recurrent graph network with local-global historical patterns for temporal knowledge graph reasoning," in *IJCAI*, 2022, pp. 2152–2158.

[6] J. Zhang, B. Hui, C. Mu, and L. Tian, "Learning multi-graph structure for temporal knowledge graph reasoning," *Expert Syst. Appl.*, vol. 255, p. 124561, 2024.

[7] K. K. Teru, E. G. Denis, and W. L. Hamilton, "Inductive relation prediction by subgraph reasoning," in *ICML*, 2020, pp. 9448–9457.

[8] Y. Liu, M. Hildebrandt, B. Blumenstein, M. Joblin, and V. Tresp, "TLogic: Temporal logical rules for explainable link forecasting on temporal knowledge graphs," in *AAAI*, 2022, pp. 5801–5809.

[9] H. Sun, J. Zhong, Y. Ma, Z. Han, and K. He, "TimeTraveler: Reinforcement learning for temporal knowledge graph forecasting," in *EMNLP*, 2021, pp. 8306–8319.

[10] C. Meilicke, M. W. Chekol, D. Ruffinelli, and H. Stuckenschmidt, "Anytime bottom-up rule learning for knowledge graph completion," in *IJCAI*, 2019, pp. 3137–3143.

[11] Z. Li et al., "Temporal knowledge graph reasoning based on evolutional representation learning," in *SIGIR*, 2021, pp. 408–417.

[12] Z. Han, P. Chen, Y. Ma, and V. Tresp, "Explainable subgraph reasoning for forecasting on temporal knowledge graphs," in *ICLR*, 2021.

[13] C. Zhu, M. Chen, C. Fan, G. Cheng, and Y. Zhang, "Learning from history: Modeling temporal knowledge graphs with sequential copy-generation networks," in *AAAI*, vol. 35, no. 5, 2021, pp. 4732–4740.

[14] W. Jin, M. Qu, X. Jin, and X. Ren, "Recurrent event network: Autoregressive structure inference over temporal knowledge graphs," in *Findings of EMNLP*, 2020, pp. 6669–6683.

[15] Z. Han, Z. Ding, Y. Ma, Y. Gu, and V. Tresp, "Learning neural ordinary equations for forecasting future links on temporal knowledge graphs," in *EMNLP*, 2021, pp. 8352–8364.

[16] A. Bordes, N. Usunier, A. Garcia-Duran, J. Weston, and O. Yakhnenko, "Translating embeddings for modeling multi-relational data," in *NeurIPS*, 2013, pp. 2787–2795.

[17] S. S. Dasgupta, S. N. Ray, and P. Talukdar, "HyTE: Hyperplane-based temporally aware knowledge graph embedding," in *EMNLP*, 2018, pp. 2001–2011.

[18] T. Lacroix, G. Obozinski, and N. Usunier, "Tensor decompositions for temporal knowledge base completion," in *ICLR*, 2020.

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

[20] M. Zhang, Y. Xia, Q. Liu, S. Wu, and L. Wang, "Learning long- and short-term representations for temporal knowledge graph reasoning," in *WWW*, 2023, pp. 177–187.

[21] Z. Li et al., "HiSMatch: Historical structure matching based temporal knowledge graph reasoning," in *Findings of EMNLP*, 2022, pp. 7328–7338.

[22] S. Pan, L. Luo, Y. Wang, C. Chen, J. Wang, and X. Wu, "Unifying large language models and knowledge graphs: A roadmap," *IEEE Trans. Knowl. Data Eng.*, vol. 36, no. 7, pp. 3580–3599, 2024.

[23] L. Luo, J. Ju, B. Xiong, Y.-F. Li, G. Haffari, and S. Pan, "ChatRule: Mining logical rules with large language models for knowledge graph reasoning," in *PAKDD*, 2025, pp. 314–325.

[24] L. Luo, Y.-F. Li, R. Haffari, and S. Pan, "Reasoning on graphs: Faithful and interpretable large language model reasoning," in *ICLR*, 2024.

[25] A. Garcez and L. Lamb, "Neurosymbolic AI: The 3rd wave," *Artif. Intell. Rev.*, vol. 56, pp. 12387–12406, 2023.

[26] M. Galkin, X. Yuan, H. Mostafa, J. Tang, and Z. Zhu, "Towards foundation models for knowledge graph reasoning," in *ICLR*, 2024.

[27] Y. Huang, "LLM-ARK: Knowledge graph reasoning using large language models via deep reinforcement learning," *arXiv preprint arXiv:2312.11282*, 2023.

[28] J. Wei et al., "Chain-of-thought prompting elicits reasoning in large language models," in *NeurIPS*, vol. 35, 2022, pp. 24824–24837.

[29] J. Sun et al., "Think-on-Graph: Deep and responsible reasoning of large language models with knowledge graphs," in *ICLR*, 2024.

[30] H. Touvron et al., "Llama 2: Open foundation and fine-tuned chat models," *arXiv preprint arXiv:2307.09288*, 2023.

[31] W. Xu, B. Liu, M. Peng, X. Jia, and M. Peng, "Pre-trained language model with prompts for temporal knowledge graph completion," in *Findings of ACL*, 2023, pp. 7790–7803.

[32] Y. Liu, M. Hildebrandt, M. Joblin, M. Ringsquandl, R. Raissouni, and V. Tresp, "Neural multi-hop reasoning with logical rules on biomedical knowledge graphs," in *ESWC*, 2021, pp. 375–391.

[33] X. V. Li and F. S. Passino, "FinDKG: Dynamic knowledge graphs with large language models for detecting global trends in financial markets," in *ICAIF*, 2024, pp. 573–581.