

DyHist: Query-Aware Historical Dynamics Modeling for Temporal Knowledge Graph Extrapolation

Peiyuan Jiang¹, Yao Liu^{1,*}, Yubo Wen¹, Yanglei Gan², Lu Liu¹, Ruojie Zhang¹
Jiaye Yang¹, Yuewei Xue¹, Yajiao Deng¹, Xiaoyu Guo¹, Qiao Liu¹

¹University of Electronic Science and Technology of China, Chengdu, China

²Southwest Minzu University, Chengdu, China

*Correspondence: liuyao@uestc.edu.cn

Abstract

Temporal Knowledge Graph (TKG) extrapolation aims to predict future facts from observed historical events. Existing methods usually achieve this goal by jointly modeling local historical evolution and long-range historical contexts. First, local evolution modeling often summarizes local history into an overall evolving representation, where query-relevant local deviations may not be sufficiently exploited as useful predictive features. Second, existing long-range history modeling often treats historical recurrence as a stable cue, while overlooking that its reliability may vary across relations and recent contexts. To address these issues, we propose **DyHist**, a query-aware historical dynamics modeling framework for TKG extrapolation. DyHist captures local dynamics by separating recent evidence from medium-range context and explicitly modeling their query-relevant deviations as predictive features. It further uses a relation-aware historical prior to calibrate candidate scores with time-decayed support, recent trends, and relation-specific strength, enabling recurrence signals to contribute differently across relations and recent contexts. Experiments on four benchmarks show that DyHist achieves the best MRR against recent baselines. Code is available at <https://github.com/PANPANKK/DyHist>.

1 Introduction

Knowledge Graphs (KGs) represent facts as triples (s, r, o) , but many real-world facts evolve over time (Sun et al., 2020; Wang et al., 2021; Xu et al., 2023). Temporal Knowledge Graphs (TKGs) therefore extend triples into timestamped facts (s, r, o, t) , enabling reasoning over time-dependent facts (Zhu et al., 2021). TKG reasoning includes interpolation over observed timestamps and extrapolation to future timestamps. This paper focuses on extrapolation, which is important for early warning and decision-making (Li et al., 2021a).

Accurate TKG extrapolation relies on deeply exploiting historical facts. Given a future query $(s, r, ?, t)$, the model needs to infer the missing object by exploiting historical dependencies among entities and relations. Existing methods have improved this task by modeling both short-term local evolution and long-range historical recurrence (Li et al., 2022a; Chen et al., 2024c; Tang et al., 2025). For short-term local history modeling, existing methods typically encode adjacent historical snapshots with GNN-based spatial encoders and recurrent or attention-based temporal modules to capture short-range structural evolution. Beyond adjacent-snapshot encoding, recent methods further enhance local history modeling with query-related paths, subgraphs, neighborhoods, or high-order structural associations (Han et al., 2021; Chen et al., 2025; Zhang et al., 2023; Ahrabian et al., 2026). For long-range history modeling, existing methods exploit distant historical facts mainly from two perspectives. Some methods extract explicit recurrence signals from historical candidates, occurrence records, or frequency statistics, while others encode query-related global contexts to capture implicit long-range dependencies (Zhu et al., 2021; Xu et al., 2023; Ye et al., 2026). Despite these advances, existing methods still have two limitations in historical information utilization. **First**, local history is usually encoded into a unified evolving representation through graph-based message passing and recurrent temporal aggregation. Although this captures the overall temporal evolution, the difference between earlier context and recent evidence is usually only implicitly fused into the evolving state, rather than being explicitly modeled as a predictive signal. As illustrated in Figure 1, Oman mainly appears in earlier local snapshots, whereas Djibouti becomes more prominent in recent snapshots. Compressing the whole local sequence into a unified representation may obscure such query-relevant shifts between earlier context and recent evidence.

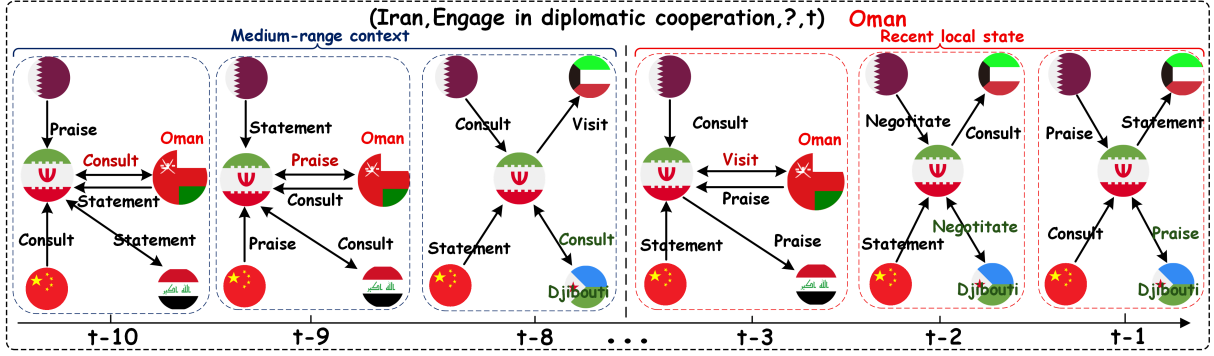


Figure 1: Illustration of local historical evidence for the query (Iran, Engage in diplomatic cooperation, $?, t$). Red links denote Oman as medium-range contextual evidence, while green links denote Djibouti as more recent evidence near the query time.

Second, existing long-range history modeling often exploits repeated co-occurrence as recurrence evidence, while underexploring how its reliability may vary across relations and recent contexts. For relatively stable relations such as *Cooperate*, recurrence may provide useful cues due to persistent diplomatic or economic ties. For more event-driven relations such as *Accuse* or *Make statement*, however, candidate entities may shift with recent events, making recurrence less consistently reliable.

To address these issues, we propose **DyHist**, a query-aware framework for modeling historical dynamics in TKG extrapolation. DyHist models historical dynamics from both local and long-range perspectives. For local history, DyHist uses query-aware snapshot descriptors to adaptively select non-overlapping recent evidence and medium-range context. The selected evidence is fused into a broader local historical context, which is then compared with the latest short-term evidence. DyHist explicitly models their deviation through fast-slow residual injection with a learned gate, allowing query-relevant recent changes to complement the overall local representation. For long-range history, DyHist introduces a relation-aware historical prior to calibrate candidate scores. For each candidate entity, the prior aggregates previous occurrences of the complete fact (s, r, o) with temporal decay, measures whether its historical support has recently increased, and scales the resulting recurrence evidence with a learnable relation-specific strength. The main contributions of this paper are as follows:

- We propose DyHist, a query-aware historical dynamics modeling framework for TKG extrapolation that jointly models recent evidence, medium-range context, and long-range historical recur-

rence for more accurate future fact prediction.

- We introduce query-aware dual-window selection and fast-slow residual injection for local history modeling. The model adaptively selects non-overlapping query-relevant recent and medium-range evidence and explicitly captures the deviation between the latest short-term evidence and the broader local historical context.
- We introduce a relation-aware dynamic historical prior that calibrates candidate scores using time-decayed complete-fact recurrence, recent support trends, and learnable relation-specific strength. Extensive experiments on four datasets demonstrate the effectiveness of DyHist.

2 Related Work

2.1 TKG Extrapolation

Temporal Knowledge Graph (TKG) extrapolation aims to predict future facts from observed historical facts. A common paradigm is dynamic representation learning, which encodes historical events or temporal snapshots to learn evolving entity and relation representations. Representative methods include Know-Evolve, which models temporal facts with point processes (Trivedi et al., 2017), RE-NET, which captures historical event sequences and concurrent structures (Jin et al., 2020), and RE-GCN, which combines graph neural networks with recurrent evolution over snapshots (Li et al., 2021b).

Recent studies further improve TKG extrapolation from diverse perspectives. TLogic performs interpretable reasoning using temporal logical rules (Liu et al., 2022). DiffuTKG and NADEX model future entities or events from a generative

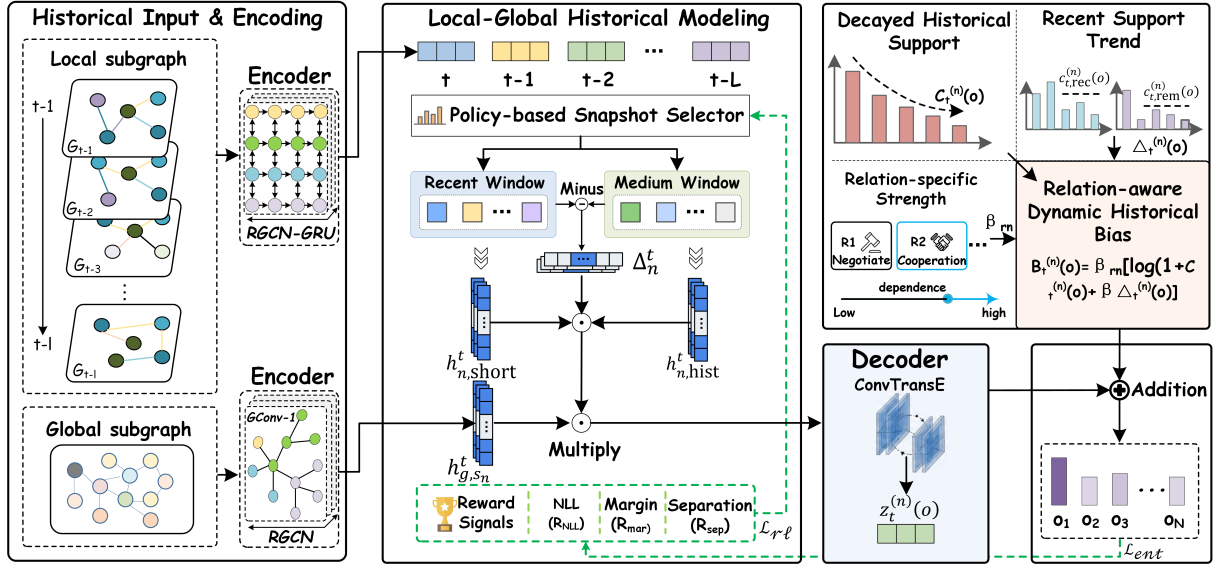


Figure 2: Overall architecture of DyHist. DyHist models query-aware local dynamics through dual-window selection and fast-slow residual injection, and calibrates decoder logits with a relation-aware dynamic historical prior.

perspective (Cai et al., 2024; Gan et al., 2026). HTKGE focuses on temporal scoring and decoder design (Yang et al., 2024), while AnRe incorporates textual semantics and textualized historical contexts for temporal reasoning (Tang et al., 2025). Although these methods differ in modeling paradigms, organizing and exploiting historical information remains central to TKG extrapolation.

2.2 Historical Information Modeling

Historical information modeling is a key component of TKG extrapolation. Existing approaches can be broadly grouped into local-history modeling and long-range history modeling. Local-history modeling focuses on facts close to the query timestamp, aiming to capture recent entity states and short-term structural evolution. Existing strategies include adjacent-snapshot encoding, temporal-neighborhood aggregation, query-related path reasoning, local-subgraph construction, and causal-subhistory modeling. Representative methods include RE-NET, RE-GCN, xERTE, TITer, CSI, and CognTKE (Jin et al., 2020; Li et al., 2021b; Han et al., 2021; Sun et al., 2021; Chen et al., 2024a, 2025). These methods provide informative local evidence for future prediction. However, recent signals and earlier contextual information are often aggregated into a unified representation, making query-relevant local changes difficult to model explicitly. Long-range history modeling exploits distant historical facts to capture repeated patterns and long-range dependencies. One line introduces

explicit recurrence signals, such as repeated facts, historical candidates, occurrence records, or frequency statistics (Zhu et al., 2021; Xu et al., 2023; Li et al., 2022a; Wang et al., 2026). Another line encodes global historical facts, query-related subgraphs, or retrieved long-range contexts into latent representations for reasoning (Chen et al., 2024c; Ye et al., 2026; Chen et al., 2025; Tang et al., 2025). These studies show that long-range history provides complementary evidence beyond recent snapshots. Nevertheless, recurrence evidence is often incorporated without explicitly calibrating its reliability across relations and recent temporal contexts.

In contrast to existing approaches, DyHist explicitly distinguishes recent evidence, medium-range local context, and long-range historical recurrence. It uses query-aware dual-window selection and fast-slow residual injection to model query-relevant local changes, while a relation-aware historical prior calibrates the influence of recurring historical patterns across relations and recent contexts.

3 Method

3.1 Overview

As shown in Figure 2, DyHist targets TKG extrapolation, which predicts the missing object in a future query $(s_n, r_n, ?, t)$ from snapshots before t . Given $\{\mathcal{G}_1, \dots, \mathcal{G}_{t-1}\}$, DyHist uses a local sequence $\mathcal{H}_t^{\text{loc}} = \{\mathcal{G}_{t-L}, \dots, \mathcal{G}_{t-1}\}$ to model recent evolution and a sampled global subgraph $\mathcal{G}_t^{\text{glob}}$ to capture long-range evidence. The local sequence

and global subgraph are encoded by RGCN-GRU and RGCN encoders, respectively. Let $\mathbf{e}_{s_n}^0$ and $\mathbf{e}_{r_n}^0$ denote the learnable base embeddings of subject entity s_n and relation r_n . DyHist constructs the query representation as $\mathbf{q}_n = W_q[\mathbf{e}_{s_n}^0 \parallel \mathbf{e}_{r_n}^0]$, where W_q is a learnable projection matrix. Using query-aware snapshot descriptors, DyHist selects recent and medium windows from the local sequence, fuses the selected evidence into a local context, and enhances it with fast-slow residual injection. The enhanced local representation is then combined with the global representation for scoring, while a relation-aware dynamic historical prior calibrates logits using recurrence evidence. Main notation is summarized in Appendix A.

3.2 Historical Representation Encoding

For each local snapshot $\mathcal{G}_\tau$, DyHist updates local entity states by combining RGCN-based structural encoding with recurrent transition:

$$\mathbf{H}_l^\tau = \text{GRU}(\text{RGCN}_{\text{loc}}(\mathcal{G}_\tau, \mathbf{H}_l^{\tau-1}), \mathbf{H}_l^{\tau-1}). \quad (1)$$

Here, $\mathbf{H}_l^\tau \in \mathbb{R}^{|\mathcal{E}| \times d}$ denotes the local entity-state matrix at timestamp τ . In addition, DyHist encodes the sampled global subgraph to obtain a global historical representation:

$$\mathbf{H}_g^t = \text{RGCN}_{\text{glob}}(\mathcal{G}_t^{\text{glob}}), \quad (2)$$

where $\mathbf{H}_g^t$ denotes the query-time global historical representation for decoding.

3.3 Query-aware Local Dynamics Modeling

Snapshot descriptor construction. For each candidate snapshot $\mathcal{G}_\tau$, DyHist constructs a compact descriptor $\mathbf{d}_\tau = [u_\tau, \ell_\tau, \delta_\tau, \alpha_\tau, m_\tau, m_\tau(1 - \ell_\tau)]$, where $u_\tau, \ell_\tau, \delta_\tau, \alpha_\tau$, and m_τ are scalar features characterizing query relevance, temporal lag, snapshot variation, latest-state consistency, and relation-context relevance, respectively. The last term highlights recent relation-aligned snapshots. Detailed definitions are provided in Appendix D.1.

Policy-based dual-window selection. Let $\mathcal{I}_t^{\text{loc}} = \{t - L, \dots, t - 1\}$ denote the local candidate snapshot indices. Based on the descriptor set $\mathcal{D}_t = \{\mathbf{d}_\tau \mid \tau \in \mathcal{I}_t^{\text{loc}}\}$, DyHist samples recent and medium window index sets:

$$\mathcal{I}_t^R \sim \pi_R(\cdot \mid \mathcal{D}_t), \quad \mathcal{I}_t^M \sim \pi_M(\cdot \mid \mathcal{D}_t), \quad (3)$$

where π_R and π_M are role-specific policies over candidate snapshot indices. The recent window

$\mathcal{I}_t^R$ is selected first, after which the medium window $\mathcal{I}_t^M$ is selected from $\mathcal{I}_t^{\text{loc}} \setminus \mathcal{I}_t^R$, ensuring non-overlapping temporal evidence. The recent window captures near-query dynamics, while the medium window provides earlier contextual evidence.

Dual-window aggregation. Given the query representation $\mathbf{q}_n$, DyHist separately aggregates the encoded subject states in each selected window:

$$\mathbf{h}_n^X = \text{Agg}_X(\{\mathbf{H}_l^\tau[s_n]\}_{\tau \in \mathcal{I}_t^X}; \mathbf{q}_n). \quad (4)$$

Here, X denotes either the recent window R or the medium window M , and Agg_X is a query-guided temporal aggregator. Its implementation is detailed in Appendix D.2. The two window representations are fused through a vector gate:

$$\begin{aligned} \mathbf{g}_n &= \sigma(W_g[\mathbf{h}_n^R \parallel \mathbf{h}_n^M]), \\ \mathbf{h}_n^L &= \mathbf{g}_n \odot \mathbf{h}_n^R + (1 - \mathbf{g}_n) \odot \mathbf{h}_n^M. \end{aligned} \quad (5)$$

where $\sigma(\cdot)$ is the sigmoid function and $\mathbf{g}_n$ is a vector gate that balances near-query dynamic evidence and medium-range contextual evidence.

Fast-slow residual injection. To explicitly capture query-relevant local changes, DyHist compares the latest short-term evidence with the fused local historical context:

$$\Delta_n^t = \mathbf{h}_{n,\text{short}}^t - \mathbf{h}_{n,\text{hist}}^t, \quad \mathbf{h}_{n,\text{hist}}^t = \mathbf{h}_n^L. \quad (6)$$

Here, $\mathbf{h}_{n,\text{short}}^t$ is obtained by applying the same query-guided aggregator to the most recent subset $\mathcal{I}_t^S \subseteq \mathcal{I}_t^R$, with details in Appendix D.2. The residual Δ_n^t characterizes the deviation between near-query evidence and the broader local historical context. The final local subject representation is obtained by gated residual injection:

$$\mathbf{h}_{l,s_n}^t = \mathbf{h}_{n,\text{hist}}^t + \sigma_g(\mathbf{h}_{n,\text{hist}}^t, \Delta_n^t) \odot \Delta_n^t, \quad (7)$$

where $\sigma_g(\cdot)$ denotes a sigmoid-based vector gate.

3.4 Dynamic Historical Prior

For query $(s_n, r_n, ?, t)$, DyHist first combines local and global historical representations for decoding:

$$\mathbf{h}_{s_n,t}^{\text{dec}} = \lambda \mathbf{h}_{l,s_n}^t + (1 - \lambda) \mathbf{H}_g^t[s_n], \quad (8)$$

where λ balances local and global historical information. The decoder then produces neural logits over all candidate objects using the context-aware relation representation $\mathbf{e}_{r_n,t}$:

$$\mathbf{x}_{n,t} = f_{\text{dec}}(\mathbf{h}_{s_n,t}^{\text{dec}}, \mathbf{e}_{r_n,t}), \mathbf{z}_t^{(n)} = \mathbf{x}_{n,t}(\mathbf{H}_l^{t-1})^\top. \quad (9)$$

Here, $\mathbf{H}_l^{t-1}$ serves as the candidate object representation bank, and $\mathbf{z}_t^{(n)}[o]$ denotes the neural logit of candidate object o . To calibrate candidate logits with recurrence evidence, DyHist first computes time-decayed support for each candidate:

$$c_t^{(n)}(o) = \sum_{\tau < t} e^{-\gamma(t-\tau)} \mathbb{I}[(s_n, r_n, o) \in \mathcal{G}_\tau]. \quad (10)$$

Let $\mathcal{I}_t^{\text{rem}} = \mathcal{I}_t^{\text{loc}} \setminus \mathcal{I}_t^R$ denote the remaining local indices. Recent support changes are captured by

$$\eta_t^{(n)}(o) = \text{ReLU} \left(\frac{c_{t,\text{rec}}^{(n)}(o)}{|\mathcal{I}_t^R|} - \frac{c_{t,\text{rem}}^{(n)}(o)}{|\mathcal{I}_t^{\text{rem}}|} \right), \quad (11)$$

where $c_{t,\text{rec}}^{(n)}(o)$ and $c_{t,\text{rem}}^{(n)}(o)$ count support in $\mathcal{I}_t^R$ and $\mathcal{I}_t^{\text{rem}}$, respectively. A learnable relation-specific strength β_r controls the prior reliability. The relation-aware historical bias is computed as

$$b_t^{(n)}(o) = \beta_{r_n} \left[\log(1 + c_t^{(n)}(o)) + \rho \eta_t^{(n)}(o) \right], \quad (12)$$

where β_{r_n} controls the recurrence strength of relation r_n and ρ controls the trend contribution. Finally, the calibrated logit is

$$\hat{z}_t^{(n)}(o) = z_t^{(n)}(o) + b_t^{(n)}(o). \quad (13)$$

3.5 Training Objective

Entity prediction loss. For the target timestamp t with N_t facts $\{(s_n, r_n, o_n, t)\}_{n=1}^{N_t}$, DyHist minimizes the object prediction negative log-likelihood:

$$\mathcal{L}_{\text{ent}} = -\frac{1}{N_t} \sum_{n=1}^{N_t} \log p(o_n | s_n, r_n, t). \quad (14)$$

Window selection loss. The dual-window policy is optimized with prediction-oriented reward feedback and temporal-role separation:

$$\mathcal{R}_t = \lambda_m \mathcal{R}_{\text{mar}} + \lambda_n \mathcal{R}_{\text{nll}} + \mathcal{R}_{\text{sep}}, \quad (15)$$

where $\mathcal{R}_{\text{mar}}$ measures the prediction margin, $\mathcal{R}_{\text{nll}} = -\mathcal{L}_{\text{ent}}$ provides loss-based feedback, and $\mathcal{R}_{\text{sep}}$ encourages distinct temporal roles. Given the joint window-selection action a_t and descriptor set $\mathcal{D}_t$, the policy-gradient loss is

$$\mathcal{L}_{\text{pg}} = -(\mathcal{R}_t - \bar{b}_t) \log \pi(a_t | \mathcal{D}_t) - \lambda_H \mathcal{H}(\pi), \quad (16)$$

where $\bar{b}_t$ is the reward baseline and $\mathcal{H}(\pi)$ is the entropy regularizer. Detailed reward definitions are given in Appendix D.3.

Overall objective. The final objective is

$$\mathcal{L} = \mathcal{L}_{\text{ent}} + \lambda_{\text{pg}} \mathcal{L}_{\text{pg}}, \quad (17)$$

where λ_{pg} balances prediction and policy losses.

4 Experiments

4.1 Experimental Setups

Datasets and Evaluation Metrics. We evaluate DyHist on four standard TKG extrapolation benchmarks: ICEWS14, ICEWS05-15, ICEWS18, and GDELT. The ICEWS datasets (Boschee et al., 2015) are derived from the Integrated Crisis Early Warning System, and GDELT (Leetaru and Schrodt, 2013) is collected from the Global Database of Events, Language, and Tone. Following previous studies (Chen et al., 2024c), we use chronological splits and report MRR and Hits@{1, 3, 10} under the time-aware filtered setting. Dataset statistics are provided in Appendix B.

Baseline Models. We compare DyHist with representative methods from three categories. Static methods include DistMult (Yang et al., 2015), ConvE (Dettmers et al., 2018), and RotatE (Sun et al., 2019). Interpolation methods include TTransE (García-Durán et al., 2018), TA-DistMult (García-Durán et al., 2018), and DE-SimpleE (Goel et al., 2020). Extrapolation methods include RE-NET (Jin et al., 2020), RE-GCN (Li et al., 2021b), CyGNet (Zhu et al., 2021), TiRGN (Li et al., 2022a), HisMatch (Li et al., 2022b), RE-TIA (Liu et al., 2023), CENET (Xu et al., 2023), CRAFT (Zhang et al., 2024), THCN (Chen et al., 2024b), LogCL (Chen et al., 2024c), CognTKE (Chen et al., 2025), NADEX (Gan et al., 2026), and MIDFA (Wang et al., 2026). Baseline results are taken from their original papers under the same evaluation setting whenever available. Baseline details are given in Appendix C.

Implementation Details. We train DyHist with Adam for up to 80 epochs and select the checkpoint with the best validation MRR for testing. The entity and relation embedding dimensions are both set to 200, and the dropout rate is 0.2. We use an RGCN-based temporal graph encoder and ConvTransE decoder. The local history length is 10, with both recent and medium-range window sizes set to 3, and the fast window size set to 2. The relation-aware historical prior is applied at the logit level, and the policy loss weight is set to 0.01. Training is conducted at the timestamp level, and DyHist

Table 1: Overall performance (%) on four datasets in terms of MRR and Hits@1/3/10 under the time-aware filtered setting. Gray cells indicate the best results, and light gray cells indicate the second-best.

Models	ICEWS14				ICEWS18				ICEWS05-15				GDELT			
	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	15.44	10.91	17.24	23.92	11.51	7.03	12.87	20.86	17.95	13.12	20.71	29.32	8.68	5.58	9.96	17.13
ConvE	35.09	25.23	39.38	54.68	24.51	16.23	29.25	44.51	33.81	24.78	39.00	54.95	16.55	11.02	18.88	31.60
RotatE	21.31	10.26	24.35	44.75	12.78	4.01	14.89	31.91	24.71	13.22	29.04	48.16	13.45	6.95	14.09	25.99
TTransE	13.72	2.98	17.70	35.74	8.31	1.92	8.56	21.89	15.57	4.80	19.24	38.29	5.50	0.47	4.94	15.25
TA-DistMult	25.80	16.94	29.74	42.99	16.75	8.61	18.41	33.59	24.31	14.58	27.92	44.21	12.00	5.76	12.94	23.54
DE-Simple	33.36	24.85	37.15	48.92	19.30	11.53	21.86	34.80	35.02	25.91	38.99	52.75	19.70	12.22	21.39	33.70
RE-NET	36.93	26.83	39.51	54.78	28.81	19.05	32.44	47.51	43.32	33.43	47.77	63.06	19.62	12.42	21.00	34.01
RE-GCN	40.39	30.66	44.96	59.21	30.58	21.01	34.34	48.75	48.03	37.33	53.85	68.27	19.64	12.42	20.90	33.69
CyGNet	35.05	25.73	39.01	53.55	24.93	15.90	28.28	42.61	36.81	26.61	41.63	56.22	18.48	11.52	19.57	31.98
TiRGN	44.04	33.83	48.95	63.84	33.66	23.19	37.99	54.22	50.04	39.25	56.13	70.71	21.67	13.63	23.27	37.60
HisMatch	46.42	35.91	51.63	66.84	33.99	23.91	37.90	53.94	52.85	42.01	59.05	73.28	22.01	14.45	23.80	36.61
RETIA	42.76	32.28	47.77	62.75	32.43	22.23	36.48	52.94	47.26	36.64	52.90	67.76	20.12	12.76	21.45	34.49
CENET	39.02	29.62	43.23	57.49	27.85	18.15	31.63	46.98	41.95	32.17	46.93	60.43	20.23	12.69	21.70	34.92
CRAFT	45.71	35.05	51.83	65.21	34.21	23.96	38.53	54.11	50.14	39.56	56.18	70.09	23.78	15.38	26.23	40.15
THCN	45.39	36.58	50.84	66.07	35.63	24.90	39.26	56.76	51.94	40.32	57.79	72.18	23.46	15.18	25.21	39.03
LogCL	48.87	37.76	54.71	70.26	35.67	24.53	40.32	57.74	57.07	46.07	63.72	77.87	23.75	14.54	25.60	42.33
CognTKE	46.06	36.49	51.11	64.49	35.24	25.21	39.93	54.71	53.13	42.62	59.42	72.70	—	—	—	—
NADEx	49.03	39.45	57.48	70.55	36.84	27.58	45.12	60.58	52.17	43.38	60.75	71.93	23.67	16.10	28.35	43.05
MIDFA	51.74	40.93	57.73	72.40	37.80	26.63	42.68	59.88	58.55	47.51	65.56	79.08	25.18	16.69	27.89	44.53
Ours	54.03	43.25	60.01	74.49	40.30	28.85	45.59	62.59	61.07	50.46	67.67	81.10	25.53	16.16	28.08	45.19

experiments are conducted on a single RTX 4090 GPU. Additional implementation details are provided in Appendix B.

4.2 Main Results

Table 1 reports overall performance on four benchmark datasets. We highlight three observations. First, temporal extrapolation methods generally outperform static and interpolation-based baselines, showing the importance of historical modeling for future fact prediction. Second, DyHist achieves the highest MRR on all four datasets and obtains competitive Hits@K results. Third, compared with recent extrapolation baselines such as LogCL, NADEx, and MIDFA, DyHist shows consistent MRR gains across datasets. These results suggest that distinguishing different types of historical evidence is beneficial. Unlike baselines that often aggregate history uniformly, DyHist separates query-relevant recent signals from medium-range context and calibrates candidate scores with a relation-aware historical prior. This enables adaptive evidence selection and history-aware score adjustment for TKG extrapolation. On the larger and noisier GDELT dataset, DyHist still achieves the best MRR and Hits@10, indicating its effectiveness under complex event streams.

4.3 Ablation Study

To examine the contribution of each component, we conduct ablation studies on ICEWS14 and

ICEWS05-15. We consider six variants: w/o QADW removes window selection and fast-slow residual injection, and instead performs query-aware aggregation over all local snapshots in $\mathcal{H}_t^{\text{loc}}$; w/o PG removes the policy-gradient loss for window selection; w/o FS removes the fast-slow residual injection; w/o Prior removes the relation-aware dynamic historical prior; w/o Decay removes time-decayed recurrence support; and w/o Trend removes the recent support trend.

The ablation results illustrate the contributions of DyHist’s components. For local history modeling, the drop of w/o QADW suggests the overall benefit of explicit window selection and fast-slow local dynamics modeling. The declines of w/o PG and w/o FS further indicate the contributions of reward-driven selection and fast-slow residual injection within this module. Among the long-range variants, w/o Prior shows the largest MRR drop, while the drops of w/o Decay and w/o Trend indicate the usefulness of decayed and trend-based recurrence signals. Overall, these findings suggest that DyHist benefits from adaptive local evidence modeling and dynamic historical prior calibration.

4.4 Analysis of Local History Modeling

We evaluate the query-aware dual-window selection (QADW) against four strategies. Latest- K , Random- K , and Top- K are single-window methods, which select snapshots by recency, random sampling, and deterministic descriptor ranking, re-

Table 2: Component ablation results on ICEWS14 and ICEWS05-15. Bold numbers denote the best results.

Variant	ICEWS14			ICEWS05-15		
	MRR	H@1	H@10	MRR	H@1	H@10
Ours	54.03	43.25	74.49	61.07	50.46	81.10
w/o QADW	51.76	40.75	72.88	58.71	48.35	79.61
w/o PG	52.94	41.43	73.80	59.48	49.76	80.74
w/o FS	52.36	41.22	73.68	59.14	49.62	80.50
w/o Prior	52.09	41.13	73.27	59.97	48.91	80.71
w/o Decay	53.49	42.57	74.18	60.84	50.24	80.66
w/o Trend	53.79	42.99	74.25	60.77	50.25	80.54

Table 3: Effectiveness and efficiency of local history selection strategies. Time and Params denote inference time and parameter size, respectively.

Strategy	ICEWS14			ICEWS05-15		
	MRR	Time	Params	MRR	Time	Params
Latest- K	52.16	10.9s	9.15M	59.24	2m39s	10.22M
Random- K	51.27	11.1s	9.15M	58.15	2m40s	10.22M
Top- K	52.34	11.5s	9.23M	60.72	2m43s	10.30M
DFW	53.35	11.1s	9.15M	60.46	2m37s	10.22M
QADW	54.03	11.2s	9.23M	61.07	2m41s	10.30M

spectively. DFW is a fixed dual-window method, which uses $\{t-1, t-2, t-3\}$ as the recent window and $\{t-4, t-5, t-6\}$ as the medium-range window, without query-aware selection.

As shown in Table 3, QADW achieves the best MRR on both datasets. Compared with Latest- K and Random- K , the gains indicate that simple temporal or random choices are insufficient. Top- K shows the usefulness of query-aware snapshot ranking, while DFW shows the benefit and limitation of fixed recent/medium partitions. QADW combines query-aware selection with adaptive dual-window roles, leading to better local history modeling with only about 0.08M additional parameters and similar inference time.

Query-aware descriptor analysis. We further examine descriptor features in the QADW selector. Time-only removes query relevance and relation-context alignment, keeping only temporal distance and snapshot-level dynamics. Q-shuf. randomly shuffles query representations across timestamp batches, and R-shuf. shuffles relation information to disrupt its alignment with the current query relations. As shown in Table 4, the full descriptor setting achieves the best results on both datasets. The competitive but lower performance of Time-only suggests that temporal and structural cues are useful but incomplete. The drops of Q-shuf. and R-shuf. further indicate that preserving query- and relation-specific correspondence provides addi-

tional guidance for adaptive local history selection.

Effect of fast-slow residual injection. We examine whether the deviation between near-query evidence and the dual-window historical context provides useful local variation signals. We compare four variants. Short-only uses only the most recent short-term representation, testing whether near-query evidence alone is sufficient. No residual uses the fused recent-medium historical context without residual injection, testing whether the dual-window context alone is sufficient. Add residual directly adds the fast-slow residual to the historical context, testing whether the deviation between short-term evidence and historical context is useful. Gated residual injects the residual through an adaptive gate, testing whether such deviations should be introduced selectively. As shown in Table 5, Short-only performs worse than No residual, indicating that relying only on near-query evidence is insufficient. The improvement of Add residual over No residual shows that fast-slow deviations contain local variation signals. Gated residual achieves slightly better overall performance, suggesting a modest benefit from adaptive residual injection.

Table 4: Analysis of descriptor features under the QADW selector.

Input	ICEWS14			ICEWS05-15		
	MRR	H@1	H@10	MRR	H@1	H@10
Time-only	52.87	41.95	73.71	60.64	50.02	80.67
Q-shuf.	53.18	42.32	73.65	60.15	49.94	80.31
R-shuf.	53.24	42.58	73.79	60.42	50.15	80.40
Full	54.03	43.25	74.49	61.07	50.46	81.10

Table 5: Analysis of fast-slow residual injection.

Variant	ICEWS14			ICEWS05-15		
	MRR	H@1	H@10	MRR	H@1	H@10
Short-only	51.51	40.27	73.42	60.00	49.41	80.06
No residual	52.36	40.92	73.68	60.66	50.01	80.74
Add residual	53.91	42.91	74.66	60.84	50.24	80.74
Gated residual	54.03	43.25	74.49	61.07	50.46	81.10

4.5 Repeated and Novel Queries

We split test queries into repeated and novel subsets according to whether the ground-truth object has appeared with the same subject-relation pair in history. Repeated queries evaluate historical recurrence, while novel queries test behavior without direct recurrence. We compare DyHist with Latest, which selects the latest- K snapshots without query-aware descriptors, and **w/o Prior**, which removes the relation-aware historical prior. As shown

Table 6: Results on repeated and novel query subsets.

Type	Model	ICEWS14			ICEWS05-15		
		MRR	H@1	H@10	MRR	H@1	H@10
Rep.	Latest	73.67	62.88	92.62	73.26	63.09	91.43
	w/o Prior	70.52	59.69	90.44	72.09	61.30	91.44
	DyHist	73.78	63.70	91.84	73.73	63.48	92.05
Nov.	Latest	30.92	19.62	53.63	32.45	21.41	55.17
	w/o Prior	31.83	20.73	54.94	33.69	22.28	57.42
	DyHist	32.33	20.76	55.43	33.81	22.43	57.49

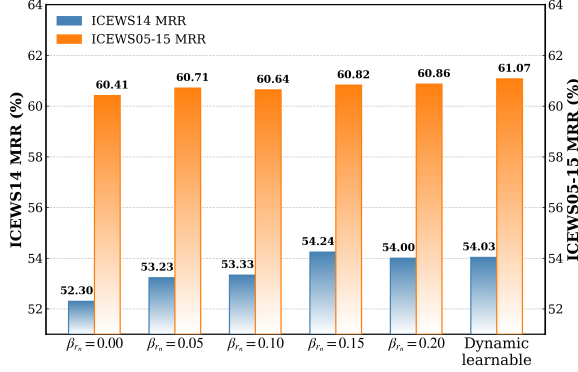


Figure 3: MRR comparison of fixed global and relation-specific learnable recurrence strengths.

in Table 6, repeated queries perform better than novel ones, confirming recurrence as a strong signal. DyHist improves over w/o Prior on repeated queries, showing that the dynamic prior helps calibrate recurrence patterns. On novel queries, DyHist slightly outperforms Latest, suggesting that query-aware descriptors can identify useful local evidence. Overall, DyHist benefits from recurrence calibration on repeated queries and query-aware local evidence on novel queries.

4.6 Analysis of Dynamic Historical Prior

Figure 3 compares fixed global prior strengths with the relation-aware learnable prior. The fixed variants use one shared scalar for all relations, while our variant learns relation-specific strengths $\{\beta_r\}$. Although a tuned fixed value such as $\beta = 0.15$ performs best on ICEWS14, the relation-aware prior achieves the best result on ICEWS05-15, suggesting that learnable relation-specific strengths offer a more adaptive alternative to manually choosing a global coefficient. We further group relations by history coverage, defined as the proportion of test queries $(s, r, ?, t)$ whose subject–relation pair (s, r) appeared before. This analysis examines whether the learned relation-specific strength β_r is related to the recurrence level of different relations. As shown in Table 7, β_r varies moderately across cov-

erage groups. On ICEWS14, it increases from low- to high-coverage relations, suggesting that relations with stronger historical recurrence tend to receive larger prior strengths. On ICEWS05-15, the trend is less monotonic, indicating that coverage alone does not fully determine the learned strength. Overall, these results suggest that the learned prior provides relation-specific adjustments beyond a uniform recurrence strength.

Table 7: Learned prior strengths under relation-level history coverage groups.

Dataset	Group	#Rel.	Avg. Cov.	Avg. β_r
ICEWS14	Low	21	49.640	0.134
ICEWS14	Medium	20	76.010	0.136
ICEWS14	High	21	89.430	0.149
ICEWS05-15	Low	42	55.770	0.134
ICEWS05-15	Medium	41	78.290	0.151
ICEWS05-15	High	41	91.080	0.143

4.7 Computational Efficiency

As shown in Figure 4, DyHist achieves the highest MRR among the compared models on both datasets. Its inference time is close to LogCL on ICEWS14 and lower than CognTKE and MIDFA on ICEWS05-15. Although DyHist uses slightly more parameters than LogCL, it remains smaller than MIDFA. The additional cost mainly comes from query-aware window selection and candidate-level prior calibration. Since the historical prior operates at the logit level and does not introduce extra deep decoding layers, the overall inference time remains comparable to efficient baselines.

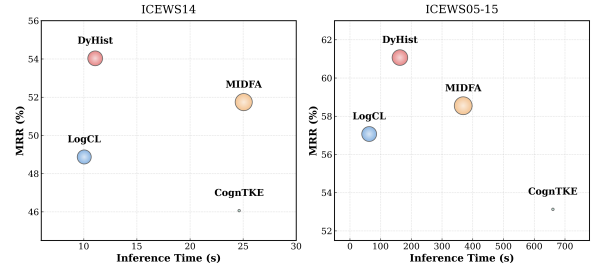


Figure 4: Effectiveness–efficiency trade-off on two datasets. Marker size denotes parameter count.

5 Conclusion

In this paper, we proposed DyHist, a query-aware historical dynamics modeling framework for temporal knowledge graph extrapolation. DyHist models historical evidence from two complementary

perspectives: query-relevant local changes and relation-dependent historical recurrence. It separates recent dynamic evidence from medium-range contextual evidence through adaptive dual-window selection and fast-slow residual modeling, and calibrates recurrence signals with a relation-aware dynamic historical prior. Experiments on four benchmarks support the usefulness of the proposed historical evidence modeling. The results suggest that TKG extrapolation requires more than aggregating historical facts: models should also estimate when historical evidence is reliable for the current query. Historical recurrence provides useful signals, while query-aware local evidence remains important when direct recurrence is weak.

6 Limitations

One limitation of this work lies in its reliance on four widely used TKG benchmarks, namely ICEWS14, ICEWS05-15, ICEWS18, and GDELT. Although standard, these datasets are mainly dominated by political and international event records, with limited diversity in domains such as science, economics, public health, or natural disasters. Therefore, the generalizability of DyHist to other types of temporal knowledge graphs remains to be further verified. In addition, DyHist explicitly exploits historical recurrence and recent support trends, which may be less effective for sparse and emerging events where direct recurrence is limited. Moreover, because event-centric TKGs may contain reporting biases, recurrence-based calibration could amplify frequently observed entities or relations if predictions are directly used in sensitive decision-making scenarios. In future work, we will investigate more effective modeling of such scenarios, where future facts rely more heavily on indirect temporal evidence.

7 Ethical Considerations

DyHist is evaluated on public temporal knowledge graph benchmarks and is intended solely for research on temporal reasoning and knowledge graph forecasting. Since event-centric datasets such as ICEWS and GDELT may contain reporting biases and uneven event coverage, predictions from the model should not be directly used for high-stakes political, security, or humanitarian decision-making without appropriate human oversight and domain-specific validation. The recurrence-based prior may amplify historically frequent entities

or relations, and future real-world applications should therefore include explicit bias assessment and uncertainty-aware evaluation.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (U22B2061), the National Key R&D Program of China (2022YFB4300603), and the Natural Science Foundation of Sichuan, China (2026NSFSC0429).

References

- Kian Ahrabian, Eric Boxer, and Jay Pujara. 2026. Toward better temporal structures for geopolitical events forecasting. *arXiv preprint arXiv:2601.00430*.
- Elizabeth Boschee, Jennifer Lautenschlager, Sean O’Brien, Steve Shellman, James Starz, and Michael Ward. 2015. [ICEWS coded event data](#). Harvard Dataverse.
- Yuxiang Cai, Qiao Liu, Yanglei Gan, Changlin Li, Xueyi Liu, Run Lin, Da Luo, and Jiaye Yang. 2024. Predicting the unpredictable: uncertainty-aware reasoning over temporal knowledge graphs via diffusion process. In *Findings of the Association for Computational Linguistics: ACL 2024*, pages 5766–5778.
- Kai Chen, Ye Wang, Xin Song, Siwei Chen, Han Yu, and Aiping Li. 2024a. Temporal knowledge graph extrapolation via causal subhistory identification. In *IJCAI*, pages 3298–3306.
- Tingxuan Chen, Jun Long, Zidong Wang, Shuai Luo, Jincai Huang, and Liu Yang. 2024b. Thcn: A hawkes process based temporal causal convolutional network for extrapolation reasoning in temporal knowledge graphs. *IEEE Transactions on Knowledge and Data Engineering*, 36(12):9374–9387.
- Wei Chen, Huaiyu Wan, Yuting Wu, Shuyuan Zhao, Jiayaqi Cheng, Yuxin Li, and Youfang Lin. 2024c. Local-global history-aware contrastive learning for temporal knowledge graph reasoning. In *2024 IEEE 40th International Conference on Data Engineering (ICDE)*, pages 733–746. IEEE.
- Wei Chen, Yuting Wu, Shuhan Wu, Zhiyu Zhang, Mengqi Liao, Youfang Lin, and Huaiyu Wan. 2025. Cognkte: A cognitive temporal knowledge extrapolation framework. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pages 14815–14823.
- Tim Dettmers, Pasquale Minervini, Pontus Stenetorp, and Sebastian Riedel. 2018. Convolutional 2d knowledge graph embeddings. In *Proceedings of the AAAI conference on artificial intelligence*, volume 32.
- Yanglei Gan, Peng He, Yuxiang Cai, Run Lin, Guanyu Zhou, and Qiao Liu. 2026. [Negative-aware diffusion](#)

- process for temporal knowledge graph extrapolation. In *Findings of the Association for Computational Linguistics: EACL 2026*, pages 3352–3367, Rabat, Morocco. Association for Computational Linguistics.
- Alberto García-Durán, 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*, pages 4816–4821.
- Rishab Goel, Seyed Mehran Kazemi, Marcus Brubaker, and Pascal Poupert. 2020. Diachronic embedding for temporal knowledge graph completion. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, pages 3988–3995.
- Zhen Han, Peng Chen, Yunpu Ma, and Volker Tresp. 2021. [Explainable subgraph reasoning for forecasting on temporal knowledge graphs](#). In *International Conference on Learning Representations*.
- Woojeong Jin, Meng Qu, Xisen Jin, and Xiang Ren. 2020. Recurrent event network: Autoregressive structure inference over temporal knowledge graphs. In *Proceedings of the 2020 conference on empirical methods in natural language processing (EMNLP)*, pages 6669–6683.
- Kalev Leetaru and Philip A. Schrod. 2013. GDELT: Global data on events, location, and tone. In *Proceedings of the International Studies Association Annual Convention*, pages 1–49.
- Yujia Li, Shiliang Sun, and Jing Zhao. 2022a. Tirgn: Time-guided recurrent graph network with local-global historical patterns for temporal knowledge graph reasoning. In *Ijcai*, pages 2152–2158.
- Zixuan Li, Zhongni Hou, Saiping Guan, Xiaolong Jin, Weihua Peng, Long Bai, Yajuan Lyu, Wei Li, Jiafeng Guo, and Xueqi Cheng. 2022b. Hismatch: Historical structure matching based temporal knowledge graph reasoning. In *Findings of the association for computational linguistics: EMNLP 2022*, pages 7328–7338.
- Zixuan Li, Xiaolong Jin, Saiping Guan, Wei Li, Jiafeng Guo, Yuanzhuo Wang, and Xueqi Cheng. 2021a. [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)*, pages 4732–4743, Online. Association for Computational Linguistics.
- Zixuan Li, Xiaolong Jin, Wei Li, Saiping Guan, Jiafeng Guo, Huawei Shen, Yuanzhuo Wang, and Xueqi Cheng. 2021b. Temporal knowledge graph reasoning based on evolutionary representation learning. In *Proceedings of the 44th international ACM SIGIR conference on research and development in information retrieval*, pages 408–417.
- Kangzheng Liu, Feng Zhao, Guandong Xu, Xianzhi Wang, and Hai Jin. 2023. Retia: relation-entity twin-interact aggregation for temporal knowledge graph extrapolation. In *2023 IEEE 39th international conference on data engineering (ICDE)*, pages 1761–1774. IEEE.
- Yushan Liu, Yunpu Ma, Marcel Hildebrandt, Mitchell Joblin, and Volker Tresp. 2022. Tlogic: Temporal logical rules for explainable link forecasting on temporal knowledge graphs. In *Proceedings of the AAAI conference on artificial intelligence*, volume 36, pages 4120–4127.
- Haohai Sun, Jialun Zhong, Yunpu Ma, Zhen Han, and Kun He. 2021. Timetraveler: Reinforcement learning for temporal knowledge graph forecasting. In *Proceedings of the 2021 conference on empirical methods in natural language processing*, pages 8306–8319.
- Tianxiang Sun, Yunfan Shao, Xipeng Qiu, Qipeng Guo, Yaru Hu, Xuan-Jing Huang, and Zheng Zhang. 2020. Colake: Contextualized language and knowledge embedding. In *Proceedings of the 28th international conference on computational linguistics*, pages 3660–3670.
- Zhiqing Sun, Zhi-Hong Deng, Jian-Yun Nie, and Jian Tang. 2019. [Rotate: Knowledge graph embedding by relational rotation in complex space](#). In *International Conference on Learning Representations*.
- Guo Tang, Zheng Chu, Wenxiang Zheng, Junjia Xiang, Yizhuo Li, Weihao Zhang, Ming Liu, and Bing Qin. 2025. Anre: Analogical replay for temporal knowledge graph forecasting. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 4632–4650.
- Rakshit Trivedi, Hanjun Dai, Yichen Wang, and Le Song. 2017. Know-evolve: Deep temporal reasoning for dynamic knowledge graphs. In *international conference on machine learning*, pages 3462–3471. PMLR.
- Jingbin Wang, Jinfan Yuan, Jinsong Lai, Fuyuan Zhang, and Kun Guo. 2026. [Temporal knowledge graph reasoning based on multidimensional information interaction and dynamic frequency awareness](#). *Neurocomputing*, 682:133496.
- Qingyun Wang, Manling Li, Xuan Wang, Nikolaus Parulian, Guangxing Han, Jiawei Ma, Jingxuan Tu, Ying Lin, Ranran Haoran Zhang, Weili Liu, Aabhas Chauhan, Yingjun Guan, Bangzheng Li, Ruisong Li, Xiangchen Song, Yi Fung, Heng Ji, Jiawei Han, Shih-Fu Chang, and 8 others. 2021. [COVID-19 literature knowledge graph construction and drug repurposing report generation](#). In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies: Demonstrations*, pages 66–77, Online. Association for Computational Linguistics.

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*, volume 37, pages 4765–4773.

Bishan Yang, Scott Wen-tau Yih, Xiaodong He, Jianfeng Gao, and Li Deng. 2015. Embedding entities and relations for learning and inference in knowledge bases. In *Proceedings of the International Conference on Learning Representations (ICLR) 2015*.

Fuqiang Yang, Yue Zhang, Xuechen Zhao, and Shengnan Pang. 2024. Unleashing the power of decoders: Temporal knowledge graph extrapolation with householder transformation. *Symmetry*, 16(9):1166.

Kailun Ye, Xiangjie Kong, Yuchao Zhang, Xuan Wang, Linan Zhu, Jiaxin Du, Guojian 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*, pages 3800–3810.

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)*, pages 12617–12631.

Shengzhe Zhang, Wei Wei, Rikui Huang, Wenfeng Xie, and Danyang Chen. 2024. Modeling historical relevant and local frequency context for representation-based temporal knowledge graph forecasting. In *Findings of the Association for Computational Linguistics: EMNLP 2024*, pages 7675–7686.

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*, volume 35, pages 4732–4740.

A Notation Summary

Table 8 summarizes the main notation used throughout DyHist. We include the symbols relevant to historical snapshot modeling, query-aware window selection, local representation aggregation, and relation-aware prior calibration. For compactness, auxiliary variables that appear only within a single module are explained where they are introduced.

B Experimental Details

Datasets and evaluation metrics. We evaluate DyHist on four standard TKG extrapolation benchmarks: ICEWS14, ICEWS18, ICEWS05-15, and GDELT. The ICEWS datasets (Boschee et al., 2015) are derived from the Integrated Crisis Early Warning System, and GDELT (Leetaru

and Schrodte, 2013) is collected from the Global Database of Events, Language, and Tone. Following previous studies (Chen et al., 2024c), we use chronological splits and report MRR and Hits@{1, 3, 10} under the time-aware filtered setting. Table 9 summarizes the dataset statistics. Compared with ICEWS datasets, GDELT has a much finer temporal granularity, with a time interval of 15 minutes. ICEWS05-15 contains the largest number of entities and timestamps among the ICEWS datasets, making it suitable for evaluating long-range temporal modeling. The unseen-event ratios show that all datasets contain a considerable proportion of test facts that do not directly appear in the training set, which provides a diagnostic setting beyond directly repeated facts.

Table 8: Main notation used in DyHist.

Symbol	Description
$(s_n, r_n, ?, t)$	Future query at timestamp t
$\mathcal{G}_\tau$	Temporal snapshot at timestamp τ
$\mathcal{H}_t^{\text{loc}}$	Local history sequence before t
$\mathcal{G}_t^{\text{glob}}$	Sampled global historical subgraph
$\mathcal{I}_t^{\text{loc}}$	Local candidate snapshot indices
$\mathcal{I}_t^R$	Selected recent-window indices
$\mathcal{I}_t^M$	Selected medium-window indices
$\mathcal{I}_t^S$	Short-term subset of the recent window
$\mathcal{I}_t^{\text{rem}}$	Remaining local indices outside $\mathcal{I}_t^R$
$\mathbf{e}_{s_n}^0$	Learnable base embedding of subject s_n
$\mathbf{e}_{r_n}^0$	Learnable base embedding of relation r_n
$\mathbf{e}_{r_n, t}$	Time-aware relation representation
$\mathbf{q}_n$	Query representation
$\mathbf{H}_t^l$	Local entity-state matrix at timestamp τ
$\mathbf{H}_g$	Global historical representation at query time
$\mathbf{d}_\tau$	Snapshot descriptor of $\mathcal{G}_\tau$
$\mathbf{h}_n^R$	Aggregated recent-window representation
$\mathbf{h}_n^M$	Aggregated medium-window representation
$\mathbf{h}_n^L$	Fused local historical context
$\mathbf{h}_{n, \text{short}}^t$	Short-term local representation
Δ_n^t	Fast-slow residual deviation
$\mathbf{h}_{l, s_n}^t$	Final local subject representation
$\mathbf{h}_{s_n, t}^{\text{dec}}$	Subject-side decoding representation
$\mathbf{z}_t^{(n)}$	Neural logits over candidate objects
$c_t^{(n)}(o)$	Time-decayed support of candidate o
$\eta_t^{(n)}(o)$	Recent support trend of candidate o
β_{r_n}	Relation-specific prior strength
$b_t^{(n)}(o)$	Relation-aware historical bias
$\hat{z}_t^{(n)}(o)$	Calibrated logit of candidate o
$\bar{b}_t$	Reward baseline for policy optimization

Additional implementation details. We use the same hyperparameter settings across all datasets unless otherwise specified. The learning rate is set to 0.001, the weight decay is 1×10^{-5} , and gradient clipping is applied with a maximum norm of 1.0. The RGCN encoder uses two layers with 100

bases. The local history length is set to 10, with both recent and medium window sizes set to 3, and the fast window size set to 2. For window selection, the policy temperature is set to 0.7, and the entropy weight is set to 0.02. The dynamic historical prior is applied at the logit level, with initial prior strength 0.11, temporal decay factor $\gamma = 0.15$, and trend weight $\rho = 0.6$. The local-global fusion weight λ is set to 0.95. Unless otherwise specified, the reported experiments use a fixed random seed of 3407.

Training and inference. Training is conducted at the timestamp level, where each batch contains facts from one timestamp. We train DyHist and its variants using the same chronological training protocol and select the checkpoint with the best validation MRR for final testing. For window selection, we sample from the learned policy distributions during training, while using top-ranked snapshots during inference to ensure stable predictions. Unless otherwise specified, all ablation and analysis experiments keep the same backbone, optimizer, and training settings, and only vary the target module or hyperparameter under study.

C Baseline Descriptions

We briefly describe the baseline models used in our experiments. The compared methods cover static knowledge graph reasoning, temporal interpolation, and temporal extrapolation models.

Static baselines. **DistMult** (Yang et al., 2015) uses a bilinear scoring function with diagonal relation matrices to model pairwise interactions between subject and object embeddings. **ConvE** (Dettmers et al., 2018) applies two-dimensional convolution over reshaped entity and relation embeddings to capture richer feature interactions for link prediction. **RotatE** (Sun et al., 2019) represents relations as rotations in the complex embedding space, enabling the model to capture relational patterns such as symmetry, antisymmetry, inversion, and composition.

Interpolation baselines. **TTransE** (García-Durán et al., 2018) extends TransE by incorporating temporal embeddings and models time-dependent facts through translation operations. **TA-DistMult** (García-Durán et al., 2018) introduces time-aware relation representations into the DistMult scoring function, allowing relation parameters to vary across temporal contexts.

DE-Simple (Goel et al., 2020) uses diachronic embeddings that represent entity embeddings as functions of time, thereby modeling temporal changes in entity representations.

Extrapolation baselines. **RE-NET** (Jin et al., 2020) models historical event sequences with recurrent neural networks and aggregates concurrent events for future link prediction. **RE-GCN** (Li et al., 2021b) combines relational graph convolution with recurrent evolution to update entity representations over temporal snapshots. **CyGNet** (Zhu et al., 2021) models temporal recurrence through a copy-generation framework that balances historical copying and candidate generation. **TiRGN** (Li et al., 2022a) encodes historical snapshots with recurrent graph neural networks to capture temporal relational evolution for extrapolation. **HisMatch** (Li et al., 2022b) matches current queries with historical facts and exploits historical consistency to improve future entity prediction. **RETIA** (Liu et al., 2023) constructs temporal interaction-aware representations by modeling relational evolution and aggregating historical neighborhood information. **CENET** (Xu et al., 2023) exploits contrastive event representation learning and historical candidate information to enhance temporal extrapolation. **CRAFT** (Zhang et al., 2024) models coarse- and fine-grained historical contexts to capture multi-level temporal dependencies for future fact prediction. **THCN** (Chen et al., 2024b) introduces temporal causal convolutional networks to model the influence of historical and concurrent facts in temporal knowledge graphs. **LogCL** (Chen et al., 2024c) learns local historical representations with contrastive learning, aiming to improve the discrimination of query-relevant temporal contexts. **CognTKE** (Chen et al., 2025) incorporates query-related cognitive subgraphs and symbolic evidence to improve interpretable temporal extrapolation. **NADEx** (Gan et al., 2026) introduces a negative-aware diffusion framework that reconstructs query objects with a Transformer denoiser conditioned on temporal-relational context and regularized by negative prototypes. **MIDFA** (Wang et al., 2026) models multidimensional information interaction and dynamic frequency awareness, jointly capturing semantic, temporal, structural, and time-sensitive frequency signals for TKG reasoning.

Table 9: Statistics of the datasets.

Dataset	ICEWS14	ICEWS18	ICEWS05-15	GDELT
Training Facts	74,845	373,018	368,868	1,734,399
Validation Facts	8,514	45,995	46,302	238,765
Test Facts	7,371	49,545	46,159	305,241
Entities	6,869	10,094	23,033	7,691
Relations	230	256	251	240
Time Interval	1 day	1 day	1 day	15 mins
Timestamps	365	365	4,017	2,975
Unseen Events	58.43%	55.69%	39.82%	43.72%

D Details of Query-aware Local Dynamics Modeling

This section provides additional details of the query-aware local dynamics modeling module, including snapshot descriptor construction, dual-window selection, and policy optimization.

D.1 Snapshot Descriptor Construction

For each target timestamp t , DyHist uses the local history $\mathcal{H}_t^{\text{loc}} = \{\mathcal{G}_{t-L}, \dots, \mathcal{G}_{t-1}\}$ as the candidate span, where $L = 10$ in our main setting. After encoding these snapshots with the local recurrent RGCN encoder, each candidate snapshot $\mathcal{G}_\tau$ is represented by the local entity-state matrix $\mathbf{H}_\tau^T \in \mathbb{R}^{|\mathcal{E}| \times d}$. Let $\mathcal{B}_t = \{(s_i, r_i, o_i, t)\}_{i=1}^{N_t}$ denote the timestamp-level batch. For each query $(s_i, r_i, ?, t)$, we construct its query representation $\mathbf{q}_i$ from the subject and relation embeddings. DyHist then builds compact snapshot descriptors to guide query-aware local window selection.

Query-snapshot relevance. The first feature measures how relevant a candidate snapshot is to the current timestamp-level queries. For each snapshot $\mathcal{G}_\tau$, we compute the average cosine similarity between the subject-side local states and the corresponding query representations:

$$\tilde{u}_\tau = \frac{1}{N_t} \sum_{i=1}^{N_t} \cos(\mathbf{H}_\tau^T[s_i], \mathbf{q}_i), \quad (18)$$

where $\mathbf{H}_\tau^T[s_i]$ is the local state of subject entity s_i at snapshot τ . This feature favors snapshots whose subject-side states are semantically aligned with the current query batch.

Temporal lag. The second feature encodes the relative temporal distance from the query timestamp:

$$\ell_\tau = \frac{t - 1 - \tau}{L - 1}, \quad \ell_\tau \in [0, 1]. \quad (19)$$

Here, $\ell_\tau = 0$ corresponds to the latest snapshot $\mathcal{G}_{t-1}$, while $\ell_\tau = 1$ corresponds to the earliest snapshot $\mathcal{G}_{t-L}$. This feature helps distinguish near-query evidence from earlier contextual evidence.

Local variation. Unlike the query-specific relevance features, the following two features describe snapshot-level structural dynamics by aggregating the local states of all entities. The third feature captures the state transition between adjacent snapshots. For $\tau > t - L$, we compute

$$\tilde{\delta}_\tau = \frac{1}{|\mathcal{E}|} \sum_{e \in \mathcal{E}} \|\mathbf{H}_\tau^T[e] - \mathbf{H}_{\tau-1}^T[e]\|_2. \quad (20)$$

For the earliest snapshot, we set $\tilde{\delta}_{t-L} = 0$. A larger $\tilde{\delta}_\tau$ indicates a stronger global state change around snapshot τ , which helps identify snapshots containing dynamic shifts relevant to future prediction.

Latest-state consistency. The fourth feature measures how consistent a candidate snapshot is with the latest local state:

$$\tilde{\alpha}_\tau = \frac{1}{|\mathcal{E}|} \sum_{e \in \mathcal{E}} \cos(\mathbf{H}_\tau^T[e], \mathbf{H}_t^T[e]). \quad (21)$$

This feature complements the local variation feature by helping the policy avoid selecting snapshots that show large changes but are no longer consistent with the latest local context.

Relation-context alignment. In addition to subject-side query relevance, DyHist computes a relation-context feature to measure whether a candidate snapshot is aligned with the relation context of the current batch. Let $\mathbf{c}_i = W_c[\mathbf{q}_i \| \mathbf{e}_{r_i}^0]$ denote the relation-aware query context representation, where W_c is a learnable projection matrix. We summarize snapshot $\mathcal{G}_\tau$ by averaging the local states of all entities:

$$\bar{\mathbf{h}}_\tau = \frac{1}{|\mathcal{E}|} \sum_{e \in \mathcal{E}} \mathbf{H}_\tau^T[e]. \quad (22)$$

The alignment score is computed as

$$\tilde{m}_\tau = \frac{1}{N_t} \sum_{i=1}^{N_t} \cos(\bar{\mathbf{h}}_\tau, \mathbf{c}_i). \quad (23)$$

After normalization, we use m_τ and its recentness-adjusted form $m_\tau(1 - \ell_\tau)$ as descriptor features.

Feature normalization. Since the descriptor features have different numerical scales, we standardize all non-lag features over the candidate span:

$$x_\tau = \frac{\tilde{x}_\tau - \mu_x}{\sigma_x + \epsilon}, \quad \tilde{x}_\tau \in \{\tilde{u}_\tau, \tilde{\delta}_\tau, \tilde{\alpha}_\tau, \tilde{m}_\tau\}, \quad (24)$$

where μ_x and σ_x are computed over $\{\mathcal{G}_{t-L}, \dots, \mathcal{G}_{t-1}\}$. The normalized features are denoted as u_τ , δ_τ , α_τ , and m_τ , while the temporal lag ℓ_τ is already normalized into $[0, 1]$. The final snapshot descriptor is written as

$$\mathbf{d}_\tau = [u_\tau, \ell_\tau, \delta_\tau, \alpha_\tau, m_\tau, m_\tau(1 - \ell_\tau)]. \quad (25)$$

For implementation, we split $\mathbf{d}_\tau$ into two parts: $\mathbf{d}_\tau = [\phi_\tau \| \psi_\tau]$, where $\phi_\tau = [u_\tau, \ell_\tau, \delta_\tau, \alpha_\tau]$ summarizes query relevance, temporal distance, local variation, and latest-state consistency, and $\psi_\tau = [m_\tau, m_\tau(1 - \ell_\tau)]$ encodes relation-context alignment and its recentness-adjusted form.

Recent and medium window scoring. Let $\mathcal{I}_t^{\text{loc}} = \{t - L, \dots, t - 1\}$ denote the candidate snapshot indices. For each branch $X \in \{R, M\}$, DyHist computes selection logits over snapshot descriptors:

$$\zeta_\tau^X = f_X([\phi_\tau \| \psi_\tau]). \quad (26)$$

The policy distribution is

$$\pi_X(\tau \mid \mathcal{D}_t) = \text{softmax}_{\tau \in \mathcal{C}_t^X}(\zeta_\tau^X), \quad (27)$$

where $\mathcal{C}_t^R = \mathcal{I}_t^{\text{loc}}$ and $\mathcal{C}_t^M = \mathcal{I}_t^{\text{loc}} \setminus \mathcal{I}_t^R$. Thus, the medium window is selected from snapshots not used by the recent window.

D.2 Query-guided Temporal Aggregation

After selecting the recent and medium windows, DyHist aggregates the encoded subject states within each window. For query $(s_n, r_n, ?, t)$ and snapshot $\mathcal{G}_\tau$, we first compute a query-guided historical representation:

$$\mathbf{a}_n^\tau = \phi_a([\mathbf{H}_l^\tau[s_n] \| \mathbf{q}_n]), \quad (28)$$

where $\phi_a(\cdot)$ is a projection network. For each window $X \in \{R, M\}$, the aggregation weight is computed as

$$\omega_{n,\tau}^X = \text{softmax}_{\tau \in \mathcal{I}_t^X}(g_X([\mathbf{a}_n^\tau \| \mathbf{q}_n \| \mathbf{d}_\tau])), \quad (29)$$

where $g_X(\cdot)$ is a window-specific scoring function. The window representation is then obtained by

$$\mathbf{h}_n^X = \sum_{\tau \in \mathcal{I}_t^X} \omega_{n,\tau}^X \mathbf{a}_n^\tau. \quad (30)$$

This implements the query-guided temporal aggregator $\text{Agg}_X(\cdot)$ used in the main text.

Short-term aggregation. For the short-term subset $\mathcal{I}_t^S \subseteq \mathcal{I}_t^R$, which contains the most recent snapshots in the recent window, DyHist uses the same aggregation form:

$$\mathbf{h}_{n,\text{short}}^t = \sum_{\tau \in \mathcal{I}_t^S} \omega_{n,\tau}^S \mathbf{a}_n^\tau, \quad (31)$$

where $\omega_{n,\tau}^S$ is computed by the recent-window scoring function $g_R(\cdot)$ restricted to $\mathcal{I}_t^S$. This short-term representation serves as the fast evidence used to compute the residual deviation from the broader recent-medium historical context.

D.3 Window Selection Loss

DyHist optimizes the recent and medium window policies with prediction-oriented feedback and temporal-role regularization. For each timestamp batch t , the policy selects a recent window $\mathcal{I}_t^R$ and a medium window $\mathcal{I}_t^M$ from the local candidate history. Let

$$a_t = (\mathcal{I}_t^R, \mathcal{I}_t^M) \quad (32)$$

denote the joint window-selection action, and let $\mathcal{D}_t$ denote the descriptor set of candidate snapshots.

The reward combines prediction feedback and temporal-role separation:

$$\mathcal{R}_t = \lambda_m \mathcal{R}_{\text{mar}} + \lambda_n \mathcal{R}_{\text{nll}} + \mathcal{R}_{\text{sep}}. \quad (33)$$

The margin reward measures whether the selected windows help separate the ground-truth entity from the strongest negative candidate:

$$\mathcal{R}_{\text{mar}} = \frac{1}{N_t} \sum_{n=1}^{N_t} \left(\hat{z}_t^{(n)}(o_n) - \max_{o \neq o_n} \hat{z}_t^{(n)}(o) \right), \quad (34)$$

where $\hat{z}_t^{(n)}(o)$ denotes the calibrated logit of candidate object o for query $(s_n, r_n, ?, t)$. The NLL reward is defined as

$$\mathcal{R}_{\text{nll}} = -\mathcal{L}_{\text{ent}}, \quad (35)$$

so lower entity prediction loss leads to a larger reward. The separation term encourages the two windows to capture different temporal roles:

$$\mathcal{R}_{\text{sep}} = \lambda_{\text{div}} \mathcal{D}_\phi + \lambda_{\text{lag}} \mathcal{D}_\ell. \quad (36)$$

Here, $\mathcal{D}_\phi$ measures the descriptor-level difference between the selected recent and medium windows:

$$\mathcal{D}_\phi = \left\| \frac{1}{|\mathcal{I}_t^R|} \sum_{\tau \in \mathcal{I}_t^R} \phi_\tau - \frac{1}{|\mathcal{I}_t^M|} \sum_{\tau \in \mathcal{I}_t^M} \phi_\tau \right\|_2, \quad (37)$$

Table 10: Statistics of the repeated/novel split.

Dataset	#Test	#Repeated	#Novel	Rep. (%)	Nov. (%)
ICEWS14	7,371	3,860	3,511	52.37	47.63
ICEWS05-15	46,159	31,567	14,592	68.39	31.61

and $\mathcal{D}_\ell$ measures their temporal lag separation:

$$\mathcal{D}_\ell = \frac{1}{|\mathcal{I}_t^M|} \sum_{\tau \in \mathcal{I}_t^M} \ell_\tau - \frac{1}{|\mathcal{I}_t^R|} \sum_{\tau \in \mathcal{I}_t^R} \ell_\tau. \quad (38)$$

Thus, the separation term encourages the recent window to focus on near-query dynamic evidence and the medium window to preserve earlier contextual information. The window-selection policy is optimized with a policy-gradient objective:

$$\mathcal{L}_{\text{pg}} = -(\mathcal{R}_t - \bar{b}_t) \log \pi(a_t | \mathcal{D}_t) - \lambda_H \mathcal{H}(\pi), \quad (39)$$

where $\bar{b}_t$ denotes the moving average of historical rewards and $\mathcal{H}(\pi)$ is an entropy regularizer. The term $\mathcal{R}_t - \bar{b}_t$ measures whether the current window selection performs better than the average reward level.

E Repeated and Novel Queries

We further examine how DyHist behaves on repeated and less repetitive queries. Following the occurrence of the ground-truth object in history, we split test queries into repeated and novel subsets. A query is considered repeated if its ground-truth object has appeared with the same subject–relation pair in historical snapshots; otherwise, it is considered novel. The split is query-level and fixed across all compared models. Table 10 summarizes the subset statistics. ICEWS14 contains 7,371 test queries, among which 3,860 are repeated and 3,511 are novel. ICEWS05-15 contains 46,159 test queries, among which 31,567 are repeated and 14,592 are novel. The repeated ratio is 52.37% on ICEWS14 and 68.39% on ICEWS05-15, indicating that ICEWS05-15 contains stronger historical recurrence. This makes ICEWS05-15 suitable for evaluating whether the dynamic historical prior can exploit recurrence patterns.

F Details of Dynamic Historical Prior Calibration

This section provides additional details of the dynamic historical prior used in the decoder. Different from representation-level historical fusion,

DyHist calibrates candidate logits with a query-conditioned historical prior. For a query $(s, r, ?, t)$, the historical prior is constructed by retrieving historical facts that share the same subject–relation pair (s, r) . Thus, a candidate object receives historical support only if the same subject–relation pair (s_n, r_n) has linked to it in previous snapshots.

Context-aware relation representation. We maintain a learnable base relation embedding $\mathbf{e}_r^0$ for each relation. At each historical snapshot $\mathcal{G}_\tau$, DyHist aggregates the states of entities associated with relation r to obtain a relation-specific context $\mathbf{c}_r^\tau$, which is fused with $\mathbf{e}_r^0$ through a sigmoid gate:

$$\mathbf{x}_r^\tau = \mathbf{e}_r^0 + \mathbf{c}_r^\tau, \quad \mathbf{g}_r^\tau = \sigma(\mathbf{x}_r^\tau W_g + \mathbf{b}_g), \quad (40)$$

where W_g and $\mathbf{b}_g$ denote the shared learnable weight matrix and bias vector of the gating function, respectively.

$$\mathbf{e}_r^\tau = \mathbf{g}_r^\tau \odot \mathbf{x}_r^\tau + (1 - \mathbf{g}_r^\tau) \odot \mathbf{e}_r^0. \quad (41)$$

The final representation is used as $\mathbf{e}_{r,t}$ in the decoder.

F.1 Query-conditioned Historical Support

Given the query $(s_n, r_n, ?, t)$ and a candidate object o , DyHist first computes its decayed historical support under the current subject–relation pair:

$$c_t^{(n)}(o) = \sum_{\tau < t} e^{-\gamma(t-\tau)} \mathbb{I}[(s_n, r_n, o) \in \mathcal{G}_\tau], \quad (42)$$

where γ is the temporal decay coefficient and $\mathbb{I}[\cdot]$ is the indicator function. This term assigns larger weights to more recent historical occurrences while still preserving long-range recurrence evidence.

To capture whether a candidate is becoming more relevant near the query timestamp, DyHist further introduces a recent trend term. Let $\mathcal{I}_t^{\text{loc}} = \{t - L, \dots, t - 1\}$ denote the local candidate snapshot indices, and let $\mathcal{I}_t^{\text{rem}} = \mathcal{I}_t^{\text{loc}} \setminus \mathcal{I}_t^R$ denote the remaining local indices. We compute the occurrence counts in the recent window and the remaining local history as

$$c_{t,\text{rec}}^{(n)}(o) = \sum_{\tau \in \mathcal{I}_t^R} \mathbb{I}[(s_n, r_n, o) \in \mathcal{G}_\tau]. \quad (43)$$

$$c_{t,\text{rem}}^{(n)}(o) = \sum_{\tau \in \mathcal{I}_t^{\text{rem}}} \mathbb{I}[(s_n, r_n, o) \in \mathcal{G}_\tau]. \quad (44)$$

The recent trend score is then defined by comparing their normalized occurrence rates:

$$\eta_t^{(n)}(o) = \text{ReLU} \left(\frac{c_{t,\text{rec}}^{(n)}(o)}{|\mathcal{I}_t^R|} - \frac{c_{t,\text{rem}}^{(n)}(o)}{|\mathcal{I}_t^{\text{rem}}|} \right). \quad (45)$$

The trend term $\eta_t^{(n)}(o)$ emphasizes candidates whose support increases near the query time, rather than simply favoring candidates that are frequent in distant history. The relation-aware historical bias is then computed as

$$b_t^{(n)}(o) = \beta_{r_n} \left[\log \left(1 + c_t^{(n)}(o) \right) + \rho \eta_t^{(n)}(o) \right], \quad (46)$$

where β_{r_n} controls the relation-specific recurrence strength and ρ controls the contribution of the recent trend term. The logarithmic transformation prevents highly frequent candidates from dominating the neural decoder.

F.2 Relation-aware Prior Strength

The relation-aware historical bias is added to the neural decoder logit as a candidate-level calibration term:

$$\hat{z}_t^{(n)}(o) = z_t^{(n)}(o) + b_t^{(n)}(o), \quad (47)$$

where $z_t^{(n)}(o)$ is the neural decoder logit and $\hat{z}_t^{(n)}(o)$ is the calibrated logit. The strength of the historical bias is controlled by the relation-specific parameter β_{r_n} . Compared with a fixed global strength shared by all relations, relation-aware strength allows DyHist to assign stronger recurrence calibration to repetitive relations and weaker calibration to event-driven relations. This design allows historical recurrence to be weighted differently across relation types.

G Case Study

We provide case studies on ICEWS14 to illustrate how DyHist uses local historical evidence and dynamic historical prior calibration. Different from quantitative ablation results, the case studies focus on individual test queries and show how the rank of the ground-truth entity changes under different model variants or scoring stages. We use filtered ranks throughout this section.

G.1 Counterfactual Ranking Analysis

We first compare DyHist with several degraded variants on the same query. For the query (*Amnesty International*, *Make statement*, $?, t$), the ground-truth object is *Israeli Defense Forces*. This

Table 11: Counterfactual case study on ICEWS14. The same query is evaluated under different model variants.

Model Variant	GT Rank	Top-3 Predictions
w/o Prior	10	Police (Hong Kong), Iran, Police (Canada)
w/o QADW	4	Police (Hong Kong), Government (Nigeria), Iran
DyHist	1	Israeli Defense Forces , Boko Haram, Iran

query is challenging because several distracting candidates, such as police-related entities and country entities, receive high neural decoding scores.

As shown in Table 11, removing the dynamic historical prior lowers the rank of the ground-truth entity from 1 to 10. This indicates that candidate-level historical calibration is important for promoting historically plausible answers. Removing QADW replaces explicit window selection and fast-slow residual injection with query-aware aggregation over the whole local history. This hurts the rank, suggesting the benefit of adaptive local evidence modeling. DyHist ranks the correct entity first, showing that local evidence modeling and dynamic prior calibration provide complementary benefits.

G.2 Detailed Evidence Utilization

We further inspect the selected historical evidence and the prediction change before and after applying the relation-aware historical prior. This analysis explains why the dynamic prior improves the final ranking and whether the selected local history contains meaningful contextual signals.

Table 12 gives a detailed view of the same query used in the counterfactual study. Before prior calibration, *Israeli Defense Forces* is ranked only 19th, while several globally frequent entities receive higher scores. The selected evidence shows complementary roles of the two windows: the recent window captures statement-related local context around the query, whereas the medium window retrieves answer-related evidence involving *Israeli Defense Forces*. After applying the dynamic historical prior, the ground-truth entity is promoted to rank 1. This suggests that DyHist benefits from combining query-aware local evidence selection with relation-aware historical calibration, rather than relying on either local context or recurrence signals alone.

Table 12: Successful-promotion case on ICEWS14. The dynamic prior improves the ground-truth rank from 19 to 1.

Query	Before Prior	After Prior
Query: <i>(Amnesty International, Make statement, ?, t)</i> Answer: <i>Israeli Defense Forces</i>	GT rank: 19 Top-5: Boko Haram, Iran, France, Thailand, China	GT rank: 1 Top-5: <i>Israeli Defense Forces</i> , Boko Haram, Iran, France, Thailand
Selected historical evidence from dual windows		
Recent window	$t - 5$: Amnesty International $\xrightarrow{\text{Make statement}}$ Criminal (Philippines); $t - 4$: Barack Obama $\xrightarrow{\text{Make statement}}$ Foreign Affairs (Trinidad and Tobago); John Kerry $\xrightarrow{\text{Make statement}}$ Iran.	
Medium window	$t - 8$: Amnesty International $\xrightarrow{\text{Make statement}}$ Police (Philippines); Media (Palestinian Territory, Occupied) $\xrightarrow{\text{Make statement}}$ <i>Israeli Defense Forces</i> . $t - 7$: John Kerry $\xrightarrow{\text{Make statement}}$ Military (Russia).	

Table 13: Partial-promotion case on ICEWS14. The dynamic prior improves the ground-truth rank from 18 to 3 without forcing it to rank first.

Query	Before Prior	After Prior
Query: <i>(Chuck Hagel, Consult, ?, t)</i> Answer: <i>Iraq</i>	GT rank: 18 Top-5: Jean-Yves Le Drian, UN Security Council, Tamar Beruchashvili, Abdelkader Messahel, Alexis Tsipras	GT rank: 3 Top-5: Jean-Yves Le Drian, UN Security Council, <i>Iraq</i> , Ashraf Ghani Ahmadzai, Tamar Beruchashvili
Selected historical evidence from dual windows		
Recent window	$t - 4$: Iran $\xrightarrow{\text{Make statement}}$ <i>Iraq</i> ; Terrorist (Islamic State of Iraq and the Levant) $\xrightarrow{\text{Threaten}}$ <i>Iraq</i> . $t - 5$: Iran $\xrightarrow{\text{Express intent to meet or negotiate}}$ <i>Iraq</i> .	
Medium window	$t - 10$: North Atlantic Treaty Organization $\xrightarrow{\text{Consult}}$ <i>Iraq</i> . $t - 7$: John Kerry $\xrightarrow{\text{Make statement}}$ <i>Iraq</i> ; Haider Al-Abadi $\xrightarrow{\text{Make statement}}$ <i>Iraq</i> .	

G.3 Additional Case: Partial Promotion

The previous case shows a successful promotion to rank 1. To avoid relying only on a single highly favorable example, we provide another case where the dynamic prior substantially improves the ground-truth rank but does not make it the top-1 prediction. This case shows that the prior acts as a calibration signal rather than a deterministic memorization rule.

As shown in Table 13, before prior calibration, *Iraq* is ranked 18th. The selected recent evidence contains Iraq-centered regional signals, such as statements, threats, and negotiation intents involving *Iraq*. These facts indicate that *Iraq* is highly active in the near-query context. The medium evidence further provides earlier contextual support, including a consultation-related event involving *Iraq*. This is consistent with the query relation *Consult* and helps connect the candidate entity to

the current prediction. After applying the dynamic historical prior, the rank of *Iraq* improves from 18 to 3. However, DyHist does not force *Iraq* to become the top-1 prediction, even though it receives a positive prior bias. This behavior suggests that the historical prior complements neural evidence rather than dominating the decoder. Therefore, the dynamic prior is better interpreted as a calibration mechanism rather than a hard recurrence-matching rule.

G.4 Discussion

These cases provide several observations. First, the selected dual-window evidence contains query-relevant historical facts, suggesting that local selection can provide interpretable support for prediction. Second, recent and medium windows capture complementary evidence, with recent windows reflecting near-query contexts and medium windows providing earlier semantic clues. Third, the dy-

dynamic historical prior calibrates candidate scores according to recurrence evidence: it can promote strongly supported ground-truth objects to the top rank, while also improving ranks without acting as a hard copy rule. Together, these examples support the main intuition of DyHist: temporal extrapolation benefits from adaptive local history modeling and dynamic candidate-level historical calibration.