SimRule: Collaborative Similarity-Driven Rule Learning for Interpretable Temporal Knowledge Graph Forecasting

Wenshuo Liu[1], Shufeng Hao[2], Ziyu Li[1], Chongyang Shi[1] *

[1]School of Computer Science, Beijing Institute of Technology, Beijing, China
{3220231443, 3120240943, cy_shi}@bit.edu.cn

[2]Taiyuan University of Technology, Taiyuan, Shanxi, China
haoshufeng@tyut.edu.cn

Abstract

Rule-based methods for temporal knowledge graph (TKG) forecasting are valued for their interpretability with the explicit logic chains, but they suffer from a severe cold-start coverage gap: exact-match closed walks terminate at sparse entities, leaving many queries without any rule evidence. Existing similarity-based remedies are unreliable because their similarity is derived from static structural co-occurrence or shallow embeddings with no temporal calibration and hand-tuned thresholds. To address these limitations, we propose SimRule, a collaborative recommendation framework for TKG rule mining. SimRule first mines both exact rules and collaborative surrogate rules with a temporally calibrated dual-view entity similarity index. Then an adaptive dual-branch recommendation fusion with evidence aggregation is designed to obtain the missing entity. Thus, each prediction remains fully auditable as an explicit rule chain with quantified time gaps. On ICEWS14, ICEWS18, and ICEWS05-15, SimRule achieves state-of-the-art MRR and Hits@K ($K \geq 3$) among rule-based methods.

1 Introduction

Temporal knowledge graphs (TKGs) represent world knowledge as timestamped quadruples (s, r, o, t), capturing how entities and their relations evolve over time. The TKG extrapolation task asks a model to predict missing objects in future queries $(e_q, r_q, ?, t_q)$. Among the paradigms proposed for this task, rule-based methods (Liu et al., 2022; Li et al., 2023; Huang et al., 2024) stand out for their interpretability: every prediction is grounded in an explicit temporal logic rule mined via closed temporal random walks. However, this imposes a hard requirement: every hop in the walk must land on a directly observed edge in the training graph.

This requirement creates a cold-start coverage gap. Real-world TKGs are structurally sparse: in ICEWS18, more than 40% of entities appear fewer than five times in training. When a walk reaches such an entity, the walk finds no qualifying outgoing edge and terminates, discarding the candidate rule. Just as a recommender system fails to generate recommendations for a user with no interaction history, a rule miner produces no rules for queries involving sparse entities. Figure 1 shows the failure of conventional rule-based models on samples.

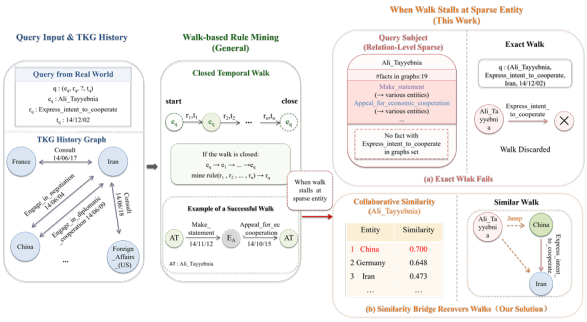

Figure 1: An illustration of the coverage gap in rule-based TKG forecasting. When the query subject is relation-level sparse, the exact-match walk is discarded and produces no rule evidence. SimRule addresses this by jumping to a collaboratively similar peer entity and continuing the walk through its history.

SimRule solves this cold-start coverage gap by treating TKG rule mining as a recommendation problem. Entities are modelled as users whose interaction histories encode behavioural profiles. Two entities are collaboratively similar if they interact with the same counterparts through the same relation types at comparable times. When a walk stalls at a sparse entity v, SimRule queries a prebuilt collaborative recommendation index and selects the highest-ranked peer v' as a recommended surrogate. The walk continues through v''s interaction history, borrowing those observed edges as indirect but principled evidence for v, exactly as a recommender system fills in missing ratings from peer users. Three design decisions follow from this framing. First, the collaborative similarity

*Corresponding author.

Findings of the Association for Computational Linguistics: EMNLP 2026, pages 38960–38974
October 24-29, 2026 ©2026 Association for Computational Linguistics

score must be temporally calibrated, because interaction profiles change over time and a static measure would recommend surrogates based on outdated behaviour. Second, each surrogate hop accumulates approximation error, so a budget of one collaborative hop per walk is enforced to balance coverage against recommendation noise. Third, exact rules and collaborative-surrogate rules carry different evidential reliability and temporal decay dynamics, requiring a dual-branch scoring architecture with disentangled parameters rather than a shared model.

Our contributions are as follows:

- We identify the cold-start coverage gap as the root failure of rule-based TKG forecasting in sparse graphs. We address it by transplanting collaborative filtering into temporal rule mining, treating entities as users and building a temporally decayed dual-view similarity index for surrogate selection.
- We propose a dual-branch temporal scoring architecture with fully disentangled parameters, resolving the gradient conflict between exact-evidence rules and collaborative-surrogate rules.
- Experiments on ICEWS14, ICEWS18, and ICEWS05-15 confirm state-of-the-art MRR and Hits@K ($K \geq 3$). Ablation and case studies validate each design decision.

2 Related Work

Extrapolation on TKGs Three paradigms dominate. Embedding-based encoders (Jin et al., 2019; Li et al., 2021, 2022a; Hu et al., 2025) trade interpretability for accuracy; recent variants RE-SEGNN (Cai et al., 2025) and TRenD (Wang et al., 2025b) refine semantic-evidence aggregation and relation-temporal modelling. Copy-and-generate methods (Zhu et al., 2021; Xu et al., 2023) exploit recurring events but fail on rare entities. Reasoning-enhanced methods include temporal logic rules (Liu et al., 2022; Li et al., 2023) via closed walks, NeSTR (Liang et al., 2026) with LLM abduction, AnRe (Tang et al., 2025) via analogical replay, and DyG-RAG (Sun et al., 2025) via dynamic-graph retrieval.

Rule learning for TKGs TKG rule learning evolved from static induction (Galárraga et al., 2013) through temporal walks (Liu et al., 2022), redundancy correction (Li et al., 2023), and learnable validity (Huang et al., 2024), with later work on denoising and transferability (Chen et al., 2025a). The 2025 line adds task-semantic RL (Chen et al., 2025b), path-search few-shot reasoning (Bai et al., 2025), and one-shot normalising-flow reasoning (Ma et al., 2025). All still rely on directly observed edges for rule instantiation; SimRule fills this orthogonal gap via collaborative-filtering-inspired similarity bridging, recovering rules for entities prior miners systematically ignore.

Entity similarity in graph forecasting Entity similarity was developed in heterogeneous networks (Sun et al., 2011; Shi et al., 2014) and graph representation learning (Perozzi et al., 2014), sharing the collaborative-filtering intuition that structural co-occurrence proxies latent affinity. HiSMatch (Li et al., 2022b) applies subgraph matching at TKG inference but does not extend rule coverage; later work explores temporally-aware retrieval via recursive decomposition and memory-enhanced retrieval (Qian et al., 2025; Wang et al., 2025a). The 2025 line enriches sparse-entity representations through global similarity weighting (Mirtaheri et al., 2025), adaptive neighbourhood enhancement (Zhang et al., 2025b), and topology-aware dynamic-graph transformers (Zhang et al., 2025a); yet all act *post hoc* at the encoder or ranker with static cues, not coupled to a rule miner. SimRule is the first to inject collaborative filtering directly into rule mining: an offline index supplies peer surrogates that extend walks for sparse entities at mining time.

3 Preliminaries

Temporal Knowledge Graphs Let $\mathcal{E}$, $\mathcal{R}$, and $\mathcal{T}$ denote the sets of entities, relations, and timestamps, respectively. A temporal knowledge graph is a collection of facts $\mathcal{G} \subset \mathcal{E} \times \mathcal{R} \times \mathcal{E} \times \mathcal{T}$, where each fact is a quadruple (s, r, o, t) indicating that relation $r \in \mathcal{R}$ holds between subject $s \in \mathcal{E}$ and object $o \in \mathcal{E}$ at timestamp $t \in \mathcal{T}$. For each edge (s, r, o, t), we define an inverse edge (o, r^{-1}, s, t) with inverse relation $r^{-1} \in \mathcal{R}$.

TKG Extrapolation Given observed history $\mathcal{G}_{<t_q} = \{(s, r, o, t) \in \mathcal{G} \mid t < t_q\}$, the extrapolation task asks the model to predict the missing object in a future query $q = (e_q, r_q, ?, t_q)$, where t_q has not been observed during training.

The model ranks all candidates $o \in \mathcal{E}$ by a score $\text{score}(e_q, r_q, o, t_q)$ and returns a ranked list. Subject prediction is handled symmetrically by reformulating it as object prediction under the inverse relation: $(e_o, r_q^{-1}, ?, t_q)$.

Temporal Logic Rules A temporal logic rule of body length l takes the form shown below.

$$(E_1, r_h, E_{l+1}, T_q) \leftarrow \bigwedge_{k=1}^{l} (E_k, r_k, E_{k+1}, T_k) \quad (1)$$

Here r_h is the head relation, $r_1, \ldots, r_l$ are body relations, and $E_1, \ldots, E_{l+1}$ are entity variables. The rule is grounded by a closed temporal walk: starting from a subject entity, following l timestamped edges in the training graph, and returning to the starting entity, with timestamps satisfying $T_1 \geq T_2 \geq \cdots \geq T_l$ and $T_q > T_1$.

4 Method

SimRule implements a three-stage collaborative recommendation pipeline for TKG rule mining. Stage (a) builds a temporally calibrated dual-view entity similarity index offline. Stage (b) uses that index to guide walk sampling, mining both exact rules and collaborative-surrogate rules. Stage (c) instantiates all mined rules against query-time history, aggregates the evidence, and scores candidates through a dual-branch architecture with a learnable fusion weight. Figure 2 gives an overview.

4.1 Collaborative Similarity Computation

The core intuition is that entities sharing interaction histories share future behaviour. We model each TKG entity as a user whose interaction history encodes a behavioural profile: two entities are collaboratively similar if they act in the same relational roles and co-occur with the same counterparts. The pre-built index serves as the recommendation catalogue from which the rule-mining walk selects surrogates when direct evidence is unavailable. Since interaction patterns shift over time, a static measure would select surrogates based on stale co-occurrence; we therefore weight each quadruple by recency.

Time-decayed interaction weighting We weight each quadruple $(s, r, o, t) \in G_{\text{train}}$ by its recency relative to the latest training timestamp t_{ref}. For decay rate $\lambda > 0$, the time weight $w(t)$ is defined as follows.

$$w(t) = \exp(-\lambda(t_{\text{ref}} - t)) \quad (2)$$

Recent quadruples receive weights close to 1, while older interactions are exponentially discounted toward zero.

Dual-view collaborative similarity Two views prevent false surrogate selection. The relational role view captures what an entity does: for entity e_i, we build a $2|\mathcal{R}|$-dimensional profile $\mathbf{v}_i$ by accumulating recency-weighted relation co-occurrence counts and apply ℓ_2-normalisation. The relational-role similarity is the cosine between two such profiles.

$$\text{sim}_r(i, j) = \mathbf{v}_i^\top \mathbf{v}_j \quad (3)$$

The structural co-occurrence view captures with whom an entity interacts: we build a time-weighted co-occurrence matrix $\mathbf{M}$ and obtain a profile $\mathbf{u}_i$ for each entity after the same normalisation. The structural similarity follows the same cosine form.

$$\text{sim}_s(i, j) = \mathbf{u}_i^\top \mathbf{u}_j \quad (4)$$

The two views are fused with mixing weight $\mu \in [0, 1]$.

$$\text{sim}(i, j) = \mu\, \text{sim}_r(i, j) + (1 - \mu)\, \text{sim}_s(i, j) \quad (5)$$

Precise profile construction For entity e_i, we construct profiles as follows:

- Relational-role profile $\mathbf{v}_i \in \mathbb{R}^{2|\mathcal{R}|}$ has separate subject and object blocks: $[\mathbf{v}_i^s]_r = \sum_{(e_i, r, *, t) \in G_{\text{train}}} w(t)$ and $[\mathbf{v}_i^o]_r = \sum_{(*, r, e_i, t) \in G_{\text{train}}} w(t)$, where $w(t)$ is from Eq. (2).
- Structural co-occurrence profile $\mathbf{u}_i$ is the i-th row of symmetric matrix $\mathbf{M}$, where $\mathbf{M}_{ij} = \sum_{(e_i, *, e_j, t) \in G_{\text{train}}} w(t)$ (direction-agnostic).

Both $\mathbf{v}_i$ and $\mathbf{u}_i$ are row-wise ℓ_2-normalized. The fusion weight is fixed to $\mu = 0.5$ on all datasets. The index is built once offline from the training set and contains no learnable parameters.

For each entity e_i we pre-compute a top-k peer list $\mathcal{S}_k(i)$ retaining the k highest-scoring entries. This index is built once offline and incurs zero cost at query time.

4.2 Recommendation-Guided Walk Sampling

Prior rule mining methods treat a dead-end walk as a failed sampling trial: for sparse entities, every walk attempt fails and the cold-start entity receives zero rule coverage regardless of how many walks are sampled. SimRule integrates the collaborative

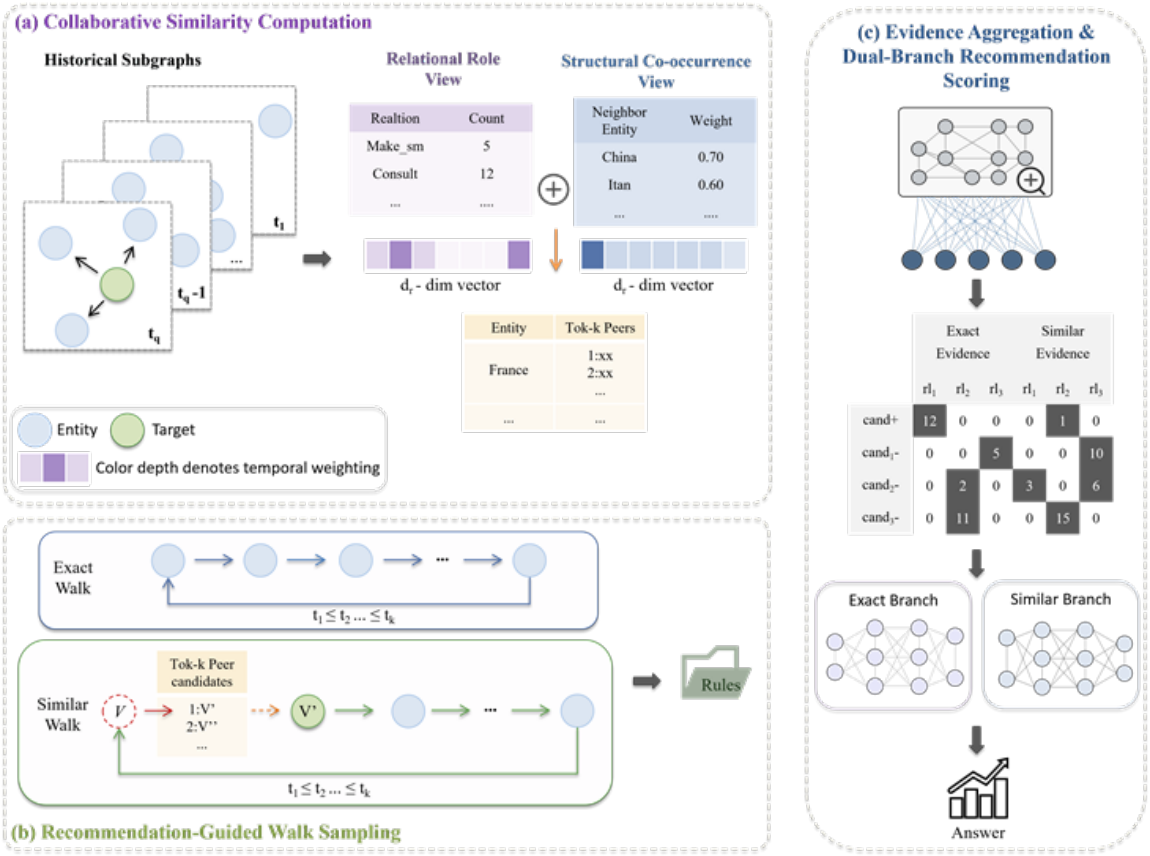

Figure 2: Overview of SimRule. **(a)** Collaborative Similarity Computation builds a temporally decayed dual-view index $\mathcal{S}_k$ offline. The historical subgraph contains only facts with $t < t_q$ (strictly prior to query time). The structural-similarity panel is an illustrative abstraction showing which entities become similar peers; formally, structural similarity is computed as cosine similarity over time-decayed co-occurrence profiles (Eq. 4). **(b)** Recommendation-Guided Walk Sampling mines exact rules $\mathcal{R}_{\text{ex}}$ and similarity rules $\mathcal{R}_{\text{sim}}$. **(c)** Evidence Aggregation and Dual-Branch Scoring produces evidence vectors and fuses two disentangled branches via γ. Shorthand labels: $\text{rl}_1, \text{rl}_2, \text{rl}_3$ denote mined rules ρ_j (Section 4.3); $\text{cand}_+, \text{cand}_-$ denote positive and negative candidate objects respectively.

recommendation index directly into the closed temporal walk to overcome this limitation. When a walk reaches a sparse entity with no qualifying edges, rather than discarding the walk, SimRule queries the recommendation index, selects the top-ranked available peer as a recommended surrogate, and continues through that surrogate's interaction history. We record each surrogate hop as the following bridge tuple.

$$J = \big(v,\ v',\ \text{sim}(v, v')\big) \tag{6}$$

Where v is the stalled entity, v' is the selected peer surrogate, and $\text{sim}(v, v')$ is the similarity score from the index. The result is a recommendation-guided walk that expands rule coverage to entities that exact-match methods cannot serve. Algorithm 1 formalises this collaborative walk sampling procedure; the walk is accepted only if it closes back to the origin s_0, so all recovered rules remain grounded in verifiable temporal patterns.

Surrogate path assumption and safeguards A surrogate-augmented walk deviates from directly observed historical groundings: each bridge substitutes entity v' for v, introducing an assumption that similar entities are behaviorally or semantically substitutable in the current reasoning context. This assumption enables coverage recovery but can introduce noise or semantic mismatch (Section 5.7, Case 3).

SimRule contains three safeguards against this noise:

- **Separation, not mixing.** Surrogate-mined rules form a distinct set $\mathcal{R}_{\text{sim}}$ with a dedicated scoring branch having learnable parameters w^{sim}, β^{sim}, and branch weight γ. If surrogate evidence is noise for a relation, training drives γ toward 1.

Algorithm 1 Collaborative Walk Sampling (Bridge-Augmented Closed Temporal Walk)

Require: Training graph G, similarity index $\mathcal{S}_k$, query relation r_h, walk length l, budget $J_{\max}$
Ensure: Rule ρ with type and confidence $\hat{c}_\rho$

Sample start edge (s_0, r_h, s_1, τ_0) from G
Initialize: $v \leftarrow s_1$, $t_{\text{ref}} \leftarrow \tau_0$, jumps $\leftarrow 0$, path $\leftarrow [\,]$
for step $k = 1, \ldots, l$ **do**
  $\mathcal{E} \leftarrow \{(v, r, u, t) \in G \mid t \leq t_{\text{ref}},$ no immediate backtrack$\}$
  if $\mathcal{E} = \emptyset$ **and** jumps $< J_{\max}$ **then**
    $v' \leftarrow \arg\max_{v'' \in \mathcal{S}_k(v)} \text{sim}(v, v'')$ such that v'' has a qualifying edge
    $v \leftarrow v'$; jumps $\leftarrow$ jumps $+ 1$; record bridge $(v, v', \text{sim}(v, v'))$
    $\mathcal{E} \leftarrow$ qualifying edges from v'
  else if $\mathcal{E} = \emptyset$ **then**
    discard walk **and return** $\emptyset$
  end if
  Sample (v, r_k, u, t_k) from $\mathcal{E}$; $t_{\text{ref}} \leftarrow t_k$; $v \leftarrow u$
  Append r_k to path
end for
if $v \neq s_0$ **then**
  return $\emptyset$ ▷ Walk is not closed
end if
Compute $\hat{c}_\rho$ via body/head support counting
return rule ρ: type EXACT if jumps $= 0$, else SIM

- **Post-hoc re-validation.** Every surrogate-mined rule is re-scored by support counting over real groundings (Eq. (8)); patterns not holding on real data receive low confidence and are pruned by standard filters.
- **Budget control.** $J_{\max} = 1$ limits approximation to at most one substitution per walk, preventing error accumulation.

Recommendation budget control Each surrogate hop introduces approximation error relative to a directly observed path. A hard budget $J_{\max}$ limits the number of collaborative hops per walk through the following constraint.

$$\sum_{k=1}^{l} \mathbb{I}\big[J_k \neq \emptyset\big] \leq J_{\max} \tag{7}$$

Where the indicator counts only those steps that actually invoked a bridge. Section 5.3 confirms that $J_{\max} = 1$ is optimal across all three datasets: one surrogate recommendation recovers the coverage benefit while preventing error accumulation.

Exact and collaborative-surrogate rule sets The procedure produces two rule sets. Exact rules $\mathcal{R}_{\text{ex}}$ use zero surrogate hops and represent fully verified paths. Collaborative-surrogate rules $\mathcal{R}_{\text{sim}}$ use at least one recommended surrogate hop and capture regularities invisible to exact-match walking, representing knowledge recoverable only through collaborative peer inference. Each rule ρ receives a recommendation confidence score defined as the smoothed ratio between supported and total body groundings.

$$\hat{c}_\rho = \frac{|\mathcal{B}^+(\rho)|}{|\mathcal{B}(\rho)| + \epsilon} \tag{8}$$

Here $\mathcal{B}(\rho)$ denotes the body groundings of ρ in G_{train}, and $\mathcal{B}^+(\rho) \subseteq \mathcal{B}(\rho)$ collects those whose head quadruple appears later in the training graph.

4.3 Evidence Aggregation and Dual-Branch Recommendation Scoring

Each mined rule is a temporal recommendation template. Instantiating each rule against the query-time history graph tests whether the rule matches observed facts that support the candidate answer. A rule that fired recently provides stronger evidential support than the same rule that last fired months ago, so a raw binary match indicator discards important temporal reliability information.

For query $q = (e_q, r_q, ?, t_q)$ and candidate o, we test each rule ρ_j against the history graph by searching for valid body groundings before t_q. The time-gap feature for rule ρ_j records the elapsed time between the most recent matching grounding and the query timestamp.

$$x_j(q, o) = t_q - \max\big\{\, t \;\big|\; \rho_j \text{ matches at } t,\ t \leq t_q \,\big\} \tag{9}$$

A smaller gap indicates more current and reliable recommendation evidence. Failed instantiations are masked out via a binary indicator.

$$m_j(q, o) = \mathbb{I} \tag{10}$$

Features for exact and collaborative-surrogate rules are concatenated into a partitioned recommendation evidence vector.

$$\mathbf{x}(q, o) = \big[\mathbf{x}^{\text{ex}}(q, o);\ \mathbf{x}^{\text{sim}}(q, o)\big] \tag{11}$$

This partition preserves the two evidence channels for the dual-branch scoring step.

Algorithm 2 Dual-Branch Recommendation Scoring

Require: Evidence vector $\mathbf{x}(q, o)$, mask $\mathbf{m}(q, o)$, parameters $(w^{\text{ex}}, \beta^{\text{ex}}, w^{\text{sim}}, \beta^{\text{sim}}, \gamma)$
Ensure: Final score $s(q, o)$
$s^{\text{ex}} \leftarrow \sum_{j=1}^{K_{\text{ex}}} w_j^{\text{ex}} \cdot m_j^{\text{ex}} \cdot \exp(-\beta_j^{\text{ex}} \cdot x_j^{\text{ex}}) + \varepsilon$
$s^{\text{sim}} \leftarrow \sum_{j=1}^{K_{\text{sim}}} w_j^{\text{sim}} \cdot m_j^{\text{sim}} \cdot \exp(-\beta_j^{\text{sim}} \cdot x_j^{\text{sim}}) + \varepsilon$
$s \leftarrow \gamma \cdot s^{\text{ex}} + (1 - \gamma) \cdot s^{\text{sim}}$
return s

Dual-branch scoring Exact and collaborative-surrogate rules differ in two ways that make a shared scoring model insufficient. In temporal decay, exact rules derive from verified paths and erode slowly, while collaborative-surrogate rules pass through a peer surrogate and are more sensitive to profile shifts, requiring a faster decay coefficient. In gradient dynamics, exact rules push the model toward conservative high-precision decisions while collaborative-surrogate rules push toward broader coverage; shared parameters receive conflicting gradient signals and converge to a suboptimal compromise. SimRule therefore maintains two fully independent branches, each with its own recommendation weight vector w^b and temporal decay vector β^b, and combines them via a learnable recommendation confidence weight γ that adapts automatically to graph density. The per-branch and fused scores are computed as follows.

$$s^{\text{ex}}(q, o) = \sum_{j=1}^{K_{\text{ex}}} w_j^{\text{ex}} m_j^{\text{ex}} e^{-\beta_j^{\text{ex}} x_j^{\text{ex}}} \tag{12}$$

$$s^{\text{sim}}(q, o) = \sum_{j=1}^{K_{\text{sim}}} w_j^{\text{sim}} m_j^{\text{sim}} e^{-\beta_j^{\text{sim}} x_j^{\text{sim}}} \tag{13}$$

$$s(q, o) = \gamma\, s^{\text{ex}}(q, o) + (1 - \gamma)\, s^{\text{sim}}(q, o) \tag{14}$$

where the two branches share no parameters and $\gamma \in (0, 1)$ is updated jointly with (w^b, β^b). Algorithm 2 summarises the procedure.

Adaptive recommendation confidence fusion The learnable coefficient $\gamma \in (0, 1)$ controls how much the model relies on exact evidence versus collaborative-surrogate evidence. In dense graphs, exact rules cover most queries and $\gamma \to 1$. In sparse graphs, collaborative-surrogate rules carry the primary recommendation signal and γ decreases accordingly. This adaptive blending is analogous to a hybrid recommender that shifts from content-based to collaborative signals as user interaction data becomes available.

Full interpretability Despite the dual-branch architecture, every non-zero score term traces to a specific rule, a specific grounded path, and a quantified time gap. SimRule remains fully auditable.

4.4 Training Objective

For query q_i with ground-truth score s_i^+ and N_{neg} negative scores $\{s_{i,n}^-\}$, we apply a temperature-scaled contrastive loss defined as follows.

$$\mathcal{L}_{\text{pair}} = -\sum_i \log \frac{\exp(s_i^+/\alpha)}{\frac{\sum_n \exp(s_{i,n}^-/\alpha) \cdot m_{i,n}}{\sum_n m_{i,n} + \varepsilon}} \tag{15}$$

Where α is a temperature hyperparameter and $m_{i,n}$ masks all-zero (uninformative) negative evidence vectors. We apply this loss per branch and weight by a stop-gradient copy $\tilde{\gamma}$ of the fusion coefficient. The total training objective takes the form shown below.

$$\mathcal{L} = \tilde{\gamma}\mathcal{L}_{\text{ex}} + (1 - \tilde{\gamma})\mathcal{L}_{\text{sim}} \tag{16}$$

Using $\tilde{\gamma}$ rather than γ directly prevents the loss weighting from interfering with the gradient update of the fusion coefficient itself. All parameters are updated jointly via Adam.

5 Experiments

5.1 Settings

5.1.1 Datasets

We evaluate on three temporal knowledge graph benchmarks derived from the Integrated Crisis Early Warning System: ICEWS14, ICEWS18, and ICEWS05-15. All datasets follow the chronological 8:1:1 train/validation/test split, where the test set strictly post-dates the training set to enforce the extrapolation requirement. Details of each dataset are provided in Appendix A.

5.1.2 Evaluation Protocol

We report MRR and Hits@K ($K \in \{1, 3, 10\}$) under the time-aware filtered setting, in percentages. All metrics are computed on the test set after training to convergence on the training set and tuning early stopping on the validation set.

5.1.3 Baselines

We compare against 16 baselines covering four paradigms: (1) static embeddings: DistMult (Yang

et al., 2015), ComplEx (Trouillon et al., 2016), ConvE (Dettmers et al., 2018), RotatE (Sun et al., 2019); (2) temporal embeddings: TTransE (Jiang et al., 2016), TA-DistMult (García-Durán et al., 2018), TNTComplEx (Lacroix et al., 2020); (3) dynamic graph encoders: CyGNet (Zhu et al., 2021), RE-GCN (Li et al., 2021), CENET (Xu et al., 2023), TiRGN (Li et al., 2022a); and (4) rule-based methods: TLogic (Liu et al., 2022), TR-Rules (Li et al., 2023), TempValid (Huang et al., 2024), CRAFT (Zhang et al., 2024), TiFuLa (Bai et al., 2026). Details of the baselines are provided in Appendix B.

5.1.4 Implementation Details

All training hyperparameters—including settings shared across the three datasets (rule mining configuration, Adam optimiser, learning rate, epochs, random seed, γ-regularisation) and dataset-specific values (history window, negative count, contrastive temperature, smoothing factor, and batch size) are listed in Appendix C.

5.2 Results and Analysis

We report entity prediction results in Table 1. SimRule achieves the best MRR, Hits@3, and Hits@10 across all three benchmarks, establishing a new state of the art within the rule-based paradigm.

SimRule consistently outperforms TempValid, the strongest prior rule-based baseline, across all three datasets. The gains are most pronounced on Hits@10, the metric most sensitive to rule coverage, and they grow with entity sparsity. This confirms that collaborative peer inference works best in scenarios where exact-match walks often fail.

Across all three benchmarks, rule-based methods consistently surpass dynamic graph encoders on MRR and Hits@K. This hierarchy underscores a fundamental advantage of explicit temporal forecasting for geopolitical event data: structured temporal patterns are more faithfully captured by interpretable logic chains than by latent embedding spaces. SimRule extends this advantage further by addressing the cold-start coverage gap that constrains all prior rule-based methods.

SimRule trails TempValid on Hits@1 for ICEWS14 and CRAFT for ICEWS18. This outcome is consistent with the known behaviour of collaborative-filtering-based methods: borrowing peer evidence occasionally introduces ambiguity at the very top of the candidate ranking, where top-1 precision requires resolving fine-grained distinctions beyond what collaborative signal can support. The consistent leadership on MRR and Hits@K ($K \geq 3$) confirms that this represents a minor and interpretable cost of the coverage extension strategy, not a general precision deficit.

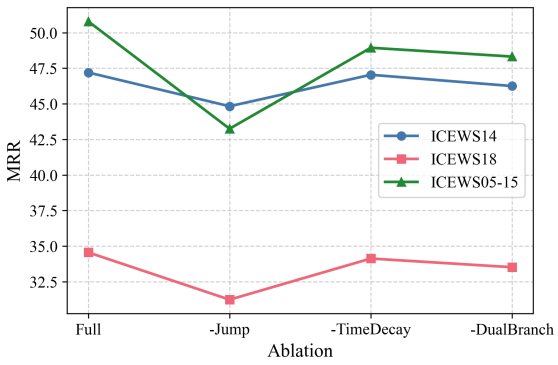

Figure 3: Component ablation results.**-SimIndexDecay:** removes offline similarity index recency weighting (Eq. (2)), leaving learnable time-gap decay (Eqs. (12)–(13)) intact. All three components contribute positively; removing the jump budget causes the largest performance drop.

5.3 Ablation Study

We construct three controlled variants that each remove one component of SimRule, with all other settings held fixed. As shown in Figure 3, all three components contribute positively to SimRule's performance. Removing the recommendation budget limit produces the largest degradation across all datasets by a clear margin. Disabling temporal decay or removing the dual-branch design both lead to consistent performance drops, validating the necessity of each core component.

Exact-rules-only baseline We additionally report an exact-rules-only variant, where all surrogate-rule evidence is removed and the scorer is retrained per-relation with identical features and negative sets as Table 1. The method achieves its most prominent gains on the largest and sparsest benchmark (ICEWS18), where exact-only reaches MRR=31.08 / H@10=53.48, compared to +3.7 MRR / +6.5 H@10 for the full model. A positive improvement is also observed on ICEWS05-15 (+0.3 MRR; 50.46 $\rightarrow$ 50.79), confirming that surrogate evidence provides substantial gains scaled by graph sparsity.

Table 1: Entity prediction results on ICEWS14, ICEWS18, and ICEWS05-15 (%). Methods are grouped into four paradigms, separated by horizontal lines. The best result in each column is **bold**; the second best is underlined.

Model	ICEWS14				ICEWS18				ICEWS05-15			
	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
ComplEx	32.54	23.43	36.13	50.73	22.94	15.19	27.05	42.11	32.63	24.01	37.50	52.81
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
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
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
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
TNTComplEx	34.05	25.08	38.50	50.92	21.23	13.28	24.02	36.91	27.54	9.52	30.80	42.86
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
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
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
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
TLogic	43.04	33.56	48.27	61.23	29.82	20.54	33.95	48.53	46.97	36.21	53.13	67.43
TR-Rules	43.32	33.96	48.55	61.17	30.41	21.10	34.58	48.92	47.64	37.06	53.80	67.57
TempValid	45.78	**35.50**	51.34	65.06	33.50	23.91	37.89	52.33	50.31	39.46	56.71	70.55
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
TiFuLa	42.91	35.18	48.90	63.51	33.12	**26.25**	36.58	51.02	49.72	**40.41**	55.80	68.12
SimRule (ours)	**47.20**	34.31	**53.53**	**73.04**	**34.56**	22.39	**38.92**	**59.60**	**50.79**	37.80	**57.67**	**76.37**

5.4 Hyperparameter Analysis

We sweep the surrogate candidate pool size k and the maximum bridge steps per walk $J_{\max}$. Both parameters exhibit stable optima across all three datasets: performance follows an inverted-U curve over k and peaks consistently at $k = 3$, while single-step bridging ($J_{\max} = 1$) is strictly better than two-step bridging. Full sweep curves and per-dataset values are reported in Appendix E.

5.5 Sparse-Entity Coverage Analysis

To validate the central claim that SimRule addresses the cold-start coverage gap, we analyze rule coverage as a function of entity sparsity.

Sparsity distribution The benchmarks are heavily sparse: 74.3% / 72.5% / 67.0% of entities appear fewer than 5 times as training subjects on ICEWS14/18/05-15.

Exact-rule coverage by subject training frequency Table 2 reports the percentage of test queries for which the gold answer receives at least one exact-rule match, stratified by query-subject training frequency.This confirms quantitatively that the cold-start coverage gap concentrates on sparse-frequency buckets.

Evidence quadrants Table 3 classifies test queries by evidence type for the gold answer.

These findings refine understanding of surrogate evidence: it primarily functions as weight-calibration for queries already covered by exact rules.

Table 2: Exact-rule coverage of gold answer by query-subject training frequency. The coverage gap is concentrated on sparse subjects.

Subject Freq.	ICEWS14	ICEWS18
0 (unseen)	17.2%	10.8%
1–4	40.3%	37.3%
5–19	67.4%	59.2%
20–99	85.4%	81.9%
≥ 100 (frequent)	94.3%	95.3%

Table 3: Evidence quadrant distribution. Surrogate evidence overwhelmingly co-occurs with exact evidence; its primary role is auxiliary weight calibration for queries already covered by exact rules.

Dataset	Both	Exact-only	Surrogate-only	Neither
ICEWS14	69.4%	15.1%	0.20%	15.3%
ICEWS18	67.4%	19.7%	0.16%	12.7%
ICEWS05-15	80.0%	12.0%	0.11%	7.9%

5.6 Analysis of Rule Library and Fusion Behaviour

To shed light on SimRule's empirical gains, we examine the structure of the mined rule library and the dynamics of the learned fusion coefficient γ.

Table 4 reports rule library statistics. Collaborative-surrogate rules are consistently outnumbered by exact rules, confirming the design intent: SimRule uses collaborative peer inference as targeted supplementation rather than a wholesale replacement of direct evidence. The collaborative-to-exact ratio is highest on ICEWS18 ($12{,}890/42{,}781 \approx 0.30$), consistent with its

Table 4: Rule library statistics ($|\mathcal{R}|$: count; $\bar{c}$: mean confidence).

Dataset	$\lvert\mathcal{R}_{\text{ex}}\rvert$	$\lvert\mathcal{R}_{\text{sim}}\rvert$	$\bar{c}_{\text{ex}}$	$\bar{c}_{\text{sim}}$
ICEWS14	28,566	5,784	0.127	0.098
ICEWS18	42,781	12,890	0.062	0.074
ICEWS05-15	61,760	12,116	0.121	0.140

greater frequency of cold-start entities.

On ICEWS14 and ICEWS18, collaborative-surrogate rules carry lower mean confidence than exact rules, as expected from the relaxed matching constraint. On ICEWS05-15, however, this ordering reverses ($\bar{c}_{\text{sim}} = 0.140 > \bar{c}_{\text{ex}} = 0.121$): over eleven years, certain cross-entity patterns stabilise into reliable templates that accumulate higher head support than shorter-lived exact rules, showing that collaborative-surrogate rules are not inherently lower quality.

The fusion coefficient γ concentrates above 0.5 across all datasets, confirming that the exact branch dominates globally. Crucially, the distribution does not degenerate to $\gamma = 1$: the collaborative branch remains active for queries where exact evidence is sparse. Beyond this global tendency, γ shows substantial per-relation dispersion: dense relations attract γ close to 1, while sparse relations exhibit lower γ to amplify the collaborative signal, emerging automatically from the disentangled training objective without any hand-tuned supervision.

5.7 Case Study

To ground the collaborative recommendation mechanism in concrete terms, we examine two successful collaborative-surrogate rules and one failure case mined from ICEWS14. Their key attributes are summarised in Table 7 of Appendix F.

Case 1: cross-actor diplomatic alignment. The walker jumps from France to Iran ($s \approx 0.55$) and continues through Iran's history along body relation *Express intent to meet or negotiate* to mine a rule with head *Consult*. The two surrogates share 14 body objects (e.g., China, Japan), and every traversed body quadruple is a genuine observation anchored on Iran. The closed walk captures a consult-behaviour regularity across two states with overlapping diplomatic counterparts—a pattern exact-match walking cannot reproduce, because closure depends on substituting Iran's meeting-intent history rather than France's own literal path.

Case 2: group-level subject substitution. Head and body coincide on *Arrest, detain, or charge with legal action*. The walker jumps from Men (South Africa) to Immigrants (South Africa) with $s \approx 0.75$. Men (South Africa) appears only as the object of arrest events, never as the arresting subject, while Immigrants (South Africa) contributes the single outgoing edge that activates the walk. Body objects overlap minimally, yet the substitution remains valid because both entities are anchored in the same national context.

Case 3: cross-national over-generalisation (failure). The walker jumps from Men (United States) to Women (Australia) with $s \approx 0.55$. Although both entities are active as objects of arrest events, their body-object sets are largely disjoint: only two objects are shared (Mexico, Police (Australia)), and the non-overlapping targets fall in different geopolitical contexts. The walk closes structurally, but the two group entities are not semantically interchangeable—a similarity threshold alone cannot prevent cross-group misalignment when shared structural signal does not reflect shared semantic scope.

Cases 1 and 2 confirm that similarity jumping recovers useful rules beyond exact-match walking when peers share context (overlapping interlocutors or the same country). Case 3 motivates object-distribution overlap as an additional filter for future work.

6 Conclusion

We presented SimRule, a collaborative framework for interpretable TKG rule mining. It treats the cold-start coverage gap as a recommendation problem: sparse entities borrow evidence from collaboratively similar peers, as recommender systems do for data-sparse users. Entities are profiled via temporally decayed dual-view similarity, and surrogate-bridged walks extend rule coverage to queries exact-match mining cannot reach.

The two-class rule library is handled by a dual-branch scoring head with disentangled parameters, which resolves the gradient conflict between exact and surrogate evidence; learned decay coefficients remain interpretable, with short-window intentions decaying fast and stable institutional patterns decaying slowly.

Experiments on three ICEWS benchmarks confirm state-of-the-art MRR and Hits@K ($K \geq 3$). Every SimRule prediction remains fully auditable: it can be traced to an explicit rule chain, a grounded path in the history graph, and a quantified time gap.

Limitations

The recommendation budget $J_{\max}$ controls the coverage-noise trade-off and is set to 1 based on empirical evaluation. While this choice is consistent across all three ICEWS datasets, its optimality may not generalise to TKGs with fundamentally different sparsity structures or denser temporal patterns such as event streams with sub-day granularity.

SimRule leads on overall ranking quality (MRR, Hits@3, Hits@10) but does not consistently outperform the strongest baselines on Hits@1. Integrating collaborative-surrogate evidence with targeted top-1 re-ranking strategies is a natural direction for future work.

The offline collaborative similarity index requires pre-computing pairwise similarities for all entities. For TKGs with millions of entities, approximate nearest-neighbour methods such as FAISS would be needed to make this step tractable.

Ethics Statement

This work studies temporal knowledge graph forecasting on publicly available event datasets (ICEWS). No personally identifiable information is involved. The datasets contain geo-political event records that may reflect geopolitical biases present in their original curation; users of SimRule should be aware that rule-based predictions may reflect and reinforce patterns present in the training data.

Acknowledgements

This work is supported by National Natural Science Foundation of China (No. 62372043)

References

Luyi Bai, Qianwen Xiao, Tongji Liu, and Kejie Guo. 2026. Temporal knowledge graph reasoning based on time layering and fuzzy logic rules. *Applied Soft Computing*, page 114598.

Luyi Bai, Han Zhang, Xuanxuan An, and Lin Zhu. 2025. Few-shot multi-hop reasoning via reinforcement learning and path search strategy over temporal knowledge graphs. *Information Processing & Management*, 62(3):104001.

Wenyu Cai, Mengfan Li, Xuanhua Shi, Yuanxin Fan, Quntao Zhu, and Hai Jin. 2025. Re-segnn: recurrent semantic evidence-aware graph neural network for temporal knowledge graph forecasting. *Science China information sciences*, 68(2):122104.

Tingxuan Chen, Liu Yang, Zidong Wang, and Jun Long. 2025a. A rule-and query-guided reinforcement learning for extrapolation reasoning in temporal knowledge graphs. *Neural Networks*, page 107186.

Tingxuan Chen, Liu Yang, Zidong Wang, and Jun Long. 2025b. A rule-and query-guided reinforcement learning for extrapolation reasoning in temporal knowledge graphs. *Neural Networks*, 185:107186.

Tim Dettmers, Pasquale Minervini, Pontus Stenetorp, and Sebastian Riedel. 2018. Convolutional 2D knowledge graph embeddings. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 1811–1818.

Luis Antonio Galárraga, Christina Teflioudi, Katja Hose, and Fabian Suchanek. 2013. AMIE: association rule mining under incomplete evidence in ontological knowledge bases. In *Proceedings of the 22nd International Conference on World Wide Web*, pages 413–422.

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.

Xinhai Hu, Ming Sun, and Chao Yang. 2025. Fskge: Fuzzy spatio-temporal knowledge graph embedding. *Neurocomputing*, page 132537.

Rikui Huang, Wei Wei, Xiaoye Qu, Peijie Wang, and Lifang Yuan. 2024. Confidence is not timeless: modeling temporal validity for rule-based temporal knowledge graph forecasting. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics*, volume 1, pages 10783–10794.

Tingsong Jiang, Tianyu Liu, Tao Ge, Lei Sha, Baobao Chang, Sujian Li, and Zhifang Sui. 2016. Encoding temporal information for time-aware link prediction. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 2350–2354.

Woojeong Jin, Meng Qu, Xisen Jin, and Xiang Ren. 2019. Recurrent event network: autoregressive structure inference over temporal knowledge graphs. *arXiv preprint arXiv:1904.05530*.

Timothée Lacroix, Guillaume Obozinski, and Nicolas Usunier. 2020. Tensor decompositions for temporal knowledge base completion. *arXiv preprint arXiv:2004.04926*.

Ningyuan Li, E Haihong, Shi Li, Mingzhi Sun, Tianyu Yao, Meina Song, Yong Wang, and Haoran Luo. 2023. TR-Rules: rule-based model for link forecasting on temporal knowledge graph considering temporal redundancy. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, pages 7885–7894.

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 *Proceedings of the 31st International Joint Conference on Artificial Intelligence*, 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, Wei Li, Saiping Guan, Jiafeng Guo, Huawei Shen, Yuanzhuo Wang, and Xueqi Cheng. 2021. Temporal knowledge graph reasoning based on evolutional representation learning. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 408–417.

Feng Liang, Weixin Zeng, Runhao Zhao, and Xiang Zhao. 2026. Nestr: A neuro-symbolic abductive framework for temporal reasoning in large language models. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 31907–31915.

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.

Ruixin Ma, Longfei Wang, Huinan Wu, Buyun Gao, Xiaoru Wang, and Liang Zhao. 2025. Historical trends and normalizing flow for one-shot temporal knowledge graph reasoning. *Expert Systems With Applications*, 260:125366.

Mehrnoosh Mirtaheri, Ryan A Rossi, Sungchul Kim, Kanak Mahadik, Tong Yu, Xiang Chen, and Mohammad Rostami. 2025. Towards improving long-tail entity predictions in temporal knowledge graphs through global similarity and weighted sampling. *arXiv preprint arXiv:2507.18977*.

Bryan Perozzi, Rami Al-Rfou, and Steven Skiena. 2014. DeepWalk: online learning of social representations. In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 701–710.

Xinying Qian, Ying Zhang, Yu Zhao, Baohang Zhou, Xuhui Sui, and Xiaojie Yuan. 2025. Plan of knowledge: Retrieval-augmented large language models for temporal knowledge graph question answering. *arXiv preprint arXiv:2511.04072*.

Chuan Shi, Xiangnan Kong, Yue Huang, Philip S Yu, and Bin Wu. 2014. HeteSim: a general framework for relevance measure in heterogeneous networks. *IEEE Transactions on Knowledge and Data Engineering*, 26(10):2479–2492.

Qingyun Sun, Jiaqi Yuan, Shan He, Xiao Guan, Haonan Yuan, Xingcheng Fu, Jianxin Li, and Philip S Yu. 2025. Dyg-rag: Dynamic graph retrieval-augmented generation with event-centric reasoning. *arXiv preprint arXiv:2507.13396*.

Yizhou Sun, Jiawei Han, Xifeng Yan, Philip S Yu, and Tianyi Wu. 2011. PathSim: meta path-based top-k similarity search in heterogeneous information networks. *Proceedings of the VLDB Endowment*, 4(11):992–1003.

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.

Théo Trouillon, Johannes Welbl, Sebastian Riedel, Éric Gaussier, and Guillaume Bouchard. 2016. Complex embeddings for simple link prediction. In *International Conference on Machine Learning*, pages 2071–2080.

Xiaoke Wang, Fu Zhang, Jingwei Cheng, Yiwen Chi, Jiashun Peng, and Yingsong Ning. 2025a. DLTKG: Denoising logic-based temporal knowledge graph reasoning. In *Findings of the Association for Computational Linguistics: EMNLP 2025*, pages 18730–18743.

Ye Wang, Binxing Fang, Shuxian Huang, Kai Chen, Yan Jia, and Aiping Li. 2025b. Extrapolation reasoning on temporal knowledge graphs via temporal dependencies learning. *CAAI Transactions on Intelligence Technology*, 10(3):815–826.

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, Wen-Tau Yih, Xiaodong He, Jianfeng Gao, and Li Deng. 2015. Embedding entities and relations for learning and inference in knowledge bases. In *International Conference on Learning Representations*.

Shengzhe Zhang, Wei Wei, Rikui Huang, Wenfeng Xie, and Dangyang 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.

Xin Zhang, Fei Cai, Jianming Zheng, Zhiqiang Pan, Wanyu Chen, Honghui Chen, and Chonghao Chen. 2025a. Triangle matters! topdyg: Topology-aware

transformer for link prediction on dynamic graphs. In *Proceedings of the ACM on Web Conference 2025*, pages 3607–3617.

Yuchao Zhang, Xiangjie Kong, Kailun Ye, Guojiang Shen, and Shangfei Zheng. 2025b. Tackling sparse facts for temporal knowledge graph completion. In *Proceedings of the ACM on Web Conference 2025*, pages 3561–3570.

Cunchao Zhu, Muhao Chen, Changjun Fan, Guangquan Cheng, and Jian Zhang. 2021. Learning from history: modeling temporal knowledge graphs with sequential copy-generation networks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 4732–4740.

A Dataset Details

The three benchmarks are derived from the Integrated Crisis Early Warning System (ICEWS), a widely-used corpus of timestamped geopolitical events.

ICEWS14. ICEWS14 contains political events occurring during 2014, with 6,869 entities (countries, organisations, political actors) and 230 relation types covering diplomatic interactions, conflicts, and cooperative actions. The dataset has 74,845 training quadruples at one-day temporal resolution.

ICEWS18. ICEWS18 applies the same event-extraction pipeline to 2018, yielding the largest entity vocabulary (23,033 entities) and 256 relation types across 373,018 training quadruples. The dataset is the most structurally sparse: over 40% of entities appear fewer than five times in the training set, providing the strongest test bed for cold-start coverage.

ICEWS05-15. ICEWS05-15 spans eleven years of events from 2005 to 2015 at daily granularity, with 10,094 entities, 251 relations, and 368,868 training quadruples. The long temporal span makes it the strongest setting for evaluating long-range temporal dependencies and the stability of mined rules over time.

B Baselines

Static embeddings. DistMult (Yang et al., 2015) learns diagonal relation matrices for scoring triples. ComplEx (Trouillon et al., 2016) extends this with complex-valued embeddings to capture asymmetric relations. ConvE (Dettmers et al., 2018) applies 2D convolutions over reshaped embeddings. RotatE (Sun et al., 2019) models relations as rotations in a complex vector space.

Temporal embeddings. TTransE (Jiang et al., 2016) augments TransE with timestamp embeddings. TA-DistMult (García-Durán et al., 2018) encodes timestamps through recurrent processing of digit sequences. TNTComplEx (Lacroix et al., 2020) factorises a fourth-order time-relation tensor.

Dynamic graph encoders. CyGNet (Zhu et al., 2021) combines a copy mechanism over historical events with a generation network. RE-GCN (Li et al., 2021) captures temporal evolution through a recurrent graph convolutional architecture. CENET (Xu et al., 2023) introduces contrastive learning to distinguish historical from non-historical events. TiRGN (Li et al., 2022a) models both local and global historical patterns through a time-guided recurrent network.

Rule-based methods. TLogic (Liu et al., 2022) mines temporal logic rules via closed temporal random walks. TR-Rules (Li et al., 2023) extends this with redundancy-aware correction. TempValid (Huang et al., 2024), the direct baseline that SimRule builds upon, learns per-rule temporal validity coefficients to weight rule evidence by recency. CRAFT (Zhang et al., 2024) and TiFuLa (Bai et al., 2026) are recent rule-based methods that further refine confidence calibration and rule-evidence aggregation, respectively.

C Common Training Configuration

Shared settings. Rule mining uses body lengths $l \in \{1, 2, 3\}$, 200 walks per relation, $J_{\max} = 1$, $\lambda = 0.4$, and top-$k = 3$ peers per entity. All models are trained with Adam (learning rate 0.01) for up to 3,000 epochs with random seed 3407 and γ-regularisation coefficient 0.01.

Dataset-specific hyperparameters. The training hyperparameters that vary by benchmark are summarised in Table 5.

Table 5: Dataset-specific training hyperparameters: history window size w, number of negatives n_{neg}, contrastive temperature α, smoothing factor β, and batch size.

Dataset	w	n_{neg}	α	β	batch
ICEWS14	0	100	1.0	0.10	256
ICEWS18	200	70	2.0	0.05	256
ICEWS05-15	1,000	60	2.0	0.05	256

D Efficiency and Scalability Analysis

Full cost profile measured during development. Table 6 summarizes computational resources consumed throughout model development.

Key observations are as followed:

- The offline similarity index is built once and incurs negligible query-time cost (0.3–1 MB dictionary lookup only).
- Surrogate-bridged walk sampling adds only constant-factor overhead.

Table 6: Efficiency breakdown. The offline similarity index is built once; query-time overhead dominated by feature grounding, which grows with history window w and is embarrassingly parallel.

Stage	**ICEWS14 (7.1K ents)**	**ICEWS18 (23K ents)**
Similarity index build (one-off, offline)	5 s / 0.9 GB peak	90 s / 8.3 GB peak
Runtime footprint (top-k dictionary)	0.3 MB	∼1 MB
Rule mining per walk	0.16 ms (exact), 0.10 ms (surrogate)	–
Feature grounding per query	1.4 s/query	39 s/query
Training + eval (all relations)	60 min	∼15 h
Inference scoring per query	0.44 ms	0.54 ms

- Feature grounding (determining which historical edges satisfy each rule body) is the dominant computational cost and grows with temporal history window.

E Hyperparameter Sensitivity Analysis

We examine the sensitivity of two bridging parameters: the surrogate candidate pool size k and the maximum bridge steps per walk $J_{\max}$. The sweep covers $k \in \{2, 3, 5, 10\}$ and $J_{\max} \in \{1, 2\}$, with all other settings held at their dataset-specific optimal values.

As shown in Figure 4, performance follows a clear inverted-U curve over k and peaks consistently at $k = 3$: too few peers (k=2) restrict surrogate diversity, while too many (k=10) introduce noisy candidates whose behavioural profile drifts from the stalled entity. As shown in Figure 5, single-step bridging ($J_{\max} = 1$) is strictly better than two-step bridging, since each additional surrogate hop accumulates approximation error relative to a directly observed path. Both conclusions hold consistently across ICEWS14, ICEWS18, and ICEWS05-15.

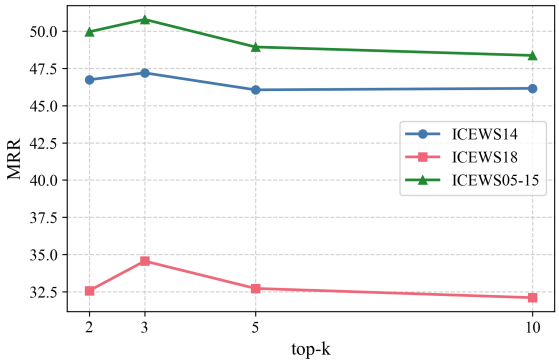

Figure 4: Sensitivity to the number of similar peers k.

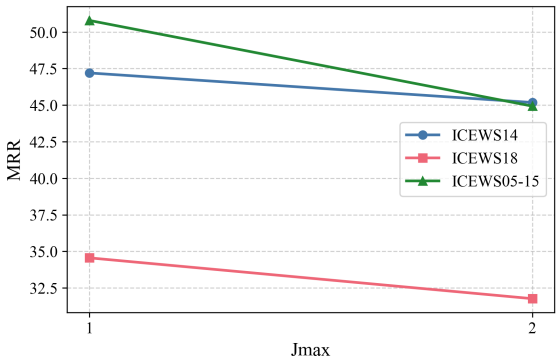

Figure 5: Sensitivity to the maximum collaborative jumps $J_{\max}$.

F Case Study Details

The detailed attributes of the three case studies discussed in Section 5.7 are summarised in Table 7.

Table 7: Two successful and one failed similarity-jump rules mined from ICEWS14. The walker jumps from query subject e to peer e' ($\Rightarrow$), then follows body relation r_1 along e''s history. Case 3 illustrates that high bsupp and $\hat{c}$ alone do not imply semantic interchangeability when body-object distributions barely overlap.

Attribute	**Case 1 (Success)**	**Case 2 (Success)**	**Case 3 (Failure)**
Query template	(France, r_h, ?, t_q)	(Men_SA, r_h, ?, t_q)	(Men_US, r_h, ?, t_q)
Head relation r_h	Consult	Arrest, detain, or charge	Arrest, detain, or charge
Body relation r_1	Express intent to meet or negotiate	Arrest, detain, or charge	Arrest, detain, or charge
Surrogate jump J	France $\Rightarrow$ Iran (s=0.55)	Men_SA $\Rightarrow$ Immigrants_SA (s=0.75)	Men_US $\Rightarrow$ Women_AU (s=0.55)
Walk trajectory	France $\Rightarrow$ Iran $\xrightarrow{r_1}$ shared diplomatic interlocutors	Men_SA $\Rightarrow$ Immigrants_SA $\xrightarrow{r_1}$ S. African institutions	Men_US $\Rightarrow$ Women_AU $\xrightarrow{r_1}$ Australian institutions
Mined confidence $\hat{c}$	0.33	0.68	0.66
Body support bsupp	467	467	453
Object overlap on r_1	high (14 shared)	low (1 shared)	low (2 shared)