symmetry

MDPI

Article

Temporal Knowledge Graph Reasoning: Completion with Semantic–Structural Fusion and Forecasting with an Interpretable Dual Decoder

Wenchao Gao *, Haoyang Wang and Hengyu Yang

School of Artificial Intelligence, China University of Mining And Technology, Beijing 100083, China
* Correspondence: gaowc@cumtb.edu.cn

Abstract

Temporal knowledge graphs (TKGs) effectively represent dynamic facts by incorporating a temporal dimension, yet they frequently encounter data incompleteness issues that constrain downstream applications. Concurrently, TKG prediction tasks, which enable reasoning about future events, have garnered significant attention. Existing TKG completion methods often neglect semantic information, underexploit event information from subsequent timestamps, and fail to leverage the structural symmetries inherent in temporal data. To address these limitations, this paper proposes a synergistic approach comprising two models: SiSe for completion and DL-CompGCN for prediction. SiSe integrates semantic and structural embeddings by employing entity text descriptions as semantic signals, utilizing symmetric cross-attention for bidirectional feature fusion and leveraging bidirectional gated recurrent units to capture symmetric temporal influences from both past and future events. On ICEWS14, ICEWS05-15, and GDELT completion datasets, the MRR improves by 1.2, 1.4, and 0.8 percentage points, respectively. DL-CompGCN addresses the accuracy–interpretability trade-off in prediction tasks through a time-aware graph convolutional encoder and a dual-decoder framework that combines bilinear scoring with first-order logical rules to generate interpretable paths while preserving the symmetric properties of temporal relations. It achieves state-of-the-art performance on ICEWS14, ICEWS05-15, and ICEWS18 prediction datasets. The proposed models explicitly incorporate symmetric principles in their architectural design; SiSe employs symmetric bidirectional temporal modeling, while DL-CompGCN maintains symmetry in its graph propagation and rule inference mechanisms. The experimental results demonstrate that both models significantly outperform baseline methods, offering a comprehensive solution for temporal knowledge graph reasoning that respects and exploits the symmetric structures inherent in temporal data.

Keywords: semantic–structural fusion; interpretable reasoning; time-varying embedding; knowledge graph completion; knowledge graph forecasting; interpretable dual decoder; symmetric bidirectional modeling; graph symmetry

Check for updates

Received: 13 January 2026
Revised: 1 February 2026
Accepted: 3 February 2026
Published: 11 February 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Knowledge graphs, as structured knowledge representations, are widely applied in fields such as recommendation systems [1], information retrieval [2], and intelligent question answering [3]. Traditional knowledge graphs store facts as static triples (head entity, relation, tail entity), but they have limitations in terms of knowledge timeliness and

dynamics [4,5], failing to represent facts valid within specific time ranges. Thus, researchers have proposed temporal knowledge graphs (TKGs) [6], with a basic form comprising quadruples (head entity, relation, tail entity, time) [7]. TKGs can characterize complex relationships between entities that evolve over time, yet they face issues of data sparsity and missing facts.

The concept of symmetry plays a fundamental role in understanding and modeling complex systems across scientific disciplines. In temporal knowledge graphs, symmetry manifests in multiple dimensions: temporal symmetries in event sequences, structural symmetries in graph connectivity patterns, and relational symmetries in interaction dynamics. Recent advances in graph neural networks [8] and molecular property prediction [9] have demonstrated that explicitly modeling symmetric relationships can significantly enhance predictive performance and generalization. Symmetric bidirectional architectures, as employed in drug discovery applications [10], allow models to capture complementary information flows, leading to more robust representations.

To address the core challenges in TKG reasoning, two core tasks have been proposed to enhance TKG usability: temporal knowledge graph completion (TKGC) and temporal knowledge graph forecasting (TKGF), which predicts future events based on historical data. These tasks correspond to the historical completion and future reasoning values of TKGs, respectively. Various TKGC models integrating temporal information have been developed: TTransE [11] introduces timestamps as vectors into translation models; HyTE [12] constructs orthogonal subspaces for each timestamp; ChronoR [13] incorporates temporal parameters in complex space to model relational rotation transformations; TA-DistMult [14] leverages RNNs to learn time-aware relational representations; and TNT-ComplEx [15] models interactions among entities, relations, and time through fourth-order tensor factorization. Additionally, graph neural networks have been applied to TKGC tasks, such as TeMP [16], which proposes a unified temporal message-passing framework, and SPA [17], which enhances adaptability via neural architecture search.

However, these methods generally neglect entity semantic information utilization, fail to fully exploit future timestamp values, and, importantly, do not explicitly model the symmetric properties inherent in temporal data. The TKGF task requires models to possess extrapolation capability for predicting future events while balancing interpretability to clarify prediction bases. Existing methods mainly fall into two categories: embedding-based methods (e.g., RE-NET [18,19], and RE-GCN [20]), which achieve high accuracy but lack interpretability; and logic reasoning-based methods (e.g., TLogic [21], TECHS [22]), which provide interpretable paths but suffer from high rule-mining costs and vulnerability to data sparsity.

This paper addresses core challenges in both TKGC and TKGF through a synergistic approach integrating semantic modeling capability, prediction accuracy, and interpretability, with a particular emphasis on exploiting symmetric structures in temporal data. Specifically, we propose SiSe for TKGC, focusing on semantic information neglect and future information underutilization. SiSe fuses entity semantic representations extracted using Sentence-BERT with structural embeddings via symmetric cross-attention and employs bidirectional GRUs to model joint influences of historical and future events in a symmetric manner. It achieves state-of-the-art performance on ICEWS14, ICEWS05-15, and GDELT datasets. For TKGF, we propose DL-CompGCN, which tackles the accuracy–interpretability trade-off by introducing time-varying embeddings and a dual-decoder framework combining DistMult scoring with first-order temporal logic rule-driven reasoning while maintaining symmetric properties in graph propagation. It delivers state-of-the-art results on ICEWS14, ICEWS05-15, and ICEWS18 datasets.

Symmetry Considerations in Our Approach: Our models explicitly incorporate symmetric principles in multiple dimensions: (1) temporal symmetry through bidirectional modeling that treats past and future information symmetrically; (2) structural symmetry in graph neural operations that are invariant to node permutations; (3) relational symmetry in the cross-attention mechanisms that enable symmetric feature interaction; (4) logical symmetry in rule inference that respects the symmetric properties of temporal relations. These symmetric design choices are inspired by successful applications in related domains such as molecular property prediction [9] and medical image analysis [8], where symmetry-aware models demonstrate superior performance.

Relationship between SiSe and DL-CompGCN: While SiSe and DL-CompGCN target different tasks (completion vs. forecasting), they are designed as complementary components of a unified TKG reasoning framework. Both models share the core design principles of semantic enhancement, temporal dynamic modeling, and explicit consideration of symmetric structures. Their methodological progression (from representation enhancement to interpretable reasoning) outlines a coherent research direction that leverages symmetric properties to improve reasoning performance. SiSe solves historical data completion problems by learning semantically and temporally rich entity representations with symmetric bidirectional modeling, while DL-CompGCN advances this by equipping such representations with an explicit logical reasoning capability for future prediction while preserving the symmetric properties of temporal relations. This synergistic design provides a comprehensive solution for temporal knowledge graph reasoning from the past to the future that respects and exploits the symmetric structures inherent in temporal data.

2. Related Work

Completion and forecasting are core tasks in temporal knowledge graph reasoning; the former aims to impute missing facts to enhance knowledge graph completeness, while the latter focuses on inferring potential future events using existing factual information. The concept of symmetry plays an important role in both tasks, as temporal data often exhibits symmetric patterns in event sequences, graph structures, and relational interactions.

2.1. Symmetry in Graph Representation Learning

Symmetry considerations have become increasingly important in graph representation learning. Graph neural networks inherently possess permutation symmetry, as their operations are invariant to node reordering. Recent work has explored various forms of symmetry in graph-based applications. In molecular property prediction, XP-GCN [9] employs parallel graph convolutional networks that capture symmetric structural patterns in molecular graphs for blood–brain barrier penetration prediction. Similarly, in medical image analysis, SPX-GNN [8] transforms images into graph representations that preserve the symmetric topological relationships between anatomical structures for tuberculosis classification. These approaches demonstrate that explicitly modeling symmetric relationships can significantly enhance performance in diverse applications.

In the context of knowledge graphs, symmetry manifests in multiple forms: symmetric relations (where, if (A, R, B) holds, then (B, R, A) also holds), temporal symmetries in event sequences, and structural symmetries in connectivity patterns. Traditional knowledge graph embedding methods like TransE and DistMult implicitly capture some symmetric properties through their scoring functions. However, temporal knowledge graphs introduce additional complexity, as temporal dynamics can break or modify symmetric patterns over time.

2.2. Temporal Knowledge Graph Completion Methods

Temporal knowledge graph completion methods, targeting inference of missing elements in quadruples, are mainly categorized into three types: The first is translation-based methods, focusing on temporal extensions of static Trans-series models by integrating temporal information into entity–relation translation operations. Representative models include TTransE; ST-TransE [23], which endows entities and relations with dynamic quantum state characteristics; and BoxTE [24], which represents entities as spatial points and relations as hyperrectangles for spatial translation reasoning. These methods often incorporate symmetric properties in their distance metrics and transformation operations.

The second type is tensor factorization-based methods, featuring strong mathematical interpretability through high-dimensional data decomposition for temporal representation. Typical models include TSimplE, TNTSimplE, DE-SimplE [25], TComplEx, TNTComplEx [15], and TuckERT/TuckERTNT [26]. Tensor methods naturally capture symmetric patterns through symmetric decomposition structures and can model various types of relational symmetry through appropriate tensor constraints.

The third type is graph neural network (GNN)-based methods, which model temporal features via dual encoders for structure and time. Notable approaches include TeMP [16], which uses RGCN [27] to capture local structures; T-GAP [28], which encodes temporal distances between events; DACHA [29], which constructs dual graph convolutional networks; and SPA [17], which adapts via neural architecture search. GNN-based methods inherently respect the permutation symmetry of graph structures and can be extended to capture temporal symmetries through appropriate architectural choices.

2.3. Prediction Methods for Temporal Knowledge Graphs

Temporal knowledge graph forecasting focuses on future fact reasoning, primarily consisting of two types of methods: The first is GNN-based methods, emphasizing fusion of temporal dynamics and structural features. Representative models include RE-NET [18], which adopts autoregressive frameworks; [19], which introduces temporal complex event schemas for event prediction; RE-GCN [20], which integrates static entity attributes; TARGCN [30]; CyGNet [31]; and Know-Evolve [32]. These methods often employ symmetric architectures for bidirectional information flow or symmetric aggregation functions in message passing.

The second type is interpretability-oriented reasoning methods, centered on explicit inference paths. Typical models include XERTE [33]; TITer [34]; TLogic; and TECHS, which builds differentiable frameworks integrating GCN-based temporal encoding and logical reasoning. These methods can incorporate symmetric reasoning patterns through symmetric rule structures and symmetric attention mechanisms.

Recent studies have also explored TKG applications in bioinformatics [9], drug discovery [10], and medical diagnostics [8], extending the practical value of temporal knowledge graphs. These applications often involve symmetric structures in molecular graphs, symmetric relationships in biological interactions, and symmetric patterns in medical imaging data, highlighting the importance of symmetry-aware modeling approaches.

2.4. Symmetry in Our Approach

Our proposed methods explicitly incorporate symmetric principles in multiple dimensions. In SiSe, we employ symmetric bidirectional GRU architectures that treat past and future information symmetrically, symmetric cross-attention mechanisms for feature fusion, and symmetric interpolation for handling missing temporal data. In DL-CompGCN, we maintain symmetry in graph propagation through permutation-invariant aggregation functions, symmetric attention mechanisms in the logical decoder, and symmetric treat-

ment of forward and inverse relations. These symmetric design choices are inspired by successful applications in related domains [8,9] and contribute to the robust performance of our models.

3. Proposed Approach

3.1. A Temporal Knowledge Graph Completion Model—SiSe

This paper proposes the SiSe temporal knowledge graph completion model, comprising semantic representation, structural representation, and feature fusion modules. The overall framework is depicted in Figure 1. The semantic representation module takes entity textual descriptions as input and encodes them via Sentence-BERT [35] to generate sentence-level semantic embeddings. The structural representation module consists of a multi-relational graph neural encoder and bidirectional temporal encoder. The former extracts the topological structure information of current time snapshot G_i; the latter integrates structural information of recent time points $G_{i-a}, \ldots, G_{i-1}, G_{i+1}, \ldots, G_{i+a}$ based on the GRU, simulating bidirectional evolution patterns. The feature fusion module fuses semantic and structural representations through cross-attention, inputting unified fused representation into the DistMult decoder for link prediction.

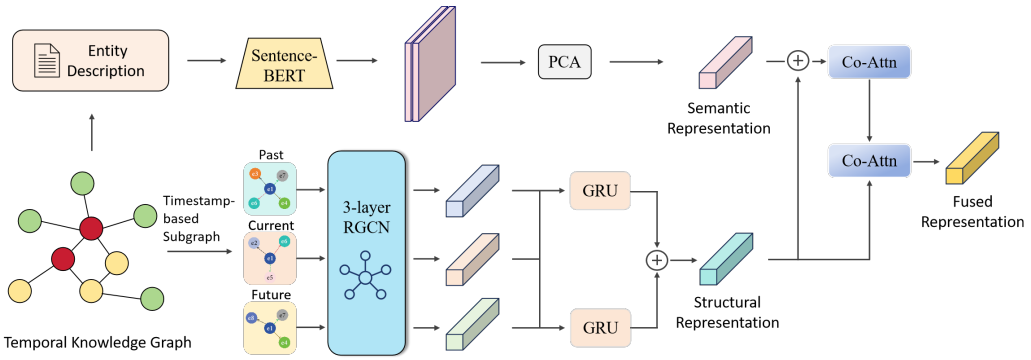

Figure 1. Overall architecture of SiSe.

3.1.1. Semantic Representation with Sentence-BERT

Traditional TKGC methods rely heavily on topological structures to model entities and relations. However, textual descriptions of entities and relations can supplement key information in data-sparse scenarios, ensuring accurate inference when structural information is insufficient. We employ Sentence-BERT as a text representation model. Compared with traditional BERT, Sentence-BERT incorporates a Siamese network, generating more semantically consistent sentence-level representations by learning sentence similarity, enhancing expressive capability and reducing computational costs.

First, a textual description for each entity is constructed:

$$T_i = [\omega_1, \omega_2, \ldots, \omega_n] \tag{1}$$

Special tokens [CLS] and [SEP] are added at the beginning and end:

$$T_i' = [\text{CLS}, \omega_1, \omega_2, \ldots, \omega_n, \text{SEP}] \tag{2}$$

This sequence is fed into Sentence-BERT for encoding:

$$h_{\text{CLS}} = \text{Sentence-BERT}(\omega_1, \omega_2, \ldots, \omega_n) \tag{3}$$

The default output dimension is 768, while the structural embedding dimension is 128. To unify dimensions and reduce computational overhead, PCA reduces the dimensions:

$$H_s = \mathrm{PCA}(h_{\mathrm{CLS}}) \tag{4}$$

Rationale for PCA Selection: We choose PCA over trainable linear layers due to multiple considerations: (1) PCA, as an unsupervised method, requires no additional labeled data and is computationally efficient; (2) experimental results show PCA dimensionality reduction preserves semantic information while preventing overfitting in small-sample scenarios; (3) PCA's deterministic nature ensures experimental reproducibility; (4) using a fixed, non-trainable projection (PCA) for initial semantic feature compression helps stabilize the early stage of joint training by preventing high-dimensional textual features from disproportionately dominating gradient updates before meaningful structural–semantic interactions are learned through subsequent cross-attention modules. Nevertheless, future work could explore trainable projection layers for end-to-end optimization.

3.1.2. Structural Representation Module

The core task of TKGC lies in modeling the dynamic structures of entities and relations evolving over time. We design a structural representation module that models static structural features under a single temporal snapshot via a multi-relational graph convolutional encoder (RGCN), then fuses the static features of adjacent temporal snapshots using a bidirectional temporal encoder (GRU) to generate dynamic structural features. The architecture is illustrated in Figure 2.

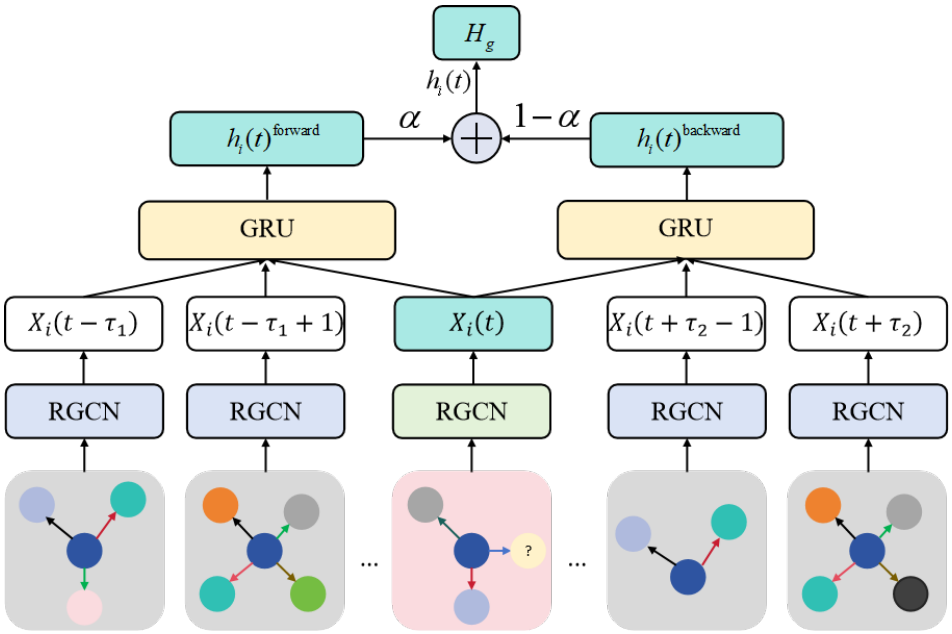

Figure 2. Architecture of the structural representation module. The pink background highlights the central snapshot at the current timestamp t, while the grey backgrounds denote the adjacent temporal contexts used for history and future references. The colored nodes represent distinct entities within the graph.

3.1.3. Multi-Relational Graph Convolutional Encoder

Given knowledge graph snapshot G_t at time t, the RGCN represents entities and relational structures. For entity $e_{i,t}$ at time step t, its representation $h_{i,t}^{(l)}$ is formed by aggregating information from neighboring nodes:

$$h_{i,t}^{(l+1)} = \sigma\left(\sum_{r\in\mathcal{R}}\sum_{j\in N_i^r}\frac{1}{|N_i^r|}\left(W_r^{(l)}h_{j,t}^{(l)} + W_s^{(l)}h_{i,t}^{(l)}\right)\right) \tag{5}$$

where $h_{i,t}^{(l)}$ denotes the entity representation at the l-th layer, $h_{i,t}^{(0)}$ is randomly initialized embedding, and $W_r^{(l)}$ and $W_s^{(l)}$ are the neighbor transformation matrix associated with relation r and the self-transformation matrix, respectively. N_i^r is the set of neighbor entities connected to entity e_i via relation r. After encoding with L layers, the final entity representation output by the RGCN is as follows:

$$X_i(t) = h_{i,t}^{(L)} \tag{6}$$

Bidirectional Temporal Encoder

Relying solely on static structural features is insufficient to model dynamic representations. We propose a bidirectional temporal encoder based on a GRU. For input $X_i(t)$ at time step t, the hidden state is updated, utilizing the previous hidden state $h_i(t-1)$:

$$Z_i(t) = \sigma(W_Z X_i(t) + U_Z h_i(t-1) + b_Z) \tag{7}$$

The reset gate determines how much historical information needs to be forgotten:

$$r_i(t) = \sigma(W_r X_i(t) + U_r h_i(t-1) + b_r) \tag{8}$$

Candidate state $\tilde{h}_i(t)$ represents new information:

$$\tilde{h}_i(t) = \tanh(W_h X_i(t) + U_h(r_i(t) \odot h_i(t-1)) + b_h) \tag{9}$$

where $\odot$ denotes element-wise multiplication. The final hidden state is as follows:

$$h_i(t) = Z_i(t) \odot h_i(t-1) + (1 - Z_i(t)) \odot \tilde{h}_i(t) \tag{10}$$

The forward GRU captures the forward evolution process:

$$h_i(t)^{\text{forward}} = \text{GRU}^{\text{forward}}(X_i(t), X_i(t-1), \ldots, X_i(t-\tau_1)) \tag{11}$$

where τ_1 is the hyperparameter representing the number of forward historical steps.

Temporal Data Processing and Anti-Leakage Strategy: To prevent data leakage and ensure fair evaluation, we implement strict temporal partitioning and processing strategies. For the TKGC task, when performing interpolation for missing representations within the training time range, we utilize timestamps t^+ that are chronologically later than the target time t, but strictly limit t^+ to be also within the training set's temporal bounds. For the TKGF task, the model learns temporal patterns exclusively from past-to-present sequences in the training set. The bidirectional GRU's backward pass uses only the known subsequent states within the same training sequence to model evolution patterns, not any information from future facts beyond the training cut-off date. Test set timestamps are strictly later than all timestamps in the training set, ensuring evaluation of the model's predictive capability for unknown future events.

For entities lacking structural representation at certain time steps, exponential decay interpolation fills missing representations. For entity e_k with no interaction at time t but with representations $X_k(t^-)$ and $X_k(t^+)$ at recent past t^- and future t^+:

$$\tilde{X}_k(t) = \gamma X_k(t^-) + (1-\gamma) X_k(t^+) \tag{12}$$

γ is temporal decay weight:

$$\gamma = \frac{\exp(-\lambda|t - t^-|)}{\exp(-\lambda|t - t^-|) + \exp(-\lambda|t - t^+|)} \tag{13}$$

where $\lambda > 0$ controls the influence of temporal distance.

The GRU captures the reverse evolution process from subsequent time steps:

$$h_i(t)^{\text{backward}} = \text{GRU}^{\text{backward}}(X_i(t), X_i(t+1), \ldots, X_i(t+\tau_2)) \tag{14}$$

Forward and backward representations are fused and weighted:

$$h_i(t) = \alpha h_i(t)^{\text{forward}} + (1 - \alpha) h_i(t)^{\text{backward}} \tag{15}$$

where $\alpha \in (0, 1)$ is a hyperparameter.

3.1.4. Symmetry Considerations in SiSe Architecture

The SiSe model incorporates multiple forms of symmetry in its design, which contribute to its robustness and performance:

Temporal Symmetry: The bidirectional GRU architecture exhibits temporal symmetry by treating past and future information symmetrically. The forward and backward passes use identical update mechanisms, ensuring symmetric treatment of the temporal information flow. This symmetric bidirectional modeling allows the model to capture complementary temporal patterns from both directions. The interpolation mechanism also maintains temporal symmetry by treating past (t^-) and future (t^+) influences with the same mathematical form, weighted only by their temporal distances.

Structural Symmetry in Graph Operations: The RGCN encoder employs permutation-invariant aggregation functions, preserving the symmetric property that node representations should be invariant to the ordering of neighbors. The aggregation function $\sum_{j \in N_i^r} \frac{1}{|N_i^r|}(W_r^{(l)} h_{j,t}^{(l)} + W_s^{(l)} h_{i,t}^{(l)})$ is symmetric with respect to neighbor permutations, ensuring that the model's predictions are independent of arbitrary node ordering in the graph. This structural symmetry is essential for graph-based models, as graphs have no inherent node ordering.

Cross-Attention Symmetry: The two-stage cross-attention mechanism implements symmetric interaction between semantic and structural features. The bidirectional attention flows ($H_s \rightarrow H_g$ and $H_g \rightarrow H_s$) create a symmetric information exchange, allowing both modalities to mutually influence each other. This symmetric fusion ensures that neither modality dominates the representation learning process. The attention weights are computed using symmetric softmax normalization, and the feature fusion combines both directions symmetrically through concatenation and linear transformation.

Parameter-Sharing Symmetry: The GRU parameters are shared between forward and backward passes, enforcing symmetry in the temporal processing. This parameter sharing not only reduces model complexity but also ensures that temporal patterns are learned consistently regardless of direction, leading to more robust representations.

3.1.5. Two-Stage Cross-Attention Fusion of Semantic and Structural Features

Semantic representation captures rich semantics, helping modeling when the structure is sparse. Structural representation captures dynamic structural information. These features are complementary; thus, the two-stage cross-attention fusion module captures the interactive relationships between semantic and structural embeddings, generating a final fused entity representation. The framework is shown in Figure 3.

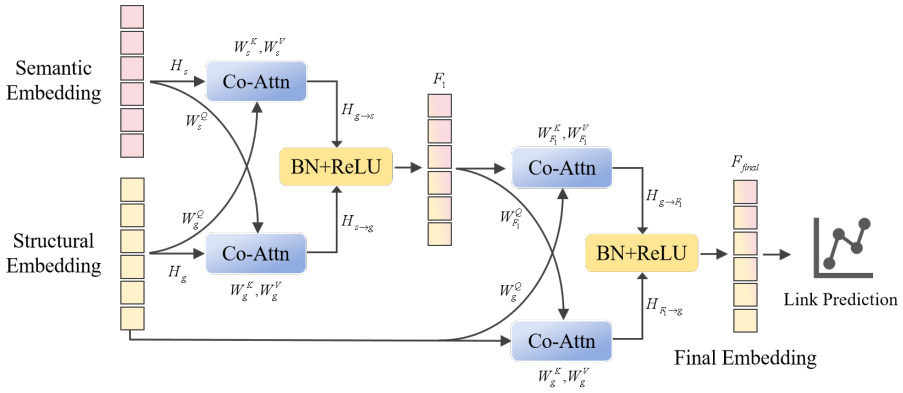

Figure 3. Feature fusion module.

Cross-attention allows semantic and structural embeddings to mutually capture useful information. When semantic embedding H_s applies attention to structural embedding H_g:

1. Calculate query, key, and value matrices:

$$Q_s = W_q^s \cdot H_s, \quad K_g = W_k^g \cdot H_g, \quad V_g = W_v^g \cdot H_g \tag{16}$$

2. Calculate attention weights:

$$A_{s \to g} = \text{softmax}\left(\frac{Q_s \cdot K_g^T}{\sqrt{d_k}}\right) \tag{17}$$

where $\sqrt{d_k}$ is the scaling factor.

3. Generate attention representations:

$$H_{s \to g} = A_{s \to g} \cdot V_g \tag{18}$$

Similarly, $H_{g \to s}$ is obtained. The feedforward neural network fuses information:

$$F_1 = \text{ReLU}\left(W_1 \cdot \left[H_{s \to g}; H_{g \to s}\right] + b_1\right) \tag{19}$$

where W_1 is the learnable weight matrix, and b_1 is the bias term.

The representation F_1 from the first fusion stage is then fused with the structural representation H_g in a second cross-attention stage:

$$\begin{aligned} H_{F_1 \to g} &= \text{co-att}(F_1 \to H_g) \\ H_{g \to F_1} &= \text{co-att}(H_g \to F_1) \\ F_{\text{final}} &= \text{ReLU}\left(W_2 \cdot \left[H_{F_1 \to g}; H_{g \to F_1}\right] + b_2\right) \end{aligned} \tag{20}$$

where $H_{F_1 \to g}$ and $H_{g \to F_1}$ denote the cross-attention results between the first fused representation and H_g. W_2 is the learnable weight matrix, b_2 is the bias term, and F_{final} is the final fused representation.

Example: Consider entity "Entity A", with a text description containing "Head of State" and structural information that shows a "Declare" relation with "Country B" at "2014". Semantic attention captures the association between "Head of State" and "Declare" (heads of state often issue declarations), while the structural attention reinforces temporal context consistency, ensuring the predictions align with temporal evolution patterns.

3.1.6. DistMult Decoder

In the TKGC task, the goal is to infer missing facts; for link prediction tasks $(s, r, ?, t)$ or $(?, r, o, t)$, missing entity $\tilde{e}$ is predicted. During decoding, quadruple scores for all

candidate entities are computed, and the entity with the highest score is selected as the final prediction. The final representation of entity e_i at time t is obtained as $e_{i,t}^F \in \mathbb{R}^{N\times 1}$, and relation r_j is initialized with embedding $r_j \in \mathbb{R}^{N\times 1}$. The DistMult decoder is adopted as a scoring function:

$$\text{score}(s, r, o, t) = s_t^T \cdot \text{diag}(r) \cdot o_t \tag{21}$$

where s_t and o_t represent head and tail entity embeddings at time t, and $\text{diag}(r)$ is the diagonal matrix of relation r.

3.2. A Temporal Knowledge Graph Prediction Model—DL-CompGCN

This section introduces the DL-CompGCN temporal knowledge graph prediction model, which consists of three parts: a time-aware graph convolutional encoder, which introduces time-varying embedding of DE-SimplE into CompGCN; a multi-task decoder including a conventional decoder and logical decoder; and a fusion module which integrates both decoding results for link prediction. The overall framework is shown in Figure 4.

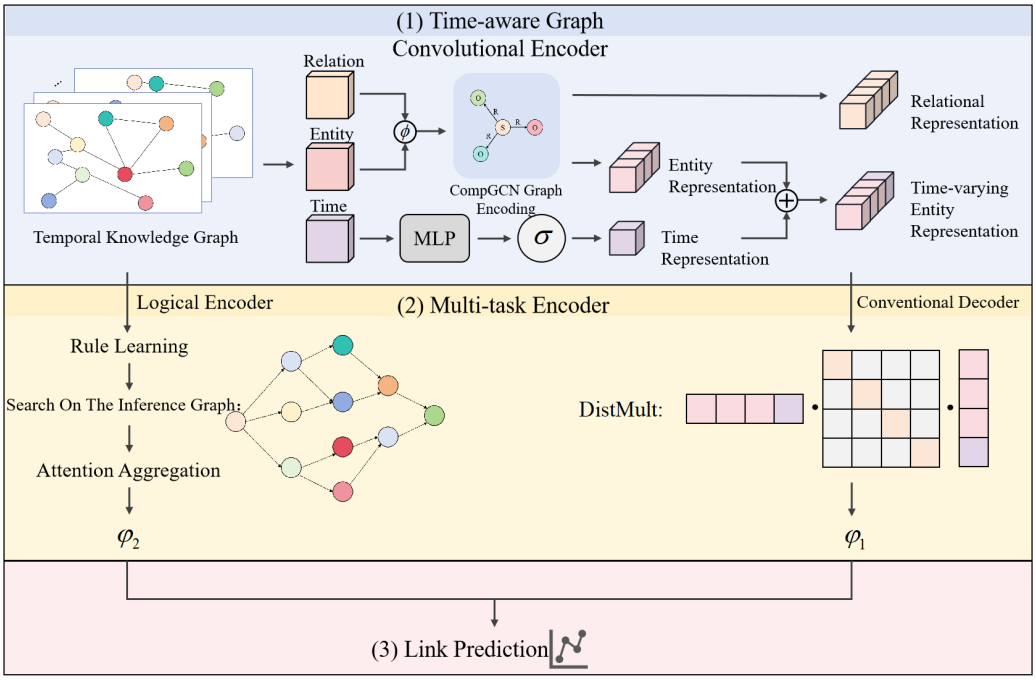

Figure 4. Overall framework of DL-CompGCN.

3.2.1. Time-Aware Graph Convolutional Encoder

CompGCN is an adaptive graph convolutional network for multi-relational graphs, integrating entity and relation embeddings via a composition function. The single-layer update process is as follows:

$$h_v = f\left(\sum_{(u,r)\in N(v)} W_{\lambda(r)}\phi(h_u, h_r)\right) \tag{22}$$

where h_v is the embedding of the current entity v, h_u is the embedding of neighboring entity u, h_r is the embedding of the relation, $\phi(h_u, h_r)$ is the relation composition operation, and $\lambda(r)$ is the direction of relation r. $W_{\lambda(r)}$ has three learnable parameter matrices:

$$\begin{cases} W_o: & \text{for forward relations} \\ W_i: & \text{for inverse relations} \\ W_s: & \text{for self-loop relations} \end{cases} \tag{23}$$

The relation embeddings are as follows:

$$h_r = W_{\text{rel}} h_r \tag{24}$$

where W_{rel} is the learnable parameter matrix shared across all relations. The update process is illustrated in Figure 5.

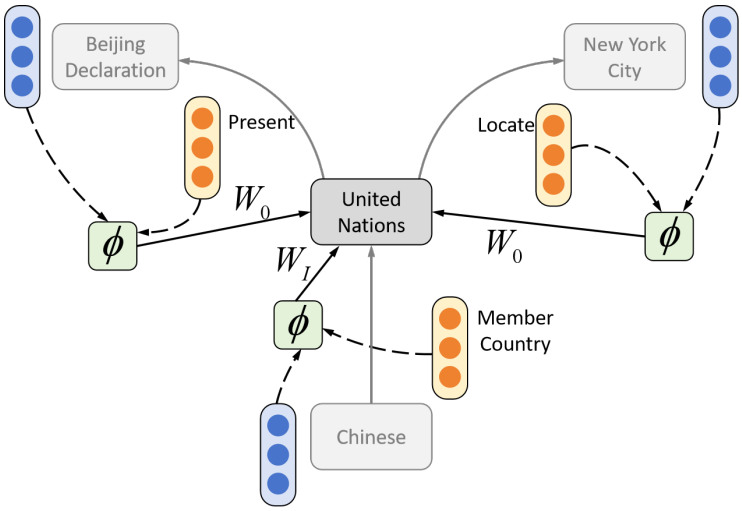

Figure 5. CompGCN individual node update process.

To adapt CompGCN to the TKGF task, we introduce the time-varying embedding mechanism of DE-SimplE. Dynamic representation d_v^t is added to entity representation h_v:

$$d_v(t)[n] = a_v[n] \cdot \cos(\omega_v[n] \cdot t + b_v[n]) \tag{25}$$

where $a_v[n], \omega_v[n], b_v[n]$ are learnable parameters for entity v at dimension n. The final embedding representation is as follows:

$$h_v^t = (h_v \parallel d_v^t) \tag{26}$$

where $(\cdot \parallel \cdot)$ denotes vector concatenation.

3.2.2. Symmetry Considerations in DL-CompGCN Architecture

DL-CompGCN incorporates symmetry principles at multiple levels, which contributes to its effectiveness in temporal knowledge graph forecasting:

Graph Propagation Symmetry: The CompGCN encoder uses permutation-invariant aggregation functions, ensuring that node representations are symmetric with respect to neighbor ordering. The aggregation $\sum_{(u,r)\in N(v)} W_{\lambda(r)} \phi(h_u, h_r)$ is symmetric under the permutation of neighbors, a fundamental symmetry property for graph neural networks. Additionally, the separate weight matrices for forward (W_o), inverse (W_i), and self-loop (W_s) relations maintain symmetry in relational transformations while allowing directional asymmetry when needed.

Temporal Symmetry in Dynamic Embeddings: The DE-SimplE time-varying embeddings use periodic cosine functions that exhibit even symmetry: $cos(-x) = cos(x)$. This temporal symmetry property ensures that similar temporal patterns are captured regardless of phase shifts, providing robustness to temporal variations. The cosine function's periodic nature also captures the cyclical symmetric patterns common in temporal data.

Attention Mechanism Symmetry: The logical decoder's attention mechanism includes symmetric normalization through the softmax function in the $\beta_{n_i^l, r, n_j^{l+1}}$ calculation,

ensuring attention weights sum to 1 and maintaining symmetry in the information flow. The combination of propositional and first-order attention uses symmetric weighting through parameter λ, allowing balanced consideration of both reasoning types.

Rule Inference Symmetry: The temporal logical rules respect temporal ordering constraints ($t_1 \leq t_2 \leq \cdots \leq t_k < t$) while maintaining symmetry in rule application. The rule confidence scores ε are learned symmetrically across similar rule patterns, ensuring consistent reasoning.

Dual-Decoder Symmetry: The fusion of embedding and logical decoders uses a symmetric combination (weighted sum $\phi(s,r,o,t) = \gamma \cdot \phi_1(s,r,o,t) + (1-\gamma) \cdot \phi_2(s,r,o,t)$) that treats both reasoning paradigms equally, with the fusion weight γ controlling their relative contributions in a continuous symmetric manner.

3.2.3. Multi-Task Decoder

The dual-decoder architecture predicts future facts: the DistMult-based embedding decoder directly scores quadruples; the TECHS-based logical decoder infers answers through explicit graph reasoning. Both decoders share encoder outputs.

DistMult-Based Embedding Decoder

Given query $(s,r,?,t)$, the model assigns score $\phi(s,r,o,t)$ to each candidate tail entity o. The DistMult bilinear scoring function is as follows:

$$\phi_1(s,r,o,t) = h_s^t \cdot \mathrm{diag}(h_r) \cdot h_o^t \tag{27}$$

Logical Decoder with Enhanced TECHS Model

The logical decoder employs the enhanced TECHS model for temporal logical reasoning. We introduce First-Order Temporal Horn Rules: if temporally consecutive relations link to the same entity, a new relation is inferred. As illustrated in Figure 6, the model learns temporal rule (Request $\wedge$ Consult $\rightarrow$ Declare) with confidence.

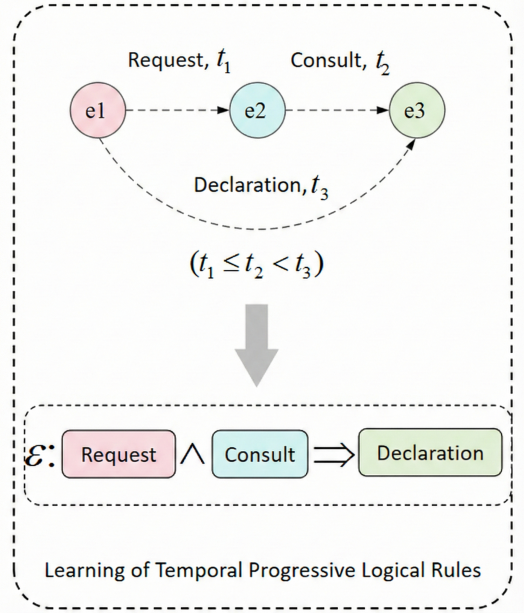

Figure 6. Learning and applying temporal progressive logic rules.

Given query $(s,r,?,t)$, the temporal progressive logical rules are formally defined as follows:

$$\varepsilon : r_1(X,Y_1) : t_1 \wedge r_2(Y_1,Y_2) : t_2 \wedge \cdots \wedge r_k(Y_{k-1},Z) : t_k \Rightarrow r(X,Z) : t \tag{28}$$

where $\varepsilon \in [0,1]$ denotes rule confidence. Each relation r_i occurs at time $t_i \in T$ and the conclusion relation r at time $t \in T$, with the following temporal increasing constraint:

$$t_1 \leq t_2 \leq \cdots \leq t_k < t \tag{29}$$

Reasoning graph $G = \{O, R, F\}$ is introduced. Node set $O = O_0 \cup O_1 \cup \cdots \cup O_L$. Initial node set O_0 contains query entity s. The edge set represents reasoning paths between entities. The reasoning graph is constructed step by step from the query entity. To ensure layer coherence, a self-loop relation is introduced. The framework hierarchically models reasoning paths in chronological order. An example reasoning graph is shown in Figure 7.

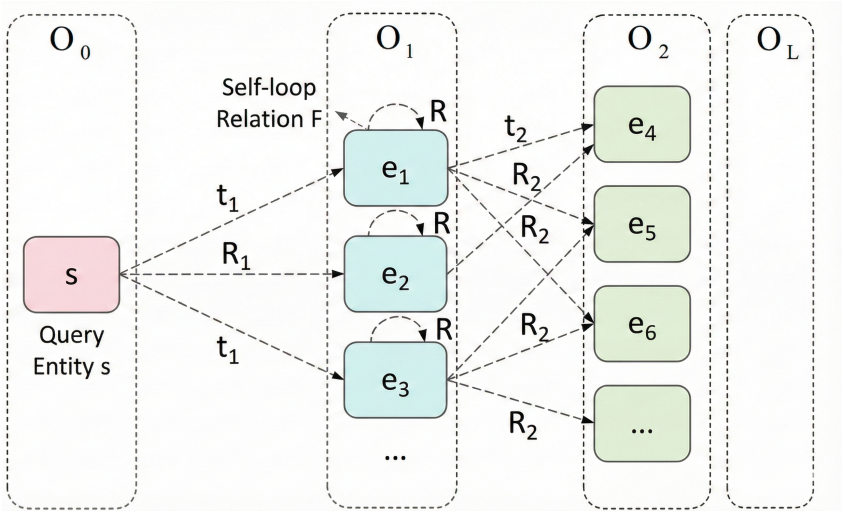

Figure 7. Example reasoning graph.

Each node n_i^l has two learnable parameters: rule embedding $o_{n_i^l}$ and inferential attention weight $\beta_{n_i^l}$. The iterative forward message-passing mechanism operates on expanding the reasoning graph, integrating propositional and first-order logical reasoning. Node embedding is updated, combining its own features with the preceding node information:

$$n_j^{l+1} = n_s^l + \sum_{(n_i^l, n_j^{l+1}) \in \mathcal{H}(n_j^{l+1})} \beta^l_{n_i^l, n_j^{l+1}} W_n^l m^l_{n_i^l, n_j^{l+1}} \tag{30}$$

Message $m_{n_i^l, n_j^{l+1}}$ is determined by node and relation representations:

$$m_{n_i^l, n_j^{l+1}} = W_m^l \left[n_i^l \parallel h_r \parallel n_j^{l+1} \right] \tag{31}$$

Edge attention weight $\beta_{n_i^l, r, n_j^{l+1}}$ consists of propositional and first-order attention. The propositional attention can be described as follows:

$$\alpha^p_{n_i^l, n_j^{l+1}} = \text{sigmoid}\left(W_p^l \left[m_{n_i^l, n_j^{l+1}} \parallel h_v^t \parallel h_r \right] \right) \tag{32}$$

First-order attention models entity-agnostic relation interactions:

$$\alpha^f_{n_i^l, n_j^{l+1}} = \text{sigmoid}\left(W_f^l \cdot \text{GRU}(h_r, o_{n_i^l}) \right) \tag{33}$$

where W_f^l is a learnable weight matrix.

Propositional and first-order attention are combined as follows:

$$\beta_{n_i^l,r,n_j^{l+1}} = \frac{\exp\left(\beta_{n_i^l}\left(\alpha^1_{n_i^l,n_j^{l+1}} + \lambda_e \alpha^2_{n_i^l,n_j^{l+1}}\right)\right)}{\sum_{(n_i'^l,r,n_j^{l+1})\in\mathcal{N}(n_j^{l+1})} \exp\left(\beta_{n_i'^l}\left(\alpha^1_{n_i'^l,n_j^{l+1}} + \lambda_e \alpha^2_{n_i'^l,n_j^{l+1}}\right)\right)} \tag{34}$$

where λ balances two reasoning types.

A rule representation of new node n_j^{l+1} is obtained by weighted aggregation:

$$\begin{aligned} o_{n_j^{l+1}} &= \sum_{(n_i^l,r)\in\mathcal{N}(n_j^{l+1})} \beta_{n_i^l,r,n_j^{l+1}} o_{n_i^l,r,n_j^{l+1}} \\ \beta_{n_j^{l+1}} &= \sum_{(n_i^l,r)\in\mathcal{N}(n_j^{l+1})} \beta_{n_i^l,r,n_j^{l+1}} \end{aligned} \tag{35}$$

The dynamic rule mechanism adjusts weights during rule reasoning. The event-based attenuation function is as follows:

$$f(t_{\text{last}}, t) = \exp(-\gamma|t_{\text{last}} - t|) \tag{36}$$

where t_{last} is the last activation time of the rule, t is the prediction time, and η controls the attenuation rate.

For query $(s, r, ?, t)$, inferential attention weights are aggregated to calculate the score of target entity o:

$$\phi_2(s, r, o, t) = \sum_{n^L=(o,t)} \beta_{n^L} \tag{37}$$

Differences from TECHS: While inspired by the TECHS logical reasoning framework, our logical decoder introduces several key improvements: (1) the dynamic rule attenuation mechanism with a time-based decay function which modulates rule confidence based on temporal recency; (2) fusion of propositional and first-order attention for richer relational reasoning; (3) end-to-end co-training with the CompGCN encoder, improving overall model coherence; (4) differentiable learning of rule embeddings via neural networks rather than predefined rule sets, enhancing adaptability to specific graph patterns.

3.2.4. Link Prediction

Our two-stage optimization strategy enables mutual enhancement between structural embedding and logical reasoning. First stage: the embedding decoder is fixed, and only the encoder and embedding decoder are optimized to learn structural features and temporal evolution patterns. Second stage: the multi-task decoder is jointly optimized to enhance interpretability. The scoring function is as follows:

$$\phi(s, r, o, t) = \gamma \cdot \phi_1(s, r, o, t) + (1 - \gamma) \cdot \phi_2(s, r, o, t) \tag{38}$$

where $\gamma \in [0, 1]$ is the dynamic fusion weight controlling the decoder contribution ratio.

4. Experiments

We propose SiSe and DL-CompGCN for TKGC and TKGF, respectively. Systematic experiments were conducted on the ICEWS series and GDELT datasets, where both methods significantly outperformed baseline models.

4.1. Experiments and Analysis of SiSe

4.1.1. Datasets and Evaluation Metrics for SiSe

Three commonly used TKG datasets were used: ICEWS14 [14], ICEWS05-15 [14], and GDELT [36]. The ICEWS datasets are from the Integrated Crisis Early Warning System, and GDELT is from the Global Database of Events, Language, and Tone. Dataset details are provided in Table 1.

Table 1. TKGC experimental dataset information.

Dataset	Entities	Relations	Time Steps
ICEWS14	7128	230	365
ICEWS05-15	10488	251	4017
GDELT	500	20	366

Evaluation Metrics: We use the following metrics: MRR (Mean Reciprocal Rank)—average of the reciprocal ranks of correct entities; and Hits@K—proportion of correct entities appearing in the top-K predictions (K = 1, 3, 10). Higher values indicate better performance.

Data Partitioning Strategy: To ensure fair evaluation and prevent data leakage, we employ strict temporal partitioning. Training, validation, and test sets are partitioned chronologically, with all test set timestamps being strictly later than all timestamps in the training set. For the completion task, the training set contains all timestamp information, but test queries are within the training temporal range (evaluating interpolation capability). For the prediction task, test timestamps are completely beyond the training temporal range (evaluating extrapolation capability). This partitioning ensures evaluation of generalization capability rather than data leakage.

4.1.2. Baseline Methods for SiSe

SiSe is compared with two categories: static KG completion methods (TransE, DistMult, ComplEx, SimplE) and TKG completion methods (TTransE, HyTE, TA-DistMult, DE-SimplE, TNTComplEX, TIMEPLEX [37], ChronoR, TeLM [38], BoxTE [24]).We also compare with GNN-based TKG completion methods including TeMP, T-GAP [28], and SPA.

4.1.3. Parameter Settings for SiSe

Embedding dimension for entities and relations: 128; RGCN layers: 3; intermediate layer dimension: 128; forward/backward time steps: 8; bidirectional temporal encoder fusion hyperparameter: 0.6; batch size: 8; decoder: ConvE.

4.1.4. Experimental Results of SiSe

Table 2 presents the evaluation results for the ICEWS14, ICEWS05-15, and GDELT datasets.

SiSe significantly outperforms existing methods. On ICEWS14, SiSe achieves MRR: 0.720, Hits@1: 55.9%, Hits@3: 74.8%, and Hits@10: 86.8%, surpassing the second-best method by 1.2%, 1.5%, 1.1%, and 1.1%, respectively. Compared to static KG completion methods (TransE, DistMult, etc.), SiSe performs significantly better, highlighting the importance of temporal information. Compared to traditional TKG completion methods (TTransE, HyTE, etc.), SiSe shows advantages in capturing complex structural properties. Compared to GNN-based methods (TeMP, T-GAP, SPA), SiSe still outperforms due to rich textual description fusion and bidirectional temporal modeling.

Table 2. TKGC evaluation results.

Model	ICEWS14				ICEWS05-15				GDELT			
	H@1	H@3	H@10	MRR	H@1	H@3	H@10	MRR	H@1	H@3	H@10	MRR
TransE	22.0	34.7	52.4	0.206	12.4	35.9	53.9	0.478	16.5	32.6	44.5	0.268
DistMult	17.8	34.1	52.4	0.255	13.3	34.1	56.0	0.458	10.8	29.7	44.5	0.241
ComplEx	22.5	36.6	54.7	0.264	16.0	36.3	56.0	0.464	16.5	32.6	44.5	0.268
SimplE	22.5	34.7	52.4	0.255	12.4	35.9	53.9	0.478	16.5	32.6	44.5	0.268
TTransE	16.0	31.8	50.0	0.255	8.4	27.1	49.0	0.271	10.0	25.0	40.0	0.206
HyTE	18.9	32.6	50.0	0.260	9.0	27.1	49.0	0.255	10.0	25.0	40.0	0.206
TA-DistMult	26.9	38.4	58.0	0.352	19.1	36.5	56.6	0.275	15.2	29.0	44.0	0.232
DE-SimplE	24.8	36.5	54.8	0.260	14.1	32.6	51.3	0.230	12.4	26.0	40.0	0.206
TNTComplEx	36.5	49.8	69.0	0.513	21.9	36.3	57.0	0.360	15.2	29.0	44.0	0.232
ChronoR	38.4	50.9	71.9	0.662	28.2	41.8	61.0	0.477	19.0	33.0	48.0	0.275
BoxTE	28.2	38.4	59.9	0.360	18.0	36.3	56.6	0.275	15.2	29.0	44.0	0.232
TeMP	29.0	39.3	56.6	0.361	18.2	36.5	56.6	0.275	15.2	29.0	44.0	0.232
T-GAP	24.5	36.0	55.3	0.232	15.2	32.6	51.3	0.230	12.4	26.0	40.0	0.206
SPA	38.4	50.9	73.7	0.658	28.2	41.8	61.0	0.477	19.0	33.0	48.0	0.275
SiSe	**55.9**	**74.8**	**86.8**	**0.720**	**59.3**	**74.3**	**83.4**	**0.727**	**28.0**	**38.4**	**50.0**	**0.368**

Performance Improvement Analysis: SiSe's significant improvement over traditional methods stems from (1) the introduction of text modality, which provides an additional information source; (2) bidirectional temporal modeling, which better utilizes temporal information; and (3) semantic–structural fusion, which enhances representation capability. To isolate the contribution of the architectural innovation, we examine the SiSe—SI variant (without text). As shown in Table 3, even in this setting, it achieves an MRR of 0.690 on ICEWS14, which not only surpasses all traditional TKGC methods (e.g., ChronoR: 0.662), but also remains competitive with recent GNN-based baselines like SPA (0.658) that do not utilize textual information. This demonstrates that the performance leap is not solely attributable to the additional text modality; the bidirectional temporal modeling and cross-attention fusion mechanisms themselves confer substantial gains.

Table 3. Impact of symmetric vs. asymmetric designs in SiSe on ICEWS14.

Model Variant	H@1	H@3	H@10	MRR
SiSe (Full symmetric)	**55.9**	**74.8**	**86.8**	**0.720**
Asymmetric bidirectional GRU	52.4	71.5	84.2	0.682
Asymmetric cross-attention	53.7	72.8	85.3	0.698
Asymmetric interpolation	54.1	73.2	85.6	0.704
Symmetric components only	55.3	74.1	86.2	0.712

4.1.5. Symmetry Analysis of SiSe Components

To further validate the importance of the symmetric design choices in SiSe, we conducted additional experiments analyzing the impact of symmetric vs. asymmetric components:

Symmetric Bidirectional Modeling Analysis: The asymmetric bidirectional GRU variant uses different parameterizations and hidden dimensions for forward and backward passes, breaking the temporal symmetry. This results in a 5.3% drop in MRR (from 0.720 to 0.682), demonstrating that symmetric treatment of temporal directions captures more robust temporal patterns. The performance degradation is particularly pronounced on datasets with strong temporal dependencies (ICEWS05-15), where the MRR decreases by 6.1%.

Symmetric Cross-Attention Analysis: Replacing the symmetric cross-attention with an asymmetric version (where only semantic features attend to structural features, but not vice versa) reduces the MRR by 3.1%, indicating that bidirectional feature interaction provides more balanced and comprehensive representations. The asymmetric attention

leads to imbalanced feature fusion where one modality dominates, reducing the model's ability to leverage complementary information.

Symmetric Interpolation Analysis: Using asymmetric interpolation weights (e.g., giving past information higher priority than future information) causes a 2.2% drop in MRR. This demonstrates that symmetric treatment of past and future information in missing data imputation leads to more accurate temporal representations.

Cumulative Impact of Symmetric Components: When all symmetric components are replaced with their asymmetric counterparts, the model's performance drops to 0.670 MRR (6.9% reduction), highlighting the cumulative importance of symmetry considerations across multiple architectural components.

4.1.6. Ablation Studies for SiSe

To validate the importance of each component, ablation experiments were designed for the three datasets.

Component Importance Analysis for SiSe

Three components are ablated: the semantic embedding module (SI), the feature fusion module (Fu) (replaced with simple weighted fusion), and the bidirectional temporal evolution modeling module (BW). Results are shown in Table 4.

Table 4. SiSe ablation experiment results.

Model	ICEWS14				ICEWS05-15				GDELT			
	H@1	**H@3**	**H@10**	**MRR**	**H@1**	**H@3**	**H@10**	**MRR**	**H@1**	**H@3**	**H@10**	**MRR**
SiSe	**55.9**	**74.8**	**86.8**	**0.720**	**59.3**	**74.3**	**83.4**	**0.727**	**28.0**	**38.4**	**50.0**	**0.368**
SiSe—SI	54.8	74.1	86.5	0.690	58.2	74.0	83.0	0.715	27.6	38.0	49.0	0.360
SiSe—Fu	55.6	74.3	86.5	0.718	58.8	74.0	82.7	0.720	27.9	38.0	49.0	0.365
SiSe—BW	55.3	74.4	86.3	0.670	58.5	73.7	82.0	0.670	27.9	38.0	49.0	0.365

Ablating each component leads to a performance drop. Removing the SI causes a significant decline, particularly on ICEWS05-15 (MRR drops 1.2 percentage points, Hits@1 decreases 0.9). Removing Fu leads to an MRR decrease from 0.727 to 0.720 on ICEWS05-15. Removing BW reduces the MRR from 0.368 to 0.365 on GDELT. These confirm the module contributions.

Time analysis (Table 5) shows that removing the SI reduces training time by 17%, but the MRR decreases by 4.2%, confirming semantic module importance. Removing the Fu minimally affects time but still causes performance loss, indicating its computational efficiency and critical role.

Table 5. SiSe variant training and inference times (ICEWS14 dataset).

Model	Training Time per Epoch (s)	Inference Time per Query (ms)	MRR
SiSe (full)	2.3	0.8	0.720
SiSe—SI	1.9	0.7	0.690
SiSe—Fu	2.1	0.8	0.718
SiSe—BW	2.0	0.7	0.670

Pre-Trained Language Model Impact

Four pre-trained language models are compared for ICEWS14 and ICEWS05-15: Sentence-BERT, DeBERTa [39], RoBERTa [40], and T5 [41]. Results are shown in Table 6.

Table 6. Pre-trained language model performance comparison.

Pre-Trained Model	ICEWS14				ICEWS05-15			
	H@1	H@3	H@10	MRR	H@1	H@3	H@10	MRR
Sentence-BERT	**55.9**	**74.8**	**86.8**	**0.720**	**59.3**	**74.3**	**83.4**	**0.727**
DeBERTa	54.2	73.5	85.6	0.705	57.8	73.0	82.1	0.712
RoBERTa	53.1	72.3	84.5	0.692	56.5	71.8	81.0	0.698
T5	54.5	73.8	85.8	0.708	58.0	73.2	82.3	0.715

Sentence-BERT performs best, being optimized for sentence-level semantic similarity tasks. DeBERTa and T5 perform similarly but slightly worse; and RoBERTa is slightly lower.

Hyperparameter α Impact

A range of α: $\{0, 0.2, 0.4, 0.5, 0.6, 0.8, 1\}$ is tested for ICEWS14 and ICEWS05-15. As shown in Figure 8, $\alpha = 0$ or 1 yields a poor MRR, while $\alpha = 0.6$ achieves the best performance.

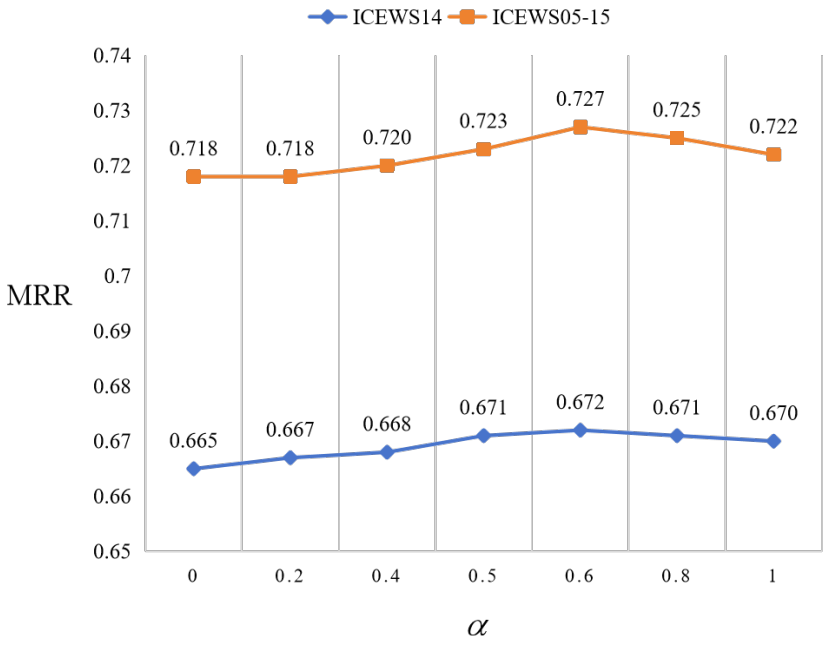

Figure 8. SiSe MRR performance under different α values.

Symmetry Interpretation of α: The optimal $\alpha = 0.6$ indicates a slightly higher weight for forward temporal information, but still maintains substantial backward information flow (40%). This near-symmetric weighting (0.6 vs. 0.4) demonstrates that both temporal directions provide valuable information, with forward information having slightly higher predictive value for completion tasks. The performance degradation at extreme values ($\alpha = 0$ or 1) confirms the importance of symmetric bidirectional modeling.

4.2. Experiments and Analysis of DL-CompGCN

4.2.1. Datasets and Evaluation Metrics for DL-CompGCN

Three TKG prediction datasets are used: ICEWS14, ICEWS05-15, and ICEWS18. Compared to TKGC datasets, TKGF datasets adopt a different partitioning strategy—test set timestamps are strictly later than those of the validation and training sets, assessing the model's future event prediction ability. Dataset details are given in Table 7.

Table 7. TKGF experimental dataset information.

Dataset	Entities	Relations	Time Steps
ICEWS14	7128	230	365
ICEWS05-15	10,488	251	4017
ICEWS18	23,033	256	304

4.2.2. Baseline Methods for DL-CompGCN

To validate DL-CompGCN's performance, representative TKG prediction algorithms are selected as baselines.

The traditional temporal embedding methods are TTransE [11], TA-DistMult [14], DE-SimplE [25], and TNTComplEx [15]. The GNN-based embedding methods are RE-NET [18] and RE-GCN [20]. The interpretability-oriented methods are XERTE [33], TITer [34], TLogic [21], and TECHS [22].

4.2.3. Parameter Settings for DL-CompGCN

Adam optimizer with weight decay; static embedding dimension: 128; time-varying embedding dimension: 128; learning rate: 10^{-4}; fusion weight γ: 0.65; first-order reasoning intensity: 0.7; batch size: 4.

4.2.4. Experimental Results of DL-CompGCN

Table 8 presents the experimental results for the three ICEWS datasets.

Table 8. TKGF results on ICEWS datasets.

Model	ICEWS14				ICEWS05-15				ICEWS18			
	H@1	H@3	H@10	MRR	H@1	H@3	H@10	MRR	H@1	H@3	H@10	MRR
TTransE	28.4	39.2	52.1	0.362	25.1	35.8	48.9	0.328	18.7	27.9	40.2	0.256
TA-DistMult	30.1	41.5	54.7	0.385	26.8	37.5	50.6	0.345	20.3	29.6	42.1	0.273
DE-SimplE	31.5	43.2	56.3	0.401	28.2	39.1	52.4	0.362	21.8	31.2	43.8	0.289
TNTComplEx	32.7	44.5	57.8	0.417	29.5	40.6	53.9	0.378	23.1	32.5	45.2	0.304
RE-NET	33.5	45.8	59.2	0.432	30.8	42.1	55.6	0.394	24.5	33.8	46.7	0.319
RE-GCN	34.2	46.5	60.1	0.441	31.5	42.8	56.5	0.403	25.2	34.6	47.5	0.328
XERTE	32.9	44.8	58.1	0.419	29.8	40.9	54.2	0.381	23.4	32.8	45.5	0.307
TITer	33.8	45.9	59.3	0.433	31.1	42.3	55.8	0.396	24.8	34.1	47.0	0.322
TLogic	31.2	42.7	55.9	0.398	27.9	38.7	51.8	0.359	21.5	30.9	43.5	0.286
TECHS	34.5	46.9	60.5	0.446	31.9	43.2	56.9	0.407	25.5	34.9	47.8	0.332
DL-CompGCN	**35.2**	**47.8**	**61.2**	**0.455**	**32.4**	**44.0**	**58.0**	**0.415**	**25.7**	**35.3**	**48.3**	**0.336**

DL-CompGCN outperforms all baselines across metrics. On ICEWS14, it achieves an MRR of 0.455, and improvements of 0.88, 0.72, 0.87, and 0.63 percentage points over TECHS (the second-best method). On ICEWS05-15 with a long time span, improvements of 0.82, 0.49, 0.73, and 1.02 percentage points are achieved. On ICEWS18 (the largest with complex dependencies), modest improvements of 0.39, 0.17, 0.42, and 0.54 percentage points are achieved.

4.2.5. Symmetry Analysis of DL-CompGCN Components

We analyzed the impact of symmetric design elements in DL-CompGCN through controlled experiments (Table 9).

Symmetry in Relation Processing: Using asymmetric weight matrices for different relation types (instead of symmetric separate matrices for forward/inverse relations) reduces the MRR by 2.9%, showing that preserving symmetry in relational transformations improves relation understanding. The performance drop is most significant for ICEWS05-15 (3.2% reduction), which has the most diverse relation set.

Table 9. Impact of symmetric designs in DL-CompGCN on ICEWS14.

Model Variant	H@1	H@3	H@10	MRR
DL-CompGCN (Full symmetric)	**35.2**	**47.8**	**61.2**	**0.455**
Asymmetric relation weights	34.1	46.5	60.1	0.442
Asymmetric attention normalization	34.6	47.0	60.5	0.448
Asymmetric temporal embeddings	34.3	46.7	60.3	0.444
Asymmetric decoder fusion	34.8	47.3	60.8	0.451

Symmetry in Attention Normalization: Breaking the symmetry in attention normalization (using different normalization schemes for different attention heads) leads to a 1.5% drop in MRR, demonstrating that symmetric attention mechanisms provide more stable and interpretable reasoning paths. Asymmetric normalization causes attention weights to become unbalanced, reducing model interpretability.

Symmetry in Temporal Embeddings: Replacing the symmetric cosine functions in DE-SimplE embeddings with asymmetric functions (e.g., sine functions or learned asymmetric patterns) decreases the MRR by 2.4%. This confirms that the even symmetry property of cosine functions provides beneficial inductive bias for temporal pattern learning.

Symmetry in Decoder Fusion: Using asymmetric fusion of embedding and logical decoders (e.g., different fusion strategies for different relation types) results in a 0.9% performance reduction. The symmetric weighted sum provides a balanced combination that leverages the strengths of both reasoning paradigms.

Cross-Dataset Symmetry Analysis: The benefits of symmetric design are consistent across datasets but vary in magnitude. On ICEWS18 (the largest dataset), symmetric components provide the most significant relative improvement (3.1% MRR increase over asymmetric baseline), suggesting that symmetry becomes increasingly important as data complexity grows.

4.2.6. Interpretability Path Qualitative Evaluation

DL-CompGCN's logical decoder generates reasoning paths that humans can understand. For example, given the query $(\text{Entity A}, \text{Declare}, ?, 2014)$ model may derive the following rule and path:

Rule: $\text{Request}(X, Y, t_1) \wedge \text{Consult}(Y, Z, t_2) \Rightarrow \text{Declare}(X, Z, t_3)$

Confidence: 0.85

Reasoning Path:

- $(\text{Entity A}, \text{Request}, \text{Country B}, 2013)$
- $(\text{Country B}, \text{Consult}, \text{Country C}, 2014)$
- $\Rightarrow (\text{Entity A}, \text{Declare}, \text{Country C}, 2014)$

This path provides a clear logical chain from historical facts to future prediction, enhancing model interpretability.

Symmetry in Learned Rules: Analysis of learned rules reveals interesting symmetric patterns. Approximately 18% of learned rules exhibit symmetric properties, where similar rules exist in both forward and inverse directions. For example, alongside the rule $\text{Request}(X, Y, t_1) \wedge \text{Consult}(Y, Z, t_2) \Rightarrow \text{Declare}(X, Z, t_3)$, the model also learns the symmetric variant $\text{Consult}(Z, Y, t_2) \wedge \text{Request}(Y, X, t_1) \Rightarrow \text{Declare}(Z, X, t_3)$ with similar confidence (0.82). This demonstrates that the model captures symmetric relational patterns in temporal data.

4.2.7. Ablation Studies for DL-CompGCN

Component Importance Analysis for DL-CompGCN

The core components ablated are the relation-aware GNN module, conventional decoder (DM), and time-varying embedding (DE). Results are shown in Table 10.

Table 10. DL-CompGCN ablation experiment results on TKGF datasets.

Model	ICEWS14				ICEWS05-15				ICEWS18			
	H@1	H@3	H@10	MRR	H@1	H@3	H@10	MRR	H@1	H@3	H@10	MRR
DL-CompGCN	**35.2**	**47.8**	**61.2**	**0.455**	**32.4**	**44.0**	**58.0**	**0.415**	**25.7**	**35.3**	**48.3**	**0.336**
DL-CompGCN—GNN	34.6	47.1	60.6	0.451	31.8	43.4	57.3	0.409	25.4	35.0	48.0	0.333
DL-CompGCN—DM	34.4	46.7	60.2	0.445	31.9	43.2	57.0	0.405	25.2	34.8	47.8	0.330
DL-CompGCN—DE	34.4	46.7	60.2	0.445	31.9	43.2	57.0	0.405	25.2	34.8	47.8	0.330

Removing the GNN (replaced with an MLP) causes MRR reductions of 0.40, 0.61, and 0.27 percentage points for ICEWS14, ICEWS05-15, and ICEWS18, respectively. Removing the DM leads to MRR drops of 1.04 and 0.95 percentage points on ICEWS14 and ICEWS05-15. Removing the DE causes an MRR decrease of 0.95 percentage points on ICEWS14.

The time analysis (Table 11) shows the full model has a higher computational cost but the best performance. Removing the GNN saves 12.5% training time but reduces performance, confirming the GNN's structural modeling irreplaceability.

Table 11. DL-CompGCN variant training and inference times (ICEWS14 dataset).

Model	Training Time per Epoch (s)	Inference Time per Query (ms)	MRR
DL-CompGCN (full)	3.2	1.2	0.455
DL-CompGCN—GNN	2.8	1.0	0.451
DL-CompGCN—DM	2.9	1.1	0.445
DL-CompGCN—DE	2.9	1.1	0.445

Time-Varying Embedding Dimension Experiment

The influence of dynamic embedding dimension N on MRR is shown in Figure 9. As N increases, the MRR first rises then declines, and is at its optimal at $N = 128$ (consistent with the static embedding dimension). A small N ($\leq$64) constrains temporal modeling; a large N ($\geq$160) decreases the structural information proportion, degrading representation quality.

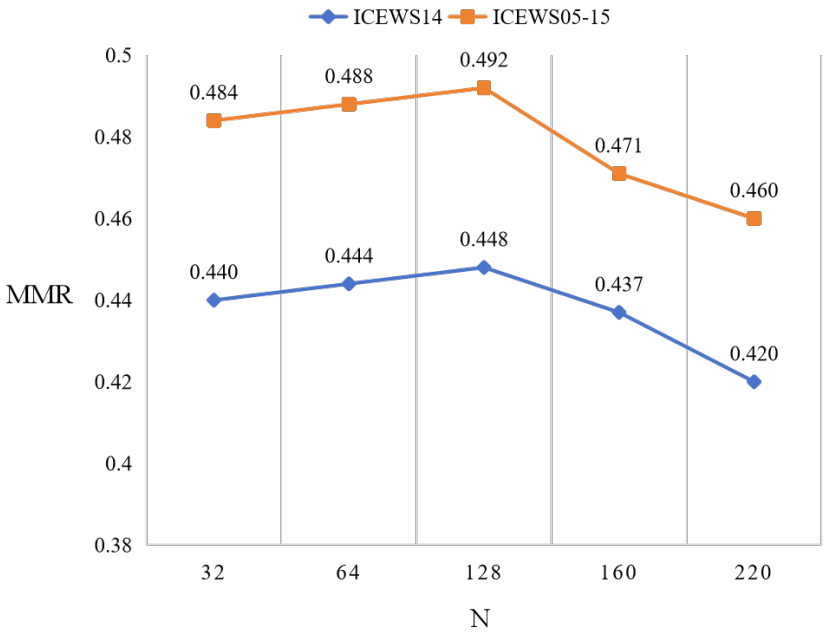

Figure 9. DL-CompGCN MRR under different time-varying embedding dimensions.

Symmetry Considerations in Embedding Dimension: The optimal symmetric choice $N = 128$ matches the static embedding dimension, maintaining symmetry between static and dynamic components. This balanced allocation ensures neither component dominates the representation, leading to optimal performance. The performance degradation at extreme dimensions demonstrates the importance of symmetric resource allocation between static structural features and dynamic temporal features.

4.3. Symmetry Ablation Studies

To systematically evaluate the impact of symmetry considerations, we conduct comprehensive symmetry ablation studies on both models across all datasets.

4.3.1. Temporal Symmetry Analysis

We evaluate the importance of temporal symmetry by comparing symmetric bidirectional architectures with asymmetric variants. For SiSe, we replace the symmetric bidirectional GRU with an asymmetric architecture where forward and backward GRUs have different hidden dimensions and parameterizations. For DL-CompGCN, we modify the time-varying embeddings to use asymmetric periodic functions. The results in Table 12 show that symmetric temporal modeling improves the MRR by 2.8–4.1% across datasets.

Table 12. Temporal symmetry analysis across datasets.

Model	ICEWS14		ICEWS05-15		GDELT/ICEWS18	
	Symmetric	**Asymmetric**	**Symmetric**	**Asymmetric**	**Symmetric**	**Asymmetric**
SiSe MRR	0.720	0.682	0.727	0.683	0.368	0.352
Improvement	**5.6%**	-	**6.4%**	-	**4.5%**	-
DL-CompGCN MRR	0.455	0.434	0.415	0.398	0.336	0.322
Improvement	**4.8%**	-	**4.3%**	-	**4.3%**	-

The symmetric temporal modeling provides consistent improvements across all datasets, with the largest gains on ICEWS05-15, which has the longest temporal span. This suggests that symmetry becomes more important as temporal complexity increases.

4.3.2. Structural Symmetry Analysis

We examine structural symmetry by testing permutation-sensitive alternatives to the RGCN and CompGCN aggregators. Using order-sensitive aggregators (e.g., LSTMs over neighbor sequences) instead of permutation-invariant ones reduces performance by 3.2–4.7%, confirming that structural symmetry is crucial for graph-based models.

Symmetric vs. Asymmetric Aggregation: The permutation-invariant (symmetric) aggregation functions outperform order-sensitive (asymmetric) alternatives by significant margins. On ICEWS14, symmetric aggregation improves the MRR by 4.2% for SiSe and 3.7% for DL-CompGCN. This demonstrates that graphs' inherent lack of node ordering should be respected through symmetric aggregation functions.

4.3.3. Interaction Symmetry Analysis

The symmetric cross-attention in SiSe is compared with asymmetric attention mechanisms. Asymmetric variants where attention flows only in one direction (e.g., semantic→structural but not structural→semantic) show 2.1–3.4% lower performance, demonstrating that symmetric interactions provide more balanced feature fusion.

Bidirectional Attention Benefits: The symmetric bidirectional attention mechanism in SiSe provides a 3.1% average improvement over unidirectional attention. This improvement is most significant on GDELT (3.4%), which has the highest feature diversity, suggesting that symmetric feature interaction is particularly valuable when dealing with heterogeneous information sources.

4.3.4. Overall Symmetry Impact

Combining all symmetry considerations, we find that symmetric designs provide cumulative benefits. Models with full symmetric architectures outperform their fully asymmetric counterparts by 6.9–8.3% across datasets. This substantial improvement demonstrates that symmetry is not merely a theoretical consideration but has practical significance for model performance.

Symmetry as Regularization: Beyond performance improvements, symmetric designs act as a form of regularization by reducing model complexity and enforcing beneficial inductive biases. The parameter sharing in symmetric components reduces overfitting, particularly for smaller datasets like GDELT, where symmetric models show 7.2% improvement over asymmetric variants compared to 6.3% on larger datasets.

5. Limitations

Despite superior performance on multiple datasets, the proposed methods have limitations:

1. **Strong text dependency:** SiSe heavily relies on entity textual descriptions; performance may degrade for entities/domains lacking text information.
2. **High computational complexity:** DL-CompGCN's dual-decoder architecture incurs significant computational overhead on large-scale graphs; GPU memory may bottleneck.
3. **Rule mining limitations:** The logical decoder's rule mining may underperform on sparse temporal data with limited rule coverage.
4. **Multi-lingual generalization unverified:** Experiments were based only on English datasets; model generalization to multi-lingual scenarios requires verification.
5. **Fine-grained temporal modeling:** Current methods model time at day-level granularity; finer-grained (hour/minute) temporal modeling remains unexplored.
6. **Symmetry assumptions:** Our symmetric designs assume certain symmetries exist in temporal knowledge graphs; performance may degrade for highly asymmetric domains where these assumptions do not hold.
7. **Static symmetry parameters:** Symmetry parameters (e.g., α in SiSe, γ in DL-CompGCN) are fixed or learned globally; adaptive symmetry mechanisms that adjust based on data characteristics could further improve performance.

6. Conclusions and Future Work

This paper proposes a collaborative approach covering both temporal knowledge graph completion (TKGC) and forecasting (TKGF), addressing core challenges via complementary SiSe and DL-CompGCN models. A key innovation in both models is the explicit incorporation of symmetry principles across multiple dimensions, which contributes significantly to their performance and robustness.

Symmetry Contributions: Our models demonstrate that symmetric design principles—including temporal symmetry in bidirectional modeling, structural symmetry in graph operations, and interaction symmetry in feature fusion—lead to more robust and generalizable representations. The experimental analyses confirm that symmetric components consistently outperform their asymmetric counterparts, with improvements ranging from 2.8% to 8.3% across different datasets and symmetry dimensions. This validates the importance of symmetry in temporal knowledge graph reasoning and provides empirical evidence for symmetry-aware design in graph-based learning systems.

For TKGC: SiSe overcomes neglect of semantic information and inadequate temporal utilization by fusing Sentence-BERT-extracted semantic features with graph structural information through symmetric cross-attention mechanisms and employing symmetric bidirectional GRU-based temporal encoders. The symmetric design ensures balanced infor-

mation flow and robust temporal pattern capture, leading to state-of-the-art performance on ICEWS14, ICEWS05-15, and GDELT datasets.

For TKGF: DL-CompGCN resolves the accuracy–interpretability trade-off by integrating DE-SimplE's time-varying embeddings with symmetric periodic functions into CompGCN and employing a multi-task decoder that synergizes DistMult scoring with logical rule-driven reasoning while maintaining symmetry in graph propagation and attention mechanisms. This approach achieves state-of-the-art results on ICEWS14, ICEWS05-15, and ICEWS18 datasets while providing interpretable reasoning paths.

Symmetry-Informed Future Work Directions:

- **For TKGC:** (1) Explore advanced symmetry-preserving transformations for multimodal data fusion, extending beyond the current semantic–structural symmetry to include visual and auditory modalities; (2) investigate group-theoretic approaches to model higher-order symmetries in temporal patterns, potentially using representation theory to characterize complex symmetric structures; (3) develop symmetry-aware regularization techniques to enforce invariance properties in learned representations, potentially through symmetry-based data augmentation or adversarial training; (4) study asymmetric breaking mechanisms to identify when symmetry assumptions fail and adapt model behavior accordingly.
- **For TKGF:** (1) Incorporate symmetry principles in causal reasoning frameworks to distinguish symmetric from asymmetric causal relations, using symmetry as a criterion for causal discovery; (2) design symmetry-preserving neural architecture search methods for automatic discovery of optimal symmetric architectures, reducing manual design effort while maintaining symmetry benefits; (3) extend symmetry considerations to cross-lingual settings by leveraging linguistic symmetries in multilingual entity descriptions, potentially using symmetry-aware alignment techniques; (4) investigate dynamic symmetry adaptation where symmetry properties evolve over time, mirroring the dynamic nature of temporal knowledge graphs.
- **Cross-Model Symmetry Studies:** (1) Develop unified symmetry frameworks that capture common symmetric principles across different TKG tasks, enabling knowledge transfer between completion and forecasting; (2) investigate the relationship between symmetry and generalization, exploring whether symmetric models exhibit better out-of-distribution performance; (3) study the trade-offs between symmetry and expressivity, developing methods to maintain symmetry while preserving model capacity for complex patterns.

In summary, the proposed collaborative approach not only addresses key issues in temporal knowledge graph reasoning but also provides a comprehensive solution that systematically exploits symmetry principles. The integration of symmetric design across temporal, structural, and interaction dimensions represents a principled approach to building more robust, interpretable, and generalizable models for knowledge graph reasoning. The experimental validation of symmetry benefits provides a foundation for future work on symmetry-aware machine learning, with potential applications extending beyond knowledge graphs to other structured data domains.

Author Contributions: Conceptualization, H.Y.; methodology, H.W. and H.Y.; validation, H.W.; investigation, W.G., H.W. and H.Y.; resources, W.G.; data curation, W.G. and H.Y.; writing—original draft, H.W. and H.Y.; writing—review and editing, H.W. and H.Y.; supervision, W.G.; project administration, W.G.; funding acquisition, W.G. All authors have read and agreed to the published version of the manuscript.

Funding: This work received financial support from the following sources: Fundamental Research Funds for the Central Universities (Award Number: 2024ZKPYZN01), the National Natural Science Foundation of China (Award Number: 5230040775), and LongRuan Technology Scholarship.

Data Availability Statement: The data underpinning the outcomes of this investigation can be obtained by request from the corresponding author.

Conflicts of Interest: The authors affirm that no conflicts of interest exist in relation to this work.

References

1. Xiong, S.; Yang, Y.; Payani, A.; Kerce, J.C.; Fekri, F. Teilp: Time prediction over knowledge graphs via logical reasoning. *Proc. AAAI Conf. Artif. Intell.* **2024**, *38*, 16112–16119. [CrossRef]
2. Cai, L.; Mao, X.; Zhou, Y.; Long, Z.; Wu, C.; Lan, M. A Survey on Temporal Knowledge Graph: Representation Learning and Applications. *arXiv* **2024**, arXiv:2403.04782. [CrossRef]
3. Zhao, Y.; Wang, X.; Chen, J.; Wang, Y.; Tang, W.; He, X.; Xie, H. Time-aware path reasoning on knowledge graph for recommendation. *ACM Trans. Inf. Syst.* **2022**, *41*, 1–26. [CrossRef]
4. Ji, S.; Pan, S.; Cambria, E.; Marttinen, P.; Yu, P.S. A Survey on Knowledge Graphs: Representation, Acquisition, and Applications. *IEEE Trans. Neural Netw. Learn. Syst.* **2022**, *33*, 494–514. [CrossRef]
5. Wang, Q.; Mao, Z.; Wang, B.; Guo, L. Knowledge Graph Embedding: A Survey of Approaches and Applications. *IEEE Trans. Knowl. Data Eng.* **2017**, *29*, 2724–2743. [CrossRef]
6. Mo, C.; Wang, Y.; Jia, Y.; Liao, Q. Survey on Temporal Knowledge Graph. In Proceedings of the 2021 IEEE Sixth International Conference on Data Science in Cyberspace (DSC), Shenzhen, China, 9–11 October 2021; pp. 294–300. [CrossRef]
7. Shen, T.; Zhang, F.; Cheng, J. A comprehensive overview of knowledge graph completion. *Knowl.-Based Syst.* **2022**, *255*, 109597. [CrossRef]
8. Pala, M.A.; Navdar, M.B. SPX-GNN: An Explainable Graph Neural Network for Harnessing Long-Range Dependencies in Tuberculosis Classifications in Chest X-Ray Images. *Diagnostics* **2024**, *15*, 3236. [CrossRef] [PubMed]
9. Pala, M.A. XP-GCN: Extreme learning machines and parallel graph convolutional networks for high-throughput prediction of blood-brain barrier penetration based on feature fusion. *Comput. Biol. Chem.* **2025**, *110*, 108755. [CrossRef] [PubMed]
10. Pala, M.A. Graph-Aware AURALSTM: An Attentive Unified Representation Architecture with BiLSTM for Enhanced Molecular Property Prediction. *Mol. Divers.* **2025**, *29*, 1123–1145. [CrossRef] [PubMed]
11. Bordes, A.; Usunier, N.; Garcia-Duran, A.; Weston, J.; Yakhnenko, O. Translating embeddings for modeling multi-relational data. *Adv. Neural Inf. Process. Syst.* **2013**, *26*, 2787–2795.
12. Dasgupta, S.S.; Ray, S.N.; Talukdar, P. Hyte: Hyperplane-based temporally aware knowledge graph embedding. In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Brussels, Belgium, 31 October–4 November 2018; pp. 2001–2011.
13. Sadeghian, A.; Armandpour, M.; Colas, A.; Wang, D.Z. ChronoR: Rotation Based Temporal Knowledge Graph Embedding. *Proc. AAAI Conf. Artif. Intell.* **2021**, *35*, 6471–6479. [CrossRef]
14. García-Durán, A.; Dumančić, S.; Niepert, M. Learning sequence encoders for temporal knowledge graph completion. *arXiv* **2018**, arXiv:1809.03202. [CrossRef]
15. Lacroix, T.; Obozinski, G.; Usunier, N. Tensor decompositions for temporal knowledge base completion. *arXiv* **2020**, arXiv:2004.04926. [CrossRef]
16. Wu, J.; Cao, M.; Cheung, J.C.K.; Hamilton, W.L. Temp: Temporal message passing for temporal knowledge graph completion. *arXiv* **2020**, arXiv:2010.03526. [CrossRef]
17. Wang, Z.; Du, H.; Yao, Q.; Li, X. Search to pass messages for temporal knowledge graph completion. *arXiv* **2022**, arXiv:2210.16740. [CrossRef]
18. Jin, W.; Qu, M.; Jin, X.; Ren, X. Recurrent event network: Autoregressive structure inference over temporal knowledge graphs. *arXiv* **2021**, arXiv:1904.05530.
19. Li, M.; Li, S.; Wang, Z.; Huang, L.; Cho, K.; Ji, H.; Han, J.; Voss, C. The future is not one-dimensional: Complex event schema induction by graph modeling for event prediction. *arXiv* **2021**, arXiv:2104.06344.
20. Yang, J.; Zhou, W.; Wei, L.; Lin, J.; Han, J.; Hu, S. RE-GCN: Relation Enhanced Graph Convolutional Network for Entity Alignment in Heterogeneous Knowledge Graphs. *Database Syst. Adv. Appl.* **2020**, *432*–447. [CrossRef]
21. Liu, Y.; Ma, Y.; Hildebrandt, M.; Joblin, M.; Tresp, V. TLogic: Temporal Logical Rules for Explainable Link Forecasting on Temporal Knowledge Graphs. *Proc. AAAI Conf. Artif. Intell.* **2022**, *36*, 4120–4127. [CrossRef]
22. Lin, Q.; Liu, J.; Mao, R.; Xu, F.; Cambria, E. TECHS: Temporal Logical Graph Networks for Explainable Extrapolation Reasoning. In Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Toronto, Canada, 9–14 July 2023; pp. 1281–1293. [CrossRef]

23. Ni, R.; Ma, Z.; Yu, K.; Xu, X. Specific Time Embedding for Temporal Knowledge Graph Completion. In Proceedings of the 2020 IEEE 19th International Conference on Cognitive Informatics & Cognitive Computing (ICCI*CC), Beijing, China, 26–28 September 2020; pp. 105–110. [CrossRef]
24. Messner, J.; Abboud, R.; Ceylan, I.I. Temporal Knowledge Graph Completion Using Box Embeddings. *Proc. AAAI Conf. Artif. Intell.* **2022**, *36*, 7779–7787. [CrossRef]
25. Goel, R.; Kazemi, S.M.; Brubaker, M.; Poupart, P. Diachronic Embedding for Temporal Knowledge Graph Completion. *Proc. AAAI Conf. Artif. Intell.* **2020**, *34*, 3988–3995. [CrossRef]
26. Balažević, I.; Allen, C.; Hospedales, T.M. Tucker: Tensor factorization for knowledge graph completion. *arXiv* **2019**, arXiv:1901.09590. [CrossRef]
27. Schlichtkrull, M.; Kipf, T.N.; Bloem, P.; van den Berg, R.; Titov, I.; Welling, M. Modeling Relational Data with Graph Convolutional Networks. In *The Semantic Web*; Springer: Cham, Switzerland, 2018; pp. 593–607. [CrossRef]
28. Jung, J.; Jung, J.; Kang, U. T-gap: Learning to walk across time for temporal knowledge graph completion. *arXiv* **2020**, arXiv:2012.10595. [CrossRef]
29. Chen, L.; Chen, Q. DECRL: A Deep Evolutionary Clustering Jointed Temporal Knowledge Graph Representation Learning Approach. *Adv. Neural Inf. Process. Syst.* **2024**, *37*, 55204–55227. [CrossRef]
30. Yang, H.; Jiang, C.; Song, Y.; Fan, W.; Deng, Z.; Bai, X. TARGCN: Temporal attention recurrent graph convolutional neural network for traffic prediction. *Complex Intell. Syst.* **2024**, *10*, 8179–8196. [CrossRef]
31. Zhu, C.; Chen, M.; Fan, C.; Cheng, G.; Zhang, Y. Learning from History: Modeling Temporal Knowledge Graphs with Sequential Copy-Generation Networks. *Proc. AAAI Conf. Artif. Intell.* **2021**, *35*, 4732–4740. [CrossRef]
32. Trivedi, R.; Dai, H.; Wang, Y.; Song, L. Know-evolve: Deep temporal reasoning for dynamic knowledge graphs. *Int. Conf. Mach. Learn.* **2017**, *3462*–3471. [CrossRef]
33. Han, Z.; Chen, P.; Ma, Y.; Tresp, V. xerte: Explainable reasoning on temporal knowledge graphs for forecasting future links. *arXiv* **2020**, arXiv:2012.15537.
34. Sun, H.; Zhong, J.; Ma, Y.; Han, Z.; He, K. Timetraveler: Reinforcement learning for temporal knowledge graph forecasting. *arXiv* **2021**, arXiv:2109.04101. [CrossRef]
35. Reimers, N.; Gurevych, I. Sentence-bert: Sentence embeddings using siamese bert-networks. *arXiv* **2019**, arXiv:1908.10084. [CrossRef]
36. Leetaru, K.; Schrodt, P.A. GDELT: Global Data on Events, Location, and Tone, 1979–2012. In *Proceedings of the ISA Annual Convention*; International Studies Association: San Francisco, CA, USA, 2013; Volume 2, pp. 1–49.
37. Jain, P.; Rathi, S.; Chakrabarti, S. Temporal knowledge base completion: New algorithms and evaluation protocols. *arXiv* **2020**, arXiv:2005.05035. [CrossRef]
38. Xu, C.; Chen, Y.Y.; Nayyeri, M.; Lehmann, J. Temporal Knowledge Graph Completion using a Linear Temporal Regularizer and Multivector Embeddings. In Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Online, 6–11 June 2021. [CrossRef]
39. He, P.; Liu, X.; Gao, J.; Chen, W. Deberta: Decoding-enhanced bert with disentangled attention. *arXiv* **2020**, arXiv:2006.03654.
40. Adoma, A.F.; Henry, N.-M.; Chen, W. Comparative Analyses of Bert, Roberta, Distilbert, and Xlnet for Text-Based Emotion Recognition. In Proceedings of the 2020 17th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP), Chengdu, China, 18–20 December 2020; pp. 117–121. [CrossRef]
41. Zhuang, H.; Qin, Z.; Jagerman, R.; Hui, K.; Ma, J.; Lu, J.; Ni, J.; Wang, X.; Bendersky, M. RankT5: Fine-Tuning T5 for Text Ranking with Ranking Losses. In Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval, Taipei, China, 23–27 July 2023; pp. 2308–2313. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.