Denoising the Future: Context-Aware Spectral Diffusion for Temporal Knowledge Graph Extrapolation

Yanglei Gan[1], Peng He[2,4], Run Lin[3], Peiyuan Jiang[2], Yifan Wang[2], Qiao Liu[2]
[1]Southwest Minzu University, Chengdu, China
[2]University of Electronic Science and Technology of China, Chengdu, China
[3]Zhejiang University, Hangzhou, China, [4]Weixin Group, Tencent, Gunagzhou, China
Correspondence: emmaahe@tencent.com, runlin@zju.edu.cn

Abstract

Temporal Knowledge Graph (TKG) extrapolation seeks to infer future facts from time-varying relational histories. Recent diffusion-based approaches improve uncertainty modeling through generative denoising, but their aggregated conditioning on subject histories may insufficiently distinguish query-specific evidence from non-salient historical facts, thereby diluting target-discriminative signals. To bridge this gap, we propose **FreqDiff**, a **Freq**uency-aware **Diff**usion framework for TKG extrapolation. Specifically, FreqDiff formulates future object prediction as query-slot denoising and develops a dual-stream denoiser that integrates temporal dependency modeling with context-aware spectral calibration. The spectral branch synthesizes history-conditioned filters from learnable bases to adaptively recalibrate denoising representations, while a frequency-domain regularizer is proposed to align the denoised target with the gold object in spectral space. Experiments on four public TKG benchmarks demonstrate that FreqDiff achieves state-of-the-art performance[1].

1 Introduction

Temporal Knowledge Graphs (TKGs) encode time-evolving facts as quadruples (s, r, o, t), where relation r links entities s and o at timestamp t (Ji et al., 2021; Liang et al., 2024). Reasoning over TKGs aims to infer missing or future facts from observed temporal histories. Existing studies typically distinguish between two reasoning settings: **interpolation**, which completes missing facts within the observed time span (Cai et al., 2023; Luo et al., 2024), and **extrapolation**, which predicts facts after the latest observed timestamp (Yang et al., 2015; Zhu et al., 2021; Li et al., 2021). This work centers on extrapolation, which enables forward-looking reasoning and supports decisions about future events.

[1]The source code is available at: https://github.com/AONE-NLP/FreqDiff.

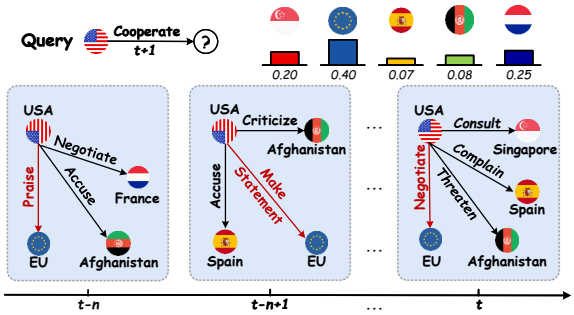


Figure 1: Illustration of subject-oriented context in TKG. Red-linked facts denote query-oriented evidence, while other subject-related facts may act as contextual noise.

Accurate TKG extrapolation requires modeling the temporal evolution of events and relational patterns. Most existing methods follow a learn-to-classify paradigm (Trivedi et al., 2017; Jin et al., 2020), which encodes historical dependencies into deterministic entity and relation representations and ranks candidate future facts with scoring functions such as TransE (Bordes et al., 2013) or DistMult (Yang et al., 2015). Recent work further improves this paradigm through GNN-based historical propagation (Li et al., 2021, 2022b), contrastive learning (Chen et al., 2024c; Xu et al., 2023) over local/global or historical/non-historical contexts, and symbolic temporal priors (Chen et al., 2025c,d) for interpretable reasoning.

Although these methods achieve strong empirical performance, their deterministic prediction mechanism makes it difficult to capture the uncertainty and diversity of future events. To address this limitation, recent studies have introduced diffusion models into TKG reasoning and shifted toward a **learn-to-generate** paradigm (Cai et al., 2024; Gan et al., 2026), where plausible target objects are generated or sampled conditioned on historical temporal contexts. Despite their empirical success, existing methods exhibit two key limitations:

Proceedings of the 2026 Conference on Empirical Methods in Natural Language Processing, pages 46895–46913
October 24-29, 2026 ©2026 Association for Computational Linguistics

- **Limited Discrimination of Query-specific Evidence.** Existing methods typically encode subject-oriented histories through unified temporal processing schemes (Zhang et al., 2023a; Chen et al., 2024a). However, a subject history may contain diverse relational trajectories (Han et al., 2021; Wang et al., 2025a), not all of which are informative for the current query. As shown in Figure 1, for the query <USA, Cooperate, ?, t+1>, accusation- or criticism-related facts are subject-relevant but weakly query-discriminative, whereas negotiation- or consultation-related facts provide more direct evidence. Without query-specific filtering, such non-salient contexts may obscure truly critical temporal evidence.
- **Reliance on Generic Loss Formulations.** Existing diffusion-based TKG reasoning methods mainly supervise denoising with generic reconstruction or ranking losses (Cai et al., 2024; Gan et al., 2026). Although effective for entity discrimination, these objectives constrain the denoised representation mostly in the embedding space and do not explicitly preserve its spectral structure. Consequently, frequency-specific cues important for reconstructing the future object may be insufficiently captured.

In this paper, we propose **FreqDiff**, a **Freq**uency-aware **Diff**usion framework for TKG extrapolation. Given a future query, FreqDiff constructs a subject-oriented event sequence as historical context and formulates missing object prediction as query-slot denoising. In the forward process, Gaussian noise is injected only into the target object representation, while historical events remain deterministic. In the reverse process, a dual-stream denoiser reconstructs the target by combining temporal dependency modeling with context-aware spectral calibration. The temporal branch captures sequential patterns from the subject history, while the spectral branch generates context-aware spectral filters from learnable bases to re-calibrate the denoising representation. Moreover, a frequency-domain consistency regularizer aligns the denoised target with the gold object embedding in spectral space, providing explicit supervision for frequency-aware reconstruction. Our contributions are three-fold:

- We propose FreqDiff, a frequency-aware diffusion framework for TKG extrapolation, which formulates future object prediction as query-slot denoising and reconstructs the target through a dual-stream denoiser with temporal modeling and context-aware spectral calibration.
- FreqDiff introduces a frequency-domain consistency regularizer that aligns the denoised target representation with the gold object embedding in spectral space, providing explicit supervision for frequency-aware reconstruction.
- Extensive experiments on four public TKG benchmarks show that FreqDiff achieves state-of-the-art performance, with further analysis validating the effectiveness of its spectral calibration and frequency-domain regularization.

2 Related Works

2.1 Temporal Knowledge Graph Reasoning

Discriminative TKG reasoning predicts future facts by learning temporal patterns from historical triples. Early continuous-time methods, such as Know-Evolve (Bordes et al., 2013; Chen et al., 2024b) and THCN (Chen et al., 2024b), model event occurrence with Hawkes processes or temporal causal convolution. Later neural approaches incorporate temporal signals into KG encoders through recurrent reasoning, graph-based propagation, cycle-aware constraints, and structural historical evidence, including RE-NET (Jin et al., 2020), RE-GCN (Li et al., 2021), CyGNet (Zhu et al., 2021), CEN (Li et al., 2022c), xERTE (Han et al., 2021), and HisMatch (Li et al., 2022d).

Beyond simply encoding all historical facts, another line of work explicitly improves the selection or bottlenecking of useful historical evidence. For example, xERTE (Han et al., 2021) extracts query-relevant temporal subgraphs with temporal relational attention, TimeTraveler (Sun et al., 2021) searches historical snapshots through reinforcement learning, and CENET-style methods distinguish historical and non-historical dependencies through contrastive learning and masking (Xu et al., 2023; Zhang et al., 2023b). In parallel, symbolic and structural methods improve interpretability and inductive generalization by deriving temporal logical rules from time-consistent random walks (Liu et al., 2022), mining relation-specific paths (Dong et al., 2023), or constructing cognitive temporal relation graphs (Chen et al., 2025d).

2.2 Generative TKG Reasoning

Generative approaches introduce uncertainty-aware modeling into TKG extrapolation. DiffuTKG (Cai et al., 2024) formulates future fact prediction as conditional denoising under a Gaussian diffusion

process, while NADEx (Gan et al., 2026) further incorporates negative-aware diffusion to sharpen decision boundaries for future entities. DPCL-Diff (Cao et al., 2025) extends this line with graph-node diffusion and dual-domain periodic contrastive learning to separate recurrent and novel temporal patterns. Beyond purely neural diffusion frameworks, Luo et al. (Luo et al., 2024) use LLMs to generate multi-step event chains, and LLM-DR (Chen et al., 2025a) combines classifier-free guided diffusion with LLM-based rule refinement. Despite their progress in generative TKG reasoning, they mainly operate in the time-domain, while the frequency properties underlying temporal dynamics remain under-explored. This motivates our frequency-aware diffusion framework, which introduces complementary spectral signals for TKG reasoning. Discussions of diffusion model and frequency modeling are provided in Appendix A.

3 Preliminary

Definition 1. Temporal Knowledge Graph. Let $\mathcal{E}$, $\mathcal{R}$, and $\mathcal{T}$ denote finite sets of entities, relation types, and timestamps, respectively. A temporal knowledge graph $\mathcal{G}$ is a collection of time-stamped quadruples:

$$\mathcal{G} = \Big\{ (s, r, o, t) \;\Big|\; s, o \in \mathcal{E},\ r \in \mathcal{R},\ t \in \mathcal{T} \Big\}, \quad (1)$$

where each tuple encodes the fact that relation r holds from subject s to object o at time t. More specifically, the TKG can be viewed as an ordered sequence of static snapshots:

$$\mathcal{G} = \{\mathcal{G}_1, \mathcal{G}_2, \ldots, \mathcal{G}_{|\mathcal{T}|}\}, \quad (2)$$

where $\mathcal{G}_t$ aggregates all triples that are valid at timestamp t. Following the standard bidirectional–relation convention (Kazemi and Poole, 2018), we augment every quadruple (s, r, o, t) with its inverse (o, r^{-1}, s, t), where r^{-1} is a distinct relation denoting the reverse semantics of r.

Definition 2. Temporal Knowledge Graph Reasoning. Let $q = (s, r, ?, t)$ be a query quadruple whose object entity is missing at timestamp t. Given the sliding history window of length L, $\mathcal{G}_{t-L-1:t-1} = \{\mathcal{G}_{t-L}, \mathcal{G}_{t-L+1}, \ldots, \mathcal{G}_{t-1}\}$. The TKG reasoning task is to learn a scoring function:

$$\text{score}_t(o) = f(s, r, o, \mathcal{G}), \hat{o} = \arg\max_{o \in \mathcal{E}} \text{score}_t(o). \quad (3)$$

Each candidate object $o \in \mathcal{E}$ is assigned a score and the highest-scoring entity completes the quadruple.

4 Method

4.1 Temporal Representation Learning

Given a query $q = (s, r_q, ?, t)$, the goal is to predict the missing object from the entity set $\mathcal{E}$ based on the recent history of the subject s. We first construct a subject-centric sequence comprising the L most recent historical events alongside the query slot:

$$\mathcal{Q}_{s,t} = \big[(s_1, r_1, o_1, \tau_1), \ldots, (s_L, r_L, o_L, \tau_L)\big]. \quad (4)$$

Following the standard sequential formulation, we project the object, relation, and relative-time components of $\mathcal{Q}_{s,t}$ into a shared h-dimensional space. Let $\mathbf{E}_o \in \mathbb{R}^{|\mathcal{E}| \times h}$, $\mathbf{E}_r \in \mathbb{R}^{|\mathcal{R}| \times h}$, and $\mathbf{E}_{\Delta t} \in \mathbb{R}^{N_t \times h}$ denote the object, relation, and relative-time embedding matrices, respectively:

$$\begin{aligned} \mathbf{o} &= [\mathbf{E}_o(o_1); \ldots; \mathbf{E}_o(o_L); \mathbf{E}_o(o_t)], \\ \mathbf{r} &= [\mathbf{E}_r(r_1); \ldots; \mathbf{E}_r(r_L); \mathbf{E}_r(r_q)], \\ \mathbf{t} &= [\mathbf{E}_{\Delta t}(\Delta\tau_1); \ldots; \mathbf{E}_{\Delta t}(\Delta\tau_L); \mathbf{E}_{\Delta t}(1)]. \end{aligned} \quad (5)$$

where $\Delta\tau_i$ represents the temporal interval between the i-th historical event and the query timestamp t.

4.2 Forward Diffusion Process

The forward process injects Gaussian noise into the target object representation at the query position. For a sampled diffusion step $m \in \{1, \ldots, M\}$, the corrupted target representation is generated as:

$$\mathbf{o}_m = \sqrt{\bar{\alpha}_m}\, \mathbf{E}_o(o_t) + \sqrt{1 - \bar{\alpha}_m}\, \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim \mathcal{N}(0, \mathbf{I}), \quad (6)$$

where $\bar{\alpha}_m = \prod_{j=1}^{m}(1 - \beta_j)$ denotes the remaining signal ratio at step m. In practice, a linearly scaled accumulated noise schedule is used:

$$1 - \bar{\alpha}_m = \delta \cdot \left(\alpha_{\min} + \frac{m-1}{M-1}(\alpha_{\max} - \alpha_{\min}) \right). \quad (7)$$

where $\delta \in [0, 1]$ is a global scaling factor that moderates the overall diffusion strength, and $\alpha_{\min}$, $\alpha_{\max}$ bound the noise levels.

4.3 Dual-stream Denoiser

During reverse denoising, the model reconstructs the target entity by conditioning on the relational and temporal context of the recent event trajectory, as shown in Figure 2. To keep the conditioning context deterministic, stochastic corruption is applied only to the query target representation, while the historical object representations remain unchanged. The denoising input is constructed as:

$$\begin{aligned} \widetilde{\mathbf{O}}_m &= [\mathbf{E}_o(o_1); \ldots; \mathbf{E}_o(o_L); \mathbf{o}_m], \\ \mathbf{X}_m &= \text{LN}\big(\text{Dropout}(\widetilde{\mathbf{O}}_m + \mathbf{R} + \mathbf{T})\big), \\ \mathbf{H}_m &= \text{LN}\big(\mathbf{X}_m + \text{Emb}(m)\big). \end{aligned} \quad (8)$$

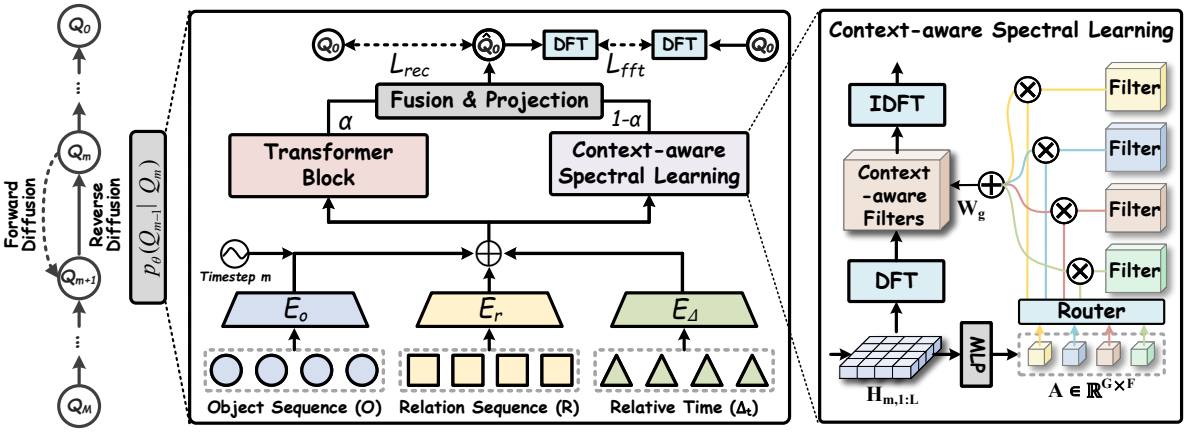


Figure 2: Overview of FreqDiff. Given a subject-oriented history, FreqDiff injects Gaussian noise into the target object representation and reconstructs it through reverse diffusion. The denoiser combines temporal modeling with context-aware spectral calibration, where history-conditioned filters re-calibrate frequency components and fuse them with time-domain representations for future object prediction.

where $\widetilde{\mathbf{O}}_m$ contains the observed historical object embeddings at the first L positions and the diffused target representation $\mathbf{o}_m$ at the query position. The relation sequence $\mathbf{R}$, relative-time sequence $\mathbf{T}$, and diffusion-step embedding $\mathrm{Emb}(m)$ are then incorporated to form $\mathbf{H}_m$. The resulting representation is processed by two complementary branches.

Temporal Dependency Learning. The temporal branch explicitly models sequential dependencies by processing the conditioned hidden states through a standard Transformer encoder:

$$\mathbf{H}_m^{\text{time}} = \text{Transformer}(\mathbf{H}_m). \tag{9}$$

Context-aware Spectral Filtering. The spectral branch complements the temporal branch by explicitly modeling frequency-domain variations in the subject-oriented event sequence, which may be less directly preserved by time-domain attention. To generate a context-aware spectral response, the historical representations are summarized by mean pooling and mapped to routing coefficients:

$$\begin{aligned} \mathbf{c} &= \text{Mean-Pool}(\mathbf{H}_{m,1:L}) \in \mathbb{R}^h, \\ \mathbf{A} &= \tanh\big(\text{MLP}(\mathbf{c})\big) \in \mathbb{R}^{G\times F}, \end{aligned} \tag{10}$$

where G denotes the number of feature groups and F denotes the number of learnable basis filters. Let $\{\mathbf{B}_f\}_{f=1}^F$ be the set of spectral basis filters, where each basis filter satisfies $\mathbf{B}_f \in \mathbb{R}^{K\times(h/G)}$, and K is the number of valid frequency bins produced by FFT. For each feature group g, the context-aware spectral filter is synthesized as:

$$\mathbf{W}_g = \sum_{f=1}^{F} \mathbf{A}_{g,f}\mathbf{B}_f, \quad \mathbf{W}_g \in \mathbb{R}^{K\times(h/G)} \tag{11}$$

The routing coefficients $\mathbf{A}_{g,f}$ instantiate a context-aware spectral filter by linearly combining the learnable basis filters. The resulting filter $\mathbf{W}_g$ is applied to the group-wise spectrum of $\mathbf{H}_m$ which is split along the feature dimension into $\mathcal{G}$ groups:

$$\begin{aligned} \mathcal{Z}'_{\text{re},g} &= \text{Re}\left(\mathcal{F}(\mathbf{H}_m)_g\right) \odot \mathbf{W}_g, \\ \mathcal{Z}'_{\text{im},g} &= \text{Im}\left(\mathcal{F}(\mathbf{H}_m)_g\right) \odot \mathbf{W}_g. \end{aligned} \tag{12}$$

where $\text{Re}(\cdot)$ and $\text{Im}(\cdot)$ denote the real and imaginary spectral components. $g \in \{1,\ldots,G\}$ indexes the feature group, and $\odot$ denotes element-wise multiplication. The calibrated groups are concatenated along the feature dimension and mapped back to the time domain via inverse FFT:

$$\mathbf{H}_m^{\text{freq}} = \mathcal{F}^{-1}(\mathcal{Z}'_{\text{re}} + j\cdot\mathcal{Z}'_{\text{im}}). \tag{13}$$

Temporal-spectral Fusion. The temporal and spectral representations are fused by interpolation:

$$\mathbf{H}_m^{\text{fuse}} = \text{LN}\left(\alpha\mathbf{H}_m^{\text{time}} + (1-\alpha)\mathbf{H}_m^{\text{freq}}\right). \tag{14}$$

where $\alpha \in [0,1]$ is a fusion coefficient. The target representation is obtained from the query position:

$$\hat{\mathbf{o}}_0 = \mathbf{H}_{m,L+1}^{\text{fuse}}. \tag{15}$$

4.4 Training Objective

The model is trained with a reconstruction objective on the denoised query representation, together with an auxiliary frequency-domain regularizer.

Reconstruction Loss. Given the denoised target $\hat{\mathbf{o}}_0$, all candidate entities are scored by dot-product matching with the entity embedding matrix $\mathbf{E}_o \in \mathbb{R}^{|\mathcal{E}|\times h}$. The reconstruction loss is defined as the negative log-likelihood of the gold object o_t:

$$\mathcal{L}_{\text{rec}} = -\log\left(\text{Softmax}\left(\frac{\mathbf{E}_o\hat{\mathbf{o}}_0}{\sqrt{h}}\right)_{o_t}\right), \tag{16}$$

Frequency-Domain Regularization. To further constrain the denoising process, a frequency-domain consistency loss is imposed between the denoised target representation and the gold object embedding. Specifically, FFT is applied along the feature dimension of both representations:

$$\mathcal{L}_{\text{fft}} = \ell\Big(\operatorname{Re}\big(\mathcal{F}(\hat{\mathbf{o}}_0)\big), \operatorname{Re}\big(\mathcal{F}(\mathbf{E}_o(o_t))\big)\Big) + \ell\Big(\operatorname{Im}\big(\mathcal{F}(\hat{\mathbf{o}}_0)\big), \operatorname{Im}\big(\mathcal{F}(\mathbf{E}_o(o_t))\big)\Big), \tag{17}$$

where $\ell(\cdot,\cdot)$ denotes a point-wise distance function[2]. This regularizer encourages the denoised representation to align with the gold object not only in the embedding space, but also in its feature-domain spectral profile, providing an explicit constraint for frequency-aware denoising.

The overall training objective is then denoted as:

$$\mathcal{L} = \mathcal{L}_{\text{rec}} + \lambda \mathcal{L}_{\text{fft}}. \tag{18}$$

where λ controls the strength of the auxiliary frequency-domain regularization.

4.5 Inference

During inference, FreqDiff initialize the unknown target representation with Gaussian noise. A standard reverse diffusion sampler would apply the denoising network $f_\theta(\cdot)$ step by step from M to 0, but this iterative process increases computational cost. Since $f_\theta(\cdot)$ is trained to recover the clean target representation from a corrupted representation o_m at any diffusion step m, we adopt an efficient inference strategy (Cai et al., 2024; Gan et al., 2026) that directly predicts the clean target from the maximum-noise representation o_M, without performing all intermediate reverse transitions:

$$\begin{aligned} \widetilde{\mathbf{O}}_M &= [\mathbf{E}_o(o_1); \ldots; \mathbf{E}_o(o_L); \mathbf{o}_M], \\ \mathbf{H}_M &= \widetilde{\mathbf{O}}_M + \mathbf{R} + \mathbf{T} + \operatorname{Emb}(M). \end{aligned} \tag{19}$$

The learned denoising network then directly predicts the clean target representation:

$$\hat{\mathbf{o}}_0 = f_\theta(\mathbf{H}_M, M). \tag{20}$$

5 Experiments

5.1 Experimental Setups

Datasets. Our experiments employ four benchmark datasets, including ICEWS14, ICEWS05-15, ICEWS18, and GDELT, to evaluate the proposed model. Specifically, the ICEWS datasets originate from the Integrated Crisis Early Warning System (Boschee et al., 2015), while the GDELT dataset is sourced from the Global Database of Events, Language, and Tone (Leetaru and Schrodt, 2013). The data statistics are summarized in Appendix C.1.

Baseline Models. We benchmark FreqDiff against three sets of approaches: **Static methods**: DistMult (Yang et al., 2015), ConvE (Dettmers et al., 2018), RotatE(Sun et al., 2019); **Interpolation methods**: TTransE (Leblay and Chekol, 2018), TA-DistMult (Garcia-Duran et al., 2018), DE-SimpIE (Goel et al., 2020); **Extrapolation methods**: RE-NET (Jin et al., 2020), Re-GCN (Li et al., 2021), CEN (Li et al., 2022c), TiRGN (Li et al., 2022b), TITer (Sun et al., 2021), RETIA (Liu et al., 2023b), CENET (Xu et al., 2023), THCN (Chen et al., 2024b), DiffuTKG (Cai et al., 2024), LogiQ (Chen et al., 2025c), CognTKE (Chen et al., 2025d), NADEx (Gan et al., 2026). We provide detailed baseline descriptions in Appendix C.3.

Evaluation Metrics. To measure temporal extrapolation performance, we cast the task as masked entity prediction, where either the subject or object is held out in quadruples of the form $(s, r, ?, t)$ or $(?, r, o, t)$. Predictions are scored and ranked, and we report Mean Reciprocal Rank (MRR) alongside Hits@1, Hits@3, and Hits@10. All results are computed under the time-aware filtering protocol.

Implementation Details. All models are optimized with Adam and trained for 100 epochs. The learning rate is set to $1e^{-3}$ on ICEWS14 and ICEWS18, and $5e^{-4}$ on ICEWS05-15 and GDELT. Entity and relation embeddings are both initialized with a dimensionality of 200. Experiments are conducted on a single NVIDIA A100 GPU with 80GB memory. Hyper-parameter configurations are provided in Appendix C.2. *The reported results are averaged with five runs with different seeds.*

5.2 Overall Performance

Table 1 summarizes FreqDiff's performance against state-of-the-art (SOTA) baselines across four benchmark datasets. From these results, we make the following key observations:

- **FreqDiff achieves the best results across all four datasets and all 16 evaluation metrics**, with relative improvements ranging from 1.14% to 16.14% over the strongest baseline. This consistent superiority demonstrates that incorporating frequency-aware modeling into the diffusion framework leads to broadly effective temporal

[2]Here we utilize point-wise L1 distance.

Table 1: Performance comparison (%) on four benchmarks with MRR and Hits@1/3/10. Best and second-best results are shown in **bold** and underlined, respectively. ♠ indicates results re-implemented using official code.

Models	ICEWS14				ICEWS18				ICEWS05-15				GDELT			
	MRR	Hit@1	Hit@3	Hit@10	MRR	Hit@1	Hit@3	Hit@10	MRR	Hit@1	Hit@3	Hit@10	MRR	Hit@1	Hit@3	Hit@10
DisMult (2015)	15.44	10.91	17.24	23.92	11.51	7.03	12.87	20.86	17.95	13.12	20.71	29.32	8.68	5.58	9.96	17.13
ConvE (2018)	35.09	25.23	39.38	54.68	24.51	16.23	29.25	44.51	33.81	24.78	39.00	54.95	16.55	11.02	18.88	31.60
RotatE (2019)	21.31	10.26	24.35	44.75	12.78	4.01	14.89	31.91	24.71	13.22	29.04	48.16	13.45	6.95	14.09	25.99
TTransE (2018)	13.72	2.98	17.70	35.74	8.31	1.92	8.56	21.89	15.57	4.80	19.24	38.29	5.50	0.47	4.94	15.25
TA-DisMult (2018)	25.80	16.94	29.74	42.99	16.75	8.61	18.41	33.59	24.31	14.58	27.92	44.21	12.00	5.76	12.94	23.54
DE-SimIE (2020)	33.36	24.85	37.15	48.92	19.30	11.53	21.86	34.80	35.02	25.91	38.99	52.75	19.70	12.22	21.39	33.70
RE-NET (2020)	36.93	26.83	39.51	54.78	28.81	19.05	32.44	47.51	43.32	33.43	47.77	63.06	19.62	12.42	21.00	34.01
RE-GCN (2021)	40.39	30.66	44.96	59.21	30.58	21.01	34.34	48.75	48.03	37.33	53.85	68.27	19.64	12.42	20.90	33.69
CyGNet (2021)	35.05	25.73	39.01	53.55	24.93	15.90	28.28	42.61	36.81	26.61	41.63	56.22	18.48	11.52	19.57	31.98
TITer (2021)	41.73	32.74	46.46	58.44	29.98	22.05	33.46	44.83	47.69	37.95	52.92	65.81	15.46	10.98	15.61	24.31
CEN (2022c)	42.20	32.08	47.46	61.31	31.50	21.70	35.44	50.59	46.84	36.38	52.45	67.01	20.39	12.96	21.77	34.97
TiRGN (2022b)	44.04	33.83	48.95	63.84	33.66	23.19	37.99	54.22	50.04	39.25	56.13	70.71	21.67	13.63	23.27	37.60
RETIA (2023b)	42.76	32.28	47.77	62.75	32.43	22.23	36.48	52.94	47.26	36.64	52.90	67.76	20.12	12.76	21.45	34.49
CENET (2023)	39.02	29.62	43.23	57.49	27.85	18.15	31.63	46.98	41.95	32.17	46.93	60.43	20.23	12.69	21.70	34.92
THCN (2024b)	45.39	36.58	50.84	66.07	35.63	24.90	39.26	56.76	51.94	40.32	57.79	72.18	23.46	15.18	25.21	39.03
DiffuTKG♠ (2024)	47.58	36.38	53.41	66.01	35.65	25.19	39.39	59.55	48.97	39.80	56.92	69.84	21.35	14.43	23.68	36.05
LogiQ (2025c)	44.71	35.72	51.03	64.21	34.94	24.76	39.57	56.32	51.04	40.71	57.55	71.00	–	–	–	–
CognTKE (2025d)	46.06	36.49	51.11	64.49	35.24	25.21	39.93	54.71	53.13	42.62	59.42	72.70	–	–	–	–
NADEx♠ (2026)	48.12	37.89	55.26	70.55	35.37	25.48	40.27	58.69	52.17	43.38	60.47	71.93	21.78	14.69	23.37	37.17
FreqDiff	**50.85**	**39.82**	**57.48**	**71.48**	**36.85**	**26.85**	**41.51**	**61.47**	**54.71**	**45.02**	**61.16**	**74.45**	**25.04**	**17.64**	**27.03**	**41.83**
Improve.	5.67%	5.09%	4.01%	1.32%	3.37%	5.37%	3.08%	3.22%	2.97%	3.78%	1.14%	2.40%	6.73%	16.20%	7.22%	7.17%

Table 2: Ablation study results ICEWS14 and ICEWS18 datasets in terms of MRR and Hit@1/10.

Settings	ICEWS14			ICEWS18		
	MRR	Hit@1	Hit@10	MRR	Hit@1	Hit@10
FreqDiff	**50.85**	**39.82**	**71.48**	**36.85**	**26.85**	**61.47**
w/o. $Freq$	48.68	38.94	70.64	35.81	26.04	58.79
w/o. $Time$	42.14	32.37	65.39	18.21	11.76	35.52
w/o. $\mathcal{L}_{\text{fft}}$	49.30	38.84	69.82	36.10	26.13	59.47
w. ℓ_2 Distance	50.21	39.13	70.77	35.14	25.53	59.22
w. global filter	48.70	39.03	70.58	34.42	24.93	58.01
w. static filter	48.61	38.89	69.81	32.07	23.10	55.18
w. maxpool	49.06	39.46	70.01	34.05	24.84	57.78

knowledge graph extrapolation.

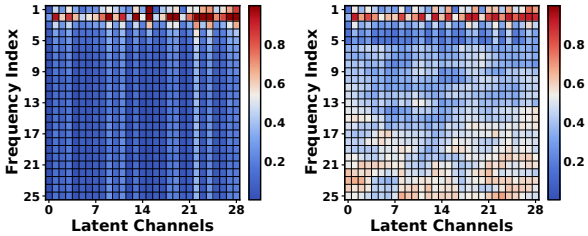
- **The largest improvement appears on GDELT**, especially with a 16.14% gain on Hit@1 and a 6.73% gain on MRR. Since GDELT involves more complex temporal evolution patterns, these gains suggest that frequency-aware modeling is beneficial in challenging extrapolation scenarios. This further supports that spectral information provides complementary signals beyond conventional temporal modeling.
- **Static and interpolation-based methods lag behind extrapolation-oriented models**, confirming that forward-looking temporal reasoning is essential for predicting future facts. Although interpolation models can exploit temporal information within observed histories, they are not explicitly optimized for future event evolution, which limits their ability in extrapolation settings.

5.3 Ablation Studies

We validate the contribution of each FreqDiff component by comparing it against seven variants:

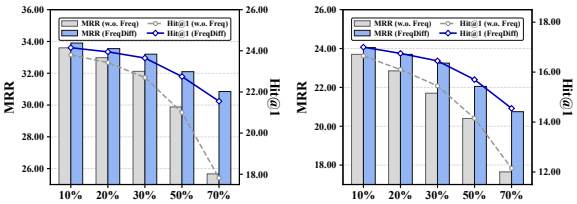
- **w/o. Freq:** omits frequency-domian modeling.
- **w/o. Time:** no time-domain modeling.
- **w/o. $\mathcal{L}_{\text{fft}}$:** removes frequency-domain loss.
- **w. ℓ_2 Distance**: replaces ℓ_1 distance in the frequency-domain regularizer with ℓ_2 distance.
- **w. global filter:** replaces context-aware spectral filter with a shared global filter.
- **w. static filter:** replaces context-aware spectral filter with a static filter.
- **w. maxpool:** replaces mean-pool with maxpool.

As shown in Table 2, both time-domain and frequency-domain designs are necessary for FreqDiff. Removing the frequency branch (w/o. $Freq$) causes consistent drops, confirming that spectral information provides useful complementary signals. Removing the time branch (w/o. $Time$) leads to the most severe degradation, indicating that time-domain modeling remains the core component for preserving event order and relation-specific temporal context. In contrast, frequency modeling works as a complementary signal. The drop after removing $\mathcal{L}_{\text{fft}}$ further verifies the value of explicit spectral supervision. Moreover, replacing the $\ell 1$ distance with ℓ_2 leads to inferior results, suggesting that ℓ_1 provides a more robust constraint for spectral alignment. Finally, replacing the context-aware filter with global or static filters consistently weakens performance, suggesting that different event histories require adaptive frequency responses rather than shared or fixed filtering. The degradation caused by max-pooling further shows that, within the frequency branch, mean-pooling better preserves the overall spectral distribution,



(a) FreqDiff w/o. Freq (b) FreqDiff

Figure 3: Visualization of learned spectral energy distributions on ICEWS14. The y-axis represents frequency indices, and the x-axis represents latent channels. Red indicates higher spectral energy.



(a) Impact on ICEWS18 (b) Impact on GDELT

Figure 4: Performance under different ratios of sequence corruption on ICEWS18 and GDELT.

whereas max-pooling may overemphasize dominant frequency components and suppress weaker but informative spectral patterns.

5.4 Analysis of Spectral Learning

We further examine how spectral learning contributes to contextual representation learning and performance under sequence corruption.

Visualization of Spectral Learning. Figure 3 visualizes the learned spectral energy distributions of FreqDiff and its variant without frequency modeling. The variant without frequency modeling shows a relatively concentrated response pattern, suggesting that its denoising representation is mainly dominated by the overall temporal patterns in the subject history. In contrast, FreqDiff exhibits a more adaptive spectral response across frequency indices and latent channels. Rather than uniformly amplifying spectral responses, it modulates the frequency-channel components of the denoising representation, attenuating generic temporal signals while retaining components more aligned with the queried relation and the target object. In this way, spectral learning introduces finer query-conditioned temporal variations into the denoising process. Additional visualized comparisons with other diffusion-based models are provided in Appendix D.1.

Table 3: Effectiveness of the frequency regularizer. Red superscripts indicate the improvement rates.

Dataset	Metric	DiffuTKG		NADEx		CENET	
		Base	+ $\mathcal{L}_{\text{fft}}$	Base	+ $\mathcal{L}_{\text{fft}}$	Base	+ $\mathcal{L}_{\text{fft}}$
ICEWS14	MRR	47.58	$48.67^{+2.29\%}$	48.12	$49.03^{+1.89\%}$	39.02	$43.39^{+11.20\%}$
	H@1	36.38	$37.73^{+3.71\%}$	37.89	$38.97^{+2.85\%}$	29.62	$32.08^{+8.31\%}$
ICEWS18	MRR	35.65	$36.24^{+1.65\%}$	35.37	$36.04^{+1.89\%}$	27.85	$29.96^{+7.58\%}$
	H@1	25.19	$26.36^{+4.64\%}$	25.48	$26.13^{+2.55\%}$	18.15	$20.85^{+14.88\%}$
GDELT	MRR	21.35	$22.90^{+7.26\%}$	21.78	$23.18^{+6.43\%}$	20.23	$21.90^{+8.26\%}$
	H@1	14.43	$15.72^{+8.94\%}$	14.69	$15.45^{+5.17\%}$	12.69	$14.01^{+10.40\%}$

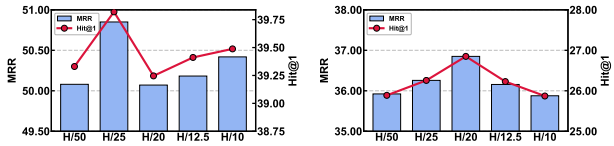
Effectiveness under Noise Condition. Figure 4 further evaluates the effectiveness of FreqDiff when the historical context is partially corrupted. Specifically, for each test query, we randomly select a given proportion (ranging from 10% to 70%) of historical events in the subject-centric sequence and perturb their event representations while keeping the trained model unchanged. As the corruption ratio increases, the performance of both FreqDiff and its variant without spectral learning declines, indicating that reliable historical context is important for TKG extrapolation. Nevertheless, FreqDiff consistently outperforms the variant without spectral modeling on both ICEWS18 and GDELT in terms of MRR and Hits@1, with a more evident advantage under higher corruption ratios. These results suggest that spectral learning improves robustness under corrupted historical observations. By introducing a frequency-aware filter, FreqDiff re-calibrates temporal-frequency representations, attenuating corrupted responses while preserving reliable denoising signals.

5.5 Generalization Analysis

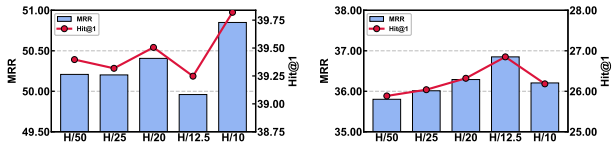
Effectiveness of Frequency Regularizer. Table 3 validates the effectiveness and *plug-and-play* applicability of the frequency-domain regularizer across different backbone models and datasets. Without changing the original architectures or training settings, incorporating $\mathcal{L}_{\text{fft}}$ consistently improves DiffuTKG, NADEx, and CENET on ICEWS14, ICEWS18, and GDELT under both MRR and H@1. Specifically, $\mathcal{L}_{\text{fft}}$ improves DiffuTKG, NADEx, and CENET by 3.73%/5.76%, 3.40%/3.52%, and 9.01%/11.20% on average in terms of MRR/H@1, respectively. These improvements confirm that penalizing frequency-domain inconsistency provides complementary guidance to conventional objectives, leading to more discriminative representations for event prediction.

Table 4: Performance of predicting unseen events in terms of MRR and Hit@1 on ICEWS14 and ICEWS18.

Models	ICEWS14		ICEWS18	
	MRR	Hit@1	MRR	Hit@1
RE-GCN	23.26	13.91	15.08	7.09
CEN	22.06	13.28	15.41	8.20
RETIA	24.17	14.67	16.62	9.08
HisMatch	27.49	19.04	17.51	11.13
DiffuTKG	25.22	15.23	16.48	8.84
NADEx	29.71	19.34	19.52	12.17
FreqDiff	**31.42**	**21.33**	**20.33**	**12.55**
Improve.	*5.75%*	*10.29%*	*4.15%*	*3.12%*



(a) Impact of G



(b) Impact of F

Figure 5: Hyper-parameter sensitivity analysis on ICEWS14 (left) and ICEWS18 (right) datasets.

Effectiveness on Unseen Events. To examine FreqDiff's generalization to unseen temporal relational facts, we evaluate unseen event prediction on ICEWS14 and ICEWS18. As shown in Table 4, FreqDiff achieves the best results across all metrics, outperforming the strongest baseline NADEx by 5.75%/10.29% on ICEWS14 and 4.15%/3.12% on ICEWS18 in terms of MRR/Hit@1. The consistent gains over conventional temporal reasoning methods and diffusion-based baselines suggest that frequency-aware modeling offers complementary signals for unseen event prediction. By exploiting spectral patterns beyond time-domain histories, FreqDiff better distinguishes plausible future facts from unseen candidates.

5.6 Sensitivity Analysis

Figure 5 shows that FreqDiff is generally stable across different hyper-parameter settings. For the feature groups G, the best results appear around H/25 on ICEWS14 and H/20 on ICEWS18, suggesting that moderate grouping provides a better balance between flexible spectral calibration and stable representation learning. For the basis number F, performance remains robust, while larger or moderate basis sets usually work better, indicating that multiple spectral bases help capture diverse frequency patterns. Overall, the results indicate that FreqDiff is not overly sensitive to hyper-parameters, and its best performance comes from a balanced temporal-spectral configuration. Additional analysis on fusion coefficient α and loss balance term λ is provided in Appendix D.2.

Table 5: Computational efficiency test on ICEWS14/18.

Model	ICEWS14		ICEWS18	
	Inf. Time	Params.	Inf. Time	Params.
RE-GCN	28.96s	26.52Mb	393.67s	42.68Mb
TiRGN	32.88s	43.35Mb	159.90s	59.59Mb
NADEx	10.91s	16.30Mb	96.95s	32.42Mb
FreqDiff	**7.28s**	17.00Mb	**81.24s**	32.83Mb

5.7 Computational Efficiency

Table 5 compares the inference efficiency and parameter scale of different models on ICEWS14 and ICEWS18 on a single NVIDIA A100 GPU under their optimal settings. FreqDiff achieves the fastest inference on both datasets, reducing inference time from 10.91s to 7.28s on ICEWS14 and from 96.95s to 81.24s on ICEWS18 compared with NADEx, corresponding to reductions of 33.27% and 16.20%, respectively. Meanwhile, FreqDiff only introduces a slight parameter increase over NADEx, from 16.30Mb to 17.00Mb on ICEWS14 and from 32.42Mb to 32.83Mb on ICEWS18. These results show that the frequency-aware design improves predictive performance while maintaining competitive efficiency, achieving a favorable balance among accuracy, inference speed, and model size.

5.8 Comparison with LLM-based Forecasters

As shown in Table 6, FreqDiff achieves the best result on 11 of the 16 dataset–metric combinations, including all four metrics on ICEWS14, three metrics on ICEWS05-15, H@10 on ICEWS18, and H@1/H@3/H@10 on GDELT. However, it is not uniformly superior. CRI performs best on ICEWS05-15 H@10 and ICEWS18 MRR/H@3, while LANTERN obtains the strongest ICEWS18 H@1 and GDELT MRR.

The results also suggest complementary strengths. Diffusion-based FreqDiff is particularly competitive when prediction depends on distributed temporal patterns and when maintaining broad candidate coverage is important, as reflected by its consistent H@10 performance. Its learned

Table 6: Performance comparison (%) with LLM-based forecasters. The best results are highlighted in **bold**.

Model	ICEWS14				ICEWS05-15				ICEWS18				GDELT			
	MRR	H@1	H@3	H@10	MRR	H@1	H@3	H@10	MRR	H@1	H@3	H@10	MRR	H@1	H@3	H@10
LLM-DA (2024)	47.10	36.90	52.60	67.10	52.10	41.60	58.60	72.80	35.70	25.50	40.30	57.00	–	–	–	–
MESH (2025)	44.36	–	49.81	64.21	48.66	–	54.26	68.57	33.96	–	38.37	54.12	–	–	–	–
ANRE (2025)	47.40	36.90	51.10	65.70	50.90	39.10	58.00	69.60	35.50	26.00	39.20	56.70	24.30	16.60	26.60	37.50
TV-LLM (2025)	44.50	36.60	50.20	64.20	50.80	41.30	56.60	72.20	33.20	22.90	37.90	54.10	–	–	–	–
CRI (2026)	49.80	38.10	55.10	70.20	53.10	40.70	59.30	**75.10**	**38.80**	27.40	**42.10**	57.80	–	–	–	–
LANTERN (2026)	48.00	37.50	51.50	66.50	51.50	40.50	58.50	70.50	36.50	**28.50**	40.00	57.50	**25.50**	17.50	27.00	38.00
FreqDiff	**50.85**	**39.82**	**57.48**	**71.48**	**54.71**	**45.02**	**61.16**	74.45	36.85	26.85	41.51	**61.47**	25.04	**17.64**	**27.03**	**41.83**

Table 7: Performance comparison on YAGO and WIKI.

Model	YAGO				WIKI			
	MRR	H@1	H@3	H@10	MRR	H@1	H@3	H@10
TITer	87.47	80.09	89.96	90.27	73.91	71.70	75.41	76.93
TiRGN	87.95	84.34	91.37	92.92	81.65	77.77	85.12	87.08
DiffuTKG	88.29	84.36	91.79	93.55	82.21	78.96	85.69	88.03
NADEx	88.31	85.14	92.18	93.72	82.83	79.02	85.83	88.34
w/o Freq	86.65	84.78	89.24	89.90	82.55	81.43	84.89	86.38
w/o $\mathcal{L}_{\text{fft}}$	88.74	86.62	91.87	91.23	84.08	83.03	86.92	87.99
FreqDiff	**90.32**	**88.60**	**93.15**	**93.91**	**85.84**	**84.76**	**88.27**	**89.83**

denoising representation can aggregate noisy or heterogeneous histories without requiring an explicit symbolic rule to cover each query. LLM-based approaches may be preferable when a query has strong symbolic or semantic support, such as a high-confidence temporal rule, a closely matched historical analogy, or a compact set of highly informative events that can be expressed in the prompt. This is consistent with the design of LLM-DA, TV-LLM, and CRI, which rely on temporal rule induction or validation, and with AnRe and LANTERN, which emphasize analogical demonstrations and carefully selected historical evidence. We provide detailed model descriptions in Appendix C.3.

5.9 Performance on Knowledge-Centric KGs

Table 7 further evaluates FreqDiff on YAGO and WIKI, whose knowledge-centric facts and temporal patterns differ from the event-driven ICEWS and GDELT datasets. FreqDiff consistently outperforms all competing methods across the eight evaluation settings. Compared with the strongest baseline, NADEx, FreqDiff improves MRR and Hit@1 by 2.01 and 3.46 percentage points on YAGO, respectively. The improvements are more pronounced on WIKI, reaching 3.01 points in MRR and 5.74 points in Hit@1. FreqDiff also achieves consistent gains in Hit@3 and Hit@10 on both datasets. These results demonstrate that its effectiveness generalizes beyond geopolitical event forecasting to knowledge-centric temporal graphs.

The ablation results further confirm the contributions of the frequency-aware components. Removing the spectral branch decreases MRR by 3.67 points on YAGO and 3.29 points on WIKI, accompanied by consistent degradation across all Hits metrics. Removing $\mathcal{L}_{\text{fft}}$ also reduces MRR by 1.58 and 1.76 points, respectively. The larger degradation caused by removing the spectral branch highlights the importance of context-aware spectral modeling, while the consistent decline without $\mathcal{L}_{\text{fft}}$ verifies the complementary role of frequency-domain supervision. Together, these results show that both components remain effective across TKGs with distinct temporal characteristics.

6 Conclusion

In this paper, we proposed FreqDiff, a frequency-aware diffusion framework for TKG extrapolation. FreqDiff formulates future object prediction as query-slot denoising and reconstructs the target representation with a dual-stream denoiser that combines temporal dependency modeling and context-aware spectral calibration. We further introduced a frequency-domain consistency regularizer to provide explicit spectral supervision for target reconstruction. Experiments on four public TKG benchmarks demonstrate that FreqDiff achieves state-of-the-art performance, while ablation studies verify the effectiveness of spectral calibration and frequency-domain regularization.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (U22B2061), the Intelligent Perception and Control Key Laboratory of Sichuan Province (2025IPCZ02), the Natural Science Foundation of Sichuan (2026NSFSC1483), and Sichuan Science and Technology Program.

Limitation

This work has two main limitations. First, although we evaluate FreqDiff on four widely used TKG benchmarks, including ICEWS14, ICEWS05–15, ICEWS18, and GDELT, these datasets are primarily centered on political and international relations events. Therefore, the generalizability of FreqDiff to other temporal knowledge graphs, such as those in scientific discovery, financial transactions, public health, or natural disasters, remains to be further examined. Second, FreqDiff introduces context-aware spectral calibration to improve frequency-aware denoising, but its current design relies on a finite set of learnable basis filters. While effective in our experiments, this design may still provide limited flexibility when modeling highly irregular or domain-specific temporal dynamics. Future work may explore more adaptive spectral parameterization strategies and evaluate frequency-aware diffusion reasoning across broader TKG domains.

Ethics Statement

This study follows established ethical standards. We use only publicly available benchmark datasets that have been collected and processed by prior research, and our work does not involve new data collection, human-subject interaction, or the use of private personal information. The proposed model is developed for scientific analysis and benchmarking in temporal knowledge graph reasoning. It is not intended for surveillance, deception, profiling, or any harmful application. We respect the data-use terms of the adopted benchmarks and aim to ensure that the research does not compromise the rights, safety, or dignity of any individual or group.

References

Antoine Bordes, Nicolas Usunier, Alberto Garcia-Duran, Jason Weston, and Oksana Yakhnenko. 2013. Translating embeddings for modeling multi-relational data. *Advances in neural information processing systems*, 26.

Elizabeth Boschee, Jennifer Lautenschlager, Sean O'Brien, Steve Shellman, James Starz, and Michael Ward. 2015. ICEWS Coded Event Data.

Borui Cai, Yong Xiang, Longxiang Gao, He Zhang, Yunfeng Li, and Jianxin Li. 2023. Temporal knowledge graph completion: a survey. In *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence*, pages 6545–6553.

Mu Cai, Hong Zhang, Huijuan Huang, Qichuan Geng, Yixuan Li, and Gao Huang. 2021. Frequency domain image translation: More photo-realistic, better identity-preserving. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 13930–13940.

Yuxiang Cai, Qiao Liu, Yanglei Gan, Changlin Li, Xueyi Liu, Run Lin, Da Luo, and Jiaye Yang. 2024. Predicting the unpredictable: Uncertainty-aware reasoning over temporal knowledge graphs via diffusion process. In *Findings of the Association for Computational Linguistics: ACL 2024*, pages 5766–5778, Bangkok, Thailand. Association for Computational Linguistics.

Yukun Cao, Lisheng Wang, and Luobin Huang. 2025. Dpcl-diff: Temporal knowledge graph reasoning based on graph node diffusion model with dual-domain periodic contrastive learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pages 14806–14814.

Bin Chen, Chunjing Xiao, and Fan Zhou. 2024a. Natural evolution-based dual-level aggregation for temporal knowledge graph reasoning. In *Findings of the association for computational linguistics: EMNLP 2024*, pages 9274–9284.

Kai Chen, Xin Song, Ye Wang, Liqun Gao, Aiping Li, Xiaojuan Zhao, Bin Zhou, and Yalong Xie. 2025a. Llm-dr: A novel llm-aided diffusion model for rule generation on temporal knowledge graphs. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pages 11481–11489.

Linwei Chen, Lin Gu, and Ying Fu. 2025b. Frequency-dynamic attention modulation for dense prediction. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 22620–22632.

Tingxuan Chen, Jun Long, Zidong Wang, Shuai Luo, Jincai Huang, and Liu Yang. 2024b. Thcn: A hawkes process based temporal causal convolutional network for extrapolation reasoning in temporal knowledge graphs. *IEEE Transactions on Knowledge and Data Engineering*.

Tingxuan Chen, Liu Yang, Zidong Wang, Shuai Luo, and Jun Long. 2025c. Enhancing extrapolation reasoning on temporal knowledge graphs with logic rules and queries. In *ICASSP 2025-2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1–5. IEEE.

Wei Chen, Huaiyu Wan, Yuting Wu, Shuyuan Zhao, Jiayaqi Cheng, Yuxin Li, and Youfang Lin. 2024c. Local-global history-aware contrastive learning for temporal knowledge graph reasoning. In *2024 IEEE 40th International Conference on Data Engineering (ICDE)*, pages 733–746. IEEE.

Wei Chen, Yuting Wu, Shuhan Wu, Zhiyu Zhang, Mengqi Liao, Youfang Lin, and Huaiyu Wan. 2025d.

Cogntke: A cognitive temporal knowledge extrapolation framework. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pages 14815–14823.

Yimin Deng, Yuxia Wu, Yejing Wang, Guoshuai Zhao, Li Zhu, Qidong Liu, Derong Xu, Zichuan Fu, Xian Wu, Zhao-Xiangyu Zheng, Yefeng, and Xueming Qian. 2025. A multi-expert structural-semantic hybrid framework for unveiling historical patterns in temporal knowledge graphs. In *Findings of the Association for Computational Linguistics: ACL 2025*, pages 20553–20565.

Tim Dettmers, Pasquale Minervini, Pontus Stenetorp, and Sebastian Riedel. 2018. Convolutional 2d knowledge graph embeddings. In *Proceedings of the AAAI conference on artificial intelligence*, volume 32.

Prafulla Dhariwal and Alexander Nichol. 2021. Diffusion models beat gans on image synthesis. *Advances in neural information processing systems*, 34:8780–8794.

Hao Dong, Zhiyuan Ning, Pengyang Wang, Ziyue Qiao, Pengfei Wang, Yuanchun Zhou, and Yanjie Fu. 2023. Adaptive path-memory network for temporal knowledge graph reasoning. In *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence*, pages 2086–2094.

Pierre Duhamel and Martin Vetterli. 1990. Fast fourier transforms: a tutorial review and a state of the art. *Signal processing*, 19(4):259–299.

Matteo Frigo and Steven G Johnson. 2005. The design and implementation of fftw3. *Proceedings of the IEEE*, 93(2):216–231.

Yanglei Gan, Peng He, Yuxiang Cai, Run Lin, Guanyu Zhou, and Qiao Liu. 2026. Negative-aware diffusion process for temporal knowledge graph extrapolation. In *Findings of the Association for Computational Linguistics: EACL 2026*, pages 3352–3367.

Alberto Garcia-Duran, Sebastijan Dumančić, and Mathias Niepert. 2018. Learning sequence encoders for temporal knowledge graph completion. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 4816–4821.

Rishab Goel, Seyed Mehran Kazemi, Marcus Brubaker, and Pascal Poupart. 2020. Diachronic embedding for temporal knowledge graph completion. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, pages 3988–3995.

Shansan Gong, Mukai Li, Jiangtao Feng, Zhiyong Wu, and Lingpeng Kong. 2022. Diffuseq: Sequence to sequence text generation with diffusion models. In *The Eleventh International Conference on Learning Representations*.

Shansan Gong, Mukai Li, Jiangtao Feng, Zhiyong Wu, and Lingpeng Kong. 2023. Diffuseq-v2: Bridging discrete and continuous text spaces for accelerated seq2seq diffusion models. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, pages 9868–9875.

Zhen Han, Peng Chen, Yunpu Ma, and Volker Tresp. 2021. Explainable subgraph reasoning for forecasting on temporal knowledge graphs. In *International Conference on Learning Representations*.

Peng He, Yanglei Gan, Tingting Dai, Run Lin, Xuexin Li, Yao Liu, and Qiao Liu. 2026a. Exploiting inter-session information with frequency-enhanced dual-path networks for sequential recommendation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pages 14820–14828.

Peng He, Yao Liu, Yanglei Gan, Run Lin, Yuxiang Cai, and Qiao Liu. 2026b. Fait: Frequency-aware inverted transformer for multivariate time series forecasting. In *Proceedings of the 32nd ACM SIGKDD Conference on Knowledge Discovery and Data Mining V. 2*, pages 1614–1625.

Shaoxiong Ji, Shirui Pan, Erik Cambria, Pekka Marttinen, and S Yu Philip. 2021. A survey on knowledge graphs: Representation, acquisition, and applications. *IEEE transactions on neural networks and learning systems*, 33(2):494–514.

Chengyuan Jin, Ao Chang, Daojian Zeng, Wenhao Teng, Xiangwen Liao, Kang Liu, Jun Zhao, and Yubo Chen. 2026. Lantern in the event stream: Training-free temporal knowledge graph forecasting by balancing inertia and shifts. In *Findings of the Association for Computational Linguistics: ACL 2026*, pages 11519–11533.

Woojeong Jin, Meng Qu, Xisen Jin, and Xiang Ren. 2020. Recurrent event network: Autoregressive structure inferenceover temporal knowledge graphs. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 6669–6683.

Seyed Mehran Kazemi and David Poole. 2018. Simple embedding for link prediction in knowledge graphs. *Advances in neural information processing systems*, 31.

Zhifeng Kong, Wei Ping, Jiaji Huang, Kexin Zhao, and Bryan Catanzaro. 2020. Diffwave: A versatile diffusion model for audio synthesis. In *International Conference on Learning Representations*.

Julien Leblay and Melisachew Wudage Chekol. 2018. Deriving validity time in knowledge graph. In *Companion proceedings of the the web conference 2018*, pages 1771–1776.

James Lee-Thorp, Joshua Ainslie, Ilya Eckstein, and Santiago Ontanon. 2022. Fnet: Mixing tokens with fourier transforms. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 4296–4313.

Kalev Leetaru and Philip A. Schrodt. 2013. Gdelt: Global data on events, location, and tone. *ISA Annual Convention*, 2:1–49.

Jiang Li, Xiangdong Su, and Guanglai Gao. 2023. Teast: Temporal knowledge graph embedding via archimedean spiral timeline. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 15460–15474.

Xiang Li, John Thickstun, Ishaan Gulrajani, Percy S Liang, and Tatsunori B Hashimoto. 2022a. Diffusion-lm improves controllable text generation. *Advances in neural information processing systems*, 35:4328–4343.

Yujia Li, Shiliang Sun, and Jing Zhao. 2022b. Tirgn: Time-guided recurrent graph network with local-global historical patterns for temporal knowledge graph reasoning. In *IJCAI*, pages 2152–2158.

Zixuan Li, Saiping Guan, Xiaolong Jin, Weihua Peng, Yajuan Lyu, Yong Zhu, Long Bai, Wei Li, Jiafeng Guo, and Xueqi Cheng. 2022c. Complex evolutional pattern learning for temporal knowledge graph reasoning. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 290–296.

Zixuan Li, Zhongni Hou, Saiping Guan, Xiaolong Jin, Weihua Peng, Long Bai, Yajuan Lyu, Wei Li, Jiafeng Guo, and Xueqi Cheng. 2022d. 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*, SIGIR '21, page 408–417, New York, NY, USA. Association for Computing Machinery.

Ke Liang, Lingyuan Meng, Meng Liu, Yue Liu, Wenxuan Tu, Siwei Wang, Sihang Zhou, Xinwang Liu, Fuchun Sun, and Kunlun He. 2024. A survey of knowledge graph reasoning on graph types: Static, dynamic, and multi-modal. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

Haohe Liu, Zehua Chen, Yi Yuan, Xinhao Mei, Xubo Liu, Danilo Mandic, Wenwu Wang, and Mark D Plumbley. 2023a. Audioldm: Text-to-audio generation with latent diffusion models. In *International Conference on Machine Learning*, pages 21450–21474. PMLR.

Kangzheng Liu, Feng Zhao, Guandong Xu, Xianzhi Wang, and Hai Jin. 2023b. Retia: relation-entity twin-interact aggregation for temporal knowledge graph extrapolation. In *2023 IEEE 39th international conference on data engineering (ICDE)*, pages 1761–1774. IEEE.

Yushan Liu, Yunpu Ma, Marcel Hildebrandt, Mitchell Joblin, and Volker Tresp. 2022. Tlogic: Temporal logical rules for explainable link forecasting on temporal knowledge graphs. In *Proceedings of the AAAI conference on artificial intelligence*, volume 36, pages 4120–4127.

Ziyang Liu and Chaokun Wang. 2025. Terdy: Temporal relation dynamics through frequency decomposition for temporal knowledge graph completion. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 9611–9622.

Ruilin Luo, Tianle Gu, Haoling Li, Junzhe Li, Zicheng Lin, Jiayi Li, and Yujiu Yang. 2024. Chain of history: Learning and forecasting with llms for temporal knowledge graph completion. *CoRR*.

Alexander Quinn Nichol, Prafulla Dhariwal, Aditya Ramesh, Pranav Shyam, Pamela Mishkin, Bob Mcgrew, Ilya Sutskever, and Mark Chen. 2022. Glide: Towards photorealistic image generation and editing with text-guided diffusion models. In *International Conference on Machine Learning*, pages 16784–16804. PMLR.

Yingsong Ning, Fu Zhang, Jingwei Cheng, Jiashun Peng, and Xiaoke Wang. 2026. Critic rule induction: Improving temporal knowledge graph forecasting with generator-critic language models. In *Findings of the Association for Computational Linguistics: ACL 2026*, pages 29436–29448.

Qihong Pan, Limin Yao, Guojiang Shen, Xiao Han, Yichuan Chen, and Xiangjie Kong. 2025. Leveraging temporal validity of rules via llms for enhanced temporal knowledge graph reasoning. *Knowledge-based systems*, page 114094.

Ali Sadeghian, Mohammadreza Armandpour, Anthony Colas, and Daisy Zhe Wang. 2021. Chronor: Rotation based temporal knowledge graph embedding. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, pages 6471–6479.

Yongliang Shen, Kaitao Song, Xu Tan, Dongsheng Li, Weiming Lu, and Yueting Zhuang. 2023. Diffusionner: Boundary diffusion for named entity recognition. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 3875–3890.

Jascha Sohl-Dickstein, Eric Weiss, Niru Maheswaranathan, and Surya Ganguli. 2015. Deep unsupervised learning using nonequilibrium thermodynamics. In *International conference on machine learning*, pages 2256–2265. pmlr.

Haohai Sun, Jialun Zhong, Yunpu Ma, Zhen Han, and Kun He. 2021. Timetraveler: Reinforcement learning

for temporal knowledge graph forecasting. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 8306–8319.

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*.

Alex Tamkin, Dan Jurafsky, and Noah Goodman. 2020. Language through a prism: A spectral approach for multiscale language representations. *Advances in Neural Information Processing Systems*, 33:5492–5504.

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.

Yuki Tatsunami and Masato Taki. 2024. Fft-based dynamic token mixer for vision. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 15328–15336.

Rakshit Trivedi, Hanjun Dai, Yichen Wang, and Le Song. 2017. Know-evolve: Deep temporal reasoning for dynamic knowledge graphs. In *international conference on machine learning*, pages 3462–3471. PMLR.

Jiapu Wang, Sun Kai, Linhao Luo, Wei Wei, Yongli Hu, Alan Wee-Chung Liew, Shirui Pan, and Baocai Yin. 2024. Large language models-guided dynamic adaptation for temporal knowledge graph reasoning. *Advances in Neural Information Processing Systems*, 37:8384–8410.

Wenjie Wang, Yiyan Xu, Fuli Feng, Xinyu Lin, Xiangnan He, and Tat-Seng Chua. 2023. Diffusion recommender model. In *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 832–841.

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, Suzhou, China. Association for Computational Linguistics.

Yulong Wang, Yushuo Liu, Xiaoyi Duan, and Kai Wang. 2025b. Filterts: Comprehensive frequency filtering for multivariate time series forecasting. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pages 21375–21383.

Chengjin Xu, Mojtaba Nayyeri, Fouad Alkhoury, Hamed Shariat Yazdi, and Jens Lehmann. 2020. Tero: A time-aware knowledge graph embedding via temporal rotation. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 1583–1593.

Yi Xu, Junjie Ou, Hui Xu, and Luoyi Fu. 2023. Temporal knowledge graph reasoning with historical contrastive learning. In *Proceedings of the AAAI conference on artificial intelligence*, volume 37, pages 4765–4773.

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

Zhengyi Yang, Jiancan Wu, Zhicai Wang, Xiang Wang, Yancheng Yuan, and Xiangnan He. 2023. Generate what you prefer: Reshaping sequential recommendation via guided diffusion. *Advances in Neural Information Processing Systems*, 36:24247–24261.

Mengqi Zhang, Yuwei Xia, Qiang Liu, Shu Wu, and Liang Wang. 2023a. Learning latent relations for temporal knowledge graph reasoning. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 12617–12631.

Mengqi Zhang, Yuwei Xia, Qiang Liu, Shu Wu, and Liang Wang. 2023b. Learning long-and short-term representations for temporal knowledge graph reasoning. In *Proceedings of the ACM web conference 2023*, pages 2412–2422.

Cunchao Zhu, Muhao Chen, Changjun Fan, Guangquan Cheng, and Yan Zhang. 2021. Learning from history: Modeling temporal knowledge graphs with sequential copy-generation networks. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, pages 4732–4740.

A Additional Related Work

A.1 Spectral Representation Learning

Spectral analysis, most commonly implemented with the Discrete Fourier Transform (DFT), decomposes signals into frequency components for efficient processing (Duhamel and Vetterli, 1990; Frigo and Johnson, 2005). Motivated by the convolution theorem, recent deep models incorporate spectral transforms to capture global dependencies and improve efficiency. This idea has been adopted in computer vision (Cai et al., 2021; Tatsunami and Taki, 2024; Chen et al., 2025b), natural language processing (Tamkin et al., 2020; Lee-Thorp et al., 2022), sequential recommendation (He et al., 2026a), and time-series forecasting (Wang et al., 2025b; He et al., 2026b).

In the realm of Temporal Knowledge Graph (TKG) reasoning, explicit spectral learning remains relatively underexplored, although several earlier TKG Completion (TKGC) models can be regarded as important precursors. Diachronic Embedding introduces sine-based temporal entity functions, and its expressivity analysis explicitly relates this parameterization to Fourier sine series (Goel et al., 2020). Along a related line, TeRo (Xu et al., 2020) and ChronoR (Sadeghian et al., 2021) model temporal evolution through rotations in complex or high-dimensional embedding spaces, while TeAST (Li et al., 2023) maps relations onto an Archimedean spiral timeline. These methods capture phase-sensitive, periodic, or geometrically regular temporal patterns, but they mainly encode temporal regularity through parametric embedding functions rather than explicitly transforming representations into the frequency domain. More recently, TeRDy (Liu and Wang, 2025) establishes a more direct connection to spectral learning by applying FFT-based low-pass and high-pass decomposition to relation embeddings, thereby separating long-term and short-term temporal relation dynamics.

A.2 Diffusion Models on Discrete Data

Diffusion models (DMs) (Sohl-Dickstein et al., 2015) have become a powerful generative paradigm, achieving strong performance in image generation (Dhariwal and Nichol, 2021; Nichol et al., 2022) and audio synthesis (Kong et al., 2020; Liu et al., 2023a). Although early diffusion models were mainly designed for continuous Euclidean spaces, recent studies have extended them to discrete symbolic data. For text generation, Diffusion-LM (Li et al., 2022a) maps word tokens into continuous embeddings and performs denoising in the latent space, while DiffuSeq (Gong et al., 2022, 2023) introduces a sequence-level corruption and denoising process to support coherent non-autoregressive generation. Diffusion has also been adapted to structured prediction tasks. DiffusionNER (Shen et al., 2023) formulates named entity recognition as a span-boundary denoising problem, gradually refining noisy boundaries into valid entity predictions.

Beyond text and structured prediction, diffusion models have been applied to symbolic interaction modeling. In recommendation, DiffRec (Wang et al., 2023) and DreamRec (Yang et al., 2023) inject noise into user–item interaction histories and learn the reverse process to capture uncertain preference distributions. These developments suggest that diffusion is not limited to dense continuous signals, but can also provide a flexible generative mechanism for discrete data, where iterative denoising helps model latent structure, uncertainty, and complex dependencies.

B Preliminary

Definition 3. Discrete Fourier Transform. The Discrete Fourier Transform (DFT) is a fundamental tool in digital signal processing. Given a length N time-domain sequence $x[n]$, the DFT maps it to the frequency domain via:

$$\mathcal{X}[k] = \sum_{n=0}^{N-1} x[n] e^{-j2\pi kn/N}, \quad k = 0, 1, ..., N-1, \tag{21}$$

where j is the imaginary unit and $\mathcal{X}[k]$ is the complex spectral coefficient associated with the discrete frequency $\omega_k = 2\pi k/N$. Each $\mathcal{X}[k]$ can be decomposed into real and imaginary parts:

$$\begin{aligned} \mathcal{X}[k] &= \textbf{Real}(\mathcal{X}[k]) + j\textbf{Imag}(\mathcal{X}[k]), \\ \textbf{Real}(\mathcal{X}[k]) &= \sum_{n=0}^{N-1} x[n] \cos\left(\frac{2\pi}{N} kn\right), \\ \textbf{Imag}(\mathcal{X}[k]) &= -\sum_{n=0}^{N-1} x[n] \sin\left(\frac{2\pi}{N} kn\right). \end{aligned} \tag{22}$$

The inverse DFT (IDFT) reconstructs the original sequence via:

$$x[n] = \frac{1}{N} \sum_{k=0}^{N-1} \mathcal{X}[k] e^{j2\pi kn/N}, \quad n = 0, 1, \ldots, N-1. \tag{23}$$

In short, we denote DFT and IDFT operators as $\mathcal{F}$, $\mathcal{F}^{-1}$, respectively.

Table 8: The statistics of the datasets. $|E|$ and $|R|$ denote the number of unique entities and event types, respectively.

Datasets	$\|E\|$	$\|R\|$	*Train*	*Valid*	*Test*	*Unseen Ratio*
ICEWS14	6,869	230	74,845	8,514	7,371	58.43%
ICEWS18	23,033	256	373,018	45,995	49,545	55.69%
ICEWS05-15	10,094	251	368,868	46,302	46,159	39.82%
GDELT	7,691	240	1,734,399	238,765	305,241	43.72%

Table 9: Implementation details for each benchmark.

Hyperparameter	**ICEWS14**	**ICEWS18**	**GDELT**	**ICEWS05-15**
Hidden dimension	200	200	200	200
Dropout	0.2	0.2	0.2	0.2
Embedding dropout	0.2	0.2	0.2	0.2
Maximum history length	128	128	128	128
Maximum input length	64	64	64	64
Number of epochs	100	100	100	100
Learning rate	$1e^{-3}$	$1e^{-3}$	$5e^{-4}$	$5e^{-4}$
Diffusion steps	200	200	200	200
α	0.7	0.7	0.7	0.7
G	8	10	4	8
F	20	8	8	10
λ	0.3	0.3	0.3	0.3
FFT loss type	L1	L1	L1	L1

C Experimental Setup

C.1 Dataset Statistics

To ensure consistency and comparability with prior TKG extrapolation studies, we follow the chronological splitting protocol adopted by (Li et al., 2021; Cai et al., 2024; Gan et al., 2026), where the earliest 80% of facts are used for training, the subsequent 10% for validation, and the latest 10% for testing. This protocol preserves the temporal order of observed facts and avoids information leakage from future timestamps during model training.

C.2 Implementation Details

Table 9 summarizes the main implementation settings. For all datasets, the hidden dimension is set to 200, with both dropout and embedding dropout fixed at 0.2. The maximum history length and input length are set to 128 and 64, respectively, and all models are trained for 100 epochs with 200 diffusion steps. Most hyperparameters are kept consistent across datasets to ensure fair comparison, while the learning rate and filterbank configuration are adjusted according to dataset characteristics. Specifically, ICEWS14 and ICEWS18 use a learning rate of $1e^{-3}$, whereas GDELT and ICEWS05-15 adopt a smaller learning rate of $5e^{-4}$. The fusion weight α and regularization weight λ are fixed at 0.7 and 0.3, respectively, and the FFT regularization adopts an L1 loss for all datasets.

C.3 Baselines

Static Baselines:

- DistMult (Yang et al., 2015), employs a bilinear scoring function, modeling triple plausibility via a relation matrix that captures pairwise interactions between subject and object embeddings.
- ConvE (Dettmers et al., 2018), applies 2D convolution over reshaped entity and relation embeddings, followed by a projection layer to learn richer feature interactions for link prediction.
- RotatE (Sun et al., 2019), represents relations as complex-valued rotations in the embedding space, enabling the model to naturally encode and infer diverse relational patterns such as symmetry and inversion.

Interpolation Baselines:

- TTransE (Leblay and Chekol, 2018), explicitly models temporal dynamics by embedding entities and relations within a continuous time framework, using translation operations along the time dimension to capture their evolution.
- TA-DistMult (Garcia-Duran et al., 2018), extends the DistMult scoring function with time-aware embeddings, allowing the model to adapt relation parameters according to temporal context.
- DE-SimplE (Goel et al., 2020), employs diachronic embeddings that parameterize entity and relation representations as functions of time, hence modeling progressive change across different timestamps.

Extrapolation Baselines:

- RE-NET (Jin et al., 2020), integrates recurrent neural architectures with graph convolution to capture the sequential evolution of entities and predict future links.
- Re-GCN (Li et al., 2021), employs a Recurrent Evolutionary GCN that recurrently updates entity and relation embeddings at each timestamp by propagating temporal signals through the KG.
- CyGNet (Zhu et al., 2021), models cyclical temporal patterns, enabling the learning of periodic behaviors and extrapolate yet-unseen links.
- TITer (Sun et al., 2021), applies hierarchical transformations to entity embeddings, iteratively tracking their evolution to anticipate future TKG states.
- CEN (Li et al., 2022c), uses length-aware convolutional filters to extract multi-scale evolutionary

patterns, with an online training strategy to handle temporal variability.

- TiRGN (Li et al., 2022b), leverages recurrent graph networks to encode dynamic relational structures, improving inference of unseen facts.
- RETIA (Liu et al., 2023b), constructs a twin hyper-relation subgraph and evolutionarily aggregates adjacent entity and relation features for enriched message passing.
- CENET (Xu et al., 2023), incorporates contrastive learning objectives to strengthen dynamic representation learning within temporal graphs.
- THCN (Chen et al., 2024b), introduces temporal causal convolutional networks grounded in Hawkes processes to distinguish the relative importance of concurrent facts.
- DiffuTKG (Cai et al., 2024), reframes TKG reasoning as a denoising diffusion process over entity embeddings to generate future links.
- LogiQ (Chen et al., 2025c), augments diffusion-based generation with logical constraints, improving both accuracy and interpretability.
- CognTKE (Chen et al., 2025d), integrates cognitively symbolic priors into embedding, enabling more transparent and reliable extrapolation.
- NADEx (Gan et al., 2026), incorporates negative-aware diffusion to better distinguish plausible future facts from spurious candidates.

LLM-based Forecasters:

- LLM-DA (Wang et al., 2024), generates temporal rules using LLMs and dynamically updates them based on recent events.
- MESH (Deng et al., 2025), employs multiple expert modules to integrate structural and semantic information for future link prediction.
- AnRe (Tang et al., 2025), combines long- and short-term historical contexts with LLM-generated analogical demonstrations to support temporal reasoning.
- TV-LLM (Pan et al., 2025), models the validity of LLM-generated rules and integrates rule-based retrieval with graph-based candidate scoring.
- CRI (Ning et al., 2026), adopts a generator-critic framework that uses fact-grounded evaluation to validate LLM-induced rules before reasoning.
- LANTERN (Jin et al., 2026), constructs reasoning prompts by balancing long-term interaction strength and short-term novelty, complemented by structure-aware analogical demonstrations.

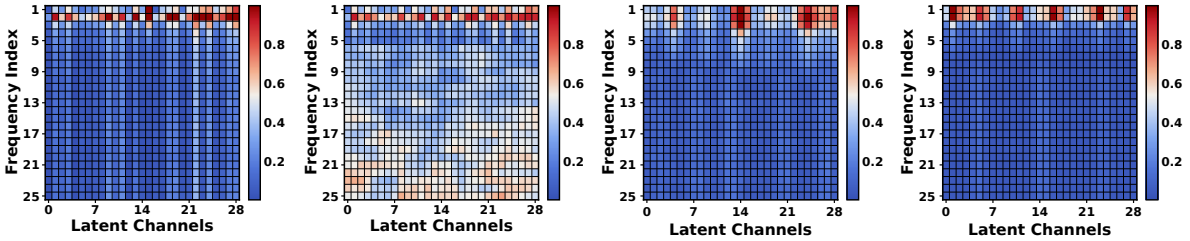
D Experimental Analysis

D.1 Spectral Energy Distributions

Figure 6 shows that the four variants learn markedly different spectral energy distributions on ICEWS14. In FreqDiff w/o. $Freq$, the energy is highly concentrated in the lowest frequency band, while most frequency–channel positions remain weakly activated, indicating that the model mainly relies on dominant low-frequency temporal signals and lacks sufficient capacity to capture diverse evolutionary patterns. DiffuTKG and NADEx exhibit a similar tendency, where spectral responses are sparse and mostly confined to a few low-frequency regions, suggesting that conventional diffusion-based TKG extrapolation models still under-utilize frequency-domain information in historical event dynamics. In contrast, FreqDiff presents a more distributed and structured energy pattern across both frequency indices and latent channels, with visible activation not only in the low-frequency region but also in middle and higher frequency bands. This indicates that the proposed frequency-aware design enables the model to preserve richer spectral components, including stable long-term trends and more fluctuating short-term relational changes. Therefore, the visualization provides intuitive evidence that spectral modeling introduces complementary temporal signals beyond standard time-domain representations, which helps explain the improved extrapolation performance of FreqDiff.

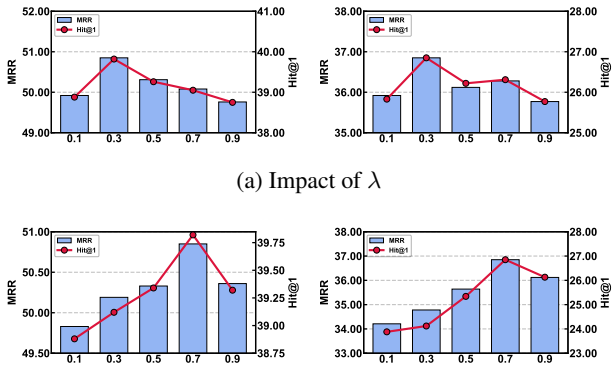
D.2 Sensitivity Analysis

Table 7 evaluates the sensitivity of the weighting coefficient λ on ICEWS14 and ICEWS18, where MRR and Hit@1 exhibit a generally consistent trend across the two datasets. When λ increases from 0.1 to 0.3, both metrics improve noticeably, indicating that a moderate strength of the corresponding regularization/objective term can effectively enhance representation learning and improve future fact prediction. The best performance is obtained at λ=0.3 on both datasets, suggesting that this setting provides the most balanced contribution between the main prediction objective and the auxiliary constraint. However, when λ continues to increase beyond 0.3, the performance gradually declines, especially on ICEWS14, where both MRR and Hit@1 show a clear downward trend from 0.5 to 0.9. This implies that an excessively large λ may overemphasize the auxiliary learning signal and weaken the model's ability to optimize the primary



(a) FreqDiff w/o. Freq (b) FreqDiff (c) DiffuTKG (d) NADEx

Figure 6: Visualization of learned spectral energy distributions on ICEWS14. The y-axis represents frequency indices, and the x-axis represents latent channels. Red indicates higher spectral energy.



(a) Impact of λ

(b) Impact of α

Figure 7: Hyper-parameter Sensitivity analysis on ICEWS14 (left) and ICEWS18 (right) datasets.

extrapolation objective. For the fusion coefficient α, both datasets achieve strong performance around $\alpha = 0.7$, showing that temporal modeling should remain dominant while spectral calibration provides complementary information.

D.3 Case Study

Table 10 offers a concrete comparison of how different models rank candidate entities under realistic temporal extrapolation scenarios. In Case 1, for the query (Democratic Party (Nigeria), Criticize or denounce, ?), FreqDiff correctly predicts Muhammadu Buhari as the top-ranked entity with a confidence score of 0.7906, which is substantially higher than the scores assigned to other candidates. This shows that FreqDiff can identify the specific political figure most likely to be involved in the future event, rather than merely selecting broad and frequent entities such as Citizen (Nigeria) or Government (Nigeria). In contrast, FreqDiff w/o. Freq ranks the correct entity second, while DiffuTKG places it third with a much lower confidence score. This comparison suggests that removing frequency modeling weakens the model's ability to capture discriminative temporal signals, causing it to rely more heavily on generic co-occurrence patterns. A similar observation can be made in Case 2. For the query (Ethiopia, Sign formal agreement, ?), FreqDiff ranks Sudan first with a confidence score of 0.7782, whereas FreqDiff w/o. Freq ranks it second and DiffuTKG fails to prioritize the correct answer, instead assigning higher ranks to countries such as Portugal, China, and Iran. These results indicate that the frequency-aware module helps FreqDiff capture relation-specific temporal regularities and distinguish the most contextually appropriate future entity from plausible but less accurate alternatives. Overall, the case study demonstrates that spectral information not only improves quantitative performance but also leads to more reliable and interpretable entity ranking in future fact prediction.

We further analyze the composition of the decoded subject histories to understand why different models favor different candidate entities. Taking Case #1 as an example, Figure 8 shows that the two competing candidates are supported by different types of historical evidence. The entity *Muhammadu Buhari*, which is ranked first by FreqDiff, is associated with only 10 input events. However, these events are highly query-relevant, as they involve direct interactions with the *Democratic Party (Nigeria)* and include several recent criticism- and accusation-related events immediately before the query timestamp. In contrast, *Citizen (Nigeria)*, which is selected as the top-1 prediction by FreqDiff w/o. $Freq$ and DiffuTKG, is linked to 23 input events. Although this candidate appears more frequently in the subject history, its associated events are more generic and broadly reflect interactions between the party and a collective political actor, rather than providing specific evidence for the masked entity in the given query. This compar-

Table 10: Case study of future entity prediction on ICEWS14. Given a query fact with the target entity masked, the table reports the top-5 predicted entities and their confidence scores produced by FreqDiff, FreqDiff w/o. $Freq$, and DiffuTKG. The red entries denote the ground-truth entities.

	FreqDiff	FreqDiff w/o. $Freq$	DiffuTKG
Case #1 **Date**: 2014-12-18 **Query**: (Democratic Party (Nigeria), *Criticize or denounce*, ?) **Gold**: Muhammadu Buhari	1. Muhammadu Buhari (0.79056861) 2. Citizen (Nigeria) (0.06997431) 3. Congress (Nigeria) (0.04214321) 4. Government (Nigeria) (0.03278727) 5. Head of Government (Nigeria) (0.00737355)	1. Citizen (Nigeria) (0.712012) 2. Muhammadu Buhari (0.145593) 3. Government (Nigeria) (0.028617) 4. Congress (Nigeria) (0.024223) 5. Boko Haram (0.011203)	1. Citizen (Nigeria) (0.231514) 2. Government (Nigeria) (0.184501) 3. Muhammadu Buhari (0.076782) 4. Chibuike Rotimi Amaechi (0.029169) 5. Ministry (Nigeria) (0.029161)
Case #2 **Date**: 2014-12-25 **Query**: (Ethiopia, *Sign formal agreement*, ?) **Gold**: Sudan	1. Sudan (0.77821444) 2. China (0.14515885) 3. Portugal (0.00858667) 4. Angola (0.00622532) 5. France (0.00528247)	1. Portugal (0.471355) 2. Sudan (0.240765) 3. China (0.106766) 4. United Arab Emirates (0.070986) 5. Angola (0.012471)	1. Portugal (0.118397) 2. China (0.081131) 3. Iran (0.068646) 4. South Africa (0.065919) 5. Ethiopia (0.055098)

ison suggests that the baselines are more easily biased toward historically frequent candidates, especially when the entity has dense but less discriminative historical links. By contrast, FreqDiff ranks *Muhammadu Buhari* first despite its fewer direct connections, indicating that the proposed spectral-aware filterbank helps to emphasize temporally informative and query-specific evidence rather than relying on raw historical frequency.

Muhammadu Buhari-linked input events

```
[01] (Muhammadu Buhari, Give ultimatum, Democratic Party (Nigeria), 2014-04-18)
[02] (Muhammadu Buhari, Make statement, Democratic Party (Nigeria), 2014-10-16)
[03] (Muhammadu Buhari, Accuse, Democratic Party (Nigeria), 2014-11-20)
[04] (Democratic Party (Nigeria), Criticize or denounce, Muhammadu Buhari, 2014-12-04)
[05] (Democratic Party (Nigeria), Make statement, Muhammadu Buhari, 2014-12-04)
[06] (Democratic Party (Nigeria), Praise or endorse, Muhammadu Buhari, 2014-12-12)
[07] (Democratic Party (Nigeria), Criticize or denounce, Muhammadu Buhari, 2014-12-15)
[08] (Democratic Party (Nigeria), Criticize or denounce, Muhammadu Buhari, 2014-12-15)
[09] (Democratic Party (Nigeria), Criticize or denounce, Muhammadu Buhari, 2014-12-15)
[10] (Democratic Party (Nigeria), Accuse, Muhammadu Buhari, 2014-12-15)
```

Citizen (Nigeria)-linked input events

```
[01] (Democratic Party (Nigeria), Use conventional military force, Citizen (Nigeria), 2014-01-13)
[02] (Citizen (Nigeria), Reject, Democratic Party (Nigeria), 2014-01-24)
[03] (Citizen (Nigeria), Reduce relations, Democratic Party (Nigeria), 2014-02-08)
[04] (Democratic Party (Nigeria), Make an appeal or request, Citizen (Nigeria), 2014-02-10)
[05] (Democratic Party (Nigeria), Appeal for military protection or peacekeeping, Citizen (Nigeria), 2014-02-10)
[06] (Democratic Party (Nigeria), Make empathetic comment, Citizen (Nigeria), 2014-02-19)
[07] (Democratic Party (Nigeria), Criticize or denounce, Citizen (Nigeria), 2014-02-25)
[08] (Citizen (Nigeria), Reject, Democratic Party (Nigeria), 2014-03-07)
[09] (Democratic Party (Nigeria), Make pessimistic comment, Citizen (Nigeria), 2014-05-02)
[10] (Democratic Party (Nigeria), Demand, Citizen (Nigeria), 2014-06-18)
[11] (Citizen (Nigeria), Accuse, Democratic Party (Nigeria), 2014-07-28)
[12] (Citizen (Nigeria), Criticize or denounce, Democratic Party (Nigeria), 2014-08-01)
[13] (Democratic Party (Nigeria), Accuse, Citizen (Nigeria), 2014-09-15)
[14] (Democratic Party (Nigeria), Threaten, Citizen (Nigeria), 2014-09-15)
[15] (Democratic Party (Nigeria), Accuse, Citizen (Nigeria), 2014-10-09)
[16] (Democratic Party (Nigeria), Threaten, Citizen (Nigeria), 2014-10-13)
[17] (Democratic Party (Nigeria), Refuse to yield, Citizen (Nigeria), 2014-10-13)
[18] (Citizen (Nigeria), Reduce relations, Democratic Party (Nigeria), 2014-10-13)
[19] (Democratic Party (Nigeria), Bring lawsuit against, Citizen (Nigeria), 2014-10-14)
[20] (Democratic Party (Nigeria), Accuse, Citizen (Nigeria), 2014-10-16)
[21] (Citizen (Nigeria), Criticize or denounce, Democratic Party (Nigeria), 2014-10-20)
[22] (Citizen (Nigeria), Reduce relations, Democratic Party (Nigeria), 2014-10-31)
[23] (Citizen (Nigeria), Make optimistic comment, Democratic Party (Nigeria), 2014-11-14)
```

Figure 8: Input event histories associated with the correct candidate and the misleading high-ranked candidate in Case #1.