

Leveraging 3D Gaussian for Temporal Knowledge Graph Embedding

Jiang Li, Xiangdong Su*, Guanglai Gao

¹ College of Computer Science, Inner Mongolia University, China

² National & Local Joint Engineering Research Center of Intelligent Information Processing Technology for Mongolian, China

³ Inner Mongolia Key Laboratory of Multilingual Artificial Intelligence Technology, China
lijiangimu@gmail.com, cssxd@imu.edu.cn

Abstract

Representation learning in knowledge graphs (KGs) has predominantly focused on static data, yet many real-world knowledge graphs are inherently dynamic. For instance, the fact (The CEO of Apple, holds position, Steve Jobs) was valid until 2011, after which it changed, emphasizing the need to incorporate temporal information into knowledge representation. In this paper, we propose 3DG-TE, a novel temporal KG embedding method inspired by 3D Gaussian Splatting, where entities, relations, and timestamps are modeled as 3D Gaussian distributions with learnable structured covariance. This approach optimizes the Gaussian distributions of entities, relations, and timestamps to improve the overall KG representation. To effectively capture temporal-relational interactions, we design structured covariances that form composite transformation operators: relations induce rotational transformations, while timestamps regulate adaptive scaling. We also design a compound scoring function that integrates mean positions and structured covariance, preserving geometric interpretability. Experimental results on three benchmark TKG datasets demonstrate that 3DG-TE outperforms state-of-the-art baselines in temporal link prediction tasks. Theoretical analysis further confirms our model's ability to capture key relation patterns.

1 Introduction

Knowledge graphs (KGs) are extensively employed as structured frameworks for modeling and organizing complex relationships among entities, and they have become foundational tools in domains such as semantic search (Sun et al., 2023), recommendation systems (Gao et al., 2023), and question answering (Wang et al., 2024). While traditional KGs are effective in many contexts, they fall short in capturing the temporal dimension of knowledge, which is essential for representing real-world

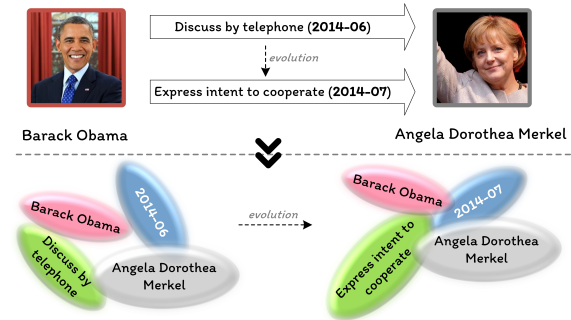


Figure 1: An illustration of temporal evolution patterns. The top section shows the temporal evolution of events between *Barack Obama* and *Angela Dorothea Merkel*. The bottom section shows how our method encodes entities, relations, and time as 3D Gaussian distributions to model temporal interactions in TKGs.

data where facts are time-dependent. For instance, the statement "*Barack Obama is the president of the United States*" is only valid between 2009 and 2017. To address this limitation, temporal knowledge graph embedding (TKGE) models have been developed. These models extend traditional KGs by incorporating temporal information to indicate when a fact is valid.

Early TKGE methods built upon translation-based KGE method TransE (Bordes et al., 2013) by encoding time as translations same as relations (Leblay and Chekol, 2018; García-Durán et al., 2018). Recent works have pointed out that TKGs inherently exhibit uncertainties and randomness. ATiSE (Xu et al., 2020b) employs an additive time series decomposition approach that models temporal uncertainty using Gaussian noise. TKGC-AGP (Zhang and Zhou, 2022) uses multivariate Gaussian processes to model temporal dependencies and uncertainties. In addition, tensor decomposition-based TKGE models (Lacroix et al., 2020; Xu et al., 2021; Li et al., 2023) have been developed. The latest research introduces mod-

* Corresponding Author

els that encode TKGs using composite structures. HGE (Pan et al., 2024) employs a hybrid space to enhance the representation of temporal dynamics, while TCompoundE (Ying et al., 2024) leverages relation-specific and time-specific compound geometric operations to model both temporal evolution and relational structures more effectively. TKGE has become a key research area for modeling the temporal KGs.

In this paper, we propose 3DG-TE, a novel TKGE approach inspired by the principles of 3D Gaussian Splatting (Kerbl et al., 2023). In computer graphics, 3D Gaussian Splatting is widely used for scene representation and rendering, where each Gaussian models a local point distribution to efficiently reconstruct complex surfaces. Inspired by 3D Gaussian Splatting, 3DG-TE encodes each quadruple (s, r, o, t) as a structured representation of a fact, where entities, relations, and timestamps are modeled as 3D Gaussian ellipsoid, as shown in Figure 1. 3DG-TE optimizes 3D Gaussian ellipsoid of entities, relations, and timestamps to more accurately represent quadruple facts, thereby enhancing link prediction performance. To more accurately model relation and timestamp interactions, we further design structured covariance representations in which relations induce rotational transformations, while timestamps regulate adaptive scaling, forming composite transformation operators. 3DG-TE focuses on 3D Gaussian space modeling over conventional probabilistic KGE modeling, such as ATiSE (Xu et al., 2020b) and TKGC-AGP (Zhang and Zhou, 2022). This enables our model to capture rich temporal interactions among entities, relations, and timestamps in TKGs, thereby facilitating more effective TKG modeling.

Specifically, ATiSE captures temporal evolution through additive time series decomposition. However, it primarily assumes independent Gaussian noise to model temporal uncertainty, which may limit its ability to capture complex interactions among time, entities, and relations. In contrast, our model is designed with a full 3D Gaussian structured covariance, but rather Gaussian noise, thus enabling accurate learning of the dynamics of relation and timestamp interactions in the TKG. TKGC-AGP models temporal evolution using multivariate Gaussian processes, leveraging kernel functions to ensure smooth temporal transitions. However, its reliance on Gaussian process regression may limit scalability when applied to large-scale knowledge graphs. Instead, our approach replaces Gaussian

process regression with a learnable 3D Gaussian that designs relation- and time-structured covariances to capture TKG interactions. This design enables our model to achieve superior performance on large-scale datasets GDEL. In addition, experimental results on other benchmark TKG datasets confirm that our approach surpasses state-of-the-art TKGE models¹.

Due to space constraints, the content of the **Related Work** section can be found in **Appendix A**.

2 Background and Notation

2.1 Problem Formulation

For a temporal knowledge graph $\mathcal{G}$, let $\mathcal{E}$, $\mathcal{R}$, and $\mathcal{T}$ denote the sets of entities, relations, and timestamps, respectively. Each fact within the TKG is represented as a quadruple (s, r, o, t) , where $s, o \in \mathcal{E}$ correspond to the subject and object entities, $r \in \mathcal{R}$ denotes the relation linking the entities, and $t \in \mathcal{T}$ specifies the temporal information of the fact. The objective of temporal knowledge graph embedding is to infer missing components in incomplete quadruples. Given a query in the form of $(s, r, ?, t)$ (where the object entity is missing) or $(?, r, o, t)$ (where the subject entity is missing), the model aims to predict the most plausible missing entity or timestamp by leveraging observed facts in the TKG. To achieve this, the score function $\phi(s, r, o, t)$ is used to evaluate the plausibility of a given quadruple. The model is trained by optimizing this function over observed facts in the dataset. During inference, missing components are predicted by ranking candidate entities or timestamps based on their scores.

2.2 3D Gaussian Splatting

3D Gaussian splatting (Kerbl et al., 2023) is a volumetric representation technique that models spatial distributions using anisotropic Gaussian functions. Unlike traditional explicit representations such as meshes or implicit neural fields like Neural Radiance Fields (NeRF) (Mildenhall et al., 2021), which rely on discrete geometry or continuous function approximation, 3D Gaussian splatting represents a scene as a collection of learnable Gaussian kernels. Each kernel encodes geometric and appearance properties, making it highly effective for modeling complex spatial structures.

¹Code is available on <https://github.com/dellixx/3DG-TE>

A 3D Gaussian is parameterized by a mean vector $\mu \in \mathbb{R}^3$ representing its spatial position and a covariance matrix $\Sigma \in \mathbb{R}^{3 \times 3}$ that defines its shape and orientation. A straightforward approach to optimizing the covariance matrix Σ would involve directly updating its elements. However, this can introduce numerical instability, as covariance matrices must remain positive semi-definite to be physically meaningful. Gradient-based optimization often struggles to enforce this constraint, leading to invalid covariance matrices. To mitigate this issue, we adopt an alternative parameterization based on structured decomposition, ensuring the validity and stability of the covariance representations. Following (Kerbl et al., 2023), we express it as:

$$\Sigma = R S S^\top R^\top, \quad (1)$$

where $R \in \mathbb{R}^{3 \times 3}$ is a rotation matrix that encodes orientation, and $S \in \mathbb{R}^{3 \times 3}$ is a diagonal scaling matrix that determines anisotropic spread. This parameterization ensures that Σ remains positive semi-definite while providing an intuitive way to control the shape and orientation of the Gaussian ellipsoids.

While 3D Gaussian splatting is primarily used in computer graphics for rendering, we adopt its core principles to construct a novel temporal knowledge graph embedding framework. Unlike rendering-based applications that focus on synthesizing photorealistic images, our approach encodes each fact (s, r, o, t) in a temporal knowledge graph as a structured 3D Gaussian ellipsoid. Here, the head entity, relation, and timestamp are represented as learnable Gaussian, parameterized by their respective means and covariance matrices. Instead of generating visual representations, we optimize these Gaussian to construct a scoring function that measures the plausibility of a fact. This formulation enables our model to dynamically capture the temporal evolution of knowledge while preserving geometric interpretability. By leveraging relation- and time-structured covariances, we facilitate efficient representation learning.

3 Methodology

3.1 3DG-TE Model

In this section, we propose 3DG-TE, a novel TKGE method inspired by 3D Gaussian Splatting (Kerbl et al., 2023). First, we construct a structured covariance that incorporates both relations and time.

Then, we formulate the score function based on the fused covariance and means.

Constructing Structured Covariance. For a quadruple (s, r, o, t) in a TKG, each entity, relation, and timestamp is modeled as a learnable 3D Gaussian $\mathcal{N}(\mu, \Sigma)$ parameterized by a mean vector μ and a structured covariance matrix Σ . The optimization objective is to refine these Gaussian distribution to best capture the factual knowledge encoded in the quadruples. In a 3D Gaussian distribution, the mean determines the center of the ellipsoid, while the covariance matrix defines its shape.

Furthermore, previous studies (Li et al., 2023; Ying et al., 2024) have shown that the relationship between the head and tail entities is dynamically influenced by both the relation and time. Therefore, we treat relation and time as a unified structured covariances and optimize their combined structured covariance in the 3DG-TE model. Specifically, we represent the relation as the rotation matrix R_r of the overall $\Sigma(r, t)$ covariance, while time is modeled as the scaling matrix S_t within the same structure. Hence, we can get

$$\Sigma_{(r,t)} = R_r S_t S_t^\top R_r^\top. \quad (2)$$

Following Kerbl et al. (2023), rotation matrix R_r is derived from a quaternion representation $q_r^{dim \times 4} = (w, x, y, z)$, where the quaternion is normalized to ensure valid rotations. The quaternion is then converted into a $(dim \times 3 \times 3)$ rotation matrix as follows:

$$R_r = \begin{bmatrix} 1 - 2(y^2 + z^2) & 2(xy - wz) & 2(xz + wy) \\ 2(xy + wz) & 1 - 2(x^2 + z^2) & 2(yz - wx) \\ 2(xz - wy) & 2(yz + wx) & 1 - 2(x^2 + y^2) \end{bmatrix}, \quad (3)$$

where (w, x, y, z) are the components of the quaternion. The time vector $s_t^{dim \times 3} = (s_1, s_2, s_3)$ is normalized and then converted into its diagonal form:

$$S_t = \text{diag}(s_t). \quad (4)$$

The scaling matrix S_t has the size $dim \times 3 \times 3$ because it is a diagonal matrix with the scaling factors placed along the diagonal.

Our approach integrates the covariance structure of both relation and time, allowing them to jointly capture entity interactions. The relation between entities is modeled as a rotation, preserving their geometric structure while enabling transformations in a multidimensional space, similar

to static KGE models such as RotatE (Sun et al., 2019) and PairRE (Chao et al., 2021). Time influences the magnitude of entity representations, reflecting changes in their significance or impact over time. The scaling matrix captures these temporal variations by adjusting the magnitude of entity embeddings, effectively modeling temporal dynamics.

After that, we can obtain the relational Gaussian and temporal Gaussian distributions represented as

$$\mathcal{N}_r(\boldsymbol{\mu}_r, \boldsymbol{\Sigma}_{(r,t)}), \quad \mathcal{N}_t(\boldsymbol{\mu}_t, \boldsymbol{\Sigma}_{(r,t)}), \quad (5)$$

where $\boldsymbol{\mu}_r$ and $\boldsymbol{\mu}_t$ represent the mean vectors corresponding to the relation r and the time t , respectively, reflecting the central positions of their embeddings in the latent space.

The Score Function. Since 3D Gaussian Splatting renders scenes based on the mean and covariance, we design a scoring function that leverages these parameters to fit the quadruple of represented facts.

First, for the 3D Gaussian information of relations and time, we adopt an information aggregation approach to fuse the mean and covariance information. Specifically, we use the strategy of adding the mean and covariance to effectively combine both aspects, capturing the dynamic translation characteristics of relations and time. Through this approach, the model can simultaneously learn spatial positions and transformation properties, enhancing its ability to model TKGs.

$$\begin{aligned} \mathcal{IA}_r &= \text{abs}(\boldsymbol{\mu}_r) + \boldsymbol{\Sigma}_{(r,t)}, \\ \mathcal{IA}_t &= \sin(\boldsymbol{\mu}_t) + \boldsymbol{\Sigma}_{(r,t)}. \end{aligned} \quad (6)$$

The absolute function is applied to the relational mean to ensure non-negative transformations, as relations often encode directed mappings. To encode temporal information, a sinusoidal function is used to capture its periodic characteristics, thereby enabling the model to effectively learn cyclical patterns over time.

After aggregating the information, we employ information distribution to ensure the geometric significance of the mean and covariance of the Gaussian ellipsoid is preserved.

$$\begin{aligned} \boldsymbol{\mu}'_r, \boldsymbol{\Sigma}'_r &= \mathcal{IA}_r[:, \text{_half_dim}], \mathcal{IA}_r[:, \text{_half_dim} :] \\ \boldsymbol{\mu}'_t, \boldsymbol{\Sigma}'_t &= \mathcal{IA}_t[:, \text{_half_dim}], \mathcal{IA}_t[:, \text{_half_dim} :] \end{aligned} \quad (7)$$

where the term ‘_half_dim’ splits the dimensional space of the Gaussian distribution into two equal parts, one for the mean and the other for the covariance.

In addition, we adopt an additive fusion strategy for the means, as mean vectors represent the central positions of entities, relations, and time in the embedding space. Adding these means preserves their relative spatial positions, enabling the model to effectively capture their interactions. In geometric space, addition corresponds to translation, allowing us to determine the positions of their 3D Gaussian ellipsoids. By adding these components, we obtain their spatial locations, which are then combined to represent the factual knowledge encoded in the quadruple. On the other hand, covariance matrices describe the shape of 3D Gaussian distributions, are typically decomposed into rotation and scaling operations. Therefore, we adopt a multiplicative fusion strategy to effectively model the interactions among these components. Hence, we can get

$$\begin{aligned} \boldsymbol{\mu}'_{(s,r,t)} &= \boldsymbol{\mu}_s + \boldsymbol{\mu}'_r + \boldsymbol{\mu}'_t, \\ \boldsymbol{\Sigma}'_{(r,t)} &= \boldsymbol{\Sigma}_s \cdot \boldsymbol{\Sigma}'_r \cdot \boldsymbol{\Sigma}'_t. \end{aligned} \quad (8)$$

As previously described, previous studies have demonstrated that entity interactions are jointly influenced by both relations and time (Li et al., 2023; Ying et al., 2024). In our model, the representation of entities does not involve covariance because we consider the position of entities in the knowledge graph to be relatively fixed. That is, their spatial shape can be represented by the mean, without the need for complex rotation or scaling transformations.

In our method, we enforce this operation by setting the covariance matrix of entities to an identity matrix $\boldsymbol{\Sigma}_s = \mathbf{I}$. This is because the physical location of an entity typically reflects its static attributes, such as name or identifier, which are independent of the dynamic changes driven by relations and time. Therefore, the representation of entities is optimized through their mean, while covariance is primarily used to capture the dynamic nature of relations and time, which differs from the fixed positions of entities themselves.

To compute the plausibility of a quadruple (s, r, o, t) , we integrate three key components: the fused mean representation $\boldsymbol{\mu}'_{(s,r,t)}$, the structured covariance $\boldsymbol{\Sigma}'_{(r,t)}$, and the object entity embedding e_o . The mean representation captures the

spatial positioning of the subject entity, relation, and timestamp, while the covariance matrix encodes the interaction dynamics between relations and time. We compute the semantic similarity between these three components, ensuring a meaningful alignment between the fused representation and the target entity. The final score function is formulated as follows:

$$\phi(s, r, o, t) = \langle \mu'_{(s,r,t)}, \Sigma'_{(r,t)}, e_o \rangle, \quad (9)$$

where e_o denotes the mean embedding of the object entity.

3.2 Loss Function

Following TeLM (Xu et al., 2021) and TeAST (Li et al., 2023), we utilize reciprocal learning to train our model. The loss function is formulated as follows:

$$\begin{aligned} \mathcal{L}_\alpha = & -\log \left(\sum_{o \in \mathcal{E}} \exp(\phi(s, r, o, t)) \right) \\ & -\log \left(\sum_{s' \in \mathcal{E}} \exp(\phi(o, r^{-1}, s', t)) \right) \\ & + \lambda_\alpha \sum_{i=1}^k \left(\|\mu'_{(s,r,t)}\|_3^3 + \|\Sigma'_{(r,t)}\|_3^3 + \|e_o\|_3^3 \right). \end{aligned} \quad (10)$$

Here, λ_α denotes the regularization coefficient for the N3 regularization weight, r^{-1} represents the inverse of the relation, and s' is the subject of the negative samples. The smoothing temporal regularizer from TNTComplEx (Lacroix et al., 2020) is employed to enforce similar representations for consecutive time points. This regularizer is expressed as:

$$\mathcal{L}_t = \frac{1}{N_t - 1} \sum_{i=1}^{N_t-1} \|\mu_{t(i+1)} - \mu_{t(i)}\|_3^3. \quad (11)$$

Finally, the complete loss function for 3DG-TE is defined as:

$$\mathcal{L} = \mathcal{L}_\alpha + \lambda_t \mathcal{L}_t, \quad (12)$$

where λ_t represents the temporal regularization weight.

3.3 Modeling Key Relation Patterns

3DG-TE can model important relation patterns, including symmetric, asymmetric, inverse and temporal evolution patterns. A comprehensive list of these properties is presented below, with detailed proofs provided in the **Appendix B**.

Proposition 1. *3DG-TE can model the symmetric relation pattern.*

Proposition 2. *3DG-TE can model the asymmetric relation pattern.*

Proposition 3. *3DG-TE can model the inverse relation pattern.*

Proposition 4. *3DG-TE can model the temporal evolution pattern.*

4 Experiments

4.1 Benchmark Datasets

We evaluate 3DG-TE on Temporal Knowledge Graph Embedding benchmark datasets: **ICEWS14** and **ICEWS05-15** (García-Durán et al., 2018) and **GDELT** (Leetaru and Schrod, 2013). The details and statistics of the datasets are provided in **Appendix C**.

4.2 Baselines

We perform a comprehensive comparison with the state-of-the-art TKGE models, including TTransE (Leblay and Chekol, 2018), DE-Simple (Goel et al., 2020), TA-DistMult (García-Durán et al., 2018), ChronoR (Sadeghian et al., 2021), TComplEx (Lacroix et al., 2020), TNTComplEx (Lacroix et al., 2020), ATiSE (Xu et al., 2020b), TeLM (Xu et al., 2021), TKGC-AGP (Zhang and Zhou, 2022) BoxTE (Messner et al., 2022), RotateQVS (Chen et al., 2022), TLT-KGE (Zhang et al., 2022) TeAST (Li et al., 2023), HGE (Pan et al., 2024) and TCompoundE (Ying et al., 2024).

Among the existing TKGE methods, TCompoundE (Ying et al., 2024) obtains SOTA results on ICEWS14, ICEWS05-15 and GDELT dataset. Furthermore, while both ATiSE (Xu et al., 2020b) and TKGC-AGP (Zhang and Zhou, 2022) incorporate Gaussian-based representations, Our model fundamentally differs from both in its theoretical foundation and methodological approach. Our focus is on 3D Gaussian space modeling, rather than probabilistic modeling. In this paper, we consider TCompoundE, ATiSE and TKGC-AGP as the primary baselines.

Models (<i>Reference</i>)	ICEWS14				ICEWS05-15				GDELT			
	MRR	H@1	H@3	H@10	MRR	H@1	H@3	H@10	MRR	H@1	H@3	H@10
TTransE (<i>WWW (2018)</i>)	0.255	0.074	-	0.601	0.271	0.084	-	0.616	0.115	0.000	0.160	0.318
TA-DisMult (<i>EMNLP (2018)</i>)	0.477	0.363	-	0.686	0.474	0.346	-	0.728	0.206	0.124	0.219	0.365
DE-Simple (<i>AAAI (2020)</i>)	0.526	0.418	0.592	0.725	0.513	0.392	0.578	0.748	0.230	0.141	0.248	0.403
TComplEx (<i>ICLR (2020)</i>)	0.610	0.530	0.660	0.770	0.660	0.590	0.710	0.800	0.340	0.294	0.361	0.498
TNTComplEx (<i>ICLR (2020)</i>)	0.620	0.520	0.660	0.760	0.670	0.590	0.710	0.810	0.349	0.258	0.373	0.502
ATiSE (<i>ISWC (2020b)</i>)	0.550	0.436	0.629	0.750	0.519	0.378	0.606	0.794	-	-	-	-
ChronoR (<i>AAAI (2021)</i>)	0.625	0.547	0.669	0.773	0.675	0.596	0.723	0.820	-	-	-	-
TeLM (<i>NAACL (2021)</i>)	0.625	0.545	0.673	0.774	0.678	0.599	0.728	0.823	0.350	0.261	0.375	0.504
TKGC-AGP (<i>COLING (2022)</i>)	0.561	0.458	0.631	0.738	0.532	0.398	0.621	0.797	-	-	-	-
BoxTE (<i>AAAI (2022)</i>)	0.613	0.528	0.664	0.763	0.667	0.582	0.719	0.820	0.352	0.269	0.377	0.511
RotateQVS (<i>ACL (2022)</i>)	0.591	0.507	0.642	0.754	0.633	0.529	0.709	0.813	0.270	0.175	0.293	0.458
TLT-KGE (<i>CIKM (2022)</i>)	0.630	0.549	0.678	0.777	0.686	0.607	0.735	0.831	0.356	0.267	0.385	0.532
TeAST (<i>ACL (2023)</i>)	0.637	0.560	0.682	0.782	0.683	0.604	0.732	0.829	0.371	0.283	0.401	0.544
HGE (<i>AAAI (2024)</i>)	0.634	0.550	0.685	0.788	0.688	0.608	0.740	0.835	0.371	0.277	0.402	0.556
TCompoundE (<i>ACL (2024)</i>)	0.644	0.561	0.694	0.795	0.692	0.612	0.743	0.837	0.433	0.347	0.469	0.595
3DG-TE (ours)	0.646	0.564	0.695	0.796	0.694	0.614	0.747	0.841	0.452	0.367	0.488	0.612

Table 1: Link prediction results on ICEWS14, ICEWS05-15 and GDELT. Dashes means results are not reported in the responding literature. The best score is in **bold**.

4.3 Evaluation Protocol

Following the prior works (Lacroix et al., 2020; Chen et al., 2022; Li et al., 2023), we assess the ranking performance of the predicted quadruples by considering all possible substitutions of the subject and object entities, denoted as (s', r, o, t) and (s, r, o', t) , where s' and o' belong to the set of entities, $\mathcal{E}$. The candidate quadruples are subsequently ranked based on their scores, with time-based filtering applied to ensure the temporal consistency of the predictions (Xu et al., 2021; Li et al., 2023). The model’s performance is quantified using standard evaluation metrics, including Mean Reciprocal Rank (MRR) and Hits@ n . Hits@ n measures the fraction of correct entities appearing in the top n predictions. Higher values of MRR and Hits@ n indicate superior model performance. For simplicity, we refer to Hits@ n as H@ n , where n takes values of 1, 3, and 10.

For more information about other experimental setups, please see **Appendix D**.

5 Results and Analysis

5.1 Main Results

Table 1 presents the link prediction performance of 3DG-TE on three benchmark datasets: ICEWS14, ICEWS05-15, and GDELT, covering diverse temporal knowledge graph scenarios. Our model

consistently outperforms all baselines in MRR, Hits@1, Hits@3, and Hits@10, demonstrating its effectiveness in modeling temporal knowledge graphs. Notably, 3DG-TE surpasses state-of-the-art models, including TCompoundE, ATiSE, and TKGC-AGP, highlighting its ability to jointly capture relational and temporal dependencies.

Moreover, BoxTE (Messner et al., 2022) emphasizes that GDELT requires a high degree of temporal inductive capacity for efficient encoding. This is attributed to the significant temporal variability in GDELT, where some facts persist over several consecutive timestamps, whereas others are brief and infrequent. As shown in Table 1, 3DG-TE achieves strong results on the GDELT dataset, outperforming many baseline models. A key strength of 3DG-TE is its 3D Gaussian-based representation, where entities, relations, and timestamps are parameterized as Gaussian distributions. The relation- and time-structured covariances enables the model to handle the diverse temporal nature of GDELT effectively. The structured covariance approach in our model makes it effective in capturing complex temporal dynamics and thus better modeling the complex GDELT. This highlights the effectiveness of our model in modeling complex TKGs.

		ICEWS14				ICEWS05-15			
		MRR	H@1	H@3	H@10	MRR	H@1	H@3	H@10
Variant of Formula	Formula	Ablation 1							
$\Sigma_{(r,t)} = R_r S_t S_t^\top R_r^\top$	Eq. 2	0.646	0.564	0.695	0.796	0.694	0.614	0.747	0.841
$\Sigma_{(r,t)} = R_t S_r S_r^\top R_t^\top$	Eq. 13	0.635	0.553	0.683	0.787	0.687	0.610	0.739	0.836
		Ablation 2							
$abs(\cdot)$	$\sin(\cdot)$								
✓	✓	0.646	0.564	0.695	0.796	0.694	0.614	0.747	0.841
✓	✗	0.012	0.004	0.011	0.025	0.008	0.002	0.006	0.015
✗	✓	0.638	0.560	0.681	0.784	0.683	0.605	0.736	0.832
Variant of Formula	Formula	Ablation 3							
$\mu'_{(s,r,t)} = \mu_s + \mu'_r + \mu'_t, \Sigma'_{(r,t)} = \Sigma_s \cdot \Sigma'_r \cdot \Sigma'_t$	Eq. 8	0.646	0.564	0.695	0.796	0.694	0.614	0.747	0.841
$\mu'_{(s,r,t)} = \mu_s \cdot \mu'_r \cdot \mu'_t, \Sigma'_{(r,t)} = \Sigma_s + \Sigma'_r + \Sigma'_t$	Eq. 14	0.010	0.004	0.010	0.019	0.007	0.002	0.005	0.013

Table 2: Ablation study on the impact of different covariance modeling strategies, the application of the absolute value and sine function, and fusion strategies for mean and covariance.

5.2 Ablation Studies

In this section, we perform a series of ablation experiments to analyze the contributions of different components of the 3DG-TE model. As shown in Table 2, Ablation study contains (**Ablation 1**) the impact of different covariance modeling strategies, (**Ablation 2**) the application of the absolute value and sine function, and (**Ablation 3**) fusion strategies for mean and covariance.

Ablation 1. We first investigate the impact of the structured covariance between relations and time in our 3DG-TE model as in Eq. 2. Specifically, we compare the performance of the model when the relation covariance is modeled as a rotation matrix and the time covariance is represented as a scaling matrix. To better understand the significance of this structure, we perform a comparative analysis where we reverse the roles of relation and time covariances. In this modified setup, we model the relation covariance as a scaling matrix and the time covariance as a rotation matrix. The new combined structured covariance is represented as:

$$\Sigma_{(r,t)} = R_t S_r S_r^\top R_t^\top. \quad (13)$$

As shown in Table 2, when we reverse the roles of relation and time covariances, the model’s performance drops slightly on both ICEWS14 and ICEWS05-15 datasets across all evaluation metrics. This suggests that the original design, where relation covariance is modeled as a rotation matrix and time covariance as a scaling matrix, provides a better alignment of the model’s ability to capture dynamic temporal relations and entity interactions.

The results further validate our approach, showing that integrating the structured covariance of both relations and time enables the model to effectively capture entity interactions. By using a rotation matrix for relations, the model preserves the geometric structure of entities while allowing transformations in a multidimensional space. The scaling matrix, which models time, adjusts the magnitude of entity representations, reflecting how the significance of entities evolves over time.

Ablation 2. Next, we explore the impact of different functions on the representation of relations and time in the 3DG-TE model. Specifically, we investigate the role of the absolute value function (abs) and the sine function (sin) applied to the relational and temporal mean vectors. In the original formulation of the model, the absolute value is applied to the relational mean to ensure non-negative transformations, as relations typically represent directed mappings. We use the sine function to effectively capture the inherent periodicity in temporal information.

As shown in Table 2, when the absolute value and sine functions are removed from the model (using a direct addition of relational and time means), the performance of the model declines significantly across all metrics. This demonstrates that these functions play a crucial role in ensuring that the model captures the correct directional and periodic nature of the relations and time.

Ablation 3. Finally, we examine the impact of different fusion strategies for the mean and covariance in the 3DG-TE model. Specifically, we

investigate the effect of using multiplication for the mean and addition for the covariance. Compare to Eq. 8, the mean vectors are fused multiplicatively, and the covariance matrices are combined additively. The new formulation of the model in this configuration is represented as:

$$\begin{aligned}\mu'_{(s,r,t)} &= \mu_s \cdot \mu'_r \cdot \mu'_t, \\ \Sigma'_{(r,t)} &= \Sigma_s + \Sigma'_r + \Sigma'_t.\end{aligned}\quad (14)$$

The results from Table 2 show that using multiplication for the mean and addition for the covariance results in a significant decrease in performance across all evaluation metrics on both ICEWS14 and ICEWS05-15 datasets. This suggests that the original fusion strategy, where the mean is added and the covariance is multiplied, is more effective in modeling the interactions between entities, relations, and time. Geometrically, the mean represents the center of the Gaussian distribution, while the covariance controls its shape and orientation. By adding the means, we preserve the relative positions of entities and relations, ensuring that their spatial representation remains consistent. However, multiplying the covariances allows for proper transformations and interactions between the components, modeling how relations and time influence the spatial distribution of entities. Adding the covariances, on the other hand, distorts the shape and spread of the ellipsoids, leading to less accurate representation. Therefore, the original approach of adding the means and multiplying the covariances better preserves the geometric structure of the graph and captures the dynamic interactions among entities, relations, and time.

5.3 SHAP Analysis of Feature Contributions

To further investigate the contributions of different components in our score function, we conduct SHAP (Lundberg and Lee, 2017) analysis on two benchmark datasets: ICEWS14 and ICEWS05-15. As shown in Figure 2, the SHAP values indicate that both relational and temporal transformations play a crucial role for the score function of 3DG-TE. Notably, relation rotation R_r has a stronger impact in ICEWS14, while time scaling S_t dominates in ICEWS05-15, highlighting dataset-specific variations in temporal evolution patterns. Additionally, entity embeddings significantly influence predictions, demonstrating the necessity of preserving entity-specific representations, while timestamp embeddings contribute more in ICEWS05-15,

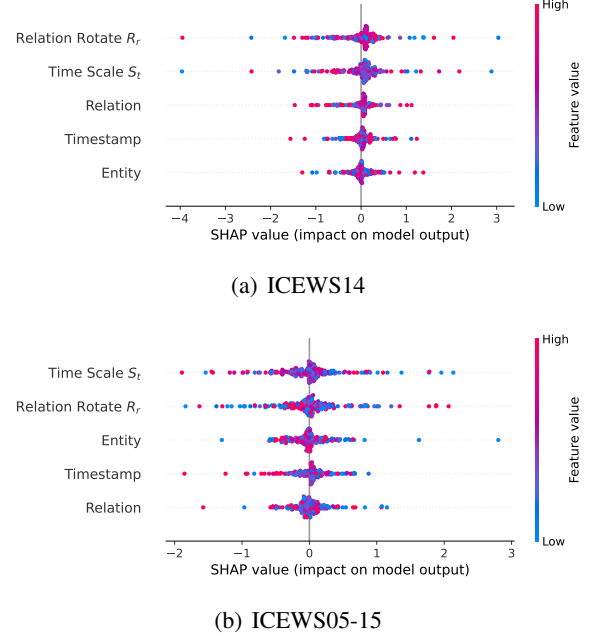


Figure 2: The SHAP values indicate the impact of different components, including entity, relation, timestamp, relation rotation R_r , and time scaling S_t , on the model’s score function on ICEWS14 and ICEWS05-15. Higher absolute SHAP values represent greater influence on the score function.

where event prediction requires stronger temporal modeling. The comparative importance of relations versus timestamps further varies, with ICEWS14 relying more on relational structure and ICEWS05-15 showing greater dependence on temporal information. These observations align with previous findings (Li et al., 2023; Ying et al., 2024), confirming that entity interactions are jointly determined by relations and time, and demonstrates the validity of constructing structured covariances in terms of relations and time in TKGs.

Additional experiments on relation and timestamp visualization and analysis are provided in Appendix E.

6 Conclusion

We presented 3DG-TE, a novel temporal knowledge graph embedding framework that leverages 3D Gaussian representations to model temporal-relational interactions. By decomposing covariance matrices into relation-driven rotations and time-specific scaling operations, our approach effectively encodes TKGs and achieves the excellent performance. Furthermore, our experiments

confirm the critical roles of relations and time in modeling temporal knowledge graphs, providing quantitative evidence to support this claim. Finally, theoretical proofs demonstrate 3DG-TE can encode important relational patterns.

Limitations

While 3DG-TE introduces a novel perspective on temporal knowledge graph embedding by leveraging 3D Gaussian distributions, similar to the majority of TKGE models (Lacroix et al., 2020; Li et al., 2023; Pan et al., 2024), etc., 3DG-TE is unable to process new entities that are not present in the training data.

Acknowledgments

This work was funded by National Natural Science Foundation of China (Grant No. 62366036), National Education Science Planning Project (Grant No. BIX230343), Outstanding Youth Fund Project of Inner Mongolia Autonomous Region (Grant No. 2025JQ010), Program for Young Talents of Science and Technology in Universities of Inner Mongolia Autonomous Region (Grant No. NJYT24033), The Central Government Fund for Promoting Local Scientific and Technological Development (Grant No. 2022ZY0198), Key R&D and Achievement Transformation Program of Inner Mongolia Autonomous Region (Grant No. 2025YFDZ0011, 2025YFDZ0026, 2025YFSH0021, 2025YFHH0073, 2023YFSH0017), Hohhot Science and Technology Project (Grant No. 2023-Zhan-Zhong-1), Science and Technology Program of the Joint Fund of Scientific Research for the Public Hospital of Inner Mongolia Academy of Medical Sciences (Grant No. 2023GLLH0035), The Project of Innovation Research in Postgraduate at Inner Mongolia University (11200-5253731).

References

- Ralph Abboud, İsmail İlkan Ceylan, Thomas Lukasiewicz, and Tommaso Salvatori. 2020. [Boxe: A box embedding model for knowledge base completion](#). In *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual*.
- Antoine Bordes, Nicolas Usunier, Alberto García-Durán, Jason Weston, and Oksana Yakhnenko. 2013. [Translating embeddings for modeling multi-relational data](#). In *Advances in Neural Information Processing Systems 26: 27th Annual Conference on Neural Information Processing Systems 2013. Proceedings of a meeting held December 5-8, 2013, Lake Tahoe, Nevada, United States*, pages 2787–2795.
- Linlin Chao, Jianshan He, Taifeng Wang, and Wei Chu. 2021. [PairRE: Knowledge graph embeddings via paired relation vectors](#). In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 4360–4369, Online. Association for Computational Linguistics.
- Kai Chen, Ye Wang, Yitong Li, and Aiping Li. 2022. [Rotategvs: Representing temporal information as rotations in quaternion vector space for temporal knowledge graph completion](#). In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), ACL 2022, Dublin, Ireland, May 22-27, 2022*, pages 5843–5857. Association for Computational Linguistics.
- John Duchi, Elad Hazan, and Yoram Singer. 2011. [Adaptive subgradient methods for online learning and stochastic optimization](#). *Journal of machine learning research*, 12:2121–2159.
- Min Gao, Jian-Yu Li, Chun-Hua Chen, Yun Li, Jun Zhang, and Zhi-Hui Zhan. 2023. Enhanced multi-task learning and knowledge graph-based recommender system. *IEEE Transactions on Knowledge and Data Engineering*.
- Alberto García-Durán, Sebastijan Dumancic, 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, Brussels, Belgium, October 31 - November 4, 2018*, pages 4816–4821. Association for Computational Linguistics.
- Rishab Goel, Seyed Mehran Kazemi, Marcus A. Brubaker, and Pascal Poupart. 2020. [Diachronic embedding for temporal knowledge graph completion](#). In *The Thirty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2020, The Thirty-Second Innovative Applications of Artificial Intelligence Conference, IAAI 2020, The Tenth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2020, New York, NY, USA, February 7-12, 2020*, pages 3988–3995. AAAI Press.
- Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 2023. 3d gaussian splatting for real-time radiance field rendering. *ACM Trans. Graph.*, 42(4):139–1.
- Timothée Lacroix, Guillaume Obozinski, and Nicolas Usunier. 2020. [Tensor decompositions for temporal knowledge base completion](#). In *8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, April 26-30, 2020*. OpenReview.net.

- Jennifer Lautenschlager, Steve Shellman, and Michael Ward. 2015. [Icews event aggregations](#).
- Julien Leblay and Melisachew Wudage Chekol. 2018. [Deriving validity time in knowledge graph](#). In *Companion of the The Web Conference 2018 on The Web Conference 2018, WWW 2018, Lyon, France, April 23-27, 2018*, pages 1771–1776. ACM.
- Kalev Leetaru and Philip A Schrod. 2013. [Gdelt: Global data on events, location, and tone, 1979–2012](#). In *ISA annual convention*, volume 2, pages 1–49. Citeseer.
- 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), ACL 2023, Toronto, Canada, July 9-14, 2023*, pages 15460–15474. Association for Computational Linguistics.
- Scott M Lundberg and Su-In Lee. 2017. [A unified approach to interpreting model predictions](#). In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems 30*, pages 4765–4774. Curran Associates, Inc.
- Johannes Messner, Ralph Abboud, and İsmail İlkan Ceylan. 2022. [Temporal knowledge graph completion using box embeddings](#). In *Thirty-Sixth AAAI Conference on Artificial Intelligence, AAAI 2022, Thirty-Fourth Conference on Innovative Applications of Artificial Intelligence, IAAI 2022, The Twelfth Symposium on Educational Advances in Artificial Intelligence, EAAI 2022 Virtual Event, February 22 - March 1, 2022*, pages 7779–7787. AAAI Press.
- Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. 2021. Nerf: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM*, 65(1):99–106.
- Jiaxin Pan, Mojtaba Nayeri, Yinan Li, and Steffen Staab. 2024. [HGE: embedding temporal knowledge graphs in a product space of heterogeneous geometric subspaces](#). In *Thirty-Eighth AAAI Conference on Artificial Intelligence, AAAI 2024, Thirty-Sixth Conference on Innovative Applications of Artificial Intelligence, IAAI 2024, Fourteenth Symposium on Educational Advances in Artificial Intelligence, EAAI 2024, February 20-27, 2024, Vancouver, Canada*, pages 8913–8920. AAAI Press.
- Ali Sadeghian, Mohammadreza Armandpour, Anthony Colas, and Daisy Zhe Wang. 2021. [Chronor: Rotation based temporal knowledge graph embedding](#). In *Thirty-Fifth AAAI Conference on Artificial Intelligence, AAAI 2021, Thirty-Third Conference on Innovative Applications of Artificial Intelligence, IAAI 2021, The Eleventh Symposium on Educational Advances in Artificial Intelligence, EAAI 2021, Virtual Event, February 2-9, 2021*, pages 6471–6479. AAAI Press.
- Jiashuo Sun, Chengjin Xu, Lumingyuan Tang, Saizhuo Wang, Chen Lin, Yeyun Gong, Heung-Yeung Shum, and Jian Guo. 2023. [Think-on-graph: Deep and responsible reasoning of large language model with knowledge graph](#). *arXiv preprint arXiv:2307.07697*.
- Zhiqing Sun, Zhi-Hong Deng, Jian-Yun Nie, and Jian Tang. 2019. [Rotate: Knowledge graph embedding by relational rotation in complex space](#). In *7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6-9, 2019*. OpenReview.net.
- Yu Wang, Nedim Lipka, Ryan A Rossi, Alexa Siu, Ruiyi Zhang, and Tyler Derr. 2024. [Knowledge graph prompting for multi-document question answering](#). In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 19206–19214.
- Chengjin Xu, Yung-Yu Chen, Mojtaba Nayeri, and Jens Lehmann. 2021. [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, NAACL-HLT 2021, Online, June 6-11, 2021*, pages 2569–2578. Association for Computational Linguistics.
- Chengjin Xu, Mojtaba Nayeri, Fouad Alkhoury, Hamed Shariat Yazdi, and Jens Lehmann. 2020a. [Tero: A time-aware knowledge graph embedding via temporal rotation](#). In *Proceedings of the 28th International Conference on Computational Linguistics, COLING 2020, Barcelona, Spain (Online), December 8-13, 2020*, pages 1583–1593. International Committee on Computational Linguistics.
- Chenjin Xu, Mojtaba Nayeri, Fouad Alkhoury, Hamed Yazdi, and Jens Lehmann. 2020b. [Temporal knowledge graph completion based on time series gaussian embedding](#). In *The Semantic Web – ISWC 2020: 19th International Semantic Web Conference, Athens, Greece, November 2–6, 2020, Proceedings, Part I*, page 654–671, Berlin, Heidelberg. Springer-Verlag.
- Rui Ying, Mengting Hu, Jianfeng Wu, Yalan Xie, Xiaoyi Liu, Zhunheng Wang, Ming Jiang, Hang Gao, Linlin Zhang, and Renhong Cheng. 2024. [Simple but effective compound geometric operations for temporal knowledge graph completion](#). In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 11074–11086, Bangkok, Thailand. Association for Computational Linguistics.
- Fuwei Zhang, Zhao Zhang, Xiang Ao, Fuzhen Zhuang, Yongjun Xu, and Qing He. 2022. [Along the time: timeline-traced embedding for temporal knowledge graph completion](#). In *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*, pages 2529–2538.

Linhai Zhang and Deyu Zhou. 2022. [Temporal knowledge graph completion with approximated Gaussian process embedding](#). In *Proceedings of the 29th International Conference on Computational Linguistics*, pages 4697–4706, Gyeongju, Republic of Korea. International Committee on Computational Linguistics.

A Related Work

A.1 Temporal Knowledge Graph Embedding

Temporal knowledge graphs (TKGs) extend conventional knowledge graphs by incorporating time as a dimension in entity and relation representations. A TKG consists of multiple quadruples, each representing a unique fact in the form of (s, r, o, t) , where s is the subject entity, r is the relation, o is the object entity, and t is the timestamp. Many TKGE models build on existing static KGE frameworks to capture the temporal dynamics of evolving data. For example, TTransE (Leblay and Chekol, 2018) is extensions of the TransE (Bordes et al., 2013) model that introduce a temporal aspect to both entity and relation embeddings. The score function of TransE and TransE are $\|s + r - o\|_{l_1/l_2}$ and $\|s + r + t - o\|_{l_1/l_2}$. DistMult (García-Durán et al., 2018) adapts the DistMult model by integrating temporal information, allowing for more accurate temporal representations. DE-Simple (Goel et al., 2020) extends the Simple model by adding diachronic entity embeddings to capture temporal shifts in the entity representations. ChronoR (Sadeghian et al., 2021) and TeRo (Xu et al., 2020a) introduce rotational operations, similar to RotatE, for subject and object entities to evaluate the semantic validity of temporal facts. Additionally, T(NT)ComplEx (Lacroix et al., 2020) and TeLM (Xu et al., 2021) utilize fourth-order tensor decomposition to extend ComplEx for modeling TKGs, providing a more comprehensive framework for temporal relations. RotateQVS (Chen et al., 2022) and TLT-KGE (Zhang et al., 2022), inspired by QuatE, represent temporal knowledge in quaternion space, improving the representation of complex relations. BoxTE (Messner et al., 2022) employs box embeddings to model TKGs, extending the BoxE (Abboud et al., 2020) model for static KGs. Recently, TeAST (Li et al., 2023) proposed mapping relations onto an Archimedean spiral to model time evolution. HGE (Pan et al., 2024) embeds temporal facts into a product space of heterogeneous geometric subspaces and employs a temporal geometric attention mechanism. TCompoundE (Ying et al., 2024) employs both time-

specific and relation-specific geometric operations to capture the intricate temporal dynamics of temporal knowledge graphs.

A.2 3D Gaussian Splatting

3D Gaussian Splatting (Kerbl et al., 2023) is an innovative technique in computer graphics and vision that represents scenes using Gaussian distributions. This method models objects as a collection of Gaussian kernels, each defined by its mean, covariance matrix, transparency, and color, and then projects them onto 2D images using volume rendering techniques. It offers an efficient and high-quality approach to scene reconstruction, surpassing traditional grid-based or implicit neural representations such as NeRF (Mildenhall et al., 2021). Unlike conventional methods, 3D Gaussian Splatting utilizes multiple Gaussian kernels to represent a scene, enabling it to capture complex shapes and blurred boundaries more naturally. This results in improved computational efficiency compared to grid-based models. Recent studies have demonstrated that 3D Gaussian Splatting excels in scene reconstruction and image synthesis, particularly in capturing objects with intricate geometry.

Inspired by 3D Gaussian Splatting, we model entities, relations, and timestamps as 3D Gaussian distributions. This approach optimizes both the mean and covariance of these distributions, effectively capturing temporal changes and representing real-world facts. By integrating time into the Gaussian representation, the model enhances its ability to capture dynamic relationships within TKGs, offering a novel framework for processing temporal data.

B The Proofs of 3DG-TE for Relation Patterns

To prove that the 3DG-TE model can capture important relation patterns, including symmetric, asymmetric, inverse, and temporal evolution patterns, we will analyze the score function mathematically and show how it models each of these relations. The scoring function of 3DG-TE is defined as:

$$\phi(s, r, t, o) = (\mu_s + \mu'_r + \mu'_t) \cdot (\Sigma'_r \cdot \Sigma'_t) \cdot e_o. \quad (15)$$

We follow the method section: all fused terms are vectors in $\mathbb{R}^d$ and the score uses an element-wise tri-linear form. Let $(\cdot)$ denote the Hadamard

product and define

$$\langle \mathbf{a}, \mathbf{b}, \mathbf{c} \rangle \triangleq \sum_{i=1}^d a_i b_i c_i. \quad (16)$$

Recall that the score is

$$\phi(s, r, t, o) = \langle \boldsymbol{\mu}'_{(s,r,t)}, \boldsymbol{\Sigma}'_{(r,t)}, \mathbf{e}_o \rangle, \quad (17)$$

where (from Eq. (8)) $\boldsymbol{\mu}'_{(s,r,t)} = \boldsymbol{\mu}_s + \boldsymbol{\mu}'_r + \boldsymbol{\mu}'_t$ and $\boldsymbol{\Sigma}'_{(r,t)} = \boldsymbol{\Sigma}'_r \cdot \boldsymbol{\Sigma}'_t$ after the information aggregation/distribution step. We also use the mean embedding for entities, i.e., $\mathbf{e}_x = \boldsymbol{\mu}_x$. For convenience, write

$$\mathbf{c}_{r,t} \triangleq \boldsymbol{\mu}'_r + \boldsymbol{\mu}'_t, \quad \mathbf{d}_{r,t} \triangleq \boldsymbol{\Sigma}'_r \cdot \boldsymbol{\Sigma}'_t. \quad (18)$$

Then

$$\phi(s, r, t, o) = \langle \boldsymbol{\mu}_s + \mathbf{c}_{r,t}, \mathbf{d}_{r,t}, \boldsymbol{\mu}_o \rangle. \quad (19)$$

B.1 Symmetric Relation Pattern

Definition. A relation r is symmetric if $\forall s, o, \tau, (s, r, o, \tau) \in \mathcal{G} \Rightarrow (o, r, s, \tau) \in \mathcal{G}$.

Sufficient condition and proof. If

$$\mathbf{c}_{r,t} = \mathbf{0} \quad (\text{i.e., } \boldsymbol{\mu}'_r + \boldsymbol{\mu}'_t = \mathbf{0}), \quad (20)$$

then

$$\begin{aligned} \phi(s, r, t, o) &= \langle \boldsymbol{\mu}_s, \mathbf{d}_{r,t}, \boldsymbol{\mu}_o \rangle \\ &= \langle \boldsymbol{\mu}_o, \mathbf{d}_{r,t}, \boldsymbol{\mu}_s \rangle = \phi(o, r, t, s). \end{aligned} \quad (21)$$

Hence, 3DG-TE can model symmetric relation pattern.

B.2 Asymmetric Relation Pattern

Definition. A relation r is asymmetric if $\forall s, o, \tau, (s, r, o, \tau) \in \mathcal{G}$ and $(o, r, s, \tau) \notin \mathcal{G}$.

Sufficient condition and proof. Consider the difference

$$\Delta\phi \triangleq \phi(s, r, t, o) - \phi(o, r, t, s). \quad (22)$$

Expanding both terms and using $\langle \boldsymbol{\mu}_s, \mathbf{d}, \boldsymbol{\mu}_o \rangle = \langle \boldsymbol{\mu}_o, \mathbf{d}, \boldsymbol{\mu}_s \rangle$, we obtain

$$\Delta\phi = \langle \mathbf{c}_{r,t}, \mathbf{d}_{r,t}, \boldsymbol{\mu}_o - \boldsymbol{\mu}_s \rangle. \quad (23)$$

Thus, whenever $\mathbf{c}_{r,t} \neq \mathbf{0}$ and for some (s, o) one has

$$\langle \mathbf{c}_{r,t}, \mathbf{d}_{r,t}, \boldsymbol{\mu}_o - \boldsymbol{\mu}_s \rangle \neq 0, \quad (24)$$

the two directions receive different scores, so anti-symmetric is representable (existence statement). Hence, 3DG-TE can model antisymmetric relation pattern.

B.3 Inverse Relation Pattern

Definition. Relations r_1, r_2 are inverse if, for relevant (s, o, τ) ,

$$\phi(s, r_1, t, o) = \phi(o, r_2, t, s). \quad (25)$$

Sufficient tying and proof. Tie the elementwise interaction vectors and negate the mean offsets:

$$\mathbf{d}_{r_1,t} = \mathbf{d}_{r_2,t} \triangleq \mathbf{d}_t, \quad \mathbf{c}_{r_2,t} = -\mathbf{c}_{r_1,t} \triangleq -\mathbf{c}_t. \quad (26)$$

Then

$$\begin{aligned} \Delta\phi &\triangleq \phi(s, r_1, t, o) - \phi(o, r_2, t, s) \\ &= \langle \boldsymbol{\mu}_s + \mathbf{c}_t, \mathbf{d}_t, \boldsymbol{\mu}_o \rangle - \langle \boldsymbol{\mu}_o - \mathbf{c}_t, \mathbf{d}_t, \boldsymbol{\mu}_s \rangle \\ &= \underbrace{\langle \boldsymbol{\mu}_s, \mathbf{d}_t, \boldsymbol{\mu}_o \rangle - \langle \boldsymbol{\mu}_o, \mathbf{d}_t, \boldsymbol{\mu}_s \rangle}_{=0} + \langle \mathbf{c}_t, \mathbf{d}_t, \boldsymbol{\mu}_o + \boldsymbol{\mu}_s \rangle \\ &= \langle \mathbf{c}_t, \mathbf{d}_t, \boldsymbol{\mu}_o + \boldsymbol{\mu}_s \rangle. \end{aligned} \quad (27)$$

Hence the inverse constraint holds on any set of pairs where

$$\langle \mathbf{c}_t, \mathbf{d}_t, \boldsymbol{\mu}_o + \boldsymbol{\mu}_s \rangle = 0. \quad (28)$$

A global sufficient condition is $\mathbf{c}_t \cdot \mathbf{d}_t = \mathbf{0}$, which makes $\Delta\phi \equiv 0$ for all (s, o) . Hence, 3DG-TE can model inverse relation pattern.

B.4 Temporal Evolution Pattern

Definition. Relation r_1 and r_2 are *evolving* over time from timestamp τ_1 to timestamp τ_2 , if $\forall s, o, (s, (r_1, \tau_1)) \wedge (s, (r_2, \tau_2)) \in \mathcal{G}$.

Sufficient condition and proof. Because the score depends on (r, t) only through $(\mathbf{c}_{r,t}, \mathbf{d}_{r,t})$, the equality holds whenever

$$\mathbf{c}_{r_1,t_1} = \mathbf{c}_{r_2,t_2} \quad \text{and} \quad \mathbf{d}_{r_1,t_1} = \mathbf{d}_{r_2,t_2}. \quad (29)$$

Indeed,

$$\begin{aligned} \phi(s, r_1, t_1, o) &= \langle \boldsymbol{\mu}_s + \mathbf{c}_{r_1,t_1}, \mathbf{d}_{r_1,t_1}, \boldsymbol{\mu}_o \rangle \\ &= \langle \boldsymbol{\mu}_s + \mathbf{c}_{r_2,t_2}, \mathbf{d}_{r_2,t_2}, \boldsymbol{\mu}_o \rangle \\ &= \phi(s, r_2, t_2, o). \end{aligned} \quad (30)$$

Thus, when the above conditions are satisfied, 3DG-TE can model temporal evolution pattern.

C Benchmark Datasets

ICEWS14 and **ICEWS05-15** are both extracted from the *Integrated Crisis Early Warning System (ICEWS)* dataset (Lautenschlager et al., 2015), which consists of temporal sociopolitical facts starting from 1995. ICEWS14 consists of sociopolitical

events in 2014, while ICEWS05-15 involves events occurring from 2005 to 2015.

GDELT is a subset of the larger *Global Database of Events, Language, and Tone (GDEL*T) (Leetaru and Schrod, 2013) TKG dataset. GDELT contains facts with daily timestamps between April 1, 2015, and March 31, 2016, focusing on the 500 most common entities and 20 most frequent relations. Notably, GDELT holds a large number of quadruples (approximately 2 million) but describes a limited number of entities (500), requiring a strong temporal inductive capacity. Further details about the datasets are provided in Table 3.

	ICEWS14	ICEWS05-15	GDELT
$\#\mathcal{E}$	7,128	10,488	500
$\#\mathcal{R}$	230	251	20
$\#\mathcal{T}$	365	4,017	366
$\#Train$	72,826	386,962	2,735,685
$\#Dev$	8,963	46,092	31,961
$\#Test$	8,941	46,275	31,961

Table 3: Statistics of the TKGE datasets used in the experiment. $\#\mathcal{E}$, $\#\mathcal{R}$, and $\#\mathcal{T}$ represent the number of entities, relations, and timestamps, respectively.

D Experimental Setup

We implement our proposed model 3DG-TE based on the TeAST (Li et al., 2023) training framework. All experiments are conducted on a single NVIDIA Tesla A100 GPU with 80GB of memory. The model is optimized using the Adagrad (Duchi et al., 2011) optimizer, and hyperparameters are tuned via grid search based on validation set performance. The learning rate is set to 0.1. The batch size b is fixed as 2000. The maximum embedding size k no more than 1600. Early stopping is applied to prevent overfitting, with a maximum of 200 training epochs. The reported results correspond to the average performance across five runs. The optimal hyperparameter configurations for 3DG-TE on each dataset are provided in Table 4.

E Visualization of Temporal Evolution Patterns

We visualize the temporal evolution patterns by randomly selecting four facts from the dataset, in-

Dataset	λ_μ	λ_t	k
ICEWS14	0.0045	0.001	1500
ICEWS05-15	0.002	0.1	1600
GDELT	0.001	0.001	1500

Table 4: Optimal hyperparameter configurations for 3DG-TE on different datasets.

cluding the following quadruples: (*Nuri al-Maliki, Make a visit, Iraq, 2014-06-13*), (*Nuri al-Maliki, Consult, Iraq, 2014-06-23*), (*Nuri al-Maliki, Make statement, Iraq, 2014-06-29*), and (*Nuri al-Maliki, Mobilize or increase police power, Iraq, 2014-08-11*). As shown in Figure 3, the red ellipsoids represent the relation 3D Gaussian distributions, while the blue ellipsoids correspond to the time 3D Gaussian distributions. These four facts, visualized in 3D, exemplify how the positions of the ellipsoids formed by relations and time influence the representation of the facts. In each plot, the relation ellipsoid (red) and time ellipsoid (blue) evolve spatially, reflecting the changing temporal dynamics between the entities involved. As the relation and time evolve, the relative positions of the ellipsoids shift, highlighting the distinctive role played by both relational and temporal transformations in modeling dynamic knowledge over time. This again supports our experimental conclusion that relationships and time play an important role in TKG evolution.

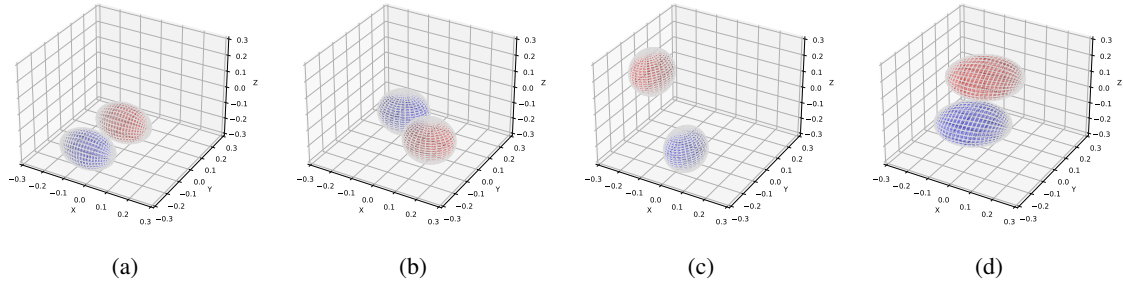


Figure 3: Visualization of Temporal Evolution Patterns in 3D-GTE. The four selected quadruples illustrate the dynamics of relations and timestamps in a temporal knowledge graph. The red ellipsoids represent the relation distributions, while the blue ellipsoids correspond to the time distributions. Four existing facts (a), (b) (c) and (d) are *(Nuri al-Maliki, Make a visit, Iraq, 2014-06-13)*, *(Nuri al-Maliki, Consult, Iraq, 2014-06-23)*, *(Nuri al-Maliki, Make statement, Iraq, 2014-06-29)*, and *(Nuri al-Maliki, Mobilize or increase police power, Iraq, 2014-08-11)*, respectively.