A Allen's Relations

The definitions of 13 Allen's Relations between two-time intervals $\tau_1 = [m_1, n_1], \tau_2 = [m_2, n_2]$ are defined in Figure 4.

Allen's Relations	Pictoral Example	Chronological Sequence
Precedes(τ_1, τ_2)	τ_1 τ_2	$n_1 < m_2$
Preceded_by(τ_1, τ_2)	τ_2 τ_1	$n_2 < m_1$
Meets(τ_1, τ_2)	τ_1 τ_2	$n_1 = m_2$
Met_by(τ_1, τ_2)	τ_2 τ_1	$n_2 = m_1$
Overlaps(τ_1, τ_2)	τ_1 τ_2	$m_1 < m_2 < n_1 < n_2$
Overlapped_by(τ_1, τ_2)	τ_2 τ_1	$m_2 < m_1 < n_2 < n_1$
Starts(τ_1, τ_2)	τ_1 τ_2	$m_1 = m_2 < n_1 < n_2$
Started_by(τ_1, τ_2)	τ_2 τ_1	$m_1 = m_2 < n_2 < n_1$
During(τ_1, τ_2)	τ_1 τ_2	$m_2 < m_1 < n_1 < n_2$
Contains(τ_1, τ_2)	τ_2 τ_1	$m_1 < m_2 < n_2 < n_1$
Finishes(τ_1, τ_2)	τ_2 τ_1	$m_1 < m_2 < n_1 = n_2$
Finished_by(τ_1, τ_2)	τ_1 τ_2	$m_2 < m_1 < n_1 = n_2$
Equal(τ_1, τ_2)	τ_1 τ_2	$m_1 = m_2 < n_1 = n_2$

Figure 4: 13 relations in Allen algebra calculus.

B Extending Backbones to HGE methods

TeRo TeRo (Xu et al. 2020) represents $\boldsymbol{s}, \boldsymbol{p}, \boldsymbol{o}, \boldsymbol{\tau}$ in Complex space as Equation 1. Similarly, we represent $\boldsymbol{s}, \boldsymbol{p}, \boldsymbol{o}, \boldsymbol{\tau}$ in Split-complex space and Dual space as Equation 2 and 3. Following (Xu et al. 2020), we represent $\boldsymbol{s_t}$ and $\boldsymbol{o_t}$ as:

$$\boldsymbol{s_{t\mathbb{M}_i}} = \boldsymbol{s}_{\mathbb{M}_i} \circ \boldsymbol{t}_{\mathbb{M}_i}, \boldsymbol{o_{t\mathbb{M}_i}} = \boldsymbol{o}_{\mathbb{M}_i} \circ \boldsymbol{t}_{\mathbb{M}_i} \tag{9}$$

For temporal relational attention, we set $\boldsymbol{p} = \boldsymbol{p}_s = \boldsymbol{p}_c$, so the dynamic relational information is captured by $\boldsymbol{p} * \boldsymbol{t}$.

TLT-KGE TLT-KGE (Zhang et al. 2022) represents $\boldsymbol{s}, \boldsymbol{p}, \boldsymbol{o}, \boldsymbol{\tau}$ in Complex space as:

$$\boldsymbol{s}_{\mathbb{C}} = \boldsymbol{e}_s + \boldsymbol{t}_\tau^e i, \boldsymbol{p}_{\mathbb{C}} = \boldsymbol{r}_p + \boldsymbol{t}_\tau^r i, \boldsymbol{o}_{\mathbb{C}} = \boldsymbol{e}_o + \boldsymbol{t}_\tau^e i, \tag{10}$$

Similarly, we represent $\boldsymbol{s}, \boldsymbol{p}, \boldsymbol{o}, \boldsymbol{\tau}$ in Dual space and Split-complex space as:

$$\begin{aligned} \boldsymbol{s}_{\mathbb{D}} &= \boldsymbol{e}_s + \boldsymbol{t}_\tau^e \epsilon, \boldsymbol{p}_{\mathbb{C}} = \boldsymbol{r}_p + \boldsymbol{t}_\tau^r \epsilon, \boldsymbol{o}_{\mathbb{C}} = \boldsymbol{e}_o + \boldsymbol{t}_\tau^e \epsilon, \\ \boldsymbol{s}_{\mathbb{S}} &= \boldsymbol{e}_s + \boldsymbol{t}_\tau^e j, \boldsymbol{p}_{\mathbb{S}} = \boldsymbol{r}_p + \boldsymbol{t}_\tau^r j, \boldsymbol{o}_{\mathbb{S}} = \boldsymbol{e}_o + \boldsymbol{t}_\tau^e j, \end{aligned} \tag{11}$$

For temporal relational attention, we adopt $\boldsymbol{r_p}$ and $\boldsymbol{r_{compr}}$ in Equation 12 of original paper as $\boldsymbol{p}_s$, and $\boldsymbol{p}_c * \boldsymbol{\tau}_\tau$ respectively, where $\boldsymbol{r_{compr}} = \boldsymbol{r_p} * \boldsymbol{t_{compr}}$.

C Embeddings in Complex, Dual and Split-Complex Subpaces

Keeping other settings fixed, we train 3 model variants TNTComplEx+complex, TNTComplEx+split, TNTComplEx+dual which use a single geometric space to optimal MRR scores on ICEWS14. We randomly select 100 entities from the entity set and analyze the similarity of their embeddings on different geometric spaces by cosine similarity:

$$S_C(A, B) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\| \|\mathbf{B}\|} = \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n A_i^2 \cdot \sum_{i=1}^n B_i^2}} \tag{12}$$

We concat the real part and imaginary part of one entity when calculating the cosine similarity. From Figure 5, we could find out that in every sub-graph, the values on the diagonal, which represent the cosine similarity between entity embeddings of the same entity on different geometric subspace, are always the highest in a row and exceed 0.95. Therefore, although their geometric interpretations are diverse, real and imaginary vectors in different subspaces are almost unanimous when trained to optimal settings.

D Dataset Overview

Dataset statistics are described in Table 5.

E Temporal Pattern Statistics

We calculate the occurrence of each temporal pattern introduced in Section 4 to give an overview distribution of the temporal patterns. Table 4 shows the statistics on ICEWS14, ICEWS05-15 and GDELT. If a group of quadruples, such as the examples shown in the 2nd column in Table 4, meets the definition in Section 4, we calculate it as one occurrence.

F Modeling Various Temporal Patterns

Proposition 5. *HGE can model (anti-)symmetry patterns introduced in Definitions 3 and 5.*

Proof. Let p be a relation with temporal symmetry. One condition for modeling this pattern is $S(s, p, \bar{o}, \tau) = S(o, p, \bar{s}, \tau)$. For simplicity of representation, we use $p_t = p_{s\tau}$. Without loss of generality, we assume that we have only a one-dimensional split-complex vector. Therefore, we have the following equality to fulfill temporal symmetry:

$$\begin{aligned} &\big(s_a p_{ta} o_a + s_b p_{tb} o_a - s_a p_{tb} o_b - s_b p_{ta} o_b\big) = \\ &\big(o_a p_{ta} s_a + o_b p_{tb} s_a - o_a p_{tb} s_b - o_b p_{ta} s_b\big). \end{aligned}$$

This leads to the following equality $s_b p_{tb} o_a = s_a p_{tb} o_b$. To hold this equality, we need to have either $p_{tb} = 0$ or $s_b o_a = s_a o_b$. So far, we show for a given grounded quadruple (s, p, o, τ), if our model learns (s, p, o, τ) to be true, it can also hold its temporal symmetry (o, p, s, τ) as true. To generalize this to the universal quantifier (every grounded quadruple), we can add one extra dimension to model temporal symmetry for the extra pair of entities. In the extended dimension for the new pair (s,o), we should have $p_{tb} = 0$ or $s_b o_a = s_a o_b$ to hold temporal symmetry. In this way, all pairs (s, o) which are connected by temporal symmetry relation will be held as true by the model. A similar procedure can be done for Dual and ComplEx spaces. Therefore, there exist assignments for embeddings of entities and relations that fulfill the encoding of the temporal symmetric pattern. $\square$

Table 4: Real examples and statistics of each pattern in the train set of ICEWS14, ICEWS05-15, GDELT.

Patterns	Examples	ICEWS14	ICEWS05-15	GDELT
static symmetric	(Iraq, sign formal agreement, Iran, 2014-04-06) (Iran, sign formal agreement, Iraq, 2014-04-06)	6,506	36,537	366,830
static inverse	(Fiji, host a visit, Julie Bishop, 2014-11-04), (Julie Bishop, make a visit, Fiji, 2014-11-04)	10,361	63,092	552,280
dynamic symmetric	(France, engage in negotiation, Poland, 2014-04-04) (Poland, engage in negotiation, France, 2014-02-20)	78,473	3,817,343	17,265,293
dynamic inverse	(Angela Merkel, discuss by telephone, Ukraine, 2014-03-14), (Ukraine, consult, Angela Merkel, 2014-03-27)	768,586	48,641,730	104,909,248
dynamic evolve	(South Korea, demand, Japan, 2014-07-15), (South Korea, reject judicial cooperation, Japan, 2014-07-18)	971,055	63,733,447	112,653,245

Proposition 6. *HGE can model inverse patterns introduced in Definitions 4 and 6.*

Proof. Let temporal relation p_1 be the inverse of the temporal relation p_2 at all time points (temporalinv). One condition to model this pattern is to fulfill $S(s, p_1, \bar{o}, \tau) = S(o, p_2, \bar{s}, \tau)$. Without loss of generality, we assume that we have only a one-dimensional split-complex vector. Therefore, we have the following equality to fulfill temporal inverse relationships:

$$\big(s_a p_{1ta} o_a + s_b p_{1tb} o_a - s_a p_{1tb} o_b - s_b p_{1ta} o_b\big) = \big(o_a p_{2ta} s_a + o_b p_{2tb} s_a - o_a p_{2tb} s_b - o_b p_{2ta} s_b\big).$$

If we set $p_{1ta} = p_{2ta},\ p_{1tb} = -p_{2tb}$, the above equality holds. This means there exist assignments for embeddings of entities, relations, and times that fulfill the encoding of temporal inverse patterns. Our proof can be generalized to d dimensional product space by adding one dimension per each grounded atom. For the pattern in delayedtemporalinv, the proof procedes likewise. The only difference is that the time embedding will be different at the two times τ_1, τ_2 to hold $p_{1ta} = p_{2ta},\ p_{1tb} = -p_{2tb}$. ☐

Proposition 7. *Let us assume that relation p_1 evolve to relation p_2 as formalized in evolvepattern. HGE can model this pattern.*

Proof. Given that p_1 evolves to p_2, and also given the two times τ_1 and τ_2 with $\tau_1 \prec\leq \tau_2$, to model the pattern, we need to have $S(s, p_1, \bar{o}, \tau_1) = S(s, p_2, \bar{o}, \tau_2)$. Without loss of generality, we assume that we have only a one-dimensional split-complex vector. Then, we must fulfill the following equality:

$$\big(s_a p_{1t_1a} o_a + s_b p_{1t_1b} o_a - s_a p_{1t_1b} o_b - s_b p_{1t_1a} o_b\big) = \big(s_a p_{2t_2a} o_a + s_b p_{2t_2b} o_a - s_a p_{2t_2b} o_b - s_b p_{2t_2a} o_b\big).$$

For this equality to hold, it must be the case that $p_{1t_1a} = p_{2t_2a},\ p_{1t_1b} = p_{2t_2b}$. Note that these equality conditions do not necessarily mean that the embedding of static and temporal relations in Equation 5 should be the same because different convex combinations can create the same vector for temporal relations. Considering the universal quantifier, we can add one extra dimension for each grounded atom to fulfill equality. A similar consideration can be applied to Dual and ComplEx spaces. Therefore, there exist assignments for embeddings of entities and relations that encode the patterns. ☐

Proposition 8. *Let p be a temporary relation in time as defined in temprelationt1t2. HGE can model this relation.*

Proof. Let p be a temporary relation as in temprelationt1t2. To follow this pattern in the embedding space, for a given grounded atom (s, p, o, τ_1), there exist τ_0, τ_2 and also the embedding vectors for $s, p, o, \tau_0, \tau_1, \tau_2$ such that we have $S(s, p, \bar{o}, \tau_1) \neq S(s, p, \bar{o}, \tau_2)$ and $S(s, p, \bar{o}, \tau_1) \neq S(s, p, \bar{o}, \tau_0)$ as one possible condition to fulfill the pattern. Similar to the previous proofs, let us assume that we have only a one-dimensional split-complex vector. To fulfill the first condition (the second one will be similar), we have

$$\big(s_a p_{t_1a} o_a + s_b p_{t_1b} o_a - s_a p_{t_1b} o_b - s_b p_{t_1a} o_b\big) \neq \big(s_a p_{t_2a} o_a + s_b p_{t_2b} o_a - s_a p_{t_2b} o_b - s_b p_{t_2a} o_b\big).$$

This can be simply fulfilled if we set $p_{t_1a} \neq p_{t_2a}, p_{t_1b} \neq p_{t_2b}$. In addition, we can have a large value for $S(s, p, \bar{o}, \tau_1)$ and a small value for $S(s, p, \bar{o}, \tau_2)$ (or vice versa) by properly setting the temporal relation close to zero at time τ_1 and high value at time τ_2 (and vice versa). A similar calculation can be done for Dual and ComplEx spaces. Therefore, there exist assignments for embeddings of entities and relations that encode the pattern. ☐

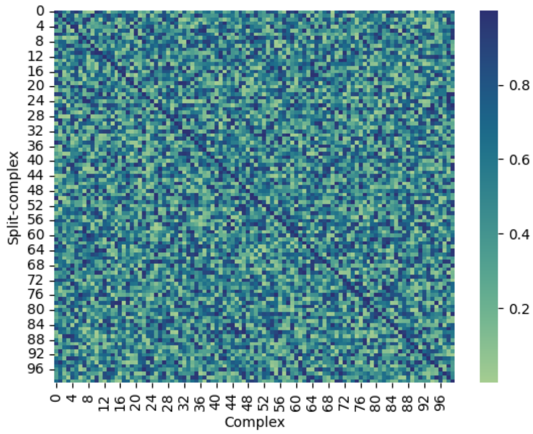
G Experiment Details

All experiments in the paper were run on the same NVIDIA A100 GPU device(40G GPU/100G CPU) with Ubuntu system 22.0. We implement a grid search to select the best regularizer weight from [5e-4, 3e-3, 5e-3, 3e-3, 1e-3, 3e-2, 1e-2, 1e-1]. A detailed list of hyperparamters is provided in hyperparamter.pdf file in the code folder of supplement material.

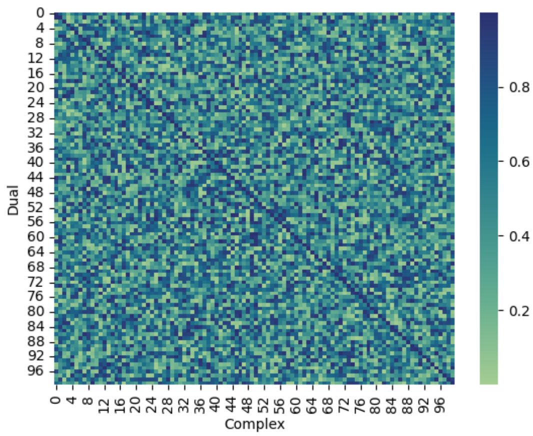
H Temporal Structural Patterns on Geometric subspaces

We consider symmetric patterns belonging to structural patterns too and define two other types of temporal structural patterns:

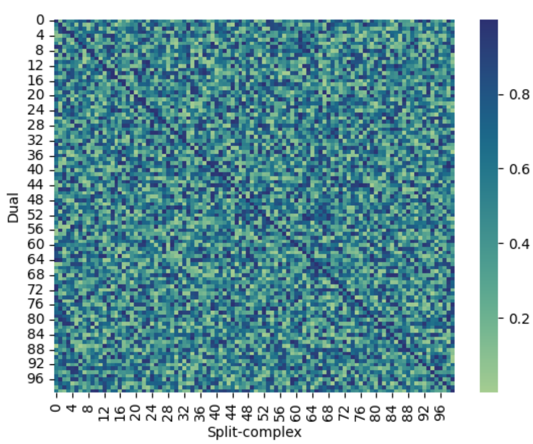
Definition 9. *Relation p forms a temporal star of size $n \in \mathbb{N}$ iff* $\forall s : \exists o_1, \tau_1 \dots, o_n, \tau_n : Precedes(\tau_1, \tau_2)\ \&$



(a) Cosine similarity score between trained entity embeddings in Complex space and Split-complex space.



(b) Cosine similarity score between trained entity embeddings in Complex space and Split-complex space.



(c) Cosine similarity score between trained entity embeddings in Split-complex space and Dual space.

Figure 5: Cosine similarity scores between entity embeddings from different geometric space. x-axis and y-axis show the entity id on relevant geometric space

Table 5: Statistics for ICEWS14, ICEWS05-15, GDELT and Wikidata12k.

Dataset	ICEWS14	ICEWS05-15	GDELT	Wikidata12k
Entities	7,128	10,488	500	12,554
Relations	230	251	20	24
Times	365	4017	366	1,726
Train	72,826	386,962	2,735,685	32,497
Validation	8,941	46,275	341,961	4,062
Test	8,963	46,092	341,961	4,062

$\ldots \ \& \ Precedes(\tau_(n-1), \tau_n) \ \& \ (s, p, o_1, \tau_1) \ \& \ (s, p, o_2, \tau_2) \ \& \ \ldots \ \& \ (s, p, o_n, \tau_n)$.

Definition 10. *A relation p forms a temporal hierarchy iff* $\forall v_1, v_2, v_3, \tau_1, \tau_2 : (v_1, p, v_2, \tau_1) \ \& \ (v_2, p, v_3, \tau_2) \rightarrow \tau_1 \prec \tau_2$

We investigate if heterogeneous geometric subspaces could represent different kinds of structural patterns. We extract 3 subsets for static symmetry, temporal hierarchy, and temporal star structural patterns from the test set of ICEWS14 and ICEWS05-15. Four variants of TNTComplEx+HGE model are tested in these subsets: 1) complex: only complex space is used. 2) split-complex: only split-complex space is used. 3) dual: only dual space is used. 4) HGE: the full model with three heterogeneous subspaces.

Table 6 shows that models using complex space perform best on static symmetric structural patterns. Models using split-complex space performs best on temporal hierarchy pattern while models using dual space perform best on temporal star pattern. This observation supports our core assumption that multiple geometric spaces may benefit temporal knowledge graph representation. Moreover, TNTComplEx+HGE performs better than all variants with single geometric spaces, demonstrating that the proposed product space with temporal geometric attention mechanism could integrate the advantages of individual subspaces.

I HGE's Time and Space Usage

As HGE reuses vectors for different geometric subspaces, the increased parameters to implement an HGE module will be $2|\mathcal{R}| * d$, which is the attention weights for two proposed attention mechanisms. We demonstrate the HGE's efficiency by comparing the number of parameters and running times of the original backbone with HGE-extended backbones. All models are trained with 200 epochs and we calculate the average running time of training epochs for each model. From Table 7, we observe that with the same embedding dimension d=1200 for entities and relations, the increased number of parameters and running time are rather moderate for HGE-extensions. Specifically, when TNTComplEx is extended by HGE, its performance is comparable to TLT-KGE with only half as many parameters and a shorter running time. Even if we decrease d of TNTComplEx+HGE to 1100, it still outperforms backbone TNTComplEx(d=1200) with fewer parameter numbers. This demonstrates that HGE's improvements do not come from the increased number of parameters, but rather from its representational approach.

Table 6: MRR performance of heterogeneous geometric spaces on diverse structural pattern subsets.

Datasets	Structural Patterns	Statistics	TNTComplEx	complex	split-complex	dual	HGE
ICEWS14	static symmetric	1352	98.8	**99.5**	99.3	98.3	**99.5**
	temporal hierarchy	1193	69.5	70.4	**71.8**	71.0	**71.8**
	temporal star	6197	70.5	71.6	71.9	<u>72.9</u>	**73.0**
ICEWS05-15	static symmetric	7240	99.7	**99.8**	99.7	99.6	**99.8**
	temporal hierarchy	16703	72.7	72.8	<u>73.7</u>	72.5	**74.3**
	temporal star	39724	73.8	73.8	72.4	<u>74.7</u>	**75.4**

Table 7: Parameter number and average runtime for original backbones and backbones extended by HGE.

Datasets	Model	Rank(d)	Parameter number	Average epoch time(s)	MRR
ICEWS14	TNTComplEx	1200	20,191,200	1.80	60.7
	TLT-KGE	1200	38,693,400	2.25	63.0
	TNTComplEx+HGE	1100	19,520,600	2.10	62.9
	TNTComplEx+HGE	1200	21,295,200	2.19	63.0
ICEWS05-15	TNTComplEx	1200	37,221,600	11.79	66.6
	TLT-KGE	1200	81,360,600	13.91	68.6
	TNTComplEx+HGE	1100	35,224,200	11.52	67.7
	TNTComplEx+HGE	1200	38,426,400	12.13	68.1

Table 8: Results of LCGE in original paper and by our implementation

Model	ICEWS14				ICEWS05-15			
	MRR	Hits@1	Hits@3	Hits@10	MRR	Hits@1	Hits@3	Hits@10
LCGE(Niu and Li 2023)	92.5	91.6	92.9	93.7	91.2	90.3	91.6	92.5
LCGE	61.6	53.2	66.7	77.5	61.8	51.4	68.1	81.2

J Baseline Selection

LCGE We found out the commonsense reasoning score introduced in equation 11 of LCGE(Niu and Li 2023) was considered during the training time but missed during the test time, which causes bias to final scores and rankings. We re-implemented the codes and attached our implementation in the supplementary material's code/LCGE_new folder. Table 8 shows the comparison of reported results in (Niu and Li 2023) and results by our implementation.

DyERNIE We do not include the baseline of DyERNIE(Han et al. 2020) since this paper reports the results using the static filtered setting. Moreover, the code released by the authors is not complete to implement hyperbolic spaces, making it hard to report time-aware filtering results.

HSAE HSAE(Ren et al. 2023) adopts a hierarchy self-attention mechanism to incorporate information from different time shots. We do not include the baseline of HSAE because the author does not publish the codes.