MAGIC: Deep Geometric Evolution with Structural Consensus for Temporal Knowledge Graph Reasoning

Chengao Liu[1], **Yuan Li**[2,*], **Yingze Wang**[1], **Jianbin Jiao**[1]
[1]School of Emergency Management Science and Engineering,
University of Chinese Academy of Sciences, Beijing, 100049, China
[2]School of Electronics, Electrical and Communication Engineering,
University of Chinese Academy of Sciences, Beijing, 100049, China
{liuchengao21, wangyingze24}@mails.ucas.ac.cn
{liyuan23, jiaojb}@ucas.ac.cn

Abstract

Temporal Knowledge Graph (TKG) reasoning remains challenging to characterize with conventional flat representations due to its intrinsic heterogeneous structure. Existing multi-geometry approaches face two key bottlenecks: 1) the Riemannian depth barrier driven by numerical instability, which restricts models to shallow architectures; and 2) gate collapse, where adaptive fusion mechanisms suffer from gradient starvation and degenerate into single-geometry solutions. To this end, we propose MAGIC (**M**ulti-geometry **A**nnealing **G**raph **I**nteraction with **C**onsensus). Our framework introduces a Tangent-Residual Engine in multi-geometric spaces, which enables the first stable 8-layer geometric evolution and reveals a phenomenon termed Geometric Annealing, where manifold curvature spontaneously evolves from semantic flatness in shallow layers to structural complexity in deeper layers. We further design an explicit reasoning module with structural consensus, leveraging geometric invariants and structural priors to regulate gradient flow, prevent collapse, and ensure robust synergy across Hyperbolic, Spherical, and Euclidean spaces. Experiments show that MAGIC achieves state-of-the-art performance in TKG reasoning, improving MRR by up to 2.9 points.

1 Introduction

Temporal Knowledge Graphs (TKGs) incorporate the dynamic evolution of factual knowledge and entity semantics into structured representations, providing crucial support for NLP applications from question answering to event forecasting. However, behind this temporality lie highly heterogeneous topological structures: hierarchical relations (e.g., organizational frameworks) require models capable of tree-like expansion, while periodic events (e.g., elections, diplomatic visits) correspond to cyclic and spherical geometries.

Traditional TKG reasoning approaches mostly embed data into a single Euclidean space (Xu et al., 2021) or simply extend static embedding techniques to the temporal dimension, leading to severe distortion of heterogeneous geometric structures, commonly referred to as the "Euclidean bottleneck" (Nickel and Kiela, 2017). As illustrated in Figure 1A, flat Euclidean spaces inevitably flatten tree-like hierarchies and break periodic cycles. In recent years, multi-geometric representation learning has emerged as a promising direction to address this geometric mismatch by integrating hyperbolic, spherical, and Euclidean spaces to jointly model hierarchical and periodic patterns (Wang et al., 2024). However, these methods remain hindered by two fundamental challenges that limit their reasoning capability (Figure 1B):

1) *The Riemannian Depth Barrier*—due to numerical instability in exponential/logarithmic maps and the accumulation of projection errors across layers (Wang et al., 2021), existing geometric architectures are typically confined to shallow designs (with depth $L \leq 2$), which prevents them from capturing long-range temporal dependencies and multi-hop reasoning paths.

2) *Gate Collapse*—current adaptive geometric fusion mechanisms (e.g., attention gating) are also prone to "gradient starvation" (Wang et al., 2020) early in training. This is because we observed that different datasets can cause gradient norms in a particular space to become significantly larger. This leads to the greedy optimization of that space and the suppression of other geometric spaces, ultimately causing the model to degenerate into a single-geometry representation and lose the expressive advantage of multi-geometry fusion.

To resolve these limitations, we propose **MAGIC** (**M**ulti-geometry **A**nnealing **G**raph **I**nteraction with **C**onsensus). First, we introduce multiple geometric spaces (spherical, hyperbolic, and Euclidean) to model TKGs, aiming to break

* Corresponding Author.

Proceedings of the 64th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 22122–22133

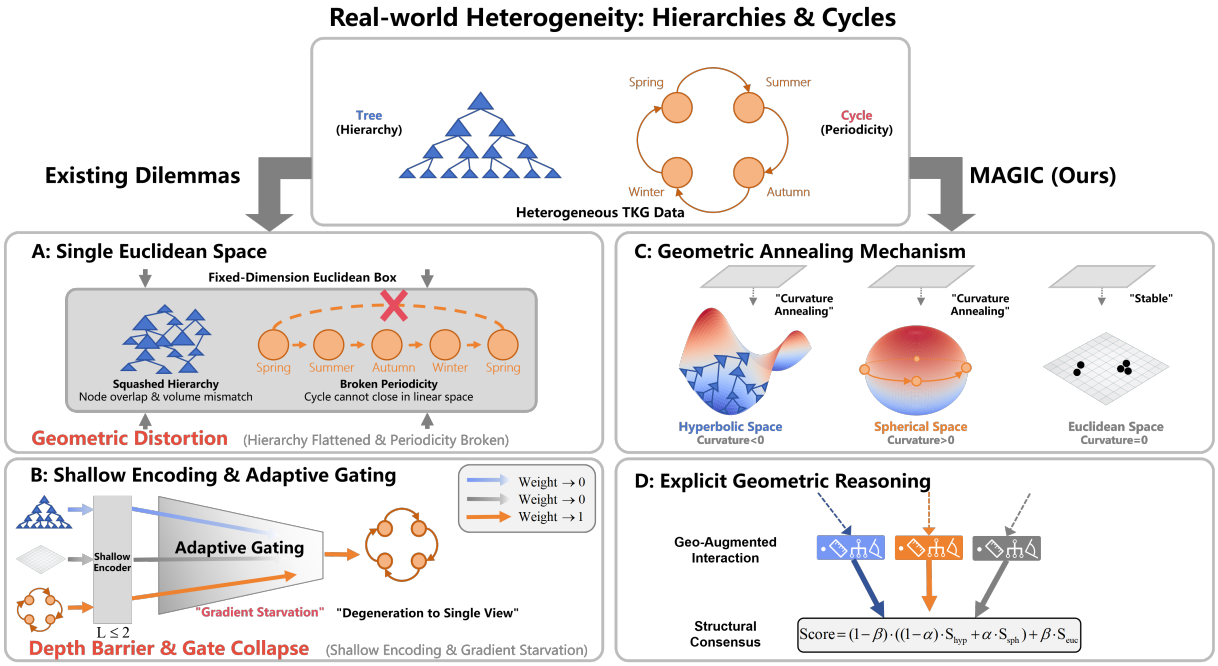

Figure 1: **Motivation and Methodological Comparison. (Left) Existing Dilemmas:** (A) *Geometric Distortion*: Single Euclidean space flattens hierarchies and breaks periodic cycles. (B) *Instability & Collapse*: Riemannian instability limits architectures to shallow encoders ($L \leq 2$), while adaptive gating suffers from gradient starvation, leading to single-view degeneration. **(Right) MAGIC (Ours):** (C) *Deep Evolution*: Our Tangent-Residual Engine enables stable 8-layer evolution, inducing *Geometric Annealing* from semantic flatness to structural complexity. (D) *Robust Reasoning*: Explicit invariant injection and fixed structural priors regulate gradient flow to prevent gate collapse.

the limitations of single-space modeling. To enable deep geometric evolution, we design a Tangent-Residual Engine to address the depth barrier. It constructs a stable Euclidean gradient pathway in the tangent space, enabling the first stable deep geometric evolution (up to $L = 8$) and inducing a phenomenon termed geometric annealing: the model tends toward flat geometry in shallow layers to align semantics, while spontaneously increasing curvature in deeper layers to accommodate complex topological structures (Figure 1C).

Second, to counter gate collapse, we propose explicit geometric reasoning with structural consensus. Instead of relying on unstable implicit fusion, we explicitly inject geometric invariants—geodesic distance (Locality Filter), norms (Hierarchy Depth), and angles (Semantic Compass). Furthermore, we enforce a structural consensus strategy. We reframe fusion weights not as learnable parameters prone to overfitting, but as fixed structural priors. This mechanism regulates gradient flow, compelling the model to optimize representations across all manifolds simultaneously and preventing the suppression of non-dominant geometries (e.g., Spherical cycles) (Figure 1D).

Our contributions are summarized as follows:

- **Diagnosis (Gate Collapse):** We demonstrate that existing adaptive fusion methods degenerate into single-view solutions due to gradient starvation, ignoring crucial patterns like periodic events.
- **Mechanism (Deep Evolution):** The MAGIC model achieves geometric annealing via a Tangent-Residual Engine, evolving from shallow semantic to deep structural reasoning, while its fixed-prior consensus mechanism regulates gradient flow to prevent gate collapse.
- **Performance (SOTA):** MAGIC achieves SOTA results (e.g., 82.5% MRR on ICEWS05-15), demonstrating superior modeling of long-term dynamics.

2 Related Work

Temporal Knowledge Graph Completion. Existing TKG reasoning methods largely extend static embedding techniques to the temporal domain. Early translation-based models, such as TTransE (Leblay and Chekol, 2018) and HyTE (Dasgupta et al., 2018), project entities onto temporal hyper-

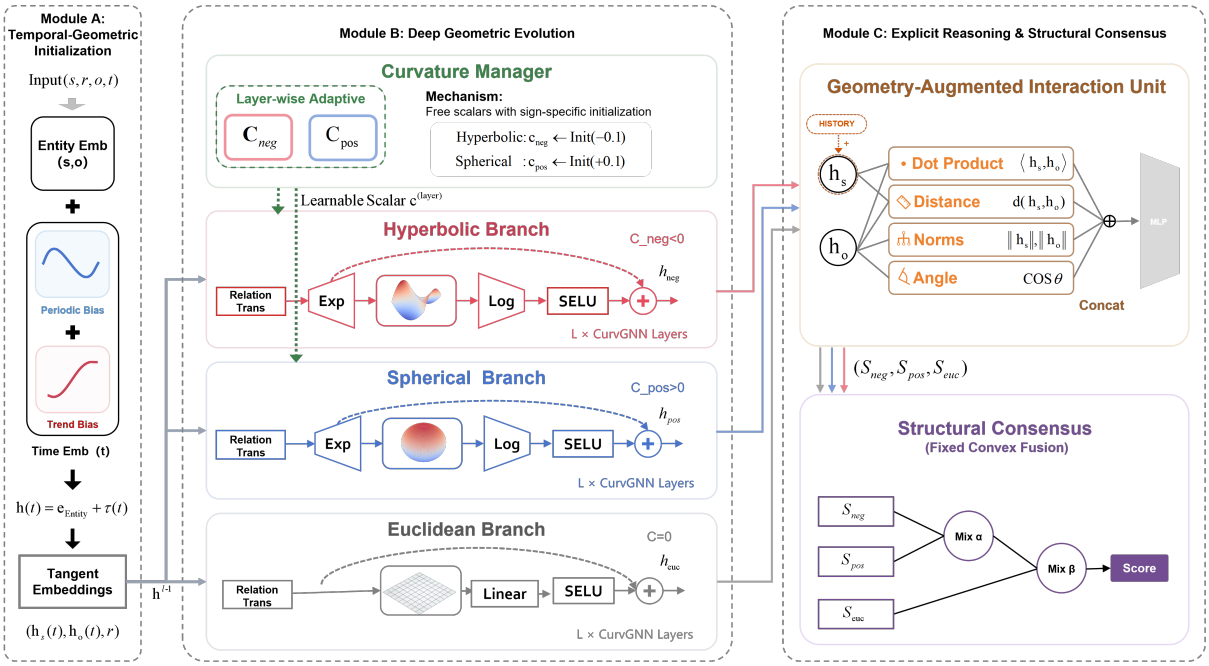

Figure 2: Overview of MAGIC.

planes. Rotation-based approaches like ChronoR (Sadeghian et al., 2021) and RotatE (Sun et al., 2019) model relations as complex space rotations, while tensor decomposition methods (e.g., TComplEx (Lacroix et al., 2020), TeLM (Xu et al., 2021)) capture high-order interactions. Other works like QDN (Wang et al., 2023) and BoxTE (Messner et al., 2022) further address temporal uncertainty. However, these methods primarily operate in flat Euclidean space, inherently imposing significant distortion when modeling heterogeneous topologies like hierarchies and cycles (Nickel and Kiela, 2017).

Geometric Graph Representation Learning. To mitigate Euclidean distortion, geometric learning leverages constant curvature manifolds. For static graphs, MuRMP (Wang et al., 2021) and ATTH (Chami et al., 2020) pioneered the use of product manifolds to simultaneously model hierarchies and cycles. In the temporal domain, recent works have integrated geometric bias: DyERNIE (Han et al., 2020) utilizes velocity vectors for evolution, while BiQCap (Zhang et al., 2023) combines biquaternions with hyperbolic space. More recently, Liu et al. (2025) introduced PoinKI, which leverages Poincaré embeddings to capture evolving hierarchical structures in continuous-time dynamic graphs. Most relevant to our work is IME (Wang et al., 2024), which fuses parallel Hyperbolic, Spherical, and Euclidean spaces. However, it relies on simple pooling mechanisms and lacks the deep structural evolution required for robust reasoning.

Positioning against LLM-based Approaches. While LLMs have revolutionized semantic reasoning, recent works like TCompoundE (Ying et al., 2024) and TeAST (Li et al., 2023) demonstrate that *rigorous geometric modeling* remains indispensable for capturing complex topological structures (e.g., spirals) that text-based models often overlook. Unlike LLM-based paradigms that are prone to hallucination, structure-centric frameworks like MAGIC provide deterministic reasoning paths and superior efficiency.

3 Method

We present MAGIC, a unified framework tailored for the heterogeneous topology (e.g., hierarchies and cycles) of TKGs. To overcome the limitations of shallow architectures and adaptive collapse, MAGIC orchestrates a deep geometric evolution across Hyperbolic, Spherical, and Euclidean manifolds. As illustrated in Figure 2, MAGIC operates in three strategic phases. (A) *Temporal-Geometric Initialization* decouples entity dynamics into static and periodic components within a unified tangent space. (B) *Deep Geometric Evolution* utilizes a Tangent-Residual Engine to enable stable deep reasoning ($L = 8$) and induce curvature annealing. (C) *Explicit Reasoning* injects geometric invariants

and enforces structural consensus to prevent gate collapse.

3.1 Problem Definition

A TKG is denoted as $\mathcal{G} = \{(s, r, o, t)\} \subseteq \mathcal{E} \times \mathcal{R} \times \mathcal{E} \times \mathcal{T}$, where $\mathcal{E}, \mathcal{R}, \mathcal{T}$ represent the sets of entities, relations, and timestamps. The goal is to predict missing links, i.e., $(s, r, ?, t)$.

We operate on three distinct manifolds (forming a product manifold signature (Gu et al., 2019; Dong et al., 2025)): Hyperbolic space $\mathbb{H}^d_{c_{neg}}$ (Poincaré ball with curvature $c_{neg} < 0$) for hierarchies, Spherical space $\mathbb{S}^d_{c_{pos}}$ (Hypersphere with $c_{pos} > 0$) for cycles, and Euclidean space $\mathbb{E}^d$ ($c = 0$) for flat regularities. Operations are defined via the exponential map $\exp^c_{\mathbf{x}}(\mathbf{v}) : T_{\mathbf{x}}\mathcal{M} \to \mathcal{M}$, and the logarithmic map $\log^c_{\mathbf{x}}(\mathbf{y}) : \mathcal{M} \to T_{\mathbf{x}}\mathcal{M}$.(Detailed definitions of these operations and numerical stability constraints are provided in Appendix B.) Here, the exponential map $\exp^c_{\mathbf{x}}(\mathbf{v})$ projects a tangent vector $\mathbf{v}$ from the tangent space $T_{\mathbf{x}}\mathcal{M}$ to a point $\mathbf{y}$ on the manifold $\mathcal{M}$, while the logarithmic map $\log^c_{\mathbf{x}}(\mathbf{y})$ performs the inverse.

3.2 Temporal-Geometric Initialization

Effective geometric evolution requires a rich semantic initialization. Unlike static KGs, entities in TKGs exhibit heterogeneous time-dependent dynamics. To capture these patterns without prematurely distorting them in curved spaces, we propose a multi-granularity temporal decoupling strategy within a unified tangent space (approximated as $\mathbb{E}^d$ at the origin). We construct time-aware representations in a unified tangent space (approximated as $\mathbb{E}^d$) at the origin before geometric projection. For a query $(s, r, ?, t)$, the input embedding $\mathbf{h}_s(t)$ is decoupled into static, periodic, and trend components:

$$\mathbf{h}_s(t) = \mathbf{e}_s + \underbrace{\text{Norm}\left(\mathbf{\Phi}_{\text{per}}(t) \oplus \mathbf{\Phi}_{\text{trend}}(t)\right)}_{\text{Temporal Bias } \mathbf{\Phi}(t)} \quad (1)$$

where $\mathbf{e}_s \in \mathbb{R}^d$ denotes the static entity identity, and $\text{Norm}(\cdot)$ denotes Layer Normalization, applied to stabilize the scale of the temporal bias before addition. To model time at different frequencies, we discretize t into Year (y), Month (m), and Day (d) granularities.

The periodic bias $\mathbf{\Phi}_{\text{per}}(t)$ utilizes sinusoidal functions to capture seasonality:

$$\mathbf{\Phi}_{\text{per}}(t) = [\sin(\mathbf{w}_{y1} \odot \mathbf{t}_y),\ \sin(\mathbf{w}_{m1} \odot \mathbf{t}_m),\ \sin(\mathbf{w}_{d1} \odot \mathbf{t}_d)] \quad (2)$$

Conversely, the trend bias $\mathbf{\Phi}_{\text{trend}}(t)$ employs hyperbolic tangent ($\tanh$) functions to model long-term drift and state saturation:

$$\mathbf{\Phi}_{\text{trend}}(t) = [\tanh(\mathbf{w}_{y2} \odot \mathbf{t}_y),\ \tanh(\mathbf{w}_{m2} \odot \mathbf{t}_m),\ \tanh(\mathbf{w}_{d2} \odot \mathbf{t}_d)] \quad (3)$$

Here, $\mathbf{w}_* \in \mathbb{R}^{d'}$ are learnable frequency embeddings, and $\odot$ denotes element-wise multiplication with broadcasted time indices.

To incorporate relational semantics r into the query, we parameterize each relation as a static embedding $\mathbf{r} \in \mathbb{E}^d$ in the same unified tangent space. We adopt a tangent translation strategy for fusion: the relation $\mathbf{r}$ acts as a displacement vector added to the subject embedding, formally $\mathbf{h}_{query}(t) \approx \mathbf{h}_s(t) + \mathbf{r}$. This Euclidean addition in the tangent space efficiently approximates the complex parallel transport required on curved manifolds, maintaining the translational intuition of knowledge graph embeddings (e.g., TransE) while remaining compatible with our deep geometric evolution module.

3.3 Deep Geometric Evolution

Traditional geometric GNNs face a severe depth bottleneck, typically collapsing beyond 3 layers due to the numerical instability of Riemannian operations. To overcome this, we propose a deep geometric evolution module that enables stable scaling to deeper architectures (e.g., $L = 8$).

Geometric Annealing Mechanism To strictly control topological evolution, we introduce a Curvature Manager (Figure 2B). Instead of fixing curvature globally, it parameterizes the curvature of each layer l as a learnable scalar $c^{(l)}$. Mechanism: To ensure correct geometric inductive bias while maintaining stability, we employ a *Sign-Specific Initialization* strategy: the hyperbolic curvature is initialized to a small negative value ($c_{neg} \approx -0.1$) and the spherical curvature to a small positive value ($c_{pos} \approx +0.1$). This places the model in a Quasi-Euclidean state ($|c| \approx 0$) at the onset of training, creating a flat "safe zone" for semantic alignment. Driven by task gradients without explicit constraints, we observe a spontaneous phenomenon termed Geometric Annealing (Figure 3): learned curvatures naturally evolve from near-zero in shallow layers to high absolute values in deep layers. We attribute this to the model's intrinsic

optimization strategy: 1) *Shallow Phase (Semantic Alignment):* The model retains flatness ($c \approx 0$) to linearly align heterogeneous semantic features without distortion, mirroring the manifold property that local neighborhoods are well-approximated by Euclidean tangent spaces. 2) *Deep Phase (Structural Expansion):* As the receptive field grows exponentially, the model increases curvature to expand manifold volume capacity, accommodating complex hierarchical (Hyperbolic) and cyclical (Spherical) structures that cannot be embedded in flat spaces without distortion.

The Tangent-Residual Engine. To support this deep evolution, we design a layer that performs non-linear transformations on the manifold but aggregates residuals in the tangent space. Crucially, by leveraging the tangent space at the origin ($\mathbf{x} = \mathbf{0}$) and projecting features back via $\log_{\mathbf{0}}^{c^{(l)}}$ before the residual addition, we bypass the numerically unstable Möbius addition. This constructs a Euclidean gradient pathway that facilitates direct gradient flow, effectively resolving the vanishing gradient problem in deep geometric networks. Formally, let $\mathbf{h}^{(l-1)}$ be the input in the tangent space. The update rule for layer l is defined as:

$$\begin{aligned} \mathbf{h}_{\mathcal{M}} &= \exp_{\mathbf{0}}^{c^{(l)}}(\mathbf{h}^{(l-1)}) \\ \mathbf{h}_{linear} &= \mathbf{W}^{(l)} \otimes_{c^{(l)}} \mathbf{h}_{\mathcal{M}} \\ \mathbf{h}_{tan} &= \log_{\mathbf{0}}^{c^{(l)}}(\text{Project}(\mathbf{h}_{linear})) \\ \mathbf{h}^{(l)} &= \sigma(\mathbf{h}_{tan}) + \mathbf{h}^{(l-1)} \end{aligned} \tag{4}$$

3.4 Explicit Geometric Reasoning

Explicit Geometric Interaction. Standard dot-product decoders discard topological information. We explicitly inject geometric invariants into the reasoning process. For each branch, we construct an interaction feature vector $\mathbf{f}_{s,o}$:

$$\mathbf{f}_{s,o} = \left[\underbrace{\langle \mathbf{h}_s, \mathbf{h}_o \rangle}_{\text{Similarity}}, \underbrace{d_{\mathcal{M}}(\mathbf{h}_s, \mathbf{h}_o)}_{\text{Locality Filter}}, \underbrace{\|\mathbf{h}_s\|, \|\mathbf{h}_o\|}_{\text{Hierarchy Depth}}, \underbrace{\cos(\mathbf{h}_s, \mathbf{h}_o)}_{\text{Semantic Compass}} \right] \tag{5}$$

Specifically, the Locality Filter ($d_{\mathcal{M}}$) enforces spatiotemporal locality by ensuring interacting events are geometrically close; Hierarchy Depth ($\|\mathbf{h}\|$) quantifies semantic specificity by mapping radial distance to hierarchical levels (distinguishing broad concepts near the origin from specific actors near the boundary); and the Semantic Compass (cos) determines relation orientation (e.g., ally vs. enemy). These features are subsequently fused via a lightweight MLP to produce geometry-specific scores.

Structural Consensus. A common approach to fusing heterogeneous geometries is using learnable attention weights (e.g., Softmax gating). However, we identify a critical optimization pathology which we term Gate Collapse: since the Hyperbolic branch typically provides larger initial gradients for hierarchical data, adaptive gates often greedily optimize the hyperbolic weight to 1.0, effectively "starving" the Spherical branch of gradients (visualized in Section 4.4).

To mitigate this, we propose structural consensus as a form of strong structural regularization. We argue that in highly heterogeneous manifolds, purely data-driven gating is prone to shortcut learning (i.e., Gate Collapse). Therefore, we treat the fusion weights α, β as rigid structural priors rather than learnable parameters. Aligning with our hierarchical architecture (Figure 2C), the final score is computed via a two-stage convex combination:

$$S_{final} = (1-\beta) \cdot \underbrace{((1-\alpha)S_{hyp} + \alpha S_{sph})}_{\text{Curved Space Consensus}} + \beta \cdot S_{euc} \tag{6}$$

Here, S_{hyp}, S_{sph}, and S_{euc} represent the distinct reasoning scores computed within the Hyperbolic (hierarchical), Spherical (cyclical), and Euclidean (flat) manifolds, respectively. The hyperparameter $\alpha \in [0, 1]$ balances the internal trade-off between hierarchy and periodicity in curved spaces, while $\beta \in [0, 1]$ controls the global regularization from the Euclidean tangent space. By fixing these priors (e.g., $\alpha = 0.5, \beta = 0.2$), we enforce a gradient flow constraint, compelling the encoder to optimize representations across all manifolds simultaneously. This ensures the decision boundary respects all geometries to prevent degeneration, effectively providing a robust inductive bias against mode collapse.

3.5 Optimization and Complexity Analysis

We train MAGIC using binary cross-entropy loss with negative sampling:

$$\begin{aligned} \mathcal{L} = &- \log \sigma(S_{final}(s, o)) \\ &- \sum_{i=1}^{N_{neg}} \log(1 - \sigma(S_{final}(s, o_i'))) \end{aligned} \tag{7}$$

To ensure numerical stability near the manifold boundary, we enforce a strict Riemannian projec-

tion $\|\mathbf{x}\|_2 \leq \frac{1}{\sqrt{|c|}} - \epsilon$ before any logarithmic mapping. Notably, unlike methods requiring expensive Riemannian SGD, our Tangent-Residual Engine allows efficient optimization via standard Adam in the tangent space. Consequently, the time complexity remains $\mathcal{O}(L \cdot |\mathcal{E}| \cdot d)$, which is linear w.r.t the graph size and scalable to large datasets.

4 Experiments

4.1 Experimental Setup

Datasets. We evaluate MAGIC on three standard benchmarks (Lin et al., 2023): ICEWS14, ICEWS05-15, and GDELT.

Baselines. We compare against: (1) *Static* methods (TransE, DistMult, SimplE, RotatE); (2) *Temporal* methods (TeRo, ChronoR, TeLM, EvoExplore, QDN); and (3) *Geometric* models (DyERNIE, BiQCap), focusing on the SOTA multi-curvature model IME, See Appendix C.

Implementation. Implemented in PyTorch, MAGIC uses dimension $d = 200$ for ICEWS05-15 and $d = 100$ otherwise. Models are optimized via Adam (lr = 1e-3), reporting filtered MRR and Hits@$1/3/10$.(Refer to Appendix A for the detailed computational budget, model scale, and dataset statistics.)

4.2 Main Results

Table 1 summarizes the performance comparison. SOTA Performance: MAGIC consistently outperforms all baselines. On the long-range ICEWS05-15, it achieves an MRR of 0.825, significantly surpassing the best baseline IME (0.796). On GDELT, MAGIC achieves a remarkable 0.668 MRR, demonstrating its capability to handle high-density, periodic event streams. Geometric Advantage: Compared to single-space models (e.g., TeLM, DyERNIE), our performance gains validate the necessity of multi-geometry modeling. Furthermore, compared to IME, which uses simple pooling, MAGIC's deep evolution and explicit geometric interaction prove to be superior feature extractors.

Performance Analysis on GDELT. A notable observation in Table 1 is MAGIC's substantial margin on GDELT (+4.4 points over SOTA). We attribute this robustness to three factors: 1) Spherical Contribution: GDELT is dominated by recurrent political events (cycles). While adaptive gates tend to collapse to the Hyperbolic branch (as diagnosed in Sec. 4.4), our Structural Consensus ensures the Spherical branch remains active, effectively capturing these periodic dynamics. 2) Locality Filtering: The geodesic distance in GeoInteract effectively filters out geometrically distant entities in the dense event space. 3) Protocol Verification: We strictly follow the time-aware filtered protocol (Xu et al., 2021), ensuring inverse relations at the same timestamp are excluded to prevent data leakage.

4.3 Ablation Study

We conduct comprehensive ablation studies on ICEWS14 to verify our core hypotheses. Results are summarized in Table 2.

Explicit Geometry vs. Implicit Learning. Replacing our geometry-augmented interaction with a simple dot product causes a substantial drop (MRR $0.842 \rightarrow 0.690$). To pinpoint contributions: removing geodesic distance drops MRR to 0.794 (-4.8%), confirming its role as a spatiotemporal Locality Filter; removing angles leads to 0.833 (-0.9%), validating the need for directional semantics.

Deep Evolution vs. Static Embedding. First, w/o Tangent Residual: Removing the residual connection drops performance to 0.839 and prevents numerical scaling beyond 2 layers. Second, w/o Curvature Manager: Fixing curvature globally causes a significant drop to 0.675. This strongly validates that Geometric Annealing (evolving from flat to curved) is essential for accommodating the exponentially growing receptive field.

Structural Consensus and Initialization. Replacing fixed priors with learnable Softmax gates triggers Gate Collapse (MRR drops to 0.675), where adaptive weights greedily overfit to the hyperbolic view (Figure 5). Unlike this degeneration, Structural Consensus acts as a regularizer to enforce effective multi-geometry utilization. Additionally, removing time-aware initialization (w/o Time Initialization) decreases MRR to 0.795, validating the necessity of decoupling temporal dynamics in the tangent space.

4.4 Mechanism Analysis

Verification of Geometric Annealing. Figure 3 visualizes the training dynamics of curvature scalars $|c|$, where line intensity encodes layer depth. As training progresses, the trajectories exhibit a distinct stratification: while all layers initialize near

Dataset	ICEWS14				ICEWS05-15				GDELT			
Metric	MRR	H@1	H@3	H@10	MRR	H@1	H@3	H@10	MRR	H@1	H@3	H@10
TransE	0.280	0.094	-	0.637	0.294	0.090	-	0.663	0.113	0.000	0.158	0.312
DistMult	0.439	0.323	-	0.672	0.456	0.337	-	0.691	0.196	0.117	0.208	0.348
SimplE	0.458	0.341	0.516	0.687	0.478	0.359	0.539	0.708	0.206	0.124	0.220	0.366
RotatE	0.418	0.291	0.478	0.690	0.304	0.164	0.355	0.595	-	-	-	-
TA-DistMult	0.477	0.363	-	0.686	0.474	0.346	-	0.728	0.206	0.124	0.219	0.365
ATiSE	0.550	0.436	0.629	0.750	0.519	0.378	0.606	0.794	-	-	-	-
TeRo	0.562	0.468	0.621	0.732	0.586	0.469	0.668	0.795	0.245	0.154	0.264	0.420
ChronoR	0.625	0.547	0.669	0.773	0.675	0.596	0.723	0.820	-	-	-	-
TeLM	0.625	0.545	0.673	0.774	0.678	0.599	0.728	0.823	-	-	-	-
TuckERTNT	0.604	0.521	0.655	0.753	0.638	0.559	0.686	0.783	0.381	0.283	0.418	0.576
BoxTE	0.613	0.528	0.664	0.763	0.667	0.582	0.719	0.820	0.352	0.269	0.377	0.511
EvoExplore	0.725	0.653	0.778	0.852	0.790	0.719	0.843	0.915	0.514	0.353	0.602	0.748
BDME	0.635	0.555	0.683	0.778	-	-	-	-	0.278	0.191	0.299	0.448
QDN	0.643	0.567	0.688	0.784	0.692	0.611	0.743	0.838	0.545	0.481	0.576	0.668
DyERNIE	0.669	0.599	0.714	0.797	0.739	0.679	0.773	0.855	0.457	0.390	0.479	0.589
BiQCap	0.643	0.563	0.687	0.798	0.691	0.621	0.738	0.837	0.273	0.183	0.308	0.469
IME	0.819	0.790	0.835	0.872	0.796	0.750	0.821	0.875	0.624	0.485	0.754	0.791
MAGIC (Ours)	**0.842**	**0.790**	**0.881**	**0.928**	**0.825**	**0.762**	**0.874**	**0.931**	**0.668**	**0.533**	**0.761**	**0.918**

Table 1: Link prediction results (MRR and Hits@1/3/10) on three TKG datasets. **Bold** denotes the best results (MAGIC), and underlined denotes the best baseline results. Baseline results are retrieved from (Wang et al., 2024).

Model Variant	**MRR**
MAGIC (Full)	**0.842**
• *Ablation on Interaction*	
w/o GeoInteract (Dot-only)	0.690
w/o Distance (Locality Filter)	0.794
w/o Angle (Semantic Compass)	0.833
• *Ablation on Geometry*	
w/o Curvature Manager (Fixed)	0.675
w/o Multi-Space (Single Euclidean)	0.677
w/o Tangent Residual	0.839
• *Ablation on Consensus*	
w/ Learnable Gate (Softmax)	0.675
w/o Time Initialization	0.795

Table 2: Ablation study on ICEWS14 (MRR).

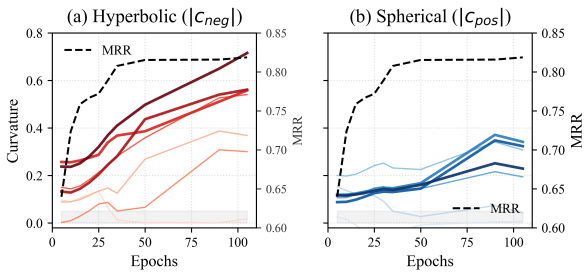

Figure 3: **Geometric Annealing on ICEWS14.** Line intensity indicates depth (Light=Layer 0 → Dark=Layer 7).

zero, the deep layers (dark lines) rapidly diverge to high curvature values, whereas shallow layers (light lines) remain effectively flat ($|c| \approx 0$). This behavior empirically validates the Geometric Annealing phenomenon, indicating that the model spontaneously organizes a topological curriculum. Crucially, it maintains quasi-Euclidean space in early layers for linear semantic alignment, while expanding manifold curvature in deeper layers to accommodate the exponentially growing structural complexity.

Breaking the Depth Barrier. Figure 4 compares the scalability of MAGIC against Single-Hyperbolic and Euclidean baselines across depths $L \in \{1, 2, 4, 8\}$. The Single-Hyperbolic model (Blue) suffers a steep collapse to < 0.50 MRR at $L = 8$, empirically confirming the Riemannian Numerical Instability inherent in deep geometric GCNs. Conversely, the Euclidean baseline (Grey) stagnates below 0.72, illustrating the Euclidean Bottleneck where flat spaces fail to encode complex topology. In sharp contrast, MAGIC (Red) maintains a stable high-performance plateau (> 0.84) regardless of depth. This stability strongly validates that our Tangent-Residual Engine effectively constructs a gradient pathway, successfully decoupling model depth from numerical instability.

Robustness of Structural Consensus. We jointly analyze the failure of adaptive gating and the stability of our proposed consensus. Figure 5

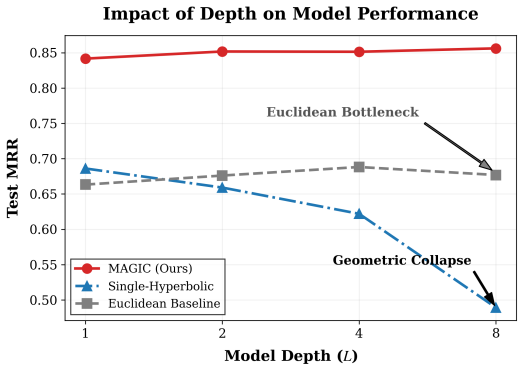

Figure 4: **Impact of Model Depth.** Comparison of MRR across varying layers. Unlike baselines that suffer from collapse (Hyperbolic) or bottlenecks (Euclidean), MAGIC maintains robust performance.

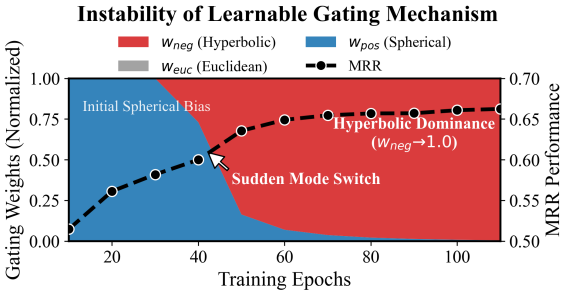

Figure 5: **Gate Collapse in Adaptive Fusion.** Evolution of learnable weights (Area) and MRR (Line). The mechanism fails to maintain diversity, degenerating to a single geometry ($w_{neg} \to 1.0$).

visualizes the instability of learnable baselines, revealing a critical Sudden Mode Switch near epoch 45. Here, the Hyperbolic weight rapidly expands to 1.0, completely suppressing the Spherical and Euclidean signals. This Gate Collapse causes performance to stagnate at a suboptimal plateau (≈ 0.675), confirming that adaptive mechanisms suffer from gradient starvation and degenerate into single-geometry solutions. In contrast, MAGIC enforces a fixed structural consensus to mitigate this issue. Figure 6 addresses potential concerns about brittleness by mapping test MRR across a hyperparameter grid. Instead of a narrow peak, the heatmap reveals a wide "basin of attraction": the model consistently achieves SOTA performance (MRR > 0.83) with minimal fluctuation ($< 1.5\%$). This confirms that our consensus acts as a robust regularizer, effectively leveraging hybrid geometries without requiring delicate manual tuning.

Case Study: Periodic Events. Analysis of cyclic elections confirms that MAGIC mitigates baseline

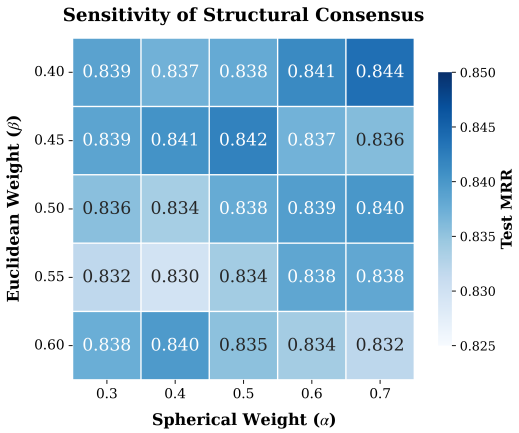

Figure 6: **Sensitivity Analysis.** The wide high-performance plateau (MRR > 0.83) across α and β confirms MAGIC's robustness to hyperparameter variations.

frequency bias and temporal hallucinations. By preserving spherical structure, our model effectively captures long-term periodicity. See Appendix D.

5 Conclusion

We presented MAGIC, a framework that transcends depth and stability bottlenecks in TKG reasoning via a Tangent-Residual Engine and Structural Consensus, enabling stable multi-layer evolution and inducing Geometric Annealing. Experiments confirm that MAGIC establishes a new SOTA by robustly capturing heterogeneous topologies. Future directions include generalizing to continuous-time manifolds via Neural ODEs and deploying MAGIC as a geometric adapter to ground LLMs. By injecting structural constraints via RAG (Lewis et al., 2020), we aim to mitigate hallucinations (Ji et al., 2023) through rigorous spatiotemporal reasoning.

Limitations

Despite establishing a new SOTA in TKG reasoning, MAGIC entails specific design trade-offs. Computational Overhead: While our Tangent-Residual Engine successfully bypasses expensive Riemannian SGD, the mandatory element-wise manifold projections introduce a slight inference latency. We view this as a necessary cost for capturing complex topology, which can be further mitigated via kernel fusion optimization. Fixed Structural Priors: To rigorously prevent the "Gate Collapse" pathology, we employ fixed consensus weights (α, β) rather than fully learnable gating. Although our sensitivity analysis confirms

a wide "basin of attraction," developing a stable, fully adaptive mechanism (e.g., via bi-level optimization) remains a theoretical challenge. Discrete Granularity: Our deep evolution focuses on snapshot-based dynamics. Extending this curvature annealing mechanism to continuous-time manifolds (e.g., integrating with Neural ODEs) represents a promising direction for handling irregular event streams.

Ethics Statement

This paper investigates the problem of temporal knowledge graph reasoning, aiming at capturing complex geometric evolutions to empower a wide range of applications, such as event forecasting and temporal information retrieval. The temporal knowledge graph datasets used in this paper (ICEWS and GDELT) are all publicly available. Therefore, we believe it does not raise any ethical issues.

Acknowledgments

This work was supported in part by the Key Deployment Program of the Chinese Academy of Sciences, China under Grant KGFZD-145-25-39.

References

Antoine Bordes, Nicolas Usunier, Alberto Garcia-Duran, Jason Weston, and Oksana Yakhnenko. 2013. Translating embeddings for modeling multi-relational data. In *Advances in Neural Information Processing Systems (NIPS)*, volume 26.

Ines Chami, Adva Wolf, Da-Cheng Juan, Frederic Sala, Sujith Ravi, and Christopher Ré. 2020. Low-dimensional hyperbolic knowledge graph embeddings. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 6901–6914.

Shib Sankar Dasgupta, Swayambhu Nath Ray, and Partha Talukdar. 2018. HyTE: Hyperplane-based temporally aware knowledge graph embedding. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 2001–2011.

Xiangyu Dong, Xingyi Zhang, Lei Chen, Mingxuan Yuan, and Sibo Wang. 2025. SpaceGNN: Multi-space graph neural network for node anomaly detection with extremely limited labels. In *Proceedings of the International Conference on Learning Representations (ICLR)*.

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

Albert Gu, Frederic Sala, Beliz Gunel, and Christopher Ré. 2019. Learning mixed-curvature representations in product spaces. In *Proceedings of the International Conference on Learning Representations (ICLR)*.

Zhen Han, Peng Chen, Yunpu Ma, and Volker Tresp. 2020. DyERNIE: Dynamic evolution of riemannian manifold embeddings for temporal knowledge graph completion. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 7301–7316.

Ziwei Ji, Nayeon Lee, Rita Frieske, Tiezheng Yu, Dan Su, Yan Xu, Etsuko Ishii, Ye Jin Bang, Andrea Madotto, and Pascale Fung. 2023. Survey of hallucination in natural language generation. *ACM Computing Surveys*, 55(12):1–38.

Seyed Mehran Kazemi and David Poole. 2018. SimplE embedding for link prediction in knowledge graphs. In *Advances in Neural Information Processing Systems (NeurIPS)*.

Timothée Lacroix, Guillaume Obozinski, and Nicolas Usunier. 2020. Tensor decompositions for temporal knowledge base completion. In *Proceedings of the International Conference on Learning Representations (ICLR)*.

Julien Leblay and Melisachew Wudage Chekol. 2018. Deriving validity time in knowledge graph. In *Proceedings of the Companion of the The Web Conference (WWW)*, pages 1771–1776.

Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen-tau Yih, Tim Rocktäschel, Sebastian Riedel, and Douwe Kiela. 2020. Retrieval-augmented generation for knowledge-intensive nlp tasks. In *Proceedings of the 34th International Conference on Neural Information Processing Systems*, NIPS '20, Red Hook, NY, USA. Curran Associates Inc.

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 (ACL)*, pages 15460–15474.

Xueyuan Lin, Haihong E, Chengjin Xu, Gengxian Zhou, Haoran Luo, Tianyi Hu, Fenglong Su, Ningyuan Li, and Mingzhi Sun. 2023. Tflex: Temporal feature-logic embedding framework for complex reasoning over temporal knowledge graph. In *Advances in Neural Information Processing Systems*, volume 36, pages 73039–73081. Curran Associates, Inc.

Chengao Liu, Yuan Li, Siheng Ning, Jianming Zhu, and Jianbin Jiao. 2025. PoinKI: A Poincaré-embedded

knowledge-integrated model for dynamic link prediction. *Knowledge-Based Systems*, page 114879.

Johannes Messner, Ralph Abboud, and Ismail Ilkan Ceylan. 2022. Temporal knowledge graph completion using box embeddings. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 7779–7787.

Maximillian Nickel and Douwe Kiela. 2017. Poincaré embeddings for learning hierarchical representations. In *Advances in Neural Information Processing Systems (NIPS)*, volume 30.

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*, pages 6471–6479.

Zhiqing Sun, Zhi-Hong Deng, Jian-Yun Nie, and Jian Tang. 2019. RotatE: Knowledge graph embedding by relational rotation in complex space. In *Proceedings of the International Conference on Learning Representations (ICLR)*.

Jiapu Wang, Zheng Cui, Boyue Wang, Shirui Pan, Junbin Gao, Baocai Yin, and Wen Gao. 2024. IME: Integrating multi-curvature shared and specific embedding for temporal knowledge graph completion. In *Proceedings of the ACM Web Conference (WWW)*, pages 1954–1962.

Jiapu Wang, Boyue Wang, Junbin Gao, Xiaoyan Li, Yongli Hu, and Baocai Yin. 2023. QDN: A quadruplet distributor network for temporal knowledge graph completion. *IEEE Transactions on Neural Networks and Learning Systems*, 35(10):14018–14030.

Shen Wang, Xiaokai Wei, Cicero Nogueira Nogueira dos Santos, Zhiguo Wang, Ramesh Nallapati, Andrew Arnold, Bing Xiang, Philip S. Yu, and Isabel F. Cruz. 2021. Mixed-curvature multi-relational graph neural network for knowledge graph completion. In *Proceedings of the Web Conference 2021*, WWW '21, page 1761–1771, New York, NY, USA. Association for Computing Machinery.

Weiyao Wang, Du Tran, and Matt Feiszli. 2020. What makes training multi-modal classification networks hard? In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 12695–12705.

Chengjin Xu, Yung-Yu Chen, Mojtaba Nayyeri, 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)*, pages 2569–2578.

Chengjin Xu, Mojtaba Nayyeri, Fouad Alkhoury, Hamed Shakeri Yazdi, and Jens Lehmann. 2020a. Temporal knowledge graph completion based on time series decomposition. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 6569–6579.

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

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

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.

Jiasheng Zhang, Shuang Liang, Yongpan Sheng, and Jie Shao. 2022. Temporal knowledge graph representation learning with local and global evolutions. *Knowledge-Based Systems*, 251:109234. (EvoExplore).

Sensen Zhang, Xun Liang, Zhiying Li, Junlan Feng, Xiangping Zheng, and Bo Wu. 2023. BiQCap: A biquaternion and capsule network-based embedding model for temporal knowledge graph completion. In *International Conference on Database Systems for Advanced Applications (DASFAA)*, pages 673–688.

A Implementation Details and Computational Budget

To facilitate reproducibility, we provide detailed statistics regarding the computational resources, dataset characteristics, and model scale. All experiments were conducted on a single **NVIDIA A100 (80GB)** GPU. The framework is implemented in PyTorch and optimized using the Adam optimizer.

Computational Cost. Table 3 summarizes the dataset topology and the associated training time. Thanks to the efficient *Tangent-Residual Engine*, MAGIC avoids the expensive Riemannian gradients typically associated with geometric GNNs. This efficiency makes training feasible even on the large-scale GDELT dataset (2.3M events) within a reasonable timeframe.

Model Scale. The total number of parameters is dominated by the entity and relation embeddings. Specifically, the model contains approximately **3.2M** parameters for ICEWS14 and **14.5M** parameters for ICEWS05-15. GDELT, despite its high event density, has a smaller entity set, resulting in fewer parameters (<0.5M).

Dataset	Graph Topology			Dataset Splits (Edges)			Cost
	$\lvert\mathcal{E}\rvert$	$\lvert\mathcal{R}\rvert$	$\lvert\mathcal{T}\rvert$	Train	Valid	Test	Time
ICEWS14	6,869	230	365	72,826	8,941	8,963	$\approx$ 4h
ICEWS05-15	10,094	251	4,017	368,962	46,275	46,092	$\approx$ 12h
GDELT	500	20	366	2,735,685	341,961	341,961	$\approx$ 28h

Table 3: Detailed statistics of the datasets and computational cost. $|\mathcal{E}|$, $|\mathcal{R}|$, and $|\mathcal{T}|$ denote the number of entities, relations, and timestamps, respectively. Training time is measured on a single NVIDIA A100 GPU.

B Manifold Operations

We strictly follow the formalism of Gyrovector spaces for operations in the product manifold $\mathcal{M} = \mathbb{H}^d \times \mathbb{S}^d \times \mathbb{E}^d$. The exponential map $\exp_{\mathbf{x}}^c : T_{\mathbf{x}}\mathcal{M} \to \mathcal{M}$ and logarithmic map $\log_{\mathbf{x}}^c : \mathcal{M} \to T_{\mathbf{x}}\mathcal{M}$ at the origin ($\mathbf{0}$) are defined as follows. These operations are applied element-wise within our Tangent-Residual Engine.

Hyperbolic Space ($\mathbb{H}_c^d$). For a component with constant negative curvature $c < 0$ (represented as $-|c|$), the maps are:

$$\exp_{\mathbf{0}}^c(\mathbf{v}) = \tanh\left(\sqrt{|c|}\|\mathbf{v}\|_2\right) \frac{\mathbf{v}}{\sqrt{|c|}\|\mathbf{v}\|_2} \quad (8)$$

$$\log_{\mathbf{0}}^c(\mathbf{y}) = \tanh^{-1}\left(\sqrt{|c|}\|\mathbf{y}\|_2\right) \frac{\mathbf{y}}{\sqrt{|c|}\|\mathbf{y}\|_2} \quad (9)$$

where $\|\cdot\|_2$ denotes the Euclidean L_2 norm.

Spherical Space ($\mathbb{S}_c^d$). For a component with constant positive curvature $c > 0$, the maps are defined as:

$$\exp_{\mathbf{0}}^c(\mathbf{v}) = \tan\left(\sqrt{c}\|\mathbf{v}\|_2\right) \frac{\mathbf{v}}{\sqrt{c}\|\mathbf{v}\|_2} \quad (10)$$

$$\log_{\mathbf{0}}^c(\mathbf{y}) = \tan^{-1}\left(\sqrt{c}\|\mathbf{y}\|_2\right) \frac{\mathbf{y}}{\sqrt{c}\|\mathbf{y}\|_2} \quad (11)$$

Numerical Stability. To prevent gradient explosion near the boundaries of the Poincaré ball or singularities in the Spherical space, we apply norm clipping. Specifically, for an input vector $\mathbf{x}$, we restrict $\|\mathbf{x}\|_2 \leq \frac{1}{\sqrt{|c|}} - \epsilon$ (with $\epsilon = 10^{-5}$) before applying the logarithmic map.

C Baseline Descriptions

We provide brief descriptions of the baselines reported in Table 1.

Static KGE Methods.

- **TransE** (Bordes et al., 2013), **DistMult** (Yang et al., 2015), **RotatE** (Sun et al., 2019): Fundamental models using translation, bilinear diagonal interaction, and complex space rotation, respectively.
- **SimplE** (Kazemi and Poole, 2018): A tensor factorization approach based on CP decomposition that allows independent embeddings for head and tail entities.

Temporal KGE Methods.

- **TA-DistMult** (García-Durán et al., 2018): Extends DistMult by incorporating temporal embeddings (often via RNNs) to capture time-sensitive relation semantics.
- **ATiSE** (Xu et al., 2020a): Models entity evolution using additive time series decomposition (trend and seasonality) within a Gaussian embedding framework.
- **TeRo** (Xu et al., 2020b), **ChronoR** (Sadeghian et al., 2021), **TeLM** (Xu et al., 2021): Methods modeling temporal evolution via complex rotation or high-order multivector tensor decomposition.
- **BoxTE** (Messner et al., 2022) & **QDN** (Wang et al., 2023): Approaches utilizing box embeddings or query-dependent variations to explicitly handle temporal uncertainty.
- **EvoExplore** (Zhang et al., 2022): A recent structural learning method that explores the interactions between static and dynamic graph structures.

Geometric & Multimodal Methods.

- **DyERNIE** (Han et al., 2020): Evolves entity embeddings using velocity vectors in Euclidean space.
- **BiQCap** (Zhang et al., 2023): Combines Hyperbolic space with quaternary representations to capture hierarchies.
- **IME** (Wang et al., 2024): The primary baseline, mapping data into product manifolds ($\mathbb{H} \times \mathbb{S} \times \mathbb{E}$) using simple pooling.

D Detailed Case Study

To qualitatively validate MAGIC's reasoning capability, we analyze a specific prediction case from the ICEWS dataset involving a recurrent political event.

Case Description.

- **Query:** *(South Korea, Legislative Election, ?, 2012-04)*.
- **Ground Truth:** *South Korea* (Targeting the domestic election event).
- **Context:** Legislative elections in South Korea typically occur every four years (e.g., 2004, 2008, 2012). Capturing this pattern requires reasoning over **long-range historical dependencies** spanning nearly a decade.

Model Comparison.

- **Baseline Failure (IME/TeLM):** The baseline models incorrectly predict high-frequency entities such as *United States* or *Japan*. We attribute this error to **Gate Collapse**. Despite the existence of cyclic signals, the adaptive gating mechanism assigns near-zero weight to the Spherical component during training. Consequently, the model fails to capture the "4-year cycle" and degenerates into predicting "hub" nodes (entities with high global degree), exhibiting a form of **temporal hallucination**.

- **MAGIC Success:** In contrast, MAGIC successfully predicts the correct tail entity. This success is driven by three factors:

  (1) **Deep Evolution ($L = 8$):** Our stable deep architecture allows the model to effectively propagate information from distant historical snapshots (2004, 2008) to the current query time (2012), overcoming the standard depth bottleneck ($L \leq 2$).

  (2) **Geometric Annealing:** The model correctly aligns the static semantics of "South Korea" in shallow layers (flat space) while evolving to capture the complex cyclic topology in deep layers (curved space).

  (3) **Structural Consensus:** By enforcing fixed priors (e.g., $\alpha_{pos} > 0.3$), MAGIC explicitly prevents the Spherical branch from being suppressed by high-frequency noise. This ensures the discovered periodicity is utilized for the final decision.

This case confirms that MAGIC's geometric constraints effectively ground the model against generic semantic dominance, enabling robust reasoning over long-range periodic dynamics.