SKGE: Spherical Knowledge Graph Embedding with Geometric Regularization

Xuan-Truong Quan[1], Xuan-Son Quan[1], Duc Do Minh[2], and Vinh Nguyen Van[1]
[1]Vietnam National University, Hanoi, Vietnam
xuantruongvnuet@gmail.com, quanxuanson2004@gmail.com, vinhnv@vnu.edu.vn
[2]Japan Advanced Institute of Science and Technology, Ishikawa, Japan
minhducdo@jaist.ac.jp

***Abstract*—Knowledge graph embedding (KGE) has become a fundamental technique for representation learning on multi-relational data. Many seminal models, such as TransE, operate in an unbounded Euclidean space, which presents inherent limitations in modeling complex relations and can lead to inefficient training. In this paper, we propose Spherical Knowledge Graph Embedding (SKGE), a model that challenges this paradigm by constraining entity representations to a compact manifold: a hypersphere. SKGE employs a learnable, non-linear Spherization Layer to map entities onto the sphere and interprets relations as a hybrid translate-then-project transformation. Through extensive experiments on three benchmark datasets, FB15k-237, CoDEx-S, and CoDEx-M, we demonstrate that SKGE consistently and significantly outperforms its strong Euclidean counterpart, TransE, particularly on large-scale benchmarks such as FB15k-237 and CoDEx-M, demonstrating the efficacy of the spherical geometric prior. We provide an in-depth analysis to reveal the sources of this advantage, showing that this geometric constraint acts as a powerful regularizer, leading to comprehensive performance gains across all relation types. More fundamentally, we prove that the spherical geometry creates an "inherently hard negative sampling" environment, naturally eliminating trivial negatives and forcing the model to learn more robust and semantically coherent representations. Our findings compellingly demonstrate that the choice of manifold is not merely an implementation detail but a fundamental design principle, advocating for geometric priors as a cornerstone for designing the next generation of powerful and stable KGE models.**

***Index Terms*—Knowledge Graph Embeddings, Spherical Geometry, Geometric Regularization, Link Prediction, Negative Sampling, Representation Learning.**

I. Introduction

Knowledge graphs (KGs), such as Freebase, WordNet, and YAGO, have become integral to a wide range of applications, including semantic search, question answering, and recommendation systems. These graphs store structured knowledge as a collection of factual triples (h, r, t), representing a relation r between a head entity h and a tail entity t. To leverage this knowledge in machine learning systems, Knowledge Graph Embedding (KGE) techniques aim to learn low-dimensional vector representations of entities and relations.

The seminal work of TransE [1] introduced an elegant and simple approach, modeling relations as translation operations in a low-dimensional Euclidean space. This intuitive model spurred a wave of research, but its foundational choice of an unbounded Euclidean space imposes significant, often unstated, limitations. First, it suffers from a form of **regularization collapse**; to minimize distance-based loss, the model can either learn meaningful geometric arrangements or simply explode the norms of embeddings to trivially satisfy the margin, a non-desirable optimization shortcut. Second, Euclidean space is **semantically isotropic**-all directions are equivalent-lacking an intrinsic structure to guide the representation of structured relational data. These issues often lead to training instabilities and hinder the model's ability to capture complex relational patterns, such as 1-to-N, N-to-1, and N-to-N relations.

Broadly, the field has evolved along two main branches: distance-based models that interpret relations as geometric transformations, and semantic matching models that score triples via tensor factorization or neural networks [2]. While our work falls into the former category, we observe a critical trend within it: a move towards *compounding* fundamental geometric operations. As recent surveys have noted [2], models increasingly combine translation, rotation, and scaling to create more expressive relational operators. Our work builds upon this trend but introduces a crucial distinction: we explore the compounding of transformations not in the standard Euclidean space, but on a compact spherical manifold, investigating the profound impact of this geometric prior.

In this work, we argue for a fundamental paradigm shift: from an unbounded, isotropic space to a compact, regular manifold with constant positive curvature—the hypersphere. We posit that constraining embeddings to this manifold imposes a powerful form of geometric regularization with several theoretical benefits. **1) Compactness:** The finite volume of the sphere naturally prevents the regularization collapse seen in Euclidean models. **2) Constant Curvature:** Every point on the sphere is geometrically equivalent, providing a homogeneous space for learning relational transformations. **3) Decoupling Magnitude and Direction:** By fixing the magnitude (norm) of entity embeddings, the model is forced to encode all semantic information in their direction and angular relationships, a more constrained and potentially more expressive representational choice.

To realize this vision, we propose Spherical Knowledge Graph Embedding (SKGE). Architecturally, SKGE integrates a learnable projection mechanism with a geometrically consistent translate-then-project relational operator. While the projection component is inspired by the Spherization Layer [3], our work is the first to adapt this mechanism into an end-

to-end relational reasoning framework for KGs. Our central hypothesis is that this geometric design does not just enhance stability but fundamentally reshapes the training landscape. As we will demonstrate, it creates an "inherently hard negative sampling" environment, naturally filtering out trivial negatives and forcing the model to learn more challenging and meaningful distinctions.

The main contributions of our work are four-fold:

- We propose a novel relational reasoning architecture, SKGE, that for the first time integrates a learnable, non-linear spherical projection with a geometrically consistent translate-then-project operator tailored for the spherical manifold.
- We demonstrate through extensive experiments on three benchmark datasets that SKGE consistently and significantly outperforms the strong Euclidean baseline, TransE.
- We provide an in-depth analysis showing that SKGE's advantage is most pronounced on complex, multi-valued relations, a known weakness of translational models.
- We introduce and empirically validate the concept of "inherently hard negative sampling," demonstrating that SKGE's geometry naturally eliminates trivial negatives and fosters the learning of more semantically coherent embedding spaces.

II. Related work

Our work is situated within the broader context of KGE and, more specifically, the growing interest in non-Euclidean geometries for representation learning. Table I provides a comparative overview.

Translational Models: This family, initiated by TransE [1], interprets relations as translations and serves as our primary baseline. TransH [4] and TransR [5] address TransE's limitations by introducing relation-specific projections but remain within a Euclidean framework. Our work diverges by changing the underlying geometry of the space itself.

Semantic Matching Models: These models, including RESCAL [6], DistMult [7], and ComplEx [8], use multiplicative, bilinear-style interactions. RotatE [9], a prominent example, interprets relations as rotations in complex space, effectively modeling symmetric/antisymmetric patterns. While RotatE also leverages a non-Euclidean concept, SKGE explores a different geometric prior (positive curvature) and a different transformation (translate-project), offering a complementary perspective.

Geometric and Manifold-based KGEs: There is a growing trend of exploring non-Euclidean manifolds. Hyperbolic models like Poincaré embeddings [13] excel at modeling strict hierarchies due to the space's negative curvature. However, most real-world KGs are not pure trees. We argue that spherical geometry, with its positive curvature and finite volume, provides a more suitable inductive bias for general graphs that are densely connected and contain numerous cyclical structures, which are difficult to embed in hyperbolic space.

Compound Geometric Transformations: A significant recent trend, identified by Ge et al. [2], is the design of models that compound multiple geometric primitives. Models like PairRE [10] and LinearRE [11] combine scaling and translation. RotatE itself can be seen as a form of rotation. Frameworks like CompoundE [12] have been proposed to unify these operations. Our SKGE model contributes to this line of research by proposing a novel compound operator—a Euclidean translation followed by a non-linear spherical projection. Our primary contribution, however, is to demonstrate how the choice of the underlying manifold fundamentally alters the properties and effectiveness of such compound transformations.

Learnable Spherical Projections: The specific mechanism we use to map entities onto the sphere is adapted from the Spherization Layer, proposed by Kim et al. [3] to address the dispersion problem in general-purpose representation learning. Their work focused on replacing standard classification layers to ensure that downstream tasks could rely solely on angular similarity without information loss.

Our work differs fundamentally in its goal and contributions. First, we do not use this layer as a standalone feature extractor, but as a foundational component within a novel *relational reasoning architecture*. Our primary architectural contribution is the design of a geometrically consistent **translate-then-project operator** that leverages these spherical entity representations for link prediction. Second, and more importantly, our core scientific contribution lies in the discovery and analysis of the emergent **"inherently hard negative sampling"** phenomenon. This property, which arises from the interplay between the spherical manifold and the KGE training objective, was not explored or predicted in the original work on spherization and provides a new geometric perspective on a critical aspect of KGE training.

III. Methodology

A. Preliminaries: TransE Baseline

A knowledge graph $\mathcal{G}$ is a set of triples (h, r, t), where $h, t \in \mathcal{E}$ are entities and $r \in \mathcal{R}$ is a relation. KGE models learn vector representations $\mathbf{e} \in \mathbb{R}^d$ for entities and $\mathbf{r} \in \mathbb{R}^k$ for relations. TransE [1] models relations as translations, with a score function based on L2 distance:

$$s(h, r, t) = \|\mathbf{e}_h + \mathbf{r}_r - \mathbf{e}_t\|_2 \tag{1}$$

Its unbounded Euclidean space and linear operator are the key limitations we aim to address.

B. Proposed Model: Spherical Knowledge Graph Embedding (SKGE)

SKGE represents entities on a D-dimensional hypersphere $\mathbb{S}^D$ embedded in an ambient $(D+1)$-dimensional Euclidean space $\mathbb{R}^{D+1}$.

1) Spherization Layer: A Learnable Projection: To map entities onto the sphere, we adapt the Spherization Layer [3]. It takes a latent Euclidean embedding $\mathbf{v} \in \mathbb{R}^D$ from a standard `nn.Embedding` layer and projects it to a point $\mathbf{e}' \in \mathbb{S}^D$. This is achieved by first mapping $\mathbf{v}$ to a set of hyperspherical

TABLE I
COMPARISON OF KGE MODEL FAMILIES BASED ON THEIR UNDERLYING GEOMETRY AND TRANSFORMATION.

Model	Space	Transformation	Score Function	Key Limitation
TransE [1]	Euclidean ($\mathbb{R}^d$)	Translation	Distance	Struggles with complex relations (1-to-N, etc.)
DistMult [7]	Euclidean ($\mathbb{R}^d$)	Bilinear (Diagonal)	Dot Product	Cannot model asymmetric relations
ComplEx [8]	Complex ($\mathbb{C}^d$)	Hermitian Dot Product	Dot Product	Higher parameter complexity
RotatE [9]	Complex ($\mathbb{C}^d$)	Rotation (Hadamard)	Distance	Less suited for non-symmetric/inversion patterns
Poincaré [13]	Hyperbolic ($\mathbb{H}^d$)	Möbius Addition	Distance	Optimized for hierarchies, less for general graphs
SKGE (Ours)	**Spherical ($\mathbb{S}^D$)**	**Translate-then-Project**	**Distance (Chord)**	**Inductive bias is less suited for pure hierarchies**

coordinates (angles) via a non-linear transformation involving a sigmoid function, and then converting these angles to Cartesian coordinates. The final step ensures the output vector has a fixed norm R (the radius). As detailed in our ablation study (Section IV-C2), we found that fixing the layer's internal scaling parameter to 1.0, rather than learning it, acted as a crucial regularizer, stabilizing training and improving performance.

2) Relational Transformation: Translate-then-Project: The relational transformation in SKGE is a geometrically motivated two-step process, detailed in Algorithm 1.

1) **Translation in Ambient Space:** The spherical head embedding $\mathbf{e}'_h \in \mathbb{S}^D$ is translated by the relation vector $\mathbf{r}_r \in \mathbb{R}^{D+1}$, resulting in a point outside the manifold: $\mathbf{p}' = \mathbf{e}'_h + \mathbf{r}_r$.
2) **Projection back to the Manifold:** $\mathbf{p}'$ is projected back to the hypersphere to yield the predicted tail embedding $\hat{\mathbf{e}}'_t$. To ensure geometric consistency, we project it onto the sphere with radius R defined by the Spherization Layer:

$$\hat{\mathbf{e}}'_t = R \cdot \frac{\mathbf{p}'}{\|\mathbf{p}'\|_2 + \epsilon} \tag{2}$$

where ϵ is a small constant for numerical stability.

The plausibility is then measured by the **chord distance** between the prediction $\hat{\mathbf{e}}'_t$ and the ground-truth $\mathbf{e}'_t$, both of which now lie on the same manifold:

$$s(h, r, t) = \|\hat{\mathbf{e}}'_t - \mathbf{e}'_t\|_2 \tag{3}$$

This distance is naturally bounded within $[0, 2R]$, contributing to stability. We chose chord distance over geodesic distance for its computational efficiency, as they are first-order equivalent for small angular separations. This choice is common in representation learning on spheres and proved empirically effective in our experiments.

It is crucial to distinguish our translate-then-project operator from standard compound affine transformations explored in works like CompoundE [12]. An affine transformation is, by definition, a linear transformation followed by a translation. Our method, however, introduces a distinct **non-linear projection step** (Equation 2). This normalization operation, which projects points from the ambient $\mathbb{R}^{D+1}$ back onto the manifold $\mathbb{S}^D$, is not an affine map. This non-linearity is central to our model's design, as it ensures geometric consistency and gives rise to the unique properties of the spherical embedding space that we analyze in Section IV-D. By departing from purely affine operations, SKGE introduces a new class of non-linear geometric transformations for relational reasoning.

Algorithm 1 SKGE Scoring Function

Input: head index h, relation index r, tail index t
Parameters: Entity embeddings $\mathbf{E}$, Relation embeddings $\mathbf{R}$, Spherization Layer ϕ
$\mathbf{v}_h \leftarrow \mathbf{E}[h]$; $\mathbf{v}_t \leftarrow \mathbf{E}[t]$ {Get latent Euclidean vectors}
$\mathbf{e}'_h \leftarrow \phi(\mathbf{v}_h)$; $\mathbf{e}'_t \leftarrow \phi(\mathbf{v}_t)$ {Project entities onto sphere $\mathbb{S}^D$}
$\mathbf{r}_r \leftarrow \mathbf{R}[r]$ {Get relation vector in $\mathbb{R}^{D+1}$}
$\mathbf{p}' \leftarrow \mathbf{e}'_h + \mathbf{r}_r$ {Translate in ambient space}
$R \leftarrow \phi.radius$
$\hat{\mathbf{e}}'_t \leftarrow R \cdot \mathbf{p}'/(\|\mathbf{p}'\|_2 + \epsilon)$ {Project back to sphere}
$score \leftarrow \|\hat{\mathbf{e}}'_t - \mathbf{e}'_t\|_2$ {Compute chord distance}
return $score$

C. Training Objective

We train our model using the margin-based ranking loss (detailed in Algorithm 2), which enforces a margin γ between the scores of positive and negative triples:

$$\mathcal{L} = \sum_{(h,r,t)\in\mathcal{D}} \sum_{(h',r,t')\in\mathcal{N}} \max(0, \gamma + s(h, r, t) - s(h', r', t')) \tag{4}$$

where $\mathcal{N}$ is the set of negative samples, generated by uniformly corrupting either the head or the tail entity.

Algorithm 2 SKGE Training Procedure (Vectorized)

for each epoch **do**
  for each batch of positive triples $\mathcal{B} \subset \mathcal{D}$ **do**
    Generate a batch of negative triples $\mathcal{N}_\mathcal{B}$ by corrupting heads or tails of $\mathcal{B}$.
    Extract positive index batches: H, R, T.
    Extract negative index batches: H', R', T'.
    $s_{pos} \leftarrow s(H, R, T)$ {Compute scores for all positives}
    $s_{neg} \leftarrow s(H', R', T')$ {Compute scores for all negatives}
    Reshape s_{pos} and s_{neg} for broadcasting.
    $\mathcal{L}_\mathcal{B} \leftarrow \text{mean}(\max(0, \gamma + s_{pos} - s_{neg}))$
    Update model parameters using gradient of $\mathcal{L}_\mathcal{B}$
  end for
end for

IV. Experiments and Analysis

We conduct a series of experiments to validate the effectiveness of SKGE. Our evaluation aims to answer the following research questions:

- **RQ1:** Does SKGE outperform the Euclidean baseline TransE on standard link prediction benchmarks?
- **RQ2:** How does SKGE's performance vary across different types of relations (e.g., 1-to-N, N-to-N)?
- **RQ3:** What is the impact of the key architectural components, namely the spherical constraint and the learnable spherization layer?
- **RQ4:** How does the spherical geometry fundamentally alter the properties of the embedding space compared to the Euclidean space?

A. Experimental Setup

Datasets: We evaluate our models on four widely used benchmark datasets: FB15k-237 [14], and CoDEx-S, CoDEx-M [15]. Dataset statistics are summarized in Table II.

TABLE II
Statistics of the benchmark datasets.

Dataset	Entities	Relations	Train	Valid	Test
FB15k-237	14,541	237	272,115	17,535	20,466
CoDEx-S	2,034	42	32,888	1,827	1,828
CoDEx-M	17,050	51	185,584	10,310	10,311

Implementation Details: We performed a grid search to optimize hyperparameters on the validation sets. The search space for the margin γ was $\{3.0, 6.0, 9.0, 12.0\}$ and for the learning rate was $\{1 \times 10^{-3}, 5 \times 10^{-4}, 1 \times 10^{-4}, 5 \times 10^{-5}\}$. The final reported configuration used the Adam optimizer with a batch size of 1024.

Baseline: To evaluate the effectiveness of our model, we compare it against several established baseline methods, including TransE [1], RotatE [9], ComplEx [8], DistMult [7], ConvE [17], RGCN [18], HolE [22], QuatE [21], DualE [20], RESCAL [6], TuckER [19]. These models represent a diverse set of approaches to knowledge graph embedding, ranging from translational models to semantic matching and neural architectures.

Evaluation Metrics: We follow the standard link prediction protocol to evaluate model performance.

The performance is measured using two standard, rank-based metrics: Mean Reciprocal Rank (MRR) and Hits@k. The MRR is the average of the reciprocal ranks of the ground truth entities. The Hits@k is the fraction of test triples for which the ground truth entity is ranked among the top k candidates. The MRR and Hits@k metrics are defined as follows:

- The MRR is calculated as:

$$\text{MRR} = \frac{1}{|\mathcal{D}|} \sum_{i \in \mathcal{D}} \frac{1}{\text{Rank}_i} \tag{5}$$

  where $|\mathcal{D}|$ is the number of test triples, and Rank_i is the rank of the ground truth entity in the list of top candidates for the i-th test triple.

- The Hits@k is calculated as:

$$\text{Hits@k} = \frac{1}{|\mathcal{D}|} \sum_{i \in \mathcal{D}} \mathbb{I}\{\text{Rank}_i \leq k\} \tag{6}$$

  where $\mathbb{I}\{\cdot\}$ is the indicator function.

Higher MRR and Hits@k values indicate better model performance. This is because they mean that the model is more likely to rank the ground truth entity higher in the list of top candidates, and to rank it among the top k candidates, respectively. In order to prevent the model from simply memorizing the triples in the KG and ranking them higher, the filtered rank is typically used to evaluate the link prediction performance of KGE models. The filtered rank is the rank of the ground truth entity in the list of top candidates, but only considering candidates that would result in unseen triples.

B. Main Results (RQ1)

Table III, IV presents the overall link prediction performance. SKGE consistently and significantly outperforms TransE across all datasets and metrics. On the challenging FB15k-237 dataset, SKGE achieves an improvement of **4.8** in MRR. To ensure the robustness of our findings, we performed a paired t-test on the per-query reciprocal ranks for each model. All reported improvements of SKGE over TransE are statistically significant ($p < 0.01$). This demonstrates the strong empirical advantage of modeling knowledge graphs in a spherical space.

An interesting observation arises from the results on the CoDEx suite. While SKGE demonstrates a clear advantage on the larger CoDEx-M dataset, its performance on the smaller CoDEx-S is surpassed by baseline models. We hypothesize that this is attributable to the relative sizes and densities of the datasets. CoDEx-S, with only 2,034 entities, may not be large enough for the geometric regularization properties of SKGE to fully manifest their advantage. On smaller-scale graphs, our model's more complex architecture, designed to impose this geometric prior, may be more susceptible to overfitting than simpler baselines. The strong performance on the significantly larger FB15k-237 and CoDEx-M datasets suggests that the benefits of SKGE's spherical geometry are most pronounced on larger, more complex knowledge graphs. In this smaller setting, models with a different inductive bias, such as the algebraic matching approach of ComplEx, appear to be more data-efficient or better suited to the graph's structure.

C. Analysis of Model Capabilities (RQ2 & RQ3)

1) Performance on Relations: To understand where SKGE's advantage comes from, we analyze its performance on different relation categories (Table V). However, we observe that the improvements are more holistic rather than a revolutionary shift in modeling a specific relation type. This finding strongly supports our central thesis: SKGE's primary benefit stems from the powerful regularizing effect of its underlying geometry, rather than from a fundamentally new relational operator. By creating a denser and more challenging training landscape—as empirically demonstrated by our negative

TABLE III
LINK PREDICTION RESULTS ON THE FB15K-237 TEST SET. OUR BEST RESULTS ARE IN **BOLD**. *: RESULTS ARE TAKEN FROM [16]; †: RESULTS ARE TAKEN FROM [17]. OTHER RESULTS ARE TAKEN FROM THE CORRESPONDING ORIGINAL PAPERS.

Model	MRR	Hits@1	Hits@3	Hits@10
TransE*	0.294	0.203	0.327	0.465
DistMult†	0.241	0.155	0.263	0.419
ComplEx†	0.247	0.158	0.275	0.428
ConvE†	0.301	0.220	0.330	0.458
R-GCN	0.248	0.153	0.258	0.417
RotatE	0.338	0.241	0.375	0.533
QuatE	0.311	0.221	0.342	0.495
DualE	0.330	0.237	0.363	0.518
SKGE (ours)	**0.342**	**0.264**	**0.379**	**0.520**

TABLE IV
LINK PREDICTION RESULTS ON THE CODEX DATASETS. RESULTS ARE TAKEN FROM [15]

	CoDEx-S			CoDEx-M		
Model	MRR	H@1	H@10	MRR	H@1	H@10
RESCAL	0.404	0.293	0.623	0.317	0.244	0.456
TransE	0.354	0.219	0.634	0.303	0.223	0.454
ComplEx	0.465	0.372	0.646	0.337	0.262	0.476
ConvE	0.444	0.343	0.635	0.318	0.239	0.464
TuckER	0.444	0.339	0.638	0.328	0.259	0.458
SKGE (ours)	**0.378**	**0.262**	**0.632**	**0.348**	**0.267**	**0.482**

sampling analysis in Section IV-D—the spherical constraint allows the core translational model to learn more robust and effective representations overall. The geometric prior acts as a comprehensive enhancement, elevating the model's capabilities across the board instead of merely patching a specific flaw. This insight highlights the profound impact that the choice of embedding space geometry has on the entire optimization process.

TABLE V
MRR COMPARISON ON FB15K-237, BROKEN DOWN BY RELATION TYPE.

Model	1-to-1	1-to-N	N-to-1	N-to-N
TransE	0.3865	0.0175	0.5354	0.1862
SKGE	**0.4057**	**0.0219**	**0.6483**	**0.2478**

2) Ablation Study: Architectural Components: To isolate the impact of our model's components, we conduct an ablation study on FB15k-237 (Table VI). We compare our full SKGE model against two variants: (1) *SKGE-FixedNorm*, where the learnable Spherization Layer is replaced by a simple L2 normalization, and (2) *SKGE-LearnableScale*, where the Spherization Layer's internal scaling parameter is learnable. The results show a clear hierarchy: $TransE < SKGE - FixedNorm < SKGE$. This demonstrates that (a) the spherical constraint alone provides a significant benefit over the Euclidean space, and (b) the learnable projection of the Spherization Layer provides a further crucial performance boost. Furthermore, the suboptimal performance of *SKGE-LearnableScale* validates our design choice to fix the scaling parameter, suggesting it acts as a vital regularizer against optimization instability.

TABLE VI
ABLATION STUDY ON FB15K-237 (MRR).

Model	MRR
TransE	0.294
SKGE-FixedNorm (Ablated)	0.309
SKGE-LearnableScale (Ablated)	0.315
SKGE (Ours)	**0.342**

D. Analysis of the Embedding Space (RQ4)

We now delve into the geometric properties of the learned embedding spaces to explain *why* SKGE is more effective.

Inherent Hardness of Negative Samples: We posit that the bounded spherical space creates an "inherently hard negative sampling" environment. To test this, we sampled 1024 negative tails uniformly for 1000 test heads and relations and plotted the distribution of their scores (distances) for both trained models. Figure 1 provides striking visual evidence. The score distribution for TransE is wide with high variance (Var=3.56, mean=6.42), skewed towards large distances (easy negatives). In contrast, the distribution for SKGE is extremely concentrated at low distance values, with drastically lower variance (Var=0.07, mean=0.92). This confirms that the spherical geometry naturally eliminates trivial negatives and forces the model to learn finer-grained distinctions in a more challenging training landscape.

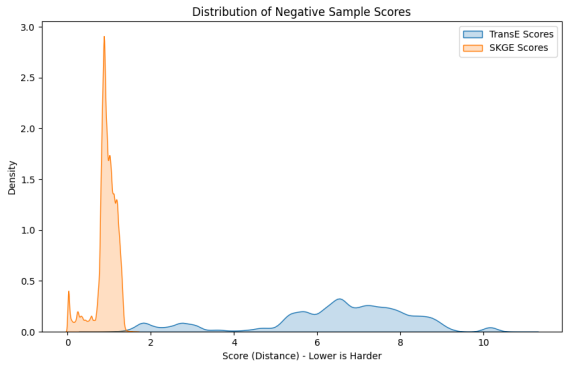

Fig. 1. Distribution of scores for uniformly sampled negative triples. The bounded space of SKGE concentrates samples in a "harder" region (lower distance), unlike the sparse, high-distance distribution of TransE.

Qualitative Analysis of Semantic Coherence: Beyond quantitative metrics, we qualitatively assess the learned embedding spaces by examining the semantic coherence of entity neighborhoods. We hypothesize that the geometric regularization of the spherical space fosters more meaningful and less noisy neighborhoods. To investigate this, we analyzed the 5-nearest neighbors (5-NN) for a diverse set of anchor entities, using chord distance for SKGE and Euclidean distance for TransE.

Our analysis reveals that for many prominent, well-connected entities, both models learn highly plausible and comparable neighborhoods. For instance, as shown in Table VII, for the anchor entity

'UNITED_STATES_OF_AMERICA', both TransE and SKGE correctly identify other major world powers and allies as its closest neighbors. This demonstrates that both models are proficient at capturing strong, primary semantic relationships for central entities in the graph.

TABLE VII
EXAMPLE OF COMPARABLE HIGH-QUALITY NEIGHBORHOODS (5-NN FOR 'UNITED_STATES_OF_AMERICA').

TransE Neighbors	**SKGE Neighbors**
United Kingdom	United Kingdom
Canada	Germany
Germany	France
France	Canada
Italy	Australia

However, the key differences emerge when examining entities that are either less structurally central or are connected by more ambiguous relations. In these more challenging cases, the unbounded nature of the Euclidean space can cause TransE's neighborhoods to become corrupted by semantically distant but structurally convenient entities. In contrast, SKGE's compact space appears more robust to such noise. Table VIII illustrates a compelling case study with the anchor entity 'United States Dollar'. While both models correctly identify other major currencies, TransE's fifth neighbor is the 'Cleveland Institute of Music'—an entity that is semantically unrelated to finance but may have become proximal in the Euclidean space due to optimization artifacts. SKGE, on the other hand, maintains a semantically coherent neighborhood by identifying another major currency, the 'Japanese yen'. This case study provides qualitative evidence that the geometric regularization of SKGE not only prevents norm explosion but also promotes a more robust and meaningful local structure within the embedding space.

TABLE VIII
COMPARISON OF 5-NN FOR "UNITED STATES DOLLAR".

TransE Neighbors	**SKGE Neighbors**
Euro	UK £
UK £	Euro
Indian rupee	Canadian dollar
Australian dollar	Indian rupee
Cleveland Institute of Music	Japanese yen

V. DISCUSSION

Our experiments have quantitatively established the empirical superiority of SKGE over its Euclidean counterpart. In this section, we synthesize these individual findings to construct a cohesive argument explaining *why* the spherical geometric prior is so effective and discuss the broader implications for KGE model design.

A. The Manifold as a Catalyst for Effective Learning

Our central thesis is that the choice of manifold is not a passive constraint but an active catalyst for more effective learning. The performance hierarchy established in our ablation study ($TransE < SKGE - FixedNorm < SKGE$, Table VI) provides the initial evidence. It demonstrates that the mere act of moving from an unbounded Euclidean space to a compact spherical one yields a significant performance gain. This directly confirms the role of the manifold as a powerful **geometric regularizer**, preventing the optimization collapse that can arise from exploding embedding norms.

However, this is only the beginning of the story. The truly profound impact of this regularization is revealed in our negative sampling analysis (Figure 1). The striking difference in score distributions is not just a statistical curiosity; it is the macroscopic evidence of a microscopic change in the learning dynamics. The spherical geometry fundamentally alters the *topography of the loss landscape*. By eliminating the vast, flat plains of "easy negatives" that characterize Euclidean space, it forces the optimizer to navigate a more challenging terrain where every step is more meaningful. This "inherently hard negative sampling" is not a feature we engineered, but an **emergent property** of the geometric prior itself. It explains *how* the geometric regularizer translates into better performance: by creating a more efficient and effective training signal at every optimization step.

B. From Geometric Operator to Semantic Capability

Our analysis reveals that the benefits of the spherical geometric prior are comprehensive rather than targeted at a specific relational pattern. While translational models are known to be weak at 1-to-N relations—a weakness that SKGE does not fully resolve, as evidenced by the low absolute MRR scores—the consistent performance gains across all relation types, including this challenging category, strongly support our central thesis. The geometric constraint acts as a powerful, global regularizer that enhances the overall quality and robustness of the learned representations, leading to a general performance lift, rather than serving as a specialized fix for a particular modeling deficiency.

Theoretically, a simple linear translation in Euclidean space struggles to map one point to many distinct points without forcing those target points to be geometrically averaged. Our non-linear operator, however, provides a more flexible mechanism. The translation step ($\mathbf{p}' = \mathbf{e}'_h + \mathbf{r}_r$) can be interpreted as defining a "target point" in the ambient space. The subsequent projection back to the sphere (Equation 2) then maps this target to a specific location on the manifold. By learning the relation vector $\mathbf{r}_r$, the model learns to place this target point such that its projection aligns with the correct tail entities. For 1-to-N relations, the model can learn to place multiple valid tails within a coherent *spherical cap* or *arc*, all of which are "close" to the projection of a single target point. This provides a geometrically intuitive mechanism for modeling multi-valued relationships that is absent in simpler translational models.

This enhanced semantic capability is qualitatively corroborated by our neighborhood analysis (Table VIII). The more semantically coherent neighborhoods produced by SKGE suggest that the model is not just learning loose associations but a more robust *semantic topology*. The structure imposed

by the manifold and the expressiveness of the operator work in concert, ensuring that geometric proximity more reliably corresponds to conceptual similarity.

C. Broader Implications and Limitations

Our findings advocate for viewing KGE model design through a geometric lens. The success of SKGE suggests that future research should not only focus on designing more complex operators but also on identifying the right geometric "container" for them. The manifold's properties (e.g., curvature, compactness) provide a powerful inductive bias that can be tailored to the assumed structure of the data.

However, SKGE's inductive bias is not a panacea. The positive curvature of the sphere, while excellent for the densely connected graphs we studied, is theoretically less suited for embedding strict, tree-like hierarchies, where the negative curvature of hyperbolic space is provably more efficient [13]. Similarly, our distance-based operator may be less adept at capturing purely logical patterns like symmetry or composition compared to algebraic models like RotatE [9]. This highlights a critical open question: how to design models that can adapt their geometry, perhaps by learning on product manifolds ($S^n \times H^m$), to match the local topology of the knowledge graph.

VI. Conclusion and Future work

In this paper, we introduced SKGE, a model leveraging the geometric properties of a hypersphere to learn robust KGEs. We demonstrated that by constraining entities to this compact manifold via a learnable Spherization Layer and employing a translate-then-project transformation, SKGE significantly outperforms its strong Euclidean baseline.

Our central finding is that imposing a spherical geometry is not merely a constraint, but an active and powerful regularizer that fundamentally reshapes the optimization landscape, leading to the emergent property of "inherently hard negative sampling." This fosters the learning of more coherent and effective representations, particularly for complex, multi-valued relations. Our findings advocate for a deeper consideration of geometric priors in KGE design.

For future work, several avenues are promising. **1) Hybrid Geometries:** A key direction is exploring product spaces, such as $S^n \times H^m$ (a sphere and a Poincaré ball). A gating mechanism could learn to determine which geometry is more suitable for a given relation type (e.g., cyclical vs. hierarchical), creating a more expressive and adaptive model. **2) Contextual Transformations:** The static relation vector $\mathbf{r}_r$ could be replaced with a dynamic transformation computed via an attention mechanism over the neighborhood of the head entity, allowing the meaning of a relation to be context-dependent. **3) Temporal KGs:** Extending the SKGE framework to temporal knowledge graphs, where entities and relations evolve, is another critical research direction. The continuous nature of the sphere may offer a suitable manifold for modeling smooth temporal evolutions.

References

[1] Bordes, Antoine, Nicolas Usunier, Alberto García-Durán, Jason Weston and Oksana Yakhnenko. "Translating Embeddings for Modeling Multi-relational Data." Neural Information Processing Systems (2013).

[2] Ge, Xiou, Yun Cheng Wang, Bin Wang and C.-C. Jay Kuo. "Knowledge Graph Embedding: An Overview." ArXiv abs/2309.12501 (2023)

[3] Kim, Hoyong and Kangil Kim. "Spherization Layer: Representation Using Only Angles." Neural Information Processing Systems (2022).

[4] Wang, Zhen, Jianwen Zhang, Jianlin Feng and Zheng Chen. "Knowledge Graph Embedding by Translating on Hyperplanes." AAAI Conference on Artificial Intelligence (2014).

[5] Yankai Lin, Zhiyuan Liu, Maosong Sun, Yang Liu, and Xuan Zhu. 2015. "Learning entity and relation embeddings for knowledge graph completion." In Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence (AAAI'15). AAAI Press, 2181–2187.

[6] Nickel, Maximilian, Volker Tresp and Hans-Peter Kriegel. "A Three-Way Model for Collective Learning on Multi-Relational Data." International Conference on Machine Learning (2011).

[7] Yang, Bishan, Wen-tau Yih, Xiaodong He, Jianfeng Gao and Li Deng. "Embedding Entities and Relations for Learning and Inference in Knowledge Bases." International Conference on Learning Representations (2014).

[8] Trouillon, Théo, Johannes Welbl, Sebastian Riedel, Éric Gaussier and Guillaume Bouchard. "Complex Embeddings for Simple Link Prediction." ArXiv abs/1606.06357 (2016)

[9] Sun, Zhiqing, Zhihong Deng, Jian-Yun Nie and Jian Tang. "RotatE: Knowledge Graph Embedding by Relational Rotation in Complex Space." International Conference on Learning Representations (2019)

[10] L. Chao, J. He, T. Wang, and W. Chu, "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), 2021, 4360–9.

[11] Y. Peng and J. Zhang, "LineaRE: Simple but Powerful Knowledge Graph Embedding for Link Prediction", in 2020 IEEE International Conference on Data Mining (ICDM), IEEE Computer Society, 2020, 422–31.

[12] X. Ge, Y.-C. Wang, B. Wang, and C.-C. J. Kuo, "CompoundE: Knowledge graph embedding with translation, rotation and scaling compound operations", arXiv preprint arXiv:2207.05324, 2022.

[13] Nickel, Maximilian and Douwe Kiela. "Poincaré Embeddings for Learning Hierarchical Representations." Advances in Neural Information Processing Systems. Vol 30. (2017)

[14] Toutanova, Kristina and Danqi Chen. "Observed versus latent features for knowledge base and text inference." Workshop on Continuous Vector Space Models and their Compositionality (2015).

[15] Safavi, Tara and Danai Koutra. "CoDEx: A Comprehensive Knowledge Graph Completion Benchmark." Conference on Empirical Methods in Natural Language Processing (2020).

[16] D. Q. Nguyen, T. D. Nguyen, D. Q. Nguyen, and D. Phung, "A novel embedding model for knowledge base completion based on convolutional neural network," arXiv preprint arXiv:1712.02121, 2017.

[17] T. Dettmers, P. Minervini, P. Stenetorp, and S. Riedel, "Convolutional 2d knowledge graph embeddings," in Proceedings of the AAAI conference on artificial intelligence, vol. 32, 2018.

[18] M. Schlichtkrull, T. N. Kipf, P. Bloem, R. Van Den Berg, I. Titov, and M. Welling, "Modeling relational data with graph convolutional networks," in The semantic web: 15th international conference, ESWC 2018, Heraklion, Crete, Greece, June 3–7, 2018, proceedings 15, pp. 593–607, Springer, 2018.

[19] Ivana Balazevic, Carl Allen, and Timothy Hospedales. 2019. "Tucker: Tensor factorization for knowledge graph completion". In EMNLP-IJCNLP.

[20] Z. Cao, Q. Xu, Z. Yang, X. Cao, and Q. Huang, "Dual quaternion knowledge graph embeddings", in Proceedings of the Thirty-Fifth AAAI Conference on Artificial Intelligence, 2021, 6894–902.

[21] S. Zhang, Y. Tay, L. Yao, and Q. Liu, "Quaternion knowledge graph embeddings", in Proceedings of the 33rd International Conference on Neural Information Processing Systems, 2019, 2735–45.

[22] M. Nickel, L. Rosasco, and T. Poggio, "Holographic embeddings of knowledge graphs", in Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 30, No. 1, 2016.