AdaGeM: Adaptive Geometric Learning on Manifolds for Hyper-Relational Knowledge Graphs

Ke-Jia Chen[1,2] , **Yuzhe Hu**[2] , **Yu Ni**[2] and **Linfeng Liu**[1,2,*]
[1]Jiangsu Key Laboratory of Big Data Security and Intelligent Processing, Nanjing University of Posts and Telecommunications, Nanjing, China
[2]School of Computer Science, Nanjing University of Posts and Telecommunications, Nanjing, China
{chenkj, 1024040810, 1024040820, liulf}@njupt.edu.cn

Abstract

Hyper-relational Knowledge Graphs (HKGs) extend traditional knowledge graphs by introducing high-order dependencies, where auxiliary qualifiers provide detailed information. However, modeling such intricate topologies (e.g., mixtures of hierarchies, cycles, and chains) is challenging. Existing methods suffer from structural distortion due to reliance on a single geometric space and overlook dynamic semantics across entity roles. To address these issues, we propose AdaGeM (Adaptive Geometric Learning on Manifolds), an innovative space-adaptive framework that integrates geometric learning with a role-aware Transformer. First, we propose an adaptive geometric hypergraph encoder that projects entities into a product manifold and design a topology-aware gating mechanism to dynamically select optimal geometric spaces for each entity. Second, we develop a role-aware Transformer equipped with role-specific projections and micro-structural bias injection to refine embeddings by distinguishing entity semantics across different roles and focusing on valid n-ary interactions, respectively. Extensive experiments on benchmark datasets demonstrate that AdaGeM outperforms state-of-the-art baselines in entity/relation prediction tasks, with ablation studies validating the necessity of multi-manifold modeling for heterogeneous HKG structures.

1 Introduction

Hyper-relational Knowledge Graphs (HKGs) formalize real-world knowledge as hyper-relational facts, each consisting of a primary triplet $(subject, relation, object)$ and a set of attribute-value qualifiers that provide auxiliary details [Rosso *et al.*, 2020]. By moving beyond the simplification of standard Knowledge Graphs (KGs) that only model binary relations, HKGs encapsulate richer and more comprehensive structural information, which is critical for downstream tasks, such as knowledge completion [Galkin *et al.*, 2020] and reasoning [Guan *et al.*, 2020].

*Corresponding Author

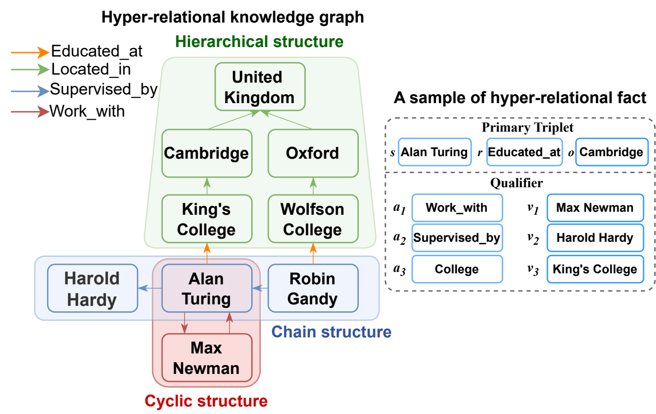


Figure 1: Illustration of intricate topological structures in HKGs. Entities are interconnected forming various topologies, including chain structures (blue), hierarchical structures (green), and cyclic structures (red).

As an emerging and pivotal research topic, HKG representation learning has evolved into two primary paradigms: sequence-based and structure-based methods. Sequence-based methods [Wang *et al.*, 2021; Chung *et al.*, 2023; Yu and Yang, 2021] often adopt Transformer [Vaswani *et al.*, 2017] architectures, serializing hyper-relational facts to capture contextual dependencies between elements. In contrast, structure-based methods [Galkin *et al.*, 2020; Luo *et al.*, 2023; Shomer *et al.*, 2023] employ Graph Neural Networks (GNNs) to explicitly model the graph topological patterns by aggregating information from both primary triplets and qualifiers. However, a critical limitation persists across most existing methods, i.e., they are confined to a single geometric space (typically Euclidean space), which fails to distinguish or uniformly model the diverse topological patterns in HKGs (e.g., hierarchical and cyclic patterns). This geometric mismatch leads to inevitable structural distortion, severely restricting the model's ability to encode complex topological nuances. Furthermore, these methods rely on static entity embeddings, neglecting the dynamic nature of entity semantics, that is, an entity's role can vary when it serves as a subject, object, or qualifier, thus leading to the loss of role-specific semantic information.

Geometric representation learning [Bronstein *et al.*, 2017;

Chami *et al.*, 2019] has emerged as a promising direction to GNNs beyond Euclidean space to non-Euclidean manifolds, which has been introduced to traditional KG research [Nickel and Kiela, 2017; Sun *et al.*, 2019] to model hierarchies and logical rules. For HKGs, a number of geometric methods have also been proposed in recent years, such as hyperbolic embedding [Yan *et al.*, 2022a], gyro-polygon modeling [Yan *et al.*, 2022b], and geometric transformations [Xiong *et al.*, 2023], aiming to capture their complex structures. Yet, these methods still suffer from structural distortion [Nickel and Kiela, 2017], as they rely solely on a single geometric space, overlooking the heterogeneous topological nature of real-world HKGs. As shown in Figure 1, the hierarchy centered on *Cambridge* aligns better with hyperbolic geometry due to its tree-like structure, while the cyclic interactions involving *Alan Turing* are better modeled in spherical space owing to their cyclic patterns.

To address the above issues, we propose **AdaGeM** (**Ada**ptive **Ge**ometric Learning on **M**anifolds), a novel framework that synthesizes global geometric structure with local semantic context through role-aware multi-manifold modeling. Our approach comprises two core components: (1) an adaptive geometric hypergraph encoder that projects each entity into multiple Riemannian manifolds to capture diverse topological patterns, followed by a topology-aware manifold gating mechanism to dynamically fuse these heterogeneous features; (2) a role-aware Transformer that incorporates role-aware projections and micro-structural biases, refining embeddings based on the specific role of each entity. Finally, we conduct comprehensive evaluations of AdaGeM on both entity and relation prediction tasks.

Our main contributions are summarized as follows: (*i*) We propose a space-adaptive geometric learning framework that resolves structural distortion in HKG embeddings by explicitly capturing heterogeneous topological patterns in n-ary facts. (*ii*) Methodologically, we design a geometric hypergraph encoder to model complex patterns across multiple manifolds and a role-aware Transformer to update entity representations based on their different roles, ensuring both structural fidelity and semantic discriminability. (*iii*) Extensive experiments on standard benchmarks demonstrate that AdaGeM outperforms state-of-the-art baselines on both entity and relation prediction tasks. Detailed analyses reveal the framework's exceptional robustness in capturing heterogeneous topologies and adapting to facts of varying arities.

2 Related Work

2.1 Hyper-relational Knowledge Graphs

Early methods of HKG modeling primarily adopt n-ary encoding schemes [Guan *et al.*, 2019; Liu *et al.*, 2021] while failing to prioritize the core semantic role of the primary triplets, leading to the loss of key relational semantics. Subsequent methods explicitly model primary triplets and attribute-value qualifiers as separate components. For instance, HINGE [Rosso *et al.*, 2020] and NeuInfer [Guan *et al.*, 2020] design dedicated encoding mechanisms for primary triplets and qualifiers, effectively preserving their unique semantics while capturing the interdependencies. From the perspective of graph topology, GNN-based methods become prevalent for capturing structural heterogeneity in HKGs. StarE [Galkin *et al.*, 2020] employs a message passing mechanism that embeds qualifiers into relation representations to iteratively update entity embeddings. QUAD [Shomer *et al.*, 2023] further enhances the paradigm with a bidirectional aggregation mechanism between primary triplets and qualifiers. HAHE [Luo *et al.*, 2023] integrates global hypergraph structures with local semantic sequences to capture complex dependencies inherent in n-ary fact representations. HyperFormer [Hu *et al.*, 2023] utilizes a mixture-of-experts mechanism to effectively mitigate noise from multi-hop neighbors. In parallel, Transformer-based methods have been introduced to capture fine-grained interactions. GRAN [Wang *et al.*, 2021] uses edge-aware attention to explicitly differentiate between various relation types. HyNT [Chung *et al.*, 2023] extends sequential architectures to effectively encode numeric attribute values in qualifiers. MAYPL [Lee and Whang, 2025] introduces a purely structure-driven framework that enables robust inductive reasoning on unseen entities.

However, these methods suffer from a fundamental limitation: they are confined to a single geometric space (typically Euclidean space). This limitation causes structural distortion as Euclidean geometry cannot faithfully preserve the diverse topologies within HKGs. Moreover, these methods rely on static entity embeddings, neglecting the dynamic nature of entity roles. We explicitly address these two issues in our work.

2.2 Geometric Graph Learning

To overcome the limitation of Euclidean space, geometric GNNs have been proposed to learn representations in non-Euclidean manifolds [Han *et al.*, 2025]. For hierarchical structures, hyperbolic GNNs leverage the negative curvature of hyperbolic space to minimize distortion when handling tree-like structures. This paradigm is pioneered by Poincaré Embeddings [Nickel and Kiela, 2017] and later extended to hyperbolic manifolds by HGCN [Chami *et al.*, 2019].

Recently, geometric learning has been introduced to HKGs, aiming to capture their complex structures. HYPER2 [Yan *et al.*, 2022a] projects entities into hyperbolic space and performs message passing in the tangent space. PolygonE [Yan *et al.*, 2022b] models n-ary facts as gyro-polygons within the Poincaré ball, using geometric centroids to represent the semantic core of each fact. Beyond hierarchies, ShrinkE [Xiong *et al.*, 2023] treats qualifiers as functional transformations to enforce logical entailment. HyCubE [Li *et al.*, 2025] introduces 3D circular convolutions to model cyclic interactions between entities and relations.

Despite these advancements, existing geometric methods for HKGs typically rely on a single geometric space (e.g., purely hyperbolic or spherical), failing to accommodate the coexistence of diverse topological patterns. This highlights an urgent need for more flexible and adaptive geometric modeling methods that can dynamically select appropriate geometric spaces based on the intrinsic structure of entities and relations, a gap our work endeavors to bridge.

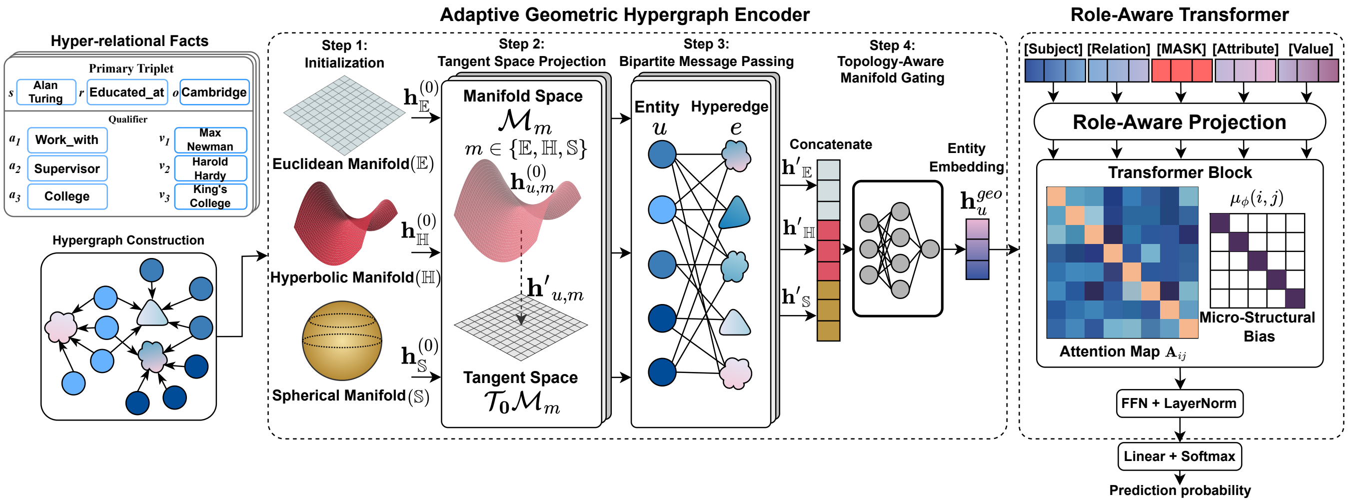


Figure 2: The overview of AdaGeM framework.

3 Preliminary

We first formalize the key concepts and mathematical foundations underlying AdaGeM, ensuring clarity for subsequent methodological descriptions.

Definition 1 (**Hyper-relational Fact**)**.** A hyper-relational fact is defined as a tuple: $f = (s, r, o, \mathcal{Q}) \in \mathcal{F}$. Here, (s, r, o) represents the primary triplet, where $s, o \in \mathcal{E}$ (the set of entities) and $r \in \mathcal{R}$ (the set of relations). The set $\mathcal{F}$ denotes the collection of all hyper-relational facts. The set $\mathcal{Q} = \{(a_i, v_i)\}_{i=1}^{k}$ consists of attribute-value qualifiers that provide auxiliary contextual details, where each attribute $a_i \in \mathcal{R}$, and the corresponding value $v_i \in \mathcal{E}$. k denotes the number of qualifiers in f.

Definition 2 (**Hyper-relational Knowledge Graph**)**.** An HKG is represented as $\mathcal{G} = (\mathcal{E}, \mathcal{R}, \mathcal{F})$. Notably, $\mathcal{E}$ and $\mathcal{R}$ encompass all the values and attributes from the qualifiers, respectively. This formulation unifies both standard triplets (where $\mathcal{Q} = \emptyset$) and hyper-relational facts (where $\mathcal{Q} \neq \emptyset$).

Definition 3 (**Hypergraph Modeling**)**.** To explicitly capture the high-order structural dependencies inherent in HKGs, we model the HKG as a hypergraph $\mathcal{G}_H = (\mathcal{V}_H, \mathcal{E}_H)$, where each hyperedge $e_f \in \mathcal{E}_H$ corresponds to a fact f and connects all entities involved in its primary triplet and qualifiers. Here, $\mathcal{V}_H$ corresponds to the entity set $\mathcal{E}$ of the HKG, and $\mathcal{E}_H$ denotes the hyperedge set derived from the fact set $\mathcal{F}$.

Definition 4 (**Product Manifold**)**.** To capture the heterogeneous topological patterns in HKGs, we formalize the embedding space as a product manifold $\mathcal{M} = \mathbb{E}^d \times \mathbb{H}^d_{c_h} \times \mathbb{S}^d_{c_s}$ [Gu *et al.*, 2018]. This manifold is constructed via the Cartesian product of three sub-manifolds, each equipped with a dimension d. Here, $\mathbb{E}^d$ represents the Euclidean subspace, and $\mathbb{H}^d_{c_h}$ and $\mathbb{S}^d_{c_s}$ denote the hyperbolic manifold (with a learnable negative curvature $c_h < 0$) and spherical manifold (with a learnable positive curvature $c_s > 0$), respectively. The Riemannian operations are detailed in Appendix B.

4 Methodology

We propose a space-adaptive geometric learning framework for HKGs, as illustrated in Figure 2. The framework consists of two innovative core components: an adaptive geometric hypergraph encoder and a role-aware Transformer. The former captures global heterogeneous topological features across multiple manifolds, while the latter refines these features by modeling the local semantic context of each hyper-relational fact with role-specific and structural bias guidance.

4.1 Adaptive Geometric Hypergraph Encoder

HKGs inherently contain heterogeneous topological patterns (e.g., hierarchies, cycles, and chains), which cannot be effectively modeled by a single geometric space. We address this issue via a multi-stage encoding process.

Hypergraph Initialization

To preserve the integrity of n-ary interactions, we propose a geometric dual-initialization strategy for both entities and hyperedges in the product manifold $\mathcal{M}$ (see Definition 4). For each component manifold $\mathcal{M}_m$ ($m \in \{\mathbb{E}, \mathbb{H}, \mathbb{S}\}$), we initialize the entity matrix $\mathbf{H}_m^{(0)} \in \mathbb{R}^{|\mathcal{E}| \times d}$ and the hyperedge matrix $\mathbf{Z}_m^{(0)} \in \mathbb{R}^{|\mathcal{E}_H| \times d}$, where $\mathcal{E}$, $\mathcal{E}_H$, and d denote the sets of entities, hyper-relational facts, and feature dimension, respectively. These tangent vectors serve as the raw learnable weights optimized by standard Euclidean optimizers, while their geometric positions on the manifold are induced via the exponential map:

$$\mathbf{h}_{u,m}^{(0)} = \exp_{\mathbf{0}}^{m}(\mathbf{h'}_{u,m}^{(0)}), \quad \mathbf{z}_{e,m}^{(0)} = \exp_{\mathbf{0}}^{m}(\mathbf{z'}_{e,m}^{(0)}). \tag{1}$$

Tangent Space Convolution

After initialization, we implement bipartite message passing on Riemannian manifolds via the two following steps:

Tangent Space Projection. For manifold-mapped embeddings in subsequent layers, we utilize the logarithmic map to project them back to the tangent space $\mathcal{T}_{\mathbf{0}}\mathcal{M}_m$ for linear operations:

$$\mathbf{h'}_{u,m} = \log_{\mathbf{0}}^{m}(\mathbf{h}_{u,m}^{(l)}). \tag{2}$$

This strategy linearizes non-Euclidean geometric features, enabling standard neural attention mechanisms while retaining structural information from the manifold.

Message Passing and Update. We adopt a bipartite message passing mechanism to model high-order entity-hyperedge dependencies. The hyperedge representations are updated by aggregating information from their constituent entities. For a hyperedge $e \in \mathcal{E}_H$, we compute the attention score $\alpha_{u,e}^m$ for each constituent entity $u \in e$ in the tangent plane to quantify its semantic contribution to e:

$$\alpha_{u,e}^m = \text{LeakyReLU}\left((\mathbf{a}_m)^\top [\mathbf{W}_{m,u}\mathbf{h'}_{u,m} \,\|\, \mathbf{W}_{m,e}\mathbf{z'}_{e,m}]\right), \tag{3}$$

where $\mathbf{a}_m$ is a learnable attention vector, $\mathbf{h'}_{u,m}$ and $\mathbf{z'}_{e,m}$ denote the projected tangent vectors of entity u and hyperedge e, respectively, $\mathbf{W}_{m,u}, \mathbf{W}_{m,e} \in \mathbb{R}^{d\times d}$ are their respective manifold-specific linear transformation matrices. $\tilde{\mathbf{z'}}_{e,m}$ is then updated by aggregating node features using normalized attention weights:

$$\tilde{\mathbf{z'}}_{e,m} = \sigma\left(\sum_{u\in e}\frac{\exp(\alpha_{u,e}^m)}{\sum_{k\in e}\exp(\alpha_{k,e}^m)}\mathbf{W}_{m,u}\mathbf{h'}_{u,m}\right), \tag{4}$$

where $\sigma(\cdot)$ denotes an activation function (e.g., ReLU). Similarly, the entity embeddings are updated by aggregating information from incident hyperedges. For an entity u, an attention score $\beta_{u,e}^m$ is computed for each hyperedge e to quantify their semantic compatibility.

$$\beta_{u,e}^m = \text{LeakyReLU}\left((\mathbf{b}_m)^\top [\mathbf{W}_{m,u}\mathbf{h'}_{u,m} \,\|\, \mathbf{W}_{m,e}\tilde{\mathbf{z'}}_{e,m}]\right), \tag{5}$$

where $\mathbf{b}_m$ is a learnable attention vector that measures the relevance of hyperedge context to the entity. Then, the entity tangent vector $\tilde{\mathbf{h'}}_{u,m}$ is updated via a weighted aggregation of these hyperedge representations:

$$\tilde{\mathbf{h'}}_{u,m} = \sigma\left(\sum_{e:u\in e}\frac{\exp(\beta_{u,e}^m)}{\sum_{k:u\in k}\exp(\beta_{u,k}^m)}\mathbf{W}_{m,e}\tilde{\mathbf{z'}}_{e,m}\right), \tag{6}$$

Finally, to ensure geometric validity, the updated tangent vectors for both entities and hyperedges are mapped back to the original manifold via the exponential map:

$$\mathbf{h}_{u,m}^{(l+1)} = \exp_{\mathbf{0}}^m(\tilde{\mathbf{h'}}_{u,m}), \quad \mathbf{z}_{e,m}^{(l+1)} = \exp_{\mathbf{0}}^m(\tilde{\mathbf{z'}}_{e,m}), \tag{7}$$

where l denotes the index of the current layer.

Topology-Aware Manifold Gating

To address redundancy and misalignment from multi-manifold modeling, we propose a Topology-Aware Manifold Gating (TAMG) mechanism that dynamically selects and fuses the most relevant geometric features without auxiliary regularization.

Let $\mathbf{h}'_{u,\mathbb{E}}, \mathbf{h}'_{u,\mathbb{H}}, \mathbf{h}'_{u,\mathbb{S}} \in \mathbb{R}^d$ denote the projected tangent vectors of entity u from the Euclidean, hyperbolic, and spherical manifolds, respectively. We first concatenate these vectors to form a unified geometric context $\mathbf{C}_u$:

$$\mathbf{C}_u = [\mathbf{h}'_{u,\mathbb{E}} \,\|\, \mathbf{h}'_{u,\mathbb{H}} \,\|\, \mathbf{h}'_{u,\mathbb{S}}]. \tag{8}$$

Subsequently, manifold-specific importance scores are computed via a learned gating network to adaptively determine the dominant manifold features:

$$\boldsymbol{\alpha}_u = \text{Softmax}(\mathbf{W}_g\mathbf{C}_u + \mathbf{b}_g), \tag{9}$$

where $\mathbf{W}_g \in \mathbb{R}^{3\times 3d}$ and $\mathbf{b}_g \in \mathbb{R}^3$ are learnable parameters. The attention vector $\boldsymbol{\alpha}_u = [\alpha_{u,\mathbb{E}}, \alpha_{u,\mathbb{H}}, \alpha_{u,\mathbb{S}}]$ represents the relevance of each geometric space for entity u, emphasizing the manifold that best aligns with the entity's topology.

The final fused representation $\mathbf{h}_u^{geo}$ is obtained via weighted aggregation:

$$\mathbf{h}_u^{geo} = \sum_{m\in\{\mathbb{E},\mathbb{H},\mathbb{S}\}} \alpha_{u,m}\cdot \mathbf{h}'_{u,m}. \tag{10}$$

4.2 Role-Aware Transformer

While the above proposed encoder distinct static global topology, entities may exhibit different roles across hyper-relational facts (e.g., subject, object, qualifier). Therefore, we design a role-aware Transformer to further refine the embeddings with local semantic context. Two key strategies are proposed as follows.

Role-Aware Projection

We first serialize a hyper-relational fact into a sequence of tokens $S = (x_1, \ldots, x_L)$ (primary triplet followed by qualifier pairs) and then define five core roles $\mathcal{C} = \{\text{SUB}, \text{REL}, \text{OBJ}, \text{ATTR}, \text{VAL}\}$. Each position index i in the sequence is mapped to a specific role via a structural mapping function $\phi : \{1, \ldots, L\} \to \mathcal{C}$.

For the token x_i corresponding to an entity u, we compute the role-enhanced representation $\mathbf{z}_i^{(0)}$ by augmenting the topological prior $\mathbf{h}_{u_i}^{geo}$ with role-specific semantics:

$$\mathbf{z}_i^{(0)} = \mathbf{h}_{u_i}^{geo} + \mathbf{e}_{\phi(i)}, \tag{11}$$

where $\mathbf{e}_{\phi(i)}$ represents a learnable role embedding. This projection distinguishes role-dependent entity semantics, resolving the ambiguity of static embeddings.

Micro-Structural Bias Injection

To guide the model toward valid n-ary interaction, we innovatively inject a micro-structural bias into the standard self-attention mechanism.

First, we categorize token pairs into five structural edge types: (s, r), (r, o), (r, a), (a, v), and "*other*" (non-structural interactions), and employ a mapping function $\phi(i, j)$ to identify the structural category of the token pair (i, j).

Then, we compute the structural attention score s_{ij} to quantify the pairwise importance with an injected bias:

$$s_{ij} = \frac{(\mathbf{x}_i\mathbf{W}_Q)(\mathbf{x}_j\mathbf{W}_K)^\top}{\sqrt{d}} + \mu_{\phi(i,j)}, \tag{12}$$

where $\mathbf{W}_Q, \mathbf{W}_K \in \mathbb{R}^{d\times d}$ are learnable projection matrices, and $\mu_{\phi(i,j)}$ is a learnable scalar bias for the edge type $\phi(i, j)$. This bias acts as a soft adjacency constraint, up-weighting meaningful interactions and suppressing noise.

The token representations are dynamically refined via weighted aggregation of value vectors:

$$\tilde{\mathbf{x}}_i = \sum_{j=1}^{L} \frac{\exp(s_{ij})}{\sum_{k=1}^{L} \exp(s_{ik})} (\mathbf{x}_j \mathbf{W}_V), \tag{13}$$

where $\mathbf{W}_V \in \mathbb{R}^{d \times d}$ is the learnable projection matrix for value vectors.

Finally, the output representations $\mathbf{x}_i^{(l+1)}$ are obtained by passing $\tilde{\mathbf{x}}_i$ through a Feed-Forward Network (FFN), followed by residual connections and layer normalization:

$$\mathbf{x}_i^{(l+1)} = \text{LayerNorm}\left(\tilde{\mathbf{x}}_i + \text{FFN}(\tilde{\mathbf{x}}_i)\right). \tag{14}$$

This process fuses global topological features with fine-grained local semantics, ensuring discriminative entity representations.

4.3 Training and Optimization

We adopt an end-to-end training strategy for AdaGeM, with the manifold curvature parameters optimized as learnable variables.

Training Objective

We employ the Binary Cross-Entropy (BCE) loss to maximize the likelihood of observed facts as much as possible, while suppressing negative samples that are generated by corrupting the subject, object, relation, attribute or value, in order to distinguish valid n-ary structures from invalid ones. The loss function is formulated as:

$$\mathcal{L} = -\frac{1}{N} \sum_{i=1}^{N} \left(y_i \log(p_i) + (1 - y_i) \log(1 - p_i) \right), \tag{15}$$

where y_i is the ground-truth label (1 for observed positive facts and 0 for generated negative ones), p_i is the model's predicted probability, and N is the batch size.

Optimization Strategy

Direct optimization on curved manifolds is computationally expensive. To balance efficiency and geometric validity, we adopt a tangent space optimization strategy. All embeddings are defined in the tangent space $\mathcal{T}_{\mathbf{0}}\mathcal{M}_m$ (isomorphic to Euclidean space $\mathbb{R}^d$), thereby allowing the employment of standard Euclidean optimizers. At each training step, parameters are updated in the tangent space and then projected back to the manifold via the exponential map $\exp_{\mathbf{0}}^{\mathbf{m}}$.

Finally, we conduct a comprehensive theoretical analysis of the proposed framework, covering the complexity analysis, the expressive capacity of product manifolds, and the geometric separability derived from inductive priors. Detailed discussions and proofs are provided in Appendix C, D, and E.

5 Experiments

We conduct extensive experiments to validate the effectiveness of AdaGeM. On three standard HKG benchmarks, we aim to answer the following research questions (RQs): **RQ1:** Does AdaGeM outperform state-of-the-art baselines on both entity and relation prediction tasks? **RQ2:** What is the individual contribution of key components to the overall performance? **RQ3:** How do different geometric manifolds contribute to capturing heterogeneous topological patterns in HKGs? **RQ4:** Can the model adaptively select appropriate geometric spaces for distinct topological patterns? **RQ5:** How does AdaGeM perform across different semantic roles, especially for qualifier values? **RQ6:** How does AdaGeM perform across hyper-relational facts with varying arities and entity degrees?

5.1 Experimental Setup

Datasets. We adopt three HKG benchmarks (JF17K [Wen *et al.*, 2016], Wikipeople [Guan *et al.*, 2020], and WD50K [Galkin *et al.*, 2020]), covering diverse scales, arity ranges, and qualifier densities, to ensure generalizability. Dataset statistics are summarized in Appendix F.

Baselines. We compare AdaGeM against a comprehensive collection of baselines, which can be categorized into four groups: (1) **Spatial Mapping-based Methods**, including m-TransH [Wen *et al.*, 2016], RAE [Zhang *et al.*, 2018], and HYPER2 [Yan *et al.*, 2022a]; (2) **CNN, FCN-based Methods**, including NaLP [Guan *et al.*, 2019], NeuInfer [Guan *et al.*, 2020], HINGE [Rosso *et al.*, 2020], and HJE [Li *et al.*, 2024]; (3) **GNN-based Methods**, including StarE [Galkin *et al.*, 2020], MSeaHKG [Di and Chen, 2023], HAHE [Luo *et al.*, 2023], and MAYPL [Lee and Whang, 2025]; and (4) **Transformer-based Methods**, including GRAN [Wang *et al.*, 2021], HyTransformer [Yu and Yang, 2021], and HyNT [Chung *et al.*, 2023]. For all baselines, we adopt the optimal hyperparameters reported in their original papers or tune them via grid search on the validation set to ensure peak performance (see Appendix G).

5.2 Overall Performance (RQ1)

As presented in Table 1, AdaGeM consistently outperforms all baselines across datasets and metrics for both subject/object prediction and all-entity prediction. The performance gains exhibit distinct patterns across datasets, reflecting the model's adaptability to varying topological characteristics. On JF17K, AdaGeM achieves the most significant margin over the runner-up GRAN, with improvements of 0.021 in MRR and 7.6% in Hits@1 for the primary triplet prediction. This can be attributed to the mixed-curvature representations, which effectively capture the hierarchical structures underestimated by Euclidean-based models. On the larger-scale and structurally sparser datasets of Wikipeople and WD50K, AdaGeM maintains a steady performance lead, surpassing strong baselines like HAHE and MAYPL. This demonstrates the model's robustness and scalability in handling high-arity relations and scattered topological information. For all-entity prediction (including qualifiers), AdaGeM shows resilient performance. For instance, on WD50K, it outperforms MAYPL by 3.7% in Hits@1, validating its ability to capture fine-grained qualifier semantics. Additionally, detailed results for relation prediction are reported in Appendix J, which further confirms the capacity of AdaGeM to model relational semantics.

5.3 Component Ablation Analysis (RQ2)

We conduct ablation studies to evaluate the contribution of each core component of AdaGeM. The variants are w/o

Model	JF17K						Wikipeople						WD50K					
	Subject/Object			All entities			Subject/Object			All entities			Subject/Object			All entities		
	MRR	H@1	H@10	MRR	H@1	H@10	MRR	H@1	H@10	MRR	H@1	H@10	MRR	H@1	H@10	MRR	H@1	H@10
m-TransH	0.206	0.206	0.462	0.102	0.069	0.168	0.063	0.063	0.300	0.137	0.130	0.308	-	-	-	-	-	-
RAE	0.215	0.215	0.466	0.310	0.219	0.504	0.058	0.058	0.306	0.172	0.102	0.320	-	-	-	-	-	-
NaLP	0.221	0.165	0.331	0.366	0.290	0.516	0.408	0.331	0.546	0.338	0.272	0.466	0.177	0.131	0.264	0.224	0.158	0.330
NeuInfer	0.449	0.361	0.624	0.473	0.397	0.618	0.476	0.415	0.585	0.333	0.259	0.477	0.243	0.176	0.377	0.228	0.162	0.341
HINGE	0.431	0.342	0.611	0.517	0.436	0.675	0.342	0.272	0.463	0.350	0.282	0.467	0.251	0.192	0.433	0.232	0.164	0.343
StarE	0.574	0.496	0.725	0.542	0.454	0.685	0.491	0.398	0.592	0.378	0.265	0.542	0.349	0.271	0.496	0.326	0.251	0.469
GRAN	0.610	0.539	0.770	0.656	0.582	0.799	0.503	0.438	0.620	0.479	0.410	0.604	0.329	0.259	0.465	0.309	0.240	0.441
MSeaHKG	-	-	-	0.577	0.481	0.711	-	-	-	0.395	0.291	0.554	-	-	-	-	-	-
HyTransformer	0.582	0.501	0.742	-	-	-	0.501	0.426	0.634	-	-	-	0.356	0.281	0.498	-	-	-
HYPER2	0.583	0.500	0.746	-	-	-	0.461	0.391	0.597	-	-	-	-	-	-	-	-	-
ShrinkE	0.589	0.506	0.749	-	-	-	0.485	0.431	0.601	-	-	-	0.345	0.275	0.482	-	-	-
HAHE	0.614	0.539	0.765	0.664	0.585	0.796	0.494	0.426	0.614	0.497	0.428	0.617	0.343	0.270	0.483	0.378	0.306	0.514
HyNT	0.562	0.481	0.714	0.588	0.50	0.728	0.459	0.377	0.597	0.457	0.376	0.597	0.308	0.240	0.439	0.337	0.271	0.464
HJE	0.450	0.375	0.582	0.590	0.507	0.729	0.471	0.387	0.608	0.472	0.387	0.609	0.322	0.252	0.455	0.345	0.274	0.477
MAYPL	0.616	0.521	0.806	0.623	0.569	0.740	0.488	0.427	0.639	0.487	0.420	0.647	0.368	0.285	0.516	0.408	0.328	0.537
Ours	**0.638**	**0.580**	**0.836**	**0.686**	**0.614**	**0.859**	**0.511**	**0.442**	**0.642**	**0.505**	**0.430**	**0.656**	**0.382**	**0.301**	**0.554**	**0.419**	**0.340**	**0.597**

Table 1: Overall entity prediction results on JF17K, Wikipeople, and WD50K datasets. The symbol "-" denotes methods inapplicable to this task. The best results are highlighted in **bold**.

Variant	JF17K						Wikipeople						WD50K					
	Subject/Object			All entities			Subject/Object			All entities			Subject/Object			All entities		
	MRR	H@1	H@10	MRR	H@1	H@10	MRR	H@1	H@10	MRR	H@1	H@10	MRR	H@1	H@10	MRR	H@1	H@10
w/o TAMG	0.615	0.548	0.805	0.662	0.590	0.825	0.478	0.410	0.615	0.482	0.398	0.620	0.355	0.270	0.510	0.388	0.302	0.555
w/o Role-aware Projection	0.608	0.535	0.798	0.655	0.582	0.816	0.472	0.405	0.608	0.476	0.392	0.612	0.350	0.265	0.508	0.382	0.296	0.548
w/o Micro-Structural Bias	0.622	0.558	0.818	0.670	0.602	0.838	0.485	0.422	0.622	0.490	0.408	0.628	0.348	0.262	0.505	0.375	0.290	0.545
w/o Curvature Learning	0.628	0.565	0.825	0.678	0.608	0.845	0.492	0.430	0.630	0.496	0.415	0.635	0.370	0.288	0.535	0.405	0.325	0.575
Full Model	**0.638**	**0.580**	**0.836**	**0.686**	**0.614**	**0.859**	**0.511**	**0.442**	**0.642**	**0.505**	**0.430**	**0.656**	**0.382**	**0.301**	**0.554**	**0.419**	**0.340**	**0.597**

Table 2: Ablation study of key components on JF17K, Wikipeople, and WD50K datasets. w/o TAMG replaces the gating mechanism with concatenation; w/o Role-aware Projection removes the role-specific entity embeddings; w/o Micro-Structural Bias removes the edge-biased attention; w/o Curvature Learning fixes curvature. The best results are highlighted in **bold**.

TAMG, w/o Role-Aware Projection, w/o Micro-Structural Bias, and w/o Curvature Learning. As shown in Table 2, all ablation variants underperform the full model. The most significant performance drop occurs when removing the Role-Aware Projection, with MRR decreasing from 0.638 to 0.608 for subject/object prediction on JF17K. This underscores the importance of distinguishing entity semantics across different roles. Similarly, replacing the TAMG mechanism with simple concatenation leads to a notable decline by 3.6% in MRR, confirming that dynamic manifold selection is more effective than rigid aggregation, as it enables the model to highlight the most relevant geometric priors. Furthermore, eliminating the micro-structural bias also degrades the performance, as the model lacks the explicit guidance to focus on valid local interactions (e.g., attribute-value coupling in qualifiers). Finally, fixing curvatures reduces the model's flexibility, indicating that adaptively fitting the curvature to the data's intrinsic structure is essential for minimizing structural distortion.

5.4 Geometric Manifold Contribution (RQ3)

To investigate the contribution of different geometric manifolds, we evaluate AdaGeM variants with single, dual, and triple manifold combinations. The results in Table 3 highlight that single-manifold variants exhibit limited performance, in contrast to dual-manifold combinations, which yield significant performance gains. The full model ($\mathbb{E}+\mathbb{H}+\mathbb{S}$) achieves the best performance across all metrics and datasets. For instance, on the JF17K dataset, it surpasses the strongest dual-manifold variant ($\mathbb{E}+\mathbb{H}$) by distinct margins, with improvements of 0.026 (4.2%) in MRR and 0.037 (6.8%) in Hits@1 for subject/object prediction. This validates the necessity of synthesizing all three geometric spaces to capture the heterogeneous topological patterns inherent in HKGs.

5.5 Geometric Adaptation Analysis (RQ4)

To verify the adaptivity of AdaGeM in selecting optimal geometric spaces for heterogeneous topological patterns, we visualize the average manifold gating weights for representative relations in Figure 3. The results reveal a strong alignment between the learned preferences and the intrinsic topology of relations. Specifically, hierarchical relations such as *subclass_of* and *parent_of* are dominated by the hyperbolic manifold ($\mathbb{H}$), with weights consistently exceeding 0.60, confirming its inherent advantage in modeling tree-like structures with minimal distortion. Cyclic relations like *spouse* and *friend_of* exhibit a distinct preference for the spherical manifold ($\mathbb{S}$) with weights reaching up to 0.70, which effectively captures closed-loop logical constraints. Moreover, attributes like *published_date* primarily rely on the Euclidean space ($\mathbb{E}$) as linear or regular patterns are best modeled in flat space. This clear differentiation validates that AdaGeM effectively captures the intrinsic topology of different relations through adaptive geometric selection.

5.6 Role-Specific Performance Analysis (RQ5)

To investigate AdaGeM's capability in perceiving different semantic roles, we analyzed the MRR performance on the

Variant	JF17K						Wikipeople						WD50K					
	Subject/Object			All entities			Subject/Object			All entities			Subject/Object			All entities		
	MRR	H@1	H@10	MRR	H@1	H@10	MRR	H@1	H@10	MRR	H@1	H@10	MRR	H@1	H@10	MRR	H@1	H@10
Only $\mathbb{E}$	0.582	0.501	0.742	0.625	0.548	0.760	0.465	0.402	0.595	0.470	0.385	0.602	0.330	0.255	0.485	0.365	0.288	0.520
Only $\mathbb{H}$	0.561	0.485	0.728	0.605	0.530	0.745	0.468	0.405	0.601	0.475	0.392	0.608	0.345	0.268	0.498	0.380	0.295	0.535
Only $\mathbb{S}$	0.540	0.462	0.710	0.582	0.505	0.725	0.442	0.380	0.575	0.450	0.365	0.585	0.305	0.230	0.460	0.340	0.260	0.495
$\mathbb{E}+\mathbb{H}$	0.612	0.543	0.795	0.658	0.585	0.810	0.490	0.428	0.625	0.492	0.412	0.630	0.365	0.285	0.525	0.402	0.320	0.565
$\mathbb{E}+\mathbb{S}$	0.605	0.531	0.782	0.645	0.570	0.805	0.482	0.415	0.618	0.485	0.405	0.625	0.355	0.272	0.510	0.390	0.305	0.550
$\mathbb{H}+\mathbb{S}$	0.598	0.518	0.774	0.638	0.562	0.790	0.485	0.420	0.620	0.488	0.410	0.628	0.360	0.280	0.518	0.395	0.315	0.558
Full Model	**0.638**	**0.580**	**0.836**	**0.686**	**0.614**	**0.859**	**0.511**	**0.442**	**0.642**	**0.505**	**0.430**	**0.646**	**0.382**	**0.301**	**0.554**	**0.419**	**0.340**	**0.597**

Table 3: Ablation study of manifold combinations on JF17K, Wikipeople, and WD50K datasets. $\mathbb{E}$, $\mathbb{H}$, and $\mathbb{S}$ denote Euclidean, hyperbolic, and spherical manifolds, respectively.

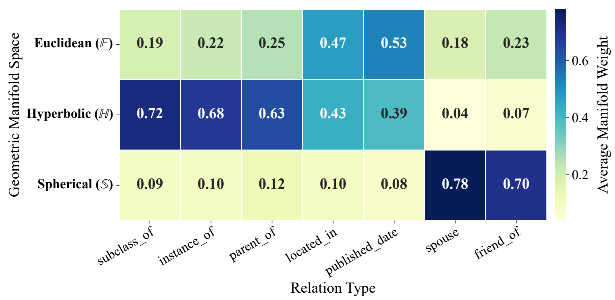


Figure 3: Correlation between heterogeneous topological patterns and manifold preferences.

JF17K dataset across three entity roles: Subject, Object, and Qualifier Value, comparing AdaGeM with NaLP and StarE.

As shown in Figure 4, AdaGeM consistently outperforms the baselines across all roles. Specifically, NaLP suffers a significant performance drop on qualifier values, indicating that simple sequence modeling struggles to distinguish qualifiers from the primary triplet, leading to semantic confusion. While StarE improves upon this by utilizing graph structures, it still lags behind our method. This validates the effectiveness of our role-aware projection, which enables the model to adaptively refine entity representations based on their specific roles. Supplementary results are presented in Appendix L.

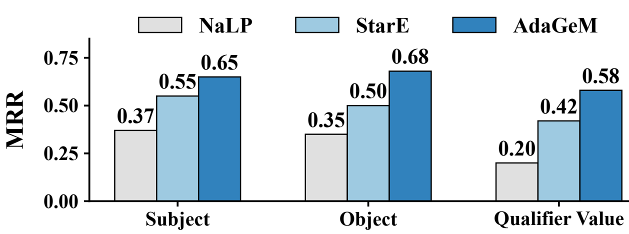
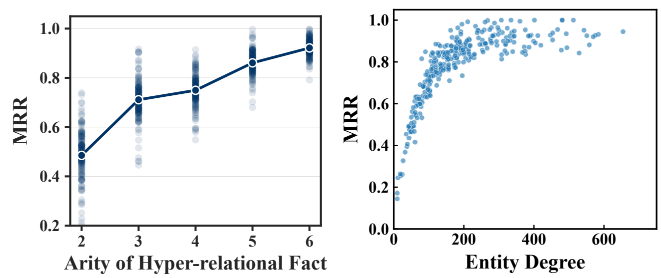


Figure 4: Performance comparison (MRR) across different entity roles (Subject, Object, and Qualifier Value) on JF17K.

5.7 Impact of Fact Arity and Entity Degree (RQ6)

To investigate AdaGeM's adaptability to structural complexity and data sparsity, we analyze its performance across diverse fact arities and entity degrees. Figure 5a illustrates the arity impact on JF17K (solid line denotes average MRR). For arity-2 facts, AdaGeM achieves a 0.51 baseline MRR, significantly improving to 0.72 at arity 3, and 0.89 at arity 6. This suggests auxiliary qualifiers provide valuable disambiguating context rather than noise. Additionally, Figure 5b shows a rapid performance saturation regarding entity degree. AdaGeM achieves competitive accuracy for low-degree tail entities and maintains stability for high-degree ones, demonstrating robust adaptability to varying fact complexity and uneven topological distributions in HKGs.



(a) (b)

Figure 5: Performance comparison (MRR) across different (a) fact arities and (b) entity degrees on JF17K.

6 Conclusion

In this paper, we address the core challenge of structural distortion and semantic ambiguity in Hyper-relational Knowledge Graph (HKG) representation learning by proposing a space-adaptive geometric learning framework. AdaGeM innovatively synergizes global topology and local semantics which dynamically fuses multi-manifold spaces via topology-aware gating to explicitly capture heterogeneous graph structures, optimizing both geometric fidelity and efficiency. A specialized role-aware Transformer then resolves semantic ambiguity by injecting role-specific and micro-structural biases. Extensive experiments on three standard benchmarks demonstrate that our proposed AdaGeM consistently outperforms state-of-the-art baselines on both entity and relation prediction tasks. This work establishes a new paradigm for HKG representation learning by bridging geometric topology modeling and semantic role awareness.

References

[Bronstein *et al.*, 2017] Michael M. Bronstein, Joan Bruna, Yann Lecun, Arthur Szlam, and Pierre Vandergheynst. Geometric Deep Learning: Going Beyond Euclidean Data. *IEEE Signal Processing Magazine*, 34(4):18–42, July 2017.

[Chami *et al.*, 2019] Ines Chami, Rex Ying, Christopher Ré, and Jure Leskovec. Hyperbolic graph convolutional neural networks. *Advances in neural information processing systems*, 32:4869, 2019.

[Chung *et al.*, 2023] Chanyoung Chung, Jaejun Lee, and Joyce Jiyoung Whang. Representation learning on hyper-relational and numeric knowledge graphs with transformers. In *Proceedings of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 310–322, 2023.

[Di and Chen, 2023] Shimin Di and Lei Chen. Message function search for knowledge graph embedding. In *Proceedings of the ACM Web Conference 2023*, page 2633–2644, 2023.

[Galkin *et al.*, 2020] Mikhail Galkin, Priyansh Trivedi, Gaurav Maheshwari, Ricardo Usbeck, and Jens Lehmann. Message passing for hyper-relational knowledge graphs. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 7346–7359, 2020.

[Gu *et al.*, 2018] Albert Gu, Frederic Sala, Beliz Gunel, and Christopher Ré. Learning mixed-curvature representations in product spaces. In *International conference on learning representations*, 2018.

[Guan *et al.*, 2019] Saiping Guan, Xiaolong Jin, Yuanzhuo Wang, and Xueqi Cheng. Link prediction on n-ary relational data. In *The World Wide Web Conference*, page 583–593, 2019.

[Guan *et al.*, 2020] Saiping Guan, Xiaolong Jin, Jiafeng Guo, Yuanzhuo Wang, and Xueqi Cheng. Neuinfer: Knowledge inference on n-ary facts. In *Proceedings of the 58th annual meeting of the association for computational linguistics*, pages 6141–6151, 2020.

[Han *et al.*, 2025] Jiaqi Han, Jiacheng Cen, Liming Wu, Zongzhao Li, Xiangzhe Kong, Rui Jiao, Ziyang Yu, Tingyang Xu, Fandi Wu, Zihe Wang, et al. A survey of geometric graph neural networks: Data structures, models and applications. *Frontiers of Computer Science*, 19(11):1911375, 2025.

[Hu *et al.*, 2023] Zhiwei Hu, Víctor Gutiérrez-Basulto, Zhiliang Xiang, Ru Li, and Jeff Z Pan. Hyperformer: Enhancing entity and relation interaction for hyper-relational knowledge graph completion. In *Proceedings of the 32nd ACM international conference on information and knowledge management*, pages 803–812, 2023.

[Lee and Whang, 2025] Jaejun Lee and Joyce Jiyoung Whang. Structure is all you need: Structural representation learning on hyper-relational knowledge graphs. In *Forty-second International Conference on Machine Learning*, 2025.

[Li *et al.*, 2024] Zhao Li, Chenxu Wang, Xin Wang, Zirui Chen, and Jianxin Li. Hje: Joint convolutional representation learning for knowledge hypergraph completion. *IEEE Transactions on Knowledge and Data Engineering*, 36(8):3879–3892, 2024.

[Li *et al.*, 2025] Zhao Li, Xin Wang, Jun Zhao, Wenbin Guo, and Jianxin Li. Hycube: Efficient knowledge hypergraph 3d circular convolutional embedding. *IEEE Trans. on Knowl. and Data Eng.*, 37(4):1902–1914, 2025.

[Liu *et al.*, 2021] Yu Liu, Quanming Yao, and Yong Li. Role-aware modeling for n-ary relational knowledge bases. *Proceedings of the Web Conference*, pages 2660–2671, 2021.

[Luo *et al.*, 2023] Haoran Luo, E Haihong, Yuhao Yang, Yikai Guo, Mingzhi Sun, Tianyu Yao, Zichen Tang, Kaiyang Wan, Meina Song, and Wei Lin. Hahe: Hierarchical attention for hyper-relational knowledge graphs in global and local level. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, pages 8095–8107, 2023.

[Nickel and Kiela, 2017] Maximilian Nickel and Douwe Kiela. Poincaré embeddings for learning hierarchical representations. In *Proceedings of the International Conference on Neural Information Processing Systems*, pages 6341–6350, 2017.

[Rosso *et al.*, 2020] Paolo Rosso, Dingqi Yang, and Philippe Cudré-Mauroux. Beyond triplets: hyper-relational knowledge graph embedding for link prediction. In *Proceedings of the web conference*, pages 1885–1896, 2020.

[Shomer *et al.*, 2023] Harry Shomer, Wei Jin, Juanhui Li, Yao Ma, and Hui Liu. Learning representations for hyper-relational knowledge graphs. In *Proceedings of the International Conference on Advances in Social Networks Analysis and Mining*, pages 253–257, 2023.

[Sun *et al.*, 2019] Zhiqing Sun, Zhi-Hong Deng, Jian-Yun Nie, and Jian Tang. Rotate: Knowledge graph embedding by relational rotation in complex space. In *International Conference on Learning Representations*, 2019.

[Vaswani *et al.*, 2017] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Proceedings of the International Conference on Neural Information Processing Systems*, page 6000–6010, 2017.

[Wang *et al.*, 2021] Quan Wang, Haifeng Wang, Yajuan Lyu, and Yong Zhu. Link prediction on n-ary relational facts: A graph-based approach. In *Findings of the Association for Computational Linguistics*, pages 396–407, 2021.

[Wen *et al.*, 2016] Jianfeng Wen, Jianxin Li, Yongyi Mao, Shini Chen, and Richong Zhang. On the representation and embedding of knowledge bases beyond binary relations. In *Proceedings of the International Joint Conference on Artificial Intelligence*, pages 1300–1307, 2016.

[Xiong *et al.*, 2023] Bo Xiong, Mojtaba Nayyeri, Shirui Pan, and Steffen Staab. Shrinking embeddings for hyper-relational knowledge graphs. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, pages 13306–13320, 2023.

[Yan *et al.*, 2022a] Shiyao Yan, Zequn Zhang, Xian Sun, Guangluan Xu, Li Jin, and Shuchao Li. Hyper2: Hyperbolic embedding for hyper-relational link prediction. *Neurocomputing*, 492:440–451, 2022.

[Yan *et al.*, 2022b] Shiyao Yan, Zequn Zhang, Xian Sun, Guangluan Xu, Shuchao Li, Qing Liu, Nayu Liu, and Shensi Wang. Polygone: Modeling n-ary relational data as gyro-polygons in hyperbolic space. *Proceedings of the AAAI Conference on Artificial Intelligence*, 36(4):4308–4317, 2022.

[Yu and Yang, 2021] Donghan Yu and Yiming Yang. Improving hyper-relational knowledge graph completion. *ArXiv*, abs/2104.08167, 2021.

[Zhang *et al.*, 2018] Richong Zhang, Junpeng Li, Jiajie Mei, and Yongyi Mao. Scalable instance reconstruction in knowledge bases via relatedness affiliated embedding. In *Proceedings of the World Wide Web Conference*, page 1185–1194, 2018.