Self-Supervised Hyperboloid Representations from Logical Queries over Knowledge Graphs

Nurendra Choudhary¹, Nikhil Rao², Sumeet Katariya², Karthik Subbian², Chandan K. Reddy¹,²
¹Department of Computer Science, Virginia Tech, Arlington, VA, USA
²Amazon, Palo Alto, CA, USA
nurendra@vt.edu,[nikhilsr, katsumee, ksubbian]@amazon.com, reddy@cs.vt.edu

ABSTRACT
Knowledge Graphs (KGs) are ubiquitous structures for information storage in several real-world applications such as web search, e-commerce, social networks, and biology. Querying KGs remains a foundational and challenging problem due to their size and complexity. Promising approaches to tackle this problem include embedding the KG units (e.g., entities and relations) in a Euclidean space such that the query embedding contains the information relevant to its results. These approaches, however, fail to capture the hierarchical nature and semantic information of the entities present in the graph. Additionally, most of these approaches only utilize multihop queries (that can be modeled by simple translation operations) to learn embeddings and ignore more complex operations such as intersection, and union of simpler queries. To tackle such complex operations, in this paper, we formulate KG representation learning as a self-supervised logical query reasoning problem that utilizes translation, intersection and union queries over KGs. We propose Hyperboloid Embeddings (HypE), a novel self-supervised dynamic reasoning framework, that utilizes positive first-order existential queries on a KG to learn representations of its entities and relations as hyperboloids in a Poincaré ball. HypE models the positive first-order queries as geometrical translation, intersection, and union. For the problem of KG reasoning in real-world datasets, the proposed HypE model significantly outperforms the state-of-the-art results. We also apply HypE to an anomaly detection task on a popular e-commerce website product taxonomy and demonstrate significant performance improvements compared to existing baseline methods. Finally, we also visualize the learned HypE embeddings in a Poincaré ball to clearly interpret and comprehend the representation space.

CCS CONCEPTS
• Computing methodologies → Reasoning about belief and knowledge; Knowledge representation and reasoning: Representation of mathematical objects.

KEYWORDS
Representation learning, knowledge graphs, hyperbolic space, reasoning queries

This paper is published under the Creative Commons Attribution 4.0 International (CC-BY 4.0) license. Authors reserve their rights to disseminate the work on their personal and corporate Web sites with the appropriate attribution.

1 INTRODUCTION
Knowledge Graphs (KGs) organize information as a set of entities connected by relations. Positive first-order existential (PFOE) queries such as translation, intersection, and union over these entities aid in effective information extraction from massive data (see Figure 1 for an example PFOE query). Efficient handling of such queries on KGs is of vital importance in a range of real-world applications including search engines, dialogue systems, and recommendation models. However, the large size of KGs and high degrees of the nodes therein makes traversal for querying a computationally challenging or, in some cases, even an impossible task [43]. One way to resolve this issue is to learn representations for the KG units (entities and relations) in a latent (generally Euclidean) space such that algebraic or logical operations can be applied to extract relevant entities. Robust representation learning of KG units has several real-world applications including KG information extraction [20], entity classification [46], and anomaly detection [22].

earlier approaches to representation learning in KGs model entities and relations as vectors in the Euclidean space [7, 31, 48]. This is suboptimal due to the constant size of a point’s answer space which does not capture the variations induced by different queries. Specifically, broad queries (Nike) should intuitively cover a larger region of the answer space compared to specific queries (Nike running shoes for men). In the recently proposed Query2Box model [35], the authors demonstrated the effectiveness of complex geometries (such as hyper-rectangles) with varying offsets that control the size of an answer space according to a query’s complexity.

Figure 1: An example of PFOE querying in the E-commerce product network. The product space of Adidas and Nike intersects with Footwear to narrow the search space. A union over these spaces yields our final set of entity results.
However, such architectures lack the ability to capture hierarchical information that is prevalent in many KGs. Furthermore, previous representation learning methods in heterogeneous graphs (or KGs) [16, 25, 26, 51] solely focus on one-hop or multi-hop reasoning over relations. Such frameworks enable static and optimized computational graphs, but lead to poor retrieval from complex intersection and union queries. Dynamic computational graphs, which are able to modify their network architecture with a switch mechanism (discussed in Section 3.5) can significantly alleviate this problem.

Although Euclidean spaces have proven to be effective for representation learning in various domains [3], several hierarchical datasets (including graph data) in the fields of network sciences and E-commerce taxonomies demonstrate a latent non-Euclidean anatomy [8]. The introduction of hyperbolic algebraic operations [17] have led to the proliferation of hyperbolic neural networks such as Hyperbolic-GCN (H-GCN) [12] and Hyperbolic Attention (HAT) networks [19]. These frameworks leverage the hyperbolic anatomy of hierarchical datasets and show a significant performance boost compared to their Euclidean counterparts. To the best of our knowledge, there is no existing work that (i) utilizes dynamic computational graphs on the hyperbolic space, (ii) applies complex hyperbolic geometries such as hyperboloids for representation learning. Additionally, the static computational graphs of H-GCN and HAT limit their learning capability to a single problem, generally, multi-hop (translation) reasoning. This severely limits their applicability to representation learning on KGs since translations can only utilize single entities. More complex intersections and unions not only use more entities, but are also more representative of real-world KG queries. While solving union and intersection queries is more challenging, they enable better representation learning [4]. Traversing over the entities in KGs facilitates an intuitive way of constructing a query-reasoning proxy task (refer Section 4.3) that enables representation learning of entities and relations. These representations, in a self-supervised framework, can further provide enriched features in downstream tasks with smaller annotated datasets (such as anomaly detection), thus alleviating the issue of data scarcity.

Motivated by the effectiveness of self-supervised learning and the need for non-Euclidean geometries in KGs, we formulate KG representation learning as a self-supervised query reasoning problem. We introduce Hyperboloid Embeddings (HypE), a self-supervised dynamic representation learning framework that utilizes PFOE queries to learn hyperboloid representations of KG units in a (non-Euclidean) Poincaré hyperball. Hyperboloids, unlike vectors in hyperbolic spaces, allow us to use dynamic sizes for KG representations. For e.g., in Figure 2, we can notice that different entities contain different number of children, and, thus learning a static vector representation is suboptimal. Hyperboloids learn an additional spatial parameter, limit (described in Section 3.4), that can model the varying entity sizes. Moreover, the dynamic nature of its computational graphs allows HypE to utilize different network layers to learn different types of operations, namely, translation, intersection, and union; and process varying number of input units depending on the learning operation. Our empirical studies include learning representations from large-scale KGs in e-commerce, web pages (DBPedia), and other widely used Knowledge bases (such as Freebase and NELL995); and evaluating the representations on the downstream task of anomaly detection. The major contributions of this paper are:

1. Formulate the KG representation learning problem as a self-supervised query reasoning problem to leverage PFOE queries.
2. Introduce Hyperboloid Embeddings (HypE), a self-supervised dynamic representation learning framework that learns hyperboloid representations of KG units in a Poincaré hyperball. This is motivated by the need for non-Euclidean geometries.
3. Perform an extensive set of empirical studies across diverse real-world datasets to evaluate the performance of HypE against several state-of-the-art baseline methods on the downstream task of Anomaly Detection.
4. Visualize the HypE embeddings to clearly interpret and comprehend the representation space.

The rest of the paper is organized as follows: Section 2 describes the related background. Section 3 formulates the representation learning problem, explains the non-Euclidean algebraic operations, and the proposed HypE model. In Section 4, we describe the real-world datasets, state-of-the-art baselines and performance metrics used to evaluate the HypE model. We demonstrate the performance results along with the visualization of HypE’s representations. Finally, Section 5 concludes the paper.

2 RELATED WORK

In this section, we review different geometries utilized for learning representations and earlier works that are adopted for reasoning over Knowledge Graphs.

**Representation Geometries**: Previous approaches to representation learning, in the context of KG, aim to learn latent representations for entities and relations. Translational frameworks [7, 32, 48] model relations using translation between entity pairs. This limits the models to only handle translation-based queries. Graph Query Embedding (GQE) [20] overcame this limitation and provided a technique for leveraging intersection queries as deep sets [49] over different queries. Furthermore, Box Lattices [41], EMQL [37] and Query2Box [35] proved the effectiveness of more complex geometries (hyper-rectangles) for queries. Word2Gauss [42] is a popular

![Figure 2: Static vs Dynamic representation. (a) hyperbolic vectors have lower precision due to static thresholds over a center, depicted by the dotted circle. (b) Dotted hyperboloids encapsulate all its child entities because of dynamic sizes. The blue and orange circles are intermediate and leaf nodes.](image)
NLP technique that learns Gaussian embeddings for words. DNGE [39] utilizes a dynamic network to learn Gaussian embeddings for entities in a graph. These Gaussian representations cannot be intuitively extended to KGs because they are not closed under more complex PFOE queries (intersection or union of Gaussians does not yield a Gaussian). Furthermore, they rely on properties of the Euclidean space to learn representations, which are proven ineffective at capturing the prevalent hierarchical features of a KG [17].

**Representation Learning on Graphs**: One of the fundamental problems in KG is to aggregate the neighbor information of nodes while learning representations. Node embedding techniques such as Node2Vec [18] and DeepWalk [34] aggregate the neighbors’ features by modeling the node’s dependence on its neighbors. ChebNet [13] uses Chebyshev polynomials and filters node features in the graph Fourier domain. GCN [23] constrains the parameters of ChebNet to alleviate overfitting and shows improved performance. Graph-BERT [51] and MAGNN [16] provide a self-supervised learning model utilizing the tasks of masking and metapath aggregation, respectively. In another line of research, Miller et al. [28] utilizes non-parametric Bayesian frameworks for link prediction on social networks. Zhu [52] further improved the approach with a max-margin framework. KGAT [44] is another popular approach that utilizes attention networks over entities and relations with a TransR [24] loss function to learn representations for user recommendation. These methods rely on relational properties and thus are effective in handling translational problems such as multi-hop reasoning. However, they are ineffective at handling more complex PFOE queries such as intersection and union.

Other popular multi-hop graph networks such as Graph Attention Network (GAT) [40] and Graph Recurrent Network (GRN) [36] have previously shown impressive results in reasoning-based QA tasks. However, hyperbolic flavors of these networks, H-GNN [17], H-GCN [11, 12] and H-GAT [19] argue that hierarchical datasets follow the anatomy of hyperbolic space and show improved performance over their Euclidean counterparts. Nonetheless, these approaches are still limited by the constant answer space that does not consider the varying fluctuations of complex queries.

Self-supervised learning [14, 21, 29, 47] utilizes large unannotated datasets to learn representations that can be fine-tuned to other tasks that have relatively smaller amount of annotated data. Traversing over the entities in KGs facilitates an intuitive way of constructing a query-reasoning proxy task (refer Section 4.3) that enables representation learning of entities and relations. These representations, in turn, are employed in downstream tasks with scarce datasets such as anomaly detection.

The proposed HypE model utilizes a self-supervised learning framework that leverages both simple and complex PFOE queries to learn hyperboloid (with varying limits) representations of KG units in a Poincaré ball to efficiently capture hierarchical information.

### 3 Proposed Framework

In this section, we first provide the standard method of querying knowledge graphs. Then, we set up the problem and describe the details of our model that learns representations of entities and relations from reasoning queries over Knowledge Graphs (KG).

#### 3.1 Querying Knowledge Graphs

KGs contain two primary units, namely, entities and relations. Entities are the basic information units that connect to each other by relation units. Heterogeneous graphs [45, 50] can be considered as a special case of KGs where the relations serve as hierarchical connections with no inherent information. PFOE queries of translation (t), intersection (∩) and union (∪) serve as the primary means of querying these KGs. Translation queries utilize an entity e and a relation r to retrieve all entities that are connected to e through r. An equivalent example of translation query for heterogeneous graphs is to retrieve all children of a node e connected by a certain edge type r. Intersection and Union operate over multiple entities E = {e₁, ..., eₙ} and correspondingly retrieve the set of all entities that are connected to all e₁ ∈ E and any eᵢ ∈ E. For heterogeneous graphs, the equivalent is to retrieve nodes connected to all nodes e₁ ∈ E and any eᵢ ∈ E. An example of PFOE querying is given in Figure 3. The widely studied problem of multi-hop traversal [16] is a more specific case of translation queries, where multiple queries are chained in a series.

![Figure 3: A simple first-order query over a Product Graph.](image)

The query Nike and Adidas footwear can be expressed as the union of the Nike and Adidas nodes intersected with the node corresponding to footwear in the product catalog.

#### 3.2 Problem Setup

We denote KG = (E, R) as a set of entities eᵢ ∈ E and relations rᵢⱼ ∈ R : eᵢ → eⱼ as Boolean functions that indicate whether a directed relation rᵢⱼ holds between eᵢ and eⱼ. Intersection (∩) and Union (∪) are positive first-order existential (PFOE) operations defined on a set of queries q₁, q₂, ..., qₙ ∈ Q:

\[ q₁([Q]) = V₁ \subseteq E \ \exists \ e₁, e₂, ..., eₙ : q₁ \cap q₂ \cap qₙ \cap qₙ \in Q \]

\[ q₁([Q]) = V₂ \subseteq E \ \exists \ e₁, e₂, ..., eₙ : q₁ \cup q₂ \cup qₙ \cup qₙ \in Q \]

where (q₁, V₁) and (q₂, V₂) is the query space and resultant entity set after the intersection and union operations over query set Q, respectively. Given a dataset of logical queries over a KG, the goal, here, is to learn hyperboloid (in a Poincaré ball) representations for its entities eᵢ ∈ ℝᵈ and relations rᵢⱼ ∈ ℝᵈ, where d is a hyper-parameter that defines the dimension of the Poincaré ball.

#### 3.3 Manifold Transformation Layer

Hierarchical structures intuitively demonstrate the latent characteristics of a hyperbolic space [17]. Thus, we utilize the Poincaré ball [9] to model our representations.

\[ eᵢ, q₁([q₁, q₂]) \text{ and } q₁([q₁, q₂]) = q₂ \text{ and } q₁([q₁, q₂]) = qₙ \text{ and } q₁([q₁, q₂]) = qₙ \]
3.3.1 Transformation from Euclidean Space. The transformation from Euclidean to hyperbolic space \((\mathbb{H}^n, g^E)\), given in [17], is defined by the manifold \(\mathbb{H}^n = \{ \mathbf{x} \in \mathbb{R}^n : \|\mathbf{x}\| < 1 \}\) with the Riemannian metric \(g^E\), where:

\[
g^E_{\mathbf{x}} = \lambda^2 \mathbf{x} \otimes \mathbf{x} = \frac{2}{1 - \|\mathbf{x}\|^2} \quad (1)
\]

\(g^E\) being the Euclidean identity metric tensor, and \(\|\mathbf{x}\|\) is the Euclidean norm of \(\mathbf{x}\). \(\lambda\) is the conformal factor between the Euclidean and hyperbolic metric tensor set to a conventional curvature of \(-1\). Eq. (1) allows us to convert a Euclidean metric to hyperbolic. Thus, the distance between points \(\mathbf{x}, \mathbf{y} \in \mathbb{H}^n\) is derived as:

\[
d_{\mathbb{H}}(\mathbf{x}, \mathbf{y}) = \cosh^{-1}\left( 1 + 2 \frac{\|\mathbf{x} - \mathbf{y}\|^2}{(1 - \|\mathbf{x}\|^2)(1 - \|\mathbf{y}\|^2)} \right) \quad (2)
\]

3.3.2 Gyrovector Spaces. Algebraic operations such as addition and scalar product which are straightforward in the Euclidean space cannot be directly applied in hyperbolic space. Gyrovector spaces allow for the formalization of these operations in hyperbolic space. Ganea et al. [17] provide the gyrovector operations relevant to training neural networks. The gyrovector operations for Poincaré ball of radius \(c\) are Möbius addition \((\odot_c)\), Möbius subtraction \((\ominus_c)\), exponential map \((\exp_c)\), logarithmic map \((\log_c)\) and Möbius scalar product \((\odot_c)\).

\[x \odot_c y := \frac{1 + 2c(x, y) + c\|y\|^2}{1 + 2c(x, y) + c^2\|x||y\|^2} x + \frac{1 - c\|x\|^2}{1 + 2c(x, y) + c^2\|x||y\|^2} y\]

\[x \odot_c y := x \odot_c y - y\]

\[\exp_c(v) := x \odot_c \left( \tanh \left( \frac{\sqrt{c^2\|v\|^2}}{2} \right) \frac{v}{\sqrt{c^2\|v\|^2}} \right)\]

\[\log_c(y) := \frac{2}{\sqrt{c^2\|y\|^2}} \tanh^{-1} \left( \frac{\sqrt{c^2\|y - x \odot_c y\|}}{\sqrt{c^2\|y\|^2}} \right) x \odot_c y\]

\[r \odot_c x := \exp_c(r \log_c(x)), \forall r \in \mathbb{R}, x \in \mathbb{H}^n\]

Here, \(\odot\) denotes assignment operation for Möbius operations. Also, the norm of \(x, y\) can subsume the scaling factor \(c\). Hence, in HypE, training can be done with a constant \(c\) or trainable \(c\). We empirically validate this assumption in our experiments (Section 4.4). Figure 4 shows an example of the manifold transformation from Euclidean space to a Poincaré ball of unit radius. HypE extends the operations to handle complex geometries, explained in Section 3.4.

3.4 Dynamic Reasoning Framework : HypE

We aim to learn hyperboloid (made of two parallel pairs of arc-aligned horocycles) embeddings for all the entities and relations in the KG. An arc-aligned horocycle (Figure 5(a)) is a partial circle that is parallel to the diameter of a Poincaré ball and orthogonally intersects its boundaries at two points. A hyperboloid embedding (see Figure 5(b)) \(e = (\text{cen}(e), \text{lim}(e)) \in \mathbb{R}^{2d}\) is characterized by:

\[H_e \equiv \{ v \in \mathbb{R}^d : \text{cen}(e) \odot_c \text{lim}(e) \leq v \leq \text{cen}(e) \odot_c \text{lim}(e) \}\]

where \(\equiv\) denotes strict equivalence and \(\leq\) is element-wise inequality and \(\text{cen}(t), \text{lim}(t) \in \mathbb{R}^d\) are center of the hyperboloid and positive border limit \((\text{lim}(t) \geq 0)\) of the enclosing horocycle from the center, respectively. The overview of the architecture is given in Figure 6. From the KG, we derive the following types of directed edge relations to build our dynamic computational graph for learning embeddings.
Intersection (t): Each relation \( r \in R \) is equipped with a relation embedding \( r = \text{Hyperboloid}_r \in \mathbb{R}^{2d} \). Given an entity embedding \( e \in E \), we model its translation \( o_t \) and distance from the result entities \( o_t \in V_t \subseteq E \) as follows:

\[
o_t = e \oplus_c r, \quad d_t = d_{\text{hyp}}(o_t, o_t)
\]

This provides us with the translated hyperboloid with a new center and larger limit (\( \lim(r) \geq 0 \)). A sample operation is illustrated in Figure 7(a).

Intersection (?): We model the intersection of a set of hyperboloid embeddings \( Q_\cap = \{ e_1, e_2, e_3, ..., e_n \} \) as \( o_\cap \) and entity distance from the result entities \( o_\cap \in V_\cap \subseteq E \) as \( d_\cap \) where:

\[
o_\cap = (\text{cen}(Q_\cap), \lim(Q_\cap)) \quad (5)
\]

\[
\text{cen}(Q_\cap) = \sum_i a_i \odot_c \text{cen}(e_i) \quad (6)
\]

\[
\lim(Q_\cap) = \min((\lim(e_1), ..., \lim(e_n))) \odot_c \sigma(DS(\{e_1, ..., e_n\})) \quad (7)
\]

\[
d_\cap(o_\cap) = d_{\text{hyp}}(o_\cap, o_\cap) \quad (8)
\]

where \( \odot_c \) is the Möbius scalar product, \( f(.) : \mathbb{R}^{2d} \rightarrow \mathbb{R}^d \) is the multilayer perceptron (MLP), \( \sigma(.) \) is the sigmoid function and \( DS(.) \) is the permutation invariant deep architecture, namely, DeepSets [49]. \( \text{Min}(.) \) and \( \exp(.) \) are element-wise minimum and exponential functions. The new center and limit are calculated by an attention layer [1] over the hyperboloid centers and DeepSets for shrinking the limits, respectively. Figure 7(b) depicts a sample intersection.

Union (∪): Unlike intersection, union operations are not closed under hyperboloids (union of hyperboloids is not a hyperboloid). Hence, the distance of entities from the union query space \( d_\cup \) is defined as the minimum distance from any hyperboloid in the union. For a set of hyperboloid embeddings \( Q_\cup = \{ e_1, e_2, e_3, ..., e_n \} \), union space is given by \( o_\cup \) and distance from result entities \( o_\cup \in V_\cup \subseteq E \) by \( d_\cup \), where

\[
o_\cup = Q_\cup \quad (9)
\]

\[
d_\cup(o_\cup) = \min \left( d_{\text{hyp}}(o_\cup, e_i) \forall e_i \in o_\cup \right) \quad (10)
\]

Note that, since union is not closed under hyperboloids it cannot be applied before the other operations. We circumvent this problem by utilizing Disjunctive Normal Form (DNF) transformation [35] on our logical queries. This allows us to push all the union operations to the end of our computational graph, thus maintaining validity for all PFOE queries. An outline of HypE’s training procedure is given in Algorithm 1.

### 3.5 Implementation Details

We implemented HypE in Pytorch [33] on two Nvidia V100 GPUs with 16 GB VRAM. For gradient descent, the model is trained using Reimannian Adam optimizer [2] with an initial learning rate of 0.0001 and standard \( \beta \) values of 0.9 and 0.999. We utilize ReLU [30] as the activation function. Also, we randomly selected 128 negative samples per positive sample in the training phase to learn better discriminative features. For our empirical studies, we learned hyperboloid embeddings of \( 2 \times 400 \) dimensions (\( d=400 \)). Due to the conditionality (i.e., if conditions in Algorithm 1) in our computational graph, we employ a switch mechanism between the network layers [15, 27]. The switch mechanism receives an operator signal that defines the operation and accordingly connects/disconnects a layer from the framework. A disconnected switch blocks back-propagation of weight updates to the disconnected layers. This enables a concurrent use of all PFOE queries to update the entity and relation embeddings. For an input query \( Q \) and resultant entities \( V \), Algorithm 1 provides the pseudocode of our overall framework to learn representations of entities \( E \) and relation \( R \). The algorithm describes the three main operations, namely, translation (lines 4-7), intersection (lines 8-11), and union (lines 12-15).

### 4 EXPERIMENTAL SETUP

This section describes the experimental setup that analyzes the performance of HypE on various problems. We aim to study the following research questions:

---

2Implementation code: https://github.com/amazon-research/hyperbolic-embeddings
Algorithm 1: HypE algorithm

**Input:** Training data $D_t, D_v, D_o$, which are set of all (query $(Q)$, result $(V)$) for translation, intersection, and union, respectively;

**Output:** Entity $E$ and Relation $R$ hyperboloids;

```plaintext
1 Randomly initialize $e \in E$ and $r \in R (e, r \in \mathbb{R}^d)$;
2 for number of epochs until convergence do
  3 \hspace{0.5em} $l = 0$; \hspace{0.5em} # Initialize loss
  4 \hspace{0.5em} for \hspace{0.5em} $\{(e, r, V_t) \in D_t\}$ do
  5 \hspace{1em} $o_t = e \oplus r$, from Eq. (4)
  6 \hspace{1em} # Update loss for translation queries
  7 \hspace{1em} $l = l + \sum_{v_t \in V_t} d_{hyp}(u_t, o_t)$
  8 \hspace{0.5em} end
  9 \hspace{0.5em} for \hspace{0.5em} $\{(Q_v, V_r) \in D_v\}$ do
 10 \hspace{1em} $o_v = (cen(Q_v), lim(Q_v))$, from Eq. (5)
 11 \hspace{1em} # Update loss for intersection queries
 12 \hspace{1em} $l = l + \sum_{v_t \in V_t} d_{hyp}(u_t, o_v)$
 13 \hspace{0.5em} end
 14 \hspace{0.5em} for \hspace{0.5em} $\{(Q_v, O_u) \in D_o\}$ do
 15 \hspace{1em} $o_u = Q_v$, from Eq. (9)
 16 \hspace{1em} # Update loss for union queries
 17 \hspace{1em} $l = l + \sum_{v_t \in V_t} \min\{d_{hyp}(u_t, e_t) \forall e_t \in o_u\}$
 18 \hspace{0.5em} end
 19 \hspace{0.5em} # Update $E$ and $R$ with backpropagation
 20 \hspace{0.5em} $E \leftarrow E - \Delta E$
 21 \hspace{0.5em} $R \leftarrow R - \Delta R$
22 end
23 return $E, R$
```

- **RQ1:** For the task of reasoning over KGs, are hyperboloid embeddings better than the baselines at learning hierarchical relations?
- **RQ2:** What is the contribution of individual components in the HypE model?
- **RQ3:** Do the representations capture relevant data features for the downstream task of anomaly detection?
- **RQ4:** Can hyperboloid embeddings leverage auxiliary semantic information from the entities?
- **RQ5:** Can we comprehend the latent representational space obtained by the proposed HypE model?

### 4.1 Datasets

We perform our experimental study on the following standard KG and hierarchical graph datasets:

1. **FB15k** [7] contains knowledge base relation triples and textual mentions of Freebase entity pairs. This dataset contains a large number of simple test triples that can be obtained by inverting the training triples.

2. **FB15k-237** [38] is a subset of FB15k where all the simple invulnerable relations are removed, so the models can learn and focus on more complex relations.

3. **NELL995** [10] is a KG dataset of relation triples constructed from the 995th iteration of the Never-Ending Language Learning (NELL) system.

4. **DBPedia Hierarchical Taxonomy** [3] is a subset extracted from Wikipedia snapshot that provides multi-level hierarchical taxonomy over 342,782 articles (leaf-nodes).

5. **E-commerce Product Network** [4] is a subsampled product taxonomy from an e-commerce platform.

To be consistent with KG terms, for the hierarchical graph datasets (DBPedia and E-commerce) we consider all the intermediate and leaf nodes as entities and the edges between them as relations. Additionally, we consider two variants for encoding edges. First, all edges are considered identical ($|R| = 1$) and second, where all edges are depth-encoded ($|R| = p$), where $p$ is maximum depth of the hierarchy ($p = 3$ for DBPedia and $p = 4$ for E-commerce dataset).

For cross-validation and evaluation, we split the graph KG into three parts: $KG_{\text{train}}$, $KG_{\text{valid}}$ and $KG_{\text{test}}$ in a 75 : 10 : 15 ratio for our experiments. More details of the datasets are given in Table 1.

| Dataset            | # entities | # relations | # edges  |
|--------------------|------------|-------------|----------|
| FB15k              | 14,951     | 2690        | 273,710  |
| FB15k-237          | 14,505     | 474         | 149,689  |
| NELL995           | 63,361     | 400         | 107,982  |
| DBPedia            | 34,575     | 3           | 240,942  |
| E-commerce         | ~118K      | 4           | ~562K    |

### 4.2 Baselines

We select our baselines based on the following two criteria:

1. The embedding geometries are closed under the intersection and translation operation, e.g., the translation or intersection of arc-aligned hyperboloids results in an arc-aligned hyperboloid.

2. The baseline can be intuitively extended to all PFOE queries over KG. This is necessary to have a fair comparison with HypE that can leverage all PFOE queries.

We adopt the following state-of-the-art baselines based on geometric diversity and our criterion to compare against HypE:

- **Graph Query Embedding (GQE)** [20] embeds entities and relations as a vector embedding in the Euclidean space.

- **Knowledge Graph Attention Network (KGAT)** [44] embeds entities and relations as a vector embedding in the Euclidean space utilizing attention networks over entities and relations with a TransR loss [24].

- **Hyperbolic Query Embeddings (HQE)** [17] utilizes manifold transformations (refer to Section 3.3) to represent entities and relations as a vector embedding in hyperbolic space.

- **Query2Box (Q2B)** [35] embeds entities and relations as axis-aligned hyper-rectangle or box embeddings in Euclidean space.

Some of the other possible baselines [7, 32, 48], solely, focus on the multi-hop (or translation) problem. They could not be naturally extended to other PFOE queries. Additionally, other geometric variants such as circular and Gaussian embeddings [39, 42] are not closed under intersection (intersection of Gaussians is not a Gaussian).

---

1https://www.kaggle.com/danofer/dbpedia-classes
2Proprietary dataset
Table 2: Performance comparison of HypE (ours) against the baselines to study the efficacy of the Query-Search space. The columns present different query structures and averages over them. The final row presents the Average Relative Improvement (%) of HypE compared to Query2Box over all datasets. E-Vector and H-Vector are vectors in Euclidean and hyperbolic space, respectively. Best results for each dataset are shown in bold.

| Metrics | Dataset | Model                | Hits@3 | Mean Reciprocal Rank |
|---------|---------|----------------------|--------|----------------------|
|         |         | H         | T         | 2t       | 3t       | 2t       | 3t       | 2t       | 3t       | 2t       | 3t       | 2t       | 3t       | 2t       | 3t       | 2t       | 3t       | 2t       | 3t       | 2t       | 3t       | 2t       | 3t       |
|         |         | Avg      |           |           |           |           |           |           |           |           |           |           |           |           |           |           |           |           |           |           |           |           |           |           |           |           |
| HypE (Hyperboloid) | FB15k | KGAT (E-Vector) | 711 | 379 | 276 | 553 | 667 | 492 | 181 | 354 | 302 | 435 | 365 | 352 | 243 | 471 | 573 | 393 | 167 | 311 | 270 | 372 | 312 | 301 | 410 |
| HypE (Hyperboloid) | FB15k | KGAT (E-Vector) | 683 | 365 | 265 | 451 | 589 | 438 | 135 | 283 | 290 | 389 | 543 | 339 | 233 | 384 | 506 | 350 | 125 | 249 | 259 | 332 | 312 | 301 | 410 |
| HypE (Hyperboloid) | FB15k | KGAT (E-Vector) | 786 | 413 | 303 | 590 | 710 | 608 | 211 | 397 | 308 | 483 | 654 | 373 | 274 | 488 | 602 | 408 | 194 | 339 | 301 | 410 | 312 | 301 | 410 |
| HypE (Hyperboloid) | FB15k | KGAT (E-Vector) | 809 | 486 | 365 | 598 | 728 | 610 | 206 | 406 | 410 | 513 | 673 | 439 | 330 | 495 | 617 | 470 | 189 | 347 | 374 | 437 | 312 | 301 | 410 |
| HypE (Hyperboloid) | NELL-995 | KGAT (E-Vector) | 436 | 227 | 167 | 293 | 422 | 202 | 069 | 135 | 174 | 236 | 373 | 205 | 165 | 280 | 384 | 179 | 068 | 130 | 170 | 217 | 312 | 301 | 410 |
| HypE (Hyperboloid) | NELL-995 | KGAT (E-Vector) | 440 | 231 | 171 | 265 | 387 | 195 | 083 | 162 | 183 | 235 | 376 | 209 | 169 | 283 | 352 | 173 | 082 | 156 | 179 | 235 | 312 | 301 | 410 |
| HypE (Hyperboloid) | NELL-995 | KGAT (E-Vector) | 467 | 240 | 186 | 324 | 453 | 239 | 050 | 108 | 193 | 293 | 486 | 220 | 171 | 305 | 427 | 214 | 083 | 156 | 179 | 235 | 312 | 301 | 410 |

4.3 RQ1: Efficacy of the Query-Search space

To analyze the efficacy of the query space obtained from the HypE model, we compare it against the state-of-the-art baselines on the following reasoning query structures:

1. **Single operator queries** include multi-level translation (1t, 2t, and 3t) multi-entity intersection (2∩, 3∩) and multi-entity union queries (2∪, 3∪). 1t, 2t, and 3t denote translation, 1 and 2 consecutive relations, respectively. 2∩ and 2∪ stand for intersection and union over two entities, respectively. 3∩ and 3∪ represent intersection and union over three entities.

2. **Compound queries** contain multiple operators chained in series to get the final result. Our experiments analyze ∩t (intersection-translation), ∩t (translation-intersection), and ∪t (union-translation). The above queries are illustrated in Figure 8.

![Figure 8: Logical query structures designed to compare HypE against baselines. The blue, red, and green units denote translation, intersection, and union operations, respectively.](image-url)
Table 4: Ablation study results. The first column presents the model variants compared against the final HypE model. Avg-1t and Avg-1,2,3t variants only utilize average center aggregation because other aggregation variants only apply when intersections are involved. HypE-TC presents the HypE variant with trainable curvature. The metrics reported in the table are averaged across evaluation on all the datasets. Finer evaluation results are provided in the Appendix B. Best results are shown in bold.

| Model Variants | Hits@3 | Mean Reciprocal Rank |
|----------------|--------|-----------------------|
|                | 1t     | 2t     | 3t     | 1t     | 2t     | 3t     | 1t     | 2t     | 3t     | Avg    |
| HypE-Avg-1t    | 651    | 535    | 337    | 538    | 459    | 350    | 385    | 311    | 431    | 399   |
| HypE-Avg-1,2,3t| 650    | 533    | 331    | 537    | 458    | 350    | 383    | 309    | 429    | 399   |
| HypE-Dict      | 650    | 533    | 331    | 537    | 458    | 350    | 383    | 309    | 429    | 399   |
| HypE-SI        | 650    | 533    | 331    | 537    | 458    | 350    | 383    | 309    | 429    | 399   |
| HypE-SC        | 650    | 533    | 331    | 537    | 458    | 350    | 383    | 309    | 429    | 399   |
| HypE (Final)   | 656    | 543    | 375    | 490    | 564    | 488    | 177    | 374    | 243    | 405   |

Reciprocal Rank (MRR). The metrics are given by:

\[
\text{HITS@K}(q_{test}) = \frac{1}{K} \sum_{k=1}^{K} f(e_{ok})
\]

\[
MRR(q_{test}) = \frac{1}{n} \sum_{i=1}^{n} \frac{f(e_{oi})}{i, \text{if } e_{oi} \in E_{result} \text{ otherwise}}
\]

From the results in Table 2, we observe that, on average, HypE outperforms the current baselines in translation, single, and compound operator queries by 15%-67%, 4%-13%, and 7%-28%, respectively. The performance improvement linearly increases with higher query depth (1t < 2t < 3t). Furthermore, we notice that unique depth encoding ([R] = d) outperforms identical depth encoding ([R] = 1) by 23% in DBPedia. Because of our subsampling strategy, the E-commerce Product Network is disjoint (i.e., there are several intermediate nodes that do not share children). Hence, the number of intersection queries are extremely low and insufficient for training HypE or baselines. However, we can still train the translation queries, and the results are shown in Table 3, where we report relative performance improvements with respect to the HQE baseline. For the sake of completeness, results on intersection and union queries are given in the Appendix C.

4.4 RQ2: Ablation Study

In this section, we empirically analyze the importance of different layers adopted in the HypE model. For this, we experiment with different variations of the center aggregation layer: Average (HypE-Avg), Attention (HypE) (refer to Eq. (6)) and Deepsets (HypE-DS) (refer to Eq. (7)). Furthermore, we test the exclusion of intersection and unions to comprehend their importance in the representation learning process. We adopt two variants of HypE: one trained on only 1t queries (HypE-Avg-1t) and the other trained on all translation queries (HypE-Avg-1,2,3t). Table 4 presents the performance metrics of different variants on the query processing task, averaged across all the datasets including the E-commerce dataset. The results are provided in Table 4.

Firstly, we observe that the exclusion of intersection and union queries results in a significant performance decrease by 25% (Avg-1,2,3t vs HypE). Furthermore, removing deeper queries such as 2t and 3t, also results in an additional decrease by 17% (Avg-1t vs Ag-1,2,3t). The tests on different aggregation layers prove that Attention is better than average and Deepsets by 23.5% and 14.5%, respectively. Additionally, we notice that employing a trainable curvature results in a slight performance improvement of 0.3%. However, given the incremental performance boost but significant increase in the number of parameters (~10K) that the trainable curvature adds to the framework, we ignore this component in the final HypE model.

As explained in Section 3.4, the final HypE model adopts a Poincaré ball manifold with non-trainable curvature, in addition to attention and Deepsets layer for center and limit aggregation, respectively. Additionally, HypE leverages all PFOE queries.

4.5 RQ3: Performance on Anomaly Detection

In this experiment, we utilize the entity and relation representations, trained on the DBPedia Hierarchical Taxonomy and E-commerce Product Network with query processing task, to identify products that might be anomalously categorized. We consider identifying the anomalous children by three levels of parents (i.e., taxonomy levels): P1, P2 and P3. The motivating application is to categorize items that are potentially mis-categorized by sellers into the more relevant (correct) part of the product taxonomy.

Table 5: Results on Misclassified Article Anomaly Detection in DBPedia dataset. Best results are shown in bold and the second best results are underlined. P, R, and F1 represent Precision, Recall, and F-score, respectively.

| Dataset | P-Level | Models | P | R | F1 |
|---------|---------|--------|---|---|----|
| DBPedia | p | HQE | 0.49 | 0.42 | 0.46 |
|         |       | HQE-DS| 0.49 | 0.43 | 0.47 |
|         |       | HQE-SC| 0.52 | 0.51 | 0.52 |
|         |       | HQE-SI| 0.46 | 0.47 | 0.46 |
|         |       | HQE-S | 0.41 | 0.42 | 0.42 |

Table 6: Results on Misclassified Product Anomaly Detection in E-commerce Product Networks. Best results are shown in bold and the second best results are underlined. The improvements are relative to the HQE baseline. P, R, and F1 represent Precision, Recall, and F-score, respectively. *

| Dataset | P-Level | Models | P | R | F1 |
|---------|---------|--------|---|---|----|
| E-commerce | p | HQE | 0.16 | 0.16 | 0.16 |
|         |       | HQE-DS| 0.16 | 0.16 | 0.16 |
|         |       | HQE-SI| 0.20 | 0.20 | 0.20 |
|         |       | HQE-S | 0.20 | 0.20 | 0.20 |

We construct a pseudo-tree, where all parent nodes are infused with 10% of noise randomly sampled anomalous leaf nodes from different parts of the dataset. The goal of the model is to learn representations from this pseudo-tree and identify anomalous leaf nodes.

*Due to the sensitive nature of the proprietary dataset, we only report relative results.
Table 7: Example of Anomalies in the E-commerce dataset. The models predict “MISCAT” and “TRUE” tags for mis-categorized and truly-categorized items, respectively. Correct and Incorrect tags are given in green and red color, respectively. HypE performs better than Query2Box (Q2B) as we consider higher level of parents because hyperbolic space is better at capturing hierarchical features. Also, HypE-SC is able to utilize semantic information to improve prediction.

| Product Title                                                                 | Parent Category | P-level | Q2B  | HypE  | HypE-SC |
|------------------------------------------------------------------------------|-----------------|---------|------|-------|---------|
| ASICS Women’s Gel-Venture 7 Trail Running Shoes, 3.5M, Graphite Grey/Dried Berry | Rompers         | 1       | TRUE | TRUE  | MISCAT |
| inktastic Family Cruise Youth T-Shirt Youth X-Large (18-20) Pacific Blue 35074       | Calvin Klein    | 1       | TRUE | MISCAT | MISCAT |
| Calvin Klein Men’s Cotton Classics Multipack Crew Neck T-Shirts                | Calvin Klein    | 1       | TRUE | MISCAT | TRUE   |
| Epic Threads Big Girls Paint Splatter Distressed Girlfriend Denim Jeans (Dark Wash, 10) | Wool & Blends   | 2       | MISCAT | MISCAT | TRUE   |
| New Balance Women’s Crag V1 Fresh Foam Trail Running Shoe, Black/Magnet/Raincloud | Customer Segment | 2       | TRUE | TRUE  | MISCAT |
| Fifth Harmony Vintage Photo Blue T Shirt (M)                                     | Customer Segment | 2       | MISCAT | TRUE  | TRUE   |
| Billy Bills Playoff Shirt Buffalo T-Shirt Let’s Go Buffalo Tee                  | Brand Stores    | 3       | TRUE | MISCAT | MISCAT |
| Kann Surf Toddler Karlie Florence Beach Sport Z-Piece Swimsuit, Azul Blue, 4T | Brand Stores    | 3       | MISCAT | TRUE  | MISCAT |
| The North Face Infant Glacier ¼ Snap, Mt. Pink, 0-3 Months US Infant            | Specialty Stores | 3      | MISCAT | TRUE  | TRUE   |
| Artisan Outfitters Mens Surfboard Shortboard Batik Cotton Hawaiian Shirt        | Specialty Stores | 3      | MISCAT | TRUE  | TRUE   |
| PUMA Unisex-Kid’s Astro Kick Sneaker, Peacoat-White-teamgold, 3 M US Little Kid | Specialty Stores | 3      | MISCAT | TRUE  | TRUE   |

![Figure 9: Visualization of HypE representations for samples from hierarchical datasets in Poincaré ball. The hyperboloids have been scaled up 10 times for better comprehension. The blue (intermediate nodes) circles are annotated with their entity names and orange circles (leaf nodes) depict articles and products in (a) and (b), respectively.](image-url)

4.6 RQ4: Leveraging Semantic Information

KGs generally also contain additional auxiliary information within the entities. In this section, we test the possibility of leveraging the semantic information in the DBPedia (article titles) and E-commerce (product titles) dataset to improve representations. We study two methods to connect HypE with FastText embeddings [6] of the corresponding titles:

- **Semantic Initiation (SI)** initiates the HypE’s entities with semantic embeddings and learns new HypE-SI embeddings with the query-processing task (given in Section 4.3).
- **Semantic Collaboration (SC)** concatenates the HypE’s pre-trained entity representations with semantic embeddings.

We investigate the performance of these methods on the task of anomaly detection. The results of the experiments are given in Tables 5 and 6. The results demonstrate that HypE-SI shows no significant performance improvement over HypE. That is, a good semantic initialization of the vectors does not result in better representations. This is reasonable, since the semantic embeddings are learnt in the Euclidean space, and several transformations occur between the initialization and final representations. This also means that the learning framework is robust to initialization. We observe a performance improvement of 2% – 8% in case of HypE-SC when compared to HypE. This suggests the ubiquity of HypE since hierarchical representations can be independently augmented with other auxiliary features to solve more complex tasks. From the examples...
given in Table 7, we can observe that HypE-SI is able to leverage semantic information from product title and category name to enrich HypE’s hierarchical information to produce better predictions. The additional semantic information is especially useful for product misclassification. In the absence of semantic information, HypE will merely learn representations based on the noisy graph and will lose discriminative information between outliers and correct nodes.

4.7 RQ5: Visualization of the Poincaré ball

We extract representative examples of different operations from our dataset and visualize them in a 2-dimensional Poincaré ball of unit radius. We employ Principal Component Analysis (PCA) [5] for dimensionality reduction ($\mathbb{R}^{2d} \rightarrow \mathbb{R}^{2}$) of the hyperboloids in Poincaré ball.

Figure 9 depicts the HypE representations in a Poincaré ball manifold. Notice that the size of nodes increases superlinearly from the center towards the circumference, which is analogous to the superlinear increase in the number of nodes from root to the leaves. Thus, HypE is able to learn a better distinction between different hierarchy levels, when compared to the Euclidean distance-based baselines, which conform to a linear increase. Furthermore, we observe that hyperboloid intersections in DBpedia Taxonomy (Figure 9(a)) capture entities with common parents. Also, the disjoint nature of E-commerce Product Networks (Figure 9(b)) is illustrated by disjoint non-intersecting hyperboloids in the latent space. In addition, we can also notice that the learnable limit parameter adjusts the size of hyperboloids to accommodate its varying number of leaf nodes. Thus, the complex geometry of HypE is able to improve its precision over vector baselines that, generally, utilize static thresholds over distance of the resultant entities from query points.

5 CONCLUSION

In this paper, we presented Hyperboloid Embeddings (HypE) model, a novel self-supervised learning framework that utilizes dynamic query-reasoning over KGs as a proxy task to learn representations of entities and relations in a hyperbolic space. We demonstrate the efficacy of a hyperbolic query-search space against state-of-the-art baselines over different datasets. Furthermore, we also show the effectiveness of hyperboloid representations in complex downstream tasks and study methods that can leverage node’s auxiliary information to enrich HypE features. Additionally, we analyze the contribution of HypE’s individual components through an ablation study. Finally, we present our hyperboloid representations in a 2-dimensional Poincaré ball for better comprehensibility.

REFERENCES

[1] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2014. Neural Machine Translation by Jointly Learning to Align and Translate. In International Conference on Learning Representations. http://arxiv.org/abs/1409.0473

[2] Gary Becigneul and Octavian-Eugen Ganea. 2019. Riemannian Adaptive Optimization Methods. In International Conference on Learning Representations. https://openreview.net/forum?id=r1esqi09K7

[3] Yoshua Bengio, Aaron Courville, and Pascal Vincent. 2013. Representation learning: A review and new perspectives. IEEE transactions on pattern analysis and machine intelligence 35, 8 (2013), 1798–1828.

[4] Yoshua Bengio, Aaron Courville, and Pascal Vincent. 2012. Unsupervised Feature Learning and Deep Learning: A Review and New Perspectives. CoRR abs/1206.5538 (2012). arXiv:1206.5538 http://arxiv.org/abs/1206.5538

[5] Christopher M Bishop et al. 1995. Neural networks for pattern recognition. Oxford university press.

[6] Piotr Bojanowski, Edouard Grave, Armand Joulin, and Tomas Mikolov. 2017. Enriching Word Vectors with Subword Information. Transactions of the Association for Computational Linguistics 5 (2017), 135–146. https://doi.org/10.1162/tacl_a_00051

[7] Antoine Bordes, Nicolas Usunier, Alberto Garcia-Duran, Jason Weston, and Oksana Yakhnenko. 2013. Translating Embeddings for Multi-relational Data. In Advances in Neural Information Processing Systems. 26: 2787–2795.

[8] Michael Bronstein, Joan Bruna, Yann LeCun, Arthur Szlam, and Pierre Vandergheynst. 2017. Geometric deep learning: going beyond euclidean data. IEEE Signal Processing Magazine 34, 4 (2017), 18–42.

[9] James W Cannon, William J Floyd, Richard Kenyon, Walter R Parry, et al. 1997. Hyperbolic geometry. Flavors of geometry 31 (1997), 59–115.

[10] Andrew Carlson, Justin Betteridge, Bryan Kisiel, Burr Settles, Estevam Hruschka Jr, and Tom M Mitchell. 2010. Toward an architecture for never-ending language learning. In Aaa. Vol. 5. Atlanta.

[11] 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. Association for Computational Linguistics, Online, 6091–6114. https://doi.org/10.18653/v1/2020.acl-main.617

[12] Ines Chami, Zhihao Ying, Christopher Ré, and Jure Leskovec. 2019. Hyperbolic graph convolutional neural networks. In Advances in neural information processing systems. 4868–4879.

[13] Michael Defferrard, Xavier Bresson, and Pierre Vandergheynst. 2016. Convolutional neural networks on graphs with fast localized spectral filtering. In Advances in neural information processing systems. 3844–3852.

[14] Carl Doersch and Andrew Zisserman. 2017. Multi-task self-supervised visual learning. In Proceedings of the IEEE International Conference on Computer Vision. 2051–2060.

[15] Yuchun Fang, Zhengyan Ma, Zhaoxiang Zhang, Xu-Yao Zhang, and Xiang Bai. 2017. Dynamic Multi-Task Learning with Convolutional Neural Network. In IJCAI 1668–1674. https://doi.org/10.24963/ijcai.2017/231

[16] Xinyu Fu, Jiani Zhang, Ziqiao Meng, and Irwin Kang. 2020. MAGNN: Metapath Aggregated Graph Neural Network for Heterogeneous Graph Embedding. In Proceedings of The Web Conference 2020. 2331–2341.

[17] Octavia Ganea, Gary Bécigneul, and Thomas Hofmann. 2018. Hyperbolic neural networks. In Advances in neural information processing systems. 5345–5355.

[18] Aditya Grover and Jure Leskovec. 2016. node2vec: Scalable feature learning for networks. In Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining. 855–864.

[19] Caglar Gulcehre, Misha Denil, Mateusz Malinowski, Ali Razavi, Razvan Pascanu, Karl Moritz Hermann, Peter Battaglia, Victor Bapst, David Raposo, Adam Santoro, and Nando de Freitas. 2019. Hyperbolic Attention Networks. In International Conference on Learning Representations.

[20] Will Hamilton, Payal Bajaj, Marinka Zitnik, Dan Jurafsky, and Jure Leskovec. 2018. Embedding logical queries on knowledge graphs. In Advances in neural information processing systems. 2026–2037.

[21] Amir Jalaludin, Timor Kadiz, and Andrew Zisserman. 2017. Self-supervised learning for spinal MRIs. In Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support. Springer, 294–302.

[22] B. Jia, C. Dong, Z. Chen, K. Chang, N. Sullivan, and G. Chen. 2018. Pattern Discovery and Anomaly Detection via Knowledge Graph. In 2018 21st International Conference on Information Fusion (FUSION), 2392–2399.

[23] Thomas N Kipf and Max Welling. 2016. Semi-supervised Classification with Graph Convolutional Networks. In International Conference on Learning Representations.

[24] Hailun Lin, Yong Liu, Weiping Wang, Yinliang Yue, and Zheng Lin. 2017. Learning entity and relation embeddings for knowledge resolution. Procedia Computer Science 108 (2017), 345–354.

[25] Xi Victoria Lin, Richard Socher, and Caiming Xiong. 2018. Multi-Hop Knowledge Graph Reasoning with Reward Shaping. In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, Brussels, Belgium, 3243–3253. https://doi.org/10.18653/v1/D18-1362

[26] Weijie Liu, Peng Zhou, Zhe Zhao, Zhiruo Wang, Qi Ju, Haotang Deng, and Ping Wang. 2020. K-BERT: Enabling Language Representation with Knowledge Graph. In AAAI. 2901–2908.

[27] Moshe Looks, Marcello Herreshoff, DeLesley Hutchins, and Peter Norvig. 2017. Deep Learning with Dynamic Computing Graphs. In International Conference on Learning Representations.

[28] Kurt Miller, Michael Jordan, and Thomas L Griffiths. 2009. Nonparametric latent feature models for link prediction. In Advances in neural information processing systems. 1276–1284.

[29] Arsha Nagrani, Joon Son Chung, Samuel Albanie, and Andrew Zisserman. 2020. Disentangled speech embeddings using cross-modal self-supervision. In ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 6829–6833.

[30] Vinod Nair and Geoffrey E. Hinton. 2010. Rectified Linear Units Improve Restricted Boltzmann Machines. In Proceedings of the 27th International Conference
A HYPERBOLIC VS EUCLIDEAN DISTANCES

To better analyze the impact of adopting hyperbolic space, we need to understand its distinction from the Euclidean space in handling hierarchy. For this, we study the intra-level and inter-level Euclidean and hyperbolic distance between entities at different levels of the dataset. Let us say $E_p$ is the set of entities at level $P_p$ in the E-commerce Product Networks dataset. For the analysis, we calculate two sets of distances; intra-level ($\Delta_{\text{intra}}$) and inter-level ($\Delta_{\text{inter}}$) distance as follows:

$$\Delta_{\text{intra}}(P_p) = \frac{\sum_{i,j \in E_p} \delta(e_i, e_j)}{|E_p| |E_p| - 1}$$  \hspace{1cm} (11)

$$\Delta_{\text{inter}}(P_{p1}, P_{p2}) = \frac{\sum_{i \in E_{p1}, j \in E_{p2}} \delta(e_i, e_j)}{|E_{p1}| |E_{p2}|}$$  \hspace{1cm} (12)

$\delta(., .)$ is replaced with Euclidean norm (on Query2Box representations) and hyperbolic distance ($d_{hyp}$ on HypE representations) to understand the difference between the hierarchical separation of entities in the two spaces.

In the $\Delta_{\text{intra}}$ results (depicted in Figure 10), we observe that, with increasing level of hierarchy the distance between entities at different levels remains constant in the case of Euclidean space and shows a clear decreasing trend for hyperbolic space. This indicates denser clustering of entities at the same level. Additionally, the $\Delta_{\text{inter}}$ results, illustrated in Figure 11, depict a linear increase in distance between inter-level entities in the Euclidean space and a superlinear growth in the hyperbolic space. This shows that hyperbolic space also learns clusters such that inter-level entities are farther apart compared to Euclidean space. This nature of inter-level discrimination and intra-level aggregation demonstrates the superior ability of hyperbolic spaces at capturing hierarchical features.

**Figure 10:** Intra-level Euclidean (Q2B) and Hyperbolic (HypE) distances. The graph presents $\Delta_{\text{intra}}$ of entity sets at different hierarchy levels, given on the x-axis.

B ABLATION STUDY: FINER RESULTS

The finer results, across all datasets, of our Ablation study, described in Section 4.4, are given in Tables 8 and 9. We observe that, on average over all types of reasoning centers, a combination of attention aggregation for centers and Deepsets aggregation for limits results in the best performance across all the datasets. Thus, we utilize this combination in our final model.
Table 8: Performance comparison of HypE (ours) against its variants to study the efficacy of its Query-Search space. The columns present different query structures and averages values. Best results for each dataset are shown in bold.

| Metrics | Dataset | Model      | Hits@3 | Mean Reciprocal Rank |
|---------|---------|------------|--------|----------------------|
|         | FB15k   | HypE-Avg-1t | 803.0  | 803.0                |
|         | FB15k   | HypE-Avg-1,2t | 803.0  | 803.0                |
|         | FB15k   | HypE-Avg    | 802.0  | 802.0                |
|         | FB15k   | HypE-DS     | 777.0  | 777.0                |
|         | HypE-Tc | 809.46     | 809.46 |                      |
|         | HypE-(final) | 809.46 | 809.46 |                      |
|         | FB15k   | HypE-Avg-1t | 567.229 | 567.229             |
|         | FB15k   | HypE-Avg-1,2t | 567.259 | 567.259             |
|         | FB15k   | HypE-Avg    | 567.36 | 567.36              |
|         | FB15k   | HypE-DS     | 549.361 | 549.361             |
|         | FB15k   | HypE-Tc     | 573.36 | 573.36              |
|         | FB15k   | HypE-(final) | 572.366 | 572.366            |
|         | NELL    | HypE-Avg-1t | 614.222 | 614.222            |
|         | NELL    | HypE-Avg-1,2t | 615.290 | 615.290            |
|         | NELL    | HypE-Avg    | 612.350 | 612.350           |
|         | NELL    | HypE-DS     | 594.350 | 594.350           |
|         | NELL    | HypE-Tc     | 616.380 | 616.380           |
|         | NELL    | HypE-(final) | 618.359 | 618.359          |
|         | DBPedia | HypE-Avg-1t | 963.473 | 963.473          |
|         | DBPedia | HypE-Avg-1,2t | 962.609 | 962.609         |
|         | DBPedia | HypE-Avg    | 961.745 | 961.745         |
|         | DBPedia | HypE-DS     | 932.745 | 932.745         |
|         | DBPedia | HypE-Tc     | 971.757 | 971.757        |
|         | DBPedia | HypE-(final) | 970.756 | 970.756       |

Table 9: Performance comparison of HypE (ours) against its variants on E-commerce dataset to study the efficacy of the model's Query-Search space. The columns present the different query structures and averages over them. Best results for each dataset are given in bold.

| Metrics | Dataset | Model      | Hits@3 | Mean Reciprocal Rank |
|---------|---------|------------|--------|----------------------|
|         | E-commerce | HypE-Avg-1t | 0.00   | 0.00                |
|         | E-commerce | HypE-Avg-1,2t | 0.00   | 0.00                |
|         | E-commerce | HypE-Avg    | 0.00   | 0.00                |
|         | E-commerce | HypE-DS     | 0.00   | 0.00                |
|         | E-commerce | HypE-Tc     | 0.00   | 0.00                |
|         | E-commerce | HypE-(final) | 0.00   | 0.00               |

Table 10: Performance comparison of HypE (ours) against the baselines on an e-commerce dataset for intersection and union queries. We fix GQE (E-vector) to be our baseline and report the relative improvements against that for all the methods. The numbers are in percentages. Best results for each dataset are shown in bold.

| Metrics | Dataset | Model      | Hits@3 | Mean Reciprocal Rank |
|---------|---------|------------|--------|----------------------|
|         | GQE     | 0.00       | 0.00   | 0.00                |
|         | KQAT    | 0.00       | 0.00   | 0.00                |
|         | HQE     | 0.00       | 0.00   | 0.00                |
|         | Q2B     | 0.00       | 0.00   | 0.00                |
|         | HypE    | 0.00       | 0.00   | 0.00                |

Figure 11: Inter-level Euclidean (Q2B) and Hyperbolic (Hype) distances. Each graph presents Δinter, between entity pairs of a source hierarchical level, given by the graph label and the other hierarchy levels in the dataset, given on the x-axis.

C E-COMMERCE RESULTS

The intersection and union results on the e-commerce datasets, described in Section 4.3, are given in Table 10.

1 Due to the sensitive nature of the proprietary dataset, we only report relative results.