

Self-Supervised Hyperboloid Representations from Logical Queries over Knowledge Graphs

Nurendra Choudhary¹, Nikhil Rao², Sumeet Katariya², Karthik Subbian², Chandan K. Reddy^{1,2}

¹Department of Computer Science, Virginia Tech, Arlington, VA

²Amazon, Palo Alto, CA

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 multi-hop 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 as well as hierarchically organized web articles 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

ACM Reference Format:

Nurendra Choudhary¹, Nikhil Rao², Sumeet Katariya², Karthik Subbian², Chandan K. Reddy^{1,2}. 2021. Self-Supervised Hyperboloid Representations from Logical Queries over Knowledge Graphs. In *Proceedings of the Web Conference 2021 (WWW '21)*, April 19–23, 2021, Ljubljana, Slovenia. ACM, New York, NY, USA, 12 pages. <https://doi.org/10.1145/3442381.3449974>

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 application domains 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].

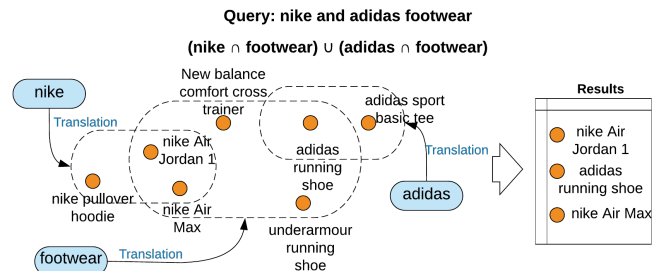


Figure 1: An example of PFOE querying in the E-commerce product network. The product space of Adidas and Nike intersects with Footwear to narrow the search space. A union over these spaces yields our final set of entity results.

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.

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.

WWW '21, April 19–23, 2021, Ljubljana, Slovenia

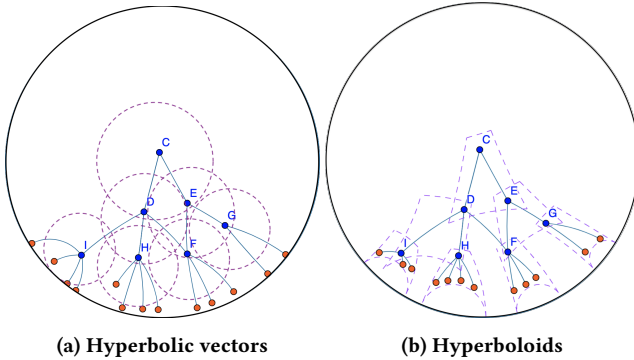
© 2021 IW3C2 (International World Wide Web Conference Committee), published under Creative Commons CC-BY 4.0 License.

ACM ISBN 978-1-4503-8312-7/21/04.

<https://doi.org/10.1145/3442381.3449974>

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.



(a) Hyperbolic vectors (b) Hyperboloids
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.

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 set of 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

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 ($\cap$) and union ($\cup$) 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 $E = \{e_1, \dots, e_n\}$ and correspondingly retrieve the set of all entities that are connected to all $e_i \in E$ and any $e_i \in E$. For heterogeneous graphs, the equivalent is to retrieve nodes connected to all nodes $e_i \in E$ and any $e_i \in E$. 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.

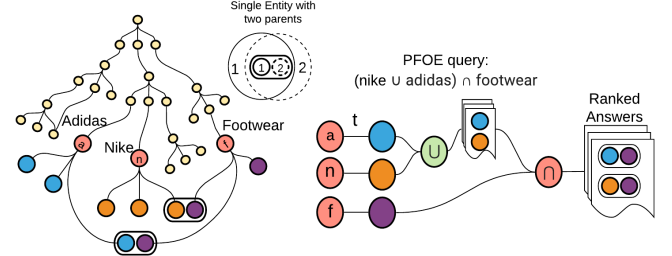


Figure 3: A simple first-order query over a Product Graph. The query Nike and Adidas footwear can be expressed as the union of the Nike and Adidas nodes intersected with the node corresponding to footwear in the product catalog.

3.2 Problem Setup

We denote $KG = (E, R)$ as a set of entities $e_i \in E$ and relations $r_{ij} \in R : e_i \rightarrow e_j$ as Boolean functions that indicate whether a directed relation r_{ij} holds between e_i and e_j . Intersection ($\cap$) and Union ($\cup$) are positive first-order existential (PFOE) operations defined on a set of queries $q_1, q_2, \dots, q_n \in Q$:

$$q_{\cap}[Q] = V_{\cap} \subseteq E \exists e_1, e_2, \dots, e_k : q_1 \cap q_2 \cap q_3 \dots \cap q_n$$

$$q_{\cup}[Q] = V_{\cup} \subseteq E \exists e_1, e_2, \dots, e_k : q_1 \cup q_2 \cup q_3 \dots \cup q_n$$

where $(q_{\cap}, V_{\cap})$ and $(q_{\cup}, V_{\cup})$ is the query space and resultant entity set after the intersection and union operations over query set Q , respectively¹. Given a dataset of logical queries over a KG, the goal, here, is to learn hyperboloid (in a Poincaré ball) representations for its entities $e_i \in \mathbb{R}^{2d}$ and relations $r_{ij} \in \mathbb{R}^{2d}$, 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.

¹e.g., if $(q_i, \{e_a, e_b\})$ and $(q_j, \{e_b, e_c\})$ are the corresponding query space and resultant entity sets, then $q_{\cap}[\{q_i, q_j\}] = \{e_b\}$ and $q_{\cup}[\{q_i, q_j\}] = \{e_a, e_b, e_c\}$.

Distance between hyperboloid and entity point (d): Given a query hyperboloid $q \in \mathbb{R}^{2d}$ and entity center $v \in \mathbb{R}^d$, the distance between them is defined as:

$$\begin{aligned} d_{hyp}(v, q) &= d_{out}(v, q) \oplus_c \gamma d_{in}(v, q) \\ d_{out}(v, q) &= \|\text{Max}(d_{\mathbb{H}}(v, q_{max}), 0) + \text{Max}(d_{\mathbb{H}}(q_{min}, v), 0)\|_1 \\ d_{in}(v, q) &= \|\text{cen}(q) \ominus_c \text{Min}(q_{max}, \text{Max}(q_{min}, v))\|_1 \\ q_{min} &= \text{cen}(q) \ominus_c \text{lim}(q), \quad q_{max} = \text{cen}(q) \oplus_c \text{lim}(q) \end{aligned} \quad (3)$$

where d_{out} represents the distance of the entity to limits of the hyperboloid and d_{in} is the distance of the entity from the hyperboloid's border to its center. γ is a scalar weight (set to 0.5 in our experiments) and $\|x\|_1$ is the $L1$ -norm of x .

Translation (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 $v_t \in V_t \subseteq E$ (d_t) as follows:

$$o_t = e \oplus_c r, \quad d_t(v) = d_{hyp}(v_t, o_t) \quad (4)$$

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

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

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

$$\text{cen}(Q_{\cap}) = \sum_i a_i \ominus_c \text{cen}(e_i); \quad a_i = \frac{\exp(f(e_i))}{\sum_j \exp(f(e_j))} \quad (6)$$

$$\text{lim}(Q_{\cap}) = \min(\{\text{lim}(e_1), \dots, \text{lim}(e_n)\}) \ominus_c \sigma(DS(\{e_1, \dots, e_n\})) \quad (7)$$

$$\begin{aligned} DS(\{e_1, \dots, e_n\}) &= f\left(\frac{1}{n} \sum_{i=1}^n f(e_i)\right) \\ d_{\cap}(v_{\cap}) &= d_{hyp}(v_{\cap}, o_{\cap}) \end{aligned} \quad (8)$$

where $\ominus_c$ is the Möbius scalar product, $f(\cdot) : \mathbb{R}^{2d} \rightarrow \mathbb{R}^d$ is the multilayer perceptron (MLP), $\sigma(\cdot)$ is the sigmoid function and $DS(\cdot)$ is the permutation invariant deep architecture, namely, DeepSets [49]. $\text{Min}(\cdot)$ and $\exp(\cdot)$ 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 ($\cup$): 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 $Q_{\cup} = \{e_1, e_2, e_3, \dots, e_n\}$, union space is given by $o_{\cup}$ and distance from result entities $v_{\cup} \in V_{\cup} \subseteq E$ by $d_{\cup}$, where

$$o_{\cup} = Q_{\cup} \quad (9)$$

$$d_{\cup}(v_{\cup}) = \min\left(\{d_{hyp}(v_{\cup}, e_i) \mid \forall e_i \in o_{\cup}\}\right) \quad (10)$$

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

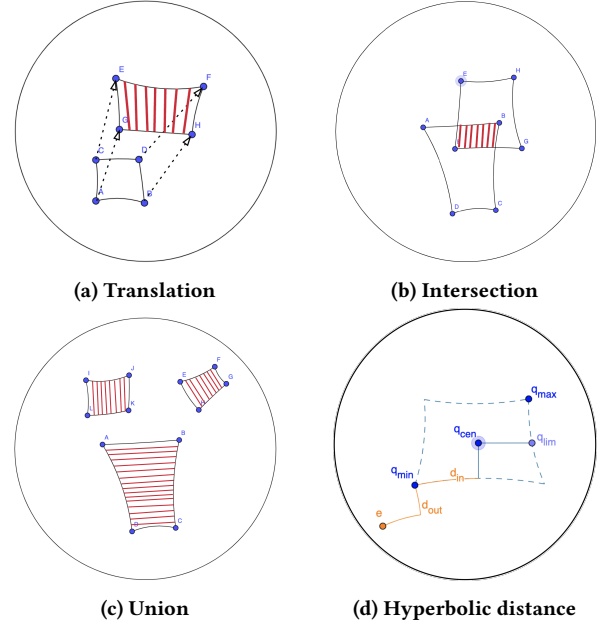


Figure 7: PFOE queries and hyperbolic distance in a Poincaré geodisc.

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 β 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 2×400 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) ².

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 $D_t, D_\cap, D_\cup$, which are set of all (query (Q), result (V)) for translation, intersection, and union, respectively;

Output: Entity E and Relation R hyperboloids;

```

1 Randomly initialize  $e \in E$  and  $r \in R$  ( $e, r \in \mathbb{H}^{2d}$ );
2 for number of epochs; until convergence do
3    $l = 0$ ; # Initialize loss
4   for  $\{(e, r, V_t) \in D_t\}$  do
5      $o_t = e \oplus r$ , from Eq. (4)
6     # Update loss for translation queries
7      $l = l + \sum_{v_t \in V_t} d_{hyp}(v_t, o_t)$ 
8   end
9   for  $\{(Q_\cap, V_\cap) \in D_\cap\}$  do
10     $o_\cap = (cen(Q_\cap), lim(Q_\cap))$ , from Eq. (5)
11    # Update loss for intersection queries
12     $l = l + \sum_{v_\cap \in V_\cap} d_{hyp}(v_\cap, o_\cap)$ 
13  end
14  for  $\{(Q_\cup, V_\cup) \in D_\cup\}$  do
15     $o_\cup = Q_\cup$ , from Eq. (9)
16    # Update loss for union queries
17     $l = l + \sum_{v_\cup \in V_\cup} Min(d_{hyp}(v_\cup, e_i) \forall e_i \in o_\cup)$ 
18  end
19  # Update E and R with backpropagation
20   $E \leftarrow E - \Delta_E l$ 
21   $R \leftarrow R - \Delta_R l$ 
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 invertible 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_{train} , KG_{valid} and KG_{test} in a 75 : 10 : 15 ratio for our experiments. More details of the datasets are given in Table 1.

Table 1: Basic statistics of the datasets including the number of unique entities, relations, and edges.

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).

³<https://www.kaggle.com/danofer/dbpedia-classes>

⁴Proprietary dataset

Table 2: Performance comparison of HypE (ours) against the baselines to study the efficacy of the Query-Search space. The columns present the 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		Hits@3										Mean Reciprocal Rank									
Dataset	Model	1t	2t	3t	2∩	3∩	2∪	∩t	t∩	∪t	Avg	1t	2t	3t	2∩	3∩	2∪	∩t	t∩	∪t	Avg
FB15k	GQE (E-Vector)	.636	.345	.248	.515	.624	.376	.151	.310	.273	.386	.505	.320	.218	.439	.536	.300	.139	.272	.244	.330
	KGAT (E-Vector)	.711	.379	.276	.553	.667	.492	.181	.354	.302	.435	.565	.352	.243	.471	.573	.393	.167	.311	.270	.372
	HQE (H-Vector)	.683	.365	.265	.451	.589	.438	.135	.283	.290	.389	.543	.339	.233	.384	.506	.350	.125	.249	.259	.332
	Q2B (Box)	.786	.413	.303	.590	.710	.608	.211	.397	.330	.483	.654	.373	.274	.488	.602	.468	.194	.339	.301	.410
	HypE (Hyperboloid)	.809	.486	.365	.598	.728	.610	.206	.406	.410	.513	.673	.439	.330	.495	.617	.470	.189	.347	.374	.437
FB15k-237	GQE (E-Vector)	.404	.214	.147	.262	.390	.164	.087	.162	.155	.221	.346	.193	.145	.250	.355	.145	.086	.156	.151	.203
	KGAT (E-Vector)	.436	.227	.167	.293	.422	.202	.069	.135	.174	.236	.373	.205	.165	.280	.384	.179	.068	.130	.170	.217
	HQE (H-Vector)	.440	.231	.171	.265	.387	.195	.083	.162	.183	.235	.376	.209	.169	.253	.352	.173	.082	.156	.179	.217
	Q2B (Box)	.467	.240	.186	.324	.453	.239	.050	.108	.193	.251	.400	.225	.173	.275	.378	.198	.105	.180	.178	.235
	HypE (Hyperboloid)	.572	.366	.255	.399	.527	.225	.145	.246	.282	.335	.490	.343	.237	.339	.440	.186	.305	.410	.260	.334
NELL995	GQE (E-Vector)	.417	.231	.203	.318	.454	.200	.081	.188	.139	.248	.311	.193	.175	.273	.399	.159	.078	.168	.130	.210
	KGAT (E-Vector)	.486	.249	.218	.331	.467	.285	.107	.200	.151	.277	.362	.208	.188	.284	.410	.227	.103	.179	.141	.234
	HQE (H-Vector)	.477	.250	.219	.270	.413	.267	.091	.153	.166	.256	.355	.209	.189	.232	.363	.213	.088	.137	.155	.216
	Q2B (Box)	.555	.266	.233	.343	.480	.369	.132	.212	.163	.306	.413	.227	.208	.288	.414	.266	.125	.193	.155	.254
	HypE (Hyperboloid)	.618	.359	.312	.400	.563	.441	.143	.227	.278	.371	.460	.306	.279	.336	.486	.318	.135	.207	.264	.310
DBPedia R = 1	GQE (E-Vector)	.673	.006	N.A.	.873	.879	.402	.160	.668	0.00	.406	.502	.005	N.A.	.749	.773	.32	.154	.597	0.00	.344
	KGAT (E-Vector)	.753	.007	N.A.	.937	.940	.526	.192	.762	0.00	.457	.561	.006	N.A.	.804	.825	.419	.185	.682	0.00	.387
	HQE (H-Vector)	.422	.003	N.A.	1.00	1.00	.138	.109	.182	.001	.296	.314	.003	N.A.	.859	.879	.110	.105	.163	.001	.270
	Q2B (Box)	.832	.007	N.A.	1.00	1.00	.649	.224	.856	0.00	.508	.619	.006	N.A.	.840	.863	.468	.212	.779	0.00	.421
	HypE (Hyperboloid)	.897	.009	N.A.	1.00	1.00	.708	.294	.935	.001	.546	.668	.008	N.A.	.840	.863	.511	.278	.853	.001	.447
DBPedia R = p	GQE (E-Vector)	.730	.565	N.A.	.873	.879	.534	.213	.705	.027	.504	.544	.421	N.A.	.651	.656	.398	.159	.526	.020	.375
	KGAT (E-Vector)	.816	.621	N.A.	.937	.940	.699	.255	.804	.030	.567	.608	.463	N.A.	.698	.700	.521	.190	.599	.022	.422
	HQE (H-Vector)	.456	.182	N.A.	1.00	1.00	.184	.143	.192	.143	.367	.339	.135	N.A.	.744	.744	.137	.106	.143	.106	.273
	Q2B (Box)	.901	.676	N.A.	1.00	1.00	.863	.297	.903	.033	.630	.670	.503	N.A.	.744	.744	.642	.221	.672	.025	.469
	HypE (Hyperboloid)	.970	.756	N.A.	1.00	1.00	.940	.388	.985	.046	.676	.722	.563	N.A.	.744	.744	.700	.289	.733	.034	.503
Avg. Improv. (%) (HypE vs Q2B)		15.0	39.2	67.0	4.1	4.7	13.0	28.5	12.9	7.3	14.4	14.9	38.4	59.8	4.4	5.0	12.9	39.5	17.9	64.9	18.5

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∪). 1t, 2t, and 3t denote translation with 1, 2 and 3 consecutive relations, respectively. 2∩ and 2∪ stand for intersection and union over two entities, respectively. 3∩ represents intersection 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.

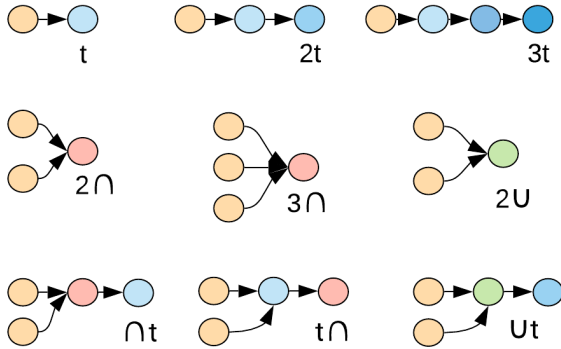


Figure 8: Logical query structures designed to compare HypE against baselines. The blue, red, and green units denote translation, intersection, and union operations, respectively.

We extract the ground truth query-entity pairs by traversing the datasets [35]. The models are trained on queries from KG_{train} and validated on KG_{valid} . The final evaluation metrics are calculated on KG_{test} . We utilize Euclidean norm and hyperbolic distance (given in Eq. (3)) to measure the distance between query embeddings and its resultant entities in Euclidean and hyperbolic spaces, respectively. The sorted query-entity distances are the ranked results for the given query.

Table 3: Performance comparison of HypE (ours) against the baselines on an e-commerce dataset for multi-level translation queries. We assume GQE (E-vector) as a baseline and report relative improvements against that for all the methods. The numbers are in percentages. Best results for each dataset are shown in bold.⁵

Metrics		Hits@3			Mean Reciprocal Rank		
Dataset	Model	1t	2t	3t	1t	2t	3t
R = 1	GQE (E-Vector)	0.0	0.0	0.0	0.0	0.0	0.0
	KGAT (E-Vector)	12.12	10.9	0.0	0.0	-36.2	-57.3
	HQE (H-Vector)	-34.8	-51.6	0.0	-34.7	-56.6	0.0
	Q2B (Box)	30.3	4.7	200	23.5	7.5	0.0
	HypE (Hyperboloid)	38.6	98.4	200	38.8	79.2	100
R = p	GQE (E-Vector)	0.0	0.0	0.0	0.0	0.0	0.0
	KGAT (E-Vector)	12.2	10.7	0.0	11.5	9.5	0.0
	HQE (H-Vector)	38.8	92.9	350	38.5	90.5	600
	Q2B (Box)	17.3	37.5	0.0	23.1	19.0	0.0
	HypE (Hyperboloid)	122.3	298.2	8350	192.3	471.4	9900

Given a test query q_{test} , let the true ranked result entities be E_{result} and model's ranked output be $E_{output} = \{e_{o1}, e_{o2}, \dots, e_{on}\}$. The evaluation metrics used in our work are Hits@K and Mean

⁵Due to the sensitive nature of the proprietary dataset, we only report relative results.

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.

Metrics		Hits@3									Mean Reciprocal Rank									
Model Variants	1t	2t	3t	2∩	3∩	2∪	∩t	t∩	∪t	Avg	1t	2t	3t	2∩	3∩	2∪	∩t	t∩	∪t	Avg
HypE-Avg-1t	.651	.274	.153	.383	.414	.484	.099	.235	.135	.312	.538	.238	.149	.309	.332	.366	.108	.216	.126	.264
HypE-Avg-1,2,3t	.650	.353	.213	.431	.474	.486	.133	.280	.187	.352	.532	.306	.204	.345	.381	.366	.140	.256	.171	.297
HypE-Avg	.650	.432	.272	.479	.535	.487	.167	.325	.239	.392	.525	.373	.258	.381	.429	.366	.172	.296	.215	.330
HypE-DS	.630	.432	.267	.488	.555	.451	.165	.356	.238	.392	.511	.376	.257	.399	.460	.344	.172	.326	.213	.334
HypE-TC	.656	.439	.275	.490	.565	.488	.177	.374	.243	.406	.530	.379	.262	.391	.458	.368	.184	.340	.216	.342
HypE (final)	.656	.438	.275	.490	.564	.488	.177	.373	.242	.405	.530	.378	.262	.390	.458	.368	.184	.340	.215	.341

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

$$HITS@K(q_{test}) = \frac{1}{K} \sum_{k=1}^K f(e_{ok}), \quad f(e_{ok}) = \begin{cases} 1, & \text{if } e_{ok} \in E_{result} \\ 0, & \text{otherwise} \end{cases}$$

$$MRR(q_{test}) = \frac{1}{n} \sum_{i=1}^n \frac{1}{f(e_{oi})}, \quad f(e_{oi}) = \begin{cases} i, & \text{if } e_{oi} \in E_{result} \\ \infty, & \text{otherwise} \end{cases}$$

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 GQE 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); P_1 , P_2 and P_3 . 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 Miscategorized 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	1			2			3		
	Models	P	R	F1	P	R	F1	P	R	F1
DBPedia $ R = p$	GQE	.512	.369	.428	.549	.446	.492	.576	.409	.479
	KGAT	.523	.375	.437	.552	.448	.495	.578	.416	.484
	HQE	.529	.385	.446	.556	.45	.497	.586	.424	.492
	Q2B	.589	.479	.527	.589	.479	.528	.597	.481	.532
	HypE	.590	<u>.479</u>	.528	<u>.648</u>	.482	.552	.650	<u>.486</u>	<u>.557</u>
	HypE-SI	.591	<u>.479</u>	.529	<u>.648</u>	<u>.483</u>	.553	.651	<u>.486</u>	<u>.557</u>
	HypE-SC	.601	.501	.546	.662	.563	.608	.705	.563	.626

Table 6: Results on Miscategorized 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 GQE baseline. P, R, and F1 represent Precision, Recall, and F-score, respectively.⁶

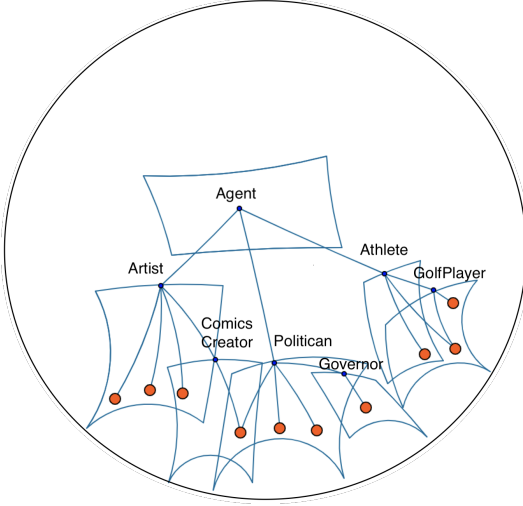
Dataset	P-Level	1			2			3		
	Models	P	R	F1	P	R	F1	P	R	F1
E-commerce $ R = p$	GQE	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	KGAT	2.2	1.8	2.1	0.6	0.5	0.7	0.2	1.6	1.1
	HQE	3.4	4.5	4.1	1.2	1.0	1.1	1.7	3.5	2.8
	Q2B	15.1	29.9	23.4	7.2	7.4	7.4	3.6	17.5	11.3
	HypE	15.3	<u>29.9</u>	23.4	<u>18.0</u>	8.4	<u>12.8</u>	16.6	18.5	<u>17.2</u>
	HypE-SI	<u>15.5</u>	<u>29.9</u>	<u>23.7</u>	<u>18.0</u>	<u>8.4</u>	12.5	16.4	<u>18.8</u>	<u>17.2</u>
	HypE-SC	17.4	35.8	27.5	20.6	26.4	23.7	22.3	37.6	30.8

We construct a pseudo-tree, where all parent nodes are infused with 10% noise of 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

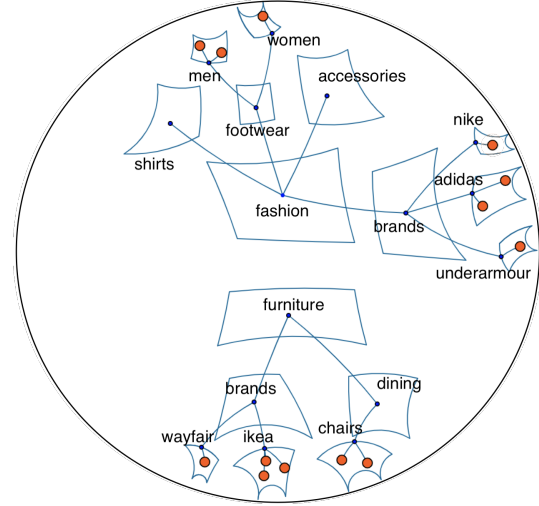
⁶Due to the sensitive nature of the proprietary dataset, we only report relative results.

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		Prediction		
	Category	P-level	Q2B	HypE	HypE-SC
ASICS Women’s Gel-Venture 7 Trail Running Shoes, 5.5M, 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	MISCAT	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	Wool & Blends	2	TRUE	MISCAT	MISCAT
Fifth Harmony Vintage Photo Blue T-Shirt (M)	Customer Segment	2	TRUE	TRUE	MISCAT
Billy Bills Playoff Shirt Buffalo T-Shirt Let’s Go Buffalo Tee	Customer Segment	2	MISCAT	TRUE	TRUE
Kanu Surf Toddler Karlie Flounce Beach Sport 2-Piece Swimsuit, Ariel Blue, 4T	Brand Stores	3	TRUE	MISCAT	MISCAT
The North Face Infant Glacier ¼ Snap, Mr. Pink, 0-3 Months US Infant	Brand 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	TRUE	TRUE	MISCAT



(a) DBpedia Taxonomy



(b) E-commerce Product Networks

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.

of the immediate parent nodes. From the set of all intermediate nodes KG_P , given a parent $p \in KG_P$ and its original set of children $C_+ = \text{children}(p)$ and randomly sampled set of anomalous children $C_- = \text{children}(KG_P \setminus \{p\})$, the aim here is to identify $c \in C_-$ from $C_+ \cup C_-$. We use Precision, Recall and F1-score as the evaluation metrics for this experiment.

The results (given in Table 5 and Table 6) show that, although HypE has comparable performance to baselines at P_1 , it outperforms the baselines by more than 5% at P_2 and P_3 . This demonstrates the robustness of HypE to noisy data and its capability of capturing relevant hierarchical features (as F1 on $P_3 > P_2 > P_1$) for downstream tasks. Furthermore, the specific task is critical in e-commerce search as irrelevant results impede a smooth user experience. Table 7 presents some qualitative examples from the E-commerce dataset.

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 miscategorization. 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 density 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 Bécigneul and Octavian-Eugen Ganea. 2019. Riemannian Adaptive Optimization Methods. In *International Conference on Learning Representations*. <https://openreview.net/forum?id=r1eiq09K7>
- [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 C. 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) <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 Modeling Multi-relational Data. In *Advances in Neural Information Processing Systems* 26. 2787–2795.
- [8] Michael M 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 R Hruschka Jr, and Tom M Mitchell. 2010. Toward an architecture for never-ending language learning.. In *Aaai*, 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, 6901–6914. <https://doi.org/10.18653/v1/2020.acl-main.617>
- [12] Ines Chami, Zhitao Ying, Christopher Ré, and Jure Leskovec. 2019. Hyperbolic graph convolutional neural networks. In *Advances in neural information processing systems*. 4868–4879.
- [13] Michaël 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 King. 2020. MAGNN: Metapath Aggregated Graph Neural Network for Heterogeneous Graph Embedding. In *Proceedings of The Web Conference 2020*. 2331–2341.
- [17] Octavian 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 Jamaludin, Timor Kadir, 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. 2017. 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 Computation Graphs. In *International Conference on Learning Representations*.
- [28] Kurt Miller, Michael I 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*

- on *International Conference on Machine Learning* (Haifa, Israel) (ICML '10). Omnipress, Madison, WI, USA, 807–814.
- [31] Maximilian Nickel, Lorenzo Rosasco, and Tomaso Poggio. 2016. Holographic Embeddings of Knowledge Graphs. In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence* (Phoenix, Arizona) (AAAI'16). AAAI Press, 1955–1961.
- [32] Maximilian Nickel, Volker Tresp, and Hans-Peter Krieger. 2011. A three-way model for collective learning on multi-relational data. In *ICML*, Vol. 11. 809–816.
- [33] Adam Paszke, Sam Gross, and et. al. Massa, Francisco. 2019. PyTorch: An Imperative Style, High-Performance Deep Learning Library. In *Advances in Neural Information Processing Systems* 32. 8024–8035.
- [34] Bryan Perozzi, Rami Al-Rfou, and Steven Skiena. 2014. Deepwalk: Online learning of social representations. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*. 701–710.
- [35] Hongyu Ren*, Weihua Hu*, and Jure Leskovec. 2020. Query2box: Reasoning over Knowledge Graphs in Vector Space Using Box Embeddings. In *International Conference on Learning Representations*. <https://openreview.net/forum?id=Bjgr4kSFD5>
- [36] Linfeng Song, Yue Zhang, Zhiguo Wang, and Daniel Gildea. 2018. A Graph-to-Sequence Model for AMR-to-Text Generation. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Association for Computational Linguistics, Melbourne, Australia, 1616–1626. <https://doi.org/10.18653/v1/P18-1150>
- [37] Haitian Sun, Andrew O Arnold, Tania Bedrax-Weiss, Fernando Pereira, and William W Cohen. 2020. Faithful Embeddings for Knowledge Base Queries. *Advances in Neural Information Processing Systems* 33 (2020).
- [38] Kristina Toutanova, Danqi Chen, Patrick Pantel, Hoifung Poon, Pallavi Choudhury, and Michael Gamon. 2015. Representing Text for Joint Embedding of Text and Knowledge Bases. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, Lisbon, Portugal, 1499–1509. <https://doi.org/10.18653/v1/D15-1174>
- [39] George Fletcher ulong Pei, Xin Du and Mykola Pechenizkiy. 2019. Dynamic Network Representation Learning via Gaussian Embedding. *NeurIPS 2019 Workshop on Graph Representation Learning* (2019).
- [40] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. 2018. Graph Attention Networks. In *International Conference on Learning Representations*. <https://openreview.net/forum?id=rjXmpikCZ>
- [41] Luke Vilnis, Xiang Li, Shikhar Murty, and Andrew McCallum. 2018. Probabilistic Embedding of Knowledge Graphs with Box Lattice Measures. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Association for Computational Linguistics, Melbourne, Australia, 263–272. <https://doi.org/10.18653/v1/P18-1025>
- [42] Luke Vilnis and Andrew McCallum. 2015. Word Representations via Gaussian Embedding. In *ICLR*. <http://arxiv.org/abs/1412.6623>
- [43] Quan Wang, Zhendong Mao, Bin Wang, and Li Guo. 2017. Knowledge graph embedding: A survey of approaches and applications. *IEEE Transactions on Knowledge and Data Engineering* 29, 12 (2017), 2724–2743.
- [44] Xiang Wang, Xiangnan He, Yixin Cao, Meng Liu, and Tat-Seng Chua. 2019. Kgat: Knowledge graph attention network for recommendation. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 950–958.
- [45] Xiao Wang, Houye Ji, Chuan Shi, Bai Wang, Yanfang Ye, Peng Cui, and Philip S Yu. 2019. Heterogeneous graph attention network. In *The World Wide Web Conference*. 2022–2032.
- [46] W. X. Wilcke, P. Bloem, V. de Boer, R. H. van t Veer, and F. A. H. van Harmelen. 2020. End-to-End Entity Classification on Multimodal Knowledge Graphs. [arXiv:2003.12383 \[cs.AI\]](https://arxiv.org/abs/2003.12383)
- [47] Xing Xu, Huimin Lu, Jingkuan Song, Yang Yang, Heng Tao Shen, and Xuelong Li. 2019. Ternary adversarial networks with self-supervision for zero-shot cross-modal retrieval. *IEEE transactions on cybernetics* 50, 6 (2019), 2400–2413.
- [48] Bishan Yang, Wen-tau Yih, Xiaodong He, Jianfeng Gao, and Li Deng. 2015. Embedding entities and relations for learning and inference in knowledge bases. In *International Conference on Learning Representations*.
- [49] Manzil Zaheer, Satwik Kottur, Siamak Ravanbakhsh, Barnabas Poczos, Russ R Salakhutdinov, and Alexander J Smola. 2017. Deep sets. In *Advances in neural information processing systems*. 3391–3401.
- [50] Chuxu Zhang, Dongjin Song, Chao Huang, Ananthram Swami, and Nitesh V. Chawla. 2019. Heterogeneous Graph Neural Network. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (Anchorage, AK, USA) (KDD '19). Association for Computing Machinery, New York, NY, USA, 793–803. <https://doi.org/10.1145/3292500.3330961>
- [51] Jiawei Zhang, Haopeng Zhang, Li Sun, and Congying Xia. 2020. Graph-Bert: Only Attention is Needed for Learning Graph Representations. *arXiv preprint arXiv:2001.05140* (2020).
- [52] Jun Zhu. 2012. Max-Margin Nonparametric Latent Feature Models for Link Prediction. In *Proceedings of the 29th International Conference on Machine Learning* (Edinburgh, Scotland) (ICML '12). Omnipress, Madison, WI, USA, 1179–1186.

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 (Δ_{intra}) and inter-level (Δ_{inter}) distance as follows:

$$\Delta_{intra}(P_p) = \frac{\sum_{i=0}^{|E_p|} \sum_{j=i}^{|E_p|} \delta(e_i, e_j)}{|E_p|(|E_p| - 1)} \quad (11)$$

$$\Delta_{inter}(P_{p1}, P_{p2}) = \frac{\sum_{i=0}^{|E_{p1}|} \sum_{j=0}^{|E_{p2}|} \delta(e_i, e_j)}{|E_{p1}| \times |E_{p2}|} \quad (12)$$

$\delta(\cdot, \cdot)$ 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 Δ_{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 Δ_{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.

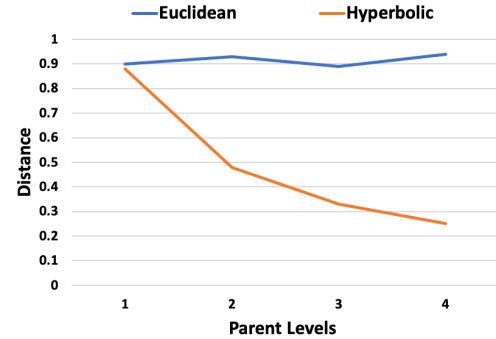


Figure 10: Intra-level **Euclidean (Q2B)** and **Hyperbolic (HypE)** distances. The graph presents Δ_{intra} of entity sets at different hierarchy levels, given on the x-axis.

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 queries, 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 the baselines to study the efficacy of its Query-Search space. The columns present different query structures and averages values. The final row presents the Average Relative Improvement (%) of HypE compared to Query2Box over all datasets. Best results for each dataset are shown in bold.

Metrics	Dataset	Model	Hits@3										Mean Reciprocal Rank									
			1t	2t	3t	2n	3n	2u	n _f	t _n	u _f	Avg	1t	2t	3t	2n	3n	2u	n _f	t _n	u _f	Avg
FB15k		HypE-Avg-1t	.803	.304	.202	.468	.534	.606	.115	.256	.228	.395	.683	.276	.188	.392	.448	.467	.111	.221	.219	.338
		HypE-Avg-1,2,3t	.803	.392	.282	.527	.612	.608	.155	.305	.317	.446	.675	.355	.257	.438	.513	.467	.144	.262	.296	.380
		HypE-Avg	.802	.479	.361	.585	.690	.609	.194	.353	.405	.497	.667	.433	.325	.483	.578	.467	.177	.302	.373	.422
		HypE-DS	.777	.479	.354	.596	.717	.564	.192	.387	.403	.496	.649	.437	.324	.506	.620	.439	.176	.333	.370	.428
		HypE-TC	.809	.486	.365	.598	.729	.610	.206	.406	.410	.513	.674	.439	.331	.495	.617	.470	.189	.348	.375	.437
		HypE (final)	.809	.486	.365	.598	.728	.610	.206	.406	.410	.513	.673	.439	.330	.495	.617	.470	.189	.347	.374	.437
FB15k-237		HypE-Avg-1t	.567	.229	.141	.312	.387	.223	.081	.155	.157	.258	.498	.216	.135	.269	.319	.185	.179	.261	.152	.258
		HypE-Avg-1,2,3t	.567	.295	.197	.351	.444	.224	.109	.185	.218	.292	.492	.277	.185	.300	.366	.185	.233	.309	.206	.291
		HypE-Avg	.567	.361	.252	.390	.500	.225	.137	.214	.279	.325	.486	.338	.234	.331	.412	.185	.286	.357	.259	.323
		HypE-DS	.549	.361	.247	.398	.519	.208	.135	.234	.277	.324	.473	.341	.233	.347	.442	.174	.285	.393	.257	.327
		HypE-TC	.573	.366	.256	.399	.528	.225	.146	.246	.283	.335	.490	.344	.238	.339	.440	.186	.306	.411	.260	.335
		HypE (final)	.572	.366	.255	.399	.527	.225	.145	.246	.282	.335	.490	.343	.237	.339	.440	.186	.305	.410	.260	.334
NELL 995		HypE-Avg-1t	.614	.225	.173	.313	.413	.438	.080	.143	.155	.286	.467	.192	.159	.266	.353	.316	.079	.131	.154	.240
		HypE-Avg-1,2,3t	.613	.290	.241	.352	.474	.439	.108	.170	.215	.323	.462	.247	.217	.297	.404	.316	.103	.156	.209	.270
		HypE-Avg	.612	.354	.309	.391	.534	.440	.135	.197	.275	.359	.456	.302	.275	.328	.455	.316	.127	.180	.263	.299
		HypE-DS	.594	.354	.303	.399	.554	.408	.134	.216	.273	.359	.444	.304	.274	.344	.488	.297	.126	.198	.261	.304
		HypE-TC	.618	.360	.313	.400	.563	.442	.144	.228	.278	.371	.461	.306	.279	.336	.487	.319	.135	.207	.264	.310
		HypE (final)	.618	.359	.312	.400	.563	.441	.143	.227	.278	.371	.460	.306	.279	.336	.486	.318	.135	.207	.264	.310
DBPedia R = p		HypE-Avg-1t	.963	.473	N.A.	.782	.734	.933	.216	.621	.026	.521	.713	.374	N.A.	.589	.540	.696	.170	.467	.020	.390
		HypE-Avg-1,2,3t	.962	.609	N.A.	.880	.841	.936	.291	.739	.036	.588	.724	.455	N.A.	.658	.619	.696	.221	.553	.027	.438
		HypE-Avg	.961	.745	N.A.	.978	.948	.938	.366	.856	.045	.655	.715	.555	N.A.	.726	.697	.696	.271	.638	.034	.486
		HypE-DS	.932	.745	N.A.	.997	.985	.869	.362	.938	.045	.654	.697	.560	N.A.	.741	.738	.654	.270	.703	.034	.490
		HypE-TC	.971	.757	N.A.	1.00	1.00	.941	.388	.986	.047	.677	.722	.563	N.A.	.745	.745	.701	.289	.733	.034	.504
		HypE (final)	.970	.756	N.A.	1.00	1.00	.940	.388	.985	.046	.676	.722	.563	N.A.	.744	.744	.700	.289	.733	.034	.503

Table 9: Performance comparison of HypE (ours) against the baselines 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. The final row presents the Average Relative Improvement (%) of HypE compared to Query2Box over all datasets. Best results for each dataset are given in bold.⁷

Metrics		Hits@3										Mean Reciprocal Rank										
Dataset	Model	1t	2t	3t	2n	3n	2U	n _f	t _n	U _f	Avg	1t	2t	3t	2n	3n	2U	n _f	t _n	U _f	Avg	
E-commerce $ R = p$	HypE-Avg-1t	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
	HypE-Avg-1,2,3t	0.0	28.6	39.4	12.5	0.0	0.5	0.0	50.0	39.8	13.0	-1.0	28.5	36.8	13.3	0.0	0.0	0.0	0.0	100	35.3	12.6
	HypE-Avg	-0.3	57.1	77.7	25.0	0.0	0.9	0.0	50.0	78.7	26.0	-2.3	57.0	72.8	23.3	0.0	0.0	0.0	0.0	100	70.6	24.2
	HypE-DS	-3.3	57.1	74.5	27.5	0.0	-6.8	0.0	50.0	76.9	26.0	-4.9	58.3	71.9	30.0	0.0	-6.1	0.0	100	69.4	26.3	
	HypE-TC	0.7	60.0	80.9	27.5	0.0	0.9	0.0	50.0	81.5	31.0	-1.3	59.6	75.4	30.0	0.0	0.6	0.0	200	70.6	28.4	
	HypE (final)	0.7	59.3	79.8	27.5	0.0	0.9	0.0	50.0	80.6	30.0	-1.3	58.9	75.4	26.7	0.0	0.6	0.0	100	70.6	28.4	

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					
			2n	3n	2u	n _f	t _n	u _f	2n	3n	2u	n _f	t _n	u _f
R = 1		GQE	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
		KGAT	53.8	2850	30.8	200	300	0.0	-36.2	-57.3	24.4	-60	-45.5	0.0
		HQE	0.0	0.0	-57.7	0.0	-50.0	1500	-58.5	-99	-61	-80	-90.9	1100
		Q2B	-1.9	-50	32.7	0.0	50	0.0	-59.6	-99	53.7	-80	-81.8	0.0
		HypE	-1.9	-50	119.2	0.0	50	1500	-59.6	-99	107.3	-80	-81.8	1100
		HypE	0.0	0.0	416.3	0.0	2600	19400	11.8	0.0	415.6	-75	100	14400
R = p		GQE	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
		KGAT	0.0	0.0	30.2	0.0	2600	0.0	-8.8	0.0	31.3	-25	0.0	0.0
		HQE	0.0	0.0	0.0	-50	0.0	19400	11.8	0.0	0.0	-75	0.0	14400
		Q2B	0.0	0.0	95.3	0.0	200	0.0	11.8	0.0	59.4	-75	100	0.0
		HypE	0.0	0.0	416.3	0.0	2600	19400	11.8	0.0	415.6	-75	100	14400
		HypE	0.0	0.0	416.3	0.0	2600	19400	11.8	0.0	415.6	-75	100	14400

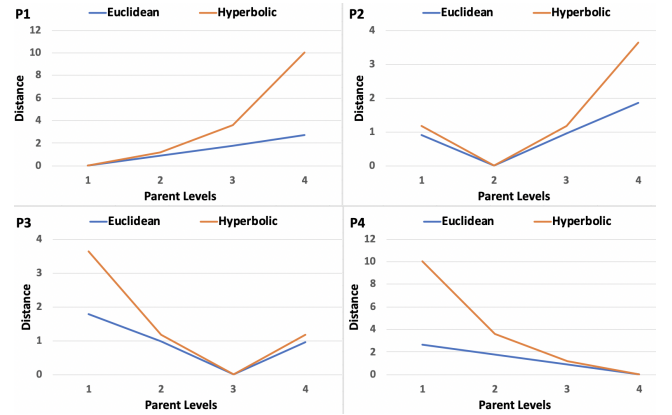


Figure 11: Inter-level Euclidean (Q2B) and Hyperbolic (HypE) distances. Each graph presents Δ_{inter} between entity pairs of a source hierarchical level, given by the graph label and the other hierarchy levels in the dataset, given on the x-axis.

C E-COMMERCE RESULTS

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

⁷Due to the sensitive nature of the proprietary dataset, we only report relative results.