

HYPON: Embedding Biomedical Ontology with Entity Sets

ZHUOYAN LI¹, Department of Computer Science, Purdue University, USA

SHENG WANG^{2*}, Paul G. Allen School of Computer Science & Engineering, University of Washington, USA

Constructing high-quality biomedical ontologies is one of the first steps to study new concepts, such as emerging infectious diseases. Manually curated ontologies are often noisy, especially for new knowledge that requires domain expertise. In this paper, we proposed a novel ontology embedding approach HYPON to automate this process. In contrast to conventional approaches, we propose to embed biomedical ontology in the hyperbolic space to better model the hierarchical structure. Importantly, our method is able to consider both graph structure and the varied-size set of entities, which are largely overlooked by existing methods. We demonstrated substantial improvement in comparison to thirteen comparison approaches on eleven biomedical ontologies, including two recently curated COVID-19 ontologies.

CCS Concepts: • **Computing methodologies** → **Neural networks**; • **Applied computing** → **Bioinformatics**; **Biomedical ontology**;

Additional Key Words and Phrases: deep learning, graph neural network, ontology embedding

ACM Reference Format:

Zhuoyan Li¹ and Sheng Wang². 2021. HYPON: Embedding Biomedical Ontology with Entity Sets. In *12th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics (BCB '21)*, August 1–4, 2021, Gainesville, FL, USA. ACM, New York, NY, USA, 7 pages. <https://doi.org/10.1145/3459930.3469515>

1 INTRODUCTION

Ontologies are crucial to biomedical research owing to the multi-scale and hierarchical structure of biomedical systems, encapsulating in processes such as evolution and cell differentiation [37]. An ontology uses controlled vocabulary to organize biomedical concepts of interests in a directed acyclic graph, where each edge represents the coarse to fine-grained relationship between concepts (Figure 1(b)). In addition to serving as human annotation standards [2, 6], ontologies are broadly used in downstream tasks such as classification [8, 21, 33], data integration [10, 23, 42], and clustering [11, 35]. Therefore, constructing high-quality ontologies in the domain of interest has become the first step to study new biological concepts including emerging infectious diseases. However, manual curation is not only time-consuming but also noisy, especially for new concepts that are poorly understood even by domain experts, thus presenting substantial challenges for utilizing these ontologies.

*Corresponding author

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

BCB '21, August 1–4, 2021, Gainesville, FL, USA

© 2021 Association for Computing Machinery.

ACM ISBN 978-1-4503-8450-6/21/08...\$15.00

<https://doi.org/10.1145/3459930.3469515>

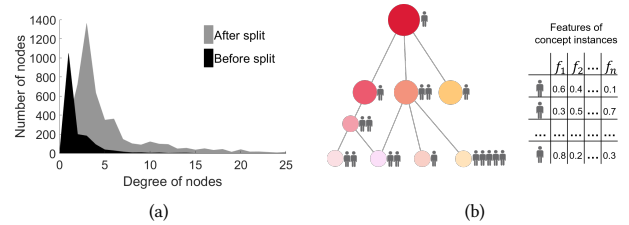


Fig. 1. **a.** Plot showing the distribution of node degrees in the COVID-19 Ontology before splitting nodes and after splitting nodes. The increase in node degrees motivates us to model the ontology graph in the hyperbolic space. **b.** Illustration of a biomedical ontology. A biomedical ontology is a directed acyclic graph which consists of the coarse to fine-grained relationship between concepts. Each concept (node) is associated with a varied-size set of entities (e.g., patients), which presents new challenge for conventional graph-based approaches. Each entity is represented by a feature vector.

In this paper, we aimed to embed biomedical ontology to alleviate such noise by formalizing the problem as a link prediction task on the ontology graph. Link prediction has been extensively studied to mitigate noisy and missing links in general graph analysis and ontology analysis [17, 20, 36]. The state-of-the-art approaches were developed based on network embedding [13] and graph convolutional neural networks [19], which embedded nodes into the low-dimensional space and then reconstructed links according to low-dimensional representations.

However, directly adapting these methods to the ontology graph overlooks two key properties of the ontology graph. Firstly, the ontology graph is a directed acyclic graph whose node degrees follow a power-law distribution (Figure 1(a)). This power-law distribution poses unique challenges for euclidean embedding methods that often do not have enough representation capacity to model the hierarchical structure [28]. To address this challenge, hyperbolic embedding methods have recently been proposed and demonstrated encouraging improvement over euclidean methods on such graphs [4, 22, 26–28]. Secondly, each node in the ontology is associated with a set of entities, such as patient samples in the disease ontology (Figure 1(b)). Due to the coarse to fine-grained structure of the ontology, the number of entities associated with each node is inevitably different, thus making the ontology graph fundamentally distinct from other graphs. We further noticed that the number of entities tends to be larger for nodes that are lower in the ontology hierarchy, which further amplifies the first challenge. Although graph convolutional neural networks can jointly model node information and graph structure, they can only handle the varied-size set of entities by averaging instance features, thus becoming less expressive for nodes with a large number of entities.

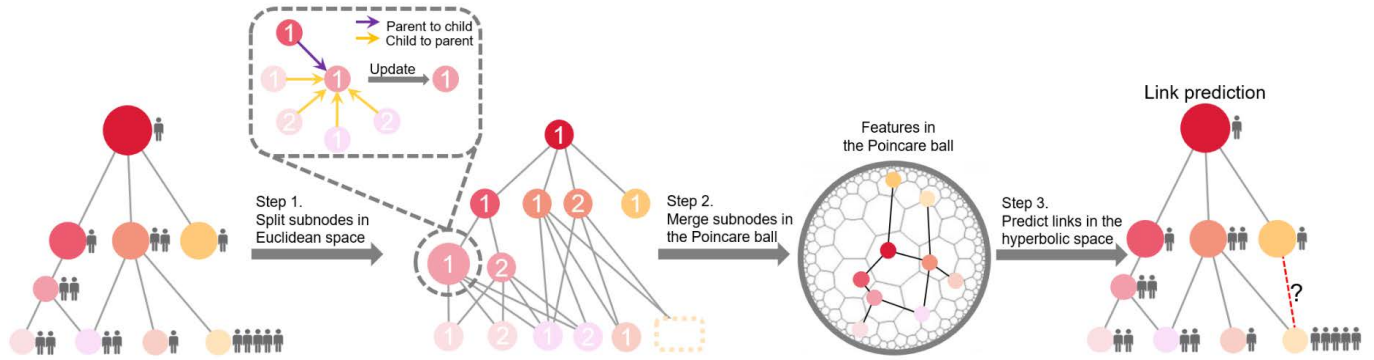


Fig. 2. **Flowchart of HYPON.** HYPON takes an ontology graph as input. Each node on the graph is associated with a varied-size set of entities (e.g., patients). HYPON first splits each node into subnodes according to the number of entities. It then performed a bidirectional message passing to aggregate information from parent nodes and child nodes separately. These subnodes are then merged together according to their embeddings in the Poincaré ball. Finally, HYPON predicts new links in the hyperbolic space using merged embeddings.

Present work. We proposed Hyperbolic Ontology neural Network (HYPON) to collectively address these challenges. HYPON embedded a biomedical ontology into the hyperbolic space with the consideration of both graph structure and the varied-size set of entities (Figure 2). HYPON consisted of a split step and a merge step. In the split step, each node was split into multiple new subnodes according to the size of the varied-size set. Since the new graph introduced larger node degrees, HYPON further embedded this graph in the hyperbolic space rather than the euclidean space. To the best of our knowledge, none of the existing graph analysis methods considered this problem of the varied-size set of entities and the arguable simplest approach is to average these features. After embedding nodes in the hyperbolic space, HYPON then merged the embeddings of subnodes together by using a permutation equivariant function. The merged embeddings are used in the downstream link prediction task. We performed experiments on eleven biomedical ontologies including two COVID-19 ontologies, and observed substantial improvement in comparison to the state-of-the-art approaches.

Our contributions can be summarized as follows:

- 1) *Conceptual:* We proposed to consider the varied-size set of entities on the ontology graph. Our proposed method HYPON is the first method that effectively models such graphs.
- 2) *Methodological:* We developed a novel method hyperbolic ontology neural network that could model varied-size set of entities associated nodes in the ontology.
- 3) *Experimental:* Our method substantially outperformed 13 comparison approaches on 11 biomedical ontologies including 2 COVID-19 ontologies by embedding the ontology into the hyperbolic space.
- 4) *COVID-19 impact:* Our method serves an important tool to accelerate the labor intensive curation of COVID-19 ontologies, which is the key step for COVID-19 research.

2 RELATED WORK

We reviewed three broad research areas that are closest to our problem setting. The key advance of our method HYPON in comparison to prior methods is the ability to model the varied-size set of entities.

Graph neural network. Graph neural Network (GNN) [7, 19, 38] has obtained the state-of-the-art performance in graph analysis through directly operating neural networks on the graph structure. For example, GCN used spectral-based graph convolution to perform classification in the graph structured data [19]. However, conventional GNNs have limited representative capacity when operating in the euclidean space, especially for hierarchical graphs that have low Gromov’s δ hyperbolicity [12]. To address this problem, hyperbolic space has been explored as an alternative to euclidean space for embedding hierarchical data. Nickel et al. proposed Poincaré embedding to learn representations of hierarchical data [28]. They further extended Poincaré embedding to Lorentz embedding which shares a more numerical stability in hyperbolic space [29]. These shallow hyperbolic methods exhibited substantially improved embedding performance over euclidean counterparts, especially when the number of dimension is small. Recently, HGCN [4] and HGNN [22] were proposed concurrently to embed graphs into the hyperbolic space and obtained substantial improvement in comparison to euclidean embedding. HGCN used attention operation for nodes aggregation in hyperbolic space and obtained substantial improvement in node classification and link prediction on graphs with low hyperbolicity [12]. HGNN generalized spectral-based graph convolution [19] from euclidean space into hyperbolic space and demonstrated promising performance in the analysis of molecule and blockchain transaction.

Despite the promising performance of these methods, none of them is able to effectively model the varied-size set of concept instances associated with each node in the ontology graph. We addressed this problem by introducing a permutation equivariant layer to model these sets and then performing message passing between sets on the graph. This set aggregation technique enables HYPON to outperform the best comparison approach by an average of 29.98% error reduction.

Learning sets of elements. Modeling sets using neural networks has gain increased interests. For example, Vinyals et al. [39] combined the seq2seq frame with chain rule to cope with disordered sets input. Zaheer et al. [41] constrained the standard feed-forward layer to be invariant to permutation and proposed a simple but

effective equivariant layer. PointNet [32] tackled with points cloud problem by stacking multiple permutation-equivariant layers and achieved the success in 3D classification and segmentation. Wagstaff et al. [40] provided a theoretical analysis for the limitation conditions of permutation invariant functions adopted in [41]. Recently, the symmetric element layer [25] generalized permutation equivariant layer [41] to model the symmetric sets and show substantial improvements in classification.

Existing efforts on learning sets of elements have not been developed into the context of graphs where nearby sets should have similar representation. To the best of our knowledge, HYPON is the first method that models sets in the context of the ontology graph. HYPON leveraged novel propagation and aggregation operations to model sets in the ontology and obtained improved performance in embedding ontology.

Ontology Embedding. In contrast to knowledge graphs (KG), constructing ontologies requires more intensive human efforts to uncover the hierarchical relationship between concepts, thus often leading to more substantial noise [30]. Several recent works extended embedding techniques to handle ontologies [16, 17, 20, 36]. Several recent works extended embedding techniques to handle ontologies. Chen et al. [5] proposed On2Vec which applied translational embeddings in the ontology to find relationship and hierarchies. Schockaert et al. [14] measured second-order proximity of ontological concepts. JOIE [16] jointly embedded ontological concepts and instances to capture the hierarchy of the ontology and the structured information of instance simultaneously by integrating ontology with a KG between instances. The difference between our work and JOIE is that the majority of the entities are only associated with one node in the ontology, making it impractical to construct a graph between entities. In addition, current efforts embedded ontology in euclidean space, which might own limited representation power for hierarchical structure. Two recent works proposed to embed ontology in the hyperbolic space [1, 24]. However, these works only utilized the structure of the ontology, whereas our proposed HYPON embedded the structure of ontology and features of nodes simultaneously in the hyperbolic space. We showed in the experiment that by embedding ontology in the hyperbolic space, HYPON outperformed methods in euclidean space by at least 37.3% of error reduction. Moreover, we observed that HYPON outperformed methods that do not considering node features by at least 53.1% of error reduction.

3 METHODS

We first reviewed the key concepts of the hyperbolic geometry to provide essential background of our method. For a more detailed description of hyperbolic geometry, please refer to [34]. We then introduced how our method embedded ontology with a varied-size set of entities into the hyperbolic space.

3.1 Hyperbolic Geometry Revisited

Poincaré ball model $(\mathcal{B}^d, g_{\mathbf{x}}^{\mathcal{B}})$ is a basic hyperbolic model. All points of the geometry are in the d -dimensional unit ball \mathcal{B}^d which is defined as:

$$\mathcal{B}^d = \{\mathbf{x} \in \mathbb{R}^d : \|\mathbf{x}\| < 1\}, \quad (1)$$

where $\|\cdot\|$ is the Euclidean norm, d is the dimension of the Poincaré ball and \mathbf{x} is a point in a d -dimension Poincaré ball. $g_{\mathbf{x}}^{\mathcal{B}}$ is the Riemannian metric tensor, which can be used to perform gradient descent in the Poincaré ball [28]. $g_{\mathbf{x}}^{\mathcal{B}}$ can be calculated as:

$$g_{\mathbf{x}}^{\mathcal{B}} = \lambda_{\mathbf{x}}^2 g_{\mathbf{x}}^E, \quad (2)$$

where $\lambda_{\mathbf{x}} = \frac{2}{1-\|\mathbf{x}\|^2}$ is the conformal factor. $g_{\mathbf{x}}^E = I_d$ is the corresponding Euclidean metric tensor, which is the dot product in the Euclidean space.

Given two points $\mathbf{x}, \mathbf{y} \in \mathcal{B}^d$, the distance between them on the Poincaré Ball can be calculated as:

$$d_{\mathcal{B}}(\mathbf{x}, \mathbf{y}) = \text{arcosh} \left(1 + 2 \frac{\|\mathbf{x} - \mathbf{y}\|^2}{(1 - \|\mathbf{x}\|^2)(1 - \|\mathbf{y}\|^2)} \right). \quad (3)$$

As \mathbf{x} is close to the boundary of the Poincaré ball ($\|\mathbf{x}\| \rightarrow 1$), the distance between \mathbf{x} and \mathbf{y} will be at infinite distance. This distance metric is therefore the key for the Poincaré ball model to embed hierarchical graphs with a low distortion in the hyperbolic space [28].

The logarithmic map $\log_{\mathbf{x}}(\mathcal{B}^d \rightarrow \mathcal{T}_{\mathbf{x}}\mathcal{B}^d)$ is used to map the point \mathbf{y} from the Poincaré ball \mathcal{B}^d to the tangent vector in the tangent space of the point \mathbf{x} , and the exponential map $\exp_{\mathbf{x}}(\mathcal{T}_{\mathbf{x}}\mathcal{B}^d \rightarrow \mathcal{B}^d)$ is used to map the point from the tangent vector in the tangent space of the point \mathbf{x} to that Poincaré ball.

$$\log_{\mathbf{x}}(\mathbf{y}) = \frac{2}{\lambda_{\mathbf{x}}} \text{arctanh}(\|-\mathbf{x} \oplus \mathbf{y}\|) \frac{-\mathbf{x} \oplus \mathbf{y}}{\|-\mathbf{x} \oplus \mathbf{y}\|}, \quad (4)$$

$$\exp_{\mathbf{x}}(\mathbf{v}) = \mathbf{x} \oplus \left(\tanh\left(\frac{\lambda_{\mathbf{x}}\|\mathbf{v}\|}{2}\right) \frac{\mathbf{v}}{\|\mathbf{v}\|} \right), \quad (5)$$

where $\mathbf{x} \in \mathcal{B}^d$ and $\mathbf{y} \in \mathcal{B}^d$ are points in the Poincaré ball and $\mathbf{v} \neq 0$ is the tangent vector in the tangent space of \mathbf{x} . $\log_{\mathbf{x}}(\mathbf{y})$ denotes the mapping of \mathbf{y} into the tangent space of point \mathbf{x} . \oplus denotes Möbius addition, which is analogous to addition in euclidean space. The Möbius addition for $\mathbf{x}, \mathbf{y} \in \mathcal{B}^d$ is defined as:

$$\mathbf{x} \oplus \mathbf{y} = \frac{(1 + \langle \mathbf{x}, \mathbf{y} \rangle + \|\mathbf{y}\|^2)\mathbf{x} + (1 - \|\mathbf{x}\|^2)\mathbf{y}}{1 + 2\langle \mathbf{x}, \mathbf{y} \rangle + \|\mathbf{x}\|^2\|\mathbf{y}\|^2}, \quad (6)$$

where $\langle \rangle$ is the inner product that $\langle \mathbf{x}, \mathbf{x} \rangle = \|\mathbf{x}\|^2$.

By leveraging (4) and (5), the linear transformation of the hyperbolic point $\mathbf{x} \in \mathcal{B}^d$ [9] can then be calculated as:

$$W \otimes \mathbf{x} = \exp_{\mathbf{o}}(W \log_{\mathbf{o}}(\mathbf{x}) + \mathbf{b}) \quad (7)$$

where \mathbf{o} is the origin in poincaré ball, W is the weight matrix and \mathbf{b} is the bias. Next, we introduced our proposed method HYPON.

3.2 Constructing new graph by splitting node features

As illustrated in Figure 1(b), a key feature of the biomedical ontology is that each node is associated with a varied-size set of entities, such as a set of patients or keyword tags. Each entity (e.g., a patient) is then associated with a feature vector. A simple feature concatenation is often impractical due to the large total number of entities. To address this challenge, we proposed to first construct a new graph by splitting each node into new subnodes, each corresponding to one entity.

Formally, let $\mathcal{G} = \{\mathcal{V}, \mathcal{E}, \mathcal{X}\}$ as a directed acyclic graph, where \mathcal{V} is the node set, $\mathcal{E} = \{(u, v) | (u, v) = 1, u, v \in \mathcal{V}\}$ is the edge

set and $\mathcal{X} = \{X_v | v \in \mathcal{V}\}$ contains the set of entities of each node. $X_v = \{\mathbf{x}_v^1, \mathbf{x}_v^2, \dots, \mathbf{x}_v^{n_v}\}$ is the set of entities of node v . n_v is the number of entity associated with node v . Each entity \mathbf{x}_v^i is associated with a d -dimensional vector. For simplicity, we use \mathbf{x}_v^i to denote the d -dimensional vector for the i -th entity of node v .

We aim at learning a function f that maps nodes $v \in \mathcal{V}$ with feature sets X_v to hyperbolic space:

$$f(\mathcal{V}, \mathcal{E}, \mathcal{X}) = \mathcal{H} \in \mathbb{R}^{|\mathcal{V}| \times d'}, \quad (8)$$

where $|\mathcal{V}|$ denotes the number of nodes, $\mathcal{H} = \{\mathbf{h}_v | v \in \mathcal{V}\}$ is the set of hyperbolic embeddings and $d' \ll d$. The hyperbolic embeddings \mathcal{H} are expected to represent hierarchical structure information and the varied-size set of entity information. Then \mathcal{H} can be used in downstream tasks such as link prediction.

We propose to construct this graph by splitting the graph \mathcal{G} at the node level so that each node $v \in V$ is split into several subnodes $\{v^1, v^2, \dots, v^{n_v}\}$, where each subnode v^i corresponds to a single feature \mathbf{x}_v^i . In addition, subnode v^i is connected to all subnodes of the child nodes of v . Accordingly, two subnodes u^j and v^i will have a link $(u^j, v^i) = 1$ if $(u, v) \in \mathcal{E}$.

Thus we construct a new expanded graph $\mathcal{G}' = \{\mathcal{V}', \mathcal{E}', \mathcal{X}'\}$, where \mathcal{V}' is the set of subnodes, $\mathcal{E}' = \{(u^j, v^i) | (u, v) \in \mathcal{E}, u^j, v^i \in \mathcal{V}'\}$ is the set of expanded edges and $\mathcal{X}' = \{\mathbf{x}_v^i | v^i \in \mathcal{V}'\}$ is the feature set of the subnodes.

3.3 Bidirectional message passing on subnodes

Since the ontology graph is a directed acyclic graph, we proposed a bidirectional aggregation approach to perform message passing from child nodes to parent nodes and parent nodes to child nodes separately. We define $\mathcal{N}_c(v^i)$ as the set of child nodes and $\mathcal{N}_p(v^i)$ as the set of parent nodes for node v^i . Given the hyperbolic features $\mathbf{x}_v^i \in \mathcal{X}'$, we calculated the hyperbolic attention [4] to compute attention weights:

$$\alpha_{u^j} = \text{SOFTMAX}_{u^j \in \mathcal{N}_c(v^i)} (W_c(\log_o(\mathbf{x}_v^i) || \log_o(\mathbf{x}_u^j))) \quad (9)$$

$$\alpha_{q^j} = \text{SOFTMAX}_{q^j \in \mathcal{N}_p(v^i)} (W_p(\log_o(\mathbf{x}_v^i) || \log_o(\mathbf{x}_q^j))) \quad (10)$$

we then perform bidirectional attention-based aggregation in the tangent space of \mathbf{x}_v^i to update \mathbf{x}_v^i as:

$$\mathbf{x}_v^i = \exp_{\mathbf{x}_v^i} \left(\sum_{u^j \in \mathcal{N}_c(v^i)} \alpha_{u^j} \log_{\mathbf{x}_v^i}(\mathbf{x}_u^j) + \sum_{q^j \in \mathcal{N}_p(v^i)} \alpha_{q^j} \log_{\mathbf{x}_v^i}(\mathbf{x}_q^j) \right), \quad (11)$$

where $||$ is the concatenation of two features and W_c, W_p are the trainable attention matrices of aggregation from child nodes and parent nodes, respectively.

3.4 Set-based aggregation

A key step of HYPON is to aggregate set information from subnodes. Although such aggregation has been extensively studied in euclidean space, it has not been applied to hyperbolic space yet. Fortunately, since the tangent space of a point in Poincaré ball model is always euclidean or a subset space of euclidean space, we generalize

the euclidean permutation equivariant functions in the Poincaré ball model. We firstly define the basic hyperbolic operations on feature set X_v of the node v :

$$\log_o(X_v) = \{\log_o(\mathbf{x}_v^1), \log_o(\mathbf{x}_v^2), \dots, \log_o(\mathbf{x}_v^{n_v})\}. \quad (12)$$

Then with hyperbolic linear transformation (7), we merge subnodes $\{v^1, v^2, \dots, v^{n_v}\}$ into a fixed dimension vector \mathbf{h}_v in the Poincaré ball ($X_v \rightarrow \mathbf{h}_v$) as:

$$\mathbf{h}_v = \sigma(\exp_o(\sum_{i=1}^{n_v} \Lambda \log_o(\mathbf{x}_v^i) - \Gamma \text{Avgpool}(\log_o(X_v)))). \quad (13)$$

where Λ and Γ are the trainable weight matrices. We leverage \log_o to map the hyperbolic representation \mathbf{x}_v^i of each subnode into the tangent space of the origin \mathbf{o} . Then we apply permutation equivariant function in X_v there.

3.5 HYPON architecture and link prediction

Finally, we summarize the architecture of HYPON and introduce how it can be used to perform link prediction. Given a graph $\mathcal{G} = (V, \mathcal{E}, \mathcal{X})$, we firstly split the nodes $v \in V$ into several subnodes $\{v^1, v^2, \dots, v^{n_v}\}$ and generate a new graph $\mathcal{G}' = (V', \mathcal{E}', \mathcal{X}')$. HYPON utilize (11) to propagate information between subnodes in the new graph \mathcal{G}' . Then the hyperbolic permutation equivariant layer is used to merge the feature set X_v into a fixed dimension feature \mathbf{h}_v^0 :

$$\mathbf{h}_v^0 = \sigma(\exp_o(\sum_{i=1}^{n_v} \Lambda \log_o(\mathbf{x}_v^i) - \Gamma \text{Avgpool}(\log_o(X_v)))). \quad (14)$$

HYPON finally stacks l hyperbolic graph convolution layers [22] as:

$$\mathbf{h}_v^l = \sigma(\exp_o(\sum_{u \in \mathcal{N}(v)} \tilde{A}_{uv} W^l \log_o(\mathbf{h}_u^{l-1}))), \quad (15)$$

where \mathbf{h}_v^l means the output of the hyperbolic graph convolution layer l . \tilde{A} is the normalized adjacent matrix.

Following previous works [4, 28], we predict the probability of an edge between nodes (u, v) by the Fermi-Dirac decoder [28]:

$$p((u, v) = 1 | \mathbf{h}_u^l, \mathbf{h}_v^l) = \frac{1}{e^{(d_{\mathcal{B}}(\mathbf{h}_u^l, \mathbf{h}_v^l)^2 - r)/t} + 1}, \quad (16)$$

where $d_{\mathcal{B}}(\cdot, \cdot)$ is the Poincaré distance and $r, t > 0$ are hyperparameters.

To investigate what kind of ontology can be better modelled by our method, we propose a growth rate score based on the increase of the number of entities. Intuitively, the ontology graph with large degree or more entities can be better embedded in the hyperbolic space.

Following Gromov's δ -hyperbolicity defined for the structure of graph, we first give the score function to represent the growth rate θ as the increase of number of entities. With $\mathcal{G} = \{\mathcal{V}, \mathcal{E}, \mathcal{X}\}$, growth rate θ is calculated as:

$$\theta = 10^{\frac{1}{|\mathcal{V}|} \sum_{u \in \mathcal{V}} \frac{n_u}{\lambda} \left(\frac{1}{|\mathcal{V}|} \sum_{u \in \mathcal{V}} \left(\frac{1}{|\mathcal{N}(u)|} \sum_{v \in \mathcal{N}(u)} \frac{n_v}{n_u} \right)^{-1} \right)} \quad (17)$$

where $|\mathcal{V}|$ is the number of nodes in \mathcal{V} , $|\mathcal{N}(u)|$ is the number of neighbour nodes, n_u is the number of a single feature in set X_u and λ is the attenuation coefficient and we set it as 2 in this paper. In

the experiment section, we showed that improvement of HYPON is correlated with growth rate θ .

Table 1. Table comparing the performance of HYPON with comparison approaches on two COVID-19 ontologies, in terms of AUROC (%). *hyperbolic*, *graph* and *feature*, represent embedding in hyperbolic space, considering graph structure, and considering entity features, respectively. The **best method** in each column is colored in blue and **second best** is colored in light blue. Error reduction is defined as the improvement of our AUROC over the AUROC of the best comparison approach. HYPON outperforms all comparison approaches with an average error reduction of 26.35% under 85% training percent and 10.4% under 50% training percent.

Dataset			COVID-19 Ontology		IDO-COVID19	
Growth rate θ			$\theta = 11.39$		$\theta = 4.02$	
Nodes			2278		395	
Training percent (%)			85	50	85	50
Poincaré	✓	✓	76.9	71.8	61.1	57.8
Euclidean		✓	73.2	72.9	52.1	50.1
Feature only			51.7	52.5	56.8	54.5
GCN		✓	74.8	63.3	58.1	56.2
GAT		✓	73.2	64.2	57.1	55.1
SAGE		✓	74.0	65.3	56.8	55.2
HNN	✓		76.1	70.1	60.2	53.4
MLP		✓	78.1	66.9	59.1	57.8
OWL2Vec		✓	74.8	72.6	65.2	58.1
Opa2Vec		✓	76.2	74.1	66.1	58.6
CANE		✓	76.8	74.9	65.8	59.6
HGCN	✓	✓	78.5	72.1	63.6	61.1
HGNN	✓	✓	76.5	73.1	62.8	60.1
Ours	✓	✓	83.3	76.2	73.9	64.7
Error Reduction (%)			-24.4%	-11.5%	-28.2%	-9.3%

4 EXPERIMENTAL RESULTS

4.1 Datasets

We collected eleven biomedical ontologies [37], which span 23203 biological concepts encompassing a variety of biological domains including cell types, diseases, chemical methods, postnatal stages, physiological and morphological measurements, and research instruments. Importantly, two of these eleven ontologies are recently developed to accelerate COVID-19 related research. The other nine general biomedical ontologies are denoted as ENRO, CHMO, CL, CMO, EMAPA, PO, TO, WBPH, XCO, using their corresponding names in the Open Biomedical Ontology Foundry [37]. We exploited the concept term to define the entities, where each word is one entity. Although we used text information as entity here, our method can be applied to other entities such as patient sample and side effect treatments.

4.2 Comparison approaches

We extensively compared our proposed method with thirteen comparison approaches. For shallow methods, we consider Euclidean embedding, Poincaré embedding [28] and *Feature only*. For feature-based methods, we consider Multi-Layer Perceptron (MLP) and Hyperbolic Neural Network (HNN) [9]. For Euclidean GNN models, we consider GCN [19], GraphSAGE (SAGE) [15] and Graph Attention Network (GAT) [38]. For hyperbolic GNN models, we

consider Hyperbolic Graph convolutional Network (HGCN) [4] and Hyperbolic Graph Neural network (HGNN) [22]. For ontology-based method, we consider OWL2Vec [17], Opa2vec [36] and CANE [20]. *Feature only* predicted the link between two nodes by calculating the Jaccard similarity of the two corresponding entity sets, which is considered as the probability. We summarized the three major difference aspects among these methods in Table 1, including the consideration of network structure, the consideration of entity features and whether embedded in hyperbolic space.

4.3 Training details

We used word embedding vectors as the feature of each word instance. In particular, we adopted 300-dim Glove [31] trained on the Wikipedia dataset. For the node feature which contains multiple words, we computed the average embeddings of all words as the input embeddings for baselines. For all methods, we performed a hyper-parameter search on a validation set over initial learning rate, weight decay, dropout rate, etc. Empirically, r and t in equation (16) were set to 2 and 1 respectively. Except shallow methods, we used 2 network layers with the same number of hidden units for HYPON and other baselines, and the dimension of final obtained embeddings was set to 16 for all methods for a fair comparison. The Poincaré embedding and HYPON were optimized with RiemannianSGD [3]. For OWL2Vec, Opa2Vec and CANE, we used their official implementations provided. The remaining methods were optimized by Adam [18] with default settings. Due to that links in most of ontologies are sparse, we added a super-node which connects to all original nodes in the ontology graph so that the graph still remains connected after removing links in experiments. Then we randomly split edges into 85%/5%/10% and 50%/5%/45% for training, validation and test sets respectively then repeated this process 5 times. The average scores of these 5 replicates were reported for all methods. Following previous work [4], we evaluated the link prediction with AUROC.

4.4 Improved link prediction

Overall, we observed that HYPON substantially outperformed the best comparison approaches on all eleven datasets with an average error reduction of 29.9% when using 85% training edges and an error reduction of 16.7% under 50 % training percent. Below, we summarized the three key observations we obtained from the comparison.

Hyperbolic embedding was better than euclidean embedding. We found that hyperbolic approaches such as HGCN, HNN and our method were in general better than euclidean approaches like GCN, GAT. We attribute such improvement to the hierarchical structure of ontologies, which can be more effectively represented in the hyperbolic space. Since our method further splits nodes based on the feature size, the average degree of each node is increased accordingly, resulting in better improvement by using a hyperbolic embedding. As a proof of concept, we found that our method achieved a at least error reduction of 12.2% in comparison to other hyperbolic approaches with 85% training edges. Moreover, we found that the improvement of our method is significantly correlated with the growth rate θ of each dataset, which again verified the merit of our approach.

Table 2. Table comparing the performance of HYPON with comparison approaches on nine biomedical ontologies, in terms of AUROC (%). The best algorithm in each column is colored in blue and second best is colored in light blue. HYPON outperforms all comparison approaches with an average error reduction of 33.1% under 85% training percent and 18.2 % under 50% training percent.

Dataset	ERO		CHMO		CL		CMO		EMAPA		PO		TO		WBPH		XCO	
Growth rate θ	$\theta = 1.90$		$\theta = 2.91$		$\theta = 2.69$		$\theta = 3.31$		$\theta = 2.79$		$\theta = 1.97$		$\theta = 1.35$		$\theta = 1.46$		$\theta = 2.43$	
Nodes	1622		2796		2231		3035		4351		1653		1501		2534		627	
Training percent (%)	85%	50%	85%	50%	85%	50%	85%	50%	85%	50%	85%	50%	85%	50%	85%	50%	85%	50%
Poincaré	65.1	67.5	61.2	56.9	65.0	62.8	69.1	54.5	57.1	62.3	60.1	59.3	68.3	63.7	63.4	62.3	62.8	60.1
Euclidean	64.5	64.7	62.6	57.7	66.2	60.4	68.5	54.2	64.8	57.4	59.5	53.7	66.5	59.1	62.9	59.0	56.4	55.3
Feature only	52.3	53.6	48.4	44.8	50.7	56.4	58.9	54.7	58.3	53.9	54.8	57.2	58.7	60.3	57.9	58.7	49.6	54.7
GCN	66.4	65.9	67.3	66.6	64.6	67.9	84.4	78.9	81.3	72.6	80.1	69.9	82.9	81.2	81.1	77.4	62.8	65.5
GAT	63.5	62.8	70.9	68.4	65.2	64.7	85.3	76.4	82.1	74.8	76.7	71.4	81.7	78.4	82.3	79.6	64.2	63.8
SAGE	65.8	64.2	64.1	64.1	63.7	62.4	80.2	73.8	75.2	71.2	78.6	68.5	80.2	77.6	78.6	73.7	61.7	64.7
HNN	65.1	64.7	62.1	58.9	74.7	67.1	83.9	74.1	66.8	63.5	69.1	67.8	77.8	74.6	68.5	64.9	66.2	62.8
MLP	63.2	63.4	61.4	60.6	71.8	63.5	74.6	72.2	69.4	67.2	66.2	66.3	70.1	69.1	64.1	62.8	66.8	63.5
OWL2Vec	66.9	64.7	65.2	62.6	74.4	67.2	75.2	70.5	75.3	72.1	76.2	71.8	71.5	68.3	68.9	65.0	71.4	67.8
Opa2Vec	67.1	66.1	68.7	66.4	76.3	71.9	72.4	67.8	78.2	75.4	78.9	73.2	68.8	65.7	70.4	66.5	72.6	67.9
CANE	67.4	65.8	69.1	67.7	74.8	69.5	73.6	71.5	79.8	74.3	74.2	73.9	70.6	69.2	68.1	66.8	71.8	68.1
HGCN	66.0	65.5	66.4	70.7	81.1	75.6	88.1	82.5	81.9	75.8	83.9	75.6	81.1	80.2	84.5	82.6	67.8	68.7
HGNN	67.6	66.2	63.3	69.8	80.5	74.9	87.5	81.3	76.2	74.4	81.0	74.1	78.2	79.4	84.9	83.1	66.4	65.6
Ours	71.1	67.4	86.5	82.4	88.2	81.6	91.1	86.1	87.2	82.6	90.8	79.4	85.9	83.9	88.1	85.6	78.2	68.9
Error Reduction(%)	-12.2%	-3.6%	-54.8%	-39.9%	-38.1%	-26.1%	-22.6%	-20.6%	-28.9%	-28.1%	-56.6%	-15.6%	-26.7%	-14.3%	-21.4%	-14.7%	-20.4%	-0.6%

Set-based bidirectional aggregation was better than average pooling. One key feature of HYPON is its ability to handle graphs with varied-size entities using a set-based bidirectional aggregation. To evaluate the effect of this aggregation, we compare HYPON to methods that simply used average pooling to aggregate varied-size entities. To this end, we observed substantial improvement of HYPON in comparison to HNN, MLP, GCN, GAT, SAGE, HGCN and HGNN. For example, under 85% training percent, HYPON obtained 0.865 AUROC on CHMO, while none of the comparison approaches obtained an AUROC better than 0.709 AUROC.

Network structure improved performance. We compared HYPON with methods that do not utilize the network structure, including both hyperbolic embedding approaches and euclidean embedding approaches. HYPON again greatly outperformed these methods with at least 17.1% error reduction. Moreover, we found that hyperbolic embedding is better than euclidean embedding on these datasets, which further raise our confidence about the superior performance of hyperbolic embedding in embedding ontology. Lastly, we observed that HYPON is better than a pure feature based baseline *Feature only*, which again ensured the importance of network structure.

Table 3. AUROC(%) for link prediction on the COVID-19 ontology with training percent as 85 % in ablation studies. Tick ✓ denotes that the corresponding component is used.

Bidirectional Agg.				
Agg.	✓	✓	✓	✓
Hyperbolic PE layer				
Average pooling	✓		✓	
AUROC	81.4	82.1	82.9	83.3

4.5 Ablation study

To further understand the importance of each component in our method, we conducted an ablation study of HYPON with and without specific components, namely hyperbolic permutation equivariant layer (Hyperbolic PE layer) and set-based bidirectional hyperbolic aggregation of child nodes and parent nodes (Bidirectional Agg.), on the COVID-19 ontology dataset. We used the simple average (Average pooling) as the alternative for Hyperbolic PE layer and the ordinary hyperbolic aggregation (Agg.) for the alternative of bidirectional bidirectional hyperbolic aggregation.

As shown in Table 3., we observed that both Hyperbolic PE layer and bidirectional agg. bring the performance improvement over HYPON with simple average pooling and ordinary hyperbolic aggregation. For instance, our method obtained 0.833 AUROC, which outperformed the 0.814 AUROC when adopting simple average pooling and ordinary hyperbolic aggregation. Compared with simple average pooling, Hyperbolic PE layer conserved important feature information and modeled sets more compatible in the hyperbolic space. Furthermore, bidirectional hyperbolic aggregation outperformed ordinary aggregation, which confirmed our intuition of ontology that parent nodes and child nodes shared different weights in the message passing.

4.6 The geometry of embeddings

In the merge step, HYPON converted the varied-size set of entities to a fixed-dimension feature. We computed the average norm of fixed-dimension features after merging and compared it to the average norm of feature sets in CL. The former is 0.9468 and the latter is 0.7314. The average norm of fixed-dimension is larger than average norm of sets. This denoted that HYPON can locate nodes with a lower hierarchy near to the boundary of poincaré ball, where the distance between two points is close to infinite. Consequently, HYPON is able to better distinguish nodes that are lower in the hierarchy.

5 CONCLUSION

In this paper, we presented HYPON, a novel method that embedded biomedical ontologies into the hyperbolic space. To model the set of varied-size entities, HYPON spilt nodes according to the entities and then performed message passing on these entities. On a collection of eleven biomedical ontologies, HYPON outperformed all thirteen comparison approaches. Two of these ontologies are recently curated COVID-19 ontologies, which raised our confidence of utilizing HYPON to accelerate COVID-19 research. Looking into the future, we plan to evaluate HYPON on other entities, such as patient samples and side effect treatments. Moreover, it is also interesting to model ordered instances such as patient infected order during pandemic.

REFERENCES

- [1] Brett K Beaulieu-Jones, Isaac S Kohane, and Andrew I Beam. 2019. Learning Contextual Hierarchical Structure of Medical Concepts with Poincaré Embeddings to Clarify Phenotypes.. In *PSB*. World Scientific, 8–17.
- [2] Olivier Bodenreider. 2004. The unified medical language system (UMLS): integrating biomedical terminology. *Nucleic acids research* 32 (2004), D267–D270.
- [3] Silvere Bonnabel. 2013. Stochastic gradient descent on Riemannian manifolds. *IEEE Trans. Automat. Control* 58, 9 (2013), 2217–2229.
- [4] Ines Chami, Zhitao Ying, Christopher Ré, and Jure Leskovec. 2019. Hyperbolic graph convolutional neural networks. In *Advances in neural information processing systems*. 4868–4879.
- [5] Muhao Chen, Yingtao Tian, Xuelu Chen, Zijun Xue, and Carlo Zaniolo. 2018. On2vec: Embedding-based relation prediction for ontology population. In *Proceedings of the 2018 SIAM International Conference on Data Mining*. SIAM, 315–323.
- [6] Wasila Dahdul, Prashanti Manda, Hong Cui, James P Balhoff, T Alexander Decchi, Nizar Ibrahim, Hilmar Lapp, Todd Vision, and Paula M Mabey. 2018. Annotation of phenotypes using ontologies: a gold standard for the training and evaluation of natural language processing systems. *Database* 2018 (2018).
- [7] 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.
- [8] María del Mar Roldán-García, María Jesús García-Godoy, and José F Aldana-Montes. 2016. Dione: An OWL representation of ICD-10-CM for classifying patients' diseases. *Journal of biomedical semantics* 7, 1 (2016), 62.
- [9] Octavian Ganea, Gary Bécigneul, and Thomas Hofmann. 2018. Hyperbolic neural networks. In *Advances in neural information processing systems*. 5345–5355.
- [10] DG Gocheva, HM Eminova, and IA Batchkova. 2016. ONTOLOGY BASED DATA AND INFORMATION INTEGRATION IN BIOMEDICAL DOMAIN. *Machines. Technologies. Materials*. 10, 2 (2016), 35–38.
- [11] Rafael S Gonçalves, Maulik R Kamdar, and Mark A Musen. 2019. Aligning biomedical metadata with ontologies using clustering and embeddings. In *European Semantic Web Conference*. Springer, 146–161.
- [12] Mikhael Gromov. 1987. Hyperbolic groups. In *Essays in group theory*. Springer, 75–263.
- [13] Aditya Grover and Jure Leskovec. 2016. node2vec: Scalable feature learning for networks. In *Proceedings of the 22nd ACM SIGKDD*. 855–864.
- [14] Víctor Gutiérrez-Basulto and Steven Schockaert. 2018. From knowledge graph embedding to ontology embedding? An analysis of the compatibility between vector space representations and rules. *arXiv preprint arXiv:1805.10461* (2018).
- [15] Will Hamilton, Zhitao Ying, and Jure Leskovec. 2017. Inductive representation learning on large graphs. In *Advances in neural information processing systems*. 1024–1034.
- [16] Junheng Hao, Muhao Chen, Wenchao Yu, Yizhou Sun, and Wei Wang. 2019. Universal representation learning of knowledge bases by jointly embedding instances and ontological concepts. In *Proceedings of the 25th ACM SIGKDD*. 1709–1719.
- [17] Ole Magnus Holter, Erik Bryhn Myklebust, et al. 2019. Embedding OWL ontologies with OWL2Vec. In *CEUR Workshop Proceedings*. Technical University of Aachen.
- [18] Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).
- [19] Thomas N Kipf and Max Welling. 2016. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907* (2016).
- [20] Sotiris Kotitsas, Dimitris Pappas, et al. 2019. Embedding Biomedical Ontologies by Jointly Encoding Network Structure and Textual Node Descriptors. In *Proceedings of the 18th BioNLP Workshop*.
- [21] Young-suk Lee, Arjun Krishnan, Rose Oughtred, Jennifer Rust, Christie S Chang, Joseph Ryu, Vessela N Kristensen, Kara Dolinski, Chandra L Theesfeld, and Olga G Troyanskaya. 2019. A computational framework for genome-wide characterization of the human disease landscape. *Cell systems* 8, 2 (2019), 152–162.
- [22] Qi Liu, Maximilian Nickel, and Douwe Kiela. 2019. Hyperbolic graph neural networks. In *Advances in Neural Information Processing Systems*. 8230–8241.
- [23] Kevin M Livingston, Michael Bada, William A Baumgartner, and Lawrence E Hunter. 2015. KaBOB: ontology-based semantic integration of biomedical databases. *BMC bioinformatics* 16, 1 (2015), 1–21.
- [24] Qiuhao Lu, Nisansa de Silva, Yunyao Li, et al. 2019. Learning Electronic Health Records through Hyperbolic Embedding of Medical Ontologies. In *Proceedings of the 10th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics*. 338–346.
- [25] Haggai Maron, Or Litany, Gal Chechik, and Ethan Fetaya. 2020. On Learning Sets of Symmetric Elements. *arXiv preprint arXiv:2002.08599* (2020).
- [26] Yu Meng, Jiaxin Huang, Guangyuan Wang, Chao Zhang, Honglei Zhuang, Lance Kaplan, and Jiawei Han. 2019. Spherical text embedding. In *Advances in Neural Information Processing Systems*. 8208–8217.
- [27] Yu Meng, Yunyi Zhang, Jiaxin Huang, Yu Zhang, Chao Zhang, and Jiawei Han. 2020. Hierarchical Topic Mining via Joint Spherical Tree and Text Embedding. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*.
- [28] Maximilian Nickel and Douwe Kiela. 2017. Poincaré embeddings for learning hierarchical representations. In *Advances in neural information processing systems*. 6338–6347.
- [29] Maximilian Nickel and Douwe Kiela. 2018. Learning continuous hierarchies in the lorentz model of hyperbolic geometry. *arXiv preprint arXiv:1806.03417* (2018).
- [30] Jeff Pasternack and Dan Roth. 2013. Latent credibility analysis. In *Proceedings of the 22nd international conference on World Wide Web*. 1009–1020.
- [31] Jeffrey Pennington, Richard Socher, and Christopher D Manning. 2014. Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*. 1532–1543.
- [32] Charles R Qi, Hao Su, Kaichun Mo, and Leonidas J Guibas. 2017. Pointnet: Deep learning on point sets for 3d classification and segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 652–660.
- [33] Ahmet Sureyya Rifaioğlu, Tunca Doğan, Maria Jesus Martin, Rengul Cetin-Atalay, and Volkan Atalay. 2019. DEEPred: automated protein function prediction with multi-task feed-forward deep neural networks. *Scientific reports* 9, 1 (2019), 1–16.
- [34] Shun Onoda Ryota Suzuki, Ryusuke Takahama. 2019. Hyperbolic Disk Embeddings for Directed Acyclic Graphs. *arXiv preprint arXiv:1902.04335* (2019).
- [35] R Sandhiya and M Sundarambal. 2019. Clustering of biomedical documents using ontology-based TF-IGM enriched semantic smoothing model for telemedicine applications. *Cluster Computing* 22, 2 (2019), 3213–3230.
- [36] Fatima Zohra Smaili, Xin Gao, and Robert Hoehndorf. 2019. Opa2vec: combining formal and informal content of biomedical ontologies to improve similarity-based prediction. *Bioinformatics* 35, 12 (2019), 2133–2140.
- [37] Barry Smith, Michael Ashburner, Cornelius Rosse, et al. 2007. The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. *Nature biotechnology* 25, 11 (2007), 1251–1255.
- [38] Petar Veličković, Guillem Cucurull, et al. 2017. Graph attention networks. *arXiv preprint arXiv:1710.10903* (2017).
- [39] Oriol Vinyals, Samy Bengio, and Manjunath Kudlur. 2015. Order matters: Sequence to sequence for sets. *arXiv preprint arXiv:1511.06391* (2015).
- [40] Edward Wagstaff, Fabian B Fuchs, Martin Engelcke, Ingmar Posner, and Michael Osborne. 2019. On the limitations of representing functions on sets. *arXiv preprint arXiv:1901.09006* (2019).
- [41] Manzil Zaheer, Satwik Kottur, et al. 2017. Deep sets. In *Advances in neural information processing systems*. 3391–3401.
- [42] Hansi Zhang, Yi Guo, Qian Li, Thomas J George, Elizabeth Shenkman, François Modave, and Jiang Bian. 2018. An ontology-guided semantic data integration framework to support integrative data analysis of cancer survival. *BMC medical informatics and decision making* 18, 2 (2018), 41.