Enhancing Hyperbolic Graph Embeddings via Contrastive Learning

Jiahong Liu[1]*, **Menglin Yang**[2]*, **Min Zhou**[3]†, **Shanshan Feng**[1], **Philippe Fournier-Viger**[4]
[1]Harbin Institute of Technology(Shenzhen); [2]The Chinese University of Hong Kong;
[3]Huawei Technologies co.ltd; [4]Shenzhen University.
jiahong.liu21@gmail.com; mlyang@cse.cuhk.edu.hk; zhoumin27@huawei.com;
victor_fengss@foxmail.com; philfv8@yahoo.com

Abstract

Recently, hyperbolic space has risen as a promising alternative for semi-supervised graph representation learning. Many efforts have been made to design hyperbolic versions of neural network operations. However, the inspiring geometric properties of this unique geometry have not been fully explored yet. The potency of graph models powered by the hyperbolic space is still largely underestimated. Besides, the rich information carried by abundant unlabelled samples is also not well utilized. Inspired by the recently active and emerging self-supervised learning, in this study, we attempt to enhance the representation power of hyperbolic graph models by drawing upon the advantages of contrastive learning. More specifically, we put forward a novel Hyperbolic Graph Contrastive Learning (HGCL) framework which learns node representations through multiple hyperbolic spaces to implicitly capture the hierarchical structure shared between different views. Then, we design a hyperbolic position consistency (HPC) constraint based on hyperbolic distance and the homophily assumption to make contrastive learning fit into hyperbolic space. Experimental results on multiple real-world datasets demonstrate the superiority of the proposed HGCL as it consistently outperforms competing methods by considerable margins for the node classification task.

1 Introduction

Hyperbolic space, which is under the mathematical framework known as *Riemannian geometry* [1], has emerged as a promising alternative for semi-supervised graph representation learning recently [2, 3, 4]. Different from the Euclidean space which expands polynomially, the hyperbolic space grows exponentially with its radius which can then be regarded as a smooth version of trees as the number of nodes in a binary tree also grows exponentially with the depth [5]. Hence, it gains natural advantages in abstracting scale-free graphs with a hierarchical organization [6].

Inspired by the recently active and flourishing self-supervised learning, we aim at enhancing the representation power of hyperbolic graph models by drawing upon the advantages of contrastive learning. A contrastive learning algorithm commonly includes a loss function with a projection-powered (i.e., dot product) softmax function to maintain the consistency of the positives and negatives counterparts, which is found to be hardness-aware in nature and gives penalties to samples according to their hardness [7]. Besides, it is also found to be able to enforce extra intra-class more compactly and inter-class discrepancy simultaneously, leading to a better discriminative power of the models [8].

Although the contrastive concept has been successfully utilized in Euclidean models [9, 10, 11], its application in hyperbolic space is hindered by the following two **challenges**. First, the embeddings optimized by traditional measure (inner product) in Euclidean space are with regularized norm, which makes it impossible to be pushed far away from the origin to utilize the spacious advantage

*Equal contribution; Work mainly done during an internship at Huawei Noah's Ark Lab.
†Corresponding author.

35th Conference on Neural Information Processing Systems (NeurIPS 2021), Sydney, Australia.

of hyperbolic space. Second, the contrastive loss aims at preserving maximal information (i.e., *uniformity*) during training [7] by pushing all different instances apart and pulling positive pairs closer. This may destroy the prior structural relations of hierarchical datasets (i.e., *tolerance*) [12], which is detrimental to downstream tasks.

In this work, We introduce a hyperbolic graph contrastive learning framework (HGCL), which brings the benefits of contrastive learning into semi-supervised hyperbolic graph neural networks. Our contributions are summarized as follows:

- A hyperbolic position consistency (HPC) constraint based on hyperbolic geometry is proposed to accommodate the two challenges mentioned above. It includes a positive sampling strategy balancing the tolerance and uniformity, and a distance-aware discriminator to properly measure the embeddings in hyperbolic space.
- Extensive experiments show that the proposed method outperforms the baseline models by large margins in the node classification task. The ablation study and analysis further gives insights into how the proposal successfully produces high-quality embeddings.
- To the best of knowledge, this is the first attempt to bridge contrastive learning with hyperbolic graph learning, shedding light to the related research topic.

2 Preliminary

Hyperbolic Geometry. Hyperbolic geometry is a non-Euclidean geometry with a constant negative curvature. In this part, we briefly review the definitions and concepts on hyperbolic geometry. A thorough and in-depth explanation can be found in [13, 14].

There are multiple hyperbolic models with different definitions and metrics that are mathematically equivalent. We here mainly consider two widely studied ones: Poincaré ball model [15] and the Lorentz model (also known as the hyperboloid model) [16], which are defined by Definition 1 and 2, respectively. The related formulas and operations, e.g., distance, maps, and parallel transport are further summarized in Table 2 (Appendix. A.1), where $\oplus_K$ and $\mathrm{gyr}[.,.]$ are the Möbius addition [17] and gyration operator [17], respectively.

Definition 1 (Poincaré Ball Model) *The n-dimensional Poincaré ball model with negative curvature $K(K<0)$ is defined as a Riemannian manifold $(\mathbb{D}_K^n, g_{\mathbf{x}}^{\mathbb{D}})$, where $\mathbb{D}_K^n = \left\{\mathbf{x} \in \mathbb{R}^n : \langle \mathbf{x}, \mathbf{x}\rangle_2 < -\frac{1}{K}\right\}$ is an open n-dimensional ball with radius $1/\sqrt{-K}$. Its metric tensor $g_{\mathbf{x}}^{\mathbb{D}_K} = (\lambda_{\mathbf{x}}^K)^2 g^{\mathbb{E}}$, where $\lambda_{\mathbf{x}}^K = \frac{2}{1+K\|\mathbf{x}\|_2^2}$ is the conformal factor and $g^{\mathbb{E}} = I$ is the Euclidean metric.*

Definition 2 (Lorentz Model) *The n-dimensional Lorentz model (also named Hyperboloid model) with negative curvature $K(K<0)$ is defined as the Riemannian manifold $(\mathbb{H}_K^n, g_{\mathbf{x}}^{\mathbb{H}})$, where $\mathbb{H}_K^n = \left\{\mathbf{x} \in \mathbb{R}^{n+1} : \langle \mathbf{x}, \mathbf{x}\rangle_{\mathcal{L}} = \frac{1}{K}\right\}$ and $g_{\mathbf{x}}^{\mathbb{H}} = diag([-1, 1, ..., 1])$. $\langle \cdot, \cdot \rangle$ is the Lorentzian inner product. Let $\mathbf{x}, \mathbf{y} \in \mathbb{R}^{n+1}$, then the Lorentzian inner product is defined as:*

$$\langle \mathbf{x}, \mathbf{y}\rangle_{\mathcal{L}} := -x_0 y_0 + \sum_{i=1}^{n} x_i y_i. \tag{1}$$

It is worth noting that the tangent space at $\mathbf{x}$ is given by a n-dimensional vector space approximating $\mathbb{H}_K^n$, that is

$$\mathcal{T}_{\mathbf{x}}\mathbb{H}_K^n := \{\mathbf{v} \in \mathcal{R}^{n+1} : \langle \mathbf{v}, \mathbf{x}\rangle_{\mathcal{L}} = 0\}. \tag{2}$$

3 Metohd

The proposed HGCL aims to enhance semi-supervised hyperbolic graph embeddings via contrastive learning, which is sketched in Figure 1. First, we encode the graph by two hyperbolic graph neural networks (HGNNs), producing different hyperbolic views, instead of tedious and tricky graph augmentations. Then, a hyperbolic contrastive loss, namely, hyperbolic position consistency (HPC) L_{hpc} is further proposed to refine the embeddings, which is further decoded according to downstream

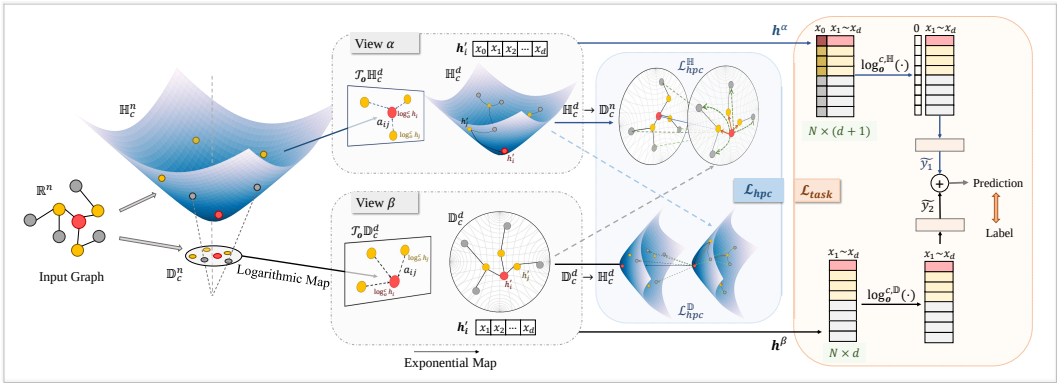
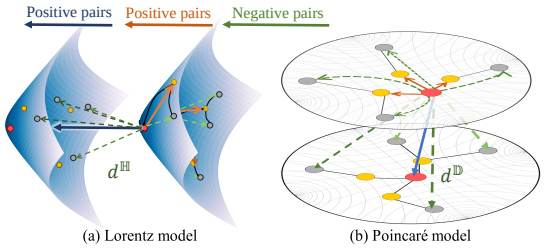


Figure 1: Framework of HGCL. The input graph in Euclidean space is firstly mapped into two hyperbolic spaces. (i) Encoder: simultaneously propagate neighbor information in two HGNNs and obtain two node embeddings, i.e. $\mathbf{h}^{\alpha}$ and $\mathbf{h}^{\beta}$; (ii) Hyperbolic Contrastive Loss: refine the embeddings by pulling the positions of semantically similar samples closer and meanwhile pushing away the negative samples; (iii) Decoder and Task-specific Loss Function: decode the embeddings and obtain the final classification result during training.

tasks and trained with the task-specific loss function. We will mainly introduce the HPC, which is the key of our method.

Hyperbolic Contrastive Loss: HPC. In contrastive learning, the negative samples are required to avoid model collapse, and the same nodes in different views are usually selected as positive samples to extract discriminative features. Then how to select the samples and evaluate the pairs' similarities is critical. Available contrastive algorithms are customized for Euclidean space and are not directly applicable in hyperbolic space as hyperbolic space possesses distinctive properties (e.g, hierarchical awareness and spacious room) compared with the Euclidean counterpart. In this work, we propose a hyperbolic contrastive loss, i.e., HPC, which takes the one-hop neighbors of the anchors (center node) as positive samples regarding the homophily assumption to strengthen the tolerance, i.e., maintain semantic information. It further adds a penalty to negative samples with the purpose of pushing embeddings away from the origin to utilize the ample space near the boundary. More importantly, a distance-aware discriminator is designed to properly measure the similarities in hyperbolic space.



(a) Lorentz model (b) Poincaré model

Figure 2: Samples selection to calculate mutual information in different manifolds. The same nodes are in red and the neighbors in orange. The negative samples are also from two sources, which are intra-view and inter-view nodes (in green).

(1) Sample selection. The main idea of sample selection is to enforce consistency between the encoded hyperbolic embeddings in different views and meanwhile maintain the semantic structure information. As illustrated in Figure 2, for each anchor (center node), the positive samples consists of two black parts, which are the same node in the other view (named consistent node) and one-hop neighbors in the same view (named tolerance node). To maximize the mutual information between similar nodes in principle, we define a local estimator $\mathrm{MI}(\cdot,\cdot)$ based on the Jensen-Shannon divergence that distinguishes the *positive* embeddings from *negative* node embeddings. Specifically, for each anchor $\mathbf{h}_i^{\alpha}$ in view α, the two parts of pairwise loss are defined by Equation (3) and Equation (4), respectively.

$$\mathrm{MI}\left(\mathbf{h}_i^{\alpha}, t_1(\mathbf{h}_i^{\beta})\right) = \mathbb{E}_{\mathbb{P}}\left[\log \mathcal{D}\left(\mathbf{h}_i^{\alpha}, t_1(\mathbf{h}_i^{\beta})\right)\right] + \lambda_n \cdot \sum_{j=1}^{m} \mathbb{E}_{\mathbb{P}\times\bar{\mathbb{P}}}\left[\log\left(1 - \mathcal{D}\left(\mathbf{h}_i^{\alpha}, t_1(\bar{\mathbf{h}}_j^{\beta})\right)\right)\right], \quad (3)$$

$$\mathrm{MI}\left(\mathbf{h}_i^{\alpha}, \mathcal{N}(\mathbf{h}_i^{\alpha})\right) = \sum_{\mathbf{h}_j \in \mathcal{N}(\mathbf{h}_i^{\alpha})} \mathbb{E}_{\mathbb{P}}\left[\log \mathcal{D}\left(\mathbf{h}_i^{\alpha}, \mathbf{h}_j\right)\right] + \lambda_n \cdot \sum_{j=1}^{m} \mathbb{E}_{\mathbb{P}\times\bar{\mathbb{P}}}\left[\log \mathcal{D}\left(1 - \left(\mathbf{h}_i^{\alpha}, \bar{\mathbf{h}}_j^{\alpha}\right)\right)\right], \quad (4)$$

where $\mathbf{h}_i^{\alpha}$ and $\mathbf{h}_i^{\beta}$ are embeddings of node v_i in views α and β, respectively; $\mathcal{N}(\mathbf{h}_i^{\alpha})$ and $\mathcal{N}(\mathbf{h}_i^{\beta})$ are the embeddings of adjacent nodes $\mathbf{h}_i^{\alpha}$ and $\mathbf{h}_i^{\beta}$ in the same view; t_1 and t_2 are the functions that transfer the vectors from the manifold of view α to the manifold of view β; m is the number of negative samples; λ_n is a hyperparameter that adds a penalty to negative samples.

(2) Distance-aware discriminator As proxy for maximizing the local MI in Equation (3) and (4), we employ a discriminator $\mathcal{D}(\cdot,\cdot) : \mathbb{R}^d \times \mathbb{R}^d \mapsto \mathbb{R}$, such that $\mathcal{D}(\mathbf{h}_i, \mathbf{h}_j)$ represents the probability scores assigned to the pair with two samples $\mathbf{h}_i$ and $\mathbf{h}_j$. Note that, we implement the discriminator as the **distance** between two representations in corresponding space, $d_{\mathbb{D}}^K(\mathbf{h}_i, \mathbf{h}_j)$ or $d_{\mathbb{H}}^K(\mathbf{h}_i, \mathbf{h}_j)$ (see Table.2) in Appendix. A.1, instead of simply using the dot product [18] or cosine similarity [19].

$$L_{hpc} = -\frac{1}{2n}\sum_{i=1}^{n}\left[\underbrace{\mathrm{MI}\left(\mathbf{h}_i^{\alpha}, t_1(\mathbf{h}_i^{\beta})\right) + \mathrm{MI}\left(\mathbf{h}_i^{\beta}, t_2(\mathbf{h}_i^{\alpha})\right)}_{P_1\text{: Consistency}} + \underbrace{\mathrm{MI}\left(\mathbf{h}_i^{\alpha}, \mathcal{N}(\mathbf{h}_i^{\alpha})\right) + \mathrm{MI}\left(\mathbf{h}_i^{\beta}, \mathcal{N}(\mathbf{h}_i^{\beta})\right)}_{P_2\text{: Tolerance}}\right]. \quad (5)$$

4 Experiments

Node classification. Following the standard practice and experimental setup in related work (HGCN) [2], we report the F1-score for DISEASE and AIRPORT datasets, and accuracy for the others in the node classification tasks. The statistics of datasets are listed in Appendix A.4. Table 1 shows the results, where the best records for each dataset have been marked in bold. Compared with baselines, the proposed HGCL achieves the best performance on all five datasets, indicating its powerful ability to embed graphs for node classification. It is thanks to the deployment of contrastive learning. Notably, our method not only performs well on datasets with lower hyperbolicity δ (e.g. DISEASE, AIRPORT), but also shows significant improvements on those with higher δ (e.g. CORA). Specifically, the accuracy is improved by 3.12% on DISEASE and 4.05% on CORA compared with the second best, i.e., HGAT [4]. What's more, the ablation study and further analysis are represented in Appendix A.2 and Appendix A.3, respectively.

Table 1: Comparisons of node classification (NC) in accuracy with the standard deviation

Dataset	DISEASE	AIRPORT	PUBMED	CITESEER	CORA
EUC	32.56 ± 1.19	60.90 ± 3.40	48.20 ± 0.76	61.28 ± 0.91	23.80 ± 0.80
HYP [15]	45.52 ± 3.09	70.29 ± 0.40	68.51 ± 0.37	61.71 ± 0.74	22.13 ± 0.97
MLP	28.80 ± 2.23	68.90 ± 0.46	72.40 ± 0.21	59.53 ± 0.90	51.59 ± 1.28
HNN [20]	41.18 ± 1.85	80.59 ± 0.46	69.88 ± 0.43	59.50 ± 1.28	54.76 ± 0.61
HGNN [3]	81.27 ± 3.53	84.71 ± 0.98	77.13 ± 0.82	69.99 ± 1.00	78.26 ± 1.19
HGCN [2]	88.16 ± 0.76	89.26 ± 1.27	76.53 ± 0.63	68.04 ± 0.59	78.03 ± 0.98
HGAT [4]	90.30 ± 0.62	89.62 ± 1.03	77.42 ± 0.66	68.64 ± 0.30	78.32 ± 1.39
HGCL(ours)	**93.42** ± 0.82	**92.35** ± 1.01	**79.14** ± 0.68	**72.11** ± 0.64	**82.37** ± 0.47

5 Conclusion

In this work, we brought the benefits of contrastive learning into hyperbolic graph learning to obtain more powerful representations. In particular, we feed the graph-structured data into two hyperbolic encoders to generate contrastive views. Then, the proposed contrastive loss HPC leverages the properties of hyperbolic space to minimize the distances of positive-paired embeddings and maximize the distances of negative-paired embeddings. The extensive experimental results show that the contrast-powered learning scheme successfully preserves the semantic and hierarchies of the dataset as it consistently outperforms the baselines across the diverse datasets and tasks. As far as we know, this is the first hyperbolic graph learning framework powered by contrastive learning, which proposes a new direction for the research community.

References

[1] Stephanie Alexander. Michael spivak, a comprehensive introduction to differential geometry. *Bulletin of the American Mathematical Society*, 84(1):27–32, 1978.

[2] Ines Chami, Zhitao Ying, Christopher Ré, and Jure Leskovec. Hyperbolic graph convolutional neural networks. In *NeurIPS*, pages 4868–4879, 2019.

[3] Qi Liu, Maximilian Nickel, and Douwe Kiela. Hyperbolic graph neural networks. In *NeurIPS*, pages 8230–8241, 2019.

[4] Yiding Zhang, Xiao Wang, Xunqiang Jiang, Chuan Shi, and Yanfang Ye. Hyperbolic graph attention network. In *AAAI*, 2019.

[5] Gregor Bachmann, Gary Bécigneul, and Octavian Ganea. Constant curvature graph convolutional networks. In *ICML*, pages 486–496. PMLR, 2020.

[6] Dmitri Krioukov, Fragkiskos Papadopoulos, Maksim Kitsak, Amin Vahdat, and Marián Boguná. Hyperbolic geometry of complex networks. *Physical Review E*, 82(3):036106, 2010.

[7] Tongzhou Wang and Phillip Isola. Understanding contrastive representation learning through alignment and uniformity on the hypersphere. In *ICML*, pages 9929–9939. PMLR, 2020.

[8] Peng Wang, Kai Han, Xiu-Shen Wei, Lei Zhang, and Lei Wang. Contrastive learning based hybrid networks for long-tailed image classification. In *CVPR*, pages 943–952, 2021.

[9] Sheng Wan, Shirui Pan, Jian Yang, and Chen Gong. Contrastive and generative graph convolutional networks for graph-based semi-supervised learning. In *AAAI*, volume 35, pages 10049–10057, 2021.

[10] Zhen Peng, Wenbing Huang, Minnan Luo, Qinghua Zheng, Yu Rong, Tingyang Xu, and Junzhou Huang. Graph representation learning via graphical mutual information maximization. In *WebConf*, pages 259–270, 2020.

[11] Yanqiao Zhu, Yichen Xu, Feng Yu, Qiang Liu, Shu Wu, and Liang Wang. Deep graph contrastive representation learning. *arXiv preprint arXiv:2006.04131*, 2020.

[12] Feng Wang and Huaping Liu. Understanding the behaviour of contrastive loss. In *CVPR*, pages 2495–2504, 2021.

[13] John M Lee. Smooth manifolds. In *Introduction to Smooth Manifolds*, pages 1–31. Springer, 2013.

[14] Maximillian Nickel and Douwe Kiela. Learning continuous hierarchies in the lorentz model of hyperbolic geometry. In *ICML*, pages 3779–3788, 2018.

[15] Maximillian Nickel and Douwe Kiela. Poincaré embeddings for learning hierarchical representations. In *NeurIPS*, pages 6338–6347, 2017.

[16] Maximillian Nickel and Douwe Kiela. Learning continuous hierarchies in the lorentz model of hyperbolic geometry. In *ICML*, pages 3779–3788. PMLR, 2018.

[17] Abraham A Ungar et al. The hyperbolic square and mobius transformations. *Banach Journal of Mathematical Analysis*, 1(1):101–116, 2007.

[18] Kaveh Hassani and Amir Hosein Khasahmadi. Contrastive multi-view representation learning on graphs. In *ICML*, pages 4116–4126. PMLR, 2020.

[19] Yuning You, Tianlong Chen, Yongduo Sui, Ting Chen, Zhangyang Wang, and Yang Shen. Graph contrastive learning with augmentations. *NeurIPS*, 33:5812–5823, 2020.

[20] Octavian Ganea, Gary Bécigneul, and Thomas Hofmann. Hyperbolic neural networks. In *NeurIPS*, pages 5345–5355, 2018.

A Appendix

A.1 Summary of Operations in Hyperbolic Models

Table 2: Summary of operations in the Poincaré ball model and the Lorentz model ($K < 0$)

	Poincaré Ball Model $(\mathbb{D}_K^n, g_{\mathbf{x}}^{\mathbb{D}})$	**Lorentz Model** $(\mathbb{H}_K^n, g_{\mathbf{x}}^{\mathbb{H}})$
Distance	$d_{\mathbb{D}}^K(\mathbf{x},\mathbf{y}) = \frac{1}{\sqrt{\lvert K\rvert}} \cosh^{-1}\left(1 - \frac{2K\lVert\mathbf{x}-\mathbf{y}\rVert_2^2}{(1+K\lVert\mathbf{x}\rVert_2^2)(1+K\lVert\mathbf{y}\rVert_2^2)}\right)$	$d_{\mathbb{H}}^K(\mathbf{x},\mathbf{y}) = \frac{1}{\sqrt{\lvert K\rvert}} \cosh^{-1}\left(K\langle\mathbf{x},\mathbf{y}\rangle_{\mathcal{L}}\right)$
Log map	$\log_{\mathbf{x}}^{K,\mathbb{D}}(\mathbf{y}) = \frac{2}{\sqrt{\lvert K\rvert}\lambda_{\mathbf{x}}^K} \tanh^{-1}\left(\sqrt{\lvert K\rvert}\,\lVert -\mathbf{x}\oplus_K \mathbf{y}\rVert_2\right) \frac{-\mathbf{x}\oplus_K \mathbf{y}}{\lVert -\mathbf{x}\oplus_K \mathbf{y}\rVert_2}$	$\log_{\mathbf{x}}^{K,\mathbb{H}}(\mathbf{y}) = \frac{\cosh^{-1}(K\langle\mathbf{x},\mathbf{y}\rangle_{\mathcal{L}})}{\sinh\left(\cosh^{-1}(K\langle\mathbf{x},\mathbf{y}\rangle_{\mathcal{L}})\right)} \left(\mathbf{y} - K\langle\mathbf{x},\mathbf{y}\rangle_{\mathcal{L}}\mathbf{x}\right)$
Exp map	$\exp_{\mathbf{x}}^{K,\mathbb{D}}(\mathbf{v}) = \mathbf{x}\oplus_K \left(\tanh\left(\sqrt{\lvert K\rvert}\frac{\lambda_{\mathbf{x}}^K\lVert\mathbf{v}\rVert_2}{2}\right)\frac{\mathbf{v}}{\sqrt{\lvert K\rvert}\lVert\mathbf{v}\rVert_2}\right)$	$\exp_{\mathbf{x}}^{K,\mathbb{H}}(v) = \cosh\left(\sqrt{\lvert K\rvert}\lVert\mathbf{v}\rVert_{\mathcal{L}}\right)\mathbf{x} + \mathbf{v}\frac{\sinh\left(\sqrt{\lvert K\rvert}\lVert\mathbf{v}\rVert_{\mathcal{L}}\right)}{\sqrt{\lvert K\rvert}\lVert\mathbf{v}\rVert_{\mathcal{L}}}$
Transport	$PT_{\mathbf{x}\to\mathbf{y}}^{K,\mathbb{D}}(\mathbf{v}) = \frac{\lambda_{\mathbf{x}}^K}{\lambda_{\mathbf{y}}^K}\,\mathrm{gyr}[\mathbf{y},-\mathbf{x}]\mathbf{v}$	$PT_{\mathbf{x}\to\mathbf{y}}^{K,\mathbb{H}}(\mathbf{v}) = \mathbf{v} - \frac{K\langle\mathbf{y},\mathbf{v}\rangle_{\mathcal{L}}}{1+K\langle\mathbf{x},\mathbf{y}\rangle_{\mathcal{L}}}(\mathbf{x}+\mathbf{y})$

A.2 Ablation Study

We conduct an ablation study to verify the effectiveness of HPC and its main components, i.e. positive sampling strategy and distance-aware discriminator. In particular, positive sampling strategy is simply removed (denoted as w/o pos), while the pairs' similarity is measured by the inner product (denoted as w/o dis). The results are summarized in Table 3.

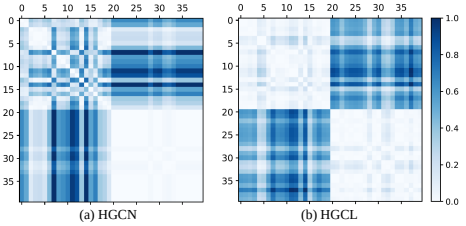
Table 3: Ablation study of the HPC on NC task.

Method	DISEASE	AIRPORT	PUBMED	CITESEER	CORA
w/o HPC	90.79 ± 1.83	91.98 ± 1.15	75.77 ± 0.47	68.89 ± 1.24	79.32 ± 0.65
w/o pos	91.84 ± 1.04	92.12 ± 1.06	75.83 ± 0.67	68.11 ± 1.74	79.18 ± 1.03
w/o dis	90.76 ± 2.19	92.00 ± 0.95	75.46 ± 0.78	68.68 ± 0.80	78.90 ± 1.21
ours	$\mathbf{93.42} \pm 0.82$	$\mathbf{92.35} \pm 1.01$	$\mathbf{79.14} \pm 0.68$	$\mathbf{72.11} \pm 0.64$	$\mathbf{82.37} \pm 0.47$

As observed, the performance decreases significantly if HPC is removed, which confirms the function of HPC in our framework. The results also reveal that both two components of HPC make essential endowment to boost the performance as any one is removed or changed from HPC is counterproductive in many cases. It is quite straightforward as dot product misjudges the similarity of hyperbolic embeddings and also hinders the utilization of the hyperbolic space.

A.3 Visualization and Analysis

Figure 3 shows the distance heatmap of inter-class and intra-class embeddings on the DISEASE dataset, where nodes 0-19 and 20-39 belong to two different classes. Compared to HGCN, it is observed that the node embeddings of HGCL in the same class are more compact, and the boundaries of different classes are less ambiguous. The observation is more evident for nodes 0-19 as the inter-class distances of the embeddings obtained by HGCN are much larger than those produced by HGCL. Thus, the contrast-enhanced hyperbolic leaning framework could pull representations of similar nodes together while pushing away the dissimilar ones, which confirms our intention that the well-designed contrastive loss is able to improve the discriminative power of hyperbolic models.



(a) HGCN (b) HGCL

Figure 3: The distance heatmap among inter-class and intra-class embeddings on DISEASE

A.4 Datasets

Three types of networks (i.e., citation network, disease spreading network, and flight network) are used. The citation networks, including CORA, CITESEER, and PUBMED, are standard benchmark datasets widely used to evaluate the performance of graph-related models. Table. 4 gives the statistics of the datasets.

Table 4: Statistics of the datasets.

DATASET	Nodes	Edges	Classes	Feature	Hyperbolicity δ
DISEASE	1044	1043	2	1000	0
AIRPORT	3188	18631	4	4	1
CITESEER	3327	4732	6	3703	3.5
PUBMED	19717	88651	3	500	3.5
CORA	2708	5429	7	1433	11