Appendix to “Expanding the Hyperbolic Kernels: A Curvature-aware Isometric Embedding View”

In the appendix, we provide the following:

- The proof of Theorem 1. This is mentioned in the methodology part (§ **3.1 Generalization of Pseudo-hyperbolic Distance in** $\mathbb{P}_c^n$) of the main paper.
- The proof of Theorem 2. This is mentioned in the methodology part (§ **3.1 Generalization of Pseudo-hyperbolic Distance in** $\mathbb{P}_c^n$) of the main paper.
- The proofs of that the curvature-aware hyperbolic kernels including the curvature-aware hyperbolic linear (CHL) kernel, the curvature-aware hyperbolic polynomial (CHPoly) kernel, the curvature-aware hyperbolic RBF (CHRBF) kernel and the curvature-aware hyperbolic Laplacian (CHLap) kernel are positive definite (PD). These are mentioned in the methodology part including (§ **3.2 RKHS with a Curvature-aware Kernel**) and (§ **4 Curvature-aware Hyperbolic Kernels**) of the main paper.
- More details about experiments. This is mentioned in the experiments part (§ **5 Experiments**) of the main paper.

1 Proof of Theorem 1 in Main Paper

Here we prove Theorem 1 from § 3.1 of the main paper. The proof follows the theories in [Rudin, 1980]. We first show the theorem of

Theorem 1. *[Rudin, 1980]. Suppose Ω is a bounded region in $\mathbb{C}^n$, $F : \Omega \to \Omega$ is holomorphic, for some $p \in \Omega$, $F(\boldsymbol{p}) = \boldsymbol{p}$ and $F'(\boldsymbol{p}) = \boldsymbol{I}$. Then $F(\boldsymbol{z}) = \boldsymbol{z}$ for all $\boldsymbol{z} \in \Omega$.*

Given the above Theorem 1, we begin to prove Theorem 1 in § 3.1 of the main paper

Proof.

Plugging $\boldsymbol{z}_j = \boldsymbol{0}$ and $\boldsymbol{z}_j = \boldsymbol{z}_i$ to the Eq. (12) in the § 3.1 of the main paper, we can get

$\varphi^c_{\boldsymbol{z}_i}(\boldsymbol{0}) = \frac{\boldsymbol{z}_i - P_{\boldsymbol{z}_i}(\boldsymbol{0}) - s_{\boldsymbol{z}_i} Q_{\boldsymbol{z}_i}(\boldsymbol{0})}{1 - c\langle \boldsymbol{0}, \boldsymbol{z}_i\rangle} = \boldsymbol{z}_i,$

$\varphi^c_{\boldsymbol{z}_i}(\boldsymbol{z}_i) = \frac{\boldsymbol{z}_i - P_{\boldsymbol{z}_i}(\boldsymbol{z}_i) - s_{\boldsymbol{z}_i} Q_{\boldsymbol{z}_i}(\boldsymbol{z}_i)}{1 - c\langle \boldsymbol{z}_i, \boldsymbol{z}_i\rangle} = \boldsymbol{0}.$

Beyond that, we need to further prove that $\varphi^c_{\boldsymbol{z}_i}$ is a involution. Put $\Phi = \varphi^c_{\boldsymbol{z}_i} \circ \varphi^c_{\boldsymbol{z}_i}$, then Φ is a holomorphic map of $\mathbb{D}_c^n$ into $\mathbb{D}_c^n$, with $\Phi(\boldsymbol{0}) = \boldsymbol{0}$.

Besides, by calculating the partial derivative of Eq. (12) in the main paper, we can obtain ${\varphi^c_{\boldsymbol{z}_i}}'(\boldsymbol{0}) = -s^2 P - sQ$ and ${\varphi^c_{\boldsymbol{z}_i}}'(\boldsymbol{z_i}) = -P/s^2 - Q/s$. This demonstrates that

$$\Phi'(\boldsymbol{0}) = {\varphi^c_{\boldsymbol{z}_i}}'(\boldsymbol{z_i}){\varphi^c_{\boldsymbol{z}_i}}'(\boldsymbol{0}) = P + Q = I \tag{1}$$

for $P^2 = P, Q^2 = Q, PQ = QP = \boldsymbol{0}$. Via Theorem 1, $\Phi(\boldsymbol{z}) = \boldsymbol{z}$ and $\varphi^c_{\boldsymbol{z}_i}(\varphi^c_{\boldsymbol{z}_i}(\boldsymbol{z}_j)) = \boldsymbol{z}_j$ is proved. $\square$

2 Proof of Theorem 2 in Main Paper

Here we prove Theorem 2 in § 3.1 of the main paper.

Proof. The pseudo-hyperbolic distance is defined as

$$\rho^n_c(\boldsymbol{z}_i, \boldsymbol{z}_j) = \sqrt{c}\|\varphi^c_{\boldsymbol{z}_i}(\boldsymbol{z}_j)\|, \tag{2}$$

where $\mathbb{D}_c^n = \{\boldsymbol{z} \in \mathbb{C}^n : c\|\boldsymbol{z}\|^2 < 1, c > 0\}$, $\boldsymbol{z}_i, \boldsymbol{z}_j \in \mathbb{D}_c^n$ and

$$\varphi^c_{\boldsymbol{z}_i}(\boldsymbol{z}_j) = \frac{\boldsymbol{z}_i - P_{\boldsymbol{z}_i}(\boldsymbol{z}_j) - s_{\boldsymbol{z}_i} Q_{\boldsymbol{z}_i}(\boldsymbol{z}_j)}{1 - c\langle \boldsymbol{z}_j, \boldsymbol{z}_i\rangle}. \tag{3}$$

To prove Theorem 2 in the main paper, we first prove that $\rho^n_c(\boldsymbol{z}_i, \boldsymbol{z}_j)$ has following properties:

(a) $1 - c\langle \varphi_{\boldsymbol{z}_i}(\boldsymbol{z}_j), \varphi_{\boldsymbol{z}_i}(\boldsymbol{z}_k)\rangle = \frac{(1-c\langle \boldsymbol{z}_i, \boldsymbol{z}_i\rangle)(1-c\langle \boldsymbol{z}_j, \boldsymbol{z}_k\rangle)}{(1-c\langle \boldsymbol{z}_j, \boldsymbol{z}_i\rangle)(1-c\langle \boldsymbol{z}_i, \boldsymbol{z}_k\rangle)}$.

(b) $1 - c|\varphi_{\boldsymbol{z}_i}(\boldsymbol{z}_j)|^2 = \frac{(1-c|\boldsymbol{z}_i|^2)(1-c|\boldsymbol{z}_j|^2)}{|1-c\langle \boldsymbol{z}_j, \boldsymbol{z}_i\rangle|^2}$.

In Eq. (a), we note that Eq. (3) can split $\varphi_{\boldsymbol{z}_i}(\boldsymbol{z}_j)$ into its $[\boldsymbol{z}_i]$ and the one that is orthogonal to $[\boldsymbol{z}_i]$. Consequently, the left side of the Eq. (a) is the equivalence of

$$1 - \frac{\langle \boldsymbol{z}_i - P_{\boldsymbol{z}_i}(\boldsymbol{z}_j), \boldsymbol{z}_i - P(\boldsymbol{z}_k)\rangle + s^2_{\boldsymbol{z}_i}\langle Q_{\boldsymbol{z}_i}(\boldsymbol{z}_j), Q(\boldsymbol{z}_k)\rangle}{(1 - c\langle \boldsymbol{z}_j, \boldsymbol{z}_i\rangle)(1 - c\langle \boldsymbol{z}_i, \boldsymbol{z}_j\rangle)}. \tag{4}$$

Besides, note $P_{\boldsymbol{z}_i}(\boldsymbol{z}_j) = \boldsymbol{z}_j$ and $Q_{\boldsymbol{z}_i}(\boldsymbol{z}_j) = \boldsymbol{z}_j$ due to P, Q are self-adjoint projections, we can also get that

$$\langle \boldsymbol{z}_j, \boldsymbol{z}_i\rangle\langle \boldsymbol{z}_i, \boldsymbol{z}_k\rangle = \langle\langle \boldsymbol{z}_j, \boldsymbol{z}_i\rangle \boldsymbol{z}_i, \boldsymbol{z}_k\rangle = \|\boldsymbol{z}_i\|^2\langle P(\boldsymbol{z}_j), \boldsymbol{z}_k\rangle, \tag{5}$$

then the Eq. (a) is easily to be get. Furthermore, the Eq. (b) is easy to derived by substituting $\boldsymbol{z}_k$ as $\boldsymbol{z}_j$ in Eq. (a).

Then the Eq. (2) can be converted to

$$\rho^n_c(\boldsymbol{z}_i, \boldsymbol{z}_j) = \sqrt{1 - \frac{(1 - c\|\boldsymbol{z}_i\|^2)(1 - c\|\boldsymbol{z}_j\|^2)}{\|1 - c\langle \boldsymbol{z}_j, \boldsymbol{z}_i\rangle\|^2}}, \tag{6}$$

which can be simplified as

$$\rho_c^n(\boldsymbol{z}_i, \boldsymbol{z}_j) = \frac{\sqrt{c}\|\boldsymbol{z}_i - \boldsymbol{z}_j\|}{\|1 - c\langle \boldsymbol{z}_j, \boldsymbol{z}_i\rangle\|}. \tag{7}$$

It is obvious that the Eq. (7) satisfies the first two properties in Theorem 2 of the main paper.

For the third property in Theorem 2 of the main paper, we prove it following the work of [Duren and Weir, 2007]. The pseudo-hyperbolic distance in the unit ball $\mathbb{D}^n$ is in the following form

$$\rho^n(\boldsymbol{z}_i, \boldsymbol{z}_j) = \sqrt{1 - \frac{(1 - \|\boldsymbol{z}_i\|^2)(1 - \|\boldsymbol{z}_j\|^2)}{\|1 - \langle \boldsymbol{z}_j, \boldsymbol{z}_i\rangle\|^2}}, \tag{8}$$

where $\mathbb{D}^n = \{\boldsymbol{z} \in \mathbb{C}^n : \|\boldsymbol{z}\|^2 < 1\}$ and $\boldsymbol{z}_i, \boldsymbol{z}_j \in \mathbb{D}^n$. $\rho^n(\boldsymbol{z}_i, \boldsymbol{z}_j)$ in the Eq. (8) satisfies the property that

$$\begin{aligned} &\frac{\|\rho^n(\boldsymbol{z}_i, \boldsymbol{z}_k) - \rho^n(\boldsymbol{z}_k, \boldsymbol{z}_j)\|}{1 - \rho^n(\boldsymbol{z}_i, \boldsymbol{z}_k)\rho^n(\boldsymbol{z}_k, \boldsymbol{z}_j)} \leq \rho^n(\boldsymbol{z}_i, \boldsymbol{z}_j) \\ \leq &\frac{\rho^n(\boldsymbol{z}_i, \boldsymbol{z}_k) + \rho^n(\boldsymbol{z}_k, \boldsymbol{z}_j)}{1 + \rho^n(\boldsymbol{z}_i, \boldsymbol{z}_k)\rho^n(\boldsymbol{z}_k, \boldsymbol{z}_j)} \quad \text{for all} \quad \boldsymbol{z}_i, \boldsymbol{z}_j, \boldsymbol{z}_k \in \mathbb{D}^n. \end{aligned}$$

Note the definition of $\mathbb{D}^n$ and $\mathbb{D}_c^n$, denote $\sqrt{c}\boldsymbol{u} = \boldsymbol{z} \in \mathbb{D}^n$, then $\boldsymbol{u} \in \mathbb{D}_c^n$ and

$$\begin{aligned} \rho^n(\boldsymbol{z}_i, \boldsymbol{z}_j) &= \rho^n(\sqrt{c}\boldsymbol{u}_i, \sqrt{c}\boldsymbol{u}_j) \\ &= \sqrt{1 - \frac{(1 - c\|\boldsymbol{u}_i\|^2)(1 - c\|\boldsymbol{u}_j\|^2)}{\|1 - c\langle \boldsymbol{u}_j, \boldsymbol{u}_i\rangle\|^2}}, \end{aligned} \tag{9}$$

the right side of the Eq. (9) is the pseudo-hyperbolic distance of $\mathbb{P}_c^n$ since $\boldsymbol{u} \in \mathbb{D}_c^n$. As a result, the third property of Theorem 2 in the main paper is proved. □

3 Proofs of Positive Definite Kernels

In this section, we prove that the CHL kernel, CHPoly kernel, CHRBF kernel and CHLap kernel are all PD kernels.

First, the definition of the PD kernel is given and then the PD property of each kernel above is proved separately.

Definition 1. *[Positive Definite Kernel [Marx, 2014]] Let $\mathcal{Z}$ be a nonempty set. A positive kernel k is a function on $\mathcal{Z} \times \mathcal{Z}$ such that*

$$\sum_{i,j=1}^{n} c_i \bar{c}_j k(\boldsymbol{z}_j, \boldsymbol{z}_i) \geq 0 \tag{10}$$

for any $\boldsymbol{z}_i, \boldsymbol{z}_j \in \mathcal{Z}$ and any finite $c_1, \ldots, c_n \in \mathbb{C}$.

3.1 Proof of Positive Definite Curvature-aware Hyperbolic Linear Kernel

To prove that the curvature-aware hyperbolic linear (CHL) kernel is a PD kernel, we first show the details of deriving it.

Theorem 2. *[Marx, 2014] Let $\mathcal{H}$ be a reproducing kernel Hilbert space on $\mathcal{X}$, and let $\{e_i\}_{i\in I}$ be any orthonormal basis for $\mathcal{H}$. Then*

$$k(\boldsymbol{w}, \boldsymbol{v}) = \sum_{i\in I} \langle e_i(\boldsymbol{v}), e_i(\boldsymbol{w})\rangle \tag{11}$$

According to Theorem 2, we define a new space $\mathcal{D}_c^n$ composed of holomorphic functions on $\mathbb{D}_c^n$, where $\mathbb{D}_c^n = \{\boldsymbol{z} \in \mathbb{C}^n : c\|\boldsymbol{z}\|^2 < 1, c > 0\}$. The inner product of $\mathcal{D}_c^n$ is defined as

$$\langle f, g\rangle = \sum_{n=0}^{\infty} \frac{\hat{f}(n)\overline{\hat{g}(n)}}{c^n}, \quad c > 0, \tag{12}$$

where $\{\hat{g}(n)\}_{n=0}^{\infty}$ and $\{\hat{f}(n)\}_{n=0}^{\infty}$ are the Taylor coefficients of f and g respectively. Therefore, the orthonormal basis on $\mathcal{D}_c^n$ is given by $\{(\sqrt{c}\boldsymbol{z})^n\}_{n=0}^{\infty}$. According to (11), the kernel in $\mathcal{D}_c^n$ is computed as

$$k^{\mathcal{D}_c^n}(\boldsymbol{z}_j, \boldsymbol{z}_i) = \sum_{n=0}^{\infty} c^n \langle \boldsymbol{z}_i, \boldsymbol{z}_j\rangle^n = \frac{1}{1 - c\langle \boldsymbol{z}_i, \boldsymbol{z}_j\rangle} \tag{13}$$

where $\boldsymbol{z}_i, \boldsymbol{z}_j \in \mathbb{D}_c^n$. Consequently, $k_{\boldsymbol{z}}^{\mathcal{D}_c^n} = \{(\sqrt{c}\boldsymbol{z})^n\}_{n=0}^{\infty}$ and $k^{\mathcal{D}_c^n}(\boldsymbol{z}_j, \boldsymbol{z}_i) = \langle k_{\boldsymbol{z}_i}^{\mathcal{D}_c^n}, k_{\boldsymbol{z}_j}^{\mathcal{D}_c^n}\rangle$.

The CHL kernel is given as the kernel in $\mathcal{D}_c^n$ in the form of

$$k^{\mathrm{CHL}}(\boldsymbol{z}_j, \boldsymbol{z}_i) = k^{\mathcal{D}_c^n}(\boldsymbol{z}_j, \boldsymbol{z}_i) = \frac{1}{1 - c\langle \boldsymbol{z}_i, \boldsymbol{z}_j\rangle}, \tag{14}$$

which is a PD kernel.

Proof. Note $k^{\mathrm{CHL}}(\boldsymbol{z}_j, \boldsymbol{z}_i) = \langle k_{\boldsymbol{z}_i}^{\mathcal{D}_c^n}, k_{\boldsymbol{z}_j}^{\mathcal{D}_c^n}\rangle$, therefore

$$\begin{aligned} \sum_{i,j}^{n} c_i \bar{c}_j k(\boldsymbol{z}_j, \boldsymbol{z}_j) &= \sum_{i,j=1}^{n} c_i \bar{c}_j \langle k_{\boldsymbol{z}_i}^{\mathcal{D}_c^n}, k_{\boldsymbol{z}_j}^{\mathcal{D}_c^n}\rangle \\ &= \langle \sum_{i=1}^{n} c_i k_{\boldsymbol{z}_i}^{\mathcal{D}_c^n}, \sum_{j=1}^{n} c_j k_{\boldsymbol{z}_j}^{\mathcal{D}_c^n}\rangle \\ &= \|\sum_{i=1}^{n} c_i k_{\boldsymbol{z}_i}^{\mathcal{D}_c^n}\|^2 \geq 0 \end{aligned}$$

□

3.2 Proof of Positive Definite Curvature-aware Hyperbolic Polynomial Kernel

The CHPoly kernel is defined as

$$k^{\mathrm{CHPoly}}(\boldsymbol{z}_i, \boldsymbol{z}_j) = (\langle k_{\boldsymbol{z}_j}^{\mathcal{D}_c^n}, k_{\boldsymbol{z}_i}^{\mathcal{D}_c^n}\rangle + b)^d \tag{15}$$

where $\langle k_{\boldsymbol{z}_j}^{\mathcal{D}_c^n}, k_{\boldsymbol{z}_i}^{\mathcal{D}_c^n}\rangle = (\frac{1}{1 - c\langle \boldsymbol{z}_j, \boldsymbol{z}_i\rangle} + b)^d$ and $b > 0, d > 0$.

In order to prove that the CHPoly kernel is PD, we first demonstrate a theorem of valid PD kernels.

Theorem 3. *Given valid PD kernel $k_1(\boldsymbol{z}_i, \boldsymbol{z}_j)$, the following new kernel will be also valid PD:*

$$k(\boldsymbol{z}_i, \boldsymbol{z}_j) = k_1(\langle \phi(\boldsymbol{z}_j), \phi(\boldsymbol{z}_i)\rangle) \tag{16}$$

where $\phi(\boldsymbol{z})$ is a function from $\boldsymbol{z}$ to $\mathbb{C}^n$ and $k_1(.,.)$ is a valid PD kernel in $\mathbb{C}^n$.

Having the Theorem 3 at hand, we can prove that the CHPoly kernel is PD.

Proof. We note that $k_{\boldsymbol{z}}^{\mathcal{D}_c^n}$ is a mapping function from $\boldsymbol{z}$ to the RKHS $\mathcal{D}_c^n$ and the polynomial kernel $k^{\mathrm{Poly}}(\boldsymbol{x}_i, \boldsymbol{x}_j) = (\langle \boldsymbol{x}_j, \boldsymbol{x}_i\rangle + b)^d$ is a valid PD kernel, where $b > 0, d > 0$.

Therefore, take $k_{\boldsymbol{z}}^{\mathcal{D}_c^n}$ as $\phi(\boldsymbol{z})$ and $k^{\mathrm{Poly}}(.,.)$ as $k_1(.,.)$ in the Eq. (16), the CHPoly kernel is also a valid PD kernel. □

3.3 Proof of Positive Definite Curvature-aware Hyperbolic RBF Kernel

The CHRBF kernel is defined as

$$k^{\mathrm{CHRBF}}(\boldsymbol{z}_i,\boldsymbol{z}_j)=\exp\big(-\frac{\|k_{\boldsymbol{z}_i}^{\mathcal{D}_c^n}-k_{\boldsymbol{z}_j}^{\mathcal{D}_c^n}\|^2}{2\tau^2}\big), \qquad (17)$$

where $\tau>0$ and $\langle k_{\boldsymbol{z}_i}^{\mathcal{D}_c^n},k_{\boldsymbol{z}_j}^{\mathcal{D}_c^n}\rangle=k^{\mathrm{CHL}}(\boldsymbol{z}_j,\boldsymbol{z}_i)$.

To prove that the kernel (17) is PD, we first provide the definition and a theorem of the negative definite (ND) kernel. In addition, a relationship between the PD kernels and ND kernels are also given.

Definition 2. *[Negative Definite Kernel [Cowling, 1983]] Let $\mathcal{Z}$ be a nonempty set. A positive kernel k is a function on $\mathcal{Z}\times\mathcal{Z}$ such that*

$$\sum_{i,j=1}^{n}c_i\bar{c}_jk(\boldsymbol{z}_j,\boldsymbol{z}_i)\le 0 \qquad (18)$$

for any $\boldsymbol{z}_i,\boldsymbol{z}_j\in\mathcal{Z}$ and any finite $c_1,\dots,c_n\in\mathbb{C}$.

Theorem 4. *[Fang* et al.*, 2021] Let $\mathcal{Z}$ be a non-empty set. An injective function $f(.):\mathcal{Z}\to\mathbb{C}^n$, maps each vector in $\mathcal{Z}$ onto an inner product space $\mathbb{C}^n$. Then $k(\boldsymbol{z}_i,\boldsymbol{z}_j):=\|f(\boldsymbol{z}_i)-f(\boldsymbol{z}_j)\|^2$ is negative definite.*

Theorem 5. *[Fang* et al.*, 2021; Cowling, 1983] Let $\mathcal{Z}$ be a non-empty set and k be a kernel on $(\mathcal{Z}\times\mathcal{Z})$. The kernel $k(\boldsymbol{z}_j,\boldsymbol{z}_i)=\exp(-\lambda\phi(\boldsymbol{z}_i,\boldsymbol{z}_j))$ is positive definite for all $\lambda>0$ if and only if $\phi(.,.)$ is negative definite.*

We note that $\|k_{\boldsymbol{z}_i}^{\mathcal{D}_c^n}-k_{\boldsymbol{z}_j}^{\mathcal{D}_c^n}\|^2$ is ND by the Theorem 4 since $k_{\boldsymbol{z}}^{\mathcal{D}_c^n}=\{(\sqrt{c}\boldsymbol{z})^n\}_{n=0}^{\infty}$ maps $\boldsymbol{z}\in\mathcal{Z}$ to the RKHS $\mathcal{D}_c^n$. Therefore, according to Theorem 5 , we can further obtain that the CHRBF kernel in Eq. (17) is PD.

3.4 Proof of Positive Definite Curvature-aware Laplacian Kernel

The CHLap kernel is defined by the form of

$$k^{\mathrm{CHLap}}(\boldsymbol{z}_i,\boldsymbol{z}_j)=\exp\big(-\frac{\|k_{\boldsymbol{z}_i}^{\mathcal{D}_c^n}-k_{\boldsymbol{z}_j}^{\mathcal{D}_c^n}\|}{\tau}\big), \qquad (19)$$

for $\tau>0$ and $\langle k_{\boldsymbol{z}_i}^{\mathcal{D}_c^n},k_{\boldsymbol{z}_j}^{\mathcal{D}_c^n}\rangle=k^{\mathrm{CHL}}(\boldsymbol{z}_j,\boldsymbol{z}_i)$. The Theorem 6 is utilized to prove that the CHLap kernel is PD.

Theorem 6. *[Fang* et al.*, 2021; Cowling, 1983] If k is a negative kernel defined on $(\mathcal{Z}\times\mathcal{Z})$ and $k(\boldsymbol{z}_i,\boldsymbol{z}_j)\ge 0$, then k^α is also negative definite for $0<\alpha<1$.*

We know that $\|k_{\boldsymbol{z}_i}^{\mathcal{D}_c^n}-k_{\boldsymbol{z}_j}^{\mathcal{D}_c^n}\|^2$ is ND and $\|k_{\boldsymbol{z}_i}^{\mathcal{D}_c^n}-k_{\boldsymbol{z}_j}^{\mathcal{D}_c^n}\|=\|k_{\boldsymbol{z}_i}^{\mathcal{D}_c^n}-k_{\boldsymbol{z}_j}^{\mathcal{D}_c^n}\|^{2\times\frac{1}{2}}$. According to Theorem 6, $\|k_{\boldsymbol{z}_i}^{\mathcal{D}_c^n}-k_{\boldsymbol{z}_j}^{\mathcal{D}_c^n}\|$ is also ND. Further, combining it with the Theorem 5, we can prove that the CHLap kernel is PD.

4 More Details of Experiments

In this section, we demonstrate more details of experiments including the graph learning and the zero-shot learning.

4.1 Details in Graph Learning

We show more details about graph learning in this section.

Details of Datasets in Graph Learning

Five real-world graph datasets including Facebook [Rozemberczki *et al.*, 2019], Terrorist [Zhao *et al.*, 2006], Wiki [Cucerzan, 2007], Amazon Electronics Computers (AC) [Shchur *et al.*, 2018], Cora ML [Bojchevski and Günnemann, 2017] are used in this study. Both the hyperbolic space and the Euclidean space are employed as the embedding space with 2, 5, 10, 25 dimensions. Poincaré embedding algorithm [Nickel and Kiela, 2017] and Euclidean embedding method Deepwalk [Perozzi *et al.*, 2014] are utilized to endow datasets to hyperbolic and Euclidean space, respectively.

Facebook is a mutual like social network among blue verified Facebook with 4 classes. We choose 2489 instances in our experiments.

Terrorist represents the dataset consist of terrorism related information with 645 nodes and 6 labels. 3 catergories with the largest sample size are chosen in our work.

Wiki includes the article networks collected form the English Wikipedia. We kept the ten categories and 2123 instances Wiki in our experiments .

AC means the Amazon Electronics Computers dataset, which is a part of Amazon co-purchase graph [McAuley *et al.*, 2015]. We choose 3 categories with the largest number of instances, and choose 2500 instances from these classes in proportion.

Cora ML, as a graph extracted from Cora [McCallum *et al.*, 2000], is a popular citation network. We select the 5 classes with the largest instance size, thus 2507 instances remain in this dataset .

Kernel SVMs Implemented in Graph Learning

Let $\mathcal{Z}$ be an instance space, a training set is given as $\{(\boldsymbol{z}_i,y_i)\}_{i=1}^m$, where $\boldsymbol{z}_i\in\mathcal{Z}$ and $y_i\in\pm 1$ represents the label of class. For PD kernels and a given $C>0$, SVM is the minimum of the following regularized empirical risk function [Burges, 1998; Loosli *et al.*, 2015]:

$$\mathcal{L}(\boldsymbol{w},\beta,\boldsymbol{\alpha})=\frac{1}{2}\|\boldsymbol{w}\|^2+C\sum_{i=1}^{l}\max(0,1-y_if(\boldsymbol{z})), \quad (20)$$

where $f(\boldsymbol{z})=\boldsymbol{w}^Tk_{\boldsymbol{z}}+\beta$, $k_{\boldsymbol{z}}\in\mathcal{H}$ means the kernel in an RKHS $\mathcal{H}$. Further more, $f(\boldsymbol{z})=0$ represents the classification hyperplane in this RKHS. Let the partial derivative of $\mathcal{L}(\boldsymbol{w},\beta,\boldsymbol{\alpha})$ with respect to $\boldsymbol{w}$ and b be 0, we can obtain the dual form of (20) in the form of

$$\max_{\boldsymbol{\alpha}}\sum_{i=1}^{m}\alpha_i-\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\alpha_i\alpha_jy_iy_jk^{\mathcal{H}}(\boldsymbol{z}_i,\boldsymbol{z}_j), \qquad (21)$$

$$s.t.\sum_{i=1}^{m}\alpha_iy_i=0, 0\le\alpha_i\le C, i=1,2,\dots,m.$$

where $k^{\mathcal{H}}(\boldsymbol{z}_i,\boldsymbol{z}_j)=\langle k_{\boldsymbol{z}_i},k_{\boldsymbol{z}_j}\rangle$ represents the PD kernels in $\mathcal{H}$.

Furthermore, the decision model can also be written as:

$$f(\boldsymbol{z})=\sum_{i=1}^{m}\alpha_iy_ik^{\mathcal{H}}(\boldsymbol{z}_i,\boldsymbol{z}_j)+\beta. \qquad (22)$$

Kernel / Dataset	Dim	Hyperbolic Embedding										Euclidean Embedding				
		CHL	CHPoly	CHRBF	CHLap	CHSig	HTang	HPoly	HRBF	HLap	HBin	EL	EPoly	ERBF	ELap	ESig
Facebook	#2	$86.5_{1.0}$	$85.0_{1.3}$	$87.1_{0.6}$	$89.2_{0.5}$	$\mathbf{89.8}_{1.0}$	$61.6_{5.9}$	$64.1_{3.0}$	$81.4_{1.8}$	$86.1_{0.3}$	$77.6_{1.3}$	$55.9_{4.5}$	$70.5_{2.6}$	$76.1_{1.7}$	$81.0_{0.8}$	$79.9_{2.5}$
	#5	$94.5_{0.5}$	$94.2_{0.4}$	$94.6_{0.7}$	$\mathbf{95.6}_{0.3}$	$95.5_{0.3}$	$74.7_{4.6}$	$89.0_{0.4}$	$94.2_{0.3}$	$95.4_{0.4}$	$92.4_{0.8}$	$65.2_{5.4}$	$91.2_{0.3}$	$92.2_{0.6}$	$93.9_{0.5}$	$92.9_{0.5}$
	#10	$95.5_{0.3}$	$95.0_{0.5}$	$94.7_{0.7}$	$\mathbf{96.4}_{0.3}$	$95.9_{0.4}$	$81.1_{4.3}$	$92.2_{0.4}$	$94.7_{0.6}$	$95.9_{0.4}$	$93.3_{0.3}$	$83.4_{4.0}$	$93.4_{0.4}$	$94.4_{0.5}$	$95.1_{0.2}$	$94.9_{0.4}$
	#25	$95.8_{0.3}$	$95.4_{0.4}$	$95.6_{0.4}$	$\mathbf{96.4}_{0.3}$	$96.2_{0.3}$	$91.6_{0.3}$	$94.3_{0.4}$	$95.1_{0.4}$	$96.1_{0.4}$	$94.0_{0.4}$	$92.6_{0.4}$	$94.8_{0.7}$	$95.9_{0.4}$	$96.3_{0.4}$	$95.1_{0.5}$
Terrorist	#2	$71.1_{3.3}$	$67.5_{4.5}$	$70.1_{2.5}$	$\mathbf{72.6}_{2.2}$	$69.8_{2.3}$	$49.2_{2.2}$	$53.8_{1.5}$	$67.4_{1.6}$	$72.0_{1.8}$	$64.1_{2.5}$	$49.8_{3.6}$	$54.3_{2.7}$	$64.5_{2.5}$	$68.1_{1.9}$	$65.6_{2.5}$
	#5	$\mathbf{78.2}_{1.9}$	$75.3_{3.9}$	$75.5_{2.0}$	$75.7_{1.4}$	$77.4_{1.6}$	$49.8_{1.5}$	$58.4_{3.1}$	$74.1_{2.1}$	$76.5_{2.1}$	$76.2_{2.4}$	$55.6_{2.7}$	$65.2_{1.3}$	$71.8_{2.7}$	$74.9_{2.8}$	$69.0_{1.8}$
	#10	$\mathbf{78.4}_{3.6}$	$75.4_{3.7}$	$76.2_{2.0}$	$78.2_{1.6}$	$77.1_{2.2}$	$53.2_{4.9}$	$65.5_{2.2}$	$76.0_{1.2}$	$77.1_{1.5}$	$77.2_{2.0}$	$63.5_{2.2}$	$72.8_{1.5}$	$73.9_{3.3}$	$77.2_{1.6}$	$72.5_{1.4}$
	#25	$76.7_{1.7}$	$75.9_{3.0}$	$76.6_{1.5}$	$\mathbf{77.2}_{1.7}$	$77.0_{1.5}$	$58.0_{2.1}$	$69.7_{1.0}$	$75.8_{1.3}$	$76.9_{1.1}$	$74.9_{3.2}$	$64.7_{2.2}$	$73.5_{1.2}$	$72.0_{3.7}$	$74.9_{2.4}$	$72.0_{2.3}$
Wiki	#2	$81.8_{0.5}$	$81.1_{5.0}$	$81.9_{1.1}$	$\mathbf{85.8}_{0.7}$	$85.2_{0.2}$	$48.7_{4.2}$	$52.2_{3.1}$	$76.9_{0.8}$	$82.9_{0.7}$	$75.3_{1.7}$	$52.8_{5.5}$	$62.7_{2.7}$	$64.6_{2.3}$	$72.2_{0.7}$	$71.9_{1.3}$
	#5	$89.2_{0.4}$	$88.1_{0.4}$	$90.4_{0.5}$	$\mathbf{92.4}_{0.2}$	$91.8_{0.5}$	$68.5_{3.8}$	$79.8_{2.6}$	$89.9_{0.4}$	$91.8_{0.3}$	$86.3_{1.5}$	$65.8_{3.6}$	$81.6_{2.0}$	$85.3_{0.8}$	$88.4_{0.6}$	$87.8_{0.5}$
	#10	$90.8_{0.3}$	$89.8_{0.4}$	$91.8_{0.3}$	$\mathbf{93.2}_{0.3}$	$92.8_{0.5}$	$75.7_{3.0}$	$87.9_{1.9}$	$91.5_{0.3}$	$93.0_{0.2}$	$88.4_{0.4}$	$78.5_{2.9}$	$89.3_{0.9}$	$90.4_{0.5}$	$92.4_{0.3}$	$89.9_{0.3}$
	#25	$91.5_{0.5}$	$90.4_{0.6}$	$92.5_{0.5}$	$\mathbf{93.8}_{0.3}$	$93.1_{0.4}$	$83.3_{3.0}$	$89.1_{1.9}$	$92.0_{0.5}$	$93.6_{0.4}$	$89.1_{0.5}$	$84.2_{2.1}$	$89.3_{0.4}$	$91.4_{0.5}$	$92.6_{0.3}$	$90.3_{0.4}$
AC	#2	$85.1_{0.6}$	$84.9_{0.7}$	$84.0_{1.0}$	$86.5_{0.6}$	$\mathbf{86.8}_{0.5}$	$74.0_{3.6}$	$68.1_{7.9}$	$83.3_{0.7}$	$86.2_{0.5}$	$78.0_{6.1}$	$63.9_{7.8}$	$77.2_{4.8}$	$81.4_{1.2}$	$82.1_{0.9}$	$77.2_{4.9}$
	#5	$94.3_{0.5}$	$94.1_{0.5}$	$94.1_{0.6}$	$95.1_{0.4}$	$\mathbf{95.2}_{0.5}$	$81.0_{6.7}$	$89.5_{0.5}$	$93.9_{0.5}$	$94.9_{0.5}$	$92.4_{0.8}$	$75.4_{2.3}$	$85.6_{4.3}$	$89.3_{0.3}$	$90.5_{0.5}$	$85.6_{4.3}$
	#10	$96.3_{0.4}$	$96.0_{0.4}$	$96.2_{0.5}$	$\mathbf{96.8}_{0.4}$	$96.4_{0.3}$	$92.5_{0.4}$	$94.9_{0.6}$	$96.2_{0.4}$	$96.7_{0.2}$	$94.6_{0.9}$	$90.5_{0.8}$	$92.1_{0.7}$	$95.1_{0.4}$	$95.2_{0.4}$	$92.1_{0.7}$
	#25	$96.4_{0.5}$	$96.0_{0.4}$	$96.1_{0.2}$	$\mathbf{96.7}_{0.2}$	$96.5_{0.1}$	$93.9_{0.3}$	$93.5_{0.6}$	$95.9_{0.3}$	$96.1_{0.5}$	$94.9_{0.6}$	$91.9_{0.4}$	$92.8_{0.9}$	$96.6_{0.4}$	$96.1_{0.3}$	$92.8_{0.9}$
Cora ML	#2	$88.2_{0.3}$	$\mathbf{88.7}_{0.7}$	$85.7_{0.5}$	$88.2_{0.3}$	$88.6_{0.4}$	$57.8_{3.0}$	$61.9_{4.0}$	$85.8_{0.4}$	$88.5_{0.5}$	$85.2_{0.4}$	$58.7_{4.6}$	$62.8_{3.9}$	$76.9_{0.7}$	$79.7_{1.0}$	$81.3_{0.5}$
	#5	$94.3_{0.2}$	$94.3_{0.3}$	$94.0_{0.3}$	$95.1_{0.2}$	$94.6_{0.3}$	$86.6_{0.1}$	$88.7_{0.3}$	$93.8_{0.5}$	$\mathbf{95.2}_{0.4}$	$93.5_{0.3}$	$67.8_{3.1}$	$85.4_{2.1}$	$88.1_{0.6}$	$90.0_{0.3}$	$89.3_{0.4}$
	#10	$95.3_{0.2}$	$95.4_{0.2}$	$95.3_{0.2}$	$\mathbf{96.5}_{0.2}$	$95.9_{0.3}$	$88.7_{0.1}$	$93.6_{0.3}$	$95.3_{0.3}$	$96.1_{0.3}$	$94.9_{0.3}$	$83.9_{3.3}$	$90.1_{0.4}$	$92.1_{0.3}$	$93.0_{0.2}$	$91.4_{0.3}$
	#25	$95.6_{0.3}$	$95.9_{0.4}$	$95.6_{0.3}$	$\mathbf{96.6}_{0.2}$	$96.2_{0.2}$	$93.0_{0.1}$	$95.6_{0.3}$	$95.7_{0.3}$	$\mathbf{96.6}_{0.3}$	$95.6_{0.4}$	$88.9_{0.2}$	$91.6_{0.9}$	$93.6_{0.4}$	$93.6_{0.2}$	$92.8_{0.3}$
Avg. ACC.		88.8	87.9	88.4	**89.9**	89.6	73.1	79.1	87.4	89.4	85.9	71.7	81.8	84.3	86.4	84.2
Top1 Times		2	1	0	13	3	0	0	0	2	0	0	0	0	0	0

Table 1: Mean AUC of node classification on graph datasets including Facebook, Terrorist, Wiki, AC and Cora ML. The subscript of each number indicates the corresponding standard deviation. We use **bold** to indicate the best result.

For indefinite hyperbolic kernels $k^{\mathcal{K}}(\boldsymbol{z}_i, \boldsymbol{z}_j)$ in an RKKS $\mathcal{K}$, we capitalize on the KSVM solver in Kreĭn spaces to process the classification tasks [Loosli *et al.*, 2015]. This algorithm converts the non-convex stabilization problem to the convex dual problem. The dual form of the the KSVM solver in Kreĭn spaces is in the form of:

$$\max_{\widetilde{\boldsymbol{\alpha}}} \sum_{i=1}^{m} \widetilde{\alpha}_i - \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} \widetilde{\alpha}_i \widetilde{\alpha}_j y_i y_j \bar{k}(\boldsymbol{z}_i, \boldsymbol{z}_j) \tag{23}$$

$$s.t. \sum_{i=1}^{m} \widetilde{\alpha}_i y_i = 0, 0 \leq \widetilde{\alpha}_i \leq C, i = 1, 2, \ldots, m,$$

where $\bar{k} = k_+ + k_-$ is the PD kernel of associated RKHSs. We denote $\boldsymbol{K}$ the indefinite kernel matrix such that $\boldsymbol{K}^{\mathcal{K}}(i,j) = k^{\mathcal{K}}(\boldsymbol{z}_i, \boldsymbol{z}_j)$, and transform the indefinite kernel matrix $\boldsymbol{K}$ to the PD Gram matrix $\bar{\boldsymbol{K}}$ by taking the absolute value of the eigenvalues of $\boldsymbol{K}$: $\bar{\boldsymbol{K}} = \boldsymbol{USDU}^T$ with $\boldsymbol{S}$ =sign($\boldsymbol{D}$). Then $\boldsymbol{\alpha} = \boldsymbol{USU}^T\widetilde{\boldsymbol{\alpha}}$ and the decision boundary is in the same form as (22) except that the kernel is indefinite.

Except the our CHSig kernel, compared HPoly kernel [Cho *et al.*, 2019] and Sig kernel [Bishop and Nasrabadi, 2006] are indefinite, all other kernels are PD. To solve the dual form of all kernel SVMs, the the LIBSVM library are utilized [Chang and Lin, 2011].

More Performance Metrics in Graph Learning

Two performance metrics including area under the ROC curve (AUC) and macro-averaged area under precision recall curve (AUPR) are reported in this section. From the Table 1 and Table 2, we observe that our proposed curvature-aware kernels also perform well under the evaluation metrics of AUC and AUPR, especially the CHLap kernel achieved the highest average AUC and AUPR.

4.2 Details in Zero-shot Learning

Details of Datasets in Zero-shot Learning

Datasets **CUB** [Wah *et al.*, 2011], **AWA1** [Lampert *et al.*, 2013] and **AWA2** [Akata *et al.*, 2015] are applied to evaluate the performance of ZSL tasks. The dimensions of semantic features for CUB, AWA1 and AWA2 are 312, 85 and 85, respectively. All datasets extract the visual feature from the ImageNet pre-trained ResNet-101 and the dimensions are 2048.

The CUB is a fine-grained dataset which contains 11788 images of 200 different bird species, with 312 attributes. The 200 bird species are split into 150 seen classes and 50 unseen classes. In addition, we use a total of 7075 images in CUB as training and validation sets , and the rest as our test set. In the test set, 1,764 images are seen and 2,679 images are unseen.

The AWA1 is a coarse-grained dataset which contains 30475 animal images with 50 classes which are annotated by 85 attributes. In our experiments, all images are devided into 40 seen classes and 10 unseen classes. Besides, 19832 images in the AWA1 are used as training and validation sets, and the rest are used as test set in which 1764 images are seen and 2679 images are unseen.

Similar to AWA1 , the AWA2 is also a coarse-grained dataset which contains 37322 animal images with the same classes and attributes as AWA1. In addition, the dataset is also devided into 40 seen classes and 10 unseen classes in the experiments. The training and validation sets consist of 23527 images totally, and the rest images are used as test set. Moreover, there are 5882 seen images and 7913 unseen images in the test set.

Loss Function in Zero-shot Learning

In the baseline, we follow the loss function in [Fang *et al.*, 2021] to optimize the network. Specifically, the loss function

Kernel / Dataset	Dim	Hyperbolic Embedding										Euclidean Embedding				
		CHL	CHPoly	CHRBF	CHLap	CHSig	HTang	HPoly	HRBF	HLap	HBin	EL	EPoly	ERBF	ELap	ESig
Facebook	#2	$72.2_{1.7}$	$69.9_{1.7}$	$72.1_{1.1}$	$75.1_{1.1}$	$\mathbf{76.4}_{1.2}$	$36.7_{3.3}$	$39.3_{2.2}$	$63.8_{2.1}$	$68.5_{1.3}$	$49.8_{3.7}$	$32.2_{2.0}$	$47.2_{3.5}$	$55.5_{2.4}$	$60.7_{0.8}$	$59.4_{2.1}$
	#5	$88.2_{0.5}$	$87.1_{0.8}$	$88.4_{0.9}$	$\mathbf{89.8}_{0.5}$	$89.4_{1.0}$	$51.4_{3.0}$	$74.7_{0.8}$	$87.2_{0.9}$	$89.1_{0.5}$	$82.6_{1.9}$	$44.0_{3.4}$	$75.8_{0.6}$	$82.2_{0.6}$	$83.9_{0.7}$	$79.7_{1.4}$
	#10	$89.4_{0.8}$	$88.9_{0.9}$	$88.6_{0.9}$	$\mathbf{90.7}_{0.4}$	$90.5_{0.9}$	$64.6_{2.5}$	$83.5_{1.0}$	$88.1_{0.6}$	$89.9_{0.6}$	$84.4_{0.8}$	$68.1_{3.0}$	$82.5_{1.2}$	$87.3_{1.0}$	$88.3_{0.3}$	$85.8_{1.0}$
	#25	$90.8_{0.3}$	$90.1_{0.3}$	$89.8_{1.1}$	$\mathbf{91.3}_{0.5}$	$90.9_{0.7}$	$83.0_{0.7}$	$87.5_{1.1}$	$89.3_{0.5}$	$90.7_{0.7}$	$87.0_{1.2}$	$83.4_{0.4}$	$84.9_{1.8}$	$88.8_{0.7}$	$89.8_{0.8}$	$87.5_{0.9}$
Terrorist	#2	$52.7_{2.2}$	$49.2_{2.5}$	$54.7_{1.7}$	$\mathbf{56.0}_{1.7}$	$52.4_{2.9}$	$33.8_{1.4}$	$36.5_{1.4}$	$49.2_{2.4}$	$53.3_{2.3}$	$46.2_{1.2}$	$34.1_{1.7}$	$37.0_{1.8}$	$47.4_{2.1}$	$50.0_{1.6}$	$47.7_{2.3}$
	#5	$\mathbf{60.5}_{2.1}$	$58.5_{2.7}$	$59.8_{2.3}$	$60.5_{1.7}$	$60.2_{2.5}$	$33.6_{0.9}$	$40.5_{2.2}$	$59.1_{1.7}$	$\mathbf{61.1}_{2.6}$	$58.1_{2.5}$	$38.5_{1.9}$	$47.7_{1.1}$	$54.6_{2.3}$	$58.1_{1.4}$	$51.4_{2.5}$
	#10	$61.8_{2.2}$	$58.6_{2.0}$	$61.3_{1.8}$	$\mathbf{63.0}_{2.2}$	$59.7_{2.8}$	$37.7_{3.4}$	$48.4_{2.5}$	$60.4_{1.5}$	$61.2_{2.5}$	$59.9_{2.0}$	$45.3_{1.0}$	$55.7_{2.0}$	$57.1_{2.7}$	$59.5_{2.3}$	$55.0_{2.4}$
	#25	$58.9_{2.2}$	$58.3_{2.6}$	$60.9_{2.1}$	$61.7_{2.0}$	$60.6_{2.3}$	$40.3_{1.5}$	$52.9_{1.6}$	$59.9_{1.8}$	$\mathbf{61.8}_{1.9}$	$57.2_{1.9}$	$48.6_{1.6}$	$56.8_{1.7}$	$56.3_{2.2}$	$59.3_{2.3}$	$55.0_{2.5}$
Wiki	#2	$55.0_{1.1}$	$53.7_{6.4}$	$56.5_{1.9}$	$\mathbf{61.4}_{1.2}$	$61.2_{0.5}$	$11.2_{0.8}$	$14.3_{2.2}$	$48.8_{1.4}$	$54.5_{1.1}$	$42.1_{1.0}$	$13.8_{1.9}$	$23.2_{0.8}$	$27.5_{1.4}$	$33.1_{1.0}$	$30.5_{0.6}$
	#5	$71.4_{0.8}$	$66.6_{0.9}$	$73.5_{1.0}$	$\mathbf{76.7}_{0.5}$	$75.3_{0.8}$	$38.4_{3.5}$	$59.7_{1.2}$	$72.5_{0.8}$	$75.4_{0.7}$	$60.7_{1.4}$	$37.4_{2.4}$	$55.2_{1.7}$	$66.1_{0.9}$	$69.0_{0.7}$	$64.8_{1.2}$
	#10	$74.8_{0.7}$	$71.0_{0.7}$	$76.8_{0.7}$	$\mathbf{79.0}_{0.4}$	$77.0_{1.0}$	$50.9_{1.6}$	$68.3_{0.9}$	$76.3_{0.9}$	$77.9_{0.6}$	$66.7_{0.6}$	$54.0_{1.9}$	$62.9_{0.9}$	$74.9_{1.1}$	$76.7_{0.6}$	$69.9_{1.1}$
	#25	$76.2_{0.9}$	$72.6_{1.2}$	$77.8_{0.7}$	$\mathbf{79.9}_{0.6}$	$77.5_{1.1}$	$60.2_{1.2}$	$71.9_{0.9}$	$77.1_{0.8}$	$79.3_{0.7}$	$68.2_{1.0}$	$64.4_{1.4}$	$65.1_{1.1}$	$77.3_{1.3}$	$78.7_{0.6}$	$72.5_{1.2}$
AC	#2	$74.3_{1.0}$	$73.9_{1.2}$	$74.2_{1.4}$	$77.7_{0.8}$	$\mathbf{77.8}_{0.7}$	$58.3_{1.7}$	$56.9_{7.2}$	$73.1_{2.1}$	$76.4_{0.8}$	$71.0_{1.9}$	$47.9_{3.6}$	$58.8_{4.5}$	$67.5_{2.2}$	$68.6_{1.2}$	$58.8_{4.5}$
	#5	$89.9_{1.1}$	$89.5_{1.1}$	$90.3_{1.0}$	$\mathbf{91.6}_{0.6}$	$91.5_{1.0}$	$69.2_{4.3}$	$78.8_{1.3}$	$89.9_{0.9}$	$91.2_{0.8}$	$86.3_{1.6}$	$60.5_{1.9}$	$74.0_{3.6}$	$81.0_{0.7}$	$82.7_{0.6}$	$74.0_{3.6}$
	#10	$93.1_{0.9}$	$93.0_{0.6}$	$93.3_{0.7}$	$\mathbf{94.2}_{0.6}$	$93.5_{0.4}$	$86.4_{0.5}$	$91.3_{1.1}$	$93.3_{0.7}$	$93.9_{0.4}$	$90.2_{2.1}$	$83.2_{1.2}$	$84.4_{1.9}$	$91.2_{0.9}$	$91.4_{0.7}$	$84.4_{1.9}$
	#25	$93.1_{0.9}$	$92.9_{0.7}$	$93.4_{0.3}$	$\mathbf{94.1}_{0.3}$	$93.5_{0.2}$	$90.2_{0.5}$	$92.1_{0.8}$	$92.9_{0.5}$	$93.4_{0.6}$	$90.8_{2.5}$	$84.6_{1.1}$	$85.5_{1.7}$	$92.2_{0.6}$	$93.1_{0.4}$	$85.5_{1.7}$
Cora ML	#2	$72.2_{0.8}$	$72.4_{1.1}$	$70.2_{0.7}$	$\mathbf{73.8}_{0.5}$	$71.0_{0.7}$	$34.4_{1.6}$	$39.4_{2.4}$	$69.9_{0.6}$	$72.9_{0.7}$	$65.5_{0.7}$	$32.1_{2.1}$	$37.0_{3.2}$	$55.8_{1.1}$	$57.9_{0.7}$	$58.9_{0.6}$
	#5	$86.3_{1.0}$	$86.2_{0.8}$	$87.0_{0.6}$	$\mathbf{88.6}_{0.4}$	$87.3_{0.8}$	$71.3_{0.2}$	$77.4_{0.5}$	$86.7_{1.0}$	$88.5_{0.7}$	$84.1_{0.7}$	$45.3_{1.7}$	$66.1_{1.4}$	$75.6_{0.7}$	$77.6_{0.5}$	$76.1_{1.1}$
	#10	$89.0_{0.4}$	$89.0_{0.7}$	$89.2_{0.4}$	$\mathbf{90.9}_{0.3}$	$89.1_{0.7}$	$77.5_{0.2}$	$85.5_{0.4}$	$89.0_{0.7}$	$90.5_{0.3}$	$87.8_{0.5}$	$66.6_{2.3}$	$75.7_{0.8}$	$82.0_{0.7}$	$83.5_{0.6}$	$80.1_{0.9}$
	#25	$89.6_{0.4}$	$89.9_{0.6}$	$89.8_{0.6}$	$\mathbf{91.2}_{0.4}$	$89.9_{0.6}$	$84.5_{0.3}$	$88.9_{0.4}$	$89.9_{0.5}$	$91.1_{0.3}$	$88.9_{0.7}$	$76.5_{0.3}$	$79.3_{2.5}$	$84.9_{0.7}$	$84.9_{0.3}$	$83.3_{0.7}$
Avg. ACC.		77	75.6	77.4	**79.4**	78.2	55.7	64.4	75.8	78.0	71.4	53.0	62.7	70.3	72.3	68.0
Top1 Times		1	0	0	16	2	0	0	0	2	0	0	0	0	0	0

Table 2: Mean AUPR of node classification on graph datasets including Facebook, Terrorist, Wiki, AC and Cora ML. The subscript of each number indicates the corresponding standard deviation. We use **bold** to indicate the best result.

is given by:

$$\mathcal{L}_{\mathrm{ZSL}} = -\frac{1}{N}\sum_{i=1}^{N}\log\big(\frac{\exp(-\|\boldsymbol{z}_i - \boldsymbol{e}^*\|)}{\sum_{j=1}^{M}\exp(-\|\boldsymbol{z}_i - \boldsymbol{e}_j\|)}\big), \tag{24}$$

where $\boldsymbol{e}*$ is the attribute feature, sharing the same label to $\boldsymbol{z}_i$. Of note, in the baseline, both the visual feature and semantic feature are in the Euclidean space, i.e., $\boldsymbol{z}_i,\ \boldsymbol{e}_i \in \mathbb{C}^n$. We can further kernelize the loss function. In this case, for $\boldsymbol{z}_i,\ \boldsymbol{e}_i \in \mathbb{P}_c^n$, the kernelized loss is formulated as:

$$\mathcal{L}_{\mathrm{ZSL}}^{\mathrm{K}} = -\frac{1}{N}\sum_{i=1}^{N}\log\big(\frac{h(k(\boldsymbol{z}_i, \boldsymbol{e}^*))}{\sum_{j=1}^{M} g(k(\boldsymbol{z}_i, \boldsymbol{e}_j))}\big), \tag{25}$$

where k is the kernel. Here, h is $\exp$ function if k is non-exponential type kernels. Otherwise, h is the identity mapping.

References

[Akata *et al.*, 2015] Zeynep Akata, Florent Perronnin, Zaid Harchaoui, and Cordelia Schmid. Label-embedding for image classification. *IEEE transactions on pattern analysis and machine intelligence*, 38(7):1425–1438, 2015.

[Bishop and Nasrabadi, 2006] Christopher M Bishop and Nasser M Nasrabadi. *Pattern Recognition and Machine Learning*, volume 4. Springer, 2006.

[Bojchevski and Günnemann, 2017] Aleksandar Bojchevski and Stephan Günnemann. Deep gaussian embedding of graphs: Unsupervised inductive learning via ranking. *arXiv preprint arXiv:1707.03815*, 2017.

[Burges, 1998] Christopher J. C. Burges. A tutorial on support vector machines for pattern recognition. *Data Min. Knowl. Discov.*, 2(2):121–167, 1998.

[Chang and Lin, 2011] Chih-Chung Chang and Chih-Jen Lin. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2:27:1–27:27, 2011.

[Cho *et al.*, 2019] Hyunghoon Cho, Benjamin DeMeo, Jian Peng, and Bonnie Berger. Large-margin classification in hyperbolic space. In *Proceedings of the International Conference on Artificial Intelligence and Statistics*, pages 1832–1840, 2019.

[Cowling, 1983] Michael G. Cowling. Harmonic analysis on semigroups. *Annals of Mathematics*, 117(2):267–283, 1983.

[Cucerzan, 2007] Silviu Cucerzan. Large-scale named entity disambiguation based on wikipedia data. In *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*, pages 708–716, 2007.

[Duren and Weir, 2007] Peter Duren and Rachel Weir. The pseudohyperbolic metric and bergman spaces in the ball. *Transactions of the American Mathematical Society*, 359(1):63–76, 2007.

[Fang *et al.*, 2021] Pengfei Fang, Mehrtash Harandi, and Lars Petersson. Kernel methods in hyperbolic spaces. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 10665–10674, October 2021.

[Lampert *et al.*, 2013] Christoph H Lampert, Hannes Nickisch, and Stefan Harmeling. Attribute-based classification for zero-shot visual object categorization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(3):453–465, 2013.

[Loosli *et al.*, 2015] Gaëlle Loosli, Stéphane Canu, and Cheng Soon Ong. Learning SVM in Kreĭn spaces. *IEEE Transactions on pattern Analysis and Machine Intelligence*, 38(6):1204–1216, 2015.

[Marx, 2014] Gregory Marx. The complete pick property and reproducing kernel hilbert spaces. 2014.

[McAuley *et al.*, 2015] Julian McAuley, Christopher Targett, Qinfeng Shi, and Anton Van Den Hengel. Image-based recommendations on styles and substitutes. In *Proceedings of the 38th international ACM SIGIR conference on research and development in information retrieval*, pages 43–52, 2015.

[McCallum *et al.*, 2000] Andrew Kachites McCallum, Kamal Nigam, Jason Rennie, and Kristie Seymore. Automating the construction of internet portals with machine learning. *Information Retrieval*, 3(2):127–163, 2000.

[Nickel and Kiela, 2017] Maximillian Nickel and Douwe Kiela. Poincaré embeddings for learning hierarchical representations. In *Proceedings of the Advances in Neural Information Processing Systems*, pages 6338–6347, 2017.

[Perozzi *et al.*, 2014] Bryan Perozzi, Rami Al-Rfou, and Steven Skiena. Deepwalk: Online learning of social representations. In *Proceedings of the International Conference on Knowledge Discovery and Data Mining*, pages 701–710. ACM, 2014.

[Rozemberczki *et al.*, 2019] Benedek Rozemberczki, Ryan Davies, Rik Sarkar, and Charles Sutton. Gemsec: Graph embedding withsself clustering. In *Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2019*, pages 65–72. ACM, 2019.

[Rudin, 1980] Walter Rudin. Function theory in the unit ball of $\mathbb{C}^n$. 1980.

[Shchur *et al.*, 2018] Oleksandr Shchur, Maximilian Mumme, Aleksandar Bojchevski, and Stephan Günnemann. Pitfalls of graph neural network evaluation. *arXiv preprint arXiv:1811.05868*, 2018.

[Wah *et al.*, 2011] Catherine Wah, Steve Branson, Peter Welinder, Pietro Perona, and Serge Belongie. The caltech-ucsd birds-200-2011 dataset. 2011.

[Zhao *et al.*, 2006] Bin Zhao, Prithviraj Sen, and Lise Getoor. Entity and elationship labeling in affiliation networks. In *ICML Workshop on Statistical Network Analysis*, 2006.