Hyperbolic Graph Convolutional Neural Networks

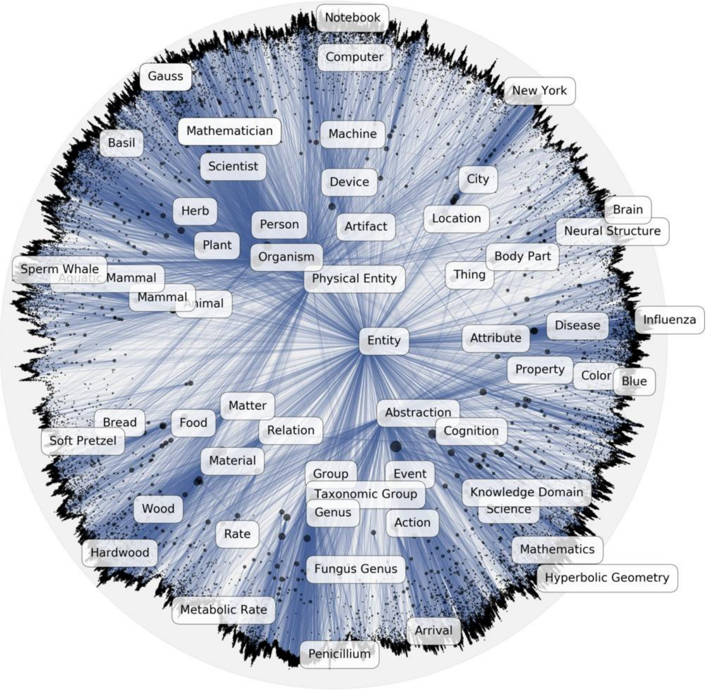
Ines Chami[*‡] Rex Ying[* †] Christopher Ré[†] Jure Leskovec[†]

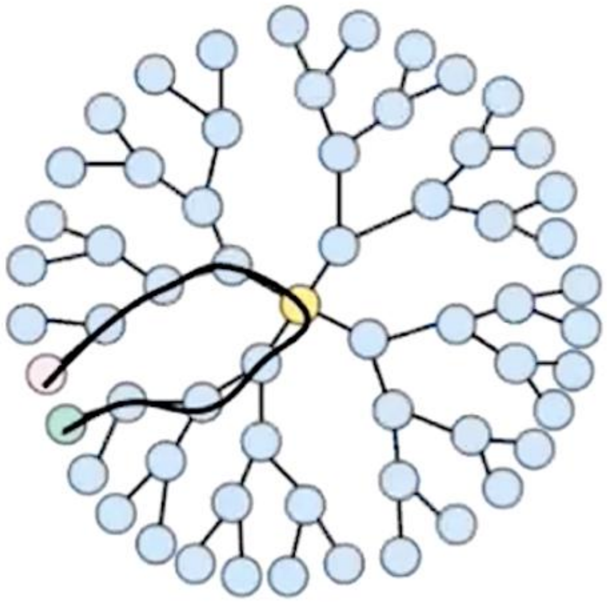
[†]Department of Computer Science, Stanford University
[‡]Institute for Computational and Mathematical Engineering, Stanford University
{chami, rexying, chrismre, jure}@cs.stanford.edu

NIPS 2019

为什么用双曲空间嵌入

- 对于具备“层级结构”的数据而言，传统的欧式嵌入会导致隐向量失真：





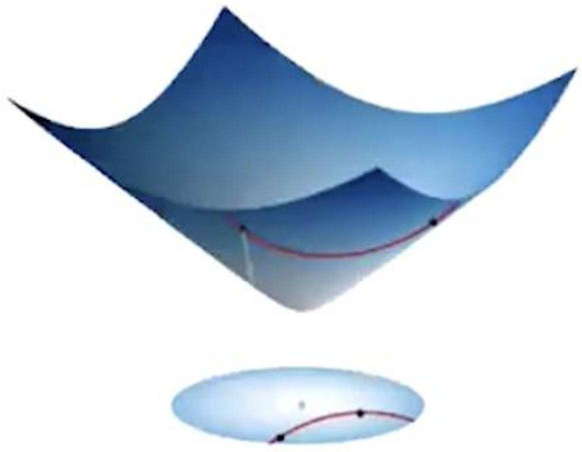
解决方案：引入曲率

$$k = \frac{1}{\text{曲率半径}}$$

双曲空间/双曲面模型

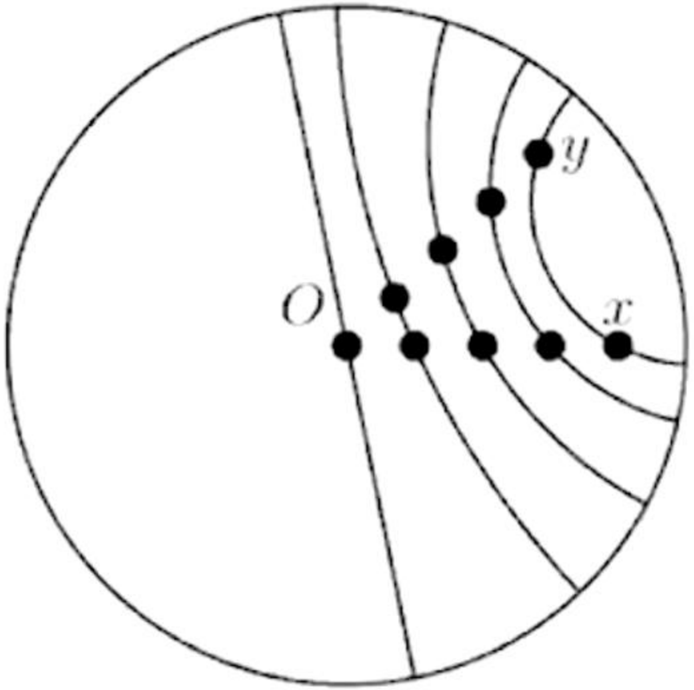
- 双曲空间：一个n维双曲空间就是一个高斯曲率处处为常数的流形。数学家认为这种形状拥有负常曲率(constant negative curvature)，而球形拥有正常曲率(constant positive curve)。

双曲面模型



庞加莱圆盘

$$x^2 - y^2 - z^2 = 1$$



度规

- 在每一点处对空间距离的定义
- 欧式空间中的度规：$dr^2 = dx^2 + dy^2 \rightarrow \begin{pmatrix} dx \\ dy \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} dx & dy \end{pmatrix}$

度规张量

- 双曲空间中的度规：$dr^2 = \frac{2dx^2}{1-x^2-y^2} + \frac{2dy^2}{1-x^2-y^2} \rightarrow$

$$\begin{pmatrix} dx \\ dy \end{pmatrix} \begin{pmatrix} \frac{2}{1-x^2-y^2} & 0 \\ 0 & \frac{2}{1-x^2-y^2} \end{pmatrix} \begin{pmatrix} dx & dy \end{pmatrix}$$

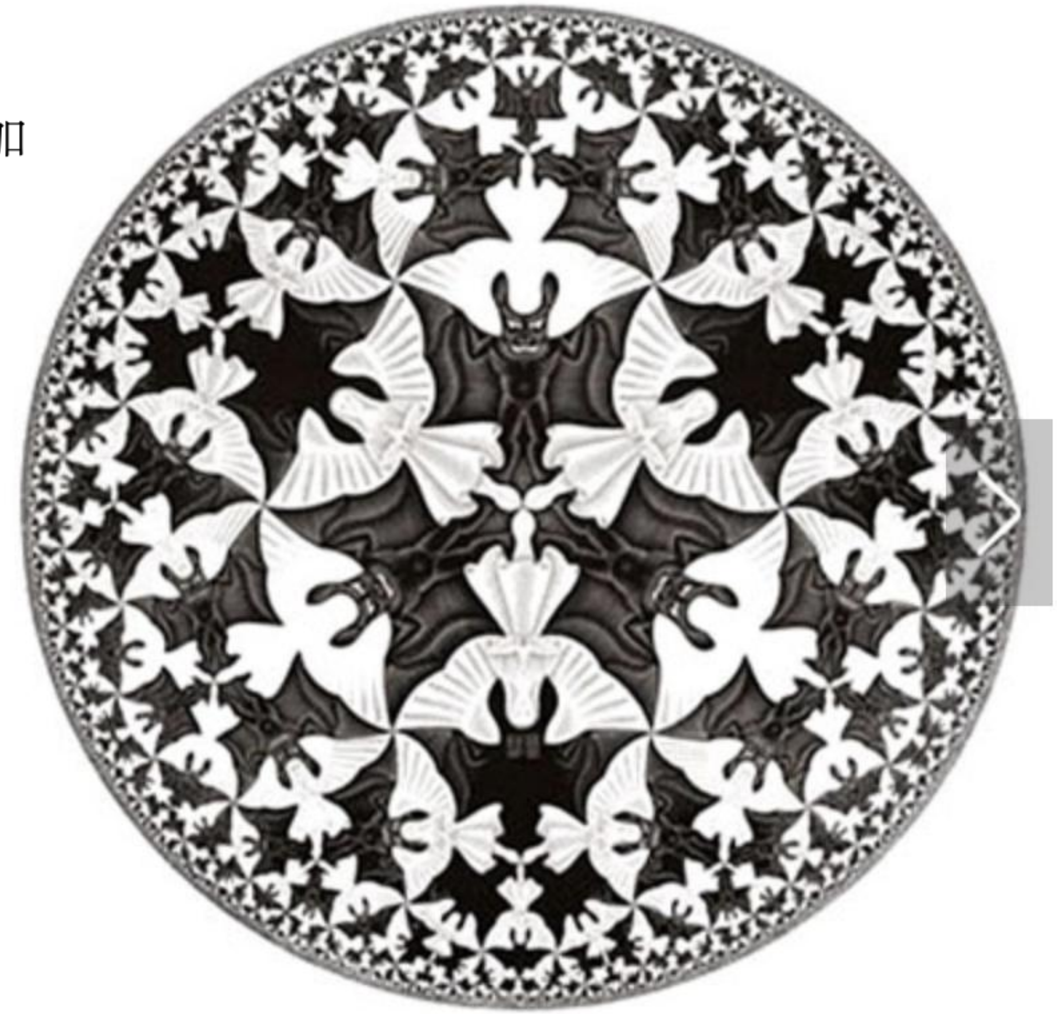
度规张量

双曲空间及基于双曲空间的嵌入

双曲空间也有维度，其维度表示比欧式空间中同维度空间增加一维，与对应的庞加莱圆盘的维度相同。

正式的：$D^n = \{x \in R^n : ||x|| < 1\}$

双曲空间的距离：$d(u,v) = arcosh(1 + \frac{2||u-v||^2}{(1-||u||^2)(1-||v||^2)})$



双曲空间下的神经网络运算

- (Hyperbolic Neural Network 2018,NIPS)

Möbius addition. The *Möbius addition* of x and y in $\mathbb{D}_c^n$ is defined as

$$x \oplus_c y := \frac{(1 + 2c\langle x, y\rangle + c\|y\|^2)x + (1 - c\|x\|^2)y}{1 + 2c\langle x, y\rangle + c^2\|x\|^2\|y\|^2}. \tag{6}$$

In particular, when $c = 0$, one recovers the Euclidean addition of two vectors in $\mathbb{R}^n$. Note that without loss of generality, the case $c > 0$ can be reduced to $c = 1$. Unless stated otherwise, we will use $\oplus$ as $\oplus_1$ to simplify notations. For general $c > 0$, this operation is not commutative nor associative. However, it satisfies $x \oplus_c \mathbf{0} = \mathbf{0} \oplus_c x = \mathbf{0}$. Moreover, for any $x, y \in \mathbb{D}_c^n$, we have $(-x) \oplus_c x = x \oplus_c (-x) = \mathbf{0}$ and $(-x) \oplus_c (x \oplus_c y) = y$ (left-cancellation law). The *Möbius substraction* is then defined by the use of the following notation: $x \ominus_c y := x \oplus_c (-y)$. See [29, section 2.1] for a geometric interpretation of the Möbius addition.

其中，$c = \frac{1}{r^2}$

双曲空间下的神经网络运算

- (Hyperbolic Neural Network 2018,NIPS)

Möbius scalar multiplication. For $c > 0$, the *Möbius scalar multiplication* of $x \in \mathbb{D}_c^n \setminus \{\mathbf{0}\}$ by $r \in \mathbb{R}$ is defined as

$$r \otimes_c x := (1/\sqrt{c}) \tanh(r \tanh^{-1}(\sqrt{c}\|x\|)) \frac{x}{\|x\|}, \tag{7}$$

and $r \otimes_c \mathbf{0} := \mathbf{0}$. Note that similarly as for the Möbius addition, one recovers the Euclidean scalar multiplication when c goes to zero: $\lim_{c \to 0} r \otimes_c x = rx$. This operation satisfies desirable properties such as $n \otimes_c x = x \oplus_c \cdots \oplus_c x$ (n additions), $(r + r') \otimes_c x = r \otimes_c x \oplus_c r' \otimes_c x$ (scalar distributivity[3]), $(rr') \otimes_c x = r \otimes_c (r' \otimes_c x)$ (scalar associativity) and $|r| \otimes_c x / \|r \otimes_c x\| = x / \|x\|$ (scaling property).

其中，$c = \frac{1}{r^2}$

双曲空间下的神经网络运算

- (Hyperbolic Neural Network 2018,NIPS)

Distance. If one defines the generalized hyperbolic metric tensor g^c as the metric conformal to the Euclidean one, with conformal factor $\lambda_x^c := 2/(1 - c\|x\|^2)$, then the induced distance function on $(\mathbb{D}_c^n, g^c)$ is given by[4]

$$d_c(x, y) = (2/\sqrt{c}) \tanh^{-1} \left(\sqrt{c}\| - x \oplus_c y\|\right). \tag{8}$$

Again, observe that $\lim_{c \to 0} d_c(x, y) = 2\|x - y\|$, *i.e.* we recover Euclidean geometry in the limit[5]. Moreover, for $c = 1$ we recover $d_{\mathbb{D}}$ of Eq. (4).

其中，$c = \frac{1}{r^2}$

双曲空间下的神经网络运算

- (Hyperbolic Neural Network 2018,NIPS)

Parallel transport. Finally, we connect parallel transport (from $T_{\mathbf{0}}\mathbb{D}_c^n$) to gyrovector spaces with the following theorem, which we prove in appendix B.

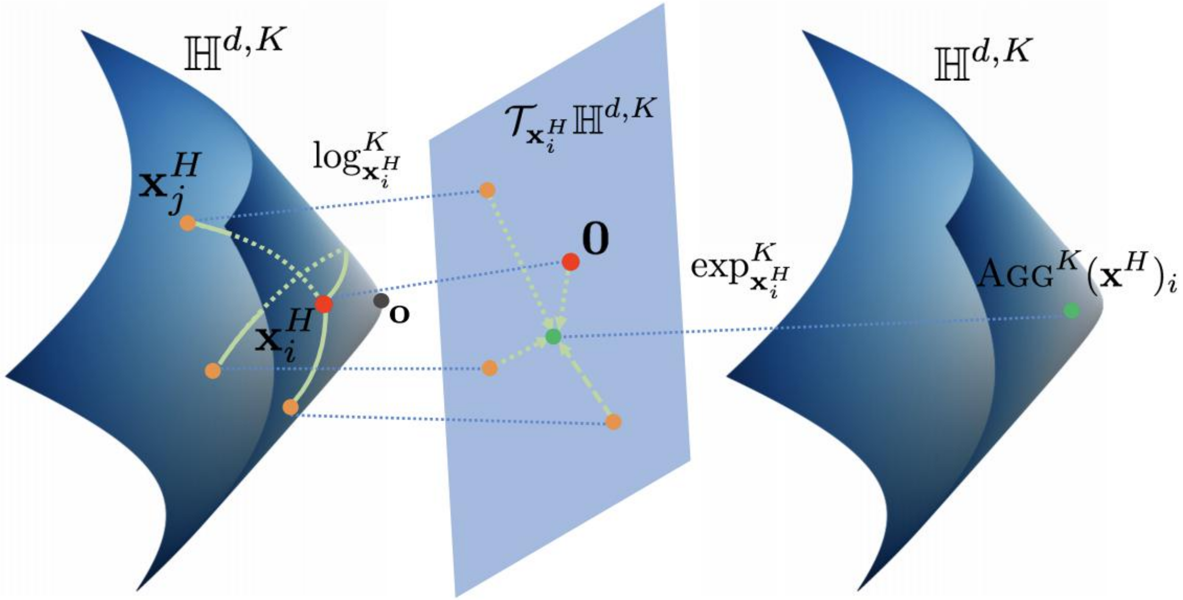
Theorem 4. *In the manifold* $(\mathbb{D}_c^n, g^c)$, *the parallel transport w.r.t. the Levi-Civita connection of a vector* $v \in T_{\mathbf{0}}\mathbb{D}_c^n$ *to another tangent space* $T_x\mathbb{D}_c^n$ *is given by the following isometry:*

$$P_{\mathbf{0}\to x}^c(v) = \log_x^c(x \oplus_c \exp_{\mathbf{0}}^c(v)) = \frac{\lambda_{\mathbf{0}}^c}{\lambda_x^c} v. \tag{16}$$

其中，$c = \frac{1}{r^2}$

双曲空间与欧式空间的转换

- log映射： $H^n \rightarrow R^n$
- exp映射: $R^n \rightarrow H^n$



Definition 3.2 (Möbius version). For $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$, we define the *Möbius version of* f as the map from $\mathbb{D}_c^n$ to $\mathbb{D}_c^m$ by:

$$f^{\otimes_c}(x) := \exp_{\mathbf{0}}^c(f(\log_{\mathbf{0}}^c(x))), \tag{26}$$

where $\exp_{\mathbf{0}}^c : T_{\mathbf{0}_m}\mathbb{D}_c^m \rightarrow \mathbb{D}_c^m$ and $\log_{\mathbf{0}}^c : \mathbb{D}_c^n \rightarrow T_{\mathbf{0}_n}\mathbb{D}_c^n$.

HGCN

- 结合图卷积和双曲嵌入，在双曲面模型上做GCN,在结合聚合上做GAT。曲率可训练。
- N维度双曲面 $x^2 - y^2 - z^2 = 1 \rightarrow x_0^2 - (x_1^2 + \cdots + x_n^2) = 1$

内积定义 $\langle \mathbf{x}, \mathbf{y} \rangle_{\mathcal{L}} := -x_0 y_0 + x_1 y_1 + \ldots + x_d y_d.$

双曲面上的距离 $d_{\mathcal{L}}^K(\mathbf{x}, \mathbf{y}) = \sqrt{K} \operatorname{arcosh}(-\langle \mathbf{x}, \mathbf{y} \rangle_{\mathcal{L}} / K).$

双曲面上的映射

$$\exp_{\mathbf{x}}^K(\mathbf{v}) = \cosh\left(\frac{||\mathbf{v}||_{\mathcal{L}}}{\sqrt{K}}\right)\mathbf{x} + \sqrt{K}\sinh\left(\frac{||\mathbf{v}||_{\mathcal{L}}}{\sqrt{K}}\right)\frac{\mathbf{v}}{||\mathbf{v}||_{\mathcal{L}}}, \quad \log_{\mathbf{x}}^K(\mathbf{y}) = d_{\mathcal{L}}^K(\mathbf{x}, \mathbf{y})\frac{\mathbf{y} + \frac{1}{K}\langle \mathbf{x}, \mathbf{y} \rangle_{\mathcal{L}}\mathbf{x}}{||\mathbf{y} + \frac{1}{K}\langle \mathbf{x}, \mathbf{y} \rangle_{\mathcal{L}}\mathbf{x}||_{\mathcal{L}}}.$$

算法框架

- 输入（欧式空间转化为双曲空间）

$$\mathbf{x}^{0,H} = \exp_{\mathbf{o}}^{K}((0, \mathbf{x}^{0,E})) = \left(\sqrt{K}\cosh\left(\frac{||\mathbf{x}^{0,E}||_2}{\sqrt{K}} \right), \sqrt{K}\sinh\left(\frac{||\mathbf{x}^{0,E}||_2}{\sqrt{K}} \right) \frac{\mathbf{x}^{0,E}}{||\mathbf{x}^{0,E}||_2} \right).$$

- 线性变换

$$W \otimes^K \mathbf{x}^H := \exp_{\mathbf{o}}^K (W \log_{\mathbf{o}}^K (\mathbf{x}^H)),$$

$$\mathbf{x}^H \oplus^K \mathbf{b} := \exp_{\mathbf{x^H}}^K (P_{\mathbf{o} \to \mathbf{x}^H}^K (\mathbf{b})).$$

算法框架

- 节点聚合

$$w_{ij} = \textsc{Softmax}_{j\in\mathcal{N}(i)}(\mathrm{MLP}(\log_{\mathbf{o}}^{K}(\mathbf{x}_i^H)||\log_{\mathbf{o}}^{K}(\mathbf{x}_j^H)))$$

$$\textsc{Agg}^K(\mathbf{x}^H)_i = \exp_{\mathbf{x}_i^H}^{K}\left(\sum_{j\in\mathcal{N}(i)} w_{ij}\log_{\mathbf{x}_i^H}^{K}(\mathbf{x}_j^H)\right).$$

- 非线性激活

$$\sigma^{\otimes^{K_{\ell-1},K_\ell}}(\mathbf{x}^H) = \exp_{\mathbf{o}}^{K_\ell}(\sigma(\log_{\mathbf{o}}^{K_{\ell-1}}(\mathbf{x}^H))).$$

- 解码

$$p((i,j)\in\mathcal{E}|\mathbf{x}_i^{L,H},\mathbf{x}_j^{L,H}) = \left[e^{(d_{\mathcal{L}}^{K_L}(\mathbf{x}_i^{L,H},\mathbf{x}_j^{L,H})^2-r)/t}+1\right]^{-1},$$

	Dataset	Disease		Disease-M		Human PPI		Airport		PubMed		Cora	
	Hyperbolicity δ	$\delta = 0$		$\delta = 0$		$\delta = 1$		$\delta = 1$		$\delta = 3.5$		$\delta = 11$	
	Method	LP	NC	LP	NC	LP	NC	LP	NC	LP	NC	LP	NC
Shallow	Euc	59.8 ± 2.0	32.5 ± 1.1	-	-	-	-	92.0 ± 0.0	60.9 ± 3.4	83.3 ± 0.1	48.2 ± 0.7	82.5 ± 0.3	23.8 ± 0.7
	Hyp [29]	63.5 ± 0.6	45.5 ± 3.3	-	-	-	-	94.5 ± 0.0	70.2 ± 0.1	87.5 ± 0.1	68.5 ± 0.3	87.6 ± 0.2	22.0 ± 1.5
	Euc-Mixed	49.6 ± 1.1	35.2 ± 3.4	-	-	-	-	91.5 ± 0.1	68.3 ± 2.3	86.0 ± 1.3	63.0 ± 0.3	84.4 ± 0.2	46.1 ± 0.4
	Hyp-Mixed	55.1 ± 1.3	56.9 ± 1.5	-	-	-	-	93.3 ± 0.0	69.6 ± 0.1	83.8 ± 0.3	73.9 ± 0.2	85.6 ± 0.5	45.9 ± 0.3
NN	MLP	72.6 ± 0.6	28.8 ± 2.5	55.3 ± 0.5	55.9 ± 0.3	67.8 ± 0.2	55.3 ± 0.4	89.8 ± 0.5	68.6 ± 0.6	84.1 ± 0.9	72.4 ± 0.2	83.1 ± 0.5	51.5 ± 1.0
	HNN[10]	75.1 ± 0.3	41.0 ± 1.8	60.9 ± 0.4	56.2 ± 0.3	72.9 ± 0.3	59.3 ± 0.4	90.8 ± 0.2	80.5 ± 0.5	94.9 ± 0.1	69.8 ± 0.4	89.0 ± 0.1	54.6 ± 0.4
GNN	GCN[21]	64.7 ± 0.5	69.7 ± 0.4	66.0 ± 0.8	59.4 ± 3.4	77.0 ± 0.5	69.7 ± 0.3	89.3 ± 0.4	81.4 ± 0.6	91.1 ± 0.5	78.1 ± 0.2	90.4 ± 0.2	81.3 ± 0.3
	GAT [41]	69.8 ± 0.3	70.4 ± 0.4	69.5 ± 0.4	62.5 ± 0.7	76.8 ± 0.4	70.5 ± 0.4	90.5 ± 0.3	81.5 ± 0.3	91.2 ± 0.1	79.0 ± 0.3	$\mathbf{93.7} \pm 0.1$	$\mathbf{83.0} \pm 0.7$
	SAGE [15]	65.9 ± 0.3	69.1 ± 0.6	67.4 ± 0.5	61.3 ± 0.4	78.1 ± 0.6	69.1 ± 0.3	90.4 ± 0.5	82.1 ± 0.5	86.2 ± 1.0	77.4 ± 2.2	85.5 ± 0.6	77.9 ± 2.4
	SGC [44]	65.1 ± 0.2	69.5 ± 0.2	66.2 ± 0.2	60.5 ± 0.3	76.1 ± 0.2	71.3 ± 0.1	89.8 ± 0.3	80.6 ± 0.1	94.1 ± 0.0	78.9 ± 0.0	91.5 ± 0.1	81.0 ± 0.1
Ours	HGCN	$\mathbf{90.8} \pm 0.3$	$\mathbf{74.5} \pm 0.9$	$\mathbf{78.1} \pm 0.4$	$\mathbf{72.2} \pm 0.5$	$\mathbf{84.5} \pm 0.4$	$\mathbf{74.6} \pm 0.3$	$\mathbf{96.4} \pm 0.1$	$\mathbf{90.6} \pm 0.2$	$\mathbf{96.3} \pm 0.0$	$\mathbf{80.3} \pm 0.3$	92.9 ± 0.1	79.9 ± 0.2
	(%) Err Red	-63.1%	-13.8%	-28.2%	-25.9%	-29.2%	-11.5%	-60.9%	-47.5%	-27.5%	-6.2%	+12.7%	+18.2%