Definition 2 (κ-left-matrix-multiplication). *Given $\mathbf{Z} \in \mathbb{R}^{N\times D}$ holding κ-stereographic vectors in its row and weights $\mathbf{A} \in \mathbb{R}^{N\times N}$, the κ-left-matrix-multiplication of $\mathbf{Z}$ by $\mathbf{A}$ is defined as follows:*

$$(\mathbf{A} \boxtimes_\kappa \mathbf{Z})_{i\bullet} := A \otimes_\kappa \mathbf{mid}_\kappa \left(\{\mathbf{Z}_{i\bullet}\}_{i=1}^n; \mathbf{A}_{i\bullet}\right), \tag{24}$$

where $A = \sum_j \mathbf{A}_{ij}$, $\mathbf{mid}_\kappa$ denotes midpoint in the κ-stereographic model.

Attentional Aggregation in κ-stereographic Model

In this part, we show that the κ-left-matrix-multiplication performs the attentional aggregation in κ-stereographic model. In other words, we prove **Theorem 1** in the subsection of *attentional aggregation layer* in our paper.

With the formal definition of the linear combination, we rewrite the **Theorem 1** equivalently as follows:

Theorem 1 (κ-left-matrix-multiplication as attentional aggregation). *Let rows of $\mathbf{H}$ hold the encoding $\boldsymbol{z}_{\mathcal{M}_i}$, (linear transformed by $\mathbf{W}$) and $\mathbf{A}$ hold the attentional weights, the κ-left-matrix-multiplication $\mathbf{A} \boxtimes_\kappa \mathbf{H}$ performs the attentional aggregation over the rows of $\mathbf{H}$, i.e., $\mathbf{A} \boxtimes_\kappa \mathbf{H}$ is the row-wise linear combination of $\mathbf{H}$ with respect to attentional weight $\mathbf{A}_{ij}$:*

$$(\mathbf{A} \boxtimes_\kappa \mathbf{H})_{i\bullet} = \boldsymbol{L}(c_{stereo}(\mathbf{A}_{ij}), \mathbf{h}_j), \tag{25}$$

where $\mathbf{h}_j$ is the j^{th} row of $\mathbf{H}$, j enumerates the index set Ψ, $\Psi = i \cup \mathcal{N}_i$ and $\mathcal{N}_i$ is the neighbors of i on the graph. $c_{stereo}(\cdot)$ is the function to output the coefficient in the κ-stereographic model.

Proof. Recall the design of the (intra-component) attentions in our paper: $\mathbf{A}$ is given as $\hat{\mathbf{A}} + \mathbf{I}$, where $\hat{\mathbf{A}}$ is filled with softmax values in its row, and $\mathbf{I}$ is the identity matrix to keep the initial information of the node itself. That is, we have the row sum, $A = 2$. Then, with the definitions of *κ-left-matrix-multiplication* and *midpoint in the κ-stereographic model*, we give the derivation as follows:

$$\begin{aligned}
&(\mathbf{A} \boxtimes_\kappa \mathbf{H})_{i\bullet} \\
=&A \otimes_\kappa \mathbf{mid}_\kappa \left(\{\mathbf{h}_j\}_{j=1}^n; \{\mathbf{A}_{ij}\}_{j=1}^n\right) \\
=&2 \otimes_\kappa \frac{1}{2} \otimes_\kappa \left(\sum_{j=1}^n \frac{\mathbf{A}_{ij}\lambda^\kappa_{\mathbf{h}_j}}{\sum_{l=1}^n \mathbf{A}_{il}(\lambda^\kappa_{\mathbf{h}_l} - 1)} \mathbf{h}_j \right) \\
=&\sum_{j=1}^n \frac{\mathbf{A}_{ij}\lambda^\kappa_{\mathbf{h}_j}}{\sum_{l=1}^n \mathbf{A}_{il}(\lambda^\kappa_{\mathbf{h}_l} - 1)} \mathbf{h}_j \\
=&\boldsymbol{L}(c_{\text{stereo}}(j), \mathbf{h}_j),
\end{aligned} \tag{26}$$

where the coefficient function is given as $c_{\text{stereo}}(j) = \frac{1}{C}\mathbf{A}_{ij}\lambda^\kappa_{\mathbf{h}_j}$, and $C = \sum_{l=1}^n \mathbf{A}_{il}(\lambda^\kappa_{\mathbf{h}_l} - 1)$. As shown above, with the well-designed attention mechanism, we eliminate the κ-scaling and make the κ-left-matrix-multiplication to be the linear combination with respect to the learnable attentions, performing the attentional aggregation. □

Table 2: The statistics of the datasets.

Dataset	#(Node)	#(Links)	#(Labels)
Citeseer	$3,327$	$4,732$	6
Cora	$2,708$	$5,429$	7
Pubmed	$19,717$	$44,338$	3
Amazon	$13,381$	$245,778$	10
Airport	$1,190$	$13,599$	4

B. Experimental Details

In this section, we give further experimental details, including data & code and implementation notes, in order to enhance the *reproducibility*.

Data and Code

Data The datasets used in this paper are publicly available, i.e., Citeseer, Cora, Pubmed, Amazon and Airport. We briefly describe these datasets as follows:

- Citeseer , Cora and Pubmed are the widely used citation networks, where nodes represent papers, and edges represent citations between them.
- The Amazon is a co-purchase graph, where nodes represent goods and edges indicate that two goods are frequently bought together.
- The Airport is an air-traffic graph, where nodes represent airports and edges indicate the traffic connection between them.

We list the statistics of the datasets in Table 2.

Code We submit the source code of an instance implementation of SELFMGNN in a ZIP named *Code*, and will publish the source code after acceptance.

Implementation Notes

In SELFMGNN, we stack the attentive aggregation layer twice to learn the component embedding. We employ a two-layer MLP$_\kappa$ in the Riemannian projector to reveal the Riemannian views for the self-supervised learning. In the experiments, we set the weight γ to be 1, i.e., the single-view and cross-view contrastive learning are considered to have the same importance. The grid search is performed over the learning rate in $[0.001, 0.003, 0.005, 0.008, 0.01]$ as well as the dropout probability in $[0, 0.8]$ with the step size of 0.1.

For all the comparison model, we perform a hyper-parameter search on a validation set to obtain the best results, and the κ-GCN is trained with positive curvature in particular to evaluate the representation ability of the spherical space. We set the dimensionality to be 24 for all the models for the fair comparison. Note that, in SELFMGNN, the component space can be set to arbitrary dimensionality, whose curvature and importance are learned from the data, and thereby we construct a mixed-curvature space of any dimensionality, matching the curvatures of any datasets.

Fugit commodi dicta, vitae fuga ipsa cumque rem vero sequi illum, qui deleniti inventore dolore ut blanditiis fugiat aliquid consectetur magni, deserunt cum soluta numquam odit?Accusantium quisquam voluptas excepturi consequuntur ducimus dolore quidem error quaerat facere, omnis tempora corporis voluptatibus aliquam