SRBGCN: Tangent space-Free Lorentz Transformations for Graph Feature Learning

Abdelrahman Mostafa[1], Wei Peng[2], Guoying Zhao[1]

[1] Center for Machine Vision and Signal Analysis, University of Oulu, Finland

[2] Department of Psychiatry and Behavioral Sciences, Stanford University, California, USA

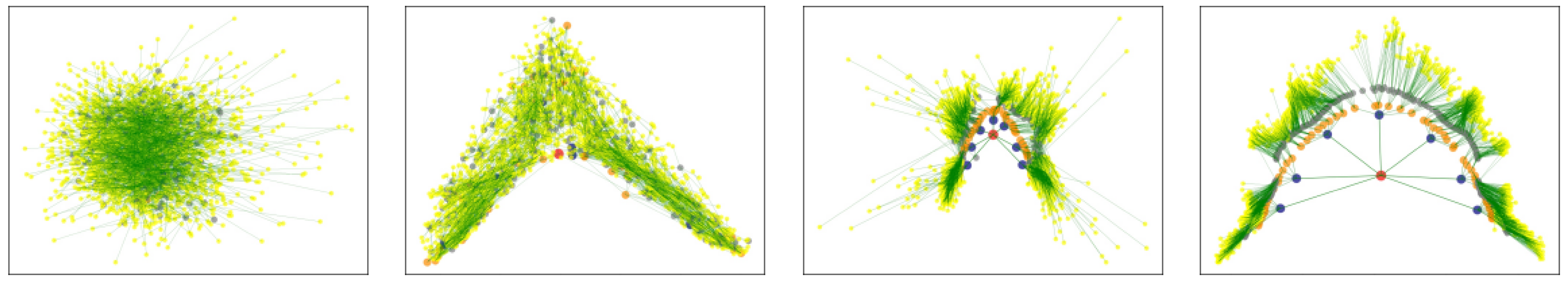
Introduction and Motivation

-A graph is a data structure that can be used to model many problems with entities (graph nodes) and the relations between these entities (graph edges). Graph Convolutional Networks (GCNs) were proposed to generalize the convolution operation to graph structures.

-However, these models embed the features into the Euclidean space which has been shown to incur a large distortion. The hyperbolic space is ideal for embedding trees as the number of tree nodes is growing exponentially with respect to the tree depth. Motivated by this, recent works built GCNs in the hyperbolic space. HGCNs achieved better performance than the corresponding Euclidean one.

-However, these works performed the network operations in the tangent space of the manifold which is a Euclidean local approximation to the manifold at a point. In this work, we propose a manifold-preserving feature transformations to build SRBGCN without resorting to the tangent space.

Keywords: Hyperbolic geometry, Lorentz transformations, graph convolutional networks.



(a) Initial embeddings (b) GCN (c) HGCN (d) Ours

Figure 1: Distortion visualization

Methods in SRBGCN

-Lorentz transformations are linear transformations which are manifold-preserving and can be used to transform features in the hyperbolic space.

-A Lorentz transformation matrix Λ should satisfy:

$$\Lambda^T g_{\mathcal{L}} \Lambda = g_{\mathcal{L}} \quad (1)$$

where $g_{\mathcal{L}} = diag(-1, 1, \ldots, 1)$ is a diagonal matrix that represents the Riemannian metric for the hyperbolic manifold. A Lorentz transformation matrix is orthogonal with respect to the Minkowski metric $g_{\mathcal{L}}$.

-The Lorentz transformation can be decomposed into two operations; The Spatial Rotation operation matrix is given by:

$$\mathbf{P} = \begin{bmatrix} 1 & 0 \\ 0 & \mathbf{Q} \end{bmatrix}_{(d+1)\times(d+1)} \quad (2)$$

where $\mathbf{Q}$ belongs to the special orthogonal group $SO(d)$ i.e. $\mathbf{Q^T Q} = \mathbf{I}$. The Boost operation matrix is given by:

$$\mathbf{L} = \begin{bmatrix} \cosh\omega & (\sinh\omega) n_d^T \\ (\sinh\omega) n_d & \mathbf{I} - (1 - \cosh\omega) n_d \otimes n_d \end{bmatrix}_{(d+1)\times(d+1)} \quad (3)$$

where $\otimes$ represents the outer product operation.

-Lorentz centroid is used for feature aggregation in SRBGCN.

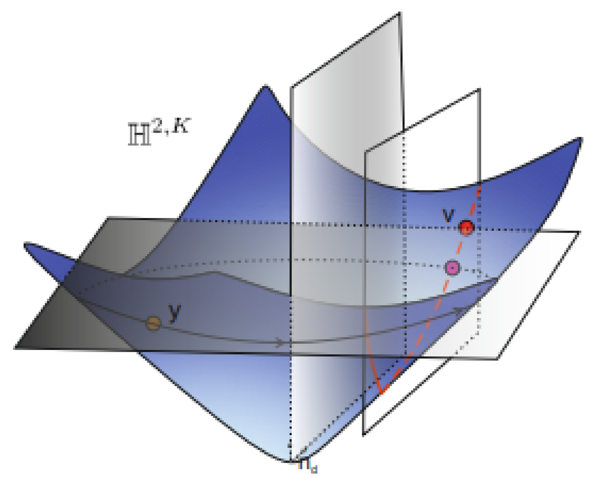


Figure 2: feature transformation

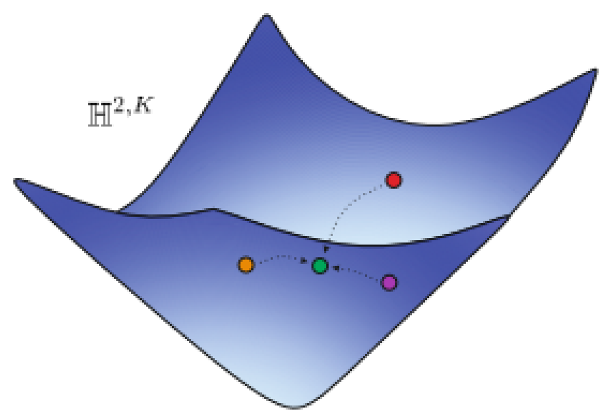


Figure 3: feature aggregation

Experiments

	Dataset	Disease ($\delta = 0$)		Airport ($\delta = 1$)		PubMed ($\delta = 3.5$)		Cora ($\delta = 11$)	
	Method	LP	NC	LP	NC	LP	NC	LP	NC
Euclidean	GCN	64.7±0.5	69.7±0.4	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	69.8±0.3	70.4±0.4	90.5±0.3	81.5±0.3	91.2±0.1	79.0±0.3	93.7±0.1	83.0±0.7
	SAGE	65.9±0.3	69.1±0.6	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	65.1±0.2	69.5±0.2	89.8±0.3	80.6±0.1	94.1±0.0	78.9±0.0	91.5±0.1	81.0±0.1
Hyberpolic	HGCN	91.2±0.6	82.8±0.8	96.4±0.1	90.6±0.2	96.1±0.2	78.4±0.4	93.1±0.4	81.3±0.6
	HAT	91.8±0.5	83.6±0.9	-	-	96.0±0.3	78.6±0.5	93.0±0.3	83.1±0.6
	LGCN	96.6±0.6	84.4±0.8	96.0±0.6	90.9±1.7	96.8±0.1	78.6±0.7	93.6±0.3	**83.3±0.7**
	HYPONET	96.8±0.4	**96.0±1.0**	97.3±0.3	90.9±1.4	95.8±0.2	78.0±1.0	93.6±0.3	80.2±1.3
	SRBGCN	**97.3±0.2**	93.0±0.4	**97.3±0.0**	**91.6±0.9**	**97.2±0.0**	**79.1±0.3**	**95.2±0.0**	82.9±0.2

Table 1: Evaluation results and comparison with other methods (dim is 16). ROC AUC results are reported for Link Prediction (LP) tasks and F1 scores are reported for Node Classification (NC) tasks

Dataset	dim	GAT	HGCN	HAT	LGCN	HYPONET	SRBGCN
Disease	4	49.4±6.3	73.2±6.5	-	87.4±3.1	91.0±3.8	**93.1±0.3**
	8	76.7±0.7	81.5±1.3	82.3±1.2	82.9±1.2	92.9±1.0	**93.3±0.4**
Cora	64	83.1±0.6	82.1±0.7	83.1±0.5	83.5±0.5	81.5±0.9	**83.8±0.3**

Table 2: Comparison between different methods using different dimensions (dim) on the Disease and Cora datasets for the node classification task

Dataset	Disease		Airport		PubMed		Cora	
Transformation	LP	NC	LP	NC	LP	NC	LP	NC
B only $\mathbf{Y}_h^l = \mathbf{X}_h^l \mathbf{L}^l$	97.2±0.3	91.4±1.9	94.4±2.8	87.3±1.4	96.9±0.1	76.6±1.7	94.1±0.2	81.7±0.7
SR only $\mathbf{Y}_h^l = \mathbf{X}_h^l \mathbf{P}^l$	97.2±0.3	92.1±0.6	**97.3±0.0**	89.7±1.4	**97.2±0.0**	78.8±0.4	**95.2±0.0**	81.4±0.4
SR and B $\mathbf{Y}_h^l = \mathbf{X}_h^l \mathbf{P}^l \mathbf{L}^l$	**97.3±0.2**	**93.3±0.4**	96.8 ±0.0	**91.6±0.9**	**97.2±0.0**	**79.1±0.3**	94.3±0.0	**83.8±0.3**

Table 3: **Ablation study** on different datasets to show the effect of using only the spatial rotation (SR) operation, the boost (B) operation and both the spatial rotation and the boost (SR and B) operations

-Four publicly available static graph datasets: Disease, Airport, PubMed and Cora. Table 1 shows the performance of different methods. For Disease dataset that has a tree structure with depth of 4, our method achieved a very good performance using dimension of 4 or 8 compared to other methods. For larger datasets with higher δ-hyperbolicity, our method achieved better performance using higher dimensional latent representation. Table 2 shows this comparison between the different methods.

-Table 3 shows an Ablation study using different operations. Table 4 shows distortion values for the Disease and Airport datasets. Our model preserves the graph structure of the dataset as can be seen in Figure 1 for the Disease dataset.

Dataset	GCN	HGCN	SRBGCN
Disease	67.92±54.91	1.04±0.55	**0.35±0.03**
Airport	175.02±216.90	1.39±0.64	**0.27±0.00**

Table 4: Distortion for Disease and Airport datasets

Contact Information

Email {Abdelrahman.Mostafa,Guoying.Zhao}@oulu.fi,wepeng@stanford.edu