Hyperbolic Graph Neural Networks

facebook Artificial Intelligence Research

Qi Liu, Maximilian Nickel, Douwe Kiela

Ground GNN with Hyperbolic Geometry

Goal: Supervised learning on entire graphs.

Motivation:

- Graph neural networks can exploit symmetries in graph-structured data, showing promises for the classification of graphs based on their structural properties.
- Typical properties of complex networks such as heterogeneous degree distributions and strong clustering can often be explained by assuming an underlying hierarchy, which is well captured in hyperbolic space (Krioukov et al., 2010).

Main Idea: We extend graph neural networks to operate on Riemannian manifolds with differentiable exponential and logarithmic maps.

Hyperbolic Graph Neural Networks

Graph neural networks can be interpreted as performing message passing between nodes (Gilmer et al., 2017).

Vanilla GCN:

$$\mathsf{h}_u^{k+1} = \sigma\left(\sum_{v\in\mathcal{I}(u)} \tilde{\mathsf{A}}_{uv}\mathsf{W}^k\mathsf{h}_v^k\right),$$

where $\tilde{\mathsf{A}} = \mathsf{D}^{-\frac{1}{2}}(\mathsf{A}+\mathsf{I})\mathsf{D}^{-\frac{1}{2}}$ captures the connectivity of the graph.

We generalize the notion of a graph convolutional network such that the network operates on Riemannian manifolds and becomes agnostic to the underlying space. Since the tangent space of a point on Riemannian manifolds always is Euclidean (or a subset of Euclidean space), functions with trainable parameters are executed there.

HGNN:

$$\mathsf{h}_u^{k+1} = \sigma\left(\exp_{\mathsf{x}'}(\textstyle\sum_{v\in\mathcal{I}(u)} \tilde{\mathsf{A}}_{uv}\mathsf{W}^k \log_{\mathsf{x}'}(\mathsf{h}_v^k))\right).$$

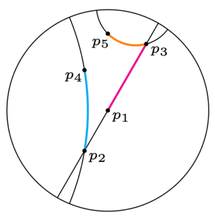
When applying the non-linearity directly on a manifold $\mathcal{M}$, we need to ensure that its application is manifold preserving, i.e., that $\sigma : \mathcal{M} \to \mathcal{M}$.

Riemannian Manifolds

A Riemannian manifold $(\mathcal{M}, g)$ is a real and smooth manifold equipped with an inner product $g_{\mathsf{x}} : \mathcal{T}_{\mathsf{x}}\mathcal{M} \times \mathcal{T}_{\mathsf{x}}\mathcal{M} \to \mathbb{R}$ at each point $\mathsf{x} \in \mathcal{M}$, which is called a Riemannian metric.

Euclidean Space: The Euclidean manifold is a manifold with zero curvature.

$$\begin{aligned}\exp_{\mathsf{x}}(\mathsf{v}) &= \mathsf{x}+\mathsf{v}\\ \log_{\mathsf{x}}(\mathsf{y}) &= \mathsf{y}-\mathsf{x}\end{aligned} \tag{1}$$

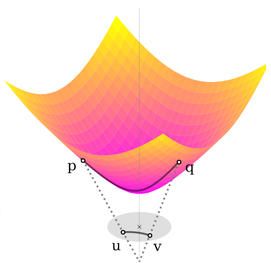


Poincaré Ball Model: The Poincaré ball model with constant negative curvature corresponds to the Riemannian manifold $(\mathcal{B}, g_{\mathsf{x}}^{\mathcal{B}})$ where $\mathcal{B} = \{\mathsf{x}\in\mathbb{R}^n : \|\mathsf{x}\| < 1\}$ is an open unit ball.

$$\begin{aligned}\exp_{\mathsf{x}}(\mathsf{v}) &= \cosh(\|\mathsf{v}\|_{\mathcal{L}})\mathsf{x} + \sinh(\|\mathsf{v}\|_{\mathcal{L}})\frac{\mathsf{v}}{\|\mathsf{v}\|_{\mathcal{L}}}\\ \log_{\mathsf{x}}(\mathsf{y}) &= \frac{\operatorname{arcosh}(-\langle\mathsf{x},\mathsf{y}\rangle_{\mathcal{L}})}{\sqrt{\langle\mathsf{x},\mathsf{y}\rangle_{\mathcal{L}}^2-1}}(\mathsf{y}+\langle\mathsf{x},\mathsf{y}\rangle_{\mathcal{L}}\mathsf{x}),\end{aligned} \tag{2}$$

Lorentz Model: The Lorentz model avoids numerical instabilities that may arise with the Poincare distance (mostly due to the division) (Nickel et al., 2018).

$$\begin{aligned}\exp_{\mathsf{x}}(\mathsf{v}) &= \cosh(\|\mathsf{v}\|_{\mathcal{L}})\mathsf{x} + \sinh(\|\mathsf{v}\|_{\mathcal{L}})\frac{\mathsf{v}}{\|\mathsf{v}\|_{\mathcal{L}}}\\ \log_{\mathsf{x}}(\mathsf{y}) &= \frac{\operatorname{arcosh}(-\langle\mathsf{x},\mathsf{y}\rangle_{\mathcal{L}})}{\sqrt{\langle\mathsf{x},\mathsf{y}\rangle_{\mathcal{L}}^2-1}}(\mathsf{y}+\langle\mathsf{x},\mathsf{y}\rangle_{\mathcal{L}}\mathsf{x}),\end{aligned} \tag{3}$$



Centroid-Based Regression and Classification

We propose an extension of the underlying idea of radial basis function networks (Poggio et al., 1990) to Riemannian manifolds.

Centroids: A list of centroids $\mathcal{C} = [\mathsf{c}_1, \mathsf{c}_2, ..., \mathsf{c}_{|\mathcal{C}|}]$, where each $\mathsf{c}_i \in \mathcal{M}$.

Pairwise Distance: $\psi_{ij} = d(\mathsf{c}_i, \mathsf{h}_j^K)$

Node-level Regression:

$$\hat{y} = \mathsf{w}_o^T(\psi_{1j}, ..., \psi_{|\mathcal{C}|j}), \tag{4}$$

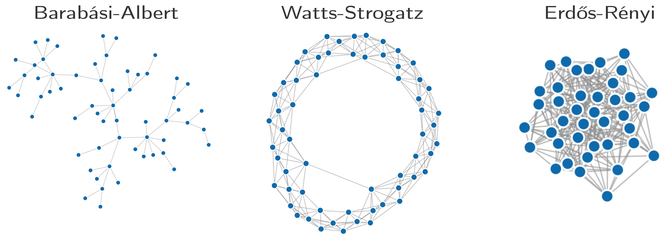
Node-level Classification:

$$p(y_j) = \operatorname{softmax}\left(\mathsf{W}_o(\psi_{1j}, ..., \psi_{|\mathcal{C}|j})\right), \tag{5}$$

Graph-level Prediction: Use average pooling to combine the distances of different nodes, obtaining $(\psi_1, ..., \psi_{|\mathcal{C}|})$, where $\psi_i = \sum_{j=1}^{|V|}\psi_{ij}/|V|$.

Synthetic Structures

We design a synthetic experiment, aiming to classify synthetically generated graphs according to the underlying generation algorithm. We choose 3 distinct graph generation algorithms: Erdős-Rényi (Erdős et al., 1959), Barabási-Albert (Barabási et al., 1999) and Watts-Strogatz (Watts et al., 1998).



	Dimensionality				
	3	5	10	20	256
Euclidean	77.2 ± 0.12	90.0 ± 0.21	90.6 ± 0.17	94.8 ± 0.25	95.3 ± 0.17
Poincare	93.0 ± 0.05	95.6 ± 0.14	95.9 ± 0.14	96.2 ± 0.06	93.7 ± 0.05
Lorentz	94.1 ± 0.03	95.1 ± 0.25	96.4 ± 0.23	96.6 ± 0.22	95.3 ± 0.28

Table 1: F1 (macro) score and standard deviation of classifying synthetically generated graphs according to the underlying graph generation algorithm (high is good).

Molecular Structures

We use ZINC (Irwin et al., 2012) for molecular property prediction, which has received attention as a reasonable benchmark for supervised learning on graphs.

	logP				
	3	5	10	20	256
Euclidean	6.7 ± 0.07	4.7 ± 0.03	4.7 ± 0.02	3.6 ± 0.00	3.3 ± 0.00
Poincare	5.7 ± 0.00	4.6 ± 0.03	3.6 ± 0.02	3.2 ± 0.01	3.1 ± 0.01
Lorentz	5.5 ± 0.02	4.5 ± 0.03	3.3 ± 0.03	2.9 ± 0.01	2.4 ± 0.02
	QED				
	3	5	10	20	256
Euclidean	22.4 ± 0.21	15.9 ± 0.14	14.5 ± 0.09	10.2 ± 0.08	6.4 ± 0.06
Poincare	22.1 ± 0.01	14.9 ± 0.13	10.2 ± 0.02	6.9 ± 0.02	6.0 ± 0.04
Lorentz	21.9 ± 0.12	14.3 ± 0.12	8.7 ± 0.04	6.7 ± 0.06	4.7 ± 0.00
	SAS				
	3	5	10	20	256
Euclidean	20.5 ± 0.04	16.8 ± 0.07	14.5 ± 0.11	9.6 ± 0.05	9.2 ± 0.08
Poincare	18.8 ± 0.03	16.1 ± 0.08	12.9 ± 0.04	9.3 ± 0.07	8.6 ± 0.02
Lorentz	18.0 ± 0.15	16.0 ± 0.15	12.5 ± 0.07	9.1 ± 0.08	7.7 ± 0.06

Table 2: Mean absolute error of predicting molecular properties: the water-octanal partition coefficient (logP); qualitative estimate of drug-likeness (QED); and synthetic accessibility score (SAS). Scaled by 100 for table formatting (low is good).

Molecular Structures

We also compare to three strong baselines on exactly the same data splits of the ZINC dataset: graph-gated neural networks (GGNN; Li et al., 2016), deep tensor neural networks (DTNN; Schütt et al., 2017) and message-passing neural networks (MPNN; Gilmer et al., 2017).

	logP	QED	SAS
DTNN (Schütt et al., 2017)	4.0 ± 0.03	8.1 ± 0.04	9.9 ± 0.06
MPNN (Gilmer et al., 2017)	4.1 ± 0.02	8.4 ± 0.05	9.2 ± 0.07
GGNN (Li et al., 2016)	3.2 ± 0.20	6.4 ± 0.20	9.1 ± 0.10
Euclidean	3.3 ± 0.00	6.4 ± 0.06	9.2 ± 0.08
Poincare	3.1 ± 0.01	6.0 ± 0.04	8.6 ± 0.02
Lorentz	2.4 ± 0.02	4.7 ± 0.00	7.7 ± 0.06

Table 3: Mean absolute error of predicting molecular properties logP, QED and SAS, as compared to current state-of-the-art deep learning methods. Scaled by 100 for table formatting (low is good).

Blockchain Transaction Graphs

We study the problem of predicting price fluctuations for the underlying asset of the Ethereum blockchain (Wood et al., 2014).

HGNN Bidirectional Extension:

$$\mathsf{h}_u^{k+1} = \sigma\left(\exp_u\Big(\sum_{v\in\mathcal{I}(u)} \tilde{\mathsf{A}}_{uv}\mathsf{W}^k \log_{\mathsf{x}'}(\mathsf{h}_v^k) + \sum_{v\in\mathcal{O}(u)} \tilde{\mathsf{A}}_{uv}\tilde{\mathsf{W}}^k \log_{\mathsf{x}'}(\mathsf{h}_v^k)\Big)\right), \tag{6}$$

	Dev	Test
Node2vec	54.10 ± 1.63	52.44 ± 1.10
ARIMA	54.50 ± 0.16	53.07 ± 0.06
Euclidean	56.15 ± 0.30	53.95 ± 0.20
Poincare	57.03 ± 0.28	54.41 ± 0.24
Lorentz	57.52 ± 0.35	55.51 ± 0.37

Table 4: Accuracy of predicting price fluctations (up-down) for the Ether/USDT market rate based on graph dynamics.

	"Whale" nodes	All nodes
Norm	0.20129	0.33178

Table 5: Average norm of influential "whale" nodes. Whales are significantly closer to the origin than average, indicating their importance.

Conclusion

- We proposed a method for generalizing graph neural networks to Riemannian manifolds, making them agnostic to the underlying space.
- We showed that the proposed architecture successfully made use of its geometrical properties in order to capture the hierarchical nature of the data.

References

Barabási, Albert-László and Réka Albert (1999). "Emergence of scaling in random networks". In: *Science* 286.5439, pp. 509–512.

Erdős, Paul and Alfréd Rényi (1959). "On random graphs, I". In: *Publicationes Mathematicae (Debrecen)* 6, pp. 290–297.

Gilmer, Justin et al. (2017). "Neural message passing for quantum chemistry". In: *Proceedings of the 34th International Conference on Machine Learning-Volume 70*. JMLR. org, pp. 1263–1272.

Irwin, John J et al. (2012). "ZINC: a free tool to discover chemistry for biology". In: *Journal of chemical information and modeling* 52.7, pp. 1757–1768.

Krioukov, Dmitri et al. (2010). "Hyperbolic geometry of complex networks". In: *Physical Review E* 82.3, p. 036106.

Li, Yujia et al. (2016). "Gated Graph Sequence Neural Networks". In: *4th International Conference on Learning Representations, ICLR 2016, San Juan, Puerto Rico, May 2-4, 2016, Conference Track Proceedings*.

Nickel, Maximilian and Douwe Kiela (2018). "Learning Continuous Hierarchies in the Lorentz Model of Hyperbolic Geometry". In: *Proceedings of the 35th International Conference on Machine Learning, ICML 2018, Stockholmsmässan, Stockholm, Sweden, July 10-15, 2018*, pp. 3776–3785.

Poggio, Tomaso and Federico Girosi (1990). "Networks for approximation and learning". In: *Proceedings of the IEEE* 78.9, pp. 1481–1497.

Schütt, Kristof T et al. (2017). "Quantum-chemical insights from deep tensor neural networks". In: *Nature communications* 8, p. 13890.

Watts, Duncan J and Steven H Strogatz (1998). "Collective dynamics of 'small-world'networks". In: *Nature* 393.6684, p. 440.

Wood, Gavin et al. (2014). "Ethereum: A secure decentralised generalised transaction ledger". In: *Ethereum project yellow paper* 151, pp. 1–32.