Computationally Tractable Riemannian Manifolds for Graph Embeddings

Calin Cruceru[1] Gary Bécigneul[1,2] Octavian-Eugen Ganea[1,2]
[1]Department of Computer Science, ETH Zurich [2]Computer Science and Artificial Intelligence Lab, MIT[3]

ETH zürich

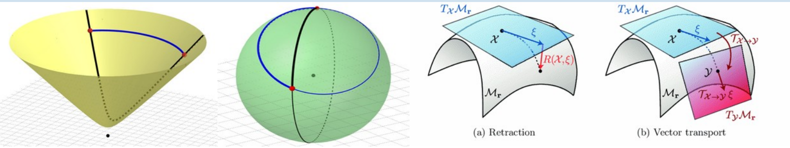
1 Problem & Context

High-level Motivation. Understand how the various ways in which spaces curve affects the types of graphs that can be accurately embedded into them.

In this Work. We study empirically how two families of Riemannian manifolds with novel curvature properties perform on the task of embedding graph nodes by matching metrics.

What We Know. (1) Cycles in graphs match elliptic geometry; (2) Trees and complex networks are intimately related to hyperbolic geometry; (3) Cartesian products of constant-curvature spaces sometimes better suit certain graphs.

On Tractability. It means being able to efficiently (i) compute distances between points on the manifold, and (ii) perform Riemannian optimization to adjust their placement.



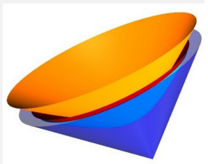
2 The Manifolds

I. $\mathcal{S}^{++}(n)$, the space of square real symmetric positive-definite matrices, with non-positive (and non-ct.) sectional curvature.

$$d(A,B) = \sqrt{\sum_{i=1}^{n} \log^2 \lambda_i(A^{-1}B)}$$ (canonical distance function)

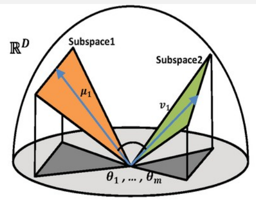
$$S(A,B) = \log\det\left(\frac{A+B}{2}\right) - \frac{1}{2}\log\det(AB)$$ (symmetric "Stein" divergence)

The latter is used as an alternative way of measuring dissimilarity, which is more computationally friendly (e.g., see gradients).



II. $Gr(k,n)$, the space of k-dimensional linear subspaces of $\mathbb{R}^n$ with non-negative (and non-ct.) sectional curvature.

Points are represented via $n \times k$ matrices with orthonormal columns. There are many possible metrics based on the principle angles $\{\theta_i\}_{i=1}^{k}$ obtained via the SVD of $A^\top B$.



We restrict to the canonical one:

$$d([A],[B]) = \sqrt{\sum_{i=1}^{k} \theta_i^2}$$

3 Methodology

- For each node u in a weighted graph G, we transductively learn an embedding $f(u) \in \mathcal{M}$ such that $d_{\mathcal{M}}(f(u), f(v)) \approx d_G(u,v)$. We propose to match them using Riemannian SNE, which in a sense subsumes the disparate loss functions used in prior work:

$$\mathcal{L}_{\text{RSNE}}(\{y_1, y_2, \dots, y_m\}) = \sum_{i=1}^{m} \mathrm{D}_{\mathrm{KL}}[p_i \parallel q_i],$$
$$p_{ij} = p(x_j \mid x_i) \propto \exp\left(-d_G^2(x_i,x_j)/T\right),$$
$$q_{ij} = q(y_j \mid y_i) \propto \exp\left(-d_{\mathcal{M}}^2(y_i,y_j)\right).$$

- We quantify the faithfulness of the embeddings via distortion metrics as well as a more fine-grained ranking metric, F1@k. For an embedding $f: G \to \mathcal{M}$, we define the precision and recall of a node u in the shortest-path tree rooted at $u \neq v$. Then, the F1 scores are defined as usual. The F1@k metric is the mean F1 score for all pairs of nodes at distance k.

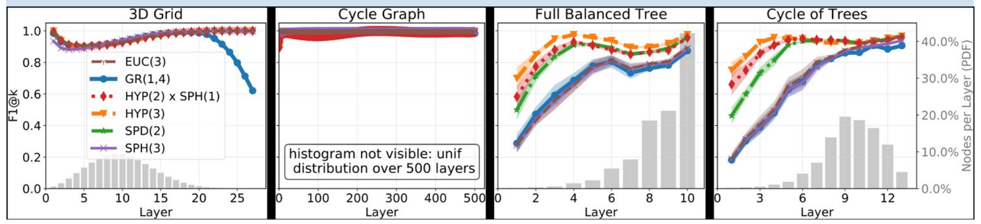
- We measure the extent to which the curved parts of an embedding space are leveraged by the learned embeddings via empirical sums of angles in geodesic triangles:

$$k_\theta(x,y,z) = \theta_{x,y} + \theta_{x,z} + \theta_{y,z},$$
$$\text{with } \theta_{x_1,x_2} = \cos^{-1}\frac{\langle u_1, u_2\rangle_{x_3}}{\|u_1\|_{x_3}\|u_2\|_{x_3}} \text{ and } u_{\{1,2\}} = \log_{x_3}(x_{\{1,2\}}).$$

The advantage: the values are bounded and easier to interpret than with previously used methods.

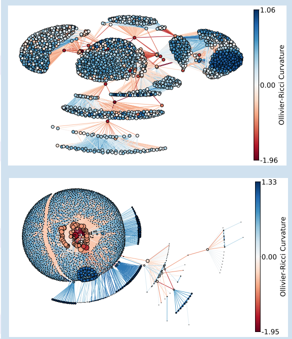
4 Experiments

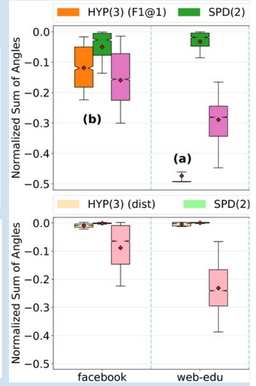
Synthetic Graphs. On the 10x10x10 grid and the 1000-nodes cycle all manifolds perform well. This is because every Riemannian manifold generalizes Euclidean space and Euclidean geometry suffices for grids and cycles (e.g., a cycle looks locally like a line). The more discriminative ones are the two other graphs which include trees. The best performing embeddings involve a hyperbolic component while the SPD ones rank between those and the non-negatively curved ones.



SPD Experiments. First, the (partial) negative curvature of the SPD and hyperbolic manifolds is beneficial: they outperform the flat Euclidean embeddings in almost all scenarios. This can be explained by the apparent scale-free nature of the input graphs. Second, we see that especially when using the S-divergence, which we attribute to the better-behaved optimization task, the SPD embeddings achieve significant improvements on the average distortion metric and are competitive and sometimes better on the ranking metrics. Finally, the rightmost figure shows a remarkably consistent pattern: the better performing embeddings yield more negatively-curved triangles.

Graph	Dim	Manifold	F1@1	AUC	Avg. Dist.
facebook	3	Euc	70.28	95.27	0.193
		Hyp	71.08	95.46	0.173
		SPD	71.09	95.26	0.170
		Stein	75.91	95.59	0.114
	6	Euc	79.60	96.41	0.090
		Hyp	81.83	96.53	0.089
		SPD	79.52	96.37	0.090
		Stein	83.95	96.74	0.061
web-edu	3	Euc	29.18	87.14	0.245
		Hyp	55.60	92.10	0.245
		SPD	29.02	88.54	0.246
		Stein	48.28	90.87	0.084
	6	Euc	49.31	91.19	0.143
		Hyp	66.23	95.78	0.143
		SPD	42.16	91.90	0.142
		Stein	62.81	96.51	0.043





Grassmann Experiments. We show here the results on two datasets: a road network in Minnesota and a dissimilarity dataset based on connectivities in a cat's cerebral cortex. Each one of them takes a column in this table. We show for each the metric that is most discriminative. While the results are less conclusive than before, we still see that the Grassmann embeddings achieve comparable or slightly better results than the other embedding spaces.

Dim	Manifold	F1@1 (road-minnesota)	Avg. Dist. (cat-cortex)
2	Euc	79.01	0.288
	Hyp	79.46	0.264
	Sphere	82.19	0.255
	$Gr(1,3)$	78.91	0.234
3	Euc	89.58	0.200
	Hyp	89.60	0.197
	Sphere	89.55	0.195
	$Gr(1,4)$	90.02	0.168
4	Euc	93.66	0.150
	Hyp	93.39	0.153
	Sphere	93.65	0.156
	$Gr(1,5)$	93.89	0.139
	$Gr(2,4)$	94.01	0.129

5 Conclusion

We proposed to use the SPD and Grassmann manifolds for learning representations of graphs and showed that they are competitive against previously considered constant-curvature spaces on the graph reconstruction task, consistently and significantly outperforming them in some cases.