Curvature and Representation Learning: Identifying Embedding Spaces for Relational Data

Melanie Weber[*]
Princeton University
mw25@math.princeton.edu

Maximilian Nickel
Facebook AI Research
maxn@fb.com

Abstract

We consider the problem of learning representations of relational data in spaces of constant sectional curvature, i.e., Euclidean, Hyperbolic, and Spherical space. In this context, we explore how to identify a suitable embedding curvature for a given relational dataset. For this task, we investigate the use of a scalable heuristic based on local graph neighborhoods and evaluate it on classic benchmark graphs.

1 Introduction

Representation learning has become an invaluable approach for learning from relational data. By modeling relationships via distances (or similarities) in an embedding space these methods enable high-quality models of relational data on a large-scale. Recently, new attention has been given to an important aspect of such methods, i.e., the geometry of the representation space. Methods such as hyperbolic embeddings [9, 10, 6] and Riemannian generative models [7, 5] showed that non-Euclidean geometries can provide significant advantages for modeling relational data. For instance, hyperbolic space is especially suited for modeling relational data with latent hierarchies. This leads to large gains in terms of representational efficiency such that tree-like graphs and complex networks can be embedded in much smaller dimensions.

However, at the same time, none of these embedding spaces are universally optimal for all relational structures and graphs (no free lunch). For instance, embeddings of quasi-cyclic graphs such as $n \times n$ square lattices and n-node cycles incur at least a multiplicative distortion of $\Omega(n/\log n)$ in hyperbolic space [14]. In spherical space, however, n-node cycles can trivially be embedded without distortion (under an appropriate constant rescaling of the distances). In the case of trees, on the other hand, this is reversed: any finite tree can be embedded arbitrarily well in hyperbolic space (up to machine precision), while they can incur large distortions in Euclidean and Spherical space [12, 6].

In this short paper, we focus therefore on identifying a suitable embedding space for a given relational dataset. We restrict our analysis to canonical Riemannian manifolds with constant sectional curvature κ, i.e., Euclidean ($\kappa = 0$), Spherical ($\kappa = 1$), and Hyperbolic space ($\kappa = -1$).

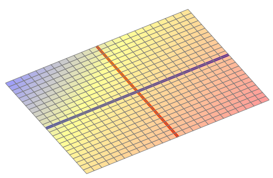
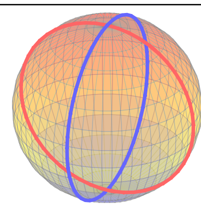
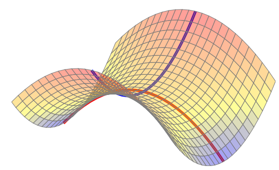
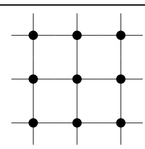
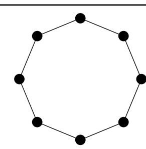
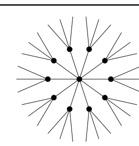
2 Methods

Let $(\mathcal{X}, d)$ be a metric space, where $\mathcal{X}$ is the underlying space and where $d : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is the distance function on the space. An embedding of $(\mathcal{X}, d_x)$ into $(\mathcal{Y}, d_y)$ is then a map $f : \mathcal{X} \to \mathcal{Y}$. Here, we are interested in embedding a graph metric (V, d_G) into a suitable metric space where d_G is the canonical shortest-path metric on graphs and where V denotes the vertices in a graph $G = (V, E)$. We will focus our attention on spaces $(\mathcal{X}, d)$ of *constant sectional curvature* κ and constant dimension n. There exist three complete, simply connected Riemannian manifolds with

[*]Work done while at Facebook AI Research, New York

32nd Conference on Neural Information Processing Systems (NIPS 2018), Montréal, Canada.

Table 1: Properties of model spaces with constant sectional curvature κ.

	Euclidean $\mathbb{R}^d$	**Spherical** $\mathbb{S}^d$	**Hyperboloid** $\mathbb{H}^d$
Space	$\mathbb{R}^n$	$\{x \in \mathbb{R}^{n+1} : \langle x, x \rangle = 1\}$	$\{x \in \mathbb{R}^{n+1} : \langle x, x \rangle = -1, x_0 > 0\}$
$\langle u, v \rangle$	$\sum_{i=1}^n u_i v_i$	$\sum_{i=1}^n u_i v_i$	$-u_0 v_0 + \sum_{i=1}^n u_i v_i$
$d(u, v)$	$\sqrt{\langle u - v, u - v \rangle}$	$\operatorname{arccos}(\langle u, v \rangle)$	$\operatorname{arccosh}(-\langle u, v \rangle)$
Curvature	$\kappa = 0$	$\kappa = 1$	$\kappa = -1$
Sum of angles	π	$> \pi$	$< \pi$
Circle length	$C(r) = 2\pi r$	$C(r) = 2\pi \sin r$	$C(r) = 2\pi \sinh r$
Disc area	$A(r) = 2\pi r^2/2$	$A(r) = 2\pi(1 - \cos r)$	$A(r) = 2\pi(\cosh r - 1)$
Principle Curvatures			
Characteristic Graph			

constant sectional curvature, i.e., Euclidean, Spherical, and Hyperbolic space. The properties of these spaces are listed in Table 1. Our goal is then to estimate a suitable embedding curvature $\kappa \in \{-1, 0, 1\}$ for a given graph G.

Deciding the curvature of a space is a difficult and computationally intensive task. For instance, Gromov's δ-*hyperbolicity* measures the "tree-likeness" of metric spaces and has been used to determine the hyperbolicity of graphs [1, 4]. However, δ-hyperbolicity is challenging to compute on large graphs (current state-of-the-art algorithms still have a worst case time complexity of ($O(|V|^4)$ [3]) and, moreover, only determines hyperbolicity and not whether Spherical or Euclidean space would be appropriate choices.

In Riemannian geometry, curvature measures how much a manifold deviates from being (locally) Euclidean. *Ricci curvature* in particular measures the amount by which the volume growth of distance balls deviates from their growth in Eucldiean space. Discretizations of Ricci curvature on graphs as proposed by [11] and [8] have recently been used to study geometric properties of networks [13, 15, 16]. However, Ollivier-Ricci curvature can be very expensive to compute even for mid-sized graphs as it involves solving an optimal transport problem for each edge in the graph. Forman-Ricci curvature, on the other hand, can be computed for large graphs but only considers very local information, i.e., the growth of 1-hop neighborhoods that are attached to an edge.

An alternative approach could be to estimate the *metric signature* of a graph, e.g., via spectral approaches as suggested in [17]. However, this method too suffers from serious challenges: First, it requires to compute the full all-pairs shortest-path matrix D_{sp} – which is infeasible for any large scale graph. Second, even in cases where computing D_{sp} is tractable, small changes in the graph can have large effects on the sign of its spectrum, making this method impractical for all graphs that are not perfectly embeddable in one of the model spaces.

In the following, we investigate an alternative approach based on graph motifs. This allows us to scale to large graphs while getting an estimate of global curvature from the volume-growth in local neighborhoods.

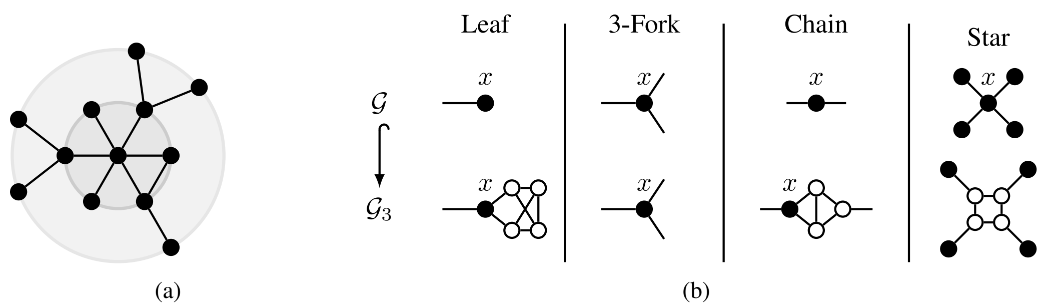


Figure 1: (a) Example neighborbood with $\delta(G) = 0$. (b) Embedding of G into a 3-regular graph G_3. Auxiliary nodes are inidicated as $\circ$.

2.1 Graph Motifs and Neighborhood Structures

We start from the observation that certain graph motifs are characteristic for certain embedding spaces, i.e., they can be embedded with small or no distortion in the matching space. These motifs are n-cycles for Spherical space, trees for Hyperbolic space and n-grids for Euclidean space (see also Table 1). An important property that makes embedding spaces suitable for these structures are their growth of area (or volume) with regard to the radius of a disc (or ball). For instance, trees benefit from a fast growing area in the embedding space, as the number of nodes in a tree grows exponentially with the level of the tree. In n-cycles on the other hand, the number of unique visited nodes stays constant for distances $d > n$. These different neighborhood growth rates correspond to the growth of disc area in hyperbolic and spherical space (see $A(r)$ in Table 1). A first approach could therefore be based on estimating the local growth rate in a neighborhood via such motifs. However, in an arbitrary input graph, motifs may be irregular and may overlap, which complicates their detection and analysis.

Thus, our first step is to *regularize* the input graph G by transforming it into a three-regular graph G_3, i.e. a graph in which every vertex has exactly three neighbors. For this purpose, we follow the approach outlined in [2]: The transformation depends on the node degree of each node and is shown in Figure 1b. In particular, we add auxiliary vertices and connections to achieve 3-regularity for *leaves* ($\deg(x) = 1$) and *chains* ($\deg(x) = 2$). The *3-fork* ($\deg(x) = 3$) is already 3-regular and therefore invariant under this transformation. For *stars* ($\deg(x) \geqslant 4$) which is characteristic for tree-like structure, we replace the center node by a ring of auxiliary nodes to gain 3-regularity. Remarkably, this transformation causes only a small distortion when the edges are appropriately re-weighted, as shown by [2]:

Theorem 2.1 ([2]). *$G \hookrightarrow^{\phi} G_3$ is a $(\epsilon + 1, \epsilon)$-quasi-isometric embedding, i.e.*

$$d_{G_3}(\phi(x), \phi(y)) \leq (\epsilon + 1) d_G(x, y) + \epsilon \ ,$$

where $\epsilon = \max_{e \in E(G)} l(e)$ denotes the maximum edge weight.

Given the regularized representation, we then estimate the growth of local neighborhoods as follows: Let $G_3 = (V, E)$ be the regularized graph and let $\mathcal{N}(v, r) = \{u \in V : d_G(u, v) = r\}$ be the set of all nodes that are exactly at distance r from v in G_3. We then say that a local neighborhood $\mathcal{N}(v) = \bigcup_{r=1}^{R} \mathcal{N}(v, r)$ is *expanding*, if

$$|\mathcal{N}(v, 1)| < |\mathcal{N}(v, 2)| < \ldots < |\mathcal{N}(v, R)| \ .$$

Otherwise, we say that $\mathcal{N}(v)$ is *contracting*.

For the characteristic graphs of a model space, the neighborhoods $\{\mathcal{N}(v)\}_{v \in V}$ will be uniformly expanding or contracting for all v. However, in real-world graphs, we will likely encounter a mixture of expanding and contracting neighborhoods. We therefore determine a suitable κ by computing the following weighted average (MOTIFCOUNT γ):

$$\gamma = \sum_{v \in V} \sigma(v) \, |\mathcal{N}(v)| \ ; \qquad \sigma(v) = \begin{cases} 1, & \mathcal{N}(v) \text{ contracting} \\ -1, & \mathcal{N}(v) \text{ expanding} \ . \end{cases}$$

Then, if $\gamma \gg 0$, we assume $\kappa = 1$; if $\gamma \ll 0$, assume $\kappa < 0$ and if $\gamma \approx 0$, assume $\kappa = 0$.

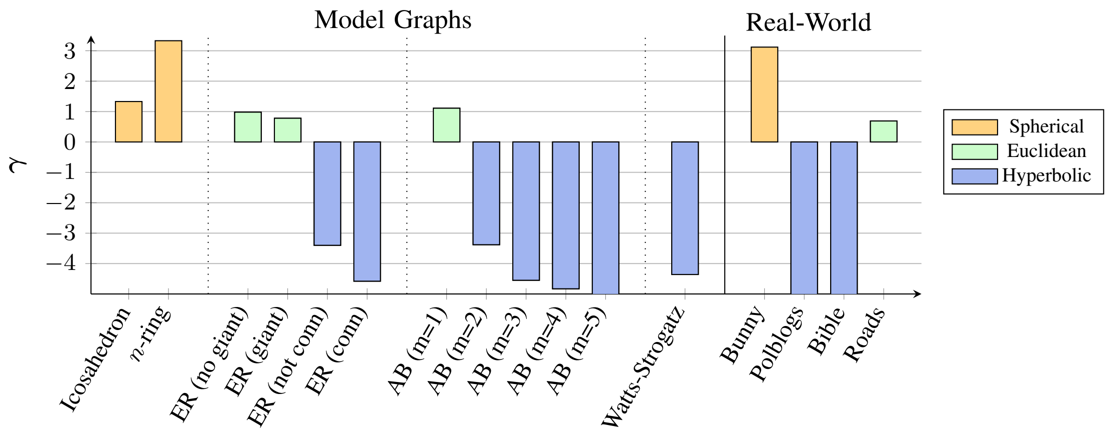


Figure 2: Experimental results MOTIFCOUNT.

The benefit of a regularized representation in this context is that it resolves overlaps between the motifs: In the transformation, auxiliary vertices are added to introduce a three-regular structure. This separates overlapping motifs, since the auxiliary vertices add to the neighborhood count, allowing for a realistic approximation of neighborhood growth rates without the need to determine multiplicities of vertices that are part of multiple motifs. Furthermore, this approach scales to large graphs as its runtime complexity is linear in the average size of a neighborhood.

3 Experiments

In this section our goal is to test the ability of MOTIFCOUNT to determine a suitable embedding curvature for relational data. For this purpose, we applied it to various model and real world graphs. The estimated values are shown in fig. 2. It can be seen that the estimated values match our theoretical expectations: For instance, regular polytope structures (icosahedron, n-cycle) are classified as *spherical*, which aligns with the fact that the corresponding 3-regular graphs are planar and can be isometrically embedded into the 3-sphere by Steinitz' theorem.

Furthermore, model networks as classically studied in network science are *hyperbolic* if they exhibit a latent hierarchical structure. To test this, we sampled ten networks each from the Albert-Barabasi, Watts-Strogatz and Erdös-Renyi models for each choice of hyperparameters and averaged the MOTIFCOUNT scores. Both the Albert-Barabasi and the Watts-Strogatz (Small world) model are classified as hyperbolic for typical hyperparameters. For the Albert-Barabasi model, we tested for different attachment parameters: Linear attachment ($m = 1$) results in a sparse graph with a large number of chains, making a Euclidean embedding space suitable. For superlinear attachment ($m > 1$), the network exhibits a community structure due to the presence of a few large hubs. In these cases, the networks are classified as hyperbolic.

For the Erdös-Renyi model, we study embeddability in the context of phase transition. The Erdös-Renyi model undergoes two phase transitions as the edge threshold increases: The emergence of a giant component at $p = 1/N$ and full connectivity at $p = \log N/N$. We generated ten networks from slightly above and slightly below each of the thresholds and averaged MOTIFCOUNT scores. We observe that in the sparse regime (i.e., below the connectivity threshold), the Euclidean space is the most suitable embedding space. As the networks become more and more connected, stars and tree-like structures emerge, creating hyperbolicity. Close to and above the connectivity threshold, the sample networks are then classified as hyperbolic. Our observations are consistent with theoretical analysis of Gromov's δ-hyperbolicity for the Erdös-Renyi and Small World models in [4].

We also tested MOTIFCOUNT on real-world networks (Figure 2). The Stanford bunny, a volumetric surface mesh commonly studied in computer graphics, is classified as spherical. The grid-like road network, resembling a lattice, is classified as Euclidean. We further tested networks with an implicit hierarchical structure, i.e., word co-occurrences in a text corpus (bible) and a network of political blogs. Both exhibit a community structure and are classified as hyperbolic.

4 Conclusion

This short paper introduced a heuristic method for identifying embedding spaces based on local volume growth. We presented preliminary experimental results that demonstrate the method on a few classic examples. Ongoing work includes a systematic test of the method on large and diverse data sets, as well as a careful analysis of its limitations. In particular, we investigate inaccuracies that are introduced by the method's local approach. For this purpose, we are extending our comparison to (theoretical) benchmarks, e.g. by comparing MOTIFCOUNT to δ-hyperbolicity. Furthermore, we are working on quantifying the *cost* (distortion) of embeddings in a "wrong" space, linking back to the importance of learning representations with suitable geometric priors.

References

[1] Aaron B Adcock, Blair D Sullivan, and Michael W Mahoney. Tree-like structure in large social and information networks. In *Data Mining (ICDM), 2013 IEEE 13th International Conference on*, pages 1–10. IEEE, 2013.

[2] Sergio Bermudo, Jose M Rodriguez, Jose M Sigarreta, and Jean-Marie Vilaire. Gromov hyperbolic graphs. *Discrete Mathematics*, 313(15):1575 – 1585, 2013.

[3] Michele Borassi, David Coudert, Pierluigi Crescenzi, and Andrea Marino. On computing the hyperbolicity of real-world graphs. In *Algorithms-ESA 2015*, pages 215–226. Springer, 2015.

[4] Wei Chen, Wenjie Fang, Guangda Hu, and Michael W Mahoney. On the hyperbolicity of small-world and treelike random graphs. *Internet Mathematics*, 9(4):434–491, 2013.

[5] Tim R Davidson, Luca Falorsi, Nicola De Cao, Thomas Kipf, and Jakub M Tomczak. Hyperspherical variational auto-encoders. *arXiv preprint arXiv:1804.00891*, 2018.

[6] Christopher De Sa, Albert Gu, Christopher Re, and Frederic Sala. Representation tradeoffs for hyperbolic embeddings. *arXiv preprint arXiv:1804.03329*, 2018.

[7] Luca Falorsi, Pim de Haan, Tim R Davidson, Nicola De Cao, Maurice Weiler, Patrick Forre, and Taco S Cohen. Explorations in homeomorphic variational auto-encoding. *arXiv preprint arXiv:1807.04689*, 2018.

[8] R. Forman. Bochner's Method for Cell Complexes and Combinatorial Ricci Curvature. *Discrete and Computational Geometry*, 29(3):323–374, 2003.

[9] Maximilian Nickel and Douwe Kiela. Poincare embeddings for learning hierarchical representations. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems 30*, pages 6338–6347. Curran Associates, Inc., 2017.

[10] Maximilian Nickel and Douwe Kiela. Learning continuous hierarchies in the lorentz model of hyperbolic geometry. In *Proceedings of the Thirty-fifth International Conference on Machine Learning (to appear)*, 2018.

[11] Y. Ollivier. A survey of Ricci curvature for metric spaces and Markov chains. *Probabilistic approach to geometry*, 57:343–381, 2010.

[12] Rik Sarkar. Low distortion delaunay embedding of trees in hyperbolic plane. In *Graph Drawing - 19th International Symposium, GD 2011, Eindhoven, The Netherlands, September 21-23, 2011, Revised Selected Papers*, pages 355–366, 2011.

[13] Emil Saucan, Areejit Samal, Melanie Weber, and Jürgen Jost. Discrete curvatures and network analysis. *MATCH*, 2017.

[14] Kevin Verbeek and Subhash Suri. Metric embedding, hyperbolic space, and social networks. *Computational Geometry*, 59:1–12, 2016.

[15] Melanie Weber, Emil Saucan, and Jürgen Jost. Characterizing complex networks with forman-ricci curvature and associated geometric flows. *Journal of Complex Networks*, 5(4):527–550, 2017.

[16] Melanie Weber, Emil Saucan, and Jürgen Jost. Coarse geometry of evolving networks. *Journal of Complex Networks*, 6(5):706–732, 2017.

[17] Richard C. Wilson, Edwin R. Hancock, Elzbieta Pekalska, and Robert P. W. Duin. Spherical and hyperbolic embeddings of data. *IEEE Trans. Pattern Anal. Mach. Intell.*, 36(11):2255–2269, 2014.