PRINCETON UNIVERSITY

Curvature and Representation Learning: Identifying Embedding Spaces for Relational Data

Melanie Weber (Princeton University) and Max Nickel (Facebook AI Research)

Download Paper as PDF:

Contact
mw25@math.princeton.edu; 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.

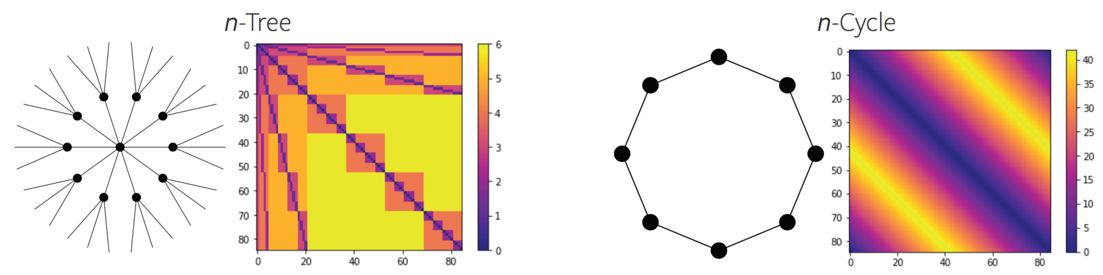
Background

Geometric Representation Learning

In embeddings, semantic similarity is captured via distances:

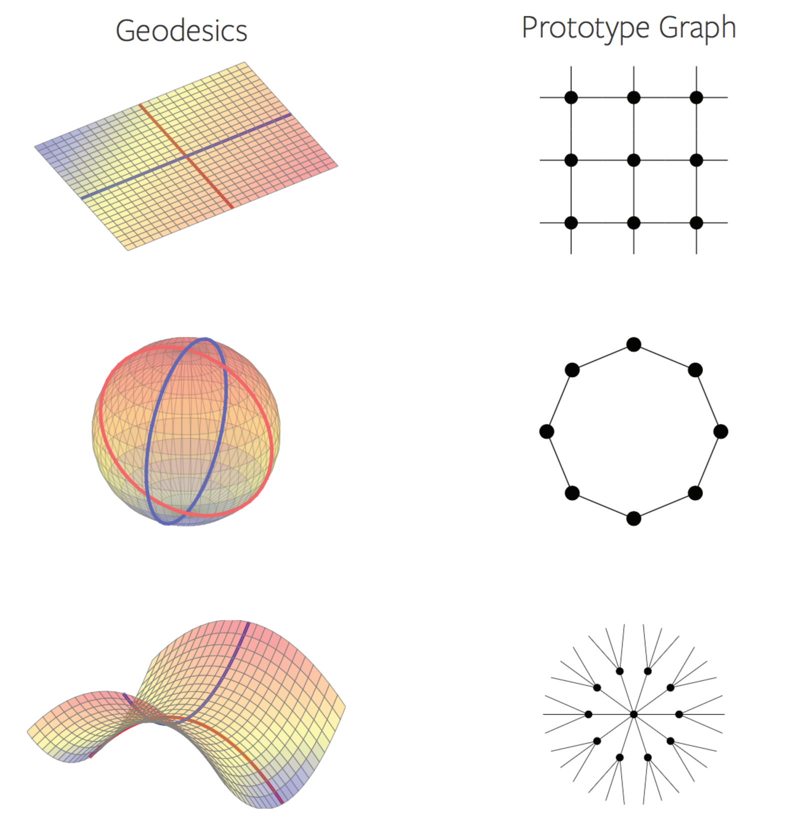
- Semantically similar entities should be close;
- Semantically dissimilar entities should be far apart;
- Geodesics in the latent space should be semantically meaningful.

Main Question: What is a good geometry to represent the global similarity structure of a graph?



Embedding Spaces

	Properties	
Euclidean	Curvature	$\kappa = 0$
	Sum of Angles	π
	Circle Length	$2\pi r$
	Disk Area	πr^2
Spherical	Curvature	$\kappa = +1$
	Sum of Angles	$> \pi$
	Circle Length	$2\pi \sin r$
	Disk Area	$2\pi(1 - \cos r)$
Hyperbolic	Curvature	$\kappa = -1$
	Sum of Angles	$< \pi$
	Circle Length	$2\pi \sinh r$
	Disk Area	$2\pi(1 - \cosh r)$



Methods

Identifying Embedding Spaces for Relational Data

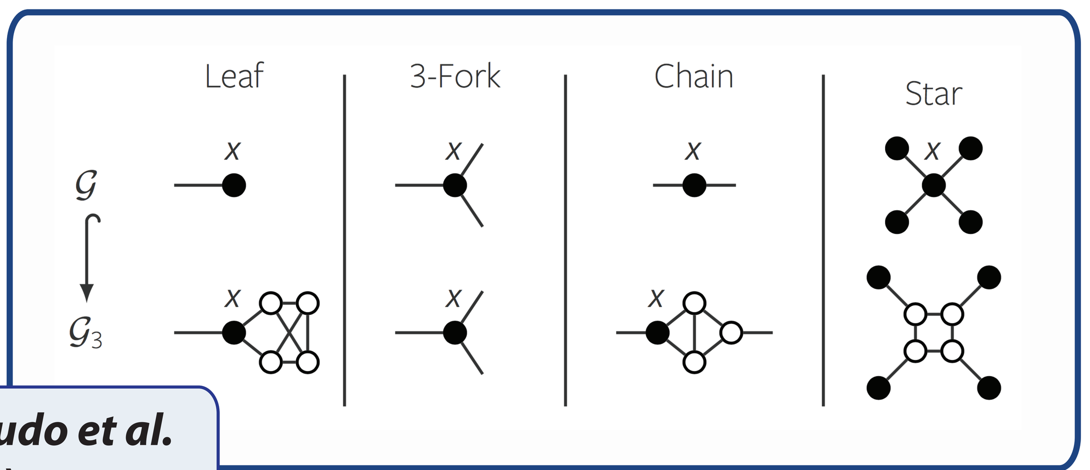
Graph motifs are characteristic for embedding spaces via local neighborhood growth rates:

- Tree-like structures (expanding neighborhoods) embed with low distortion into hyperbolic space;
- Cycles (contracting neighborhoods) embed with low distortion into spherical space.

Estimating Local Neighborhood Growth
1. Regularization
2. MotifCount

Regularization

In real graphs, motifs will be overlapping; therefore we estimate neighborhood growth rates in a regularized representation. The transformation $G \hookrightarrow G_3$ can be done with small distortion:



Bermudo et al. (2013)

MotifCount

For each regularized network, we compute the ***MotifCount*** score:

$$\gamma = \sum_{v \in V(G)} \sigma(v) |\mathcal{N}(v)|$$

$$\sigma(v) = \begin{cases} 1, & \mathcal{N}(v) \text{ contracting} \\ -1, & \mathcal{N}(v) \text{ expanding} \end{cases}$$

We call an R-hop neighborhood $\mathcal{N}(v) = \bigcup_{r=1}^{R} \mathcal{N}(v,r)$ expanding, if

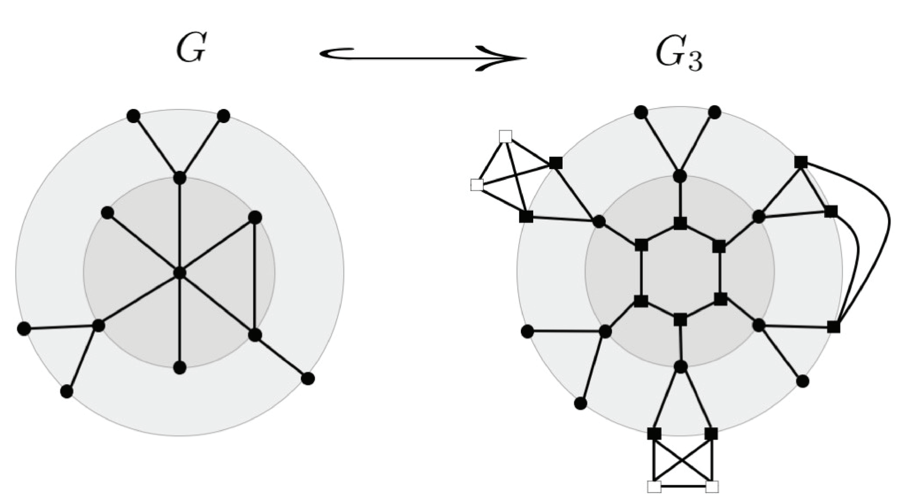
$$|\mathcal{N}(v,1)| < |\mathcal{N}(v,2)| < \ldots < \mathcal{N}(v,r)| \,.$$

Then:
- $\gamma << 0 \;\rightarrow \kappa = -1$;
- $\gamma >> 0 \;\rightarrow \kappa = 1$;
- $\gamma \approx 0 \;\rightarrow \kappa = 0$.

Experiments

Example: Regularization

Low-distortion regularization of small sample graph.



Experiments

We test the ability of ***MotifCount*** to determine a suitable embedding curvature for relational data. For this purpose, we applied our method to various model and real world graphs. The estimated values shown on the right match our theoretical expectations.