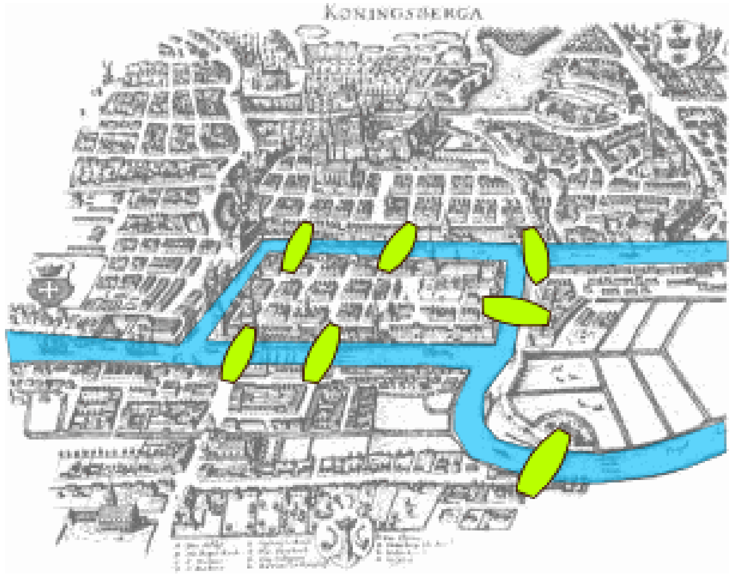
Graph Curvature

Topology vs Geometry

- **What is the difference?**
- **Topology**
  - studies properties of spaces that are invariant under any continuous deformation.
  - rubber-sheet geometry

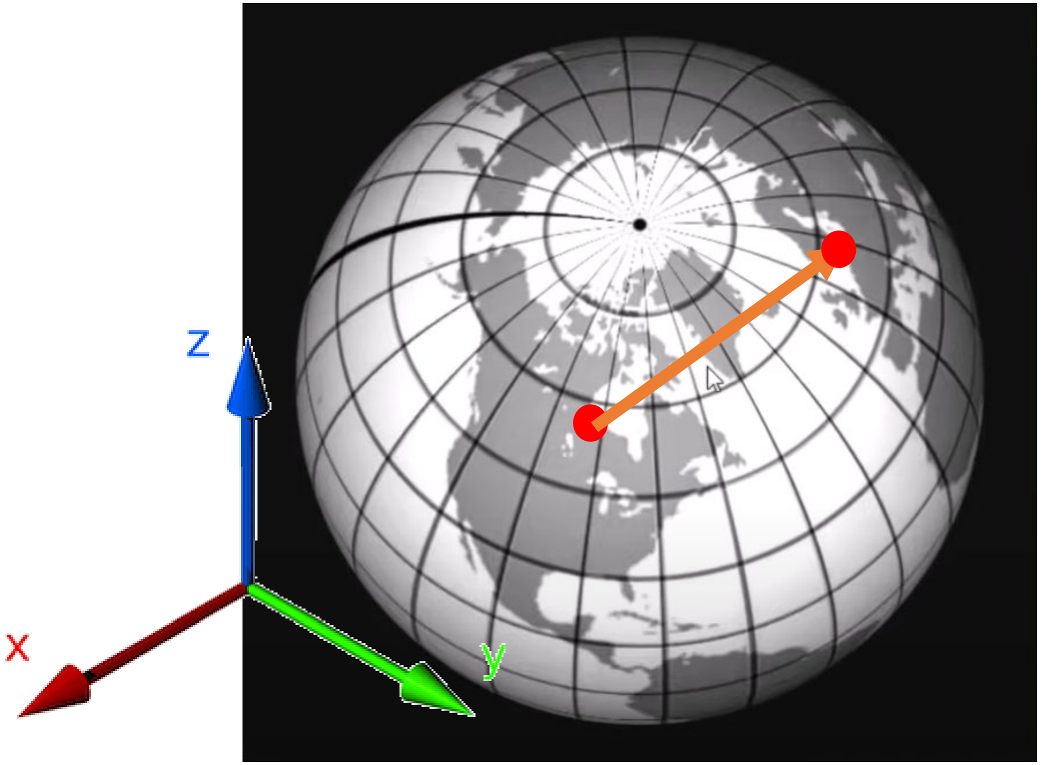


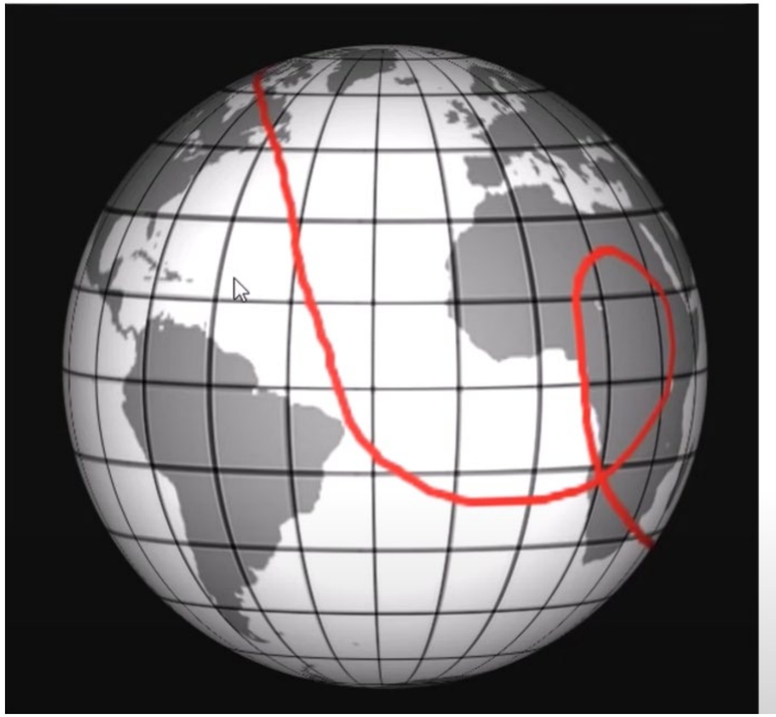
Graphs as Geometric Objects

- Geometric view: Extrinsic vs Intrinsic

Extrinsic: coordinates

Intrinsic: length and curvature

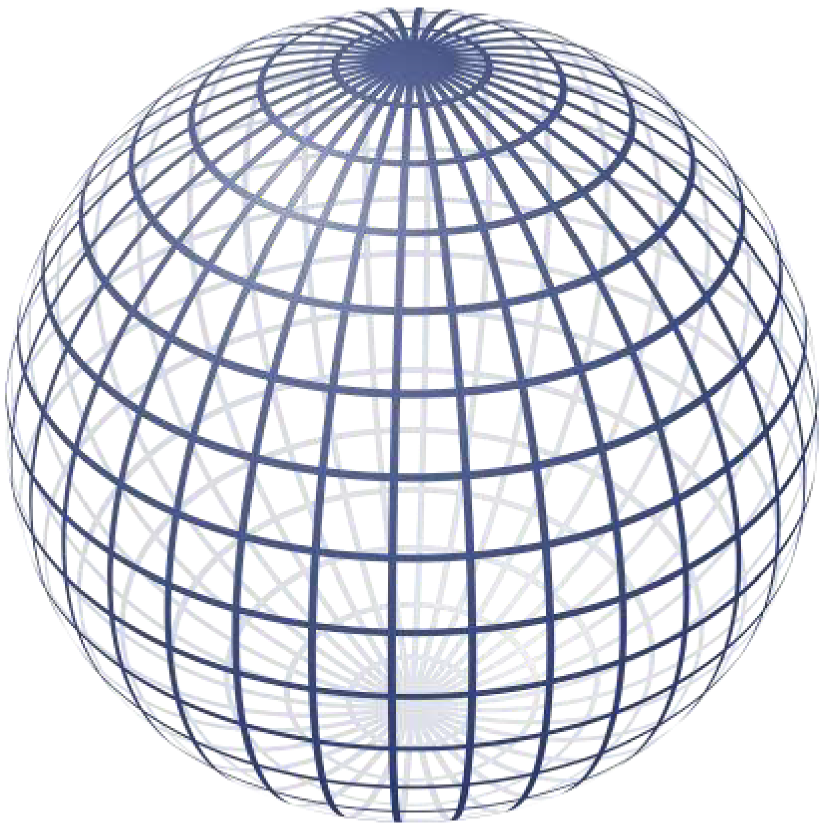


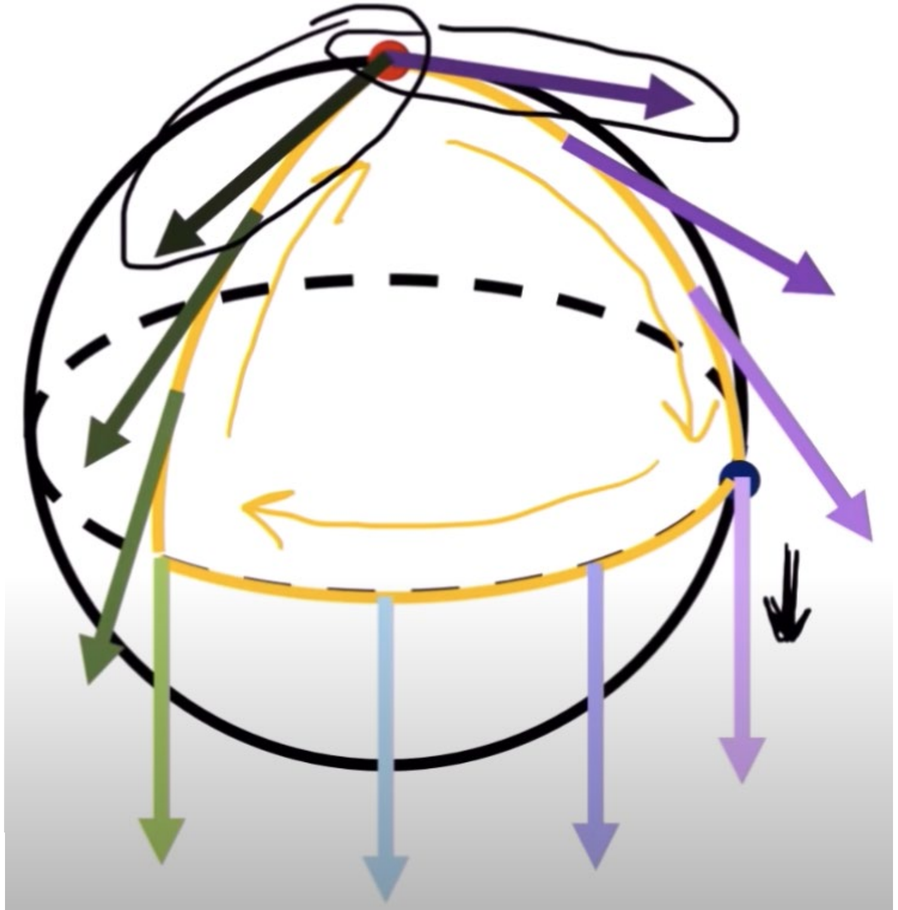


What is curvature?

Curvature in Geometry

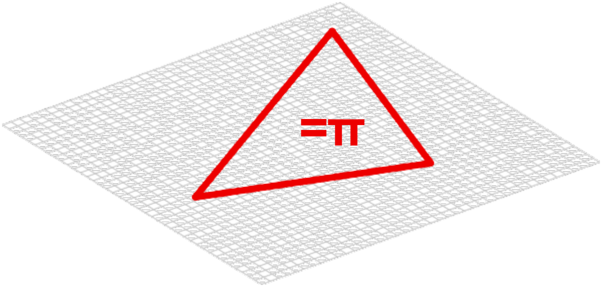
- How do we know that the earth is round?



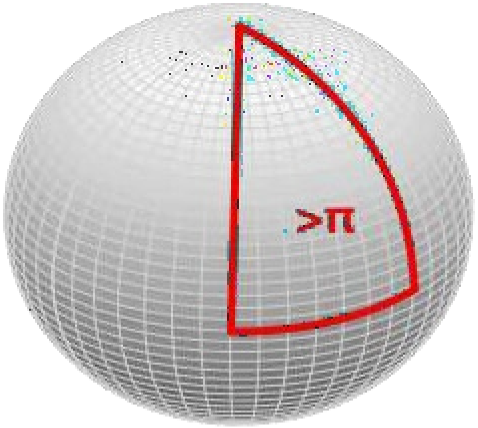


Curvature in Geometry

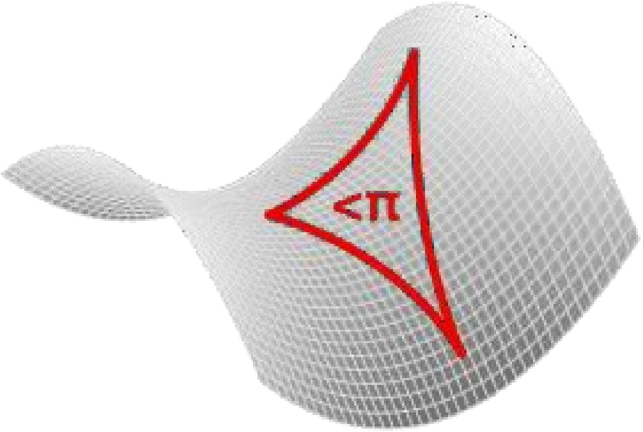
- A geometric property: Flatness of an object



2D Plane
Zero Curvature



3D Sphere
Positive Curvature

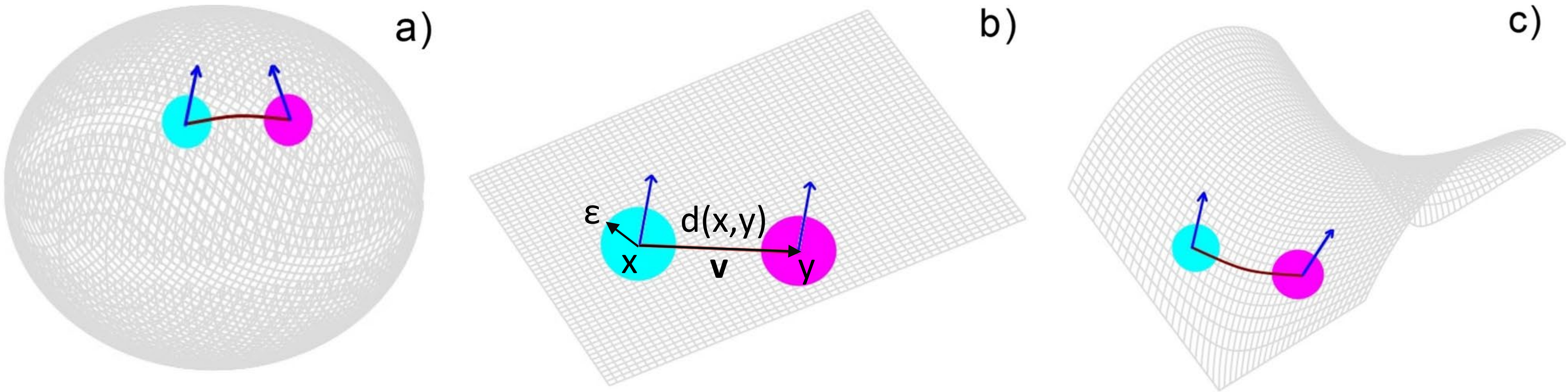


3D Saddle
Negative Curvature

Ricci Curvature

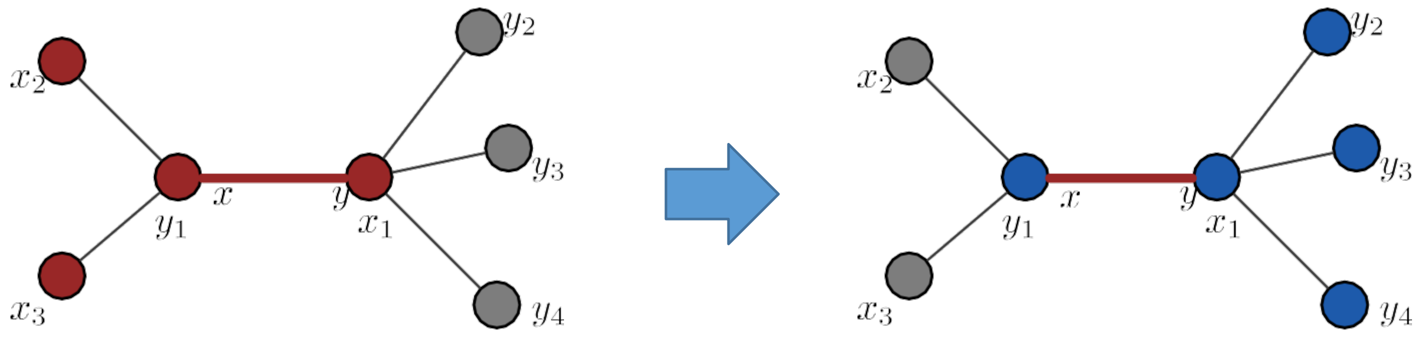
Consider $\varepsilon > 0$ and let x be an arbitrary point in a n dimensional manifold, y be the endpoint of δv where v is a tangent vector at x with $\delta = d(x, y)$ and S_x and S_y be the geodesic (blue and red) balls with radius ε. As $(\varepsilon, \delta) \to 0$:

$$\text{Ric }(x, v) \simeq \frac{2(n+2)}{\varepsilon^2}\left(1 - \frac{d(S_x, S_y)}{d(x,y)}\right)$$



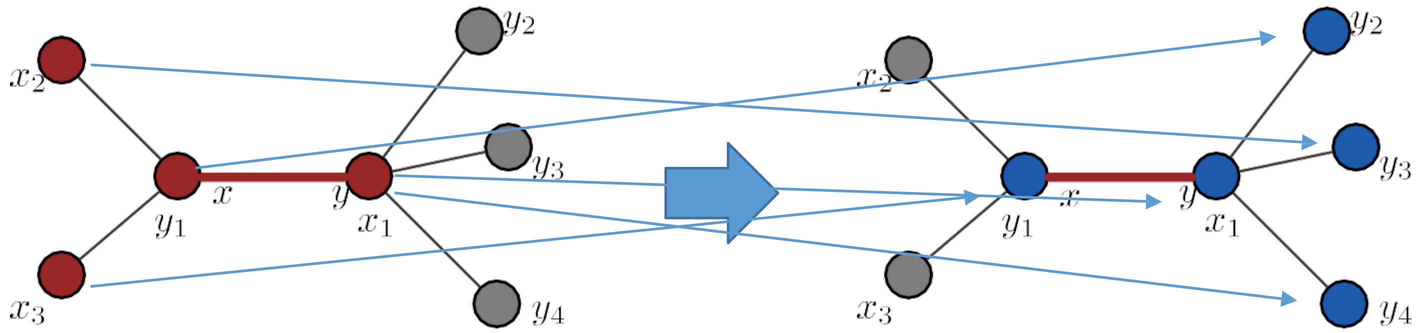
Graph Ricci Curvature

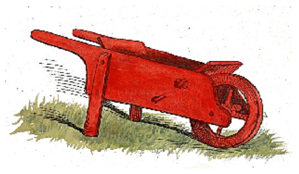
Analog: For an edge *xy*, consider the distances from *x's neighbors* to *y's neighbors* and compare it with the length of *xy*

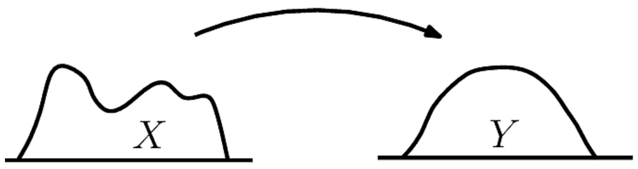


Graph Ricci Curvature

- Issue: how to match x's neighbors to y's neighbors?
  - Assign uniform distribution μ_1, μ_2 on x' and y's neighbors.
  - Use optimal transportation distance (earth-mover distance) from μ_1 to μ_2: the matching that minimize the total transport distance.







Ollivier Ricci Curvature

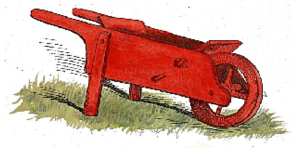
Definition (Ollivier)

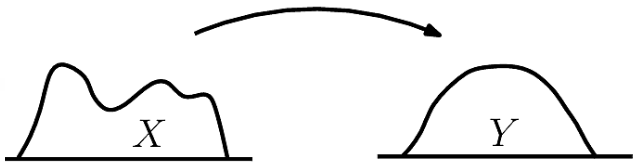
Let (X, d) be a metric space and let m_1, m_2 be two probability measures on X. For any two distinct points $x, y \in X$, the (Ollivier-) Ricci curvature along xy is defined as

$$\kappa(x,y) := 1 - \frac{W_1(m_x, m_y)}{d(x,y)},$$

$$\text{Ric}\ (x,v) \simeq \frac{2(n+2)}{\varepsilon^2}\left(1 - \frac{d(S_x,S_y)}{d(x,y)}\right)$$

where m_x (m_y) is a probability distribution defined on x (y) and its neighbors, $W_1(\mu_1, \mu_2)$ is the L_1 **optimal transportation distance** between two probability measure μ_1 and μ_2 on X:



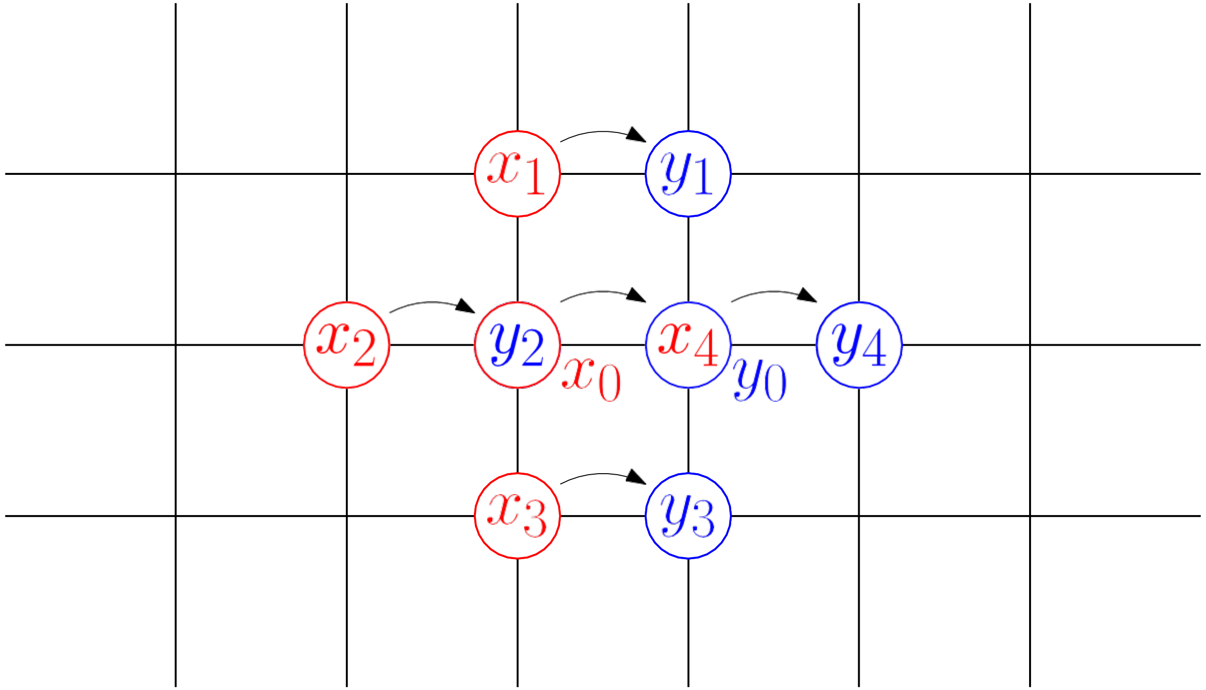


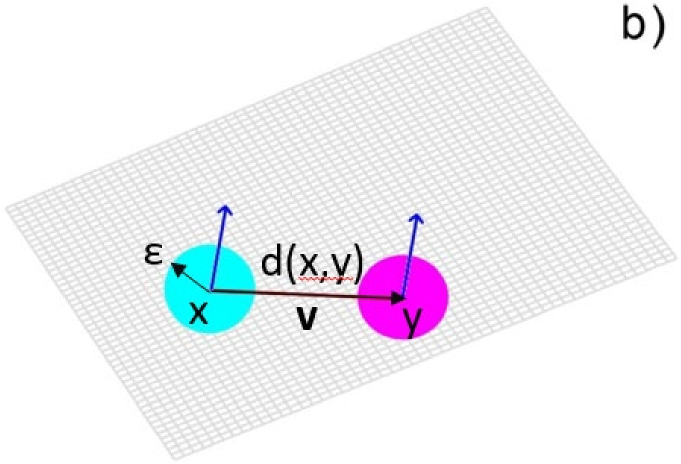
$$W_1(\mu_1,\mu_2) := \inf_{\psi \in \Pi(\mu_1,\mu_2)} \int_{(u,v)} d(u,v)\, d\psi(u,v)$$

[1] Ollivier, Y. (2007, January 31). Ricci curvature of Markov chains on metric spaces. *arXiv.org*.
[2] Lin, Y., Lu, L., & Yau, S.-T. (2011). Ricci curvature of graphs. *Tohoku Mathematical Journal*, *63*(4), 605–627.

Example: Zero Curvature

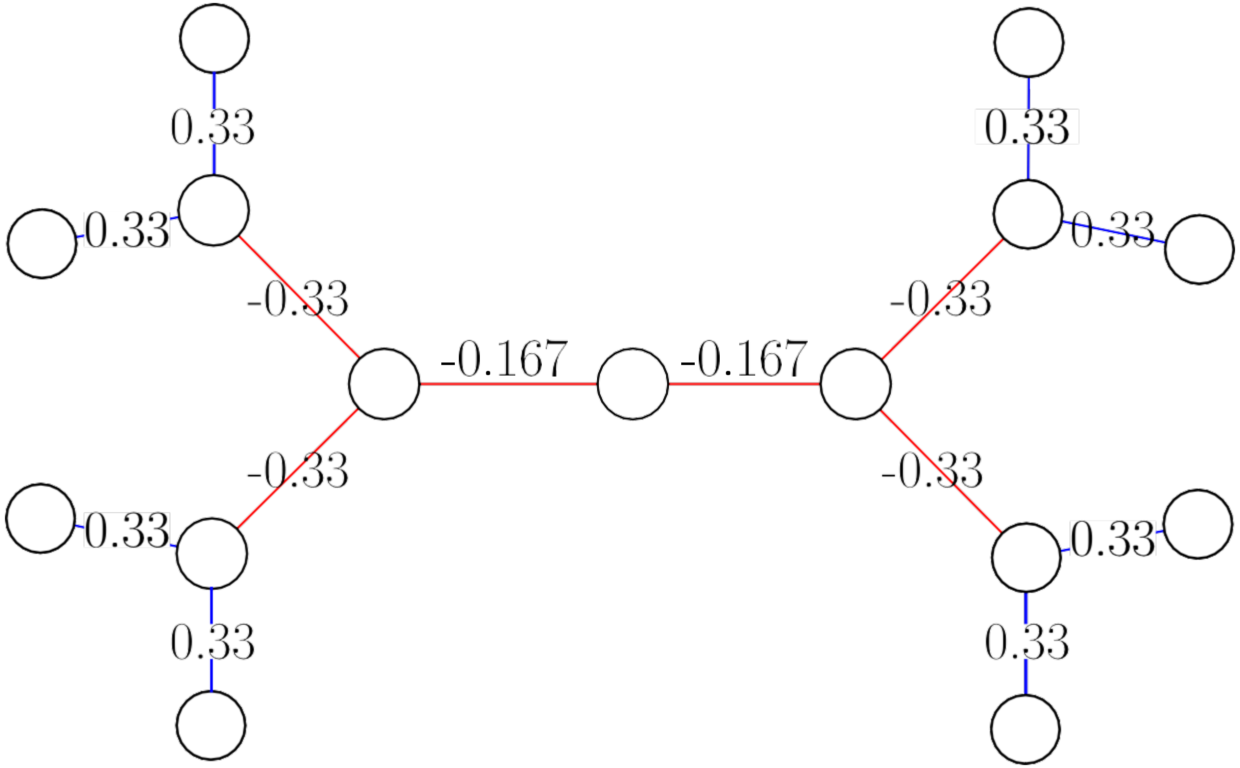
- 2D grid

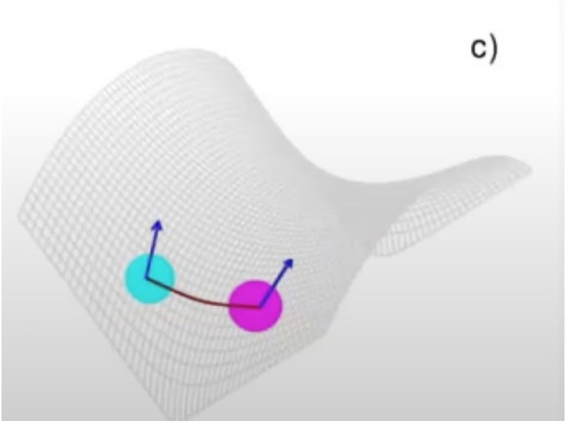




Example: Negative Curvature

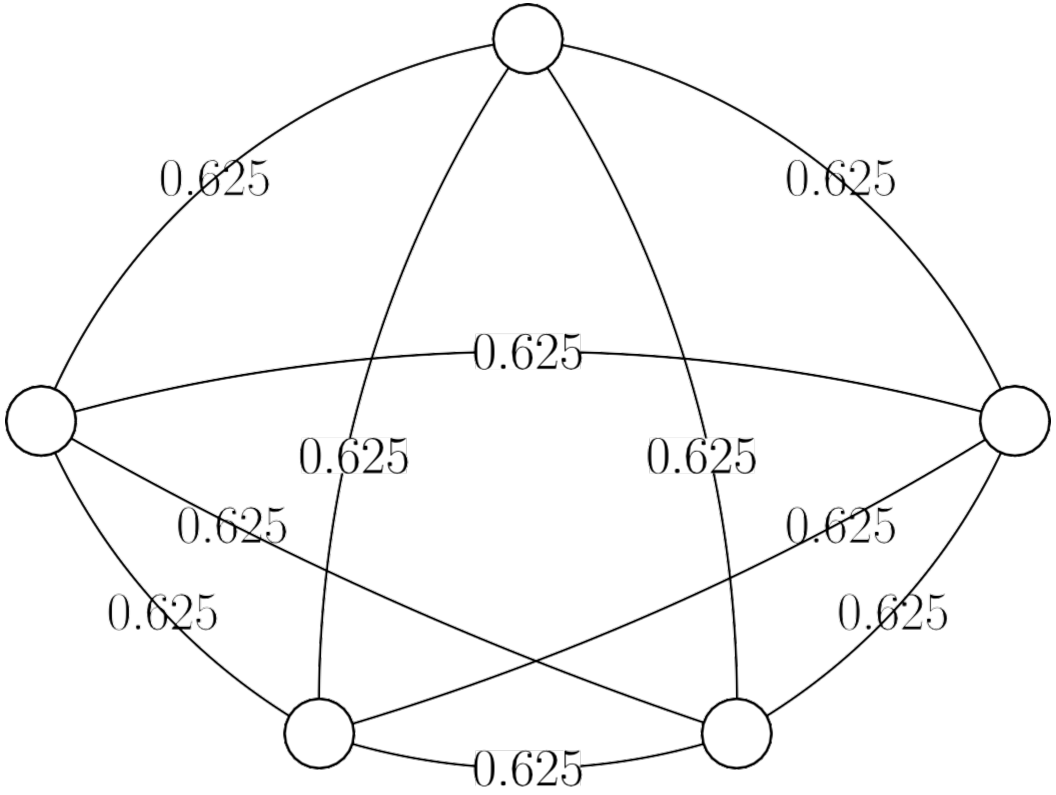
- Tree: $\kappa(x, y) = 1/d_x + 1/d_y - 1$, d_x is degree of x.

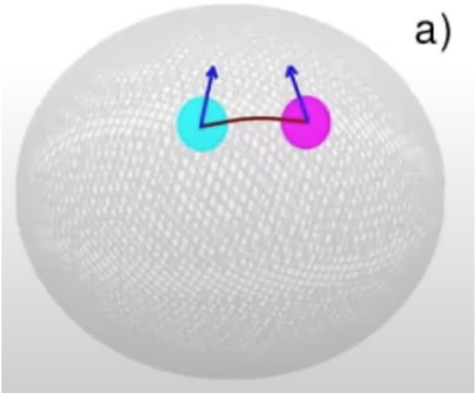




Example: Positive Curvature

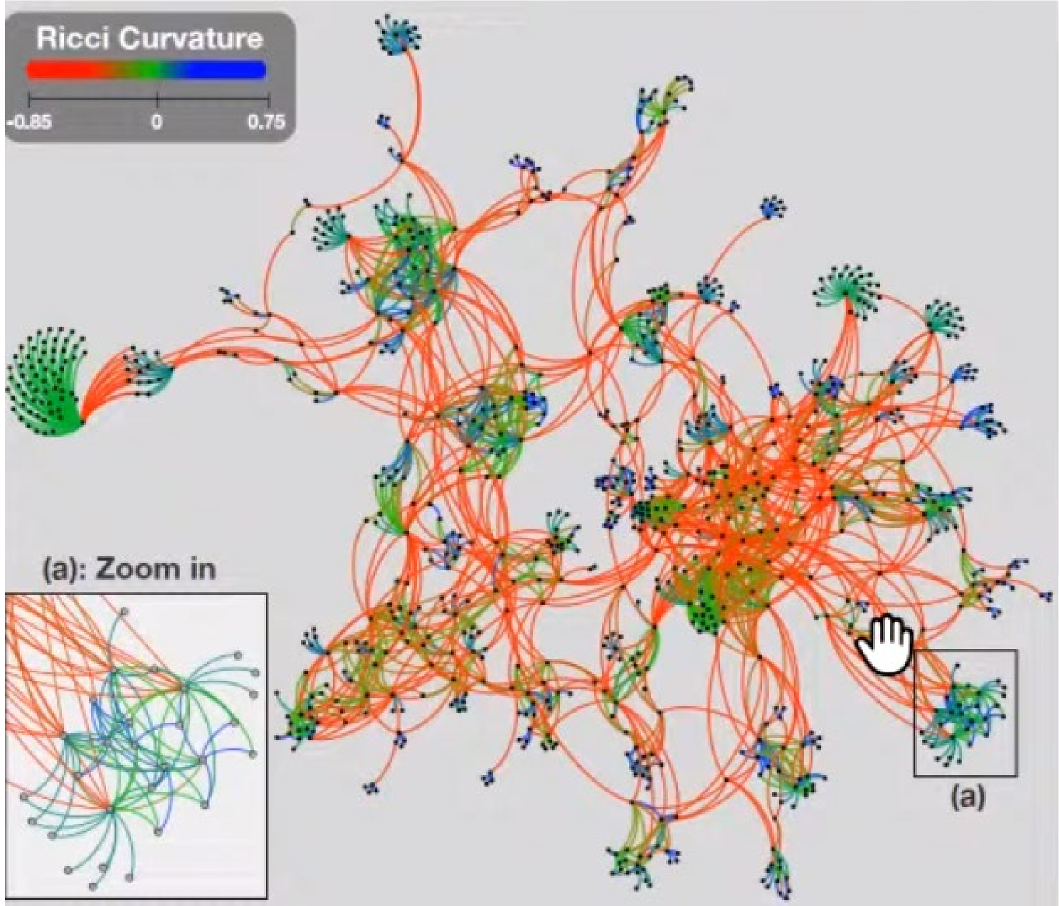
- Complete graph





Curvature Distribution

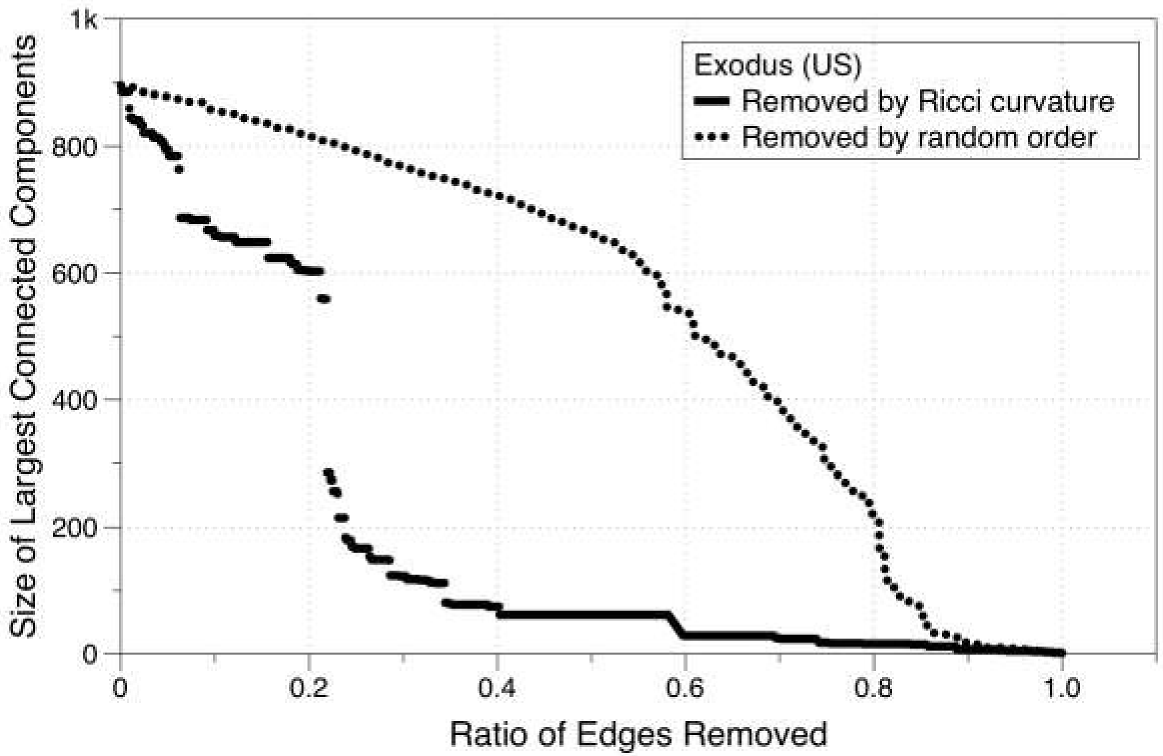
- Negatively curved edges ➔ “backbones”



Ni, Chien-Chun, et al. "Ricci curvature of the internet topology.“ INFOCOM 2015

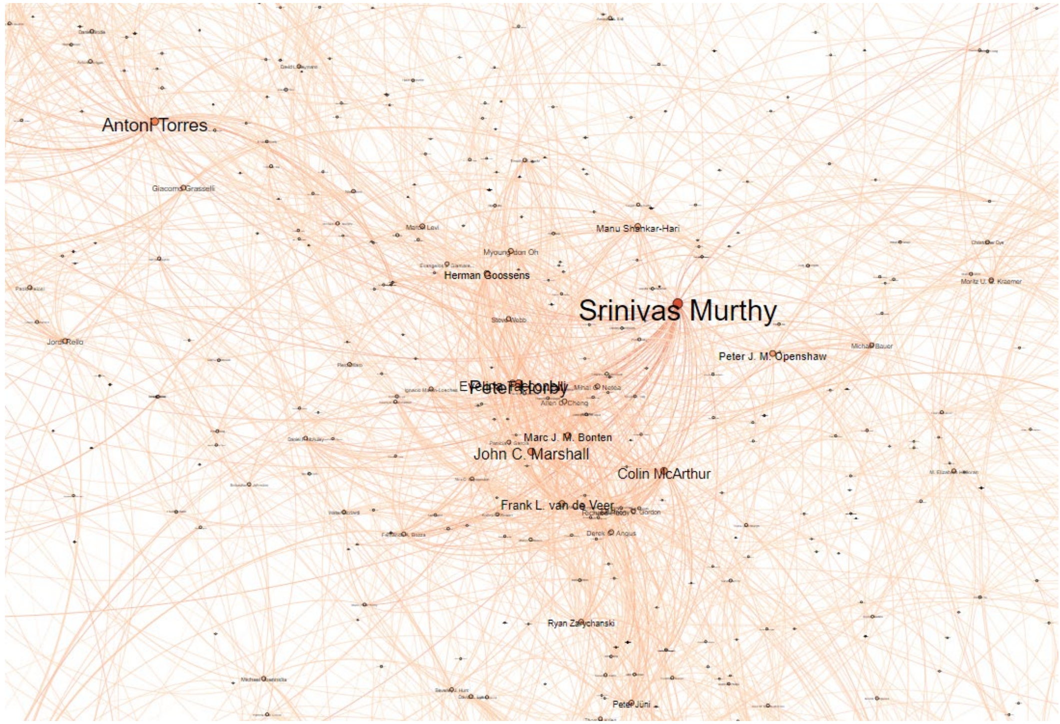
Robustness vs. Vulnerability

- Removing edges with increasing curvature



Ricci Curvature in Academic Graphs

- Finding scholars that bridge different communities

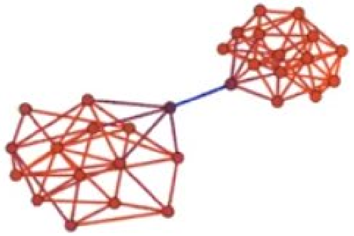


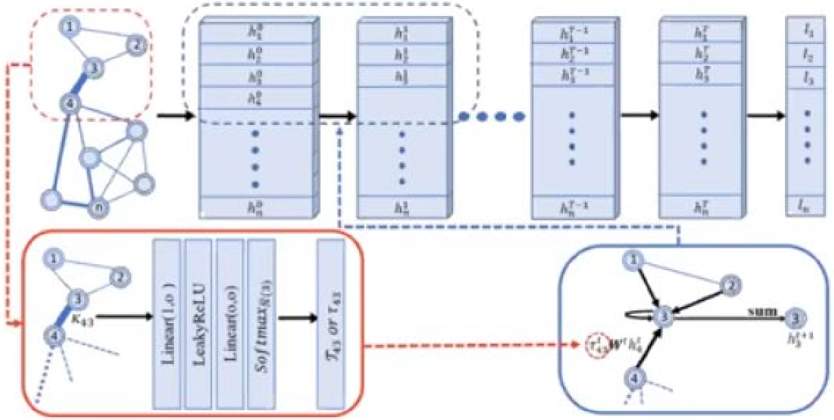
Graph Learning with Ricci Curvature

- Ricci curvature improves message passing in graph learning
  - Negative curved edges pass important messages

$$h_x^{t+1} = \sigma_t \left(\sum_{y \in \overline{\mathcal{N}}(x)} \tau_{xy}^t W^t h_y^t \right)$$

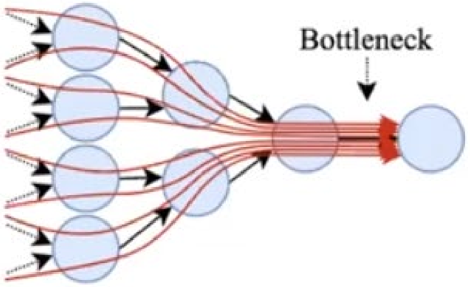
$$f^t : \kappa_{xy} \rightarrow \tau_{xy}^t$$



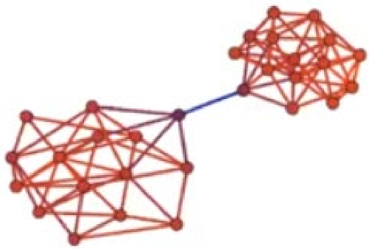


Graph Learning with Ricci Curvature

- Ricci curvature explains **oversquashing** in graph representation learning
  - negatively curved edges act as graph bottlenecks and lead to oversquashing



(b) The bottleneck of graph neural networks



MPNN

$$h_i^{(\ell+1)} = \phi_\ell\left(h_i^{(\ell)}, \sum_{j=1}^{n} \hat{A}_{ij}\psi_\ell(h_i^{(\ell)}, h_j^{(\ell)})\right)$$

Theorem 4. *Consider a MPNN as in equation 1. Let $i \sim j$ with $d_i \leq d_j$ and assume that:*

(i) $|\nabla\phi_\ell| \leq \alpha$ and $|\nabla\psi_\ell| \leq \beta$ for each $0 \leq \ell \leq L-1$, with $L \geq 2$ the depth of the MPNN.

(ii) There exists δ s.t. $0 < \delta < (\max\{d_i, d_j\})^{-\frac{1}{2}}$, $\delta < \gamma_{max}^{-1}$, and $\mathrm{Ric}(i,j) \leq -2+\delta$.

Then there exists $Q_j \subset S_2(i)$ satisfying $|Q_j| > \delta^{-1}$ and for $0 \leq \ell_0 \leq L-2$ we have

$$\frac{1}{|Q_j|}\sum_{k\in Q_j}\left|\frac{\partial h_k^{(\ell_0+2)}}{\partial h_i^{(\ell_0)}}\right| < (\alpha\beta)^2\delta^{\frac{1}{4}}. \tag{4}$$

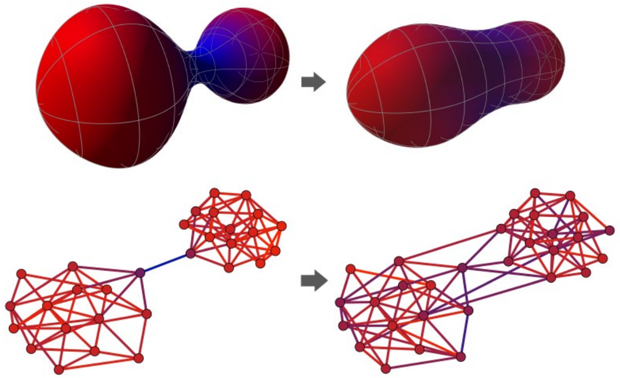
Topping, Jake, et al. "Understanding over-squashing and bottlenecks on graphs via curvature."ICLR 2022

Graph Learning with Ricci Curvature

- Ricci curvature explains **oversquashing** in graph representation learning
  - Curvature gives a control of cheeger constant (spectral gap)

Cheeger constant $$h_G := \min_{S \subset V} h_S, \quad h_S := \frac{|\partial S|}{\min\{\mathrm{vol}(S), \mathrm{vol}(V \setminus S)\}}$$

where $\partial S = \{(i,j) : i \in S,\ j \in V \setminus S\}$ and $\mathrm{vol}(S) = \sum_{i \in S} d_i$.



Cheeger inequality $$2h_G \geq \lambda_1 \geq \frac{h_G^2}{2}$$

Curvature and Cheeger constant

$$\textit{If } \mathrm{Ric}(i,j) \geq k > 0 \textit{ for all } i \sim j, \textit{ then } \lambda_1/2 \geq h_G \geq \frac{k}{2}.$$

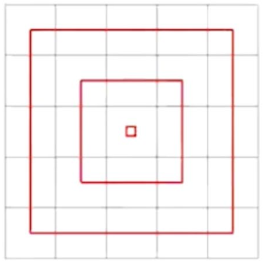
Topping, Jake, et al. "Understanding over-squashing and bottlenecks on graphs via curvature."ICLR 2022

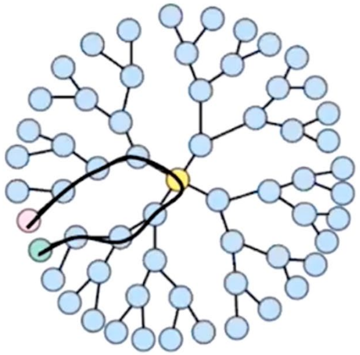
Graph Embedding in Curved Space

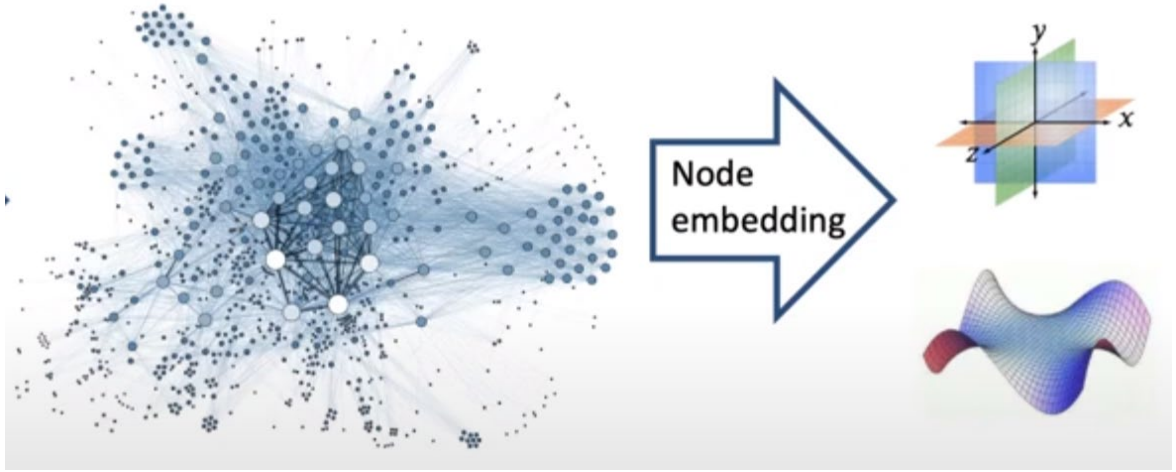
Graph Geometric Representation Learning

- Learn representations that reflect the intrinsic geometry of graphs
  - Preserve graph similarity relations
  - Low dimension

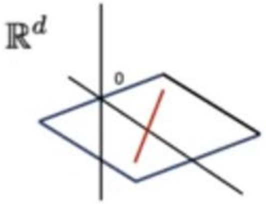
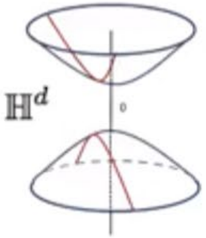
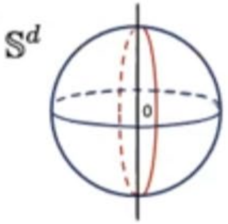
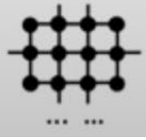
Geometry of space aligns with geometry of data





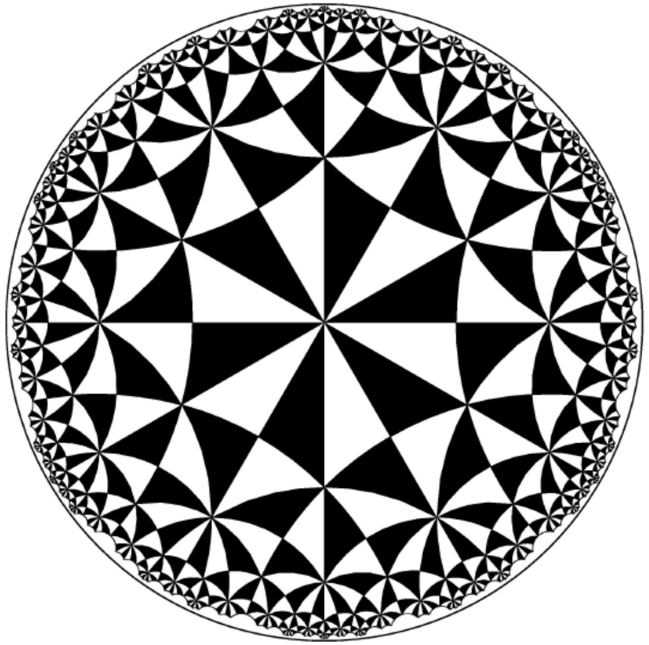


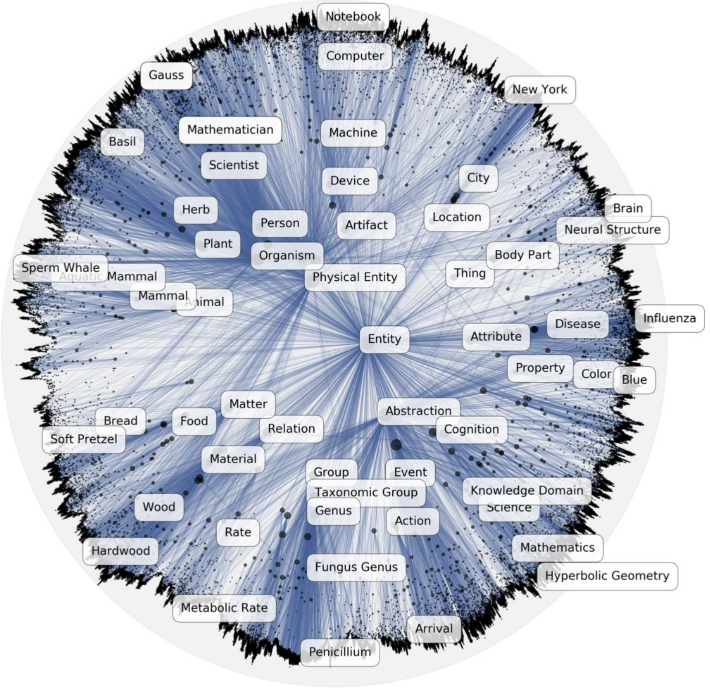
Geometric Model Spaces

	Euclidean	Hyperbolic	Spherical
			
Curvature	$\kappa = 0$	$\kappa < 0$	$\kappa > 0$
Inner product	$\langle x, y \rangle = \sum_{i=1}^{n} x_i y_i$	$\langle x, y \rangle_H = -x_0 y_0 + \sum_{i=1}^{n} x_i y_i$	$\langle x, y \rangle = \sum_{i=1}^{n} x_i y_i$
Distance	$d(x, y) = \sqrt{\langle x - y, x - y \rangle}$	$d(x, y) = \operatorname{acosh}\left(-\langle x, y \rangle_H\right)$	$d(x, y) = \arccos\left(\langle x, y \rangle\right)$
Canonical Graph			

Hyperbolic Model

- Exponential neighborhood growth rate
  - Continous space: δ - neighborhood $\mathcal{B}_\delta(x) = \{y \in \mathcal{X} : d_{\mathcal{X}}(x, y) \leq \delta\}$
    $\mathcal{X} = \mathbb{R}^d$: polynomial growth (→ lattice)
    $\mathcal{X} = \mathbb{H}^d$: exponential growth (→ tree)





Hyperbolic Graph Learning

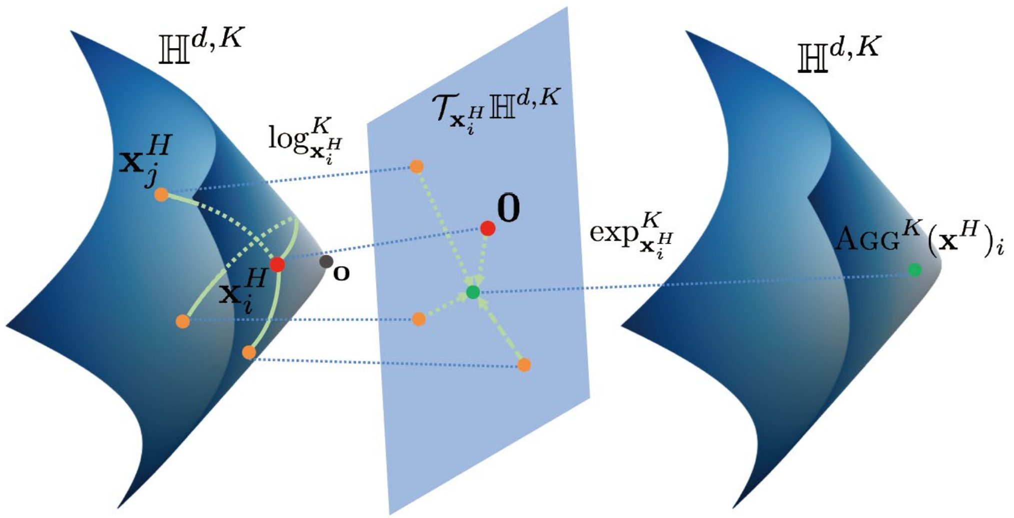
- Embed graph on the hyperbolic space
  - Adapt graph learning operations in Euclidean space to hyperbolic space

$$\mathbf{h}_i^{\ell,E} = W^\ell \mathbf{x}_i^{\ell-1,E} + \mathbf{b}^\ell$$
$$\mathbf{x}_i^{\ell,E} = \sigma(\mathbf{h}_i^{\ell,E} + \sum_{j\in\mathcal{N}(i)} w_{ij}\mathbf{h}_j^{\ell,E})$$

Euclidean space

$$\mathbf{h}_i^{\ell,H} = (W^\ell \otimes^{K_{\ell-1}} \mathbf{x}_i^{\ell-1,H}) \oplus^{K_{\ell-1}} \mathbf{b}^\ell$$
$$\mathbf{y}_i^{\ell,H} = \text{AGG}^{K_{\ell-1}}(\mathbf{h}^{\ell,H})_i$$
$$\mathbf{x}_i^{\ell,H} = \sigma^{\otimes^{K_{\ell-1},K_\ell}}(\mathbf{y}_i^{\ell,H})$$

hyperbolic space



Summary

- From an intrinsic view, graph is curved
  - Negative curved edges are key structures
  - Curvature can help to improve graph representation learning
- From an extrinsic view, graph should be embedded into best-fit space
  - Negative curved spaces can embed more
  - Non-Euclidean geometry can be further exploited

Summary of social networks

- Social networks
  - Small world
  - Scale free
  - Community
    - Girvan-Newman
    - Louvain
    - Spectral clustering
  - Curvature learning