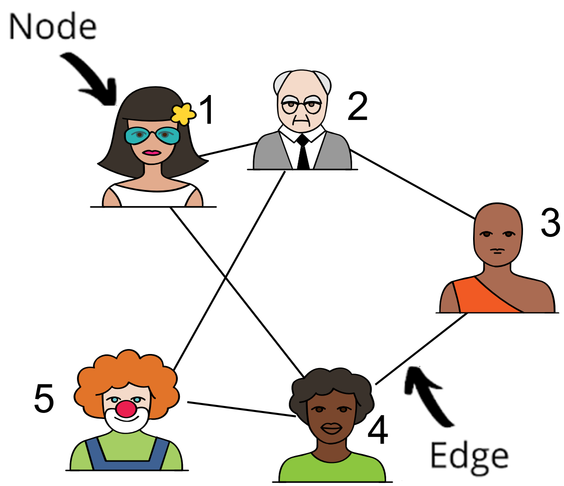
Subsampling in Large Networks

Ping Ma
Department of Statistics
University of Georgia
Malab.uga.edu

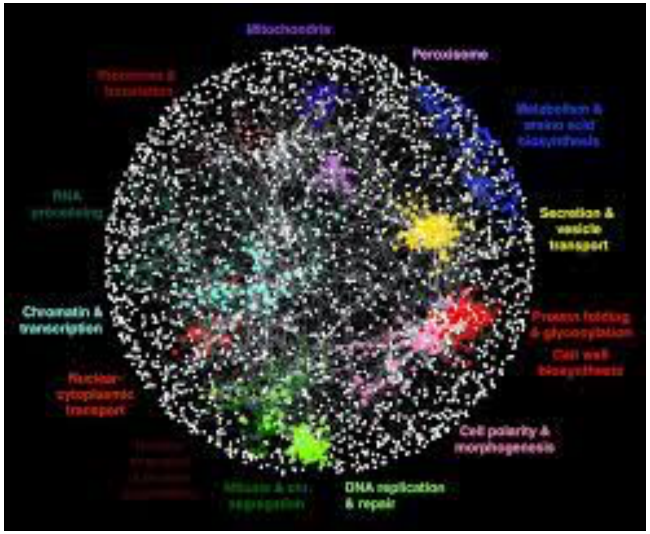
BU-Keio-Tsinghua Workshop 2023

Network

Social network



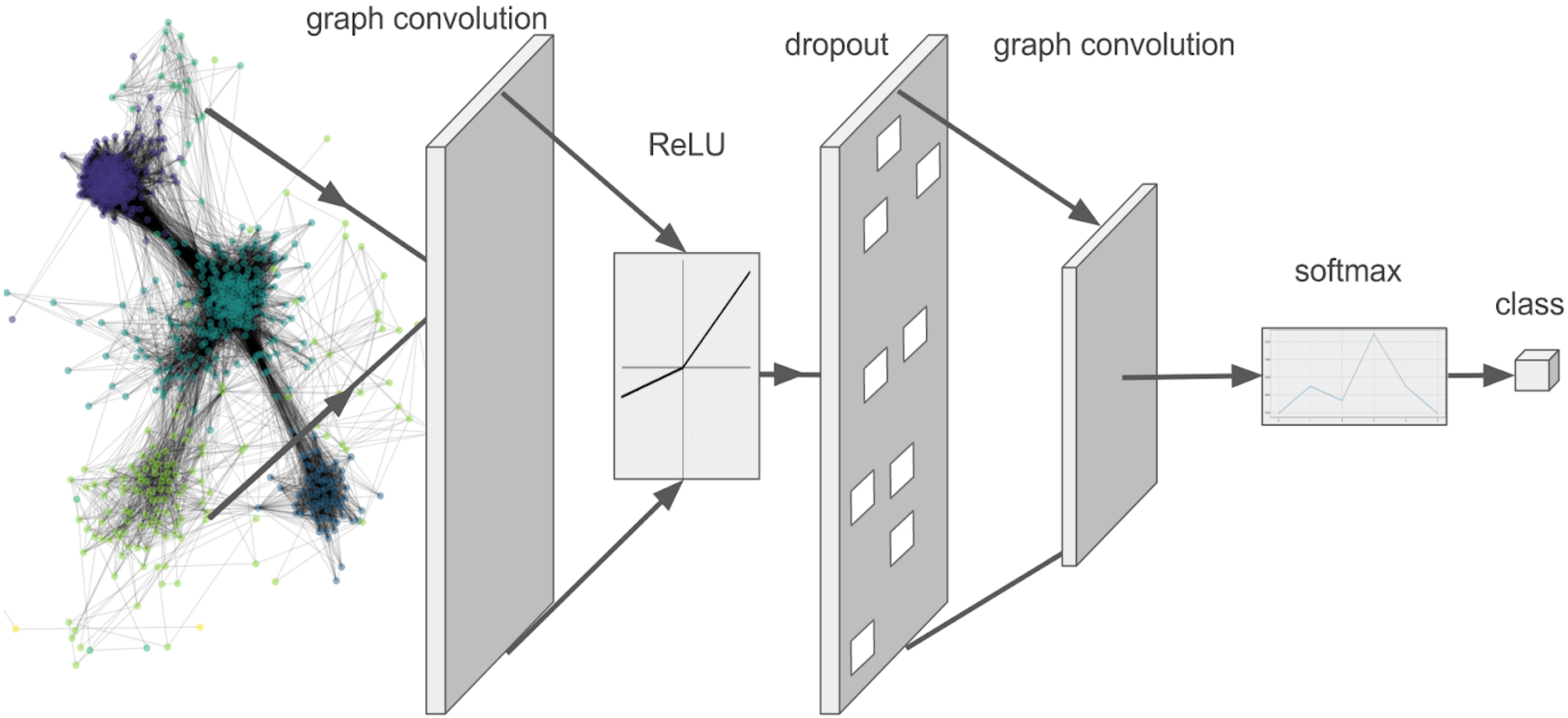
Protein-protein network



Large Networks

- Hard to visualize
- Hard to analyze
- Hard for downstream computation

Graph Neural Network



Subnetwork

- A representation (or a sketch) of the large network
- Subsampling: methods for taking subnetwork from the large network

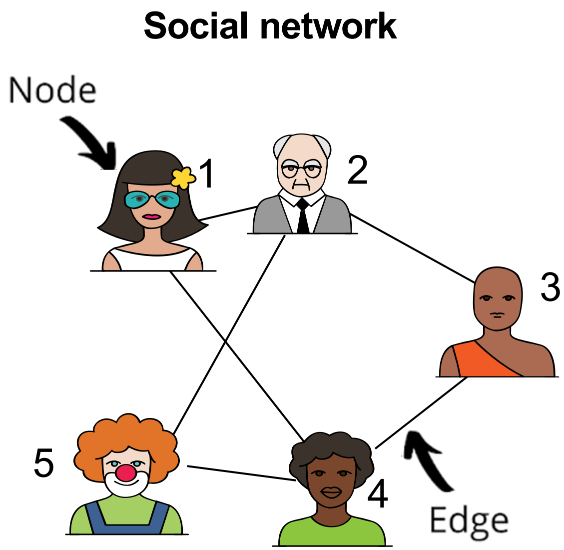
Three Settings

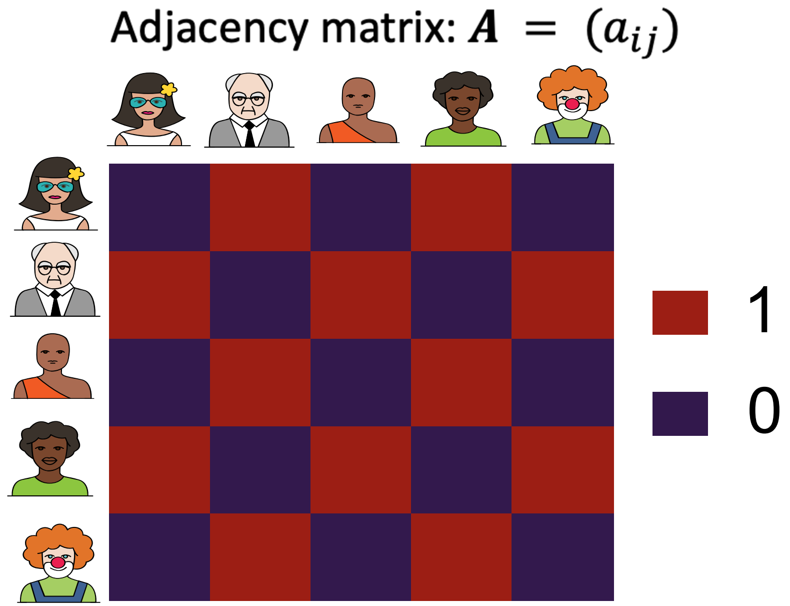
- The original large network is accessible
- The original large network is not accessible
- Something in between

Desirable Properties of Subsampling

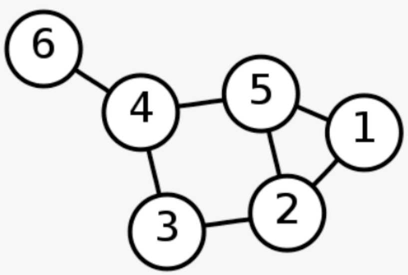
- Local to global: Importance indices of nodes and/or edges are local features with global (whole network) information
- Local computation: The subsampling methods do not need to compute the importance indices of all nodes and/or edges.

Graph

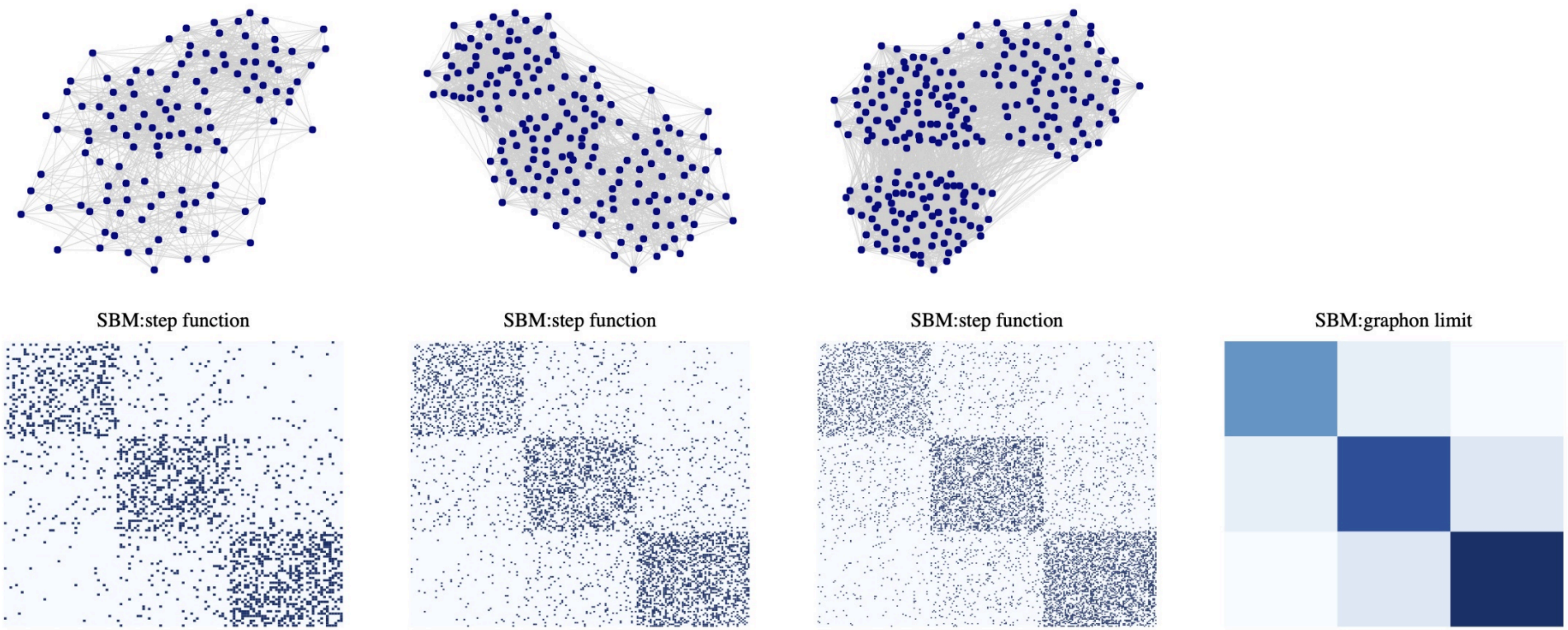

Social network
Node
1
2
3
4
5
Edge


Adjacency matrix: $\boldsymbol{A} = (a_{ij})$
1
0

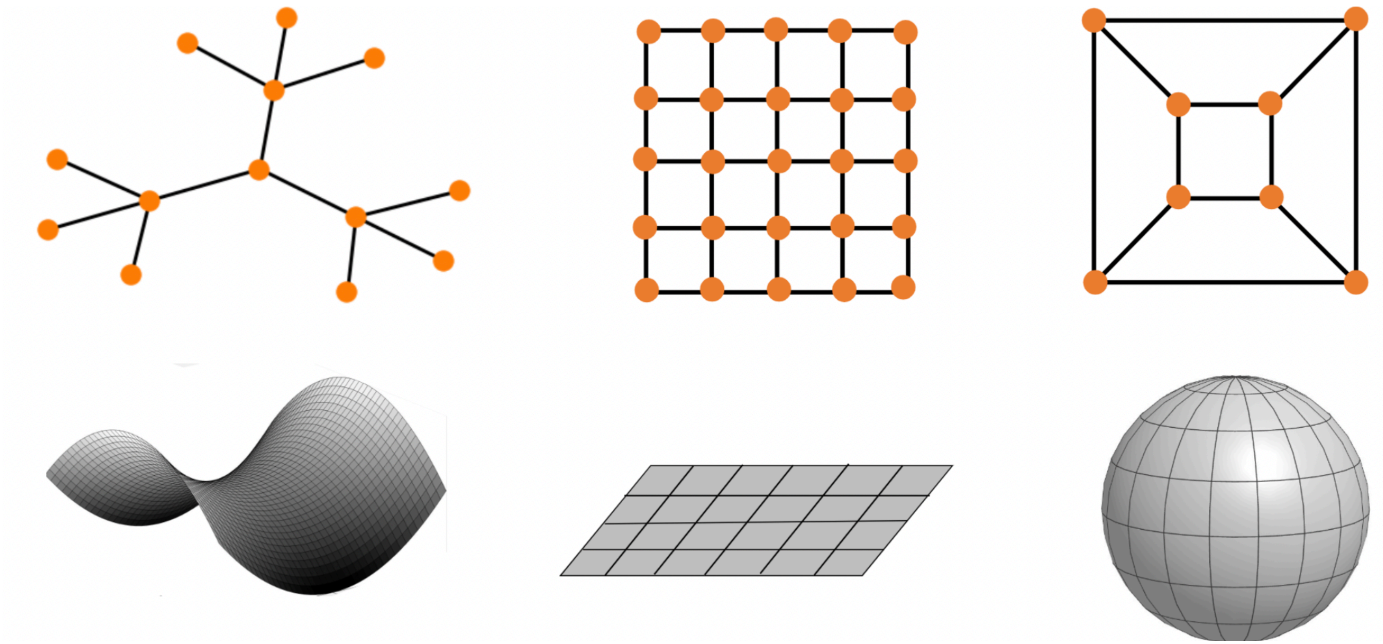
Graph and Matrix

Labelled graph	Degree matrix	Adjacency matrix	Laplacian matrix
	$\begin{pmatrix} 2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix}$	$\begin{pmatrix} 2 & -1 & 0 & 0 & -1 & 0 \\ -1 & 3 & -1 & 0 & -1 & 0 \\ 0 & -1 & 2 & -1 & 0 & 0 \\ 0 & 0 & -1 & 3 & -1 & -1 \\ -1 & -1 & 0 & -1 & 3 & 0 \\ 0 & 0 & 0 & -1 & 0 & 1 \end{pmatrix}$

Graphon and Graphex



Manifolds

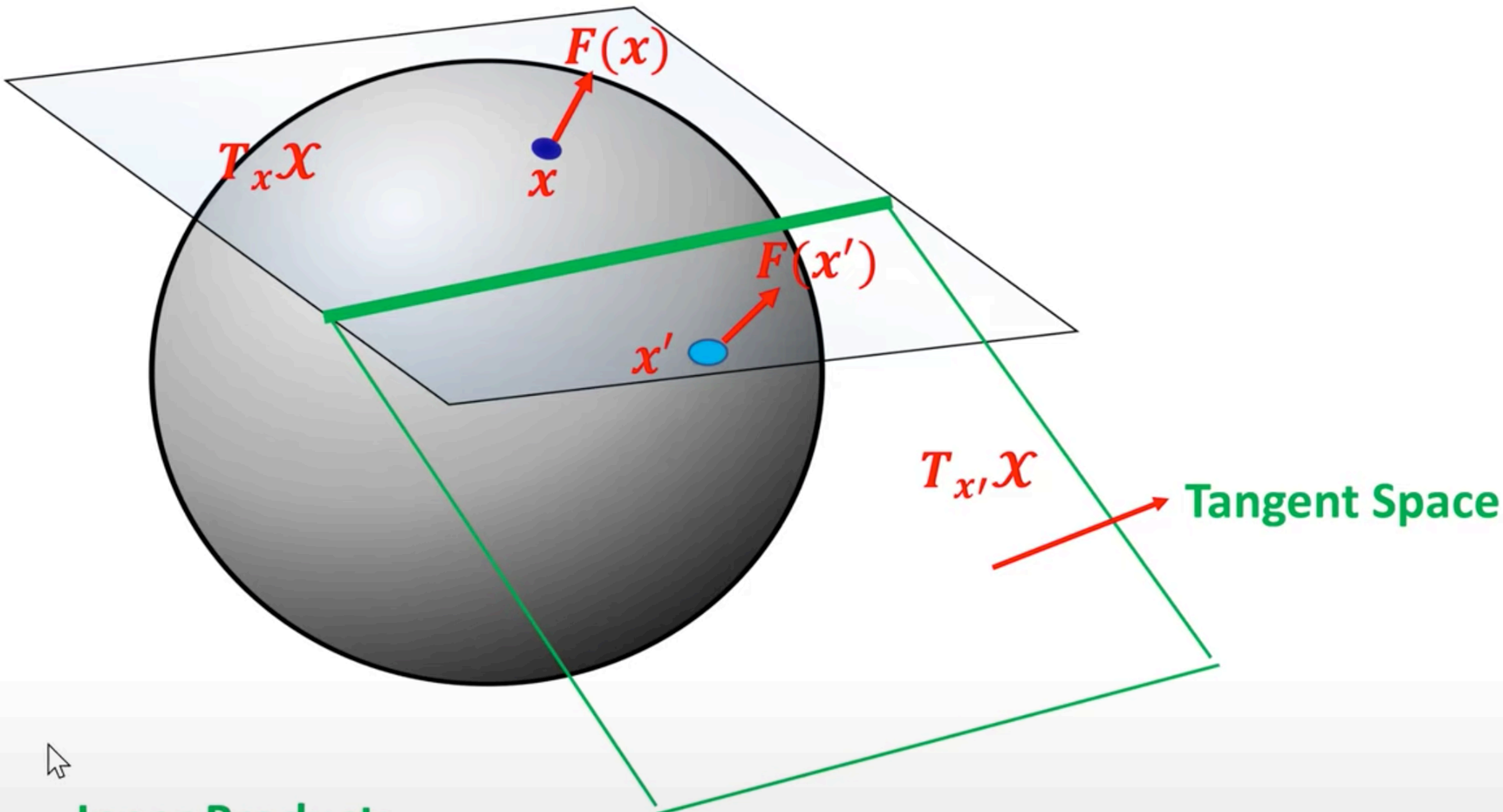


Manifolds

No predetermined coordinates

- The flexibility to choose coordinates arbitrarily
- Ensure that any objects we define globally on a manifold do not depend on a particular choice of coordinates.

Riemannian Manifolds



Inner Product:

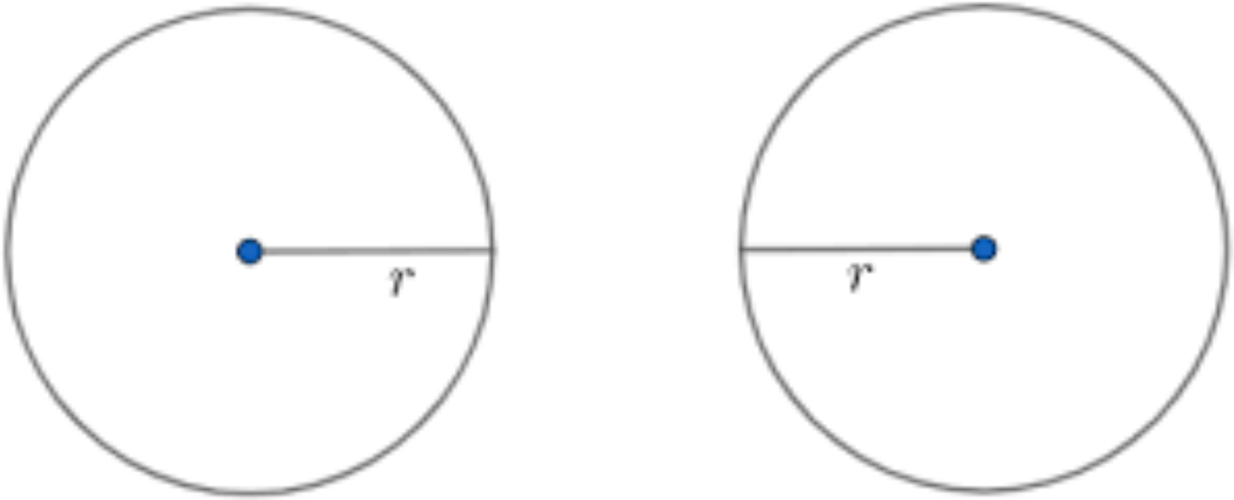
$<.,.>_{T_x\mathcal{X}}: T_x\mathcal{X} \times T_x\mathcal{X} \rightarrow R$ is called **Riemannian metric**

A manifold equipped with a metric is called a Riemannian manifold.

Xiong 2023

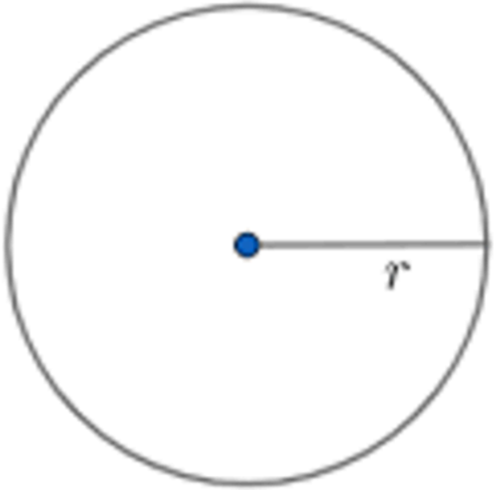
Classification Theorem of Circles

Congruence: same radius r



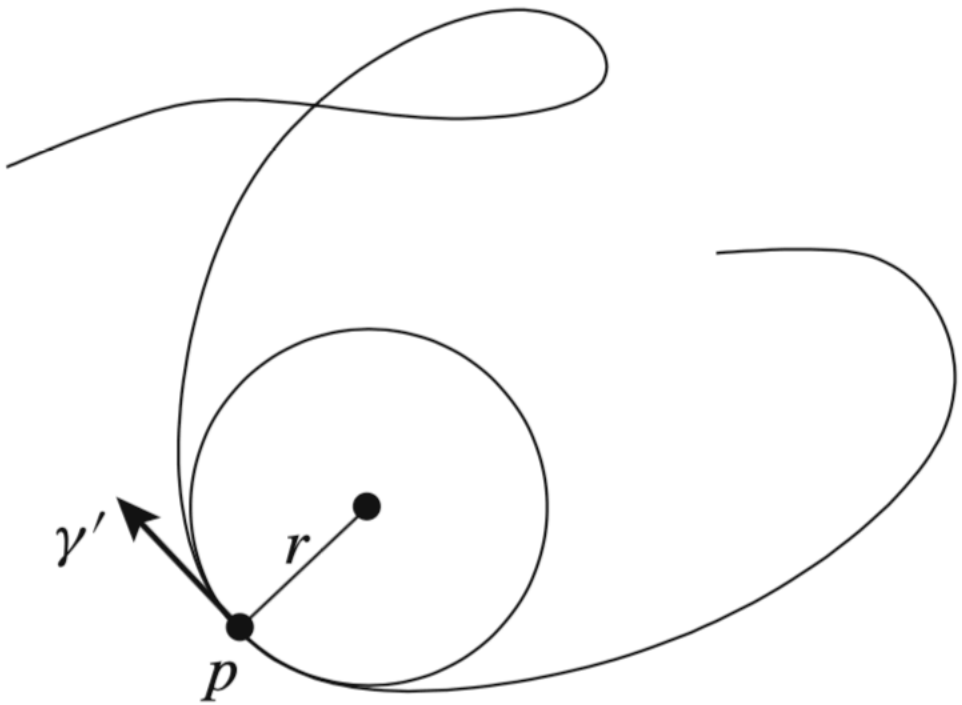
Local-to-Global Theorem of Circles

Circumference: $2\pi r$



Curvature

$$\kappa(t) = |\gamma''(t)|$$

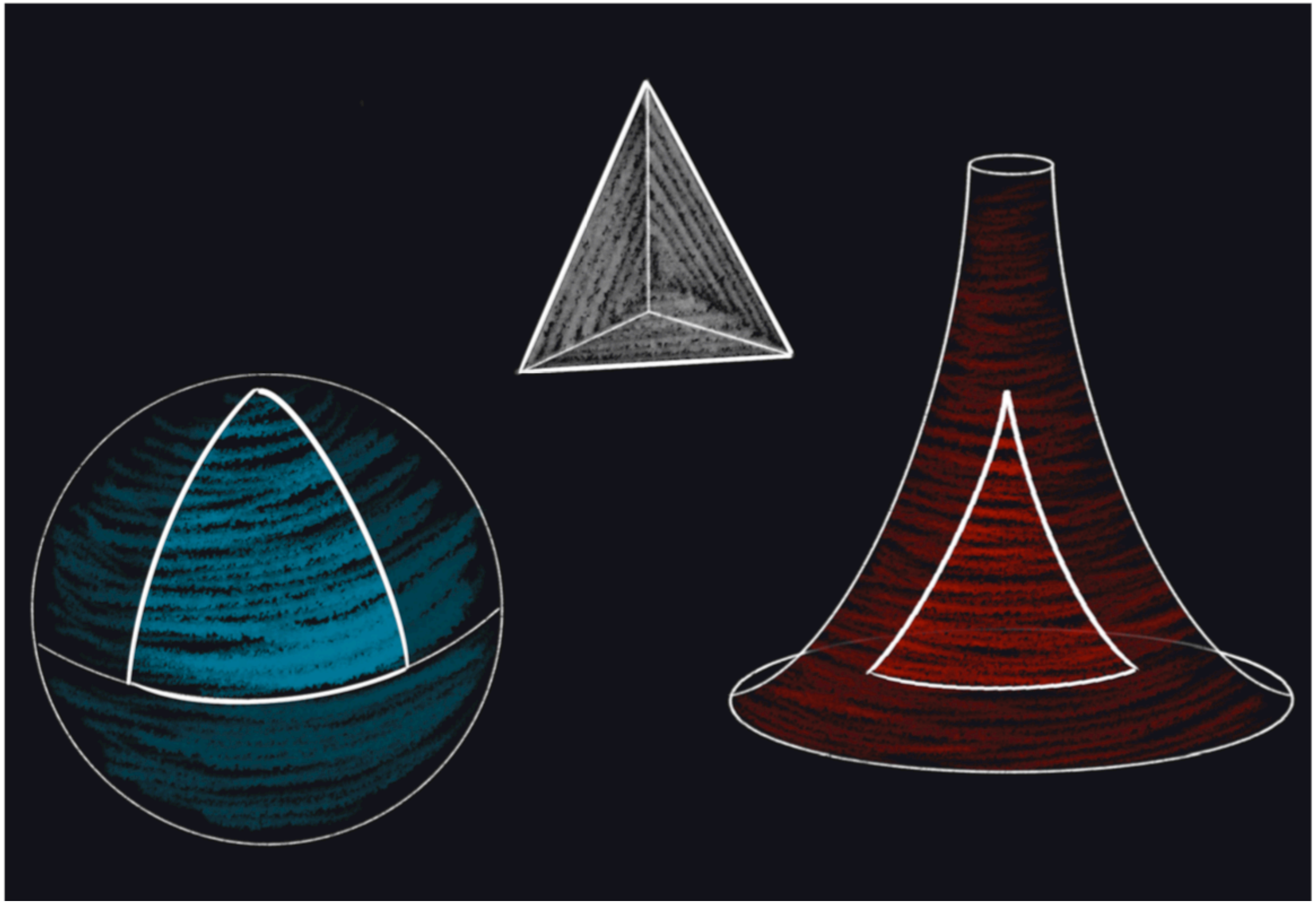


Lee (2018)

Curvature Theorems

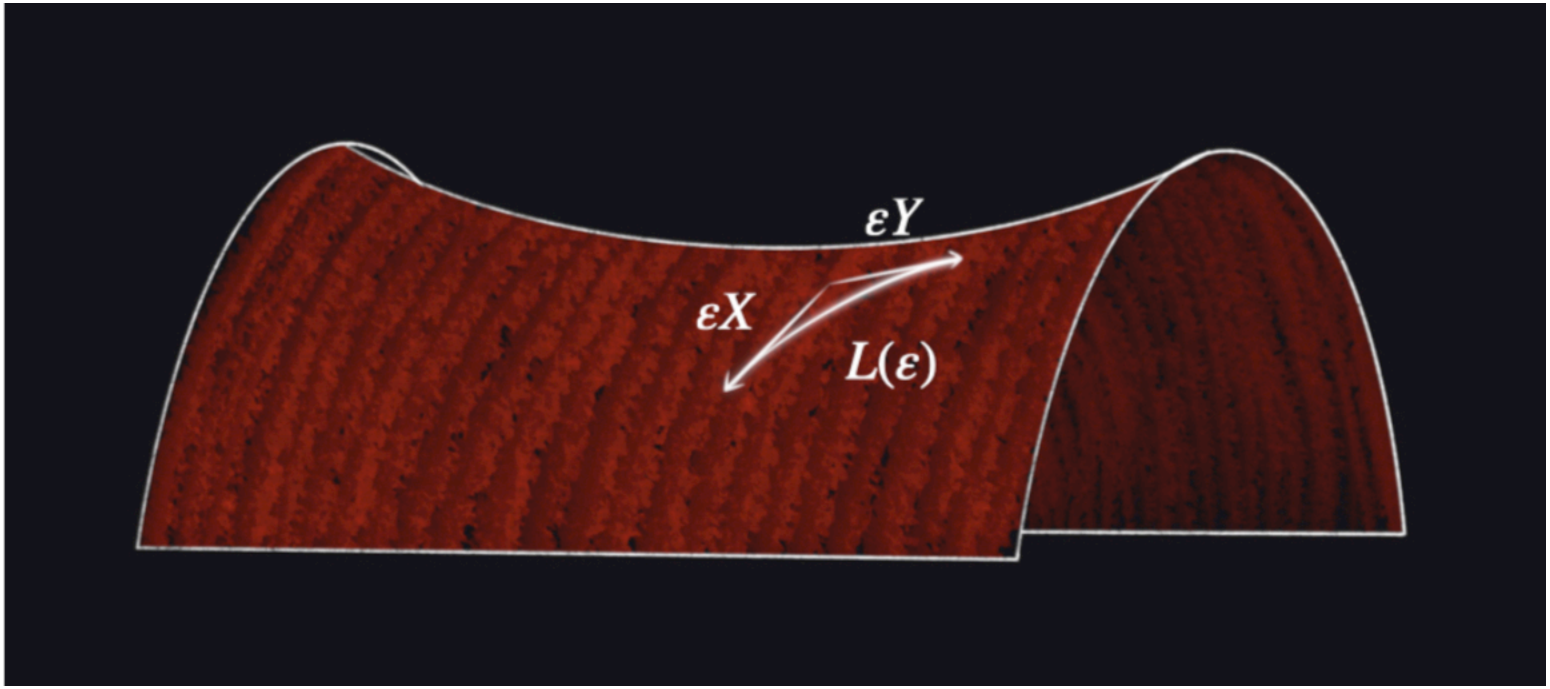
- Classification: Two curves are congruent iff their curvatures are the same.
- Local-to-global: For a simple closed curve, the integration of its curvature is 2π.

Curvature in High Dimension



Khan (2022)

Sectional Curvature



$$L(\varepsilon) = \varepsilon\|X - Y\| \left(1 - \frac{1}{12}K(X,Y)(1 + \langle X, Y\rangle)\varepsilon^2\right) + O\left(\varepsilon^4\right)$$

K(X, Y) is defined to be the sectional curvature of the tangent plane spanned by X and Y

Ricci Curvature

$$\mathrm{Ric}(X,X) = \frac{1}{2}\frac{(n-1)}{\omega\left(\mathbb{S}^{n-2}\right)}\oint_{\|Y\|=1 \text{ and } X\perp Y} K(X,Y)\,\mathrm{d}\mathbb{S}^{n-2}(Y)$$

$\omega\left(\mathbb{S}^{n-2}\right)$ is the surface area of the $(n-2)$-dimensional sphere.

The Ricci curvature Ric(X,X) is (n − 1) times the average of all of the sectional curvatures of tangent planes containing X.

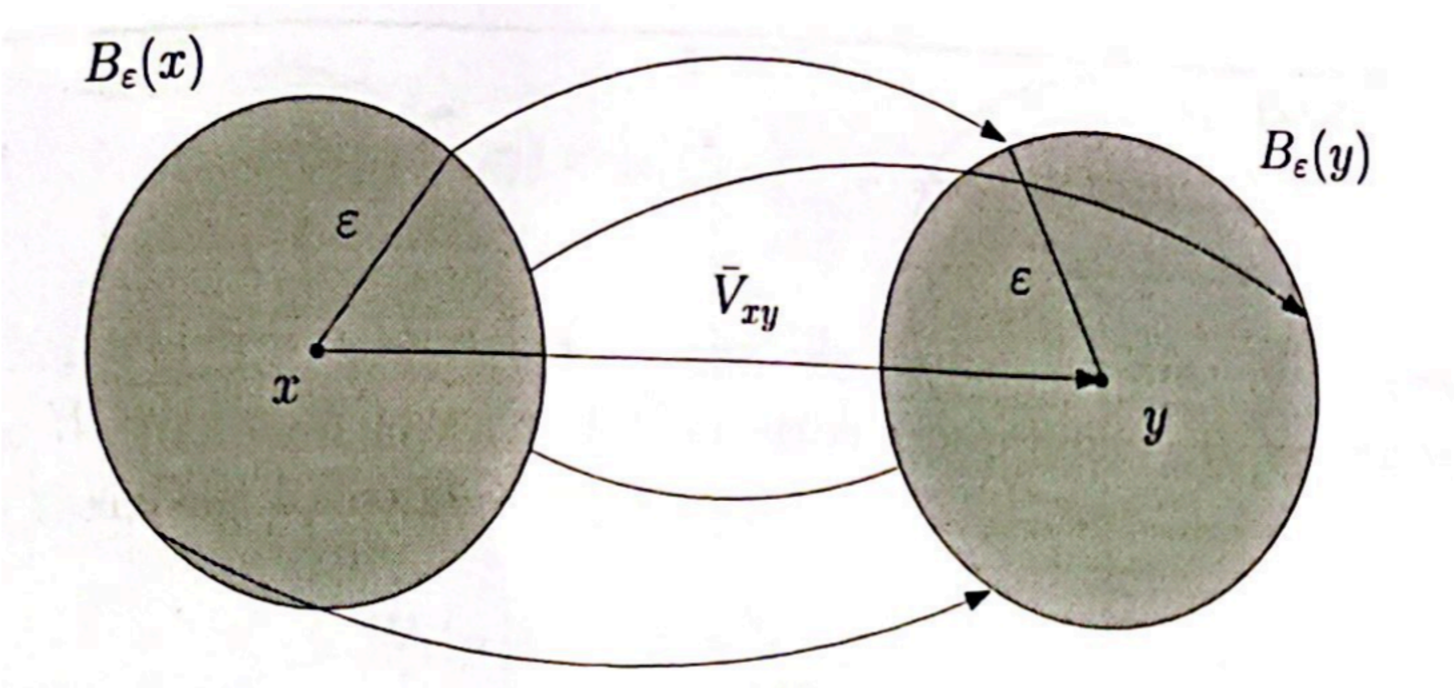
$$\mathrm{Ric}(X,Y) = \frac{1}{2}\left(\mathrm{Ric}(X+Y,X+Y) - \mathrm{Ric}(X,X) - \mathrm{Ric}(Y,Y)\right)$$

Ricci Curvature

- Measuring the degree to which the geometry determined by a given Riemannian metric might differ from that of ordinary Euclidean space

Olivier-Ricci Curvature

Transport ball B(x) to ball B(y).



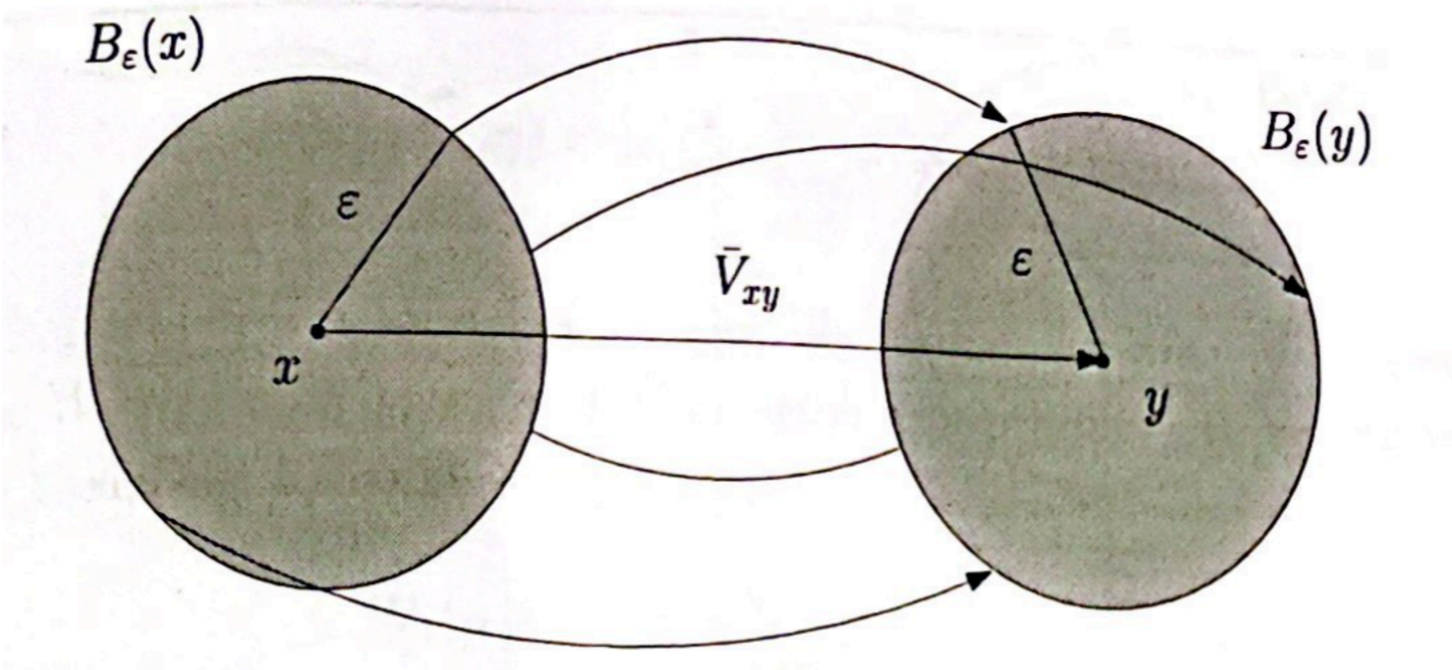
The average distance is

$$\delta\left[1-\frac{\varepsilon^2}{2(n+2)}\mathrm{Ric}(\bar{\mathbf{v}}_{xy})+O(\varepsilon^3+\varepsilon^2\delta)\right]$$

$$\delta = d(x,y).$$

Olivier-Ricci Curvature

Transport ball B(x) to ball B(y).



The average distance is

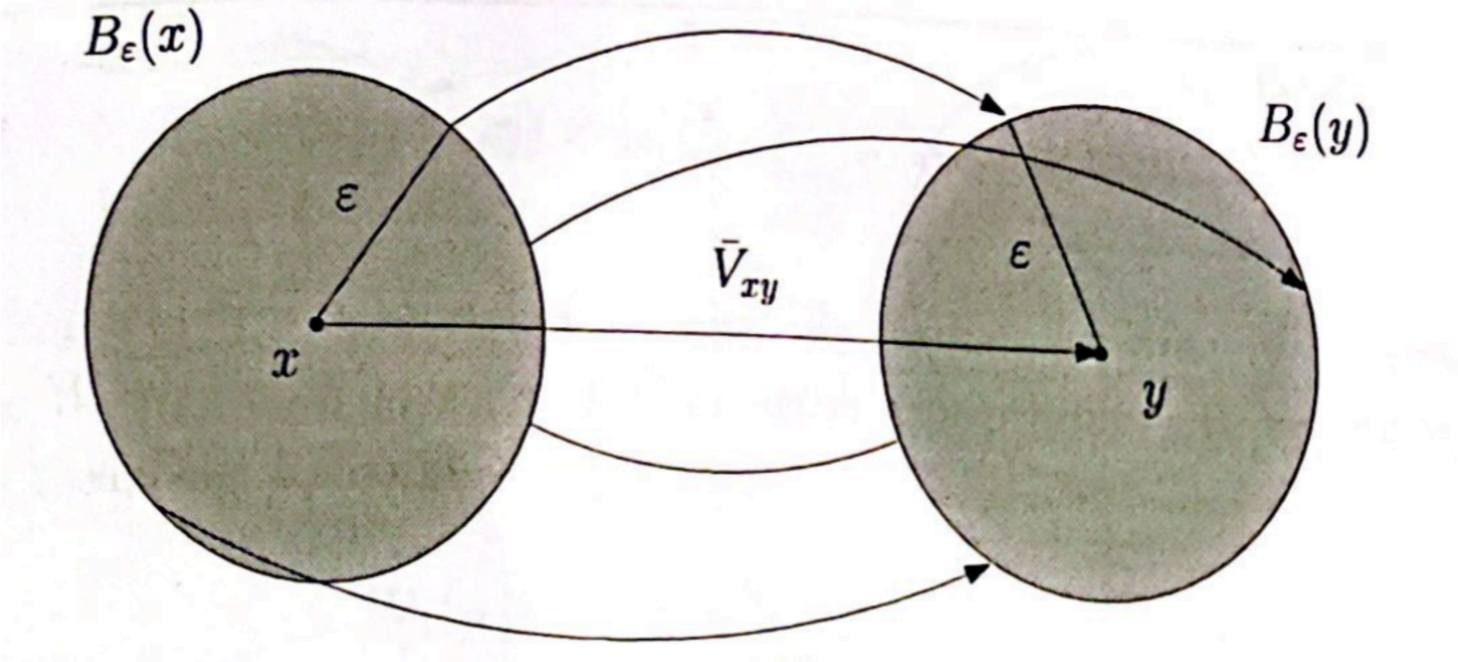
$$\delta\left[1 - \frac{\varepsilon^2}{2(n+2)}\mathrm{Ric}(\bar{\mathbf{v}}_{xy}) + O(\varepsilon^3 + \varepsilon^2\delta)\right]$$

$\delta = d(x, y)$.

$= W$

Olivier-Ricci Curvature

Transport ball B(x) to ball B(y).



The average distance is

$$\delta = d(x, y).$$

$$\delta\left[1 - \frac{\varepsilon^2}{2(n+2)}\mathrm{Ric}(\bar{\mathbf{v}}_{xy}) + O(\varepsilon^3 + \varepsilon^2\delta)\right]$$

$=W$

κ

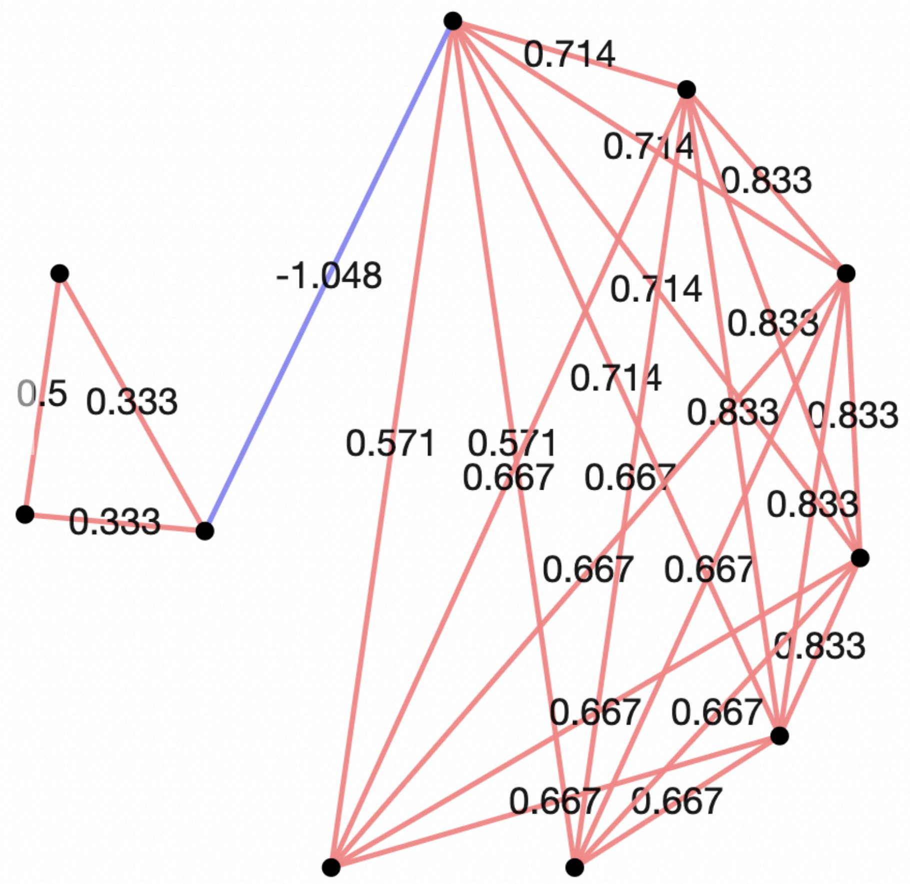
Olivier-Ricci Curvature

$$\kappa(u,v) = 1 - \frac{W(m_u^\alpha, m_v^\alpha)}{d(u,v)} \qquad W(m_u^\alpha, m_v^\alpha) = \inf_\xi \sum_{u,v \in V} \xi(u,v) d(u,v)$$

$$m_u^\alpha(x) = \begin{cases} \alpha & \text{if } x = u \\ (1-\alpha)/d_u & \text{if } x \in \delta(u) \\ 0 & \text{otherwise} \end{cases}$$

- Graphs are generated from manifold
- OR curvature on Graphs $\rightarrow$ Ricci curvature on Manifold

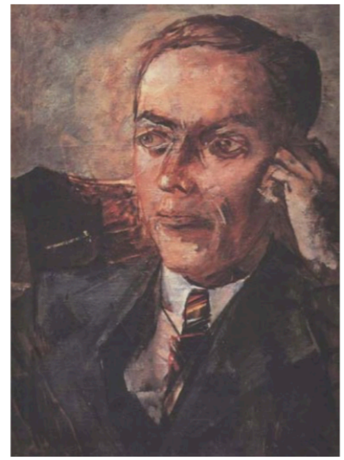
Subsampling in Graphs

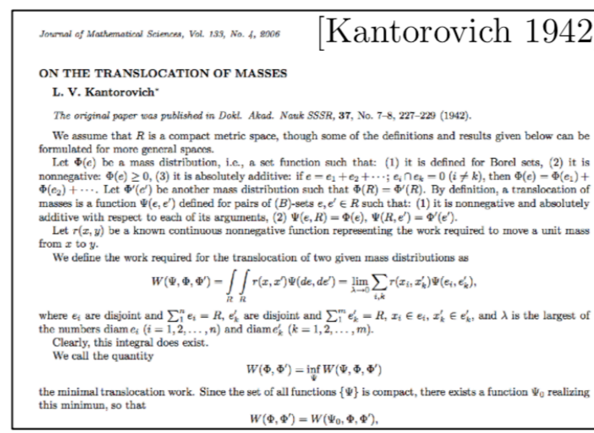


Edges with large curvature are within a community;
Edges with small curvature are between communities

Leonid Kantorovich (1912-1986)

Леонид Витальевич Канторович





Journal of Mathematical Sciences, Vol. 133, No. 4, 2006

[Kantorovich 1942]

ON THE TRANSLOCATION OF MASSES

L. V. Kantorovich*

The original paper was published in Dokl. Akad. Nauk SSSR, **37**, No. 7–8, 227–229 (1942).

We assume that R is a compact metric space, though some of the definitions and results given below can be formulated for more general spaces.

Let $\Phi(e)$ be a mass distribution, i.e., a set function such that: (1) it is defined for Borel sets, (2) it is nonnegative: $\Phi(e) \geq 0$, (3) it is absolutely additive: if $e = e_1 + e_2 + \cdots$; $e_i \cap e_k = 0$ $(i \neq k)$, then $\Phi(e) = \Phi(e_1) + \Phi(e_2) + \cdots$. Let $\Phi'(e')$ be another mass distribution such that $\Phi(R) = \Phi'(R)$. By definition, a translocation of masses is a function $\Psi(e, e')$ defined for pairs of (B)-sets $e, e' \in R$ such that: (1) it is nonnegative and absolutely additive with respect to each of its arguments, (2) $\Psi(e, R) = \Phi(e)$, $\Psi(R, e') = \Phi'(e')$.

Let $r(x, y)$ be a known continuous nonnegative function representing the work required to move a unit mass from x to y.

We define the work required for the translocation of two given mass distributions as

$$W(\Psi, \Phi, \Phi') = \int_R \int_R r(x, x')\Psi(de, de') = \lim_{\lambda \to 0} \sum_{i,k} r(x_i, x'_k)\Psi(e_i, e'_k),$$

where e_i are disjoint and $\sum_1^n e_i = R$, e'_k are disjoint and $\sum_1^m e'_k = R$, $x_i \in e_i$, $x'_k \in e'_k$, and λ is the largest of the numbers $\operatorname{diam} e_i$ $(i = 1, 2, \ldots, n)$ and $\operatorname{diam} e'_k$ $(k = 1, 2, \ldots, m)$.

Clearly, this integral does exist.

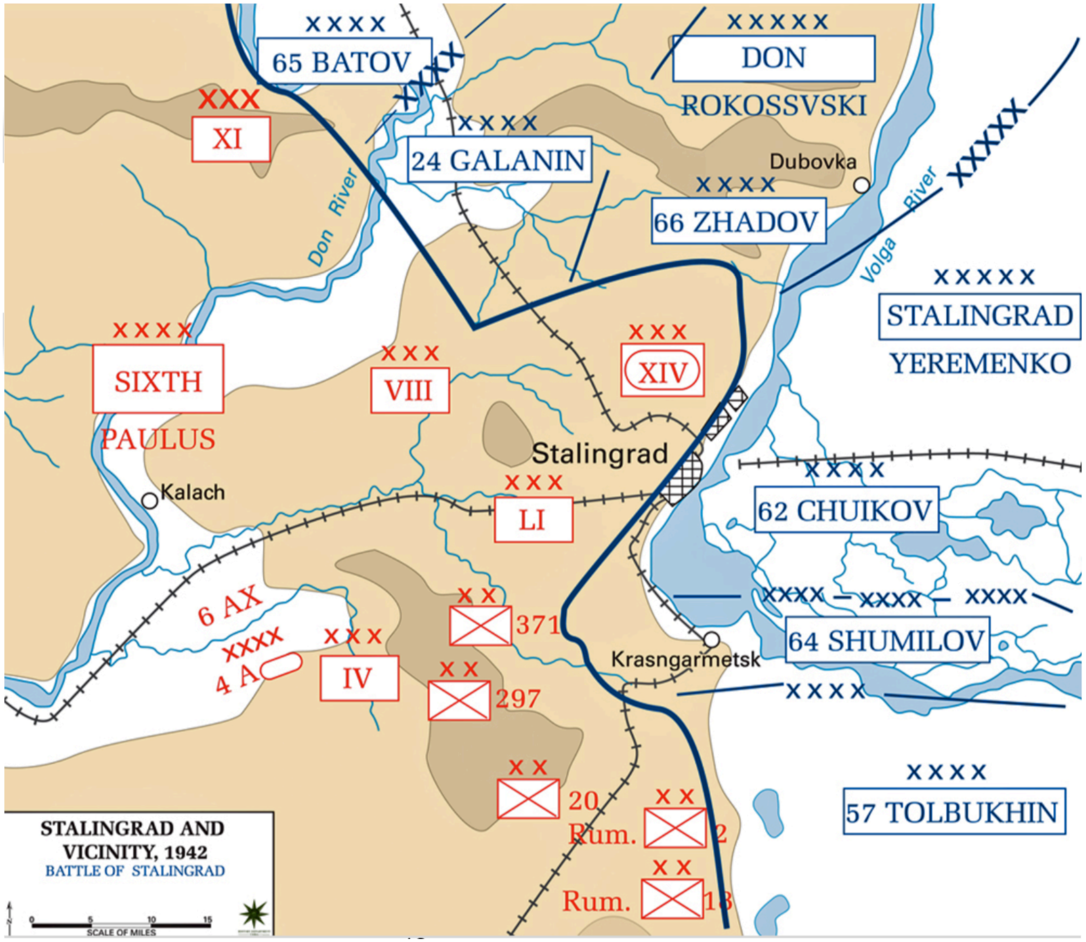
We call the quantity

$$W(\Phi, \Phi') = \inf_{\Psi} W(\Psi, \Phi, \Phi')$$

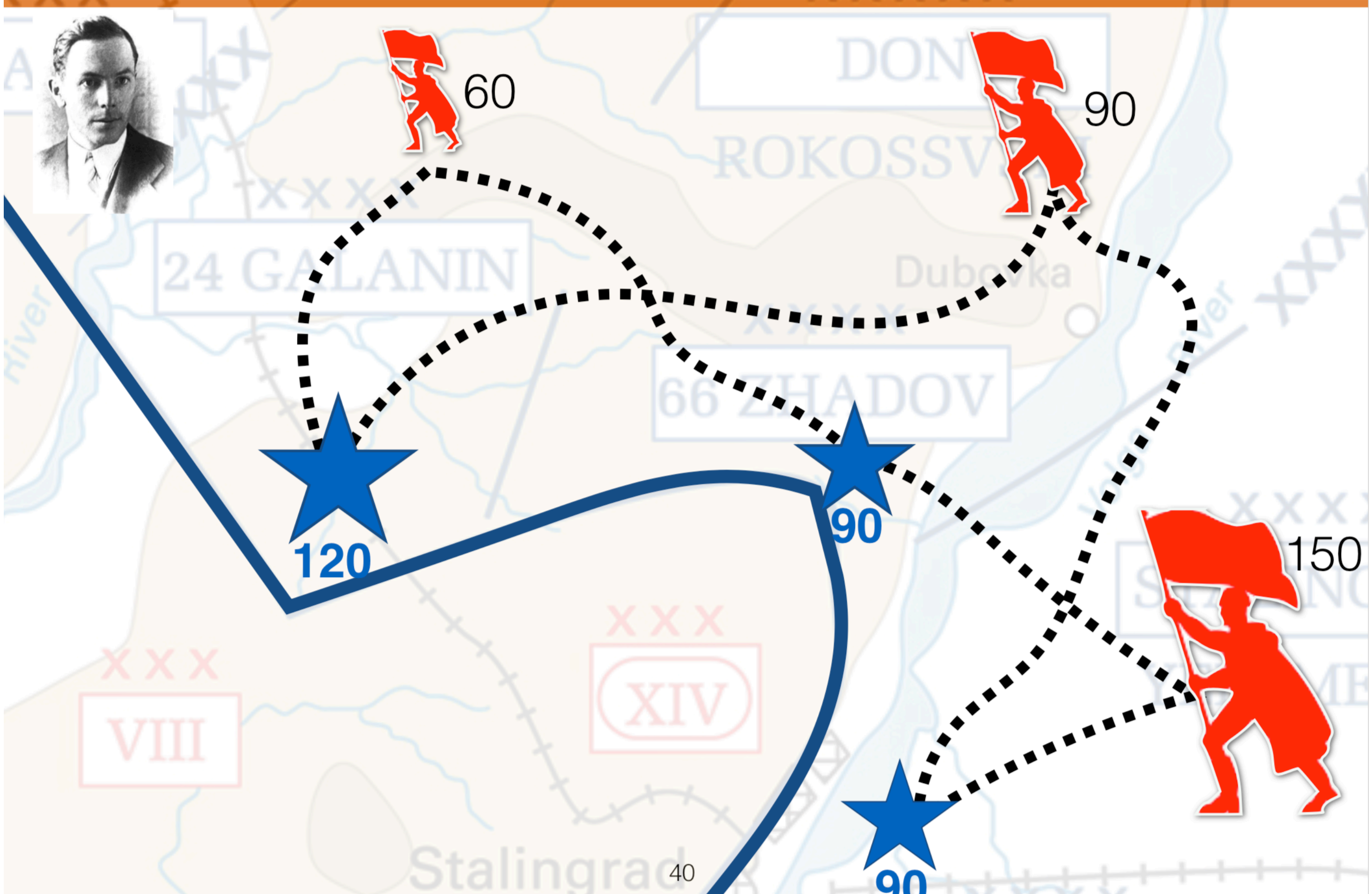
the minimal translocation work. Since the set of all functions $\{\Psi\}$ is compact, there exists a function Ψ_0 realizing this minimun, so that

$$W(\Phi, \Phi') = W(\Psi_0, \Phi, \Phi'),$$

Kantorovich Problem



Kantorovich Problem



Kantorovich Problem

Transportation matrix

60	?	?	?
90	?	?	?
150	?	?	?
	120	90	90

Distance matrix

	A	B	C
1	d_{1A}	d_{1B}	d_{1C}
2	d_{2A}	d_{2B}	d_{2C}
3	d_{3A}	d_{3B}	d_{3C}

Kantorovich Problem

Transportation matrix

a_1	p_{1A}	p_{1B}	p_{1C}
a_2	p_{2A}	p_{2B}	p_{2C}
a_3	p_{3A}	p_{3B}	p_{3C}
	b_A	b_B	b_C

Distance matrix

1	d_{1A}	d_{1B}	d_{1C}
2	d_{2A}	d_{2B}	d_{2C}
3	d_{3A}	d_{3B}	d_{3C}
	A	B	C

Constraints

$$\forall i \in \{1,2,3\}, \quad \sum_{j \in \{A,B,C\}} p_{ij} = a_i$$

$$\forall j \in \{A,B,C\}, \quad \sum_{i \in \{1,2,3\}} p_{ij} = b_j$$

$$p_{ij} \geq 0$$

Cost function

$$C(P) = \sum_{j \in \{A,B,C\}} \sum_{i \in \{1,2,3\}} p_{ij} d_{ij}$$

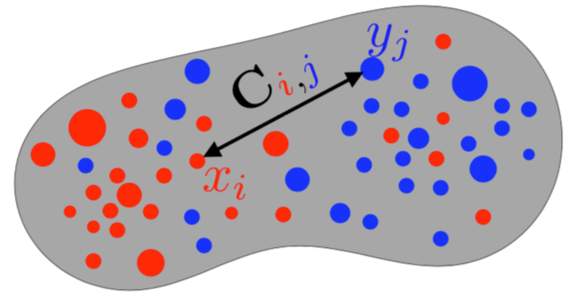
Problem

$$\min_{\text{all valid } P} C(P)$$

Kantorovitch's Formulation

Input distributions

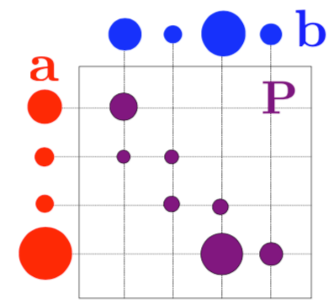
$\alpha = \sum_{i=1}^{n} \mathbf{a}_i \delta_{x_i} \quad \beta = \sum_{j=1}^{m} \mathbf{b}_j \delta_{y_j}$



Points $(x_i)_i$, $(y_j)_j$

Weights $\mathbf{a}_i \geqslant 0$, $\mathbf{b}_j \geqslant 0$.

$\sum_{i=1}^{n} \mathbf{a}_i = \sum_{j=1}^{m} \mathbf{b}_j = 1$



Couplings:

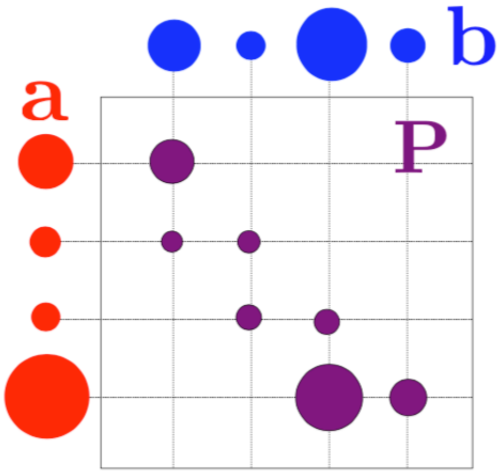
$\mathbf{U}(\mathbf{a}, \mathbf{b}) \stackrel{\text{def.}}{=} \left\{\mathbf{P} \in \mathbb{R}_+^{n \times m} \; ; \; \mathbf{P}\mathbb{1}_m = \mathbf{a}, \mathbf{P}^\top \mathbb{1}_n = \mathbf{b}\right\}$

[Kantorovich 1942]

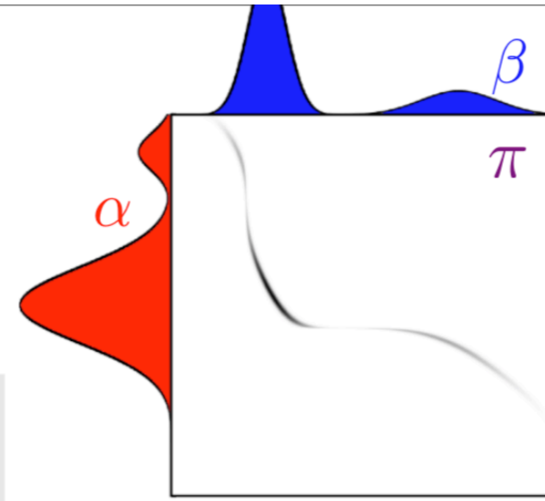
$$\min \left\{ \sum_{i,j} \mathbf{C}_{i,j} \mathbf{P}_{i,j} \; ; \; \mathbf{P} \in \mathbf{U}(\mathbf{a}, \mathbf{b}) \right\}$$

$\rightarrow$ Linear program, simplex $O(n^3 \log(n))$.

Wasserstein Distance



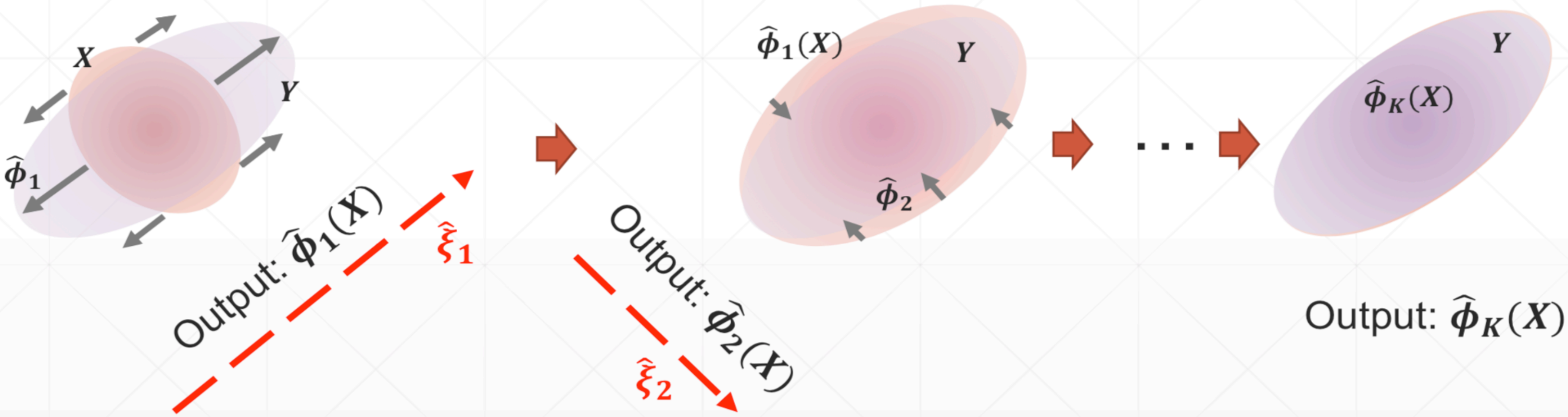
$$\pi = \sum_{i,j} \mathbf{P}_{i,j} \delta_{x_i, y_j}$$



$$c(x,y) = d(x,y)^p$$

$$\mathrm{W}_p(\alpha, \beta)^p \stackrel{\text{def.}}{=} \min_{\pi \in \mathcal{M}_+^1(\mathcal{X}^2)} \left\{ \int_{\mathcal{X}^2} d(x,y)^p \mathrm{d}\pi(x,y) \; ; \; \pi_1 = \alpha, \pi_2 = \beta \right\}$$

Projection pursuit Monge map (PPMM)



K: # *Transportations*

Computational cost: $O(Knd^2 + Knlog(n))$

Meng et al 2019 NeuriPS

Generative Models

Inpainting
NVIDIA, 2018

Super resolution
Dahl et al, CVPR 2017

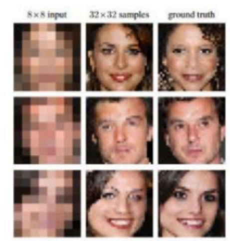
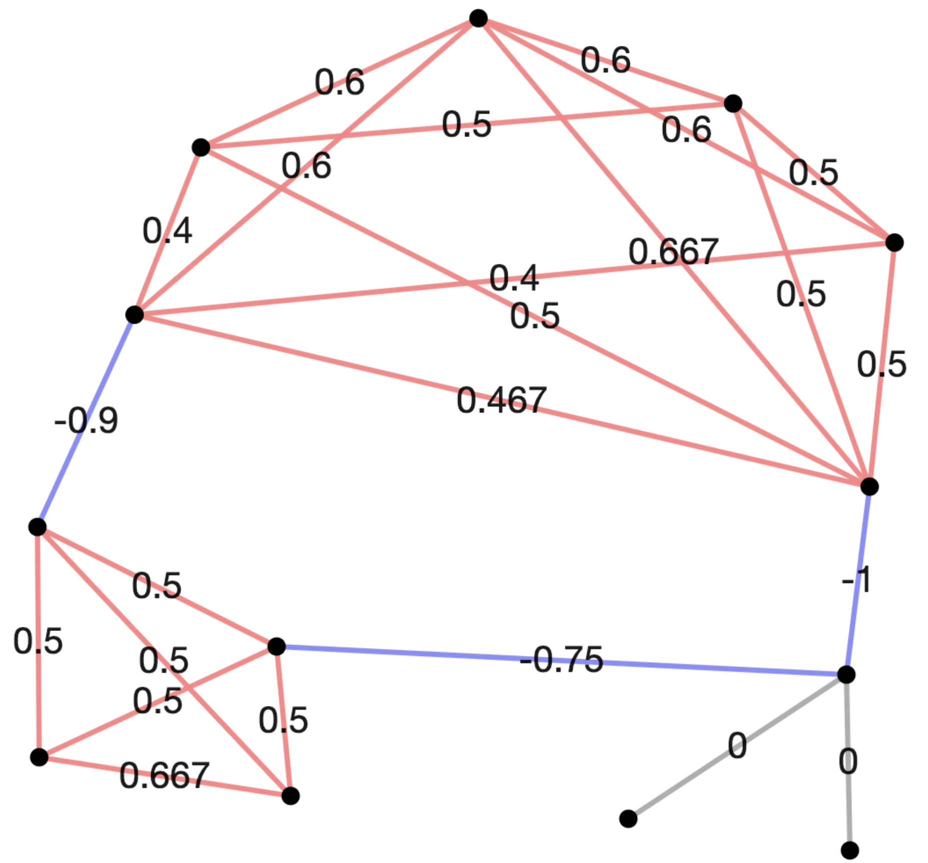


Image coloring
Isola et al., CVPR 2017

Color transfer
Arbelot et al., LJK 2015

Photos to Emojis
Taigman et al, 2016

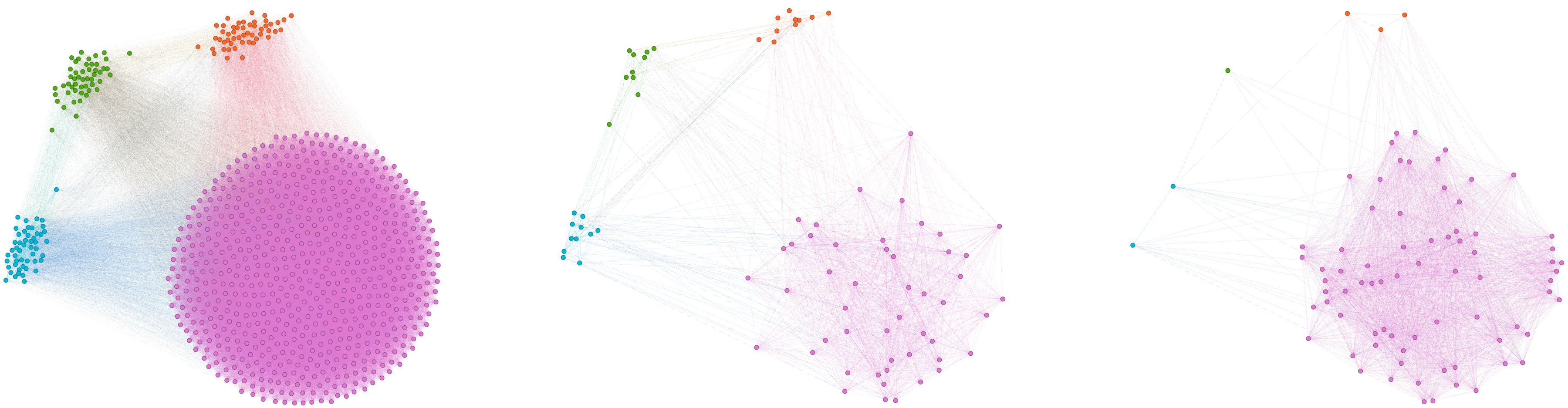
Subsampling in Graphs



Edges with large curvature are within a community;
Edges with small curvature are between communities

OR Curvature Gradient-based Subsampling

$$\left(x^{(i+1)}, y^{(i+1)}\right) = \text{argmax}_{(x,y)\in\Delta\left(\left(x^{(i)},y^{(i)}\right)\right)} \left|\kappa(x,y) - \kappa\left(x^{(i+1)}, y^{(i+1)}\right)\right|$$



Experiment Results

Dataset	Prop.	ORG-sub	MHRW	CSE	FFS	Snowball	RW	MDRW
Polbooks (T: 1.88 s)	10%	**0.00 (T: 0.10 s)**	1.20	0.62	2.68	0.48	0.33	0.00
Polblogs (T: 48.6 s)	5%	**0.00 (T: 0.23 s)**	1.87	0.90	2.00	0.43	1.03	0.30
PubMed (T: NA)	2%	**0.00 (T: 4.42 s)**	0.30	0.80	0.40	0.20	1.20	1.80

Time of estimation of M is the much lower than full sample!

Error of estimation of M is the lowest!

Acknowledgement

ICLR 2023

Shushan Wu

Huimin Cheng

Jiazhang Cai

Wenxuan Zhong