NetLSD: hearing the shape of a graph

Anton Tsitsulin[1] Davide Mottin[1] Panagiotis Karras[2]
Alex Bronstein[3] Emmanuel Müller[1]

[1] HPI
Germany

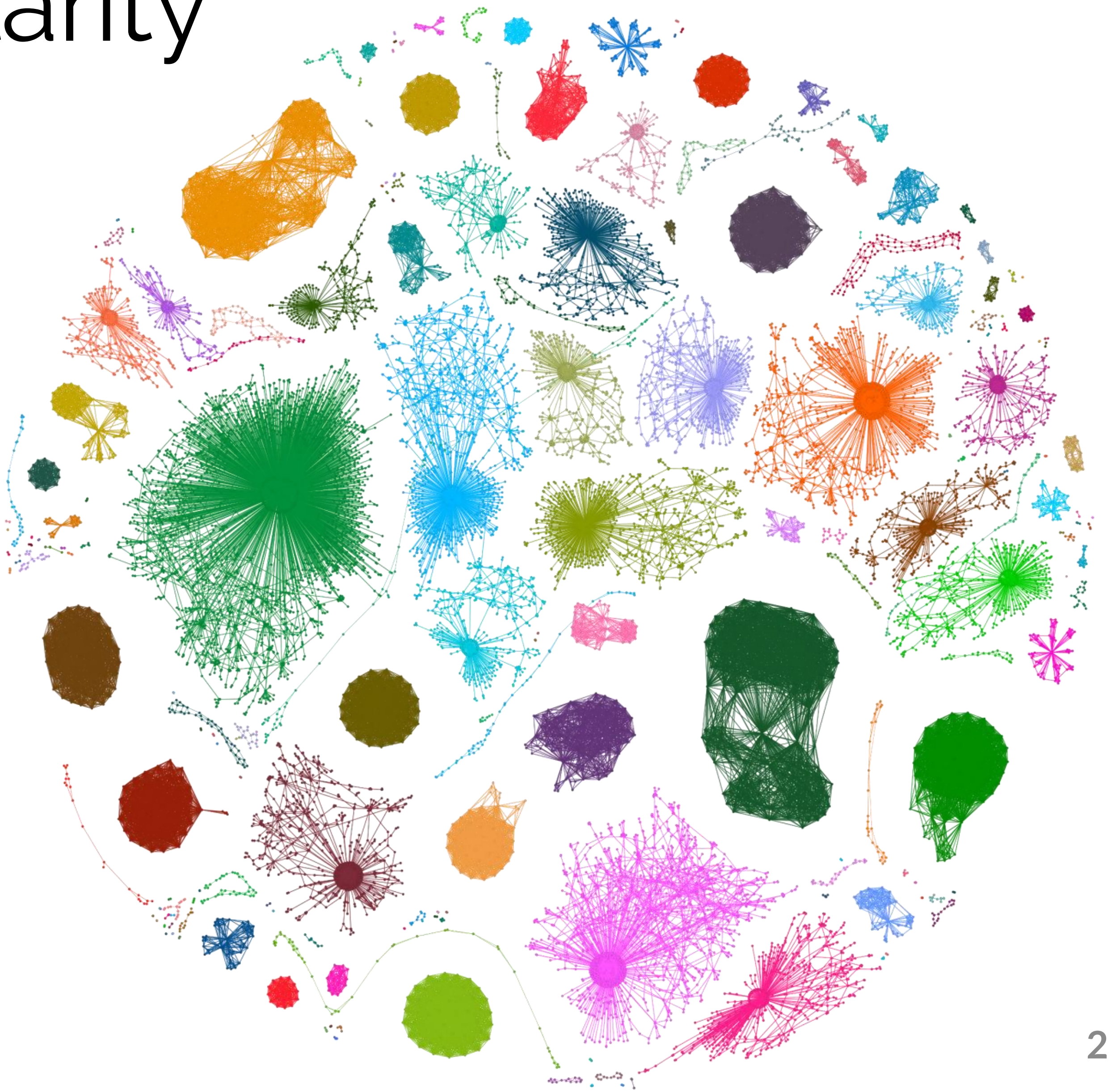
[2] Aarhus university
Denmark

[3] Technion
Israel

Defining graph similarity

With it, we can do:

- Classification
- Clustering
- Anomaly detection
- …

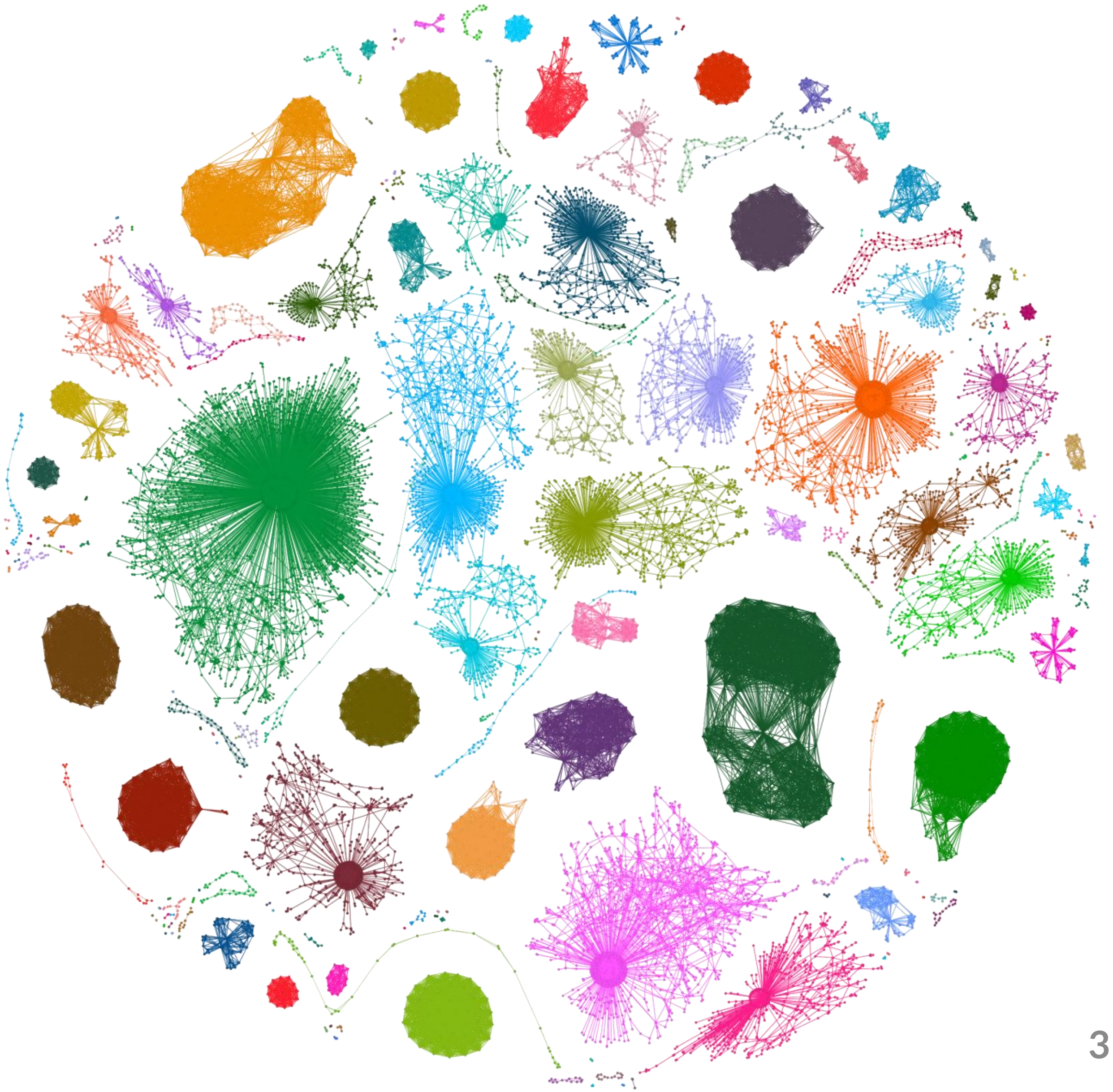


Scalability is key!

Two problem sources:

- Big graphs
- Many graphs

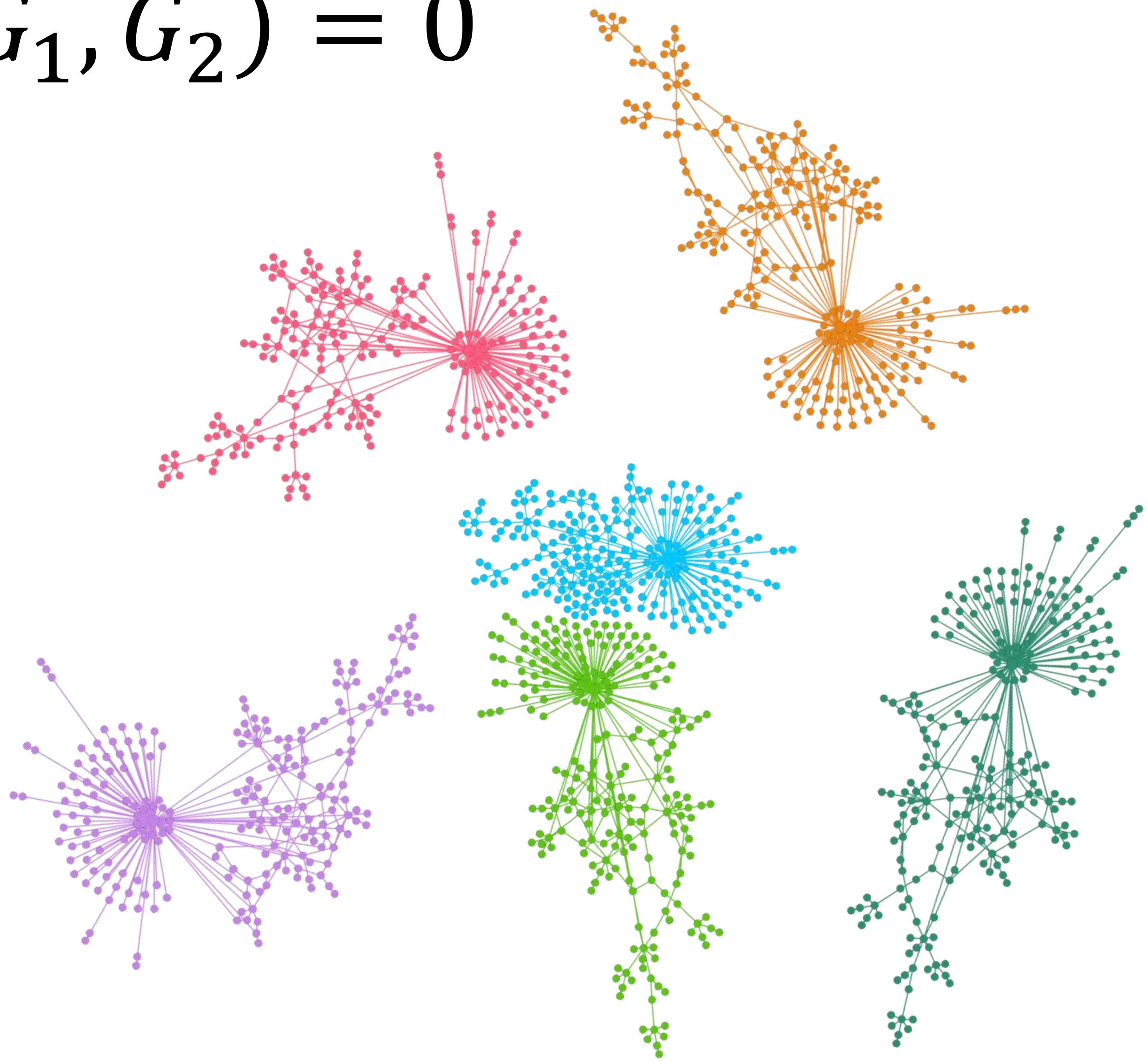
Solution: graph descriptors



Isomorphism $\Rightarrow d(G_1, G_2) = 0$

3 key properties:

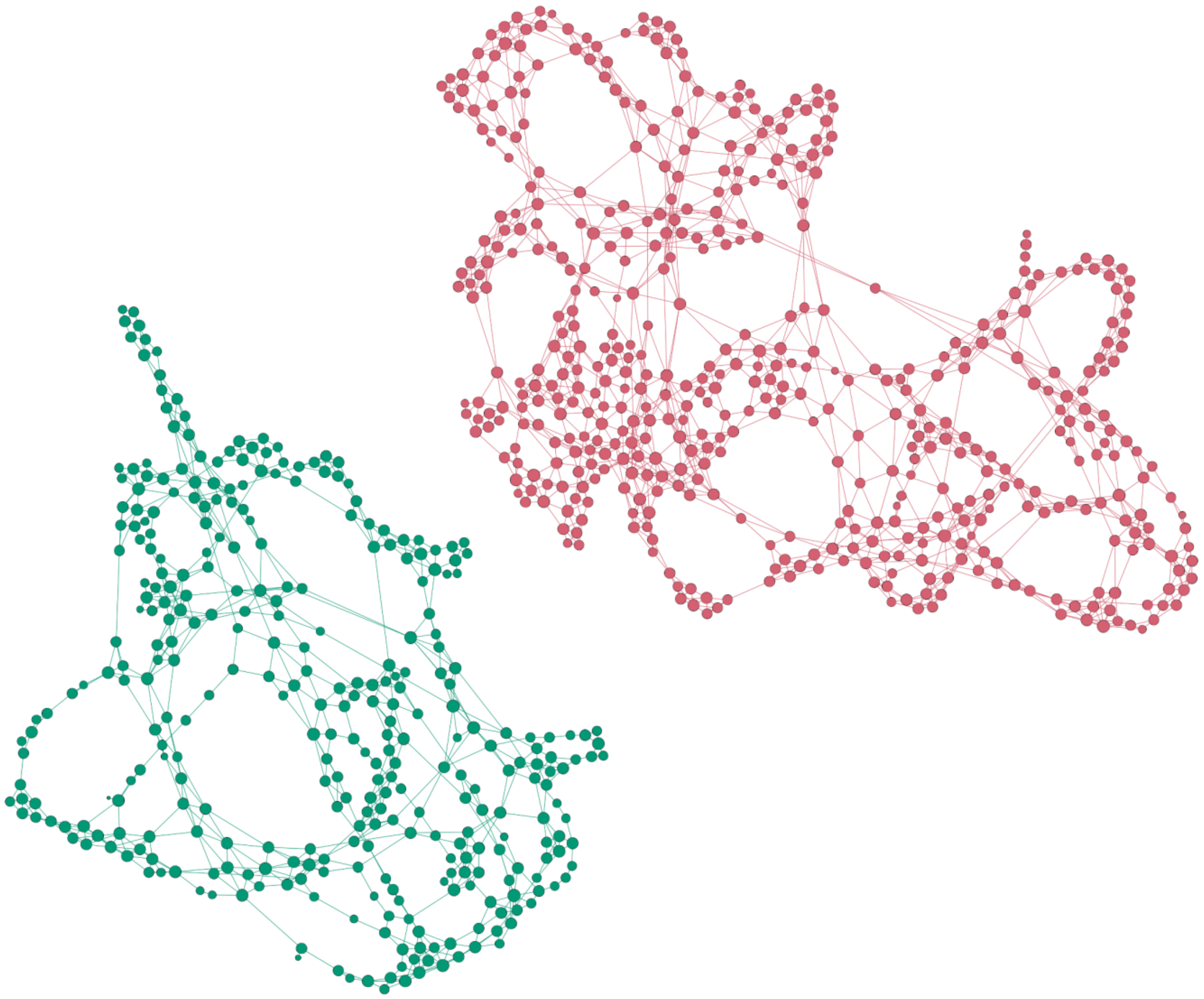
- Permutation invariance
- Scale-adaptivity
- Size invariance



Local structures are important

3 key properties:

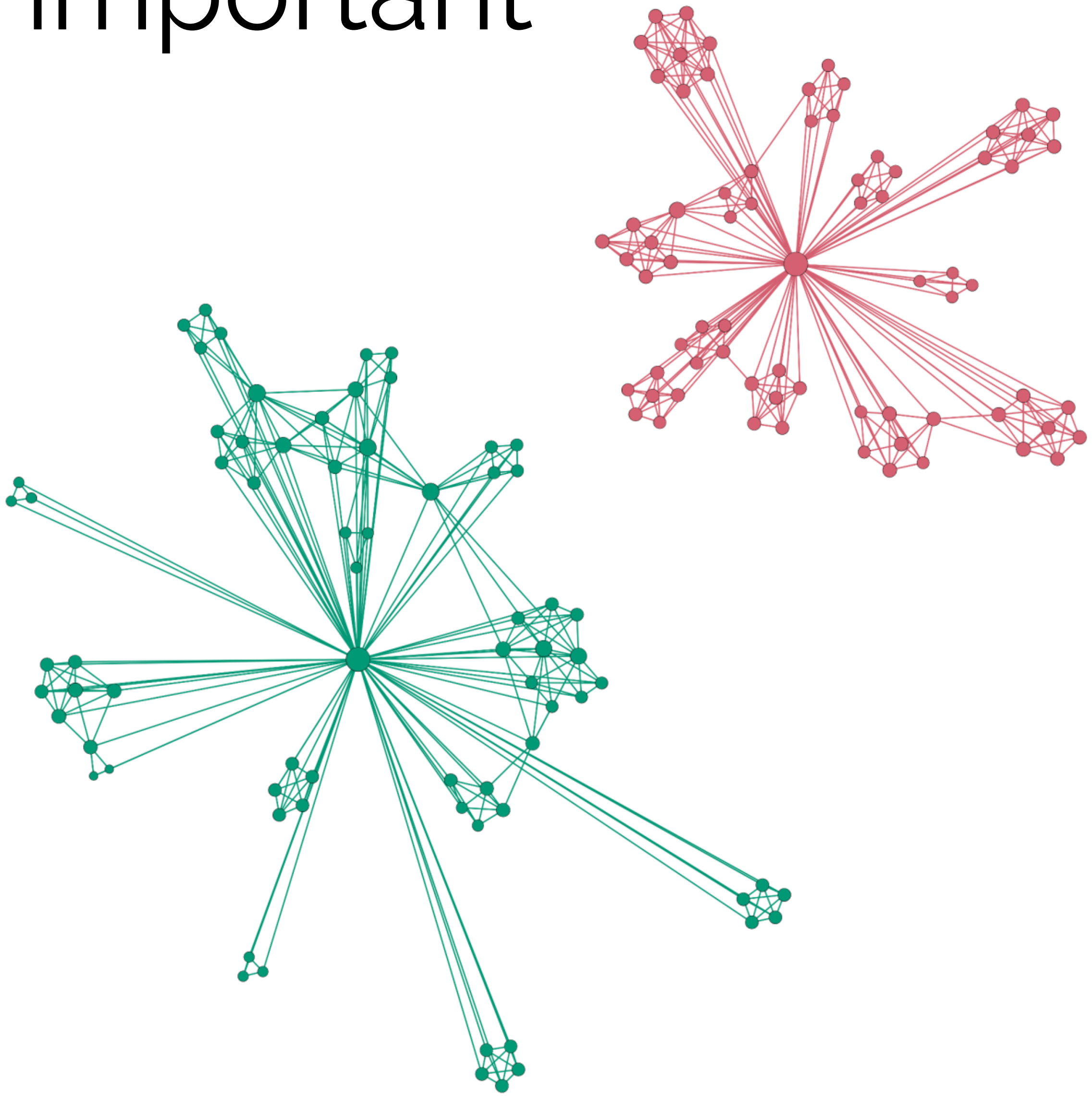
- Permutation invariance
- Scale-adaptivity
- Size invariance



Global structure is important

3 key properties:

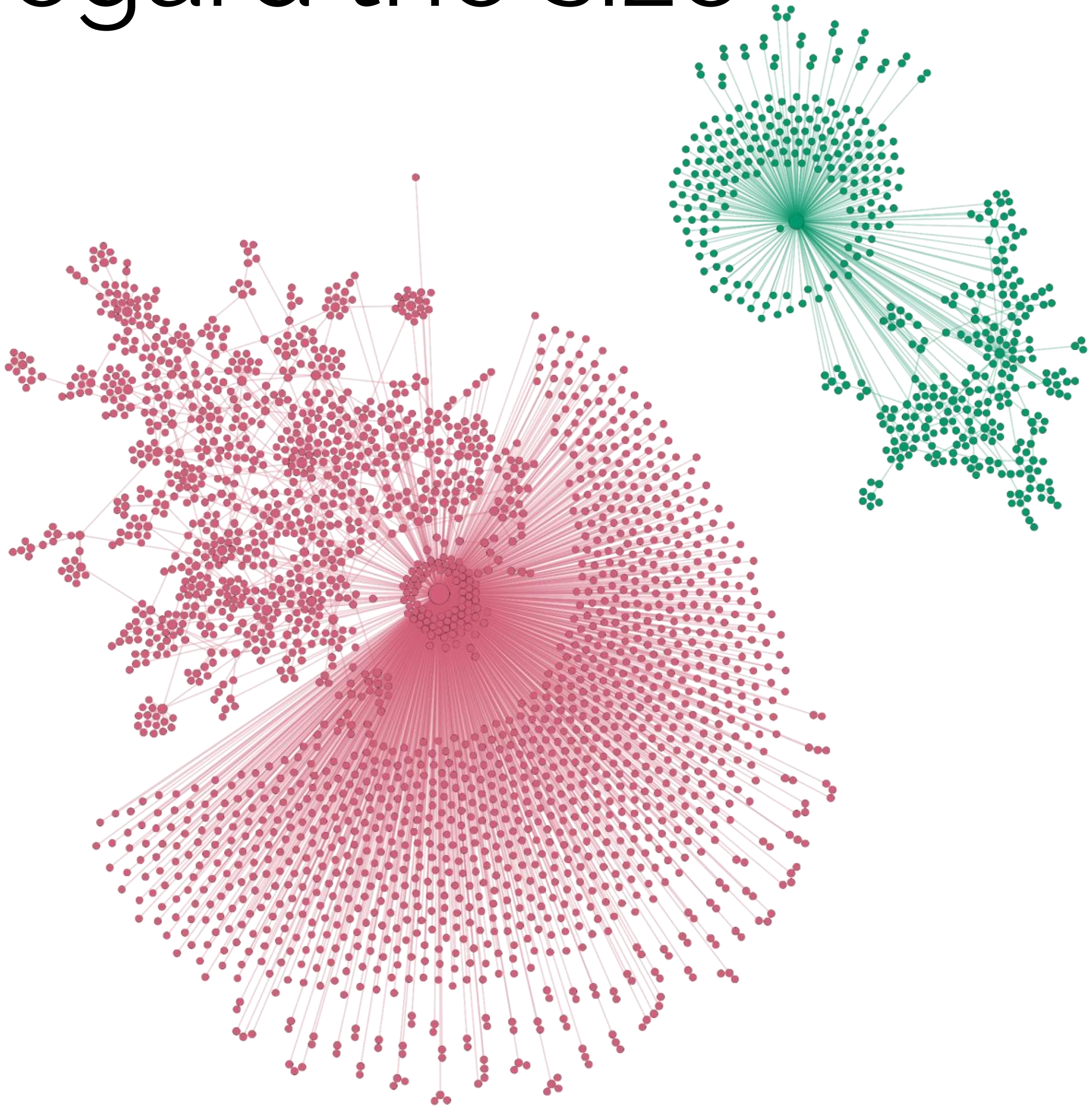
- Permutation invariance
- Scale-adaptivity
- Size invariance



We may need to disregard the size

3 key properties:

- Permutation invariance
- Scale-adaptivity
- Size invariance



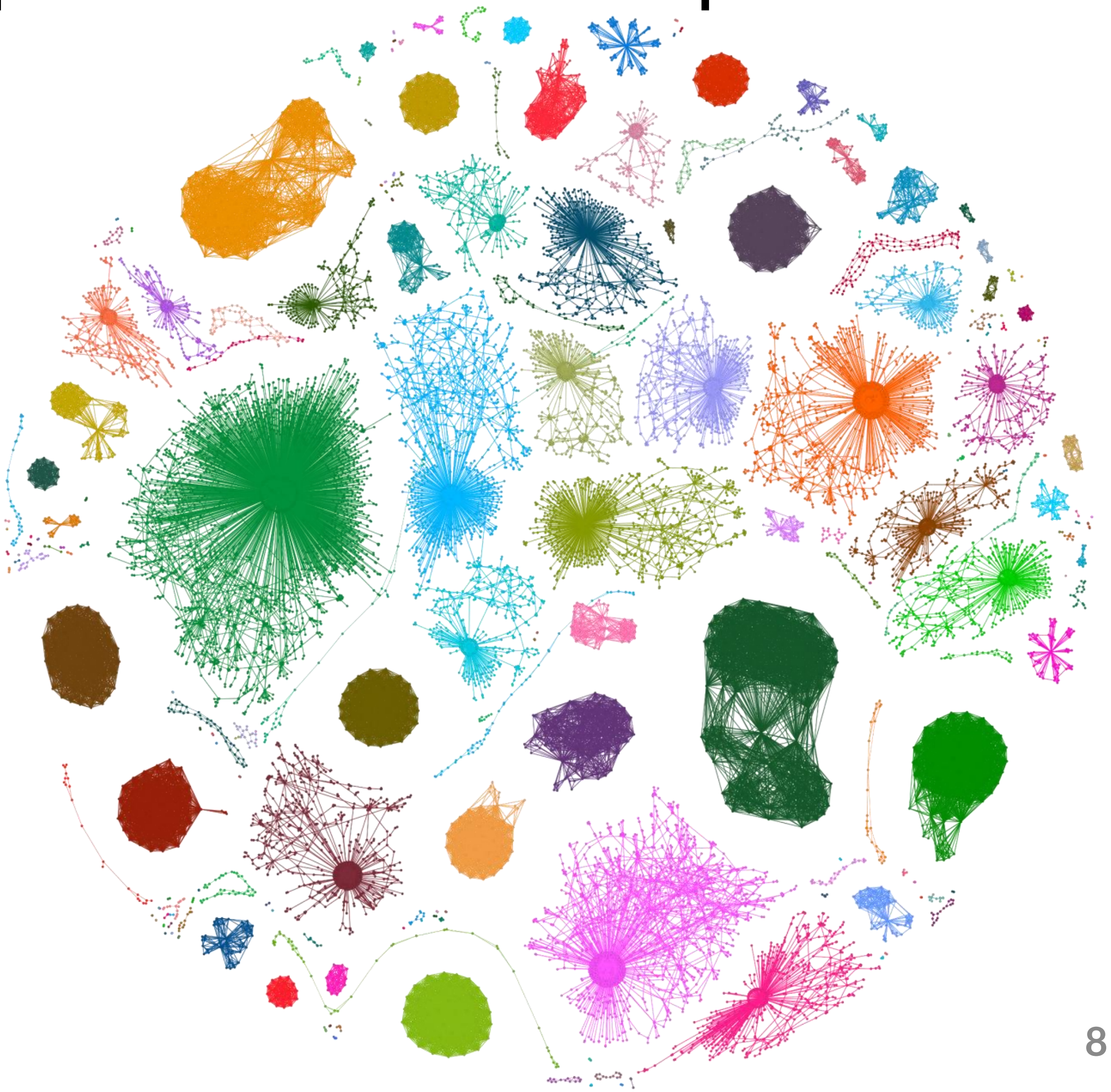
Network Laplacian Spectral Descriptors

3 key properties:

- Permutation invariance
- Scale-adaptivity
- Size invariance

+ Scalability

= **NetLSD**



Optimal Transport

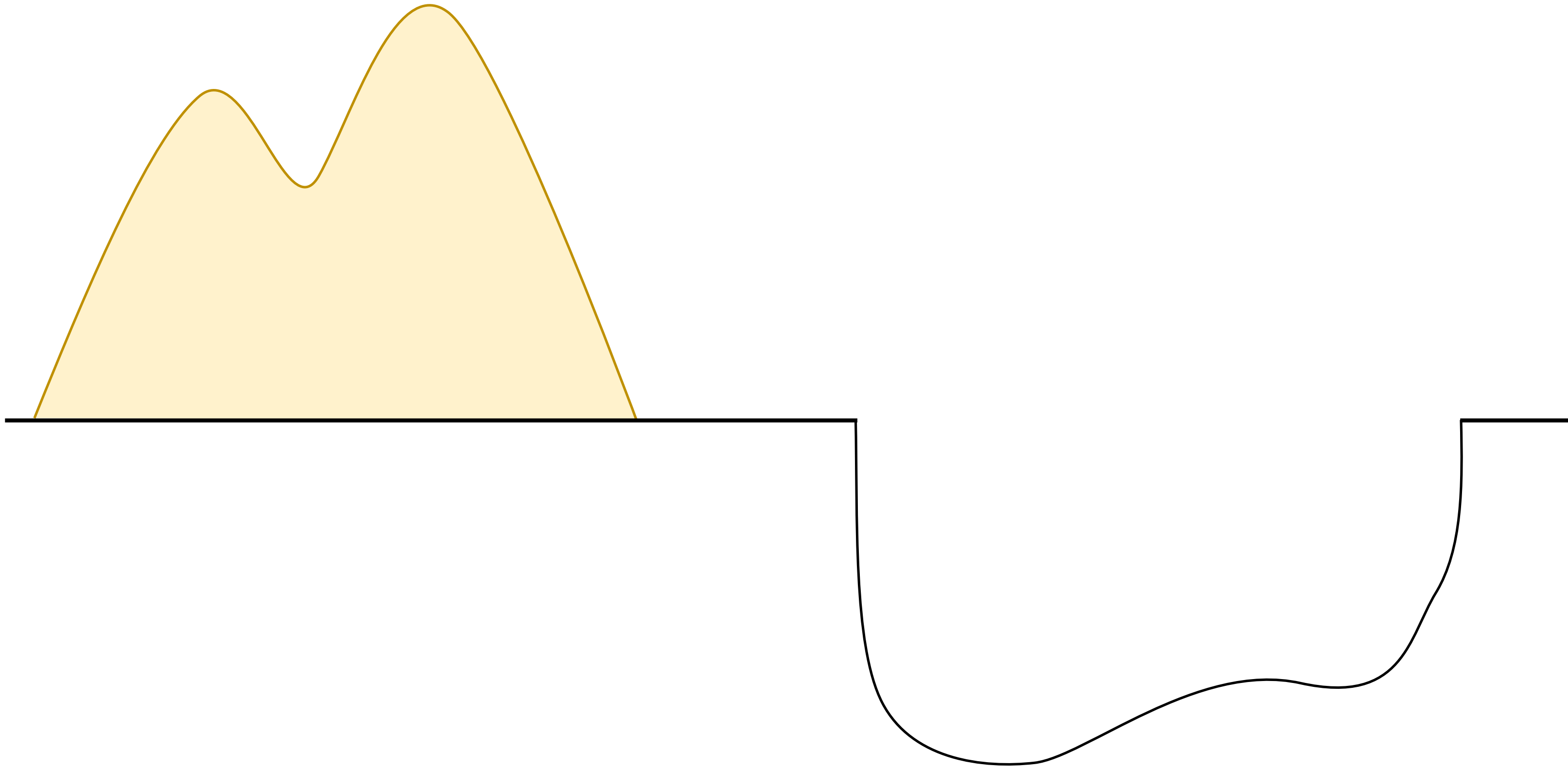
Geometry for probability measures supported on a space.

Optimal Transport

Geometry for probability measures supported on a space.

G. Monge
1781

L. Kantorovich
1939

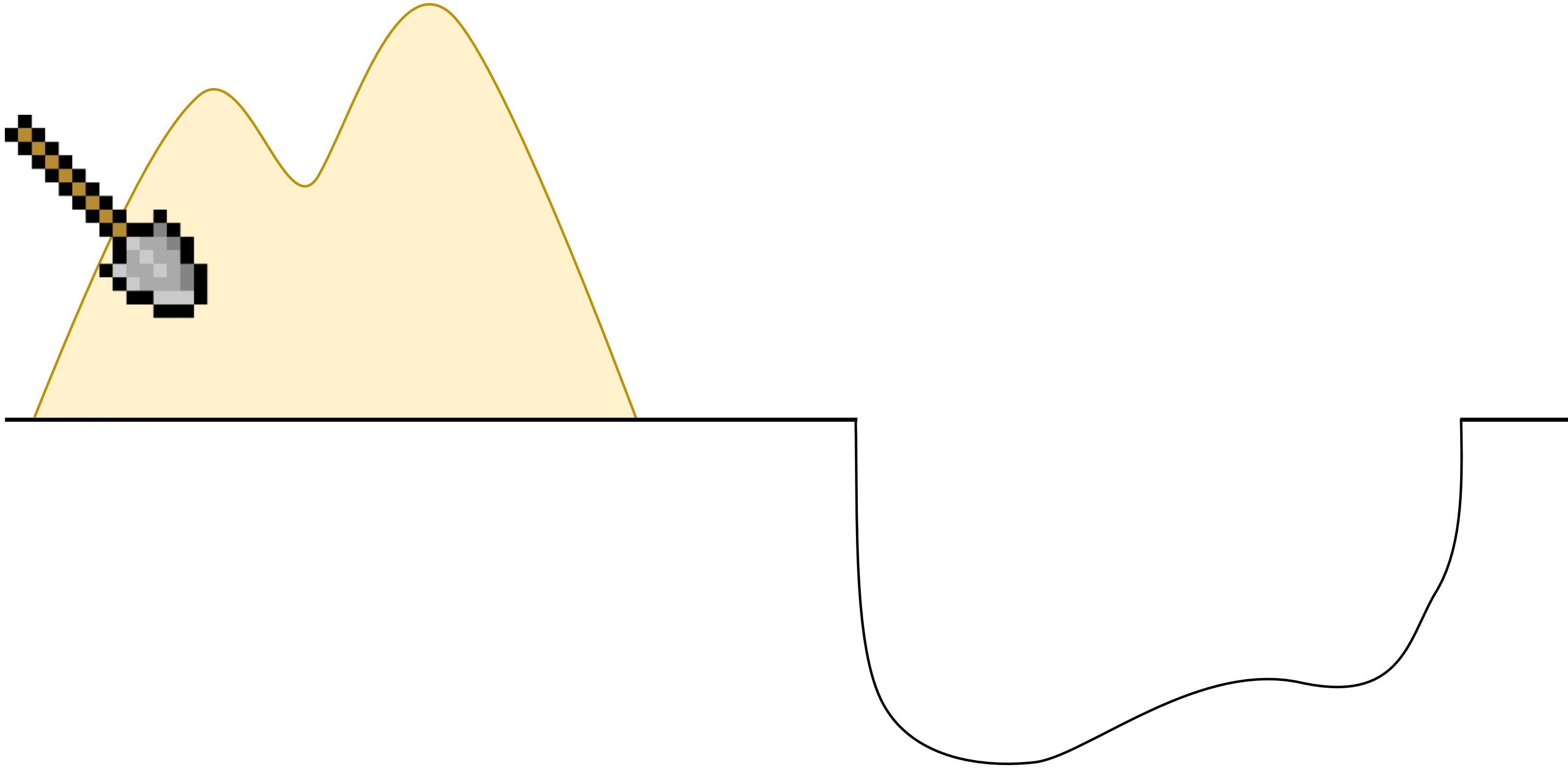


Optimal Transport

Geometry for probability measures supported on a space.

G. Monge
1781

L. Kantorovich
1939

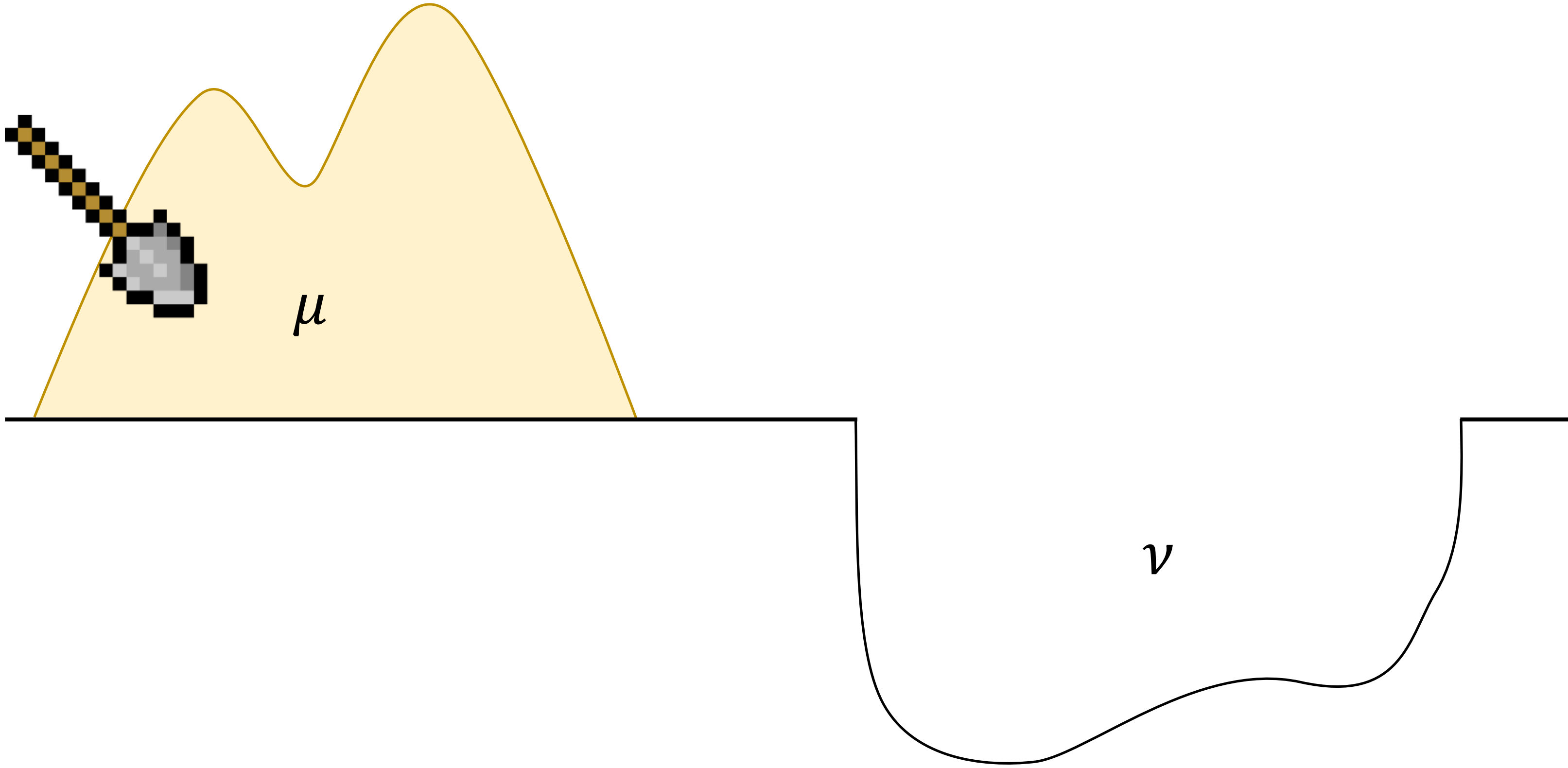


Optimal Transport

Geometry for probability measures supported on a space.

G. Monge
1781

L. Kantorovich
1939

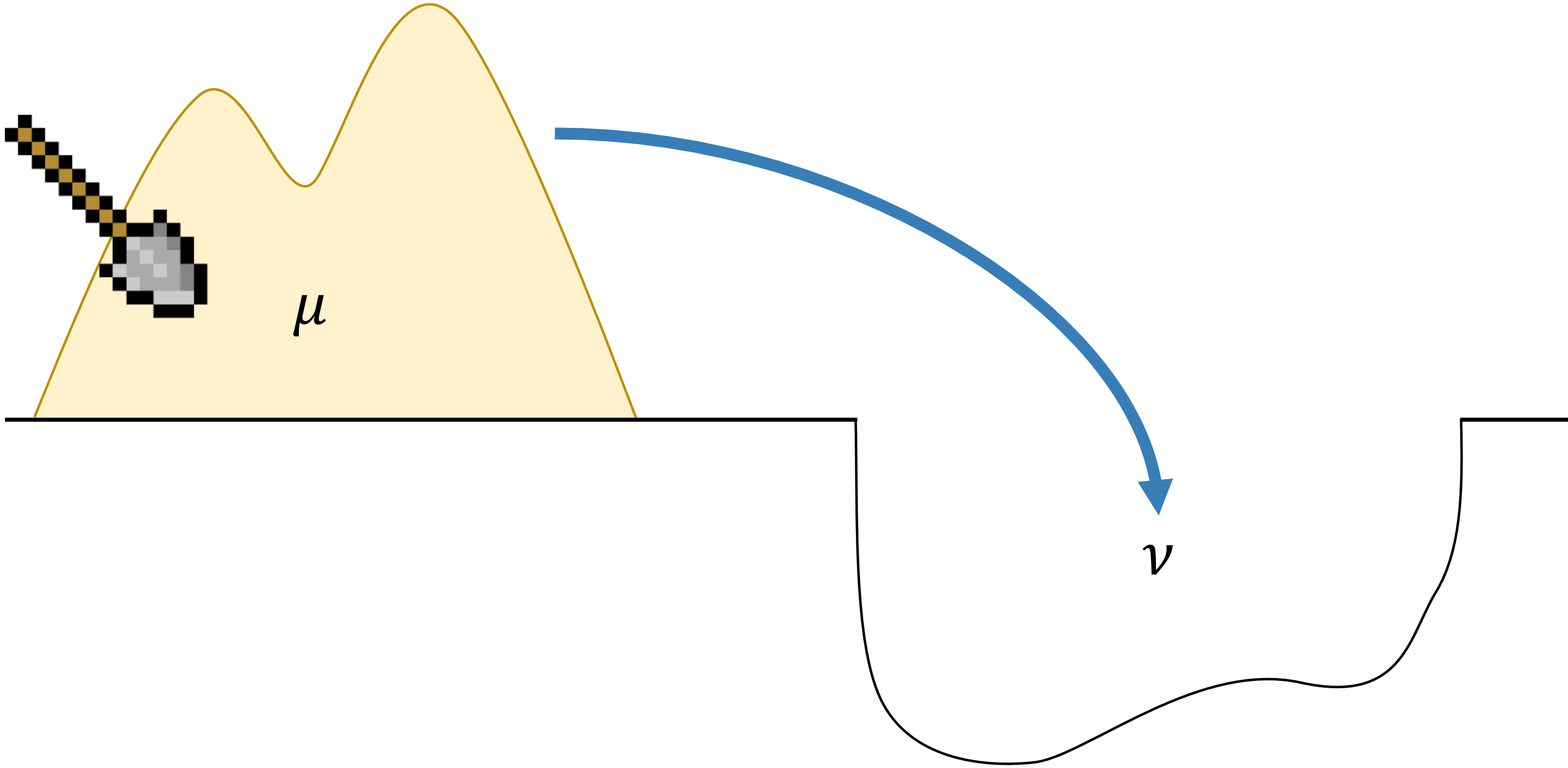


Optimal Transport

Geometry for probability measures supported on a space.

G. Monge
1781

L. Kantorovich
1939



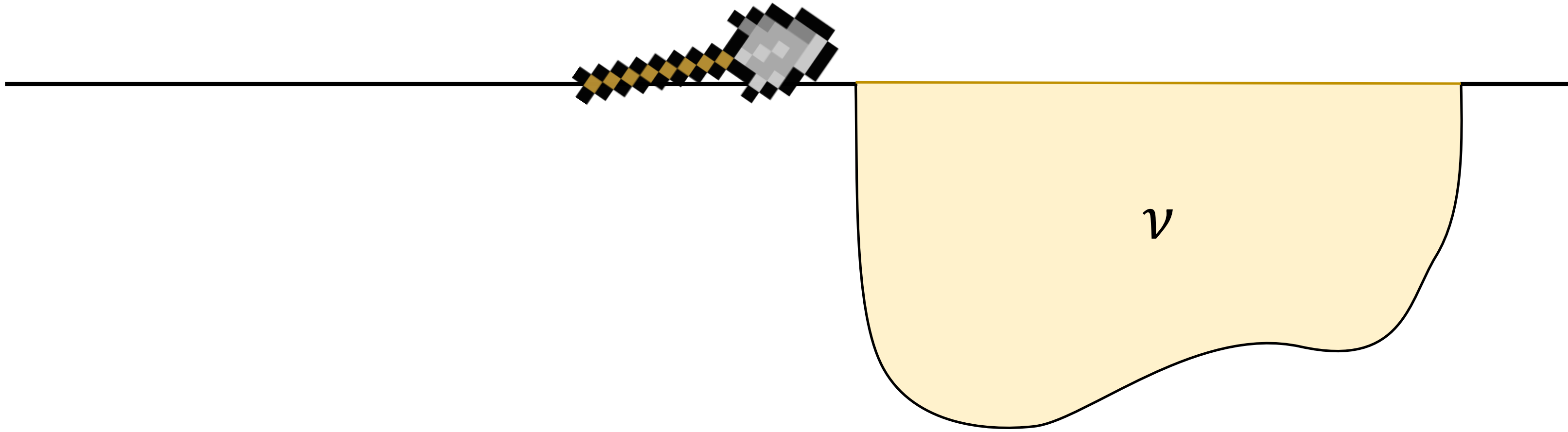
Optimal Transport

Geometry for probability measures supported on a space.

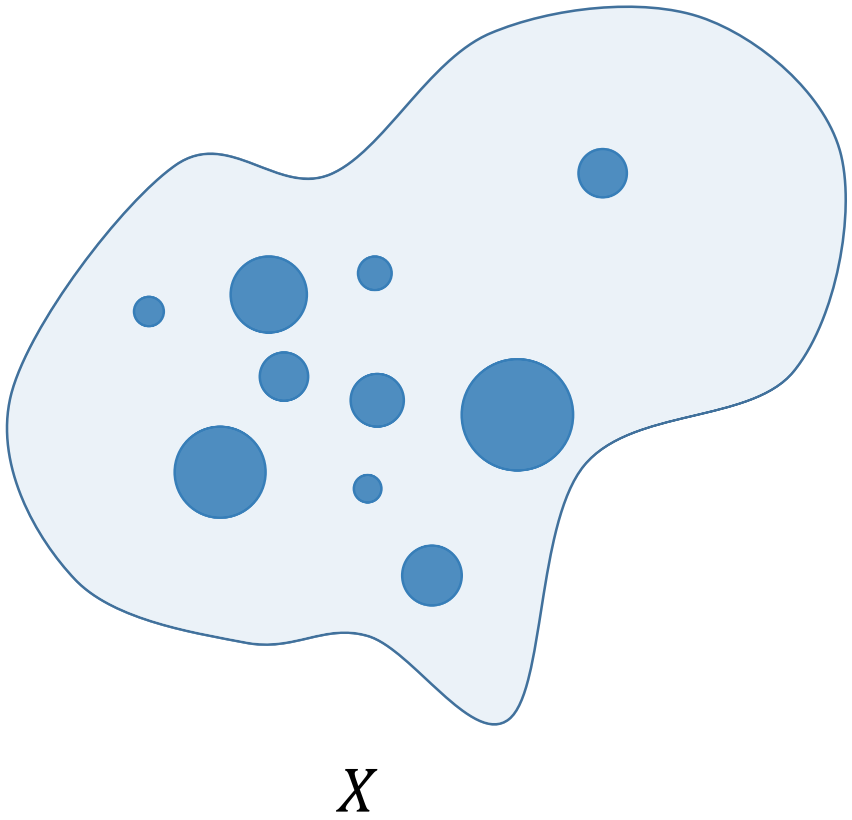
G. Monge
1781

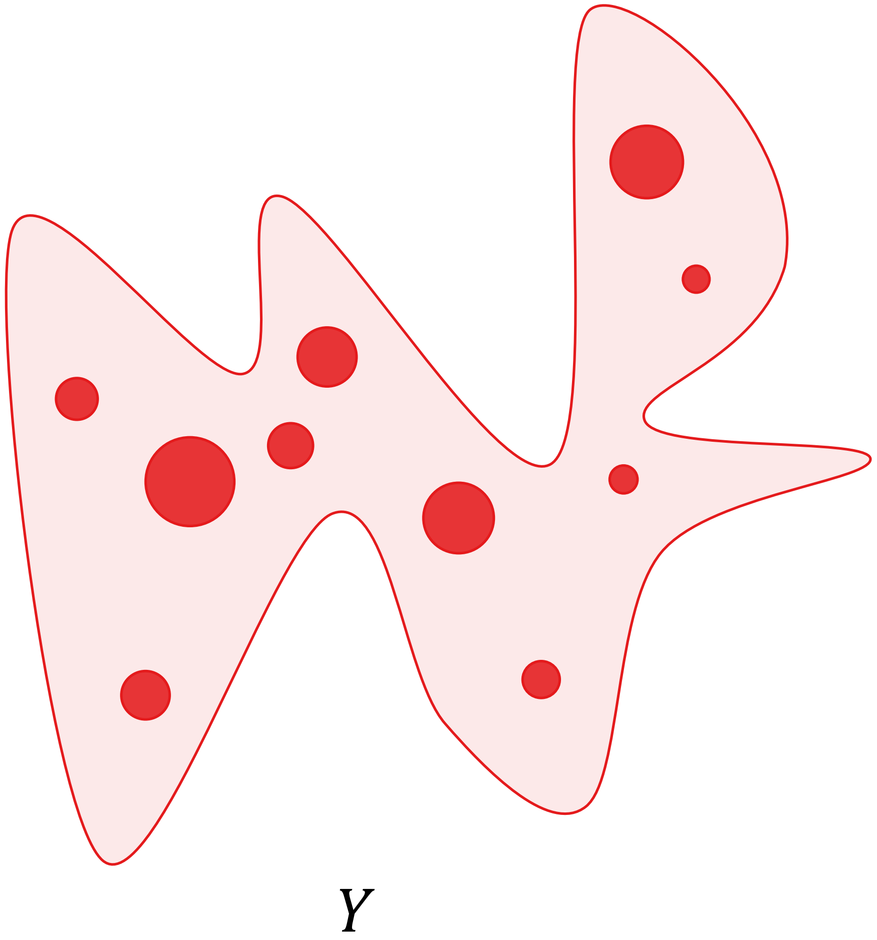
L. Kantorovich
1939

Discrete case →
Linear programming

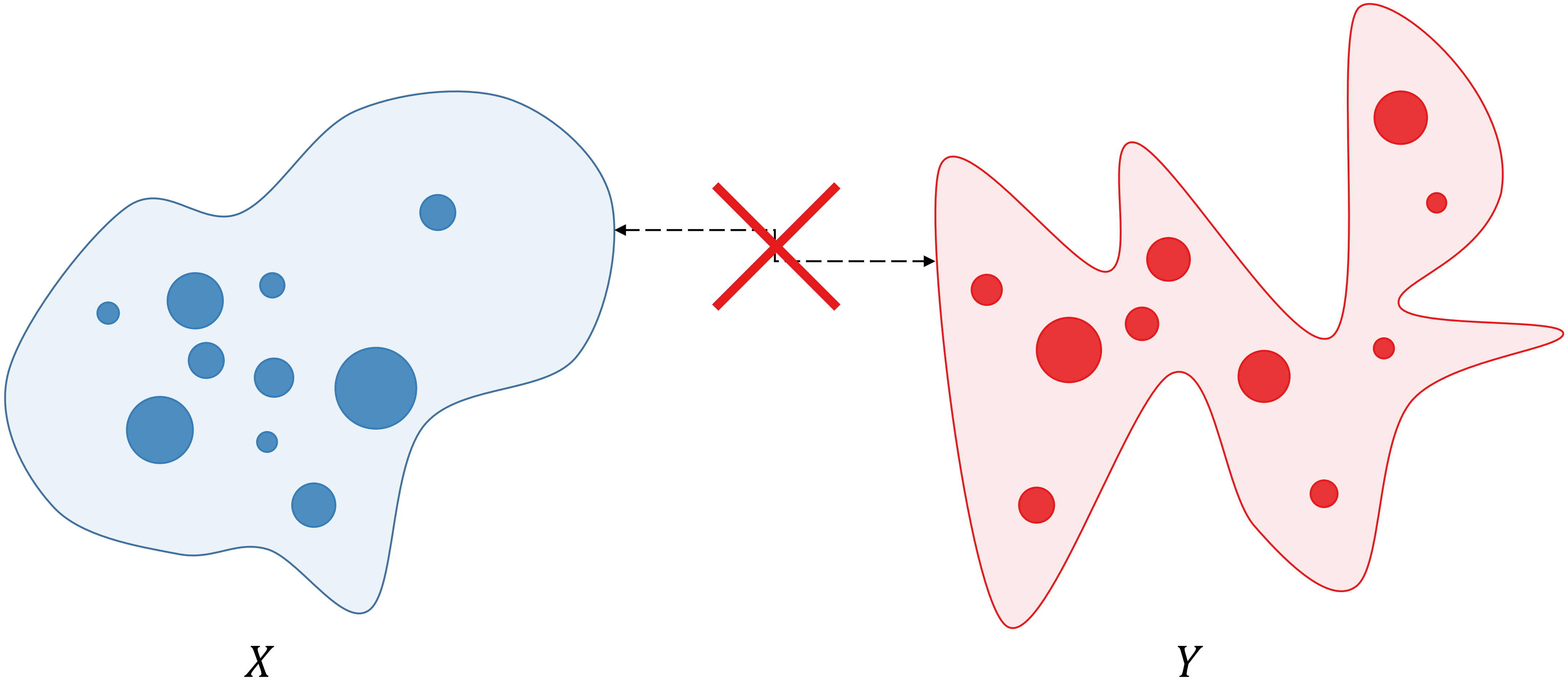


Gromov-Wasserstein distance

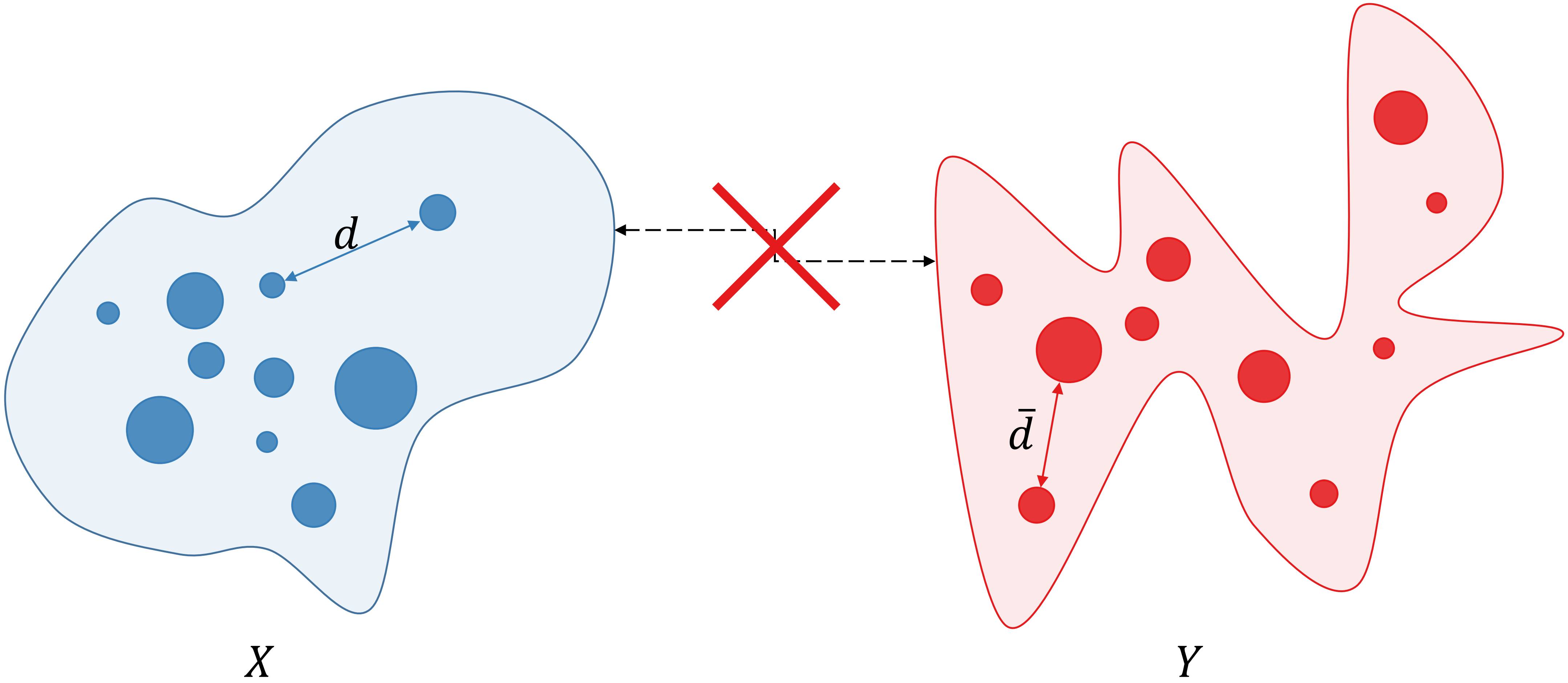




Gromov-Wasserstein distance

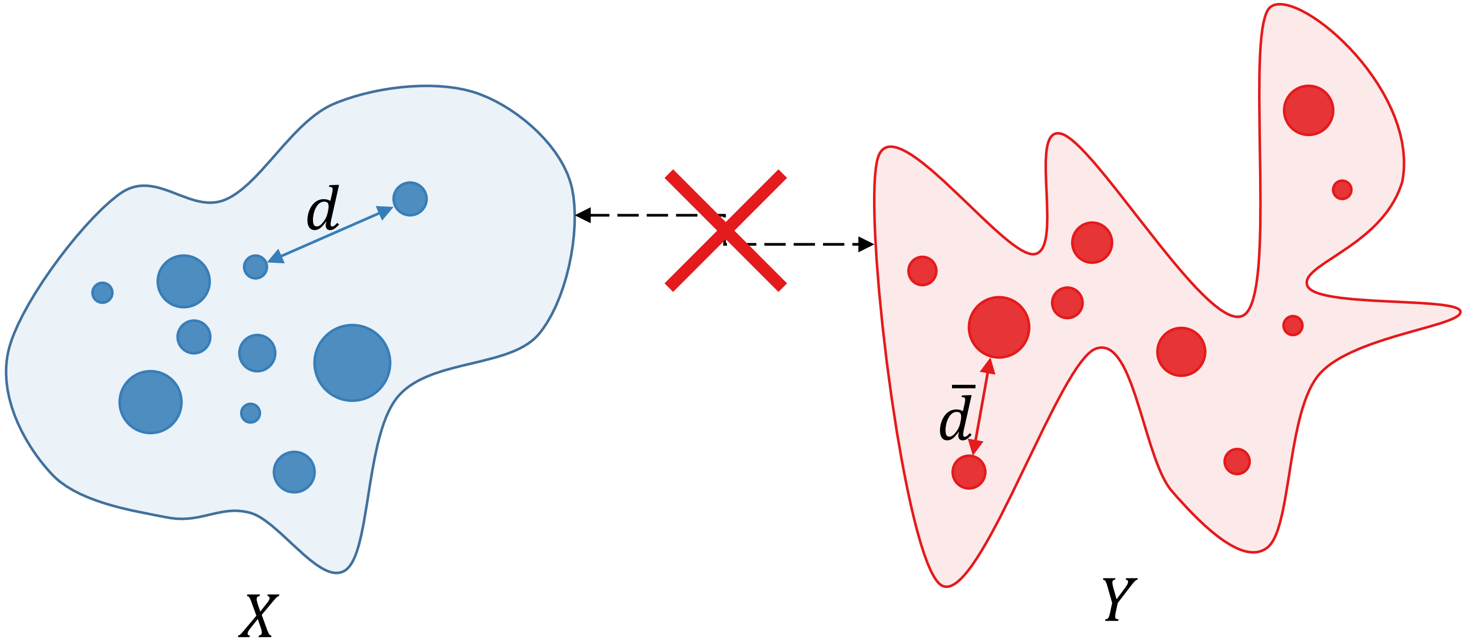


Gromov-Wasserstein distance



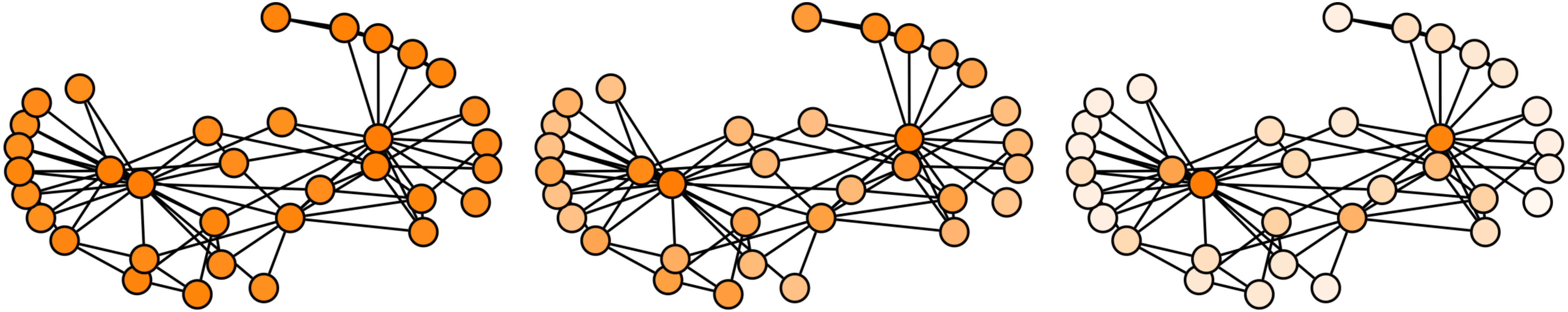
Gromov-Wasserstein distance

$$d_{\mathcal{GW},p}(X,Y) = \frac{1}{2}\left(\inf_{M}\sum_{i,j}\sum_{i',j'}\left|d(x_i,x_{i'}) - \bar{d}(y_j,y_{j'})\right|^p m_{ij}m_{i'j'}\right)^{1/p}$$



Heat diffusion has an explicit notion of scale

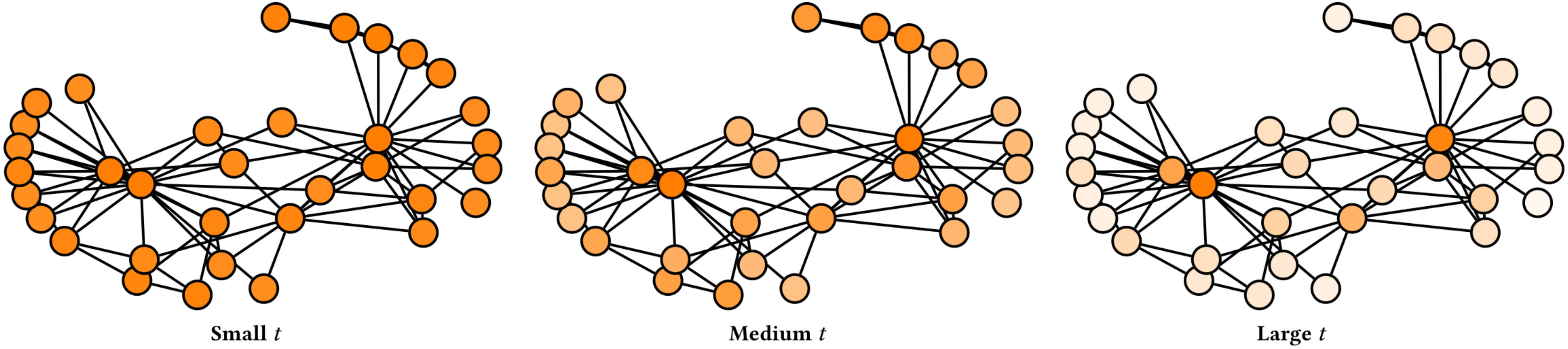
$$\frac{\partial u_t}{\partial t} = -\mathcal{L} u_t$$



Small t **Medium** t **Large** t

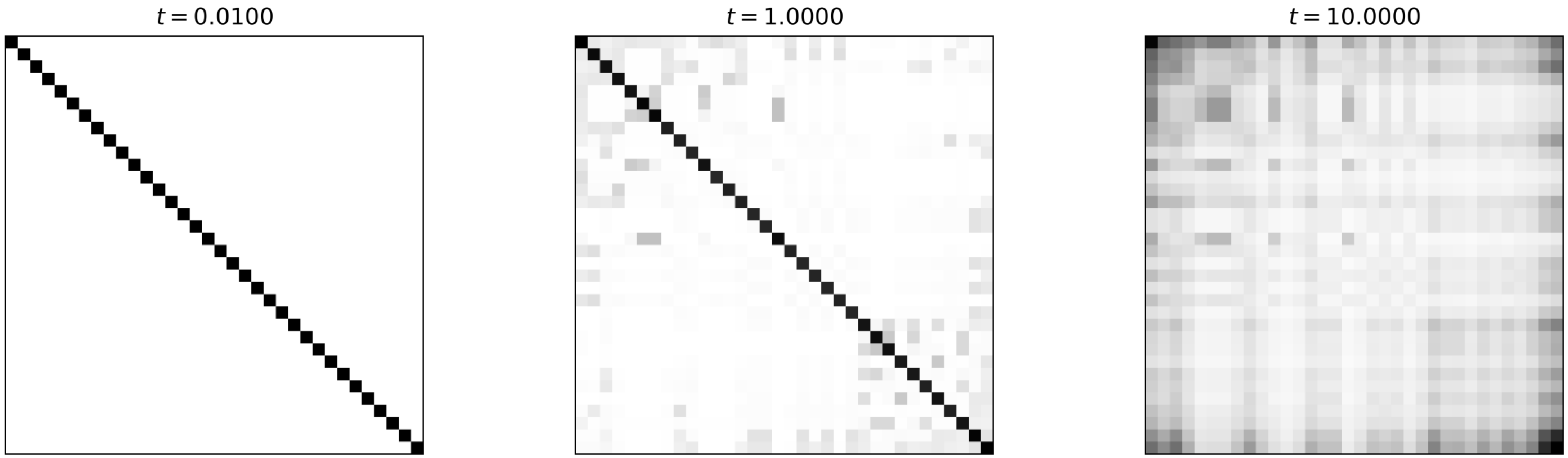
Heat kernel has an explicit notion of scale

$$H_t = e^{-t\mathcal{L}} = \Phi e^{-t\Lambda} \Phi^\top = \sum_{j=1}^{n} e^{-t\lambda_j} \phi_j \, \phi_j^\top$$



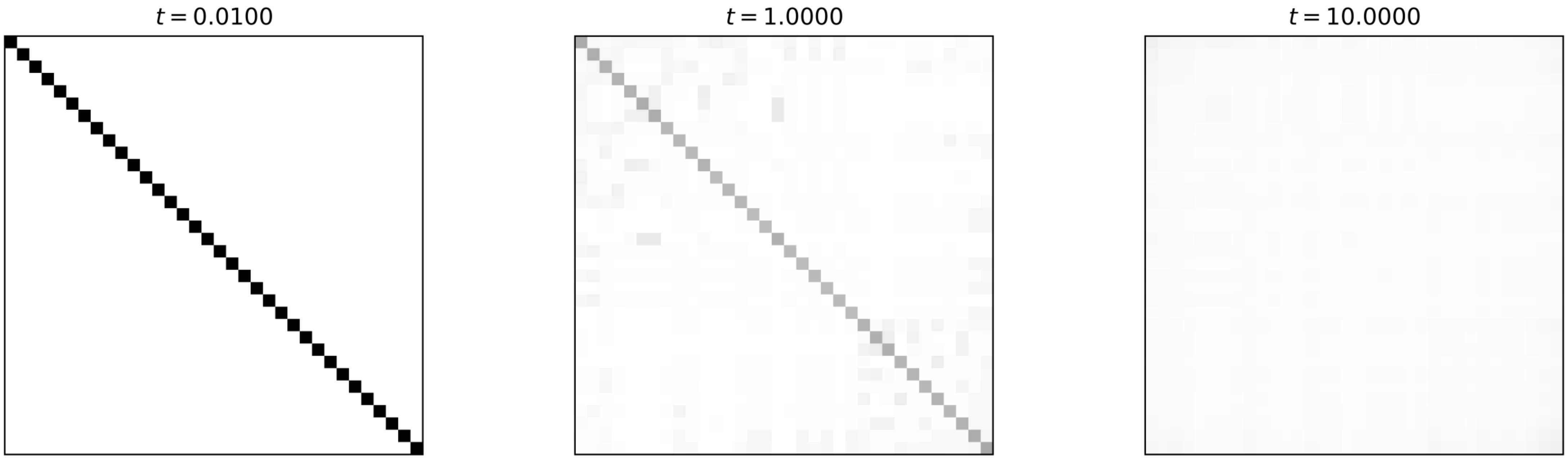
Scale corresponds to locality

$$H_t = e^{-t\mathcal{L}} = \Phi e^{-t\Lambda}\Phi^\top = \sum_{j=1}^{n} e^{-t\lambda_j}\,\phi_j\,\phi_j^\top$$



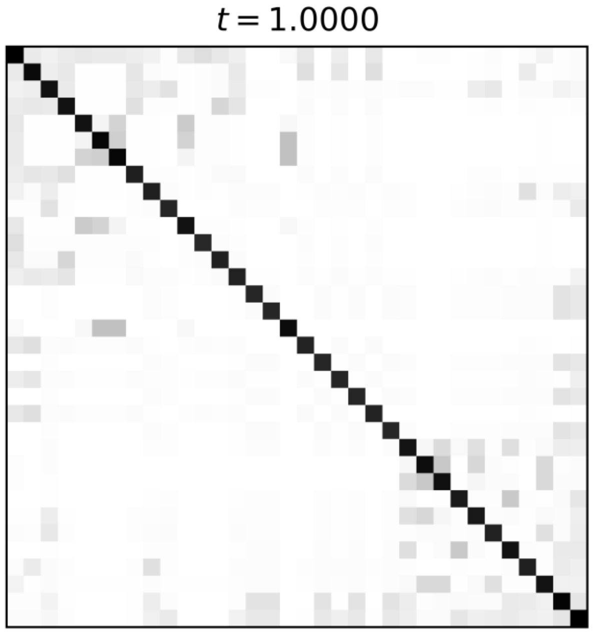
Scale corresponds to locality

$$H_t = e^{-t\mathcal{L}} = \Phi e^{-t\Lambda}\Phi^\top = \sum_{j=1}^{n} e^{-t\lambda_j}\,\phi_j\,\phi_j^\top$$



Spectral Gromov-Wasserstein = Gromov-Wasserstein + heat kernel

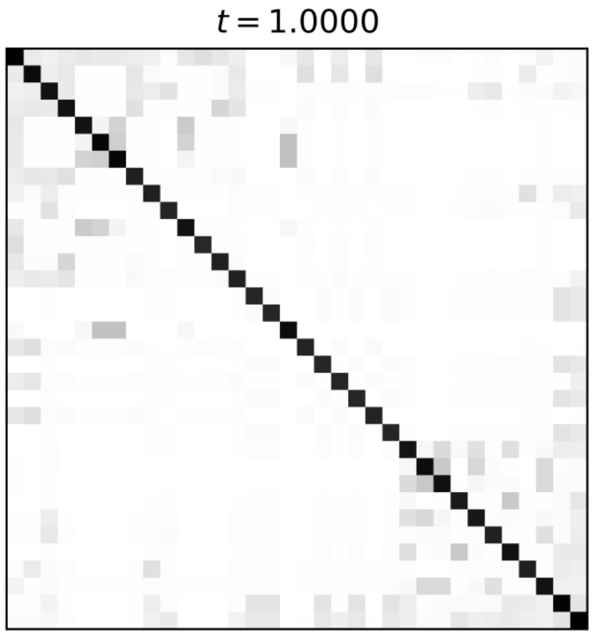
$$d^{\text{spec}}_{\mathcal{GW},p}(X,Y) = \inf_{M} \sup_{t>0} e^{-2\left(t+t^{-1}\right)} \cdot \left(\sum_{i,j} \sum_{i',j'} \left| H^X_t(x_i, x_{i'}) - H^Y_t(y_j, y_{j'}) \right|^p m_{ij} m_{i'j'} \right)^{1/p}$$



Using heat kernel at all t as a distance doesn't make our task any easier

Spectral Gromov-Wasserstein has a useful lower bound!

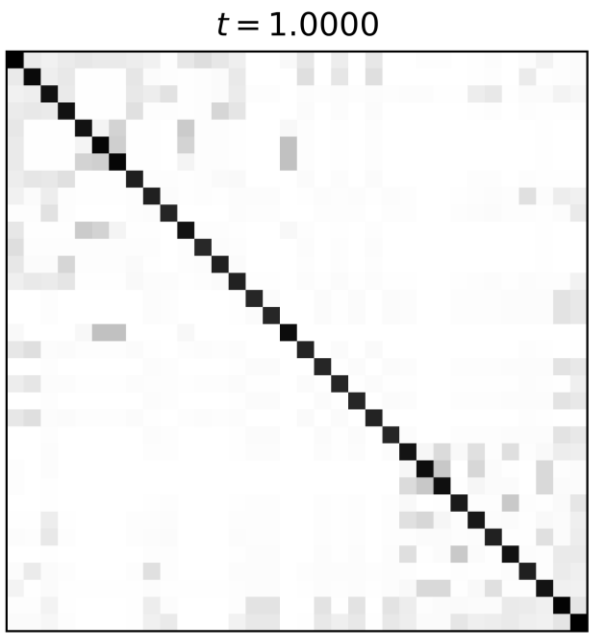
$$d^{\text{spec}}_{\mathcal{GW},p}(X,Y) = \inf_M \sup_{t>0} e^{-2\left(t+t^{-1}\right)} \cdot \left(\sum_{i,j} \sum_{i',j'} \left| H_t^X(x_i, x_{i'}) - H_t^Y(y_j, y_{j'}) \right|^p m_{ij} m_{i'j'} \right)^{1/p} \geq \sup_{t>0} e^{-2\left(t+t^{-1}\right)} \cdot \left| \operatorname{tr}(H^X) - \operatorname{tr}(H^Y) \right|$$



Using heat kernel at all t as a distance **does** make our task **way** easier!

Spectral Gromov-Wasserstein has a useful lower bound!

$$d_{\mathcal{GW},p}^{\text{spec}}(X,Y) = \inf_{M} \sup_{t>0} e^{-2\left(t+t^{-1}\right)} \cdot \left(\sum_{i,j} \sum_{i',j'} \left| H_t^X(x_i, x_{i'}) - H_t^Y(y_j, y_{j'}) \right|^p m_{ij} m_{i'j'} \right)^{1/p} \geq \sup_{t>0} e^{-2\left(t+t^{-1}\right)} \cdot \left| \operatorname{tr}(H^X) - \operatorname{tr}(H^Y) \right|$$



Using heat kernel at all t as a distance **does** make our task **way** easier!

We can just compare heat traces!

Network Laplacian Spectral Descriptors

$$h_t = \mathrm{tr}(H_t) = \sum_j e^{-t\lambda_j}$$

We sample t logarithmically, and compare h_t with L_2 distance
However, h_t is size-dependent!

Size invariance = normalization

$$h_t = \mathrm{tr}(H_t) = \sum_j e^{-t\lambda_j}$$

We can normalize by h_t of the complete (K) or empty graph $\overline{K}$
Computation of all λ is still expensive: $O(n^3)$

Scalability

We propose two options:

1. Use local Taylor expansion: $h_t = \mathrm{tr}(e^{-t\mathcal{L}}) = \sum_{k=0}^{\infty} \frac{\mathrm{tr}((-t\mathcal{L})^k)}{k!} \approx n - t\ \mathrm{tr}(\mathcal{L}) + \frac{t^2}{2}\,\mathrm{tr}(\mathcal{L}^2) + \ldots$

   Second term is degree distribution; third is weighted triangle count

Scalability

We propose two options:

1. Use local Taylor expansion: $h_t = \mathrm{tr}(e^{-t\mathcal{L}}) = \sum_{k=0}^{\infty} \frac{\mathrm{tr}((-t\mathcal{L})^k)}{k!} \approx n - t\ \mathrm{tr}(\mathcal{L}) + \frac{t^2}{2}\,\mathrm{tr}(\mathcal{L}^2) + \dots$

   Second term is degree distribution; third is weighted triangle count

2. Compute top + bottom eigenvalues, approximate the rest

   Linear extrapolation = explicit assumption on the manifold (Weyl's law)

Scalability

We propose two options:

1. Use local Taylor expansion: $h_t = \operatorname{tr}(e^{-t\mathcal{L}}) = \sum_{k=0}^{\infty} \frac{\operatorname{tr}((-t\mathcal{L})^k)}{k!} \approx n - t\ \operatorname{tr}(\mathcal{L}) + \frac{t^2}{2} \operatorname{tr}(\mathcal{L}^2) + \ldots$

   Second term is degree distribution; third is weighted triangle count

2. Compute top + bottom eigenvalues, approximate the rest

   Linear extrapolation = explicit assumption on the manifold (Weyl's law)

Other spectrum approximators can be even more efficient!

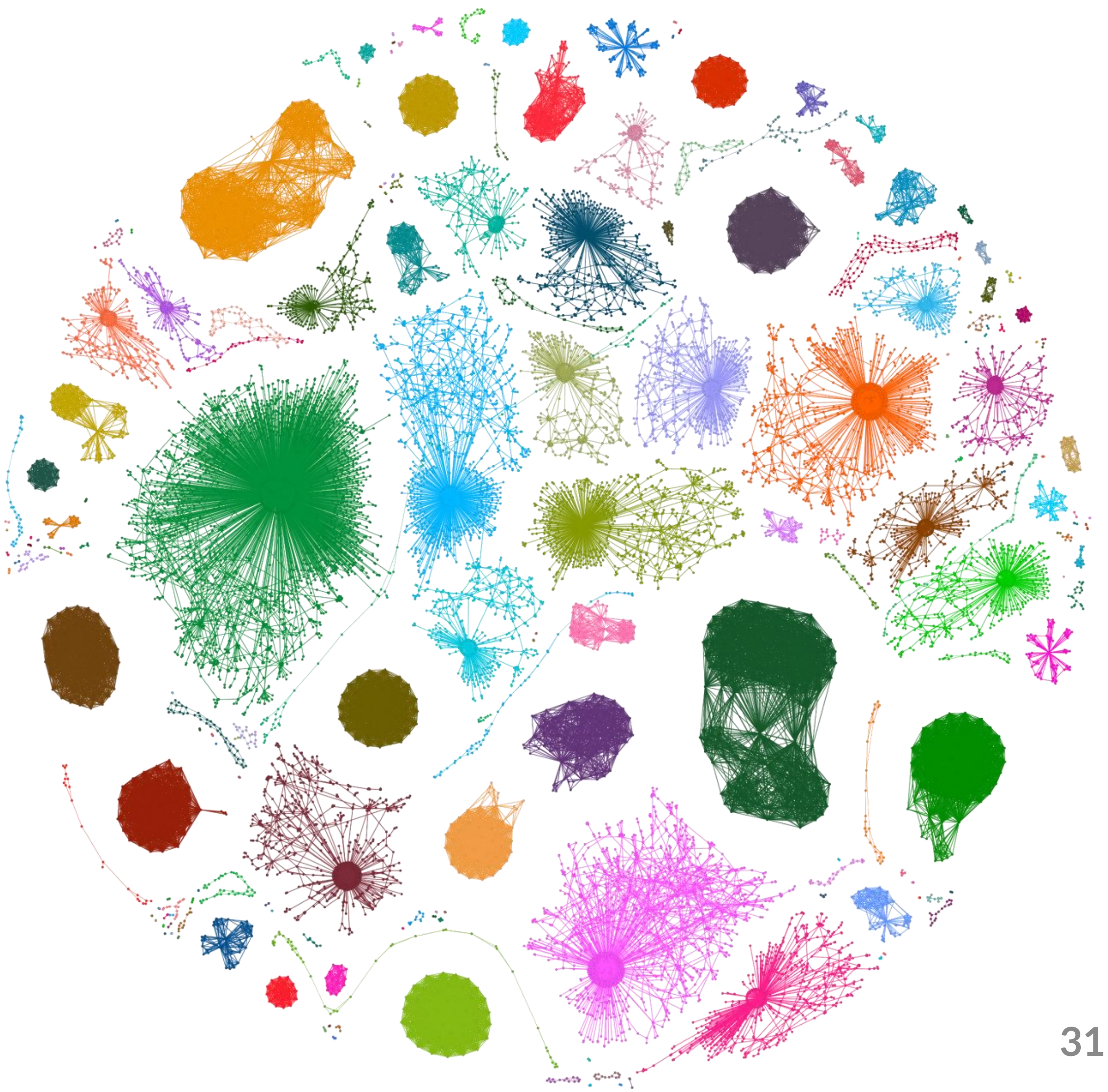
[Cohen-Steiner et al. | KDD 2018]

[Adams et al. | arXiv 1802.03451]

Experimental design

3 key properties:

- Permutation invariance
- Scale-adaptivity
- Size invariance



Detecting graphs with communities

3 key properties:

- Permutation invariance
- Scale-adaptivity
- Size invariance

	Method	$n \sim \mathcal{P}(\lambda)$ 64	128	256	512	1024
NetLSD	$h(G)$	54.39	59.01	60.82	57.99	53.80
	$h(G)/h(\bar{K})$	54.53	62.27	70.83	76.45	78.40
	$w(G)$	56.23	63.77	69.57	71.66	70.34
	$w(G)/w(\bar{K})$	55.51	63.85	72.12	77.59	79.39
NIPS'17	FGSD	55.44	54.99	53.86	52.74	50.92
ASONAM'13	NetSimile	59.55	56.57	59.41	66.23	60.58

Accuracy of classification of SBM vs Erdős–Rényi graphs

Detecting rewired graphs

3 key properties:

- Permutation invariance
- Scale-adaptivity
- Size invariance

	Method	MUTAG	PROTEINS	NCI1	ENZYMES	COLLAB	IMDB-B
		dataset					
NetLSD	$h(G)$	76.03	91.81	69.74	92.51	59.82	67.18
NetLSD	$h(G)/h(\bar{K})$	79.12	94.90	74.55	95.20	65.85	70.58
NetLSD	$w(G)$	78.18	93.04	70.54	94.03	69.01	75.26
NetLSD	$w(G)/w(\bar{K})$	79.72	89.00	74.14	90.77	70.35	75.54
NIPS'17	FGSD	77.79	60.11	64.08	53.93	55.18	56.23
ASONAM'13	NETSIMILE	77.11	85.73	58.58	87.38	54.43	54.44

Accuracy of classification of real vs rewired graphs

Classifying real graphs

3 key properties:

- Permutation invariance
- Scale-adaptivity
- Size invariance

	Method	MUTAG	PROTEINS	NCI1	ENZYMES	COLLAB	IMDB-B
		dataset					
NetLSD	$h(G)$	86.47	64.89	66.49	31.99	68.00	68.04
NetLSD	$h(G)/h(\bar{K})$	85.32	65.73	67.44	33.31	69.42	70.17
NetLSD	$w(G)$	83.35	66.80	70.78	40.41	75.77	68.63
NetLSD	$w(G)/w(\bar{K})$	81.72	65.58	67.67	35.78	77.24	69.33
NIPS'17	FGSD	84.90	65.30	75.77	41.58	73.96	69.54
ASONAM'13	NETSIMILE	84.09	62.45	66.56	33.23	73.10	69.20

Accuracy of graph classification

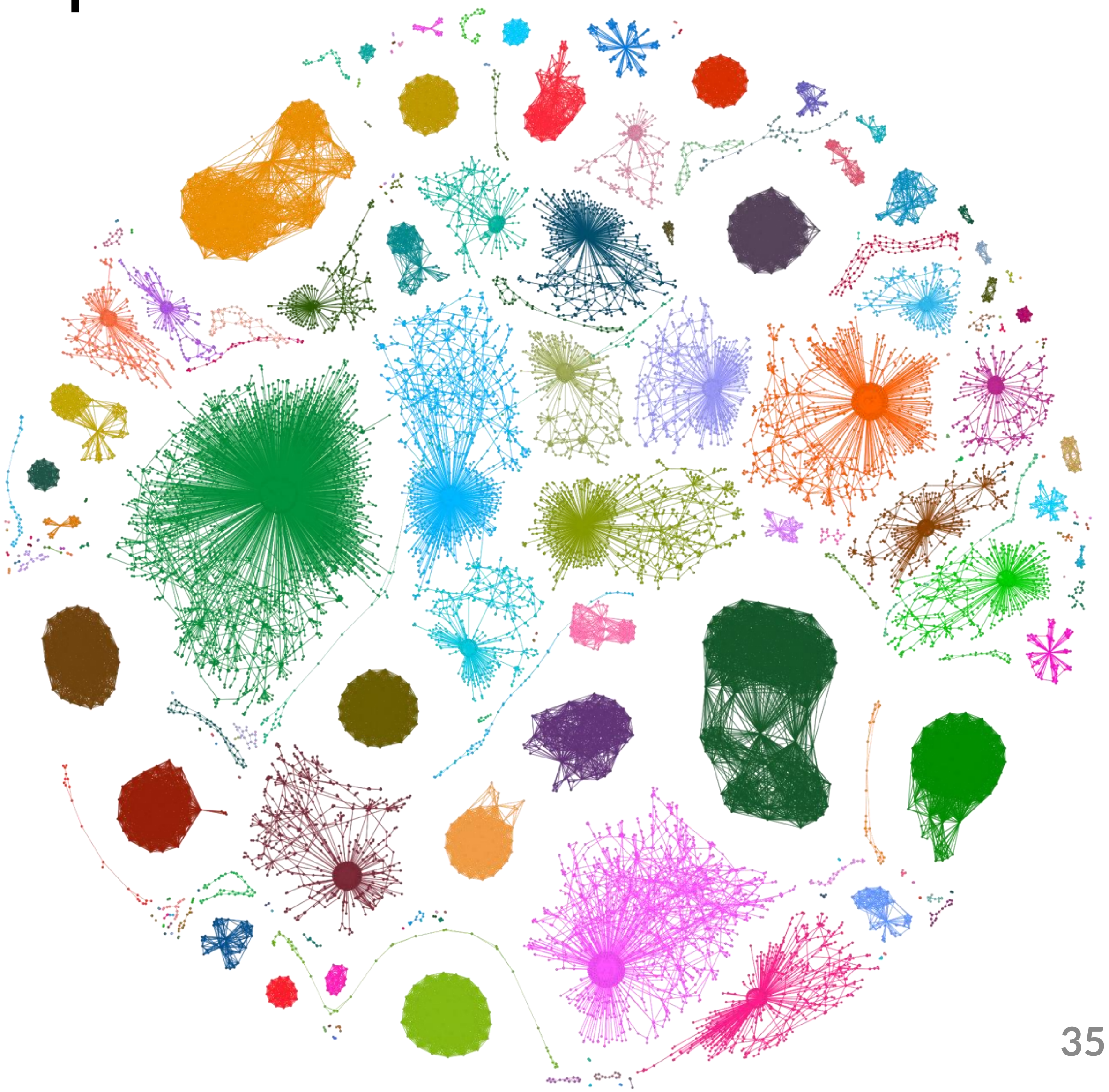
Expressive graph comparison

3 key properties:

- Permutation invariance
- Scale-adaptivity
- Size invariance

+ Scalability

= **NetLSD**



Questions?

code github.com/xgfs/netlsd
website tsitsul.in/publications/netlsd
shy? anton@tsitsul.in

Network Laplacian Spectral Descriptors: wave kernel trace

$$w_t = \mathrm{tr}(W_t) = \sum_j e^{-it\lambda_j}$$

We sample t logarithmically, and compare $\mathrm{Re}(w_t)$ with L_2 distance
w_t detects symmetries!
≈quantum random walks

Hearing the Shape of a Graph

"Can One Hear the Shape of a Drum?" – Kac 1966

No, as there are co-spectral drums (graphs)

Conjecture: # of co-spectral graphs $\rightarrow 0$ as # of nodes $\rightarrow \infty$

[Dufree, Martin 2015]