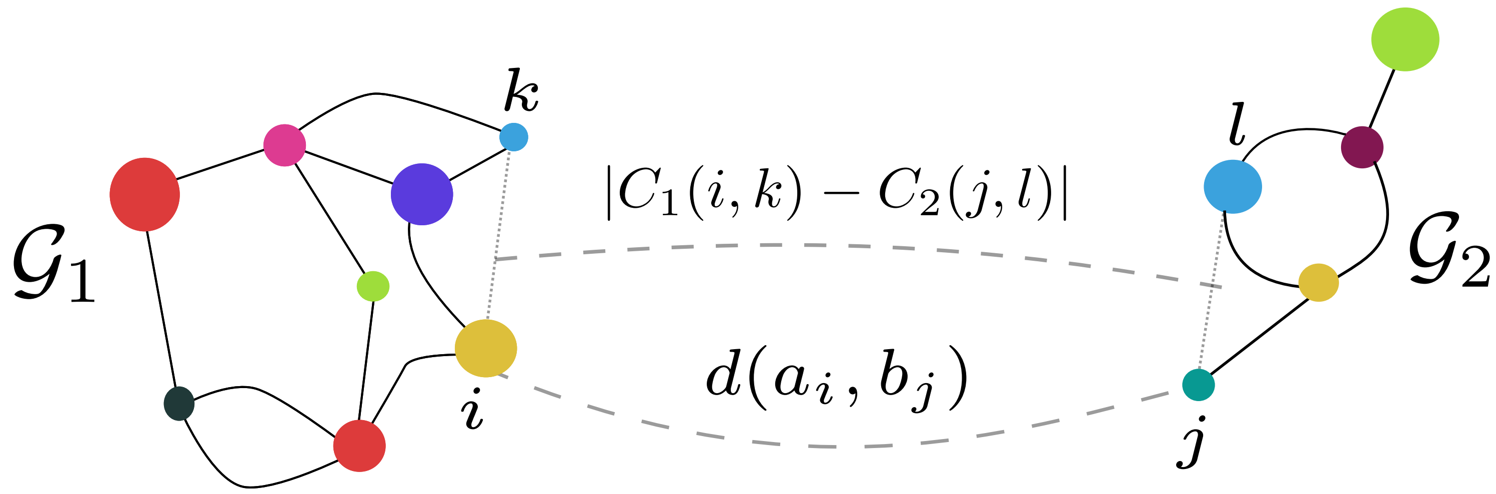
Optimal Transport for structured data with application on graphs



Titouan Vayer

Joint work with Laetitia Chapel, Remi Flamary, Romain Tavenard and Nicolas Courty

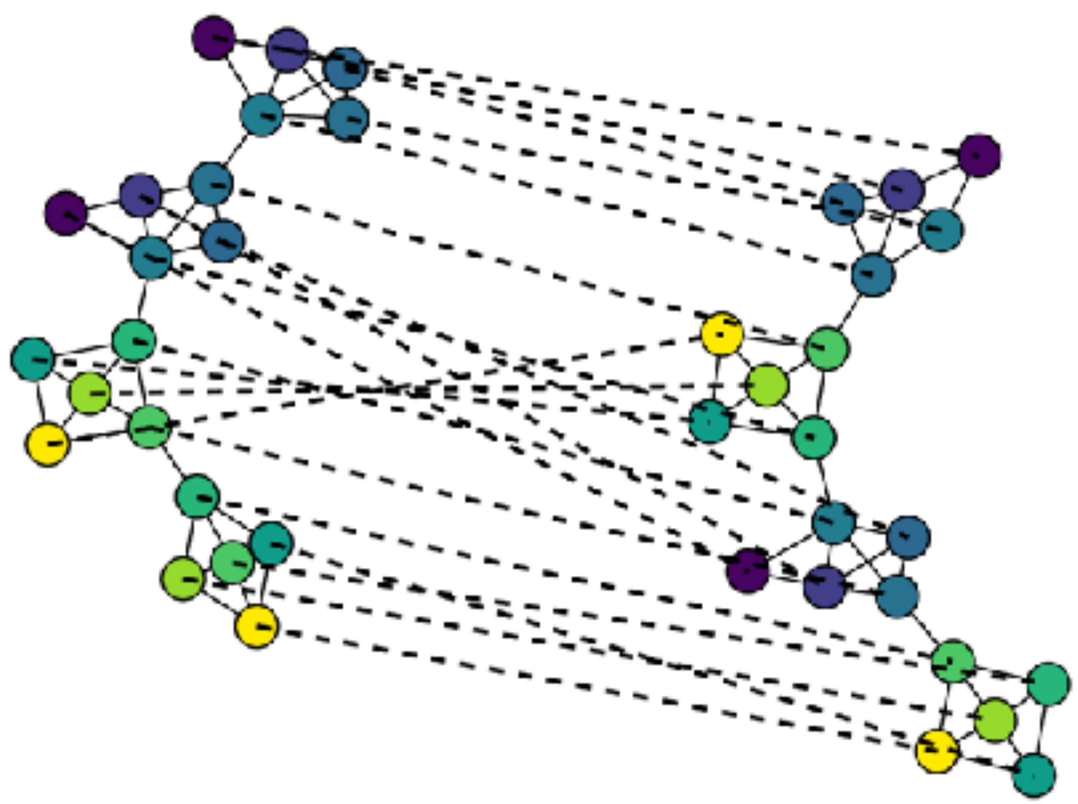
A novel distance between labeled graphs based on optimal transport

Contributions:

- **Differentiable distance between labeled graphs. Jointly considers the features and the structures**

Contributions:

- **Differentiable distance between labeled graphs. Jointly considers the features and the structures**

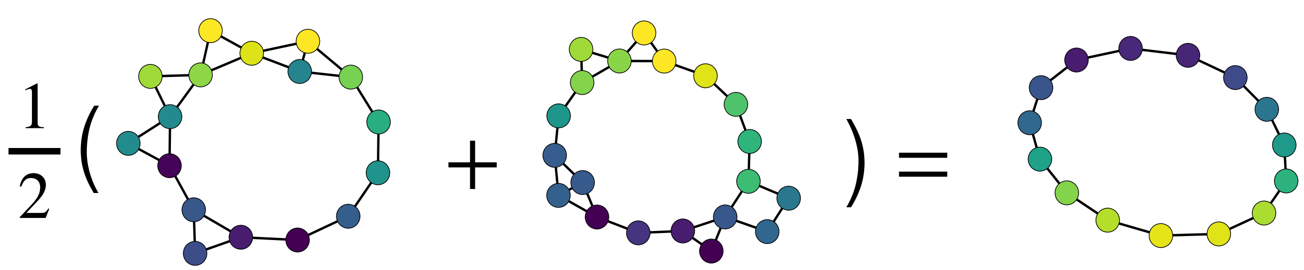


Distance = 1.41

Optimal transport: soft assignment between the nodes

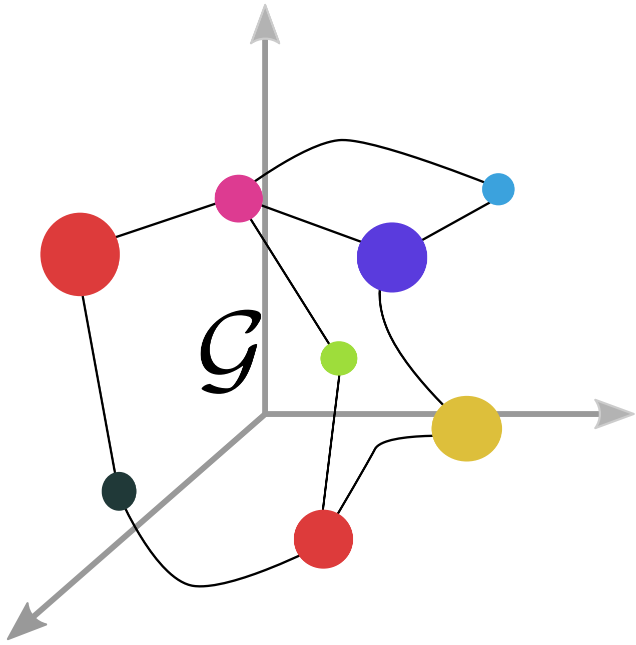
Contributions:

- **Differentiable distance between labeled graphs. Jointly considers the features and the structures**



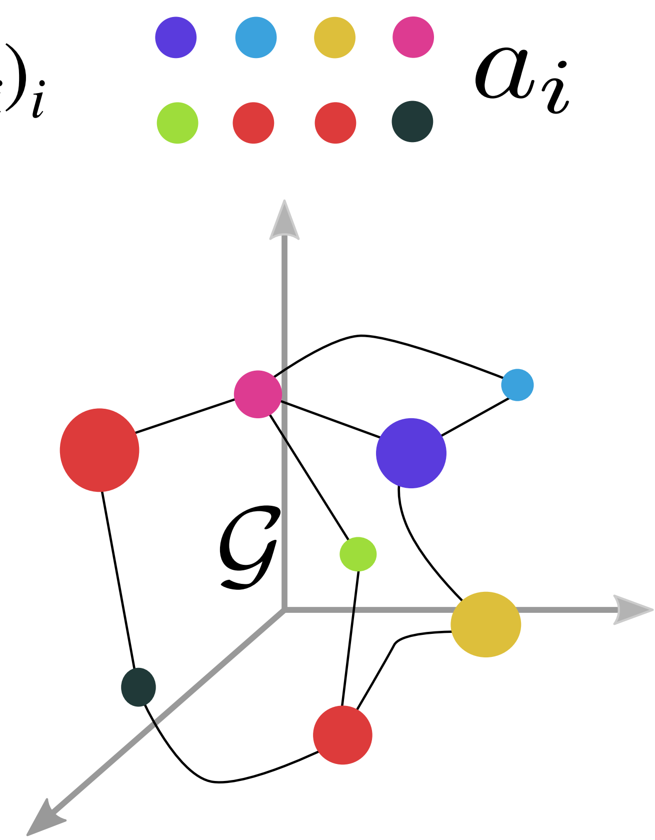
Computing **average** of labeled graphs

Structured data as probability distribution

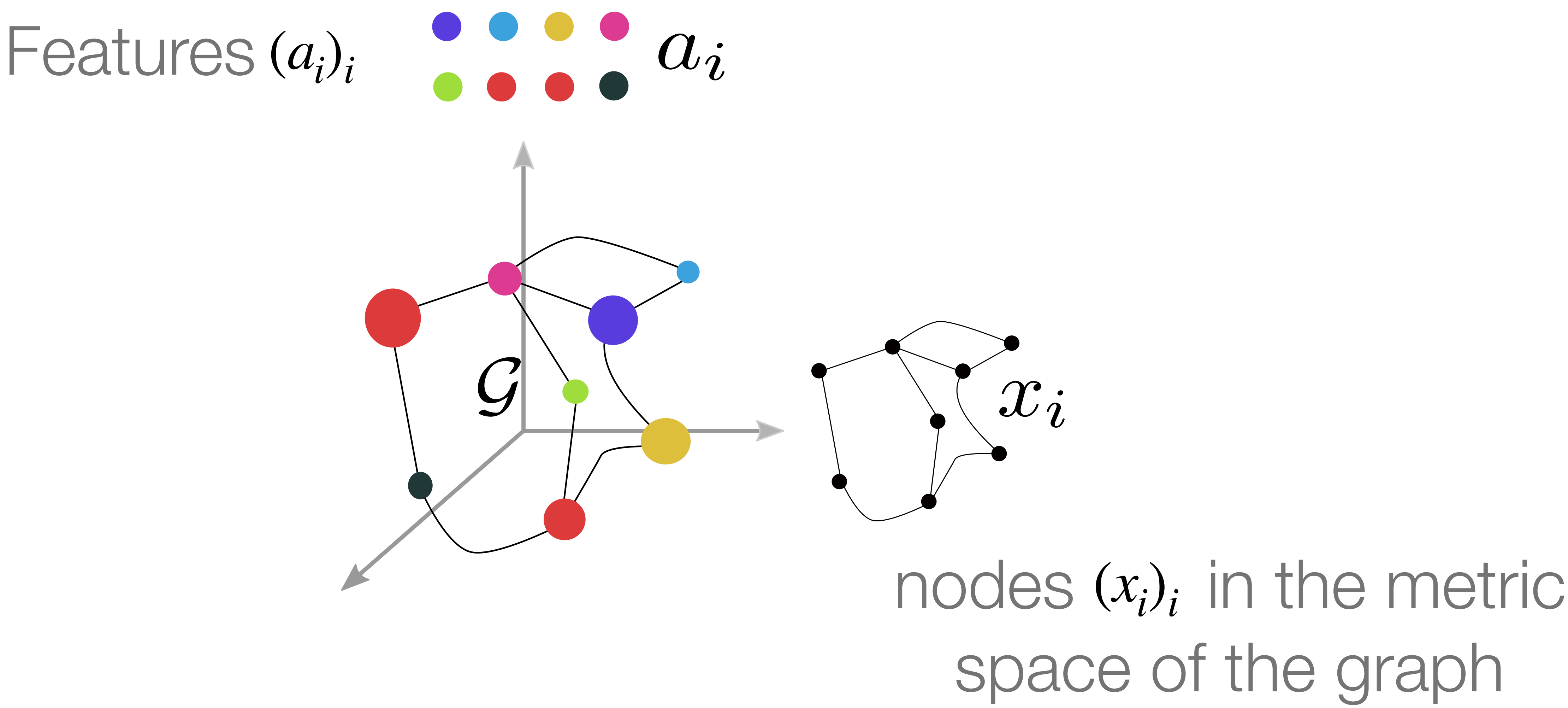


Structured data as probability distribution

Features $(a_i)_i$

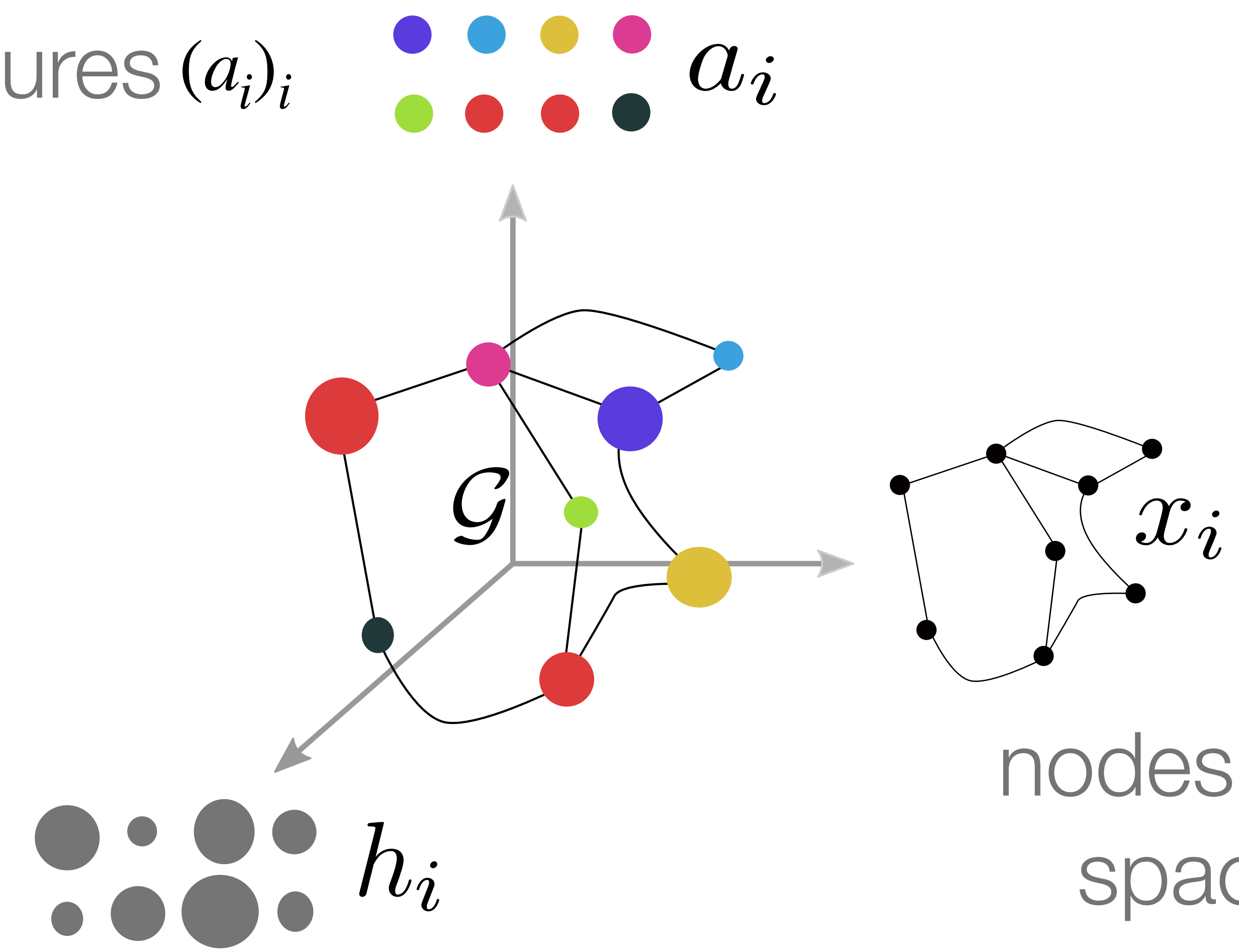


Structured data as probability distribution



Structured data as probability distribution

Features $(a_i)_i$



nodes $(x_i)_i$ in the metric space of the graph

weighted by their masses $(h_i)_i$

Optimal transport in a nutshell

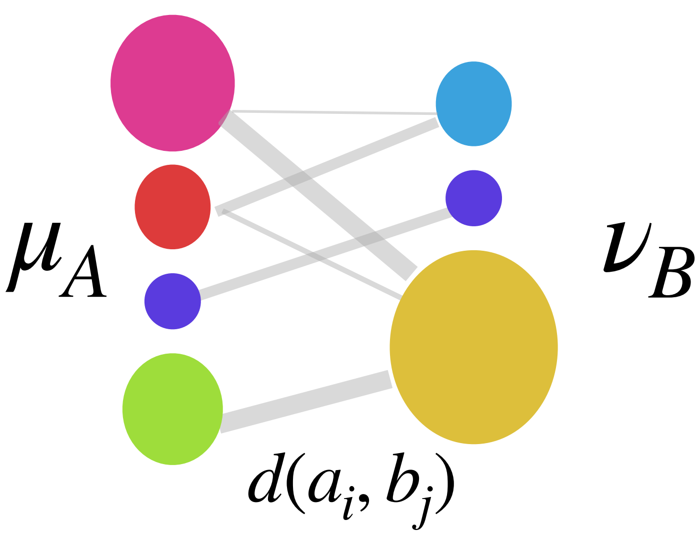
Compare two probability distributions by transporting one onto another

Wasserstein distance

Gromov-Wasserstein distance

Optimal transport in a nutshell

Compare two probability distributions by transporting one onto another

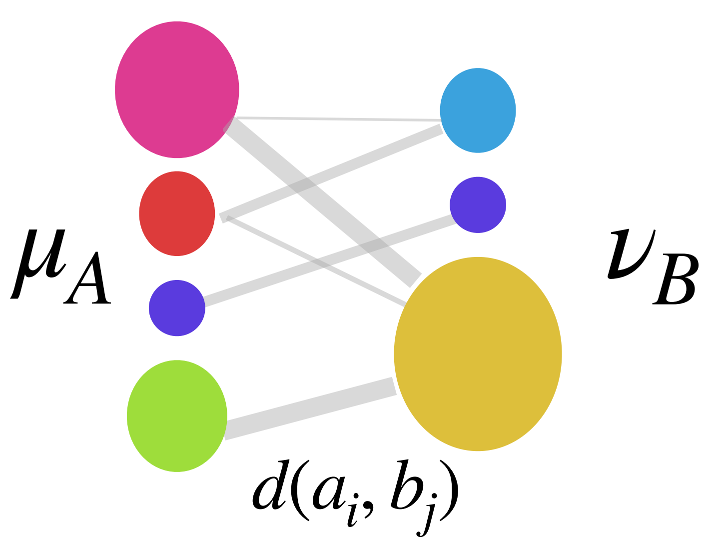


Wasserstein distance

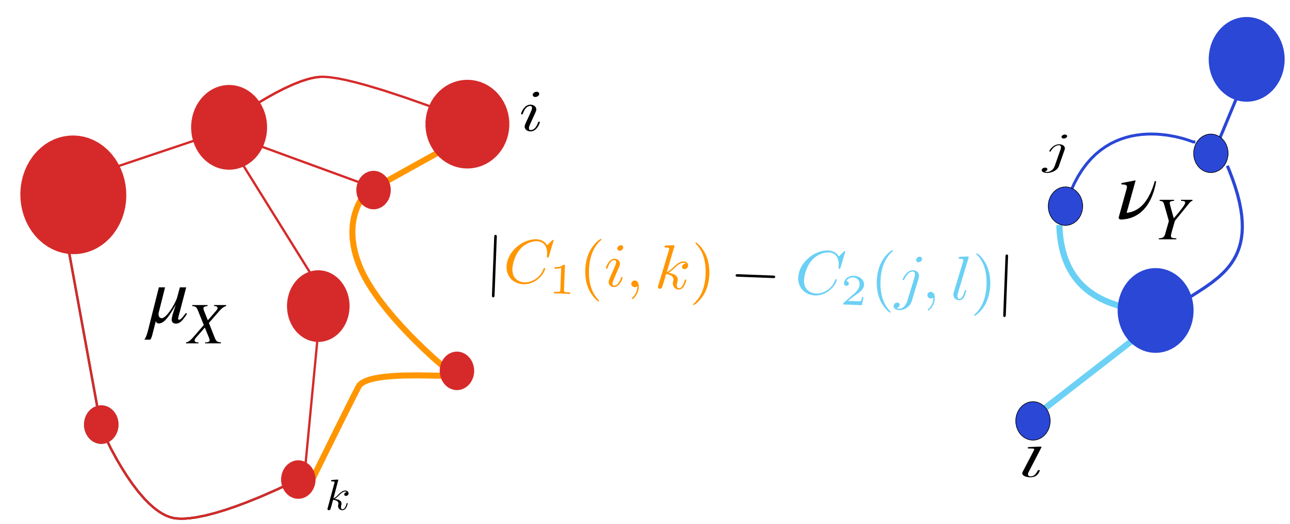
Gromov-Wasserstein distance

Optimal transport in a nutshell

Compare two probability distributions by transporting one onto another

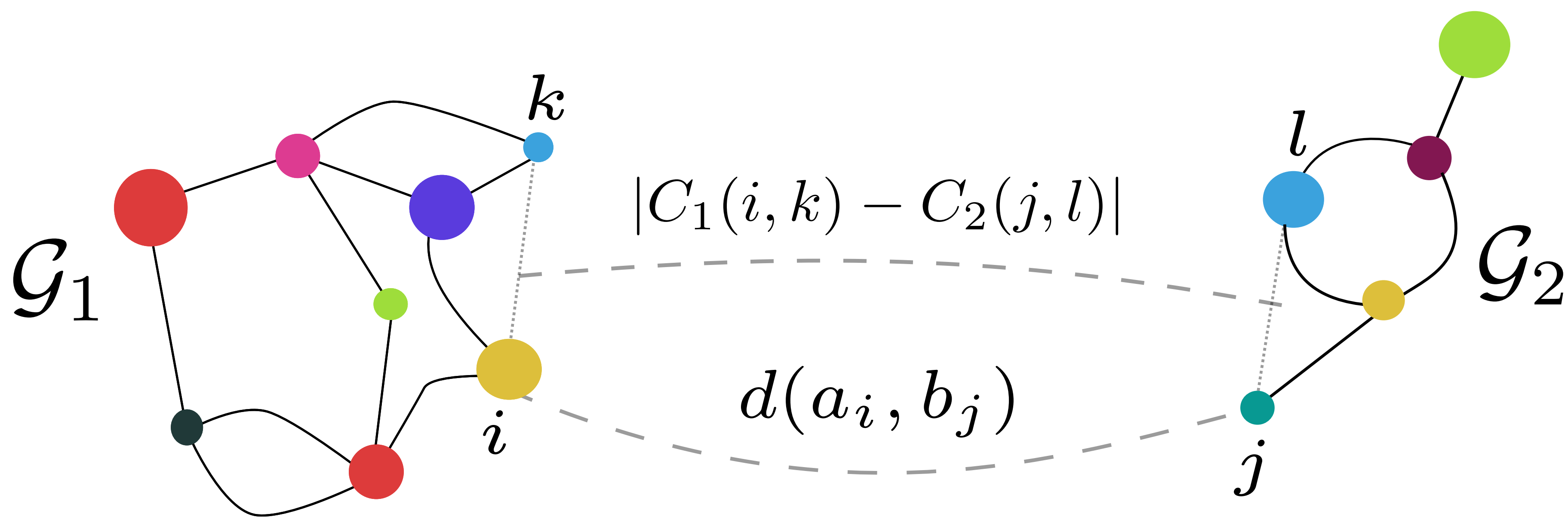


Wasserstein distance



Gromov-Wasserstein distance

Fused Gromov-Wasserstein distance

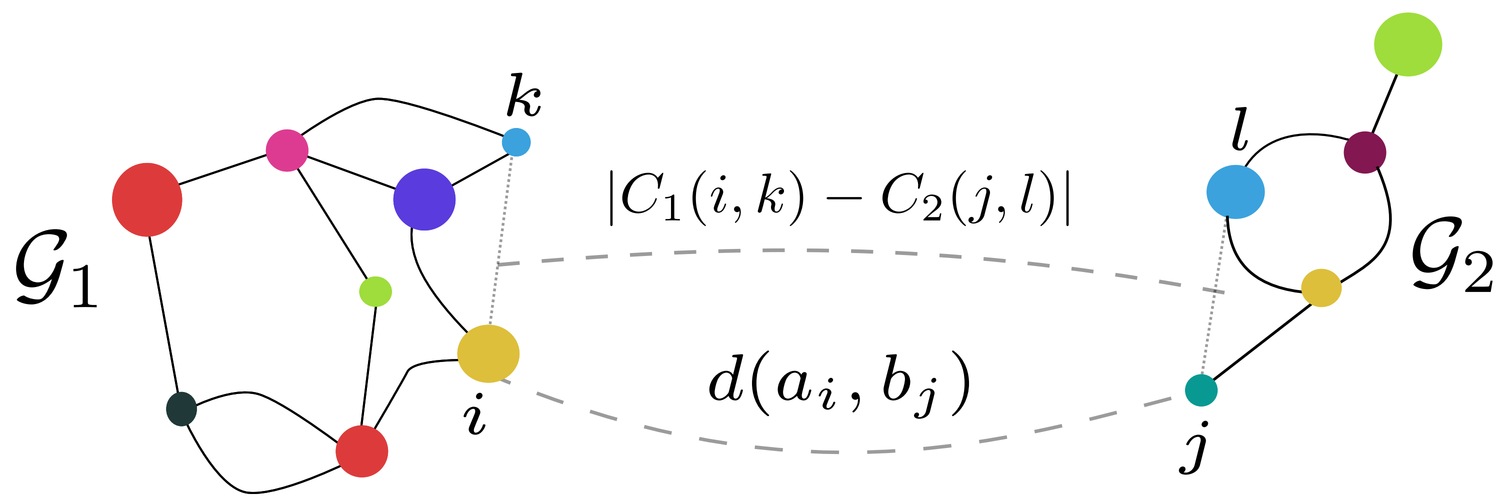


$$FGW_{q,\alpha}(\mu,\nu) = \min_{\pi \in \Pi(\mu,\nu)} \sum_{i,j,k,l} \left((1-\alpha) d(a_i, b_j)^q + \alpha \, |\, C_1(i,k) - C_2(j,l) \,|^q \right) \pi_{i,j} \pi_{k,l}$$

where $\boldsymbol{\pi}$ is the soft assignment matrix

$\boldsymbol{\alpha}$ is a trade-off features/structures

Fused Gromov-Wasserstein distance



Properties

- **Interpolate** between **Wasserstein** distance on features and **Gromov-Wasserstein** distance on the structures
- **Distance on labeled graph**: vanishes iff graphs have same labels and weights at the same place up to a permutation

Optimization problem

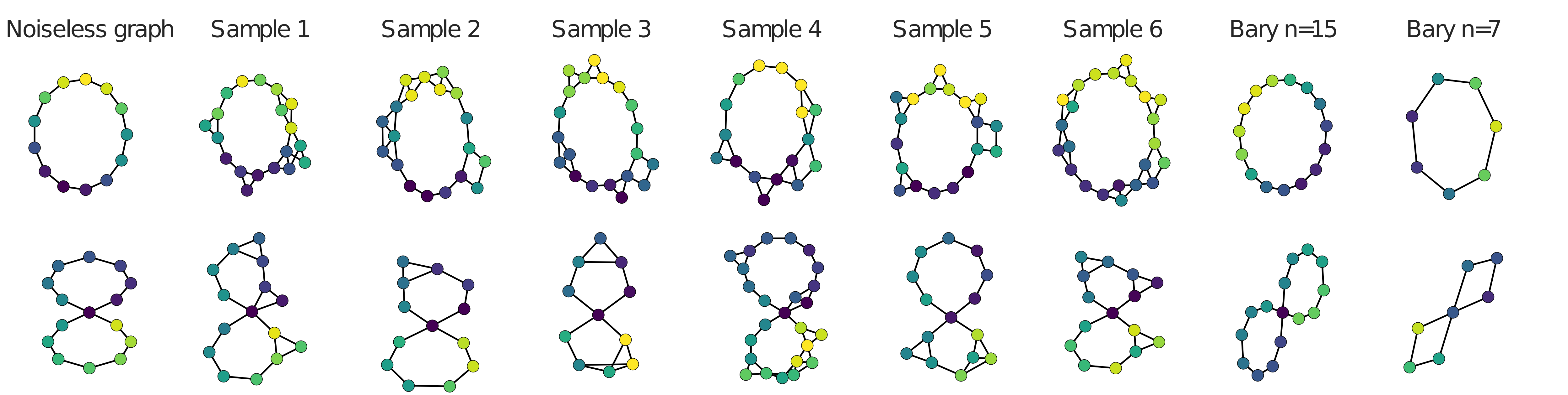
- Non convex Quadratic Program: hard !
- Conditional Gradient Descent (aka Frank Wolfe)
- Suitable for entropic regularization + Sinkhorn iteraterations

Applications

Classification

Dataset	Labeled Graphs MUTAG	PTC	NCI1	Social Graphs IMDB-B	Vector attributes Graph SYNTHETIC	PROTEIN	CUNEIFORM
WL	86.21±8.15	62.17±7.80	85.13±1.61	Unapplicable(U)	U	U	U
GK	82.42±8.40	56.46±8.03	60.78±2.48	56.00±3.61	41.13±4.68	U	U
RW	79.47±8.17	55.09±7.34	58.63±2.44	U	U	U	U
SP	85.79±2.51	58.53±2.55	73.00±0.51	55.80±2.93	38.93±5.12	U	U
HOPPER	U	U	U	U	90.67±4.67	71.96±3.22	32.59±8.73
PROPA	U	U	U	U	64.67±6.70	61.34±4.38	12.59± 6.67
PSCN $k = 10$	83.47±10.26	58.34±7.71	70.65±2.58	U	**100.00±0.00**	67.95±11.28	25.19±7.73
FGW	**88.42±5.67**	**65.31±7.90**	**86.42±1.63**	**63.80±3.49**	**100.00±0.00**	**74.55±2.74**	**76.67±7.04**

Graph Barycenter + k-means clustering of graphs



Check out our poster at Pacific Ballroom #133!!