Gromov-Wasserstein distances

Second Approach:

- GW: Generalizes OT to the non-registered case

- Main idea: compare relations instead of absolute positions

- GW defines a proper distance! [Mémoli, 2011]

- Non-convex - yet solved efficiently [Solomon et al., 2016]

- For very large problems: first solve reduced problem, then fit orthogonal mapping

$$\mathrm{GW}(\mathbf{C}, \mathbf{C}', \mathbf{p}, \mathbf{q}) = \min_{\Gamma \in \Pi(\mathbf{a}, \mathbf{b})} \sum_{i,j,k,l} \mathcal{L}(\mathbf{C}_{ik}, \mathbf{C}'_{jl}) \Gamma_{ij} \Gamma_{kl}$$

Second Approach:
Gromov-Wasserstein distances

- GW: Generalizes OT to the non-registered case
- Main idea: compare relations instead of absolute positions

$$\mathrm{GW}(\mathbf{C}, \mathbf{C}', \mathbf{p}, \mathbf{q}) = \min_{\Gamma \in \Pi(\mathbf{a}, \mathbf{b})} \sum_{i,j,k,l} \mathcal{L}(\mathbf{C}_{ik}, \mathbf{C}'_{jl}) \Gamma_{ij} \Gamma_{kl}$$

- GW defines a proper distance! [Mémoli, 2011]
- Non-convex - yet solved efficiently [Solomon et al., 2016]
- For very large problems: first solve reduced problem, then fit orthogonal mapping

Invariant OT

Application: Unsupervised Word Translation

- Have **monolingual** word embeddings, want to find translations