Data Analytics and Machine Learning Group
Department of Informatics
Technical University of Munich

Scalable Optimal Transport in High Dimensions for Graph Distances, Embedding Alignment, and More

Johannes Klicpera, Marten Lienen, Stephan Günnemann

www.daml.in.tum.de/lcn

Massively scalable optimal transport: Sparse and LCN-Sinkhorn

Entropy-regularized optimal transport

$$\bar{\boldsymbol{P}} = \arg\min_{\boldsymbol{P}} \langle \boldsymbol{P}, \boldsymbol{C} \rangle_{\mathrm{F}} - \lambda H(\boldsymbol{P})$$

Solvable via Sinkhorn algorithm (alternating matrix normalization):

$$\boldsymbol{K} = \mathrm{e}^{-\boldsymbol{C}/\lambda}$$

$$\boldsymbol{s}^{(i)} = \boldsymbol{p} \oslash (\boldsymbol{K}\boldsymbol{t}^{(i-1)}), \qquad \boldsymbol{t}^{(i)} = \boldsymbol{q} \oslash (\boldsymbol{K}^T \boldsymbol{s}^{(i)})$$

$$\bar{\boldsymbol{P}} = \mathrm{diag}(\bar{\boldsymbol{s}})\, \boldsymbol{K}\, \mathrm{diag}(\bar{\boldsymbol{t}})$$

Parallelizable, runs in $\mathcal{O}(n^2)$ → $\boldsymbol{K}$ already has that many entries!

Sparse Sinkhorn

$\boldsymbol{K}$ exponential in $\boldsymbol{C}$ → only near neighbors important!

Sparsely approximate $\boldsymbol{K}$:

$$K^{\mathrm{sp}}_{ij} = \begin{cases} K_{ij} & \text{if } x_i, x_j \text{ are near,} \\ 0 & \text{otherwise} \end{cases}$$

Calculate P via Sinkhorn algorithm: $\bar{\boldsymbol{P}} \approx \bar{\boldsymbol{P}}^{\mathrm{sp}} = \mathrm{diag}(\bar{\boldsymbol{s}})\, \boldsymbol{K}^{\mathrm{sp}}\, \mathrm{diag}(\bar{\boldsymbol{t}})$
Runs in $\mathcal{O}(n \log n)$!

Locally corrected Nyström (LCN)

How to incorporate *short- and long-range* interactions?

Correct low-rank Nyström approximation with sparse values:

$$\boldsymbol{K}_{\mathrm{Nys}} = \boldsymbol{U}\boldsymbol{A}^{-1}\boldsymbol{V}$$

$$\boldsymbol{K}_{\mathrm{LCN}} = \boldsymbol{K}_{\mathrm{Nys}} - \boldsymbol{K}^{\mathrm{sp}}_{\mathrm{Nys}} + \boldsymbol{K}^{\mathrm{sp}}$$

LCN is a general kernel approximation.
We can still use the Sinkhorn algorithm → LCN-Sinkhorn, runs in $\mathcal{O}(n \log n + nl^2)$

GTN: GNN with optimal transport -48% error for graph distance learning

Graph Transport Network (GTN)

Graph neural network (GNN) generates node embeddings
Embeddings matched using Sinkhorn

→ Learns graph distances via backpropagation

Varying numbers of nodes → **learnable unbalanced OT**
Bipartite matching (BP) matrix for optimal transport
(Scaled) norms as deletion cost

$$\boldsymbol{C}_{BP} = \begin{bmatrix} \boldsymbol{C} & \boldsymbol{C}^{(\mathbf{p},\varepsilon)} \\ \boldsymbol{C}^{(\varepsilon,\mathbf{q})} & \boldsymbol{C}^{(\varepsilon,\varepsilon)} \end{bmatrix}, \quad \boldsymbol{C}^{(\mathbf{p},\varepsilon)}_{ij} = \begin{cases} c_{i,\varepsilon} & i = j \\ 0 & i \neq j \end{cases}, \quad \boldsymbol{C}^{(\varepsilon,\mathbf{q})}_{ij} = \begin{cases} c_{\varepsilon,j} & i = j \\ 0 & i \neq j \end{cases}, \quad \boldsymbol{C}^{(\varepsilon,\varepsilon)}_{ij} = 0$$

Multi-head OT: OT with multiple embeddings & varying regularization λ

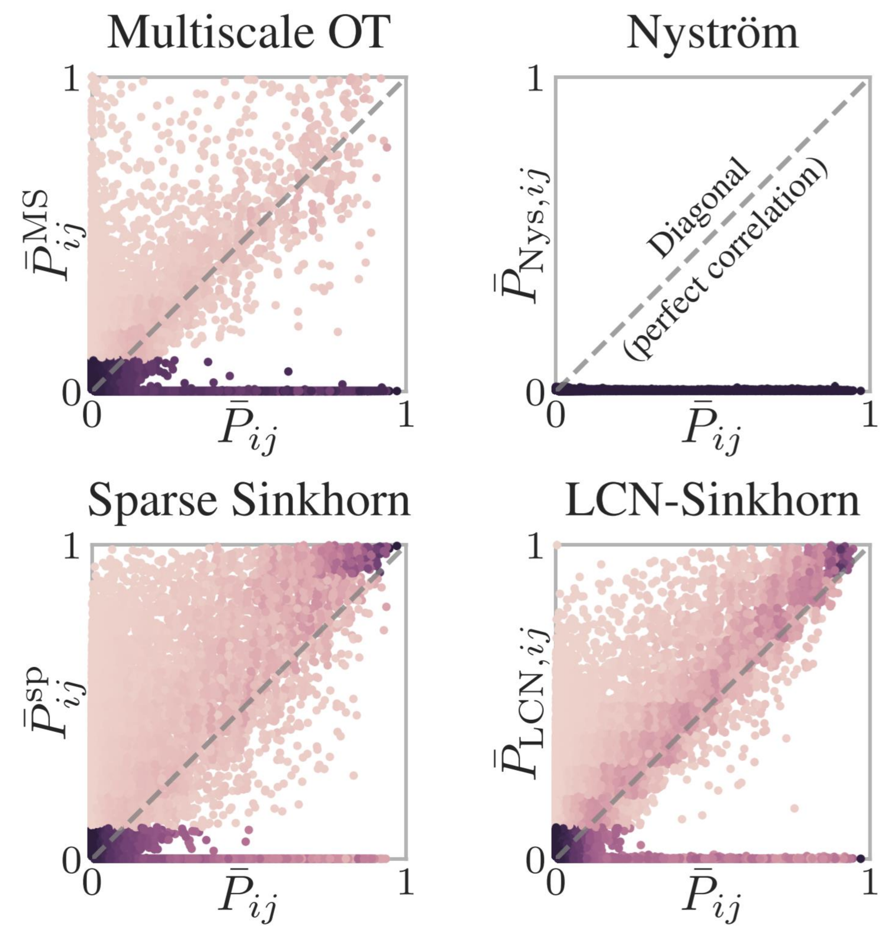
Model	Linux	AIDS30	Pref. att.
SiamMPNN	0.09	13.8	12.1
SimGNN	0.039	4.5	8.3
GMN	0.015	10.3	7.8
GTN, 1 head	0.022	3.7	4.5
8 OT heads	**0.012**	**3.2**	**3.5**
Balanced OT	0.034	15.3	27.4

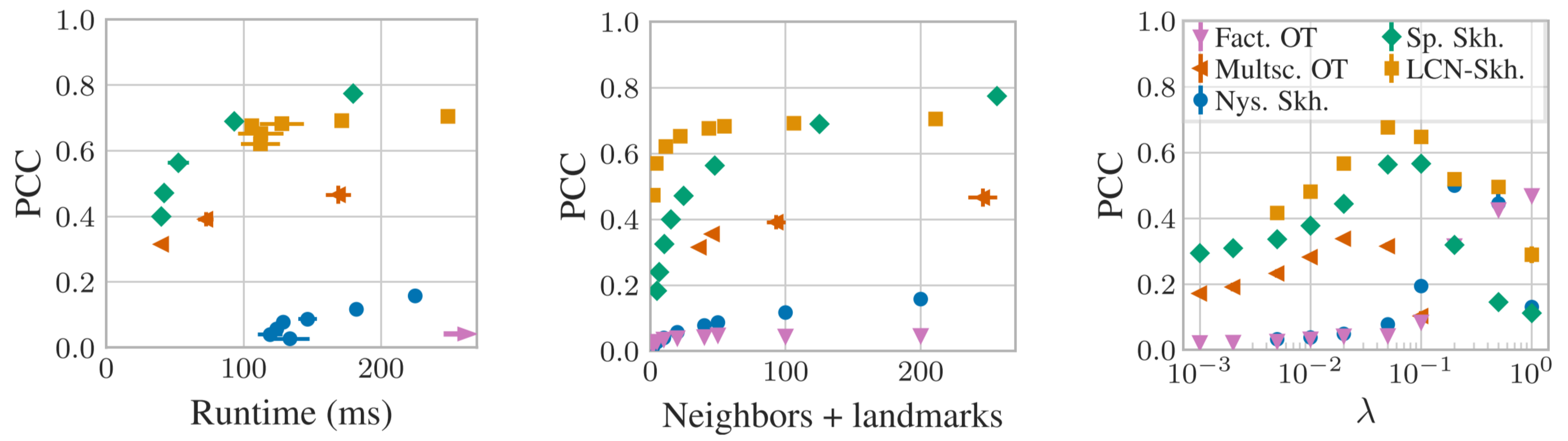
Method	AIDS 30	Pref. att.	Pref. att. 200
Full Sinkhorn	**3.7**	4.5	1.3
Nyström Skh.	**3.6**	6.2	2.4
Multiscale OT	11.2	27.4	6.7
Sparse Skh.	44	40.7	7.6
LCN-Sinkhorn	4.0	**5.1**	**1.4**

LCN: Provably better approximation, same convergence

- Sparse corrections provably reduce kernel approximation error of Nyström for uniform and clustered data [Theorem 1&2].
- Better kernel approximation directly translates to better Sinkhorn approximation [Theorem 3].
- Sparse and LCN-Sinkhorn converge in $\mathcal{O}(\ln \min \boldsymbol{K} / \varepsilon)$, like full Sinkhorn [Theorem 4].
- Sparse and LCN-Sinkhorn allow fast backpropagation via analytical gradients [Prop. 1].

Fast & accurate transport plan and embedding alignment





Embedding alignment

3x faster & 3.1pp more accurate

Method	Time (s)	Avg. accuracy
Original	268	67.9%
Full Sinkhorn	402	70.1%
Multiscale OT	88.2	26.8%
Nyström Sinkhorn	102	47.8%
Sparse Sinkhorn	**49.2**	68.4%
LCN-Sinkhorn	86.8	**71.0%**