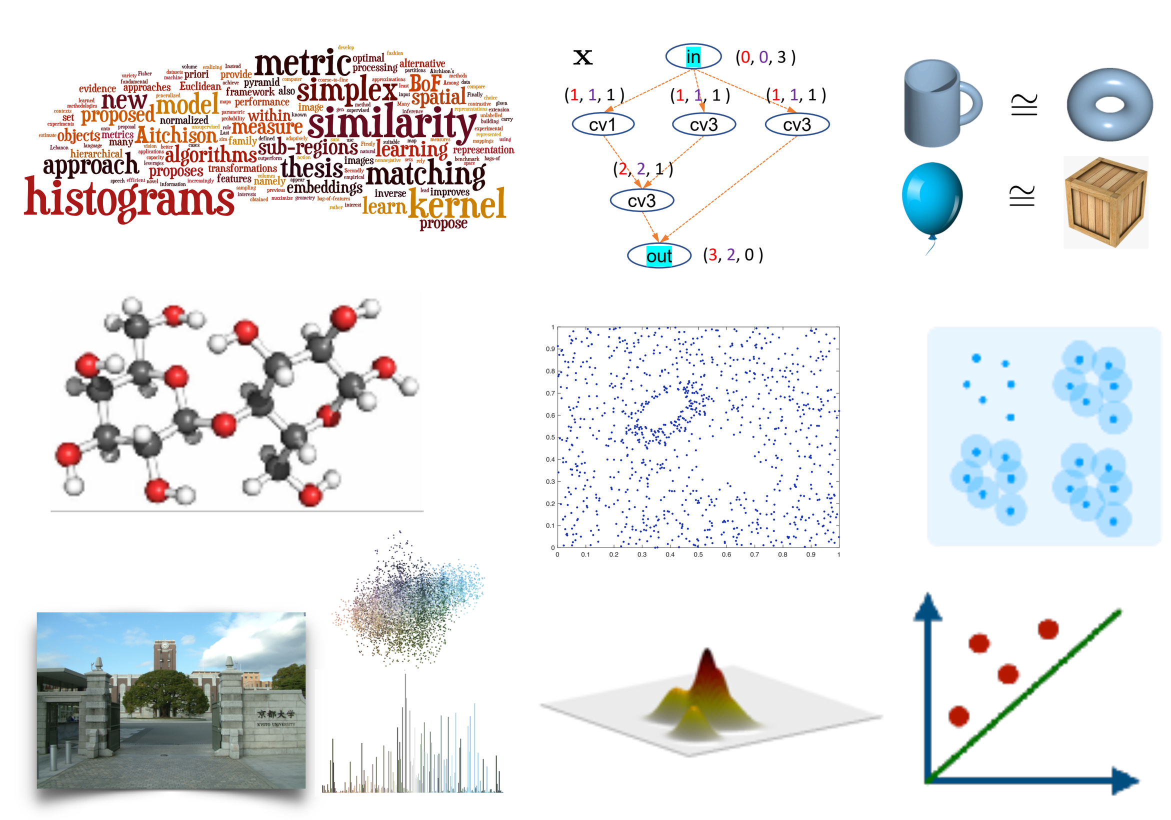
Optimal Transport for Measures on a Graph

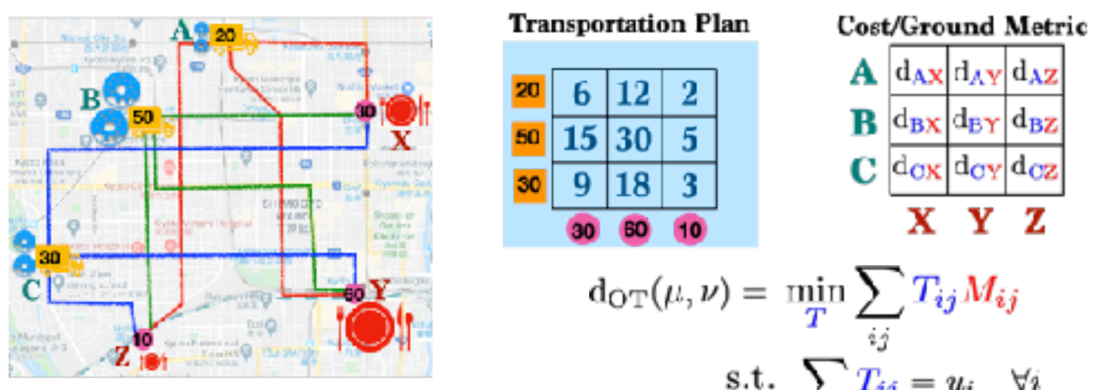
Tam Le - Assistant Professor
tam@ism.ac.jp

Optimal Transport: A Natural Geometry for Probability Distributions

✦ Several objects of interest are represented as distributions



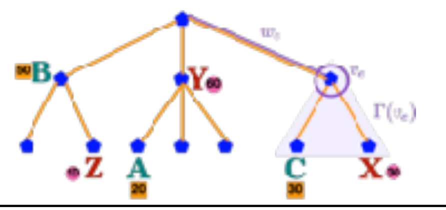
✦ Optimal transport (OT)



$$d_{OT}(\mu,\nu) = \min_{T} \sum_{ij} T_{ij} M_{ij}$$
$$\text{s.t.} \quad \sum_j T_{ij} = u_i \quad \forall i$$
$$\sum_i T_{ij} = v_j \quad \forall j$$
$$T_{ij} \geq 0 \quad \forall i, j$$

Some drawbacks:
(i) same mass
(ii) high complexity
(iii) indefiniteness

✦ Tree-(sliced) Wasserstein: OT for measures on a tree **[1]**



$$d_{TW}(\mu,\nu) = \sum_{e \in \mathcal{T}} w_e \left|\mu(\Gamma(v_e)) - \nu(\Gamma(v_e))\right|$$

✦ Entropy partial transport (EPT): unbalanced OT on a tree **[2]**

$$\mathcal{W}_m(\mu,\nu) := \inf_{\substack{\gamma \in \Pi_{\leq}(\mu,\nu) \\ \gamma(\mathcal{T}\times\mathcal{T}) = m}} \mathcal{F}_1(\gamma_1|\mu) + \mathcal{F}_2(\gamma_2|\nu) + b\int_{\mathcal{T}\times\mathcal{T}} c(x,y)\gamma(dx,dy)$$

Dual formulation:

$$\mathrm{ET}_\lambda(\mu,\nu) = \sup\left\{\int_{\mathcal{T}} f(\mu-\nu) : f \in \mathbb{L}\right\} - \frac{b\lambda}{2}\left[\mu(\mathcal{T}) + \nu(\mathcal{T})\right]$$

$$\mathbb{L} := \left\{f \in C(\mathcal{T}) : -w_2 - \frac{b\lambda}{2} \leq f \leq w_1 + \frac{b\lambda}{2}, |f(x) - f(y)| \leq b d_{\mathcal{T}}(x,y)\right\}$$

✦ Regularized EPT:

$$\widetilde{\mathrm{ET}}_\lambda^\alpha(\mu,\nu) := \sup\left\{\int_{\mathcal{T}} f(\mu-\nu) : f \in \mathbb{L}_\alpha\right\} - \frac{b\lambda}{2}\left[\mu(\mathcal{T}) + \nu(\mathcal{T})\right]$$

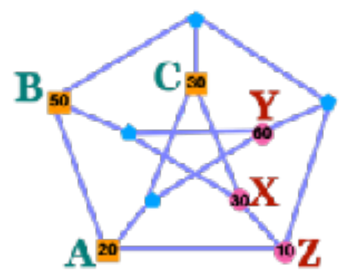
$$\mathbb{L}_\alpha := \left\{f \in C(\mathcal{T}) : f(x) = s + \int_{[r,x]} g(y)\omega(dy), \|g\|_{L^\infty(\mathcal{T})} \leq b, \; s \in \left[-w_2(r) - \frac{b\lambda}{2} + \alpha, w_1(r) + \frac{b\lambda}{2} - \alpha\right]\right\}$$

$$\widetilde{\mathrm{ET}}_\lambda^\alpha(\mu,\nu) = \int_{\mathcal{T}} \left|\mu(\Gamma(x)) - \nu(\Gamma(x))\right| \omega(dx) - \frac{b\lambda}{2}\left[\mu(\mathcal{T}) + \nu(\mathcal{T})\right] + \left[w_i(r) + \frac{b\lambda}{2} - \alpha\right]\left|\mu(\mathcal{T}) - \nu(\mathcal{T})\right|$$

$i := 1$ if $\mu(\mathcal{T}) \geq \nu(\mathcal{T})$ $\quad$ $i := 2$ if $\mu(\mathcal{T}) < \nu(\mathcal{T})$.

Sobolev Transport: A Scalable Variant of OT for Measures on a Graph

✦ OT for measures supported on a graph metric space



+ Min-cost flow problem [Beckmann form.]
+ Complexity: $\mathcal{O}(V^2 E \log V)$ [Orlin's alg.]
+ Indefiniteness

✦ Graph-based Sobolev space $W^{1,p}(\mathbb{G},\lambda)$

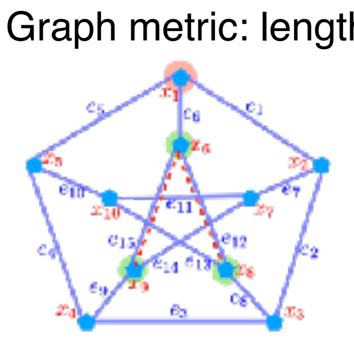
$f : \mathbb{G} \mapsto \mathbb{R}$ belongs to $W^{1,p}(\mathbb{G},\lambda)$

If there exists $h \in L^p(\mathbb{G},\lambda)$ such that

$$f(x) - f(z_0) = \int_{[z_0,x]} h(y)\lambda(dy) \quad \forall x \in \mathbb{G}$$

Such function h is unique, called graph derivative of f w.r.t. λ

✦ Graph metric: length of shortest path between nodes on graph



$$\Lambda(x) = \{y \in \mathbb{G} : x \in [z_0, y]\}$$
$$\gamma_e = \{y \in \mathbb{G} : e \subset [z_0, y]\}$$
$$\Lambda(x_6) = \gamma(e_6) = \tilde{\mathbb{G}}(\{x_6, x_8, x_9\}, \{e_{12}, e_{15}\})$$

[Q1?] How to leverage graph structure to develop scalable approach for optimal transport between measures?
[Q2?] Extend the approach for measures on a graph, but having different total mass?

✦ Sobolev transport (ST) for probability measures on a graph **[3]**

$$\mathcal{S}_p(\mu,\nu) = \begin{cases} \sup & \left[\int_{\mathbb{G}} f(x)\mu(dx) - \int_{\mathbb{G}} f(x)\nu(dx)\right] \\ \text{s.t.} & f \in W^{1,p'}(\mathbb{G},\lambda), \; \|f'\|_{L^{p'}(\mathbb{G},\lambda)} \leq 1. \end{cases}$$

✦ Closed-form expression

$$\mathcal{S}_p(\mu,\nu)^p = \int_{\mathbb{G}} |\mu(\Lambda(x)) - \nu(\Lambda(x))|^p \lambda(dx)$$

Discrete case: for measures supported on graph vertices

$$\mathcal{S}_p(\mu,\nu)^p = \sum_{e \in E} \lambda(e) \left|\mu(\gamma_e) - \nu(\gamma_e)\right|^p$$

✦ Assumptions:
(i) graph is given
(ii) it exist a root node z_0, i.e., unique shortest path from z_0 to other nodes.

✦ Special case: if the given graph is a tree, then Sobolev transport is equivalent to tree-Wasserstein.

[?] The definition of Sobolev transport is based on dual formulation of OT. So, how to extend it for measures supported on a graph, but having different total mass?

✦ Entropy partial transport on a graph

$$\mathrm{ET}(\mu,\nu) = \inf_{\gamma \in \Pi_{\leq}(\mu,\nu)} \mathcal{F}_1(\gamma_1|\mu) + \mathcal{F}_2(\gamma_2|\nu) + b\int_{\mathbb{G}\times\mathbb{G}} [c(x,y) - \beta]\,\gamma(dx,dy)$$

✦ Dual formulation:

$$\mathrm{ET}(\mu,\nu) = \sup_{f \in \mathbb{U}} \int_{\mathbb{G}} f(\mu-\nu) - \frac{b\beta}{2}[\mu(\mathbb{G}) + \nu(\mathbb{G})]$$

$$\mathbb{U} = \left\{f \in C(\mathbb{G}) : -w_2 - \frac{b\beta}{2} \leq f \leq w_1 + \frac{b\beta}{2}, |f(x) - f(y)| \leq b d_{\mathbb{G}}(x,y)\right\}$$

- Challenge to compute EPT efficiently
- Unknown how to extend *p*-order ($p > 1$) even on a tree.

✦ Regularized set $\mathbb{U}_{p'}^\alpha$ for the critic function

For $1 \leq p \leq \infty$ and $0 \leq \alpha \leq \frac{1}{2}[b\alpha + w_1(z_0) + w_2(z_0)]$, $\mathbb{U}_{p'}^\alpha$ is a set of functions: $\quad f \in W^{1,p'}(\mathbb{G},\omega) \quad \|f'\|_{L^{p'}(\mathbb{G},\omega)} \leq b$

$$f(z_0) \in \left[-w_2(z_0) - \frac{b\beta}{2} + \alpha, w_1(z_0) + \frac{b\beta}{2} - \alpha\right]$$

✦ Unbalanced Sobolev Transport **[4]**:

$$\mathrm{US}_p^\alpha(\mu,\nu) = \sup_{f \in \mathbb{U}_{p'}^\alpha} \left[\int_{\mathbb{G}} f(x)\mu(dx) - \int_{\mathbb{G}} f(x)\nu(dx)\right]$$

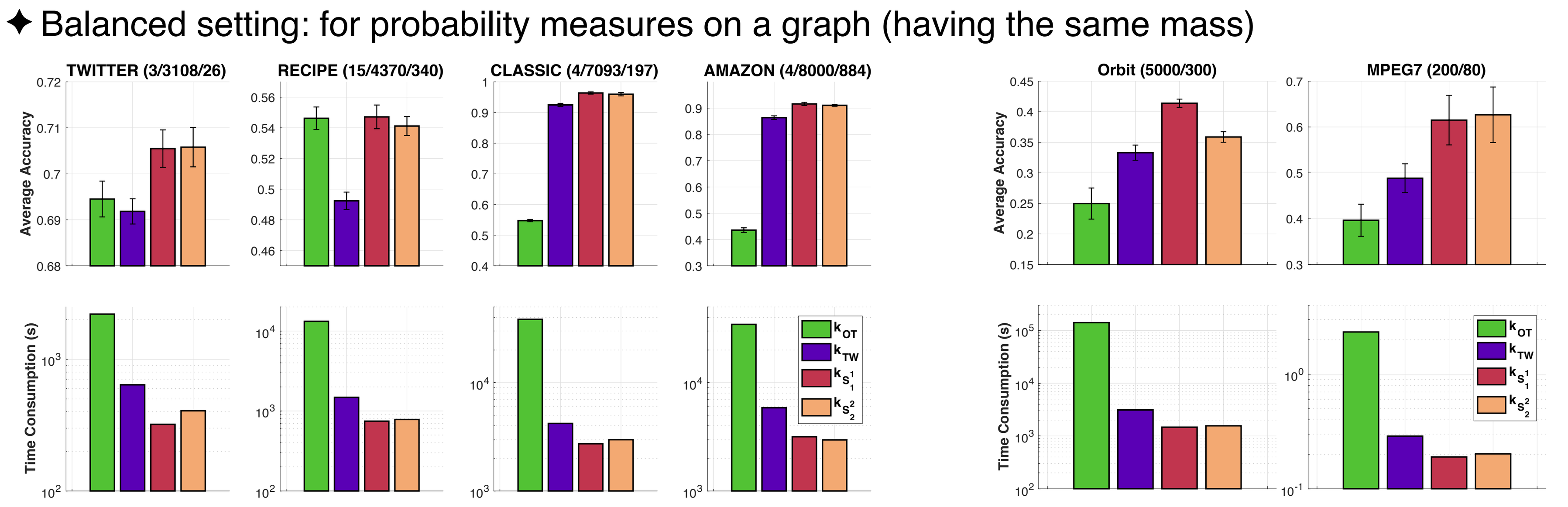
✦ Closed-form expression

$$\mathrm{US}_p^\alpha(\mu,\nu) = b\left(\sum_{e \in E} w_e \left|\mu(\gamma_e) - \nu(\gamma_e)\right|^p\right)^{\frac{1}{p}} + \Theta\left|\mu(\mathbb{G}) - \nu(\mathbb{G})\right|$$

$$\Theta = w_1(z_0) + \frac{b\beta}{2} - \alpha \qquad \text{if } \mu(\mathbb{G}) \geq \nu(\mathbb{G})$$
$$\Theta = w_2(z_0) + \frac{b\beta}{2} - \alpha \qquad \text{if } \mu(\mathbb{G}) < \nu(\mathbb{G})$$

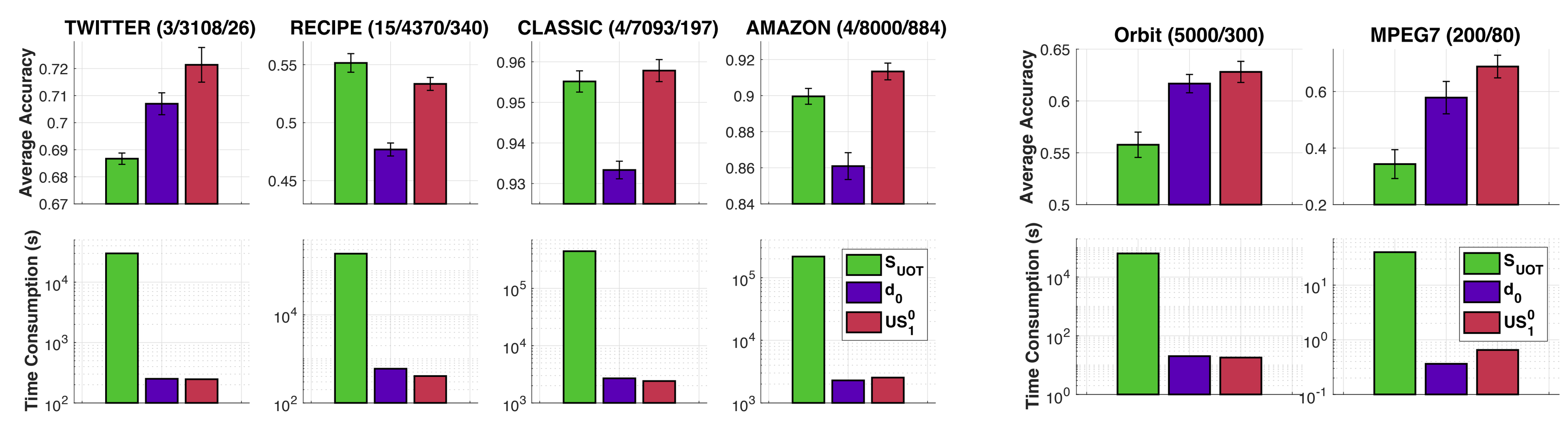
Experimental Results

✦ Balanced setting: for probability measures on a graph (having the same mass)



Document classification $\quad$ Topological Data Analysis

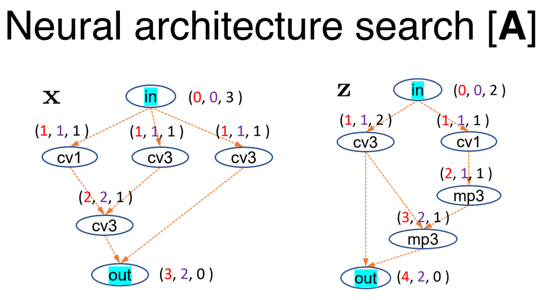
✦ Unbalanced setting: for measures on a graph and having different total mass



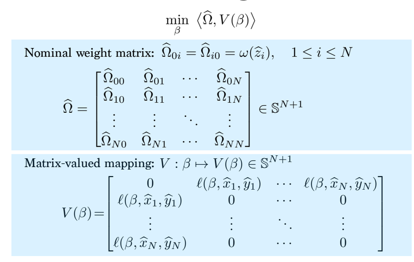
Document classification $\quad$ Topological Data Analysis

OT-based Approaches in Applications

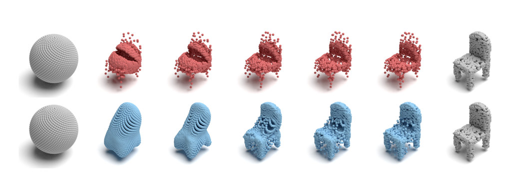
Neural architecture search **[A]**



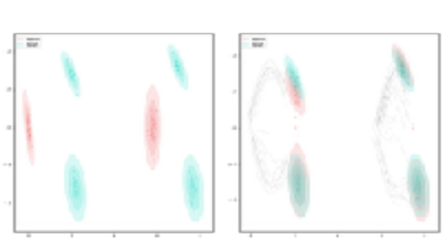
Adversarial regression **[B]**



3D Point Cloud Rep. **[C]**



Dynamic Flows **[D]**



[A] V. Nguyen*, **TL***, M. Yamada, M. Osborne. Optimal Transport Kernels for Sequential and Parallel Neural Architecture Search. **ICML'21**.
[B] **TL***, T. Nguyen*, M. Yamada, J. Blanchet, V. A. Nguyen. Adversarial Regression with Doubly Nonnegative Weight Matrices. **NeurIPS'21**.
[C] T. Nguyen, Q. H. Pham, **TL**, T. Pham, N. Ho, B. S. Hua. Point Set Distances for Learning Representation of 3D point clouds. **ECCV'21**.
[D] X. Hua, T. Nguyen, **TL**, J. Blanchet, V. A. Nguyen. Dynamic Flows on Curved Space Generated by Labeled Data. **IJCAI'23**.

References

[1] **TL**, M. Yamada, K. Fukumizu, M. Cuturi. Tree-Sliced Variants of Wasserstein Distances, **NeurIPS'19**.
[2] **TL***, T. Nguyen*. Entropy Partial Transport with Tree Metrics: Theory and Practice, **AISTATS'21**.
[3] **TL***, T. Nguyen*, D. Phung, V. A. Nguyen. Sobolev Transport: A Scalable Metric for Probability Measures with Graph Metrics. **AISTATS'22**.
[4] **TL**, T. Nguyen, K. Fukumizu. Scalable Unbalanced Sobolev Transport for Measures on a Graph. **AISTATS'23**.

✦ Code: https://tamle-ml.github.io/code.html