Generalized Sobolev Transport for Probability Measures on a Graph

Tam Le*[a,c], Truyen Nguyen*[b], Kenji Fukumizu[a]

[a]The Institute of Statistical Mathematics

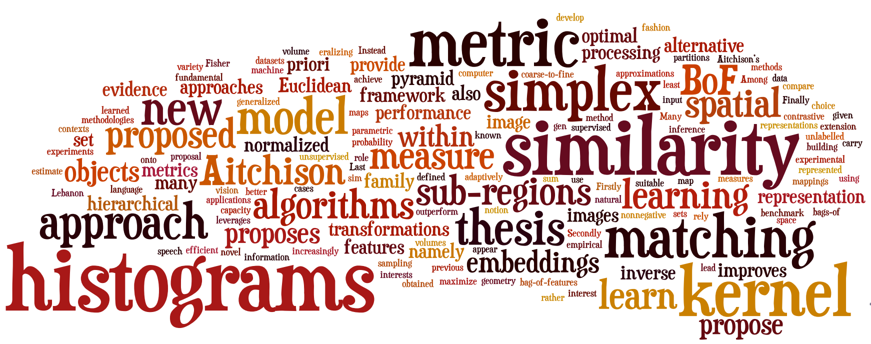
[b]The University of Akron

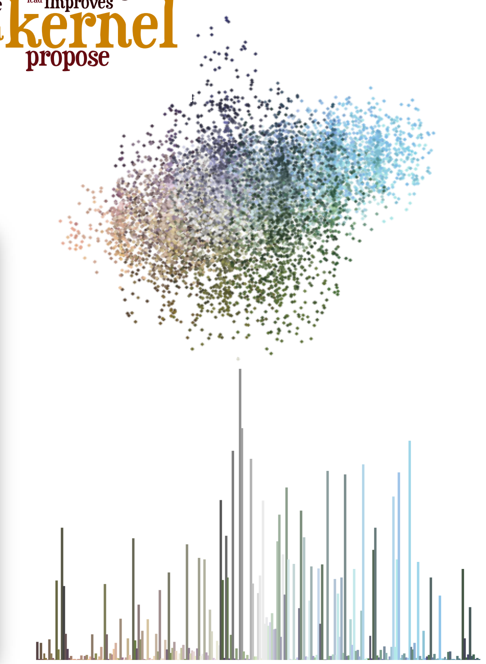
[c]RIKEN AIP

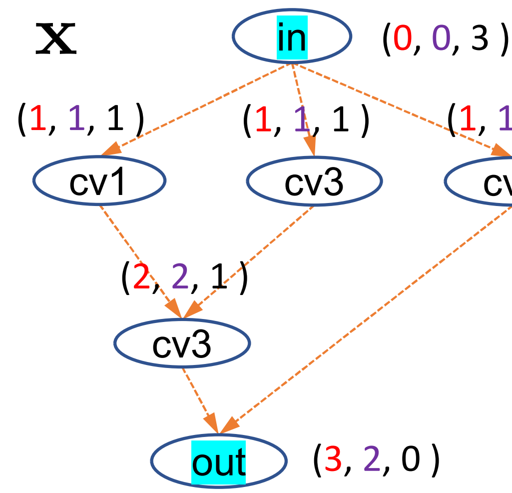
International Conference On Machine Learning

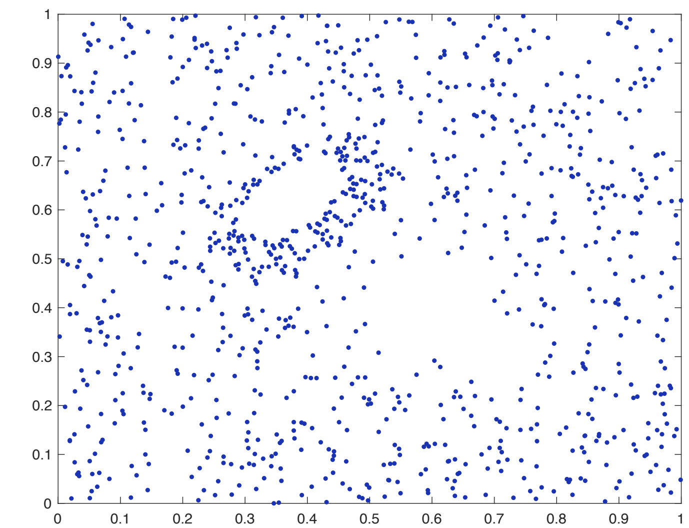
International Conference on Machine Learning (ICML), 2024

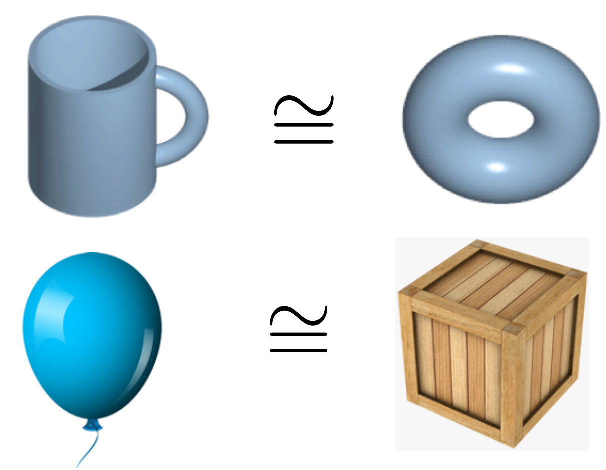
Probability Measures

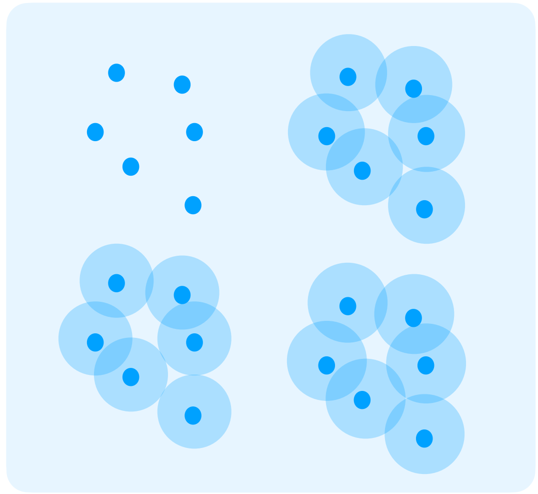


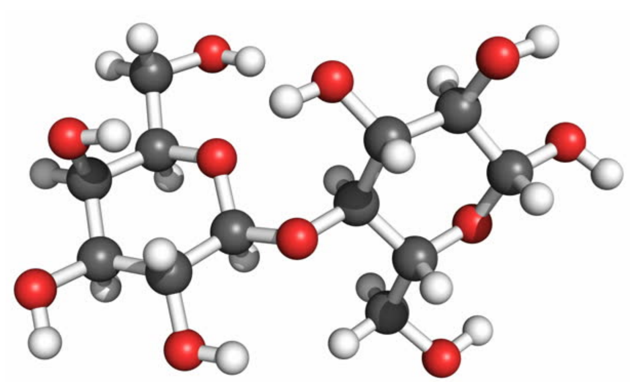


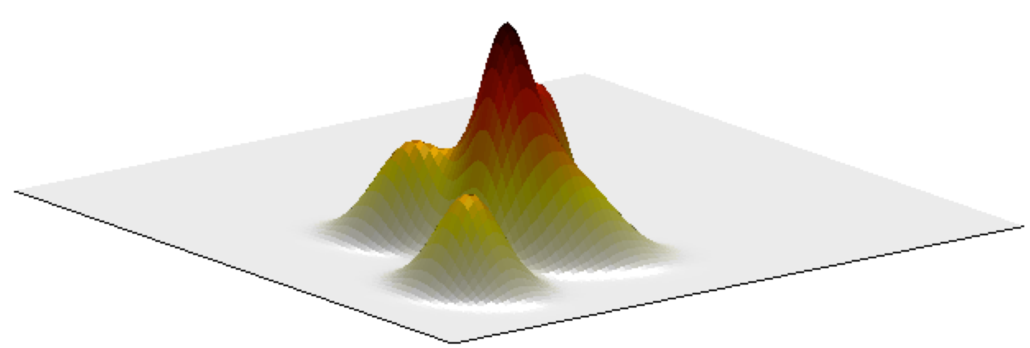


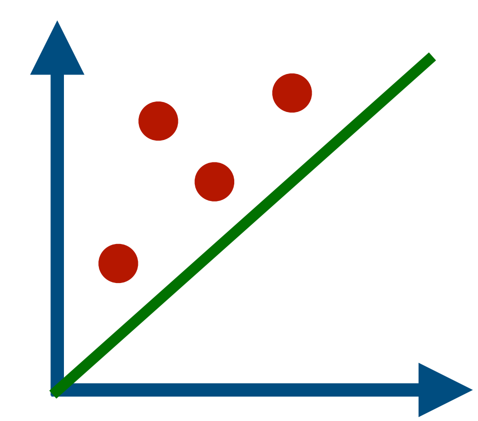












Several objects of interest can be represented as distributions.

Optimal Transport (OT)

Monge

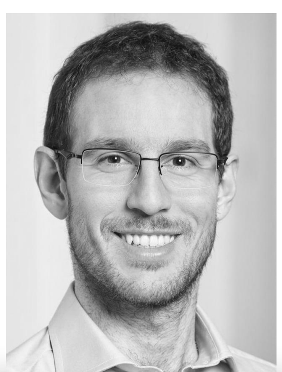
Kantorovich

Koopmans

Nobel ’75

Villani

Fields ’10



Figalli

Fields ’18

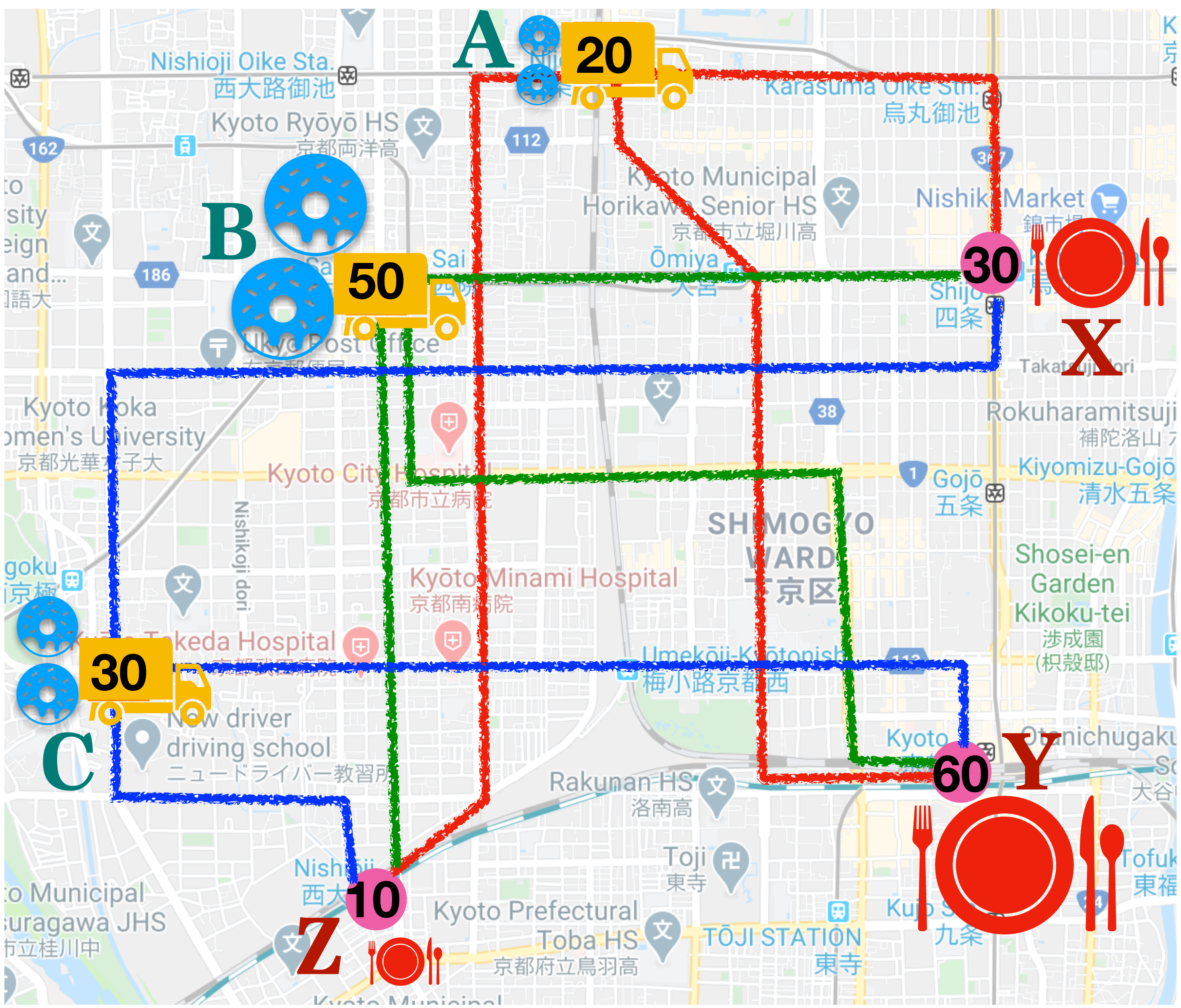
Caffarelli

Abel ’23

The natural geometry for probability measures

Optimal Transport

The natural geometry for distribution data



Cost/Ground Metric

	X	Y	Z
A	d_{AX}	d_{AY}	d_{AZ}
B	d_{BX}	d_{BY}	d_{BZ}
C	d_{CX}	d_{CY}	d_{CZ}

Transportation Plan

	30	60	10
20	6	12	2
50	15	30	5
30	9	18	3

Optimal Transport

Cost/Ground Metric M

	X	Y	Z
A	d_{AX}	d_{AY}	d_{AZ}
B	d_{BX}	d_{BY}	d_{BZ}
C	d_{CX}	d_{CY}	d_{CZ}

Transportation Plan T

u			
20	6	12	2
50	15	30	5
30	9	18	3
v	30	60	10

$$\mu = u_1\delta_A + u_2\delta_B + u_3\delta_C$$
$$\nu = v_1\delta_X + v_2\delta_Y + v_3\delta_Z$$

$$d_{\mathrm{OT}}(\mu,\nu) = \min_{T} \sum_{ij} T_{ij} M_{ij}$$

$$\text{s.t.} \quad \sum_j T_{ij} = u_i \quad \forall i$$

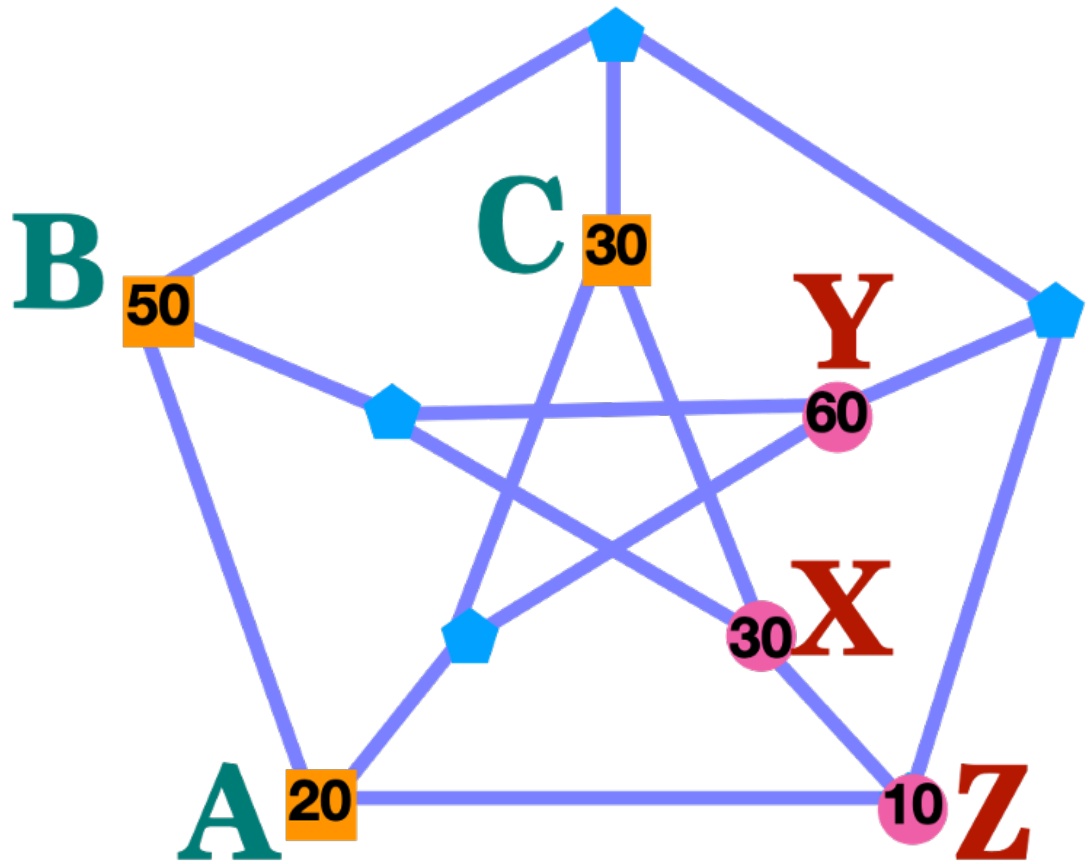
$$\sum_i T_{ij} = v_j \quad \forall j$$

$$T_{ij} \geq 0 \quad \forall i,j$$

- **High complexity: $\mathcal{O}(n^3 \log n)$**
- **Indefinite**

OT with Graph Metric

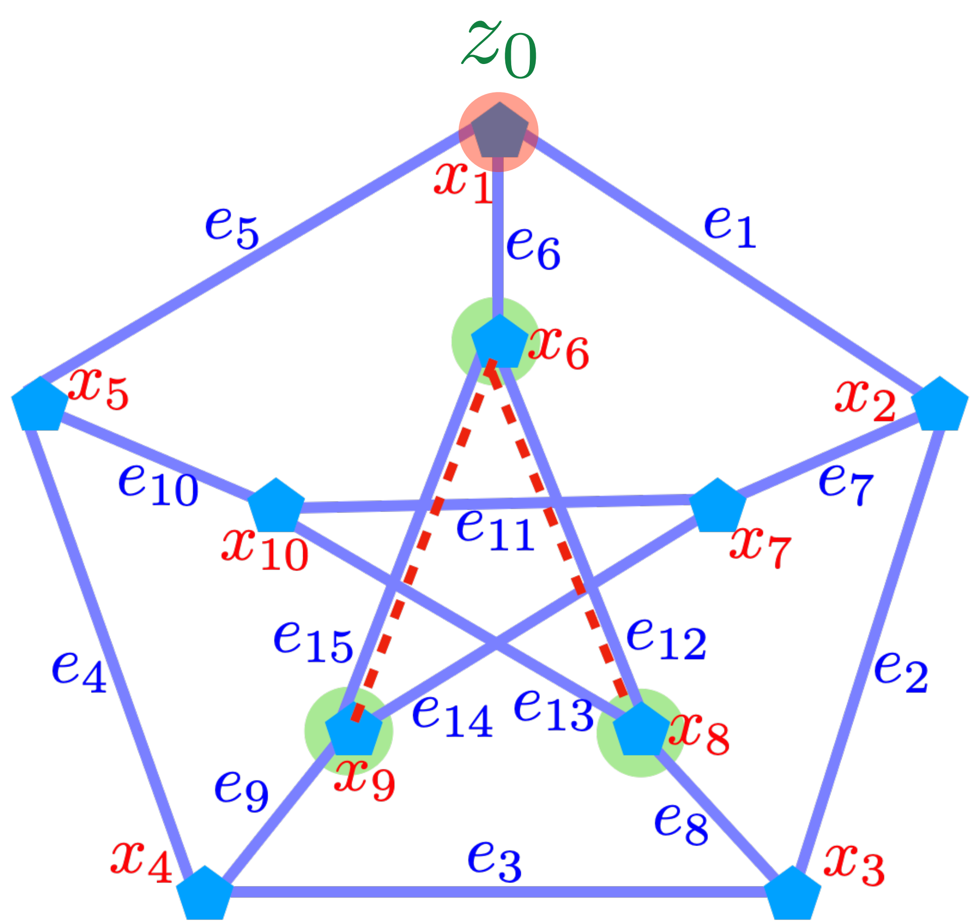
For probability measures supported on graph metric:



- Min-cost flow problem via Beckmann formulation.
- Complexity: $\mathcal{O}(V^2 E \log V)$ by Orlin's algorithm.
- OT with graph metric ground cost is indefinite.

J. B. Orlin. A Polynomial Time Primal Network Simplex Algorithm for Minimum Cost Flows. MP, 1997.

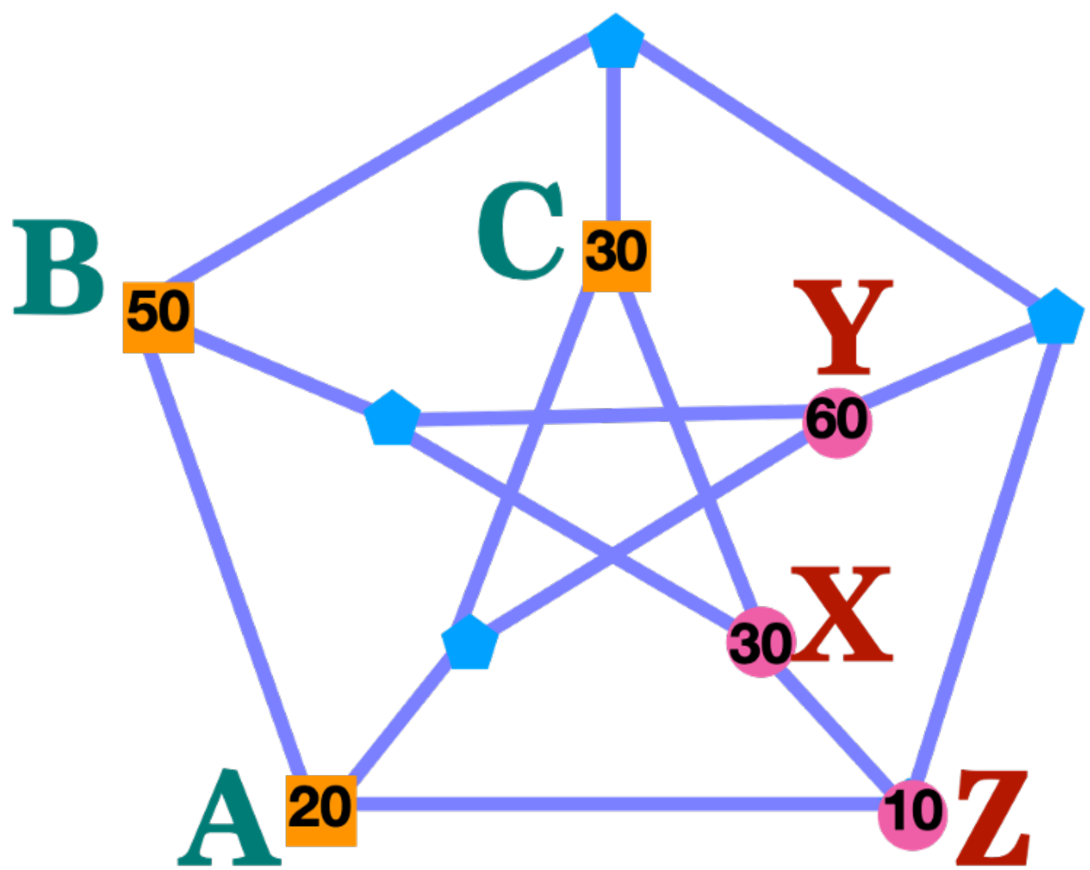
Graph Metric



- Graph metric: length of shortest path between nodes on graph.
- Assumptions: (i) graph is given, (ii) it exist a root node z_0 (i.e., unique shortest path from z_0 to other nodes).

TL*, *T. Nguyen*, D. Phung, V. A. Nguyen. Sobolev transport: a scalable metric for probability measures with graph metrics.* **AISTATS**, *2022.*

Sobolev Transport



- Closed-form solution
- Complexity: $\mathcal{O}(E + V)$
- Negative definite

TL*, *T. Nguyen*, D. Phung, V. A. Nguyen. Sobolev transport: a scalable metric for probability measures with graph metrics.* **AISTATS**, *2022.*

Graph-based Sobolev Space

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

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

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

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

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

λ : nonnegative Borel measure on graph $\mathbb{G}$

$p \geq 1$

$L^p(\mathbb{G}, \lambda)$: space of $f : \mathbb{G} \mapsto \mathbb{R}$ such that $\int_{\mathbb{G}} |f(y)|^p \lambda(\mathrm{d}y) < \infty$

TL*, *T. Nguyen*, D. Phung, V. A. Nguyen. Sobolev transport: a scalable metric for probability measures with graph metrics.* **AISTATS**, *2022.*

Sobolev Transport

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

λ : nonnegative Borel measure on graph $\mathbb{G}$

$p \geq 1$

p': conjugate of p ; $\frac{1}{p} + \frac{1}{p'} = 1$

$\mu, \nu \in \mathcal{P}(\mathbb{G})$

f' : graph derivate of f w.r.t. λ

$\|f\|_{L^p(\mathbb{G},\lambda)} = \left(\int_{\mathbb{G}} |f(y)|^p \lambda(\mathrm{d}y) \right)^{\frac{1}{p}}$

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

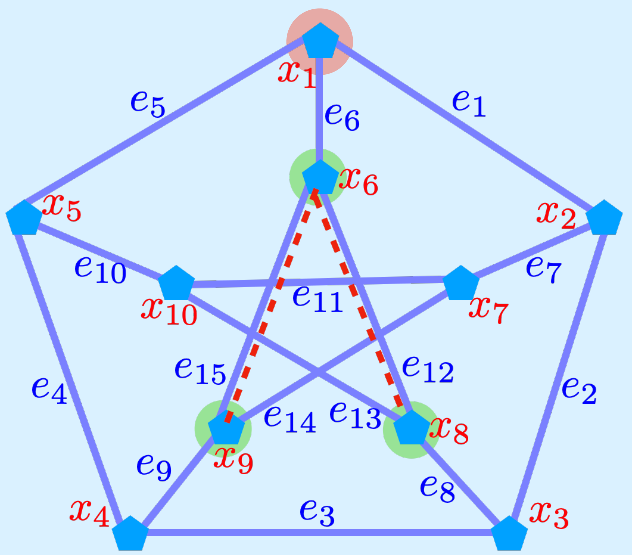
TL*, *T. Nguyen*, D. Phung, V. A. Nguyen. Sobolev transport: a scalable metric for probability measures with graph metrics.* **AISTATS**, *2022.*

Sobolev Transport

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

Discrete case: supports on graph vertices

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



$$\Lambda(x) = \{y \in \mathbb{G} : \, x \in [z_0, y]\}$$

$$\gamma_e = \{y \in \mathbb{G} : \, e \subset [z_0, y]\}$$

$$\Lambda(x_6) = \gamma(e_6) = \tilde{\mathbb{G}}(\{x_6, x_8, x_9\}, \{e_{12}, e_{15}\})$$

TL*, *T. Nguyen*, D. Phung, V. A. Nguyen. Sobolev transport: a scalable metric for probability measures with graph metrics.* **AISTATS**, *2022.*

Sobolev Transport (ST)

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

- Definition is based on dual 1-order Wasserstein.
- Coupled with L^p geometric structure.

- Challenge: nontrivial to use ST with other prior geometric structures.

Orlicz-Wasserstein (OW)

$$W_\Phi(\mu,\nu) = \inf_{\pi \in \Pi(\mu,\nu)} \inf \left[t > 0 : \int_{\mathbb{G}\times\mathbb{G}} \Phi\left(\frac{d_{\mathbb{G}}(x,z)}{t}\right) \mathrm{d}\pi(x,z) \leq 1 \right]$$

Φ: N-function on graph $\mathbb{G}$

- N-function: a strictly increasing and convex function

$$\Phi : [0,\infty) \to [0,\infty)$$

s.t. $\displaystyle \lim_{t\to 0} \frac{\Phi(t)}{t} = 0 \qquad \lim_{t\to+\infty} \frac{\Phi(t)}{t} = +\infty$

- Popular N-functions:

$\Phi(t) = t^p, \quad 1 < p < \infty \qquad \Phi(t) = \exp(t^p) - 1, \quad 1 < p < \infty$

$\Phi(t) = \exp(t) - t - 1 \qquad \Phi(t) = (1+t)\log(1+t) - t$

Orlicz-Wasserstein (OW)

$$W_\Phi(\mu,\nu) = \inf_{\pi \in \Pi(\mu,\nu)} \inf \left[t > 0 : \int_{\mathbb{G}\times\mathbb{G}} \Phi\left(\frac{d_{\mathbb{G}}(x,z)}{t}\right) \mathrm{d}\pi(x,z) \leq 1 \right]$$

OW helps to advance certain ML approaches

- For quantifying the rate of parameter convergence in infinite Gaussian mixtures [Guha et al., ICML'23]
    - OW alleviates raised concerns from the usage of 2-order Wasserstein.
    - OW helps to significantly improve the contraction rate.

A. Guha, N. Ho, X. Nguyen. On Excess Mass Behavior in Gaussian Mixture Models with Orlicz-Wasserstein Distances. ICML, 2023.

Orlicz-Wasserstein (OW)

$$W_\Phi(\mu,\nu) = \inf_{\pi\in\Pi(\mu,\nu)} \inf \left[t > 0 : \int_{\mathbb{G}\times\mathbb{G}} \Phi\left(\frac{d_{\mathbb{G}}(x,z)}{t}\right) \mathrm{d}\pi(x,z) \le 1 \right]$$

OW helps to advance certain ML approaches

- For quantifying the rate of parameter convergence in infinite Gaussian mixtures [Guha et al., ICML'23]
- Overcome longstanding challenges for proving fast convergence of hypocoercive differential equations [Altschuler & Chewi, FOCS'23]

A. Guha, N. Ho, X. Nguyen. On Excess Mass Behavior in Gaussian Mixture Models with Orlicz-Wasserstein Distances. ICML, 2023.
J. M. Altschuler, S. Chewi. Faster High-Accuracy Log-Concave Sampling via Algorithmic Warm Starts. FOCS, 2023.

Orlicz-Wasserstein (OW)

$$W_\Phi(\mu,\nu) = \inf_{\pi\in\Pi(\mu,\nu)} \inf \left[t > 0 : \int_{\mathbb{G}\times\mathbb{G}} \Phi\left(\frac{d_{\mathbb{G}}(x,z)}{t}\right) \mathrm{d}\pi(x,z) \le 1\right]$$

OW helps to advance certain ML approaches

- For quantifying the rate of parameter convergence in infinite Gaussian mixtures [Guha et al., ICML'23]
- Overcome longstanding challenges for proving fast convergence of hypocoercive differential equations [Altschuler & Chewi, FOCS'23]

Computational challenge:

- Two-level optimization formula.

A. Guha, N. Ho, X. Nguyen. On Excess Mass Behavior in Gaussian Mixture Models with Orlicz-Wasserstein Distances. ICML, 2023.
J. M. Altschuler, S. Chewi. Faster High-Accuracy Log-Concave Sampling via Algorithmic Warm Starts. FOCS, 2023.

Graph-based Orlicz-Sobolev Space

- Graph-based Orlicz-Sobolev space: $WL_\Phi(\mathbb{G}, \omega)$

$f : \mathbb{G} \mapsto \mathbb{R}$ belongs to $WL_\Phi(\mathbb{G}, \omega)$

If there exists $h \in L_\Phi(\mathbb{G}, \omega)$ such that

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

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

ω : nonnegative Borel measure on graph $\mathbb{G}$

$L_\Phi(\mathbb{G}, \omega)$: space of $f : \mathbb{G} \mapsto \mathbb{R}$ such that $\int_{\mathbb{G}} \Phi(|f(x)|)\omega(\mathrm{d}x) < \infty$

TL*, *T. Nguyen*, K. Fukumizu. Generalized Sobolev Transport for Probability Measures on a Graph.* **ICML**, *2024*

Generalized Sobolev Transport

$$\mathcal{GS}_\Phi(\mu,\nu) := \begin{cases} \sup & \left| \int_{\mathbb{G}} f(x)\mu(\mathrm{d}x) - \int_{\mathbb{G}} f(x)\nu(\mathrm{d}x) \right| \\ \text{s.t.} & f \in WL_\Psi(\mathbb{G},\omega),\ \|f'\|_{L_\Psi} \leq 1, \end{cases}$$

ω : nonnegative Borel measure on graph $\mathbb{G}$

$\mu, \nu \in \mathcal{P}(\mathbb{G})$

f' : graph derivate of f w.r.t. ω

$WL_\Phi(\mathbb{G},\omega)$: graph-based Orlicz-Sobolev space

$$\|f\|_{L_\Phi} := \inf \left\{ t > 0 \mid \int_{\mathbb{G}} \Phi\left(\frac{|f(x)|}{t}\right) \omega(\mathrm{d}x) \leq 1 \right\}.$$

$$\Psi(t) = \sup \left[at - \Phi(a) \ \mid \ a \geq 0 \right], \quad \text{for } \ t \geq 0.$$

TL*, *T. Nguyen*, K. Fukumizu. Generalized Sobolev Transport for Probability Measures on a Graph.* **ICML**, *2024*

Generalized Sobolev Transport (GST)

$$\boxed{\mathcal{GS}_\Phi(\mu,\nu) = \inf_{k>0} \frac{1}{k}\left(1 + \int_{\mathbb{G}} \Phi(k\,|\mu(\Lambda(x)) - \nu(\Lambda(x))|)\omega(\mathrm{d}x)\right),}$$

Discrete case: supports on graph vertices

$$\boxed{\mathcal{GS}_\Phi(\mu,\nu) = \inf_{k>0} \frac{1}{k}\left[1 + \sum_{e\in E} w_e \Phi(k|\mu(\gamma_e) - \nu(\gamma_e)|)\right],}$$

➡ Compute GST by simply solving univariate optimization problem.

TL*, *T. Nguyen*, K. Fukumizu. Generalized Sobolev Transport for Probability Measures on a Graph.* **ICML**, *2024*

Connection to Sobolev Transport

$$\Phi(t) := \frac{(p-1)^{p-1}}{p^p} t^p, \quad 1 < p < \infty$$

$$\boxed{\mathcal{GS}_{\Phi}(\mu, \nu) = \mathcal{S}_p(\mu, \nu)}$$

➡ GST yields a closed-form expression for this case.

TL*, *T. Nguyen*, K. Fukumizu. Generalized Sobolev Transport for Probability Measures on a Graph.* **ICML**, *2024*

Connection to Orlicz-Wasserstein

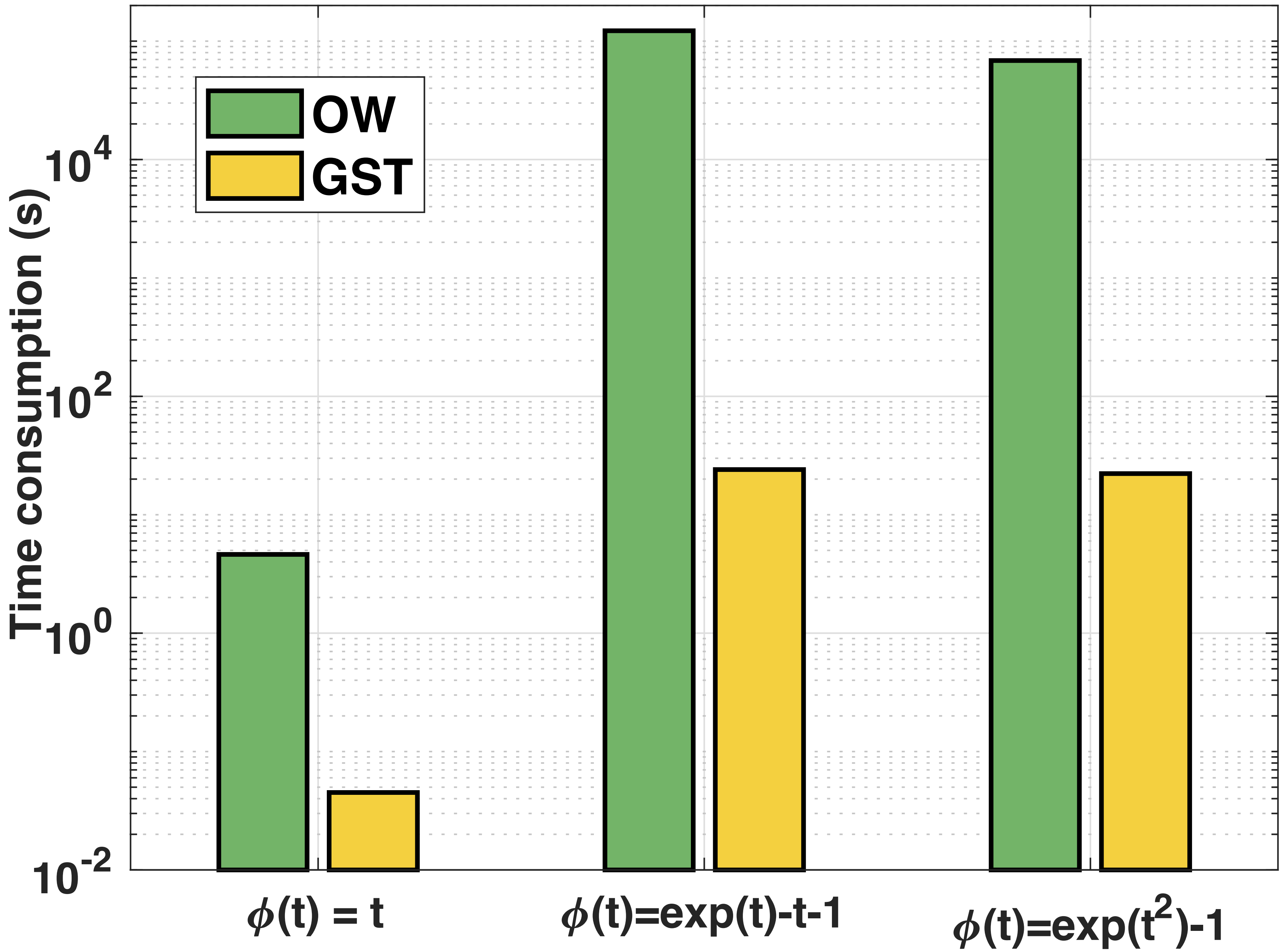
$\Phi(t) = t$

$\mathbb{G}$: a tree

$$\boxed{\mathcal{GS}_{\Phi}(\mu, \nu) = W_{\Phi}(\mu, \nu)}$$

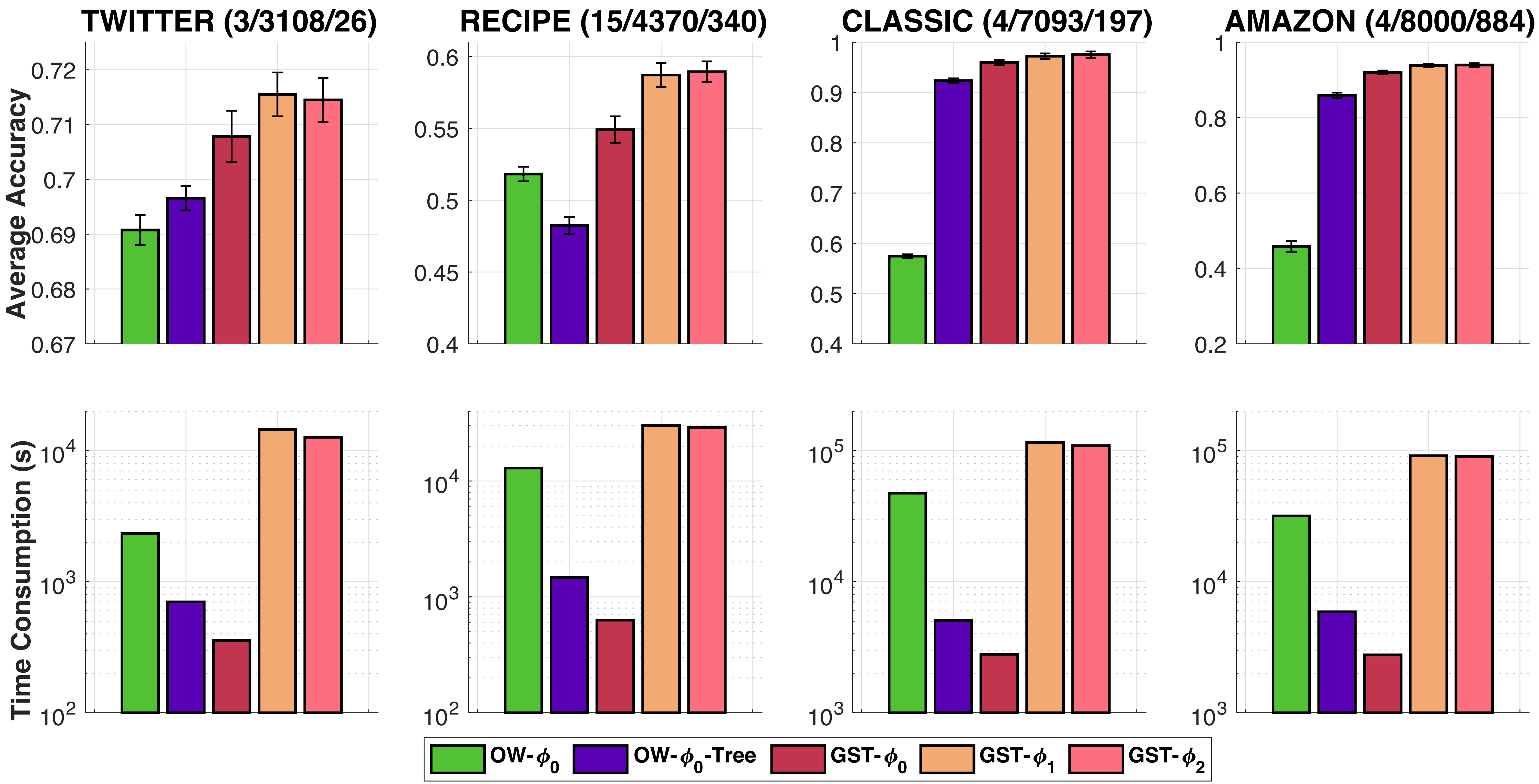
TL*, *T. Nguyen*, K. Fukumizu. Generalized Sobolev Transport for Probability Measures on a Graph.* **ICML**, *2024*

Fast computation: GST v.s. OW



TL*, *T. Nguyen*, K. Fukumizu. Generalized Sobolev Transport for Probability Measures on a Graph.* **ICML**, *2024*

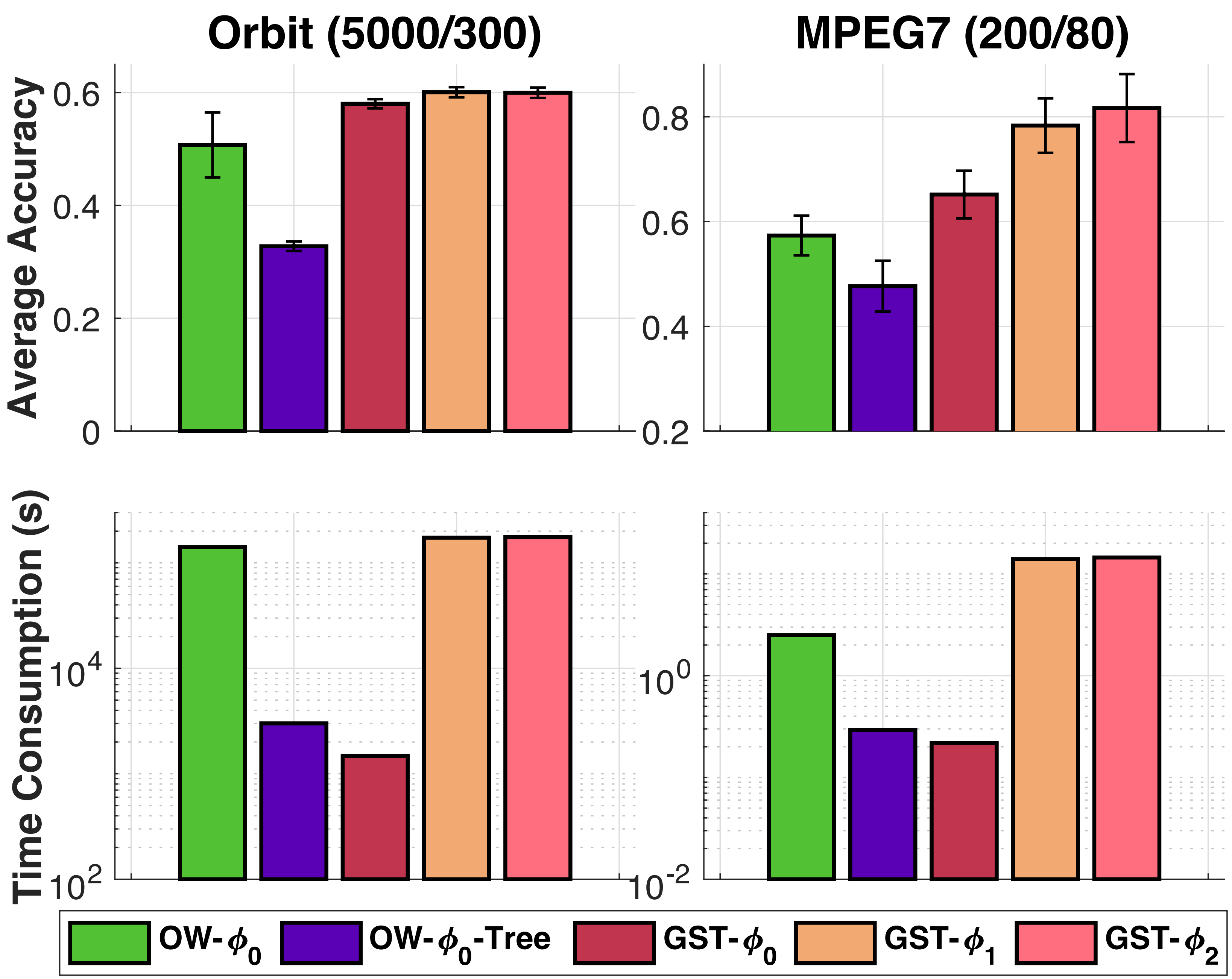
Empirical Results



$$\Phi_0(t) = t; \quad \Phi_1(t) = \exp(t) - t - 1; \quad \Phi_2(t) = \exp(t^2) - 1$$

TL*, *T. Nguyen*, K. Fukumizu. Generalized Sobolev Transport for Probability Measures on a Graph.* **ICML**, *2024*

Empirical Results



$$\Phi_0(t) = t; \quad \Phi_1(t) = \exp(t) - t - 1; \quad \Phi_2(t) = \exp(t^2) - 1$$

TL*, *T. Nguyen*, K. Fukumizu. Generalized Sobolev Transport for Probability Measures on a Graph.* **ICML**, *2024*

Thank you
for your attention