# Orlicz-Sobolev with Musielak: An Efficient Regularization Approach for Graph-based IPM

**Tam Le***(a), Truyen Nguyen*(b),
Hideitsu Hino**(a), Kenji Fukumizu**(a)
(a)The Institute of Statistical Mathematics
(b)The University of Akron




Neural Information Processing Systems (NeurIPS), 2026

# Problem Setting

We study **Sobolev IPM** problem for *graph-based measures*.

$$\mathcal{S}_p(\mu,\nu) := \sup_{f\in\mathcal{B}(p')}\left|\int_{\mathbb{G}} f(x)\mu(\mathrm{d}x) - \int_{\mathbb{G}} f(y)\nu(\mathrm{d}y)\right|$$

$$\text{where } \mathcal{B}(p') := \left\{f \in W_0^{1,p'}(\mathbb{G},\lambda) : \|f\|_{W^{1,p'}} \leq 1\right\}$$

# Problem Setting

We study **Sobolev IPM** problem for *graph-based measures*.

$$\mathcal{S}_p(\mu,\nu) := \sup_{f\in\mathcal{B}(p')}\left|\int_{\mathbb{G}} f(x)\mu(\mathrm{d}x) - \int_{\mathbb{G}} f(y)\nu(\mathrm{d}y)\right|$$

where $\mathcal{B}(p') := \left\{ f \in W_0^{1,p'}(\mathbb{G},\lambda) : \|f\|_{W^{1,p'}} \le 1 \right\}$

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

$p'$ : conjugate of $p$ , i.e., $\dfrac{1}{p}+\dfrac{1}{p'}=1$

$W_0^{1,p}(\mathbb{G},\lambda)$: Sobolev space $W^{1,p}(\mathbb{G},\lambda)$ with $f(z_0)=0$

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

$\|f\|_{W^{1,p}} := \left(\|f\|_{L^p}^p + \|f'\|_{L^p}^p\right)^{\frac{1}{p}}$

where $\|f\|_{L^p} := \left(\displaystyle\int_{\mathbb{G}} |f(x)|^p \lambda(\mathrm{d}x)\right)^{\frac{1}{p}}$.

# Problem Setting

We study **Sobolev IPM** problem for *graph-based measures*.

$$\mathcal{S}_p(\mu,\nu) := \sup_{f\in\mathcal{B}(p')}\left|\int_{\mathbb{G}} f(x)\mu(\mathrm{d}x) - \int_{\mathbb{G}} f(y)\nu(\mathrm{d}y)\right|$$

$$\text{where } \mathcal{B}(p') := \left\{f \in W_0^{1,p'}(\mathbb{G},\lambda) : \|f\|_{W^{1,p'}} \le 1\right\}$$

**Challenges:**

- Intrinsically coupled with the $L^p$ geometric structure.
- IPM formulation: nontrivial to adapt to different prior geometric structures.

# Approach & Challenge

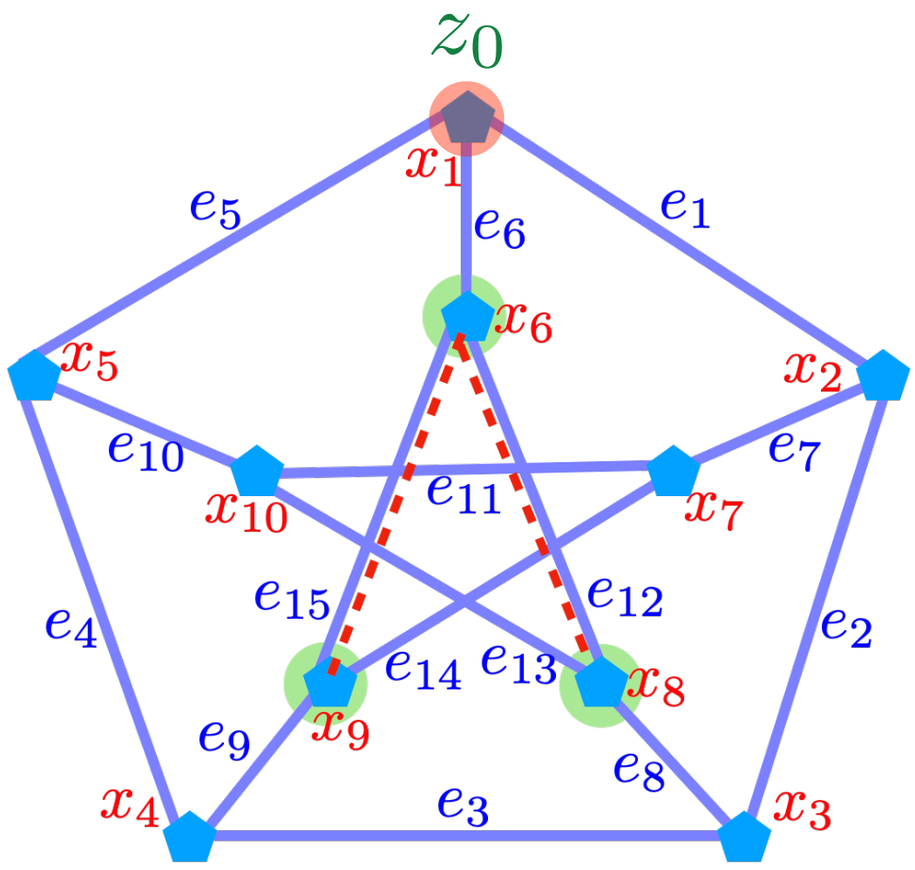
By leveraging *Orlicz geometric structure* (e.g., Orlicz-Wasserstein), we generalize Sobolev IPM beyond the $L^p$ geometry regime.

- Recent advantages of Orlicz geometric structure for machine learning problems.
- Adapt to different prior geometric structures beyond $L^p$.

**Challenge:**

- How to compute the generalized Sobolev IPM efficiently for practical applications?

# 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).

# Generalized Sobolev IPM (GSI)

**GSI:** Sobolev IPM with Orlicz geometric structure

$$\mathcal{GS}_{\Phi}(\mu,\nu) := \sup_{f\in\mathcal{B}_{\Psi}} \left| \int_{\mathbb{G}} f(x)\mu(\mathrm{d}x) - \int_{\mathbb{G}} f(y)\nu(\mathrm{d}y) \right|$$

where $\mathcal{B}_{\Psi} := \left\{ f \in WL^1_{\Psi,0}(\mathbb{G},\lambda), \|f\|_{WL^1_{\Psi}} \leq 1 \right\}$

$WL^1_{\Psi,0}(\mathbb{G},\lambda)$ **: Orlicz-Sobolev space with** $f(z_0)=0$

**A pair of complimentary N-functions:** $\Psi(t) = \sup_{a\geq 0}[at - \Phi(a)]$

**Orlicz-Sobolev norm:** $\|f\|_{WL^1_{\Phi}} := \|f\|_{L_{\Phi}} + \|f'\|_{L_{\Phi}}$

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

# Key Insights

**Theorem 1 (informal) [Equivalence]**

$$\frac{1}{2}\,\|f'\|_{L^{\hat{w}}_{\Phi}} \leq \|f\|_{WL^{1}_{\Phi}} \leq (1+\lambda(\mathbb{G}))\,\|f'\|_{L^{\hat{w}}_{\Phi}}$$

where the Musielak norm is defined:

$$\|f\|_{L^{\hat{w}}_{\Phi}} := \inf\left\{t>0 \mid \int_{\mathbb{G}} \hat{w}(x)\,\Phi\left(\frac{|f(x)|}{t}\right)\lambda(\mathrm{d}x) \leq 1\right\},$$

**with the weight function:**

$$\hat{w}(x) := 1 + \frac{\lambda(\Lambda(x))}{\lambda(\mathbb{G})}, \quad \forall x \in \mathbb{G}.$$

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

# GSI with Musielak Regularization

Generalized Sobolev IPM with Musielak regularization (**GSI-M**)

$$\widehat{\mathcal{GS}}_{\Phi}(\mu,\nu) := \sup_{f\in\mathcal{B}_{\Psi}^{\hat{w}}} \left| \int_{\mathbb{G}} f(x)\mu(\mathrm{d}x) - \int_{\mathbb{G}} f(y)\nu(\mathrm{d}y) \right|$$

where $\mathcal{B}_{\Psi}^{\hat{w}} := \left\{ f \in WL^{1}_{\Psi,0}(\mathbb{G},\lambda), \|f'\|_{L^{\hat{w}}_{\Psi}} \leq 1 \right\}$

# 1D-Optimization for GSI-M

$$\widehat{\mathcal{GS}}_{\Phi}(\mu,\nu) = \inf_{k>0} \frac{1}{k}\left\{1 + \sum_{e\in E}\left[\int_0^1 \bar{w}_t(e)\,\Phi\left(\frac{k|\bar{h}(e)|}{\bar{w}_t(e)}\right) w_e \mathrm{d}t\right]\right\},$$

where $\bar{h}(e) := \mu(\gamma_e) - \nu(\gamma_e)$ $\qquad$ $\gamma_e := \left\{y \in \mathbb{G} :\ e \subset [z_0, y]\right\}$

$$\bar{w}_t(e) := \frac{w_e}{\lambda(\mathbb{G})}t + 1 + \frac{\lambda(\gamma_e)}{\lambda(\mathbb{G})}, \quad \forall e \in E,\, t \in [0,1]$$

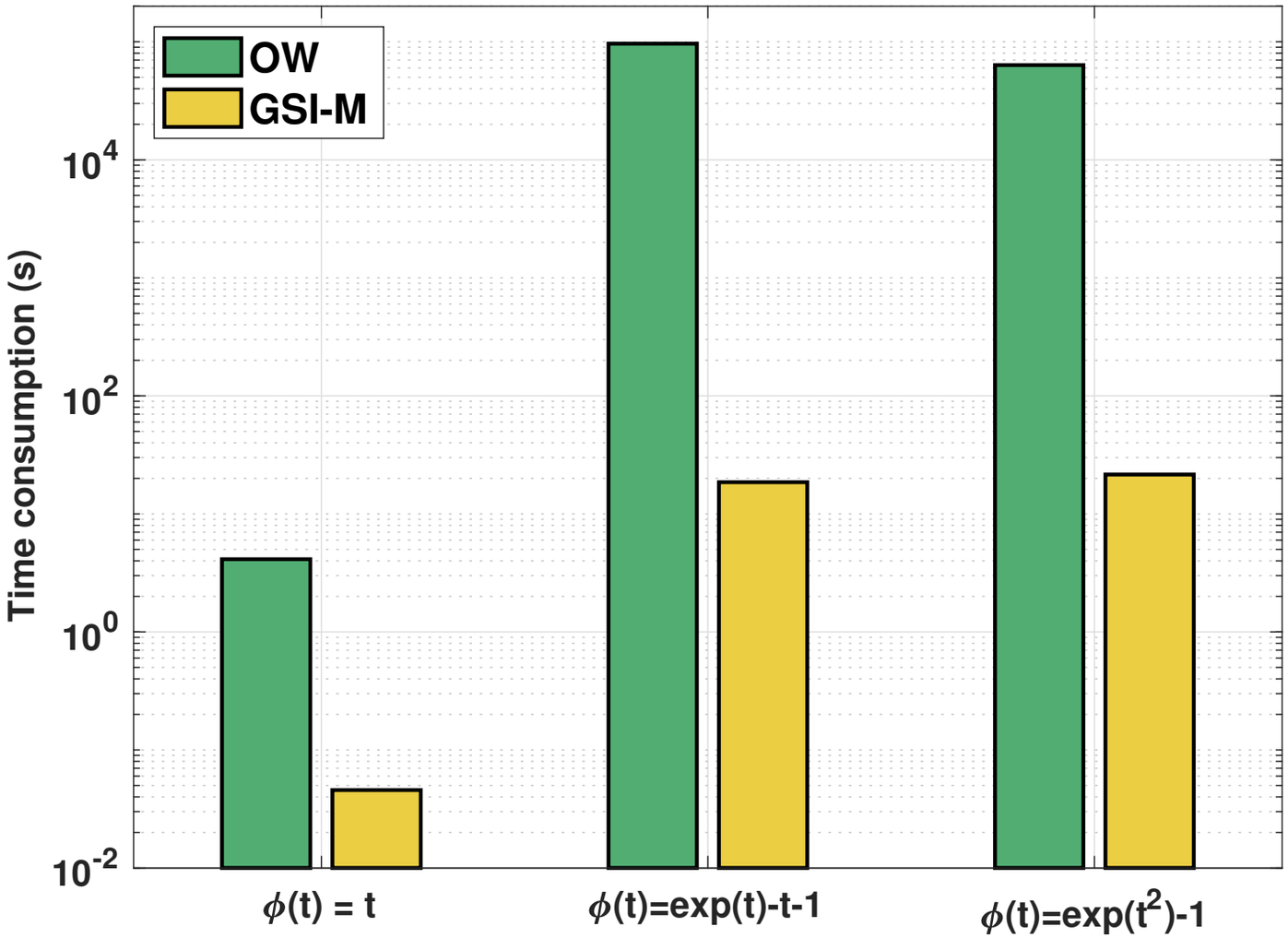
# Theoretical Properties

**Theorem 2 (informal):** GSI-M is a **metric.**

**Theorem 3 (informal)** [Equivalent metric]

$$\frac{1}{2}\,\mathcal{GS}_{\Phi}(\mu,\nu) \leq \widehat{\mathcal{GS}}_{\Phi}(\mu,\nu) \leq (1+\lambda(\mathbb{G}))\,\mathcal{GS}_{\Phi}(\mu,\nu)$$

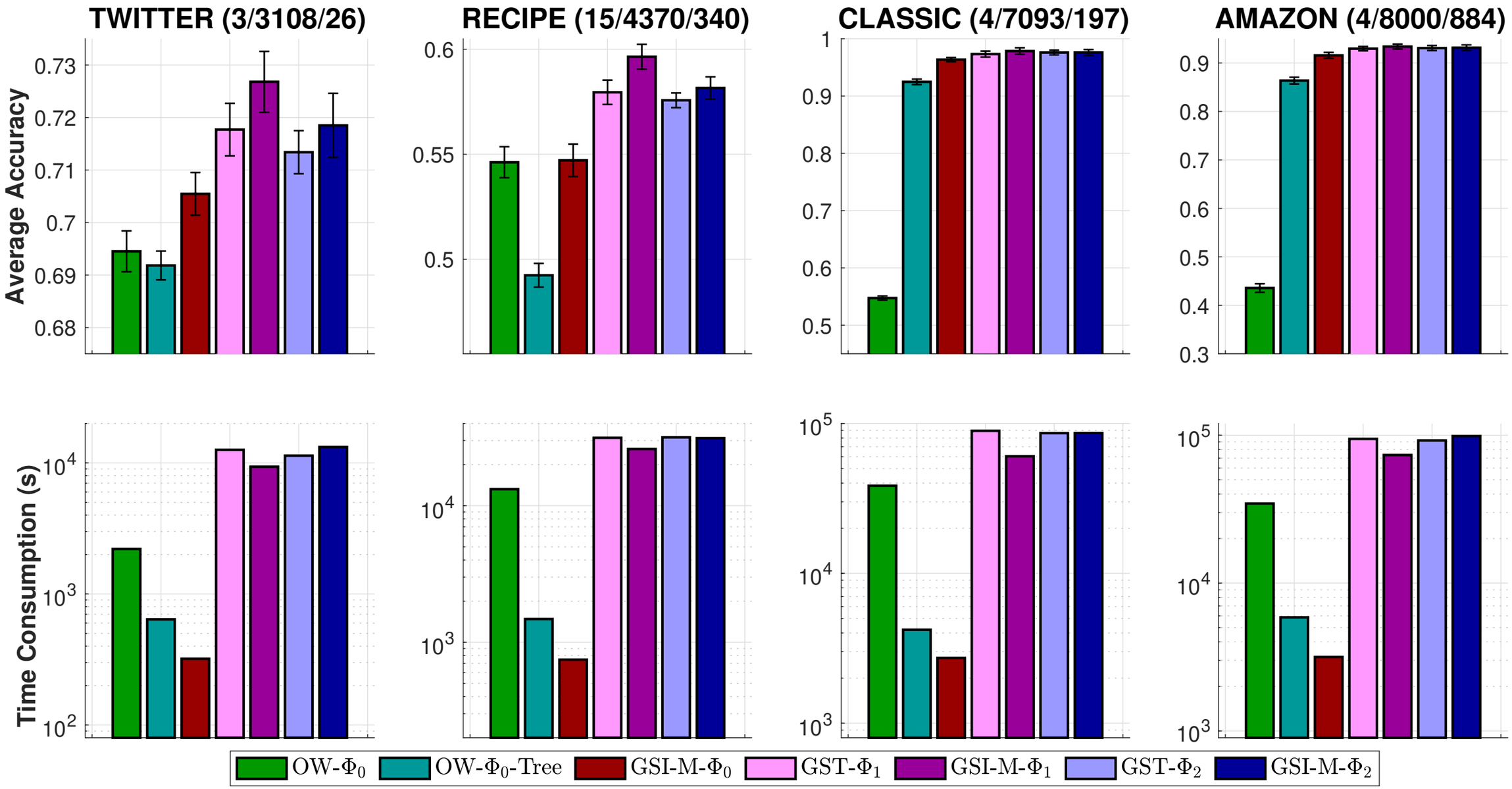
# Time Consumption




- 10K pairs (AMAZON)
- Graph with 1K nodes

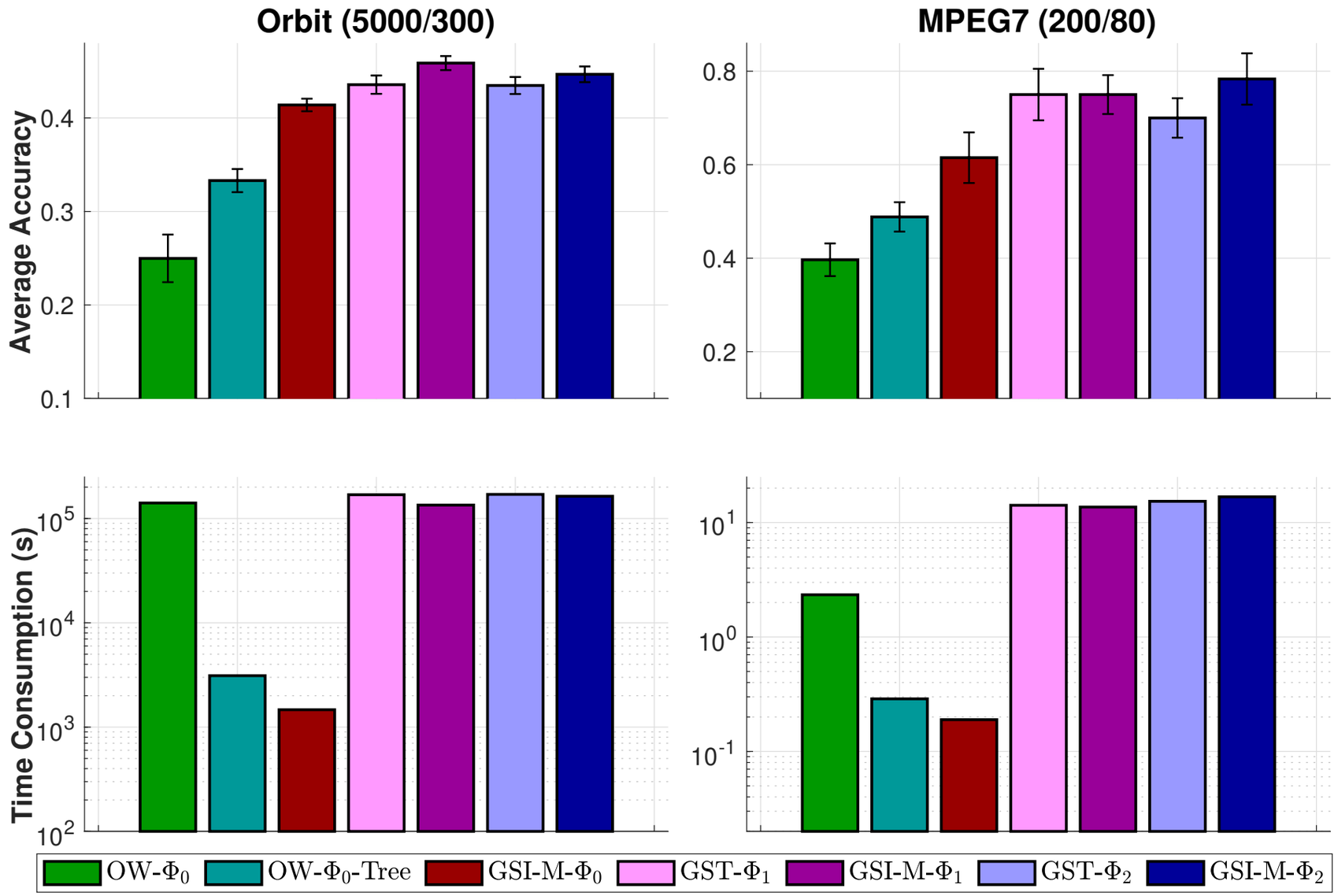
- **GSI-M: < 22 seconds**
- **Orlicz-Wasserstein (OW): > 17 hours**

# Empirical Results




- Kernel SVM classification
- Graph with 10K nodes and 100K edges

# Empirical Results




- Kernel SVM classification
- Graph with 10K nodes and 100K edges

  (MPEG7: graph 1K nodes, 7K edges due to its small size)

# Thank you for your attention

# 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. $\lambda$

$\lambda$ : 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.*

# Graph-based Orlicz-Sobolev Space

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

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

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. $\lambda$

$\omega$ : 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*

# 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.*