An Efficient Orlicz-Sobolev Approach for Transporting Unbalanced Measures on a Graph

Tam Le*[a], Truyen Nguyen*[b],
Hideitsu Hino[a], Kenji Fukumizu[a]
[a]Institute of Statistical Mathematics
[b]University of Akron

NEURAL INFORMATION
PROCESSING SYSTEMS

Neural Information Processing Systems (NeurIPS), 2025
[Spotlight]

Problem Setting

We study optimal transport (OT) for *unbalanced graph-based measures*.

- Orlicz-Wasserstein (OW) / Generalized Sobolev transport (GST) employ *Orlicz geometric structure* to mitigate the limitations of traditional L^p geometry to remarkably advance certain machine learning approaches.
- OW/GST are restricted to measures with equal total mass.

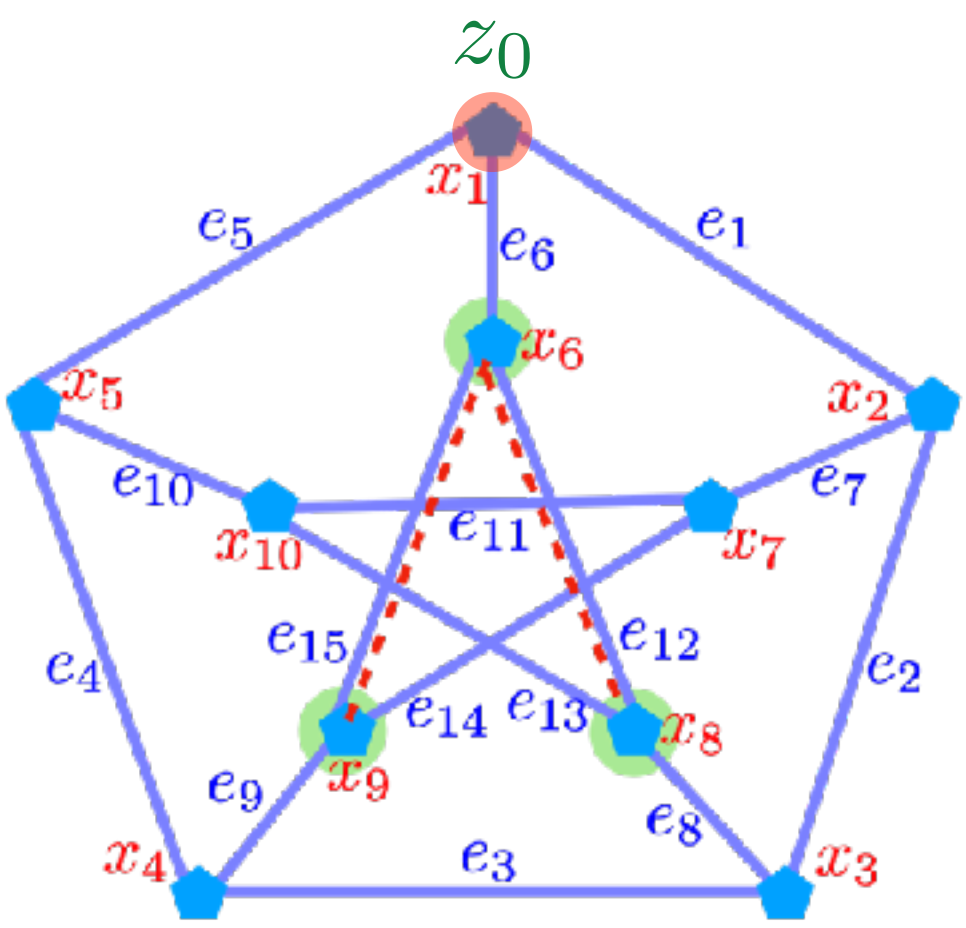
Challenges

How to efficiently *generalized OW/GST for unbalanced measures* supported on a graph metric space?

- For OW: one can incorporate mass constraints, or marginal discrepancy penalization. However, it leads to a more complex two-level optimization problem.

- For GST: it is a scalable, yet rigid framework. It poses significant challenges to extend GST to accommodate unbalanced measures.

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).

Orlicz-EPT

Orlicz-EPT: Entropy Partial Transport with Orlicz geometric structure

- We revisit EPT problem for unbalanced graph-based measures, and leverage Caffarelli & McCann's insight to reformulate EPT as a standard complete OT.
- We carefully calibrate the ground cost of the corresponding standard OT to ensure its nonnegativity.
- We establish theoretical background to solve Orlicz-EPT using a binary search algorithmic approach.

Revisit EPT on a Graph

$$\mathrm{W}_m(\mu, \nu) := \inf_{\substack{\gamma \in \Pi_{\leq}(\mu,\nu) \\ \gamma(\mathbb{G}\times\mathbb{G})=m}} \left[\underline{\mathcal{F}_1(\gamma_1|\mu) + \mathcal{F}_2(\gamma_2|\nu)} + \underline{b \int_{\mathbb{G}\times\mathbb{G}} d_{\mathbb{G}}(x,y)\gamma(\mathrm{d}x, \mathrm{d}y)} \right]$$

Revisit EPT on a Graph

$$\mathrm{W}_m(\mu,\nu) := \inf_{\substack{\gamma\in\Pi_{\leq}(\mu,\nu)\\ \gamma(\mathbb{G}\times\mathbb{G})=m}} \left[\underline{\mathcal{F}_1(\gamma_1|\mu) + \mathcal{F}_2(\gamma_2|\nu)} + \underline{b\int_{\mathbb{G}\times\mathbb{G}} d_{\mathbb{G}}(x,y)\gamma(\mathrm{d}x,\mathrm{d}y)} \right]$$

$$\Pi_{\leq}(\mu,\nu) := \{\gamma : \gamma_1 \leq \mu,\ \gamma_2 \leq \nu\}$$

f_1, f_2 : Radon-Nikodym derivatives

$$\gamma_1 = f_1\mu, \quad \gamma_2 = f_2\nu$$

F_1, F_2 : entropy functions

$$F_1(s) = F_2(s) = |s-1|$$

Nonnegative weights

$$w_1, w_2 : \mathcal{T} \to [0,\infty)$$

Weighted relative entropies

$$\mathcal{F}_1(\gamma_1|\mu) := \int_{\mathbb{G}} w_1(x)F_1(f_1(x))\mu(\mathrm{d}x)$$

$$\mathcal{F}_2(\gamma_2|\nu) := \int_{\mathbb{G}} w_2(x)F_2(f_2(x))\nu(\mathrm{d}x)$$

Caffarelli & McCann's Insight

EPT as a standard complete OT problem.

$$\boxed{\mathrm{ET}_\lambda(\mu,\nu) = (\mu(\mathbb{G})+\nu(\mathbb{G}))\left(\underline{\mathcal{W}_{\hat{c}}(\hat{\mu},\hat{\nu})} - b\lambda\right)}$$

where $\mathcal{W}_{\hat{c}}(\hat{\mu},\hat{\nu}) := \inf_{\hat{\gamma}\in\Pi(\hat{\mu},\hat{\nu})} \int_{\hat{\mathbb{G}}\times\hat{\mathbb{G}}} \hat{c}(x,y)\hat{\gamma}(\mathrm{d}x,\mathrm{d}y)$

$$\hat{\mu} = \frac{\mu + \nu(\mathbb{G})\delta_{\hat{s}}}{\mu(\mathbb{G})+\nu(\mathbb{G})} \qquad \hat{\nu} = \frac{\nu + \mu(\mathbb{G})\delta_{\hat{s}}}{\mu(\mathbb{G})+\nu(\mathbb{G})}$$

$$\hat{\mathbb{G}} := \mathbb{G}\cup\{\hat{s}\} \qquad \hat{s}\notin\mathbb{G}$$

$$\hat{c}(x,y) := \begin{cases} b\, d_{\mathbb{G}}(x,y) & \text{if } x,y\in\mathbb{G}, \\ w_1(x)+b\lambda & \text{if } x\in\mathbb{G} \text{ and } y=\hat{s}, \\ w_2(y)+b\lambda & \text{if } x=\hat{s} \text{ and } y\in\mathbb{G}, \\ b\lambda & \text{if } x=y=\hat{s}. \end{cases}$$

Orlicz-EPT

$$\mathcal{OE}_\Phi(\mu,\nu) := (\mu(\mathbb{G}) + \nu(\mathbb{G}))\left(\underline{\mathcal{W}_\Phi(\hat{\mu},\hat{\nu})} - b\lambda\right)$$

where $\mathcal{W}_\Phi(\hat{\mu},\hat{\nu})$ is the Orlicz-Wasserstein (OW), defined as:

$$\mathcal{W}_\Phi(\hat{\mu},\hat{\nu}) := \inf_{\tilde{\gamma}\in\Pi(\hat{\mu},\hat{\nu})} \inf \left[t > 0 : \int_{\hat{\mathbb{G}}\times\hat{\mathbb{G}}} \Phi\left(\frac{\hat{c}(x,y)}{t}\right) \mathrm{d}\tilde{\gamma}(x,y) \leq 1\right]$$

Orlicz-EPT with Binary Search

$$\mathcal{A}(t;\hat{\mu},\hat{\nu}) := \inf_{\tilde{\gamma}\in\Pi(\hat{\mu},\hat{\nu})} \int_{\hat{\mathbb{G}}\times\hat{\mathbb{G}}} \Phi\left(\frac{\hat{c}(x,y)}{t}\right) \mathrm{d}\tilde{\gamma}(x,y) \quad \text{for} \quad t>0$$

$t \in (0,+\infty) \longmapsto \mathcal{A}(t;\hat{\mu},\hat{\nu})$: monotone non-increasing

$$\mathcal{A}_\varepsilon(t;\hat{\mu},\hat{\nu}) := \inf_{\tilde{\gamma}\in\Pi(\hat{\mu},\hat{\nu})} \left[\int_{\hat{\mathbb{G}}\times\hat{\mathbb{G}}} \Phi\left(\frac{\hat{c}(x,y)}{t}\right) \mathrm{d}\tilde{\gamma}(x,y) - \varepsilon H(\tilde{\gamma})\right]$$

where $H(\tilde{\gamma}) := -\int_{\hat{\mathbb{G}}\times\hat{\mathbb{G}}} (\log\tilde{\gamma}(x,y) - 1)\mathrm{d}\tilde{\gamma}(x,y)$

$t \in (0,+\infty) \longmapsto \mathcal{A}_\varepsilon(t;\hat{\mu},\hat{\nu})$: monotone non-increasing

Orlicz-Sobolev Transport

Orlicz-Sobolev Transport: A scalable variant of Orlicz-EPT.

- We leverage dual EPT problem, and develop a novel regularization approach.
- OST adopts Orlicz geometric structure used in Orlicz-EPT, but offers a much more efficient computation.

Dual EPT

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

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

Key insight:

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

$$\begin{aligned} \int_{[z_0,x]} f'(y)\omega(\mathrm{d}y) &\le 2\left\|f'\right\|_{L_\Psi} \left\|\mathbf{1}_{[z_0,x]}\right\|_{L_\Phi} \\ &\le 2b\left\|\mathbf{1}_{[z_0,x]}\right\|_{L_\Phi} \\ &\le \frac{2b}{\Phi^{-1}(1/\omega(\mathbb{G}))} \end{aligned}$$

Orlicz-Sobolev Transport (OST)

$$\mathcal{OS}_{\Phi,\alpha}(\mu,\nu) := \sup_{f\in\mathbb{U}_{\Psi,\alpha}} \left[\int_{\mathbb{G}} f(x)\mu(\mathrm{d}x) - \int_{\mathbb{G}} f(x)\nu(\mathrm{d}x)\right]$$

$$\mathbb{U}_{\Psi,\alpha} := \left\{ f \in WL_\Psi(\mathbb{G},\omega) : \ \|f'\|_{L_\Psi} \leq b, f(z_0) \in \mathcal{I}_\alpha \right\}$$

$$\mathcal{I}_\alpha := \left[-w_2(z_0) - \frac{b\lambda}{2} + \alpha, w_1(z_0) + \frac{b\lambda}{2} - \alpha \right]$$

$$\alpha \in [0, \frac{1}{2}(b\lambda + w_1(z_0) + w_2(z_0))]$$

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

Univariate Optimization for OST

$$\mathcal{OS}_{\Phi,\alpha}(\mu,\nu) = \inf_{k>0} \frac{1}{k}\left(1 + \sum_{e\in E} w_e \Phi(kb\,|\mu(\gamma_e) - \nu(\gamma_e)|)\right) \\ +\Theta|\mu(\mathbb{G}) - \nu(\mathbb{G})|$$

$$\Theta := \begin{cases} w_1(z_0) + \frac{b\lambda}{2} - \alpha & \text{if} \quad \mu(\mathbb{G}) \geq \nu(\mathbb{G}), \\ w_2(z_0) + \frac{b\lambda}{2} - \alpha & \text{if} \quad \mu(\mathbb{G}) < \nu(\mathbb{G}). \end{cases}$$

Theoretical Properties

Theorem (informal): $\mathcal{OS}_{\Phi,\alpha}$ is a metric if $w_1(z_0) = w_2(z_0)$.

⋆ Connection of OST with GST $\mathcal{GS}_{\Phi}$:

For $\mu(\mathbb{G}) = \nu(\mathbb{G}), b = 1$, then $\mathcal{OS}_{\Phi,\alpha}(\mu,\nu) = \mathcal{GS}_{\Phi}(\mu,\nu)$

⋆ Connection of OST with ST $\mathcal{S}_p$:

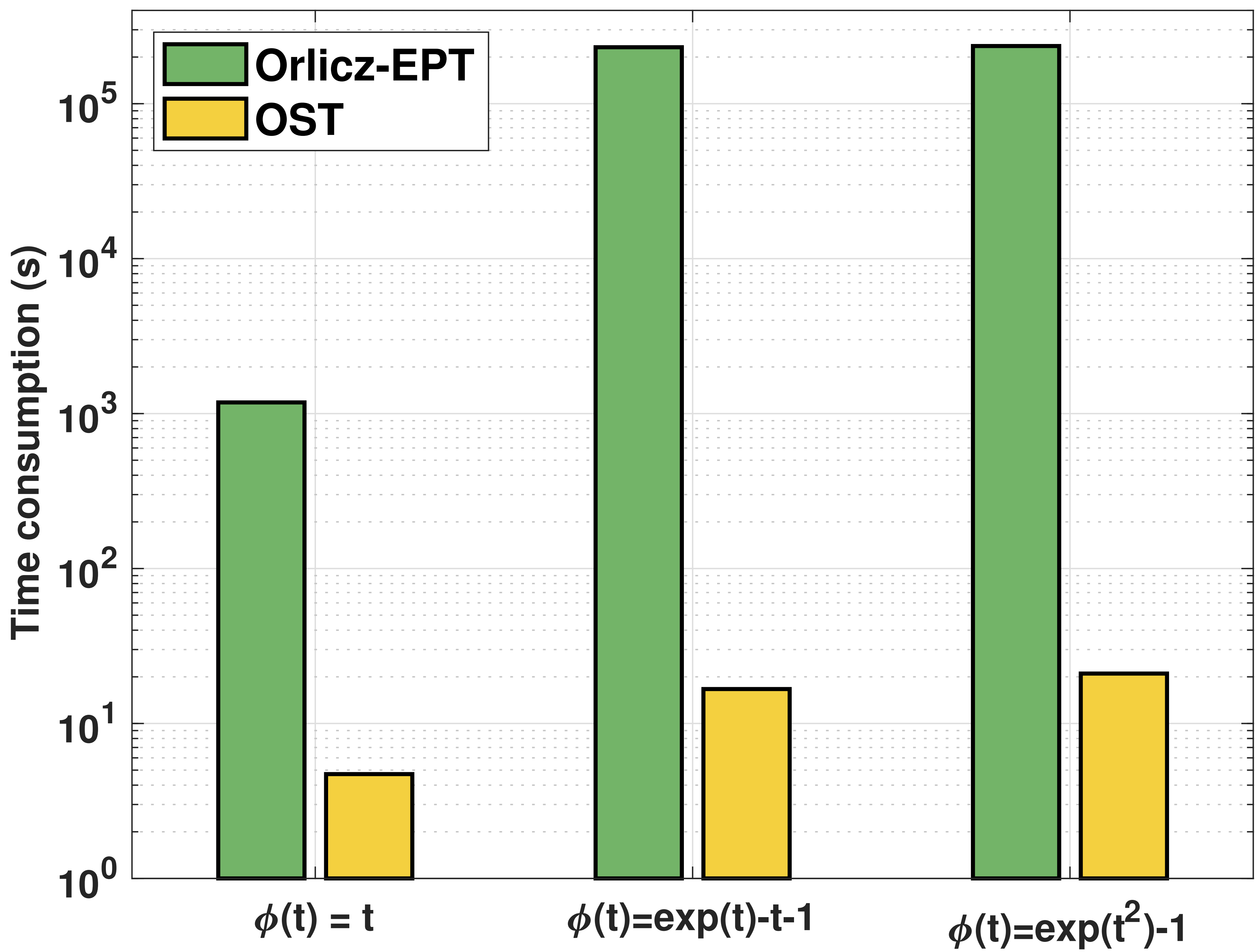
For $\mu(\mathbb{G}) = \nu(\mathbb{G}), b = 1, \Phi(t) = \dfrac{(p-1)^{p-1}}{p^p} t^p$

$$\mathcal{OS}_{\Phi,\alpha}(\mu,\nu) = \mathcal{S}_p(\mu,\nu)$$

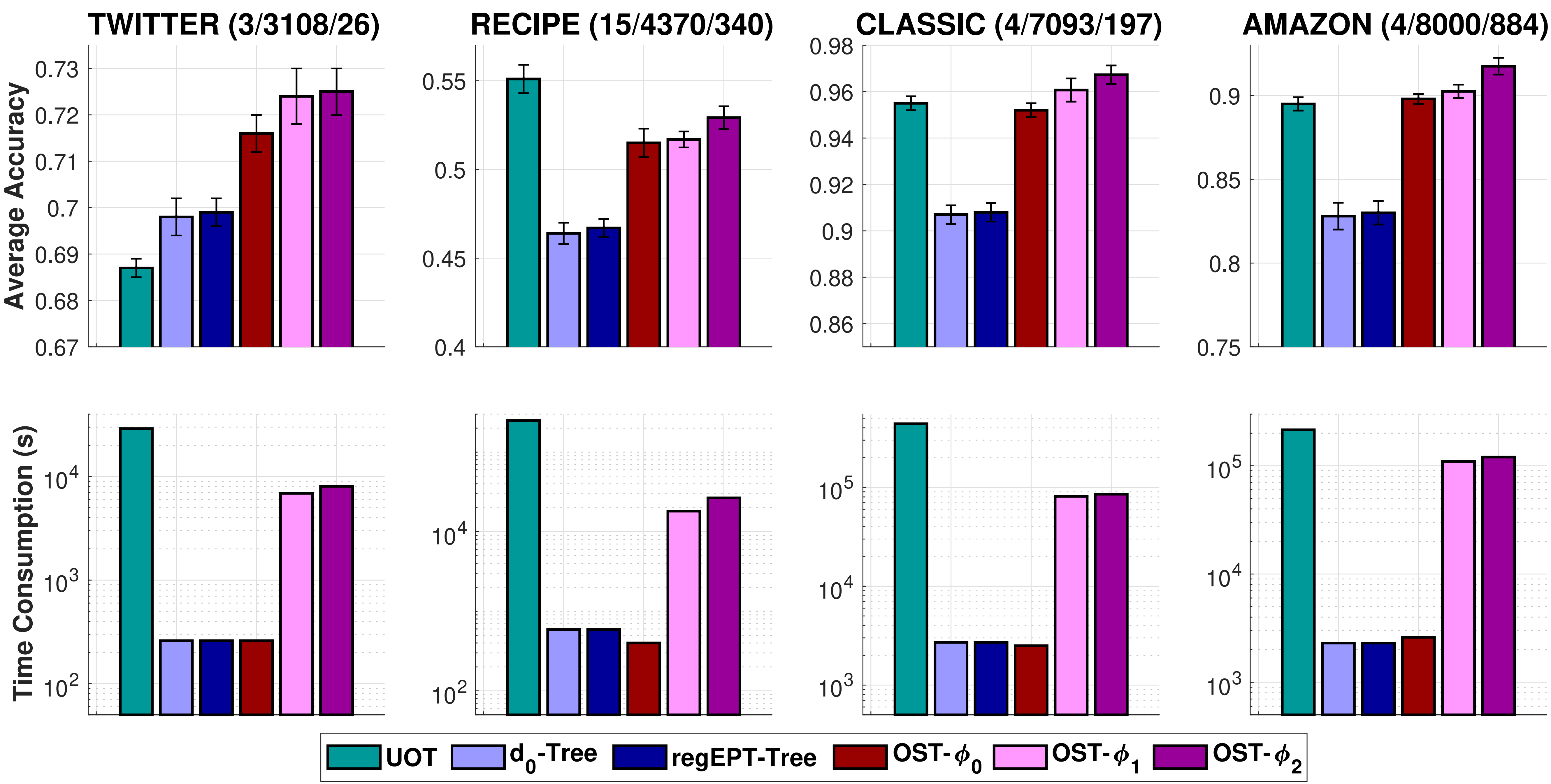
⋆ Connection of OST with UST $\mathcal{US}_{p,\alpha}$:

For $\Phi(t) = \dfrac{(p-1)^{p-1}}{p^p} t^p$, then $\mathcal{OS}_{\Phi,\alpha}(\mu,\nu) = \mathcal{US}_{p,\alpha}(\mu,\nu)$

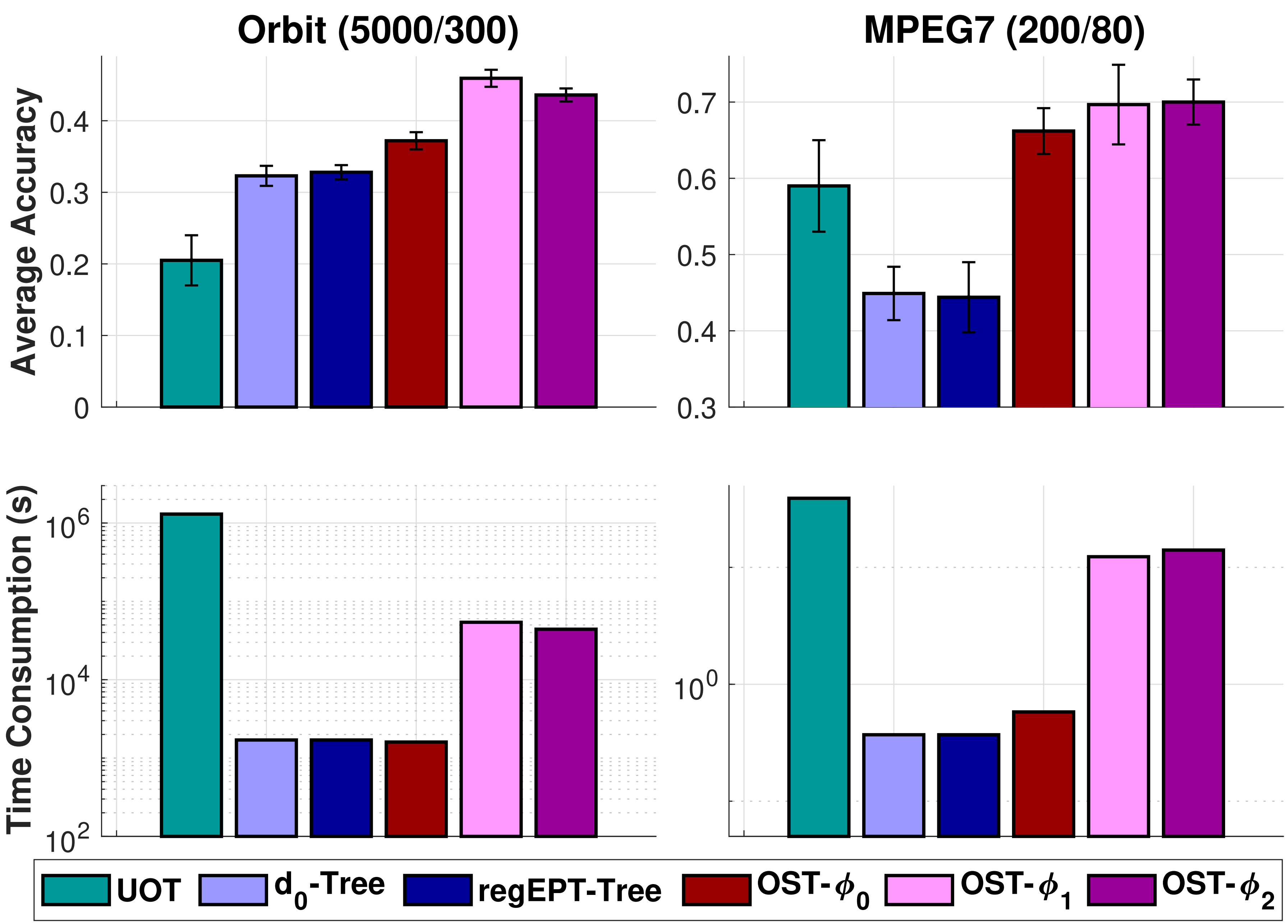
Time Consumption



Empirical Results

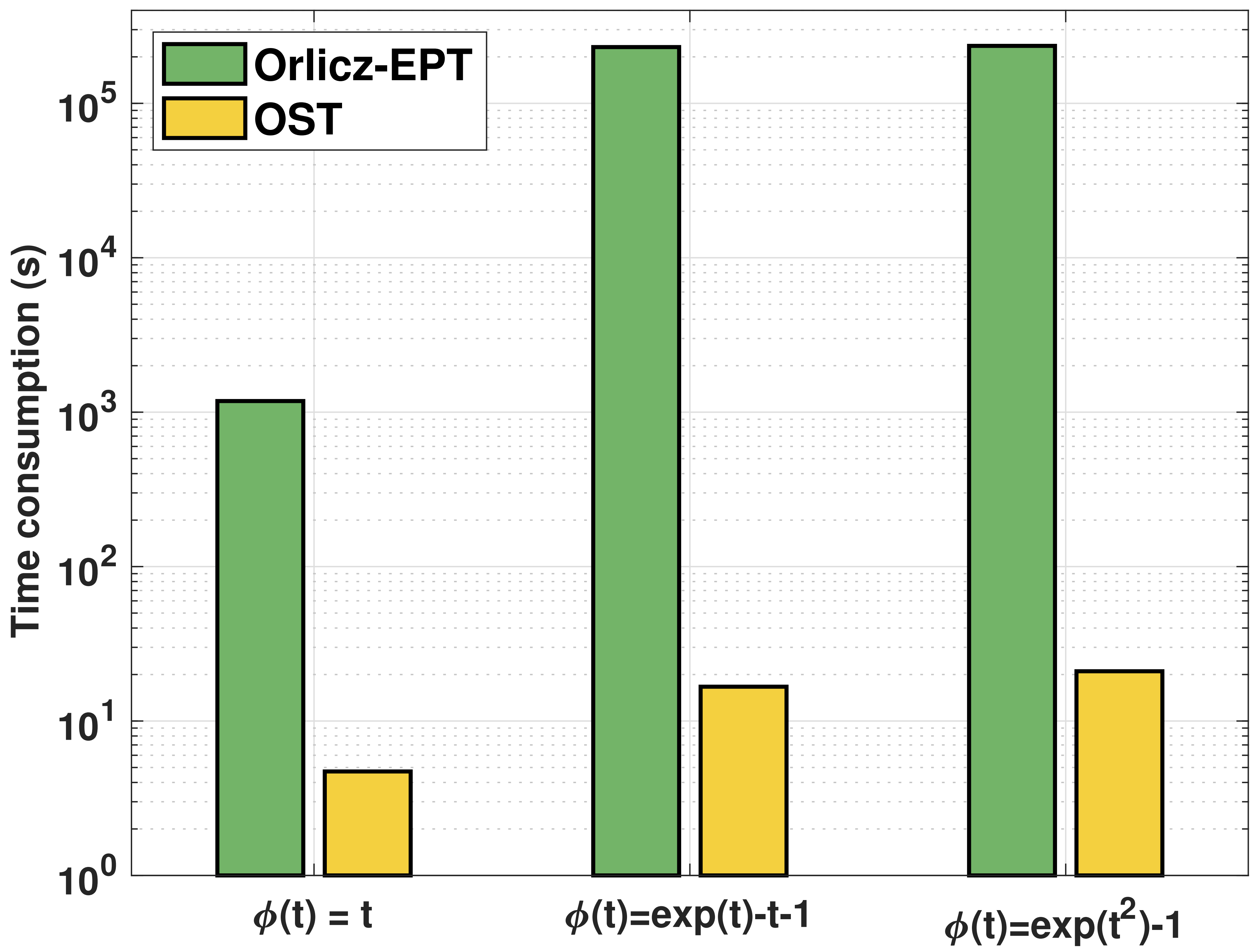


Empirical Results

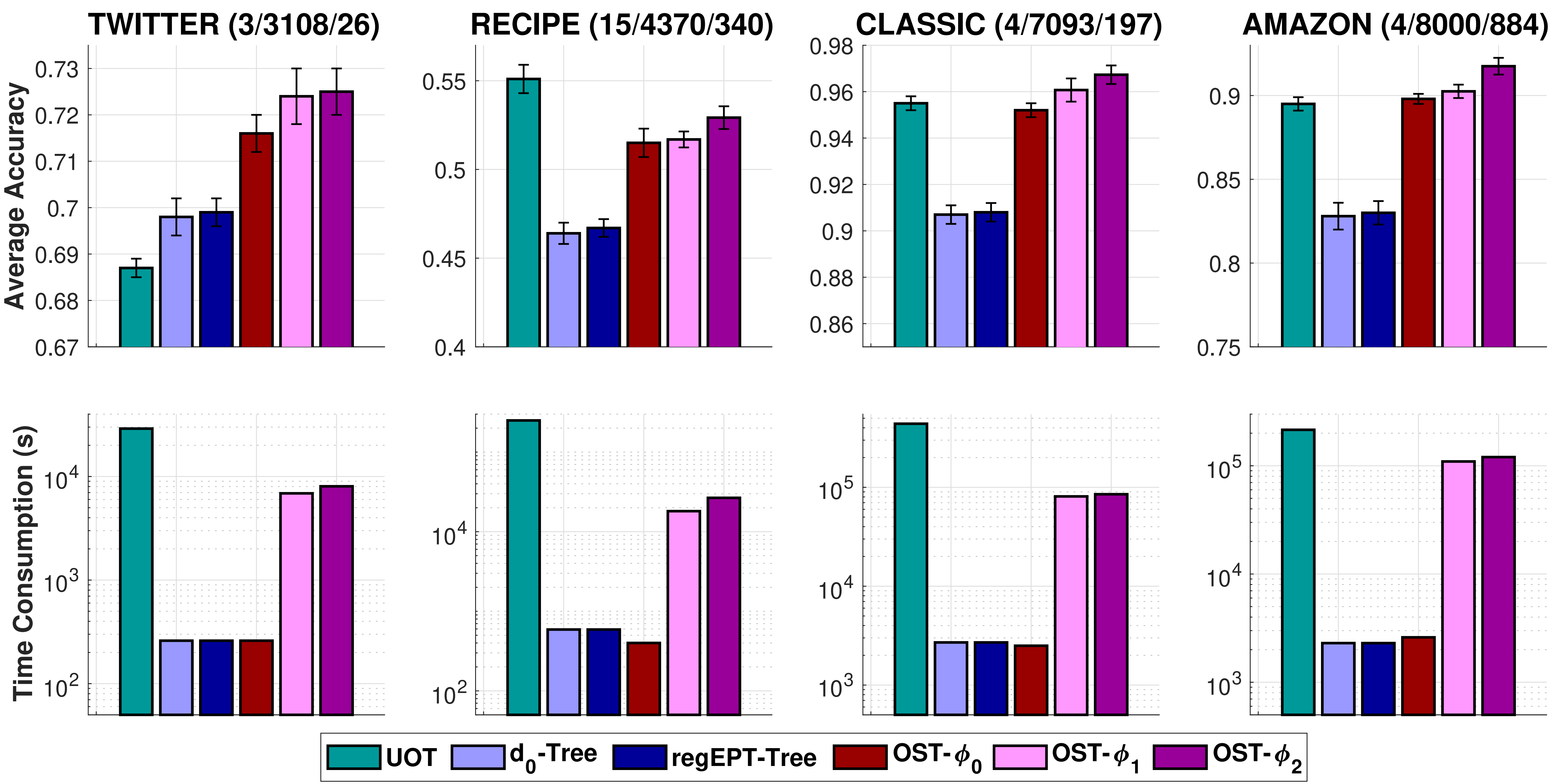


Thank you
for your attention

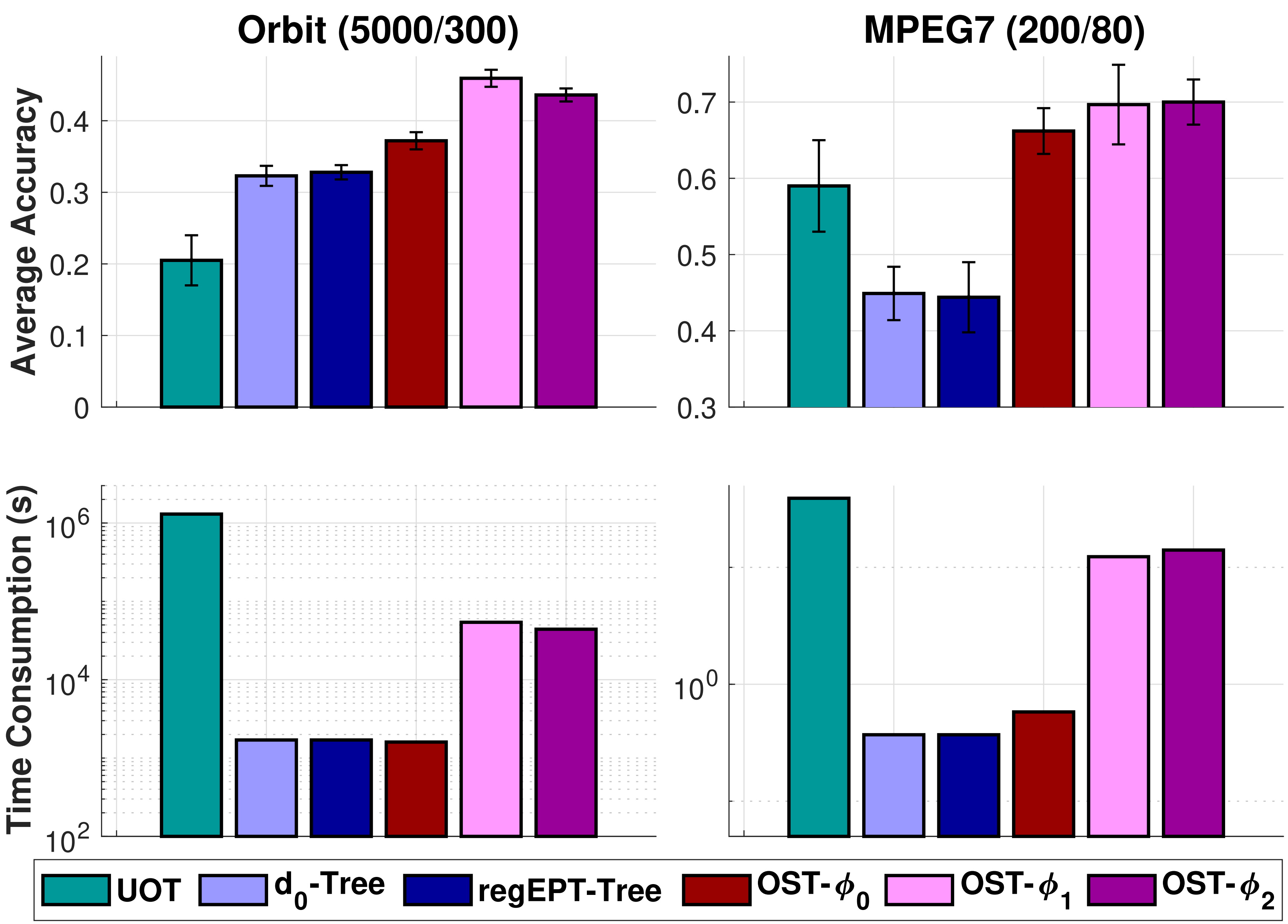
Empirical Results (G_Sqrt)



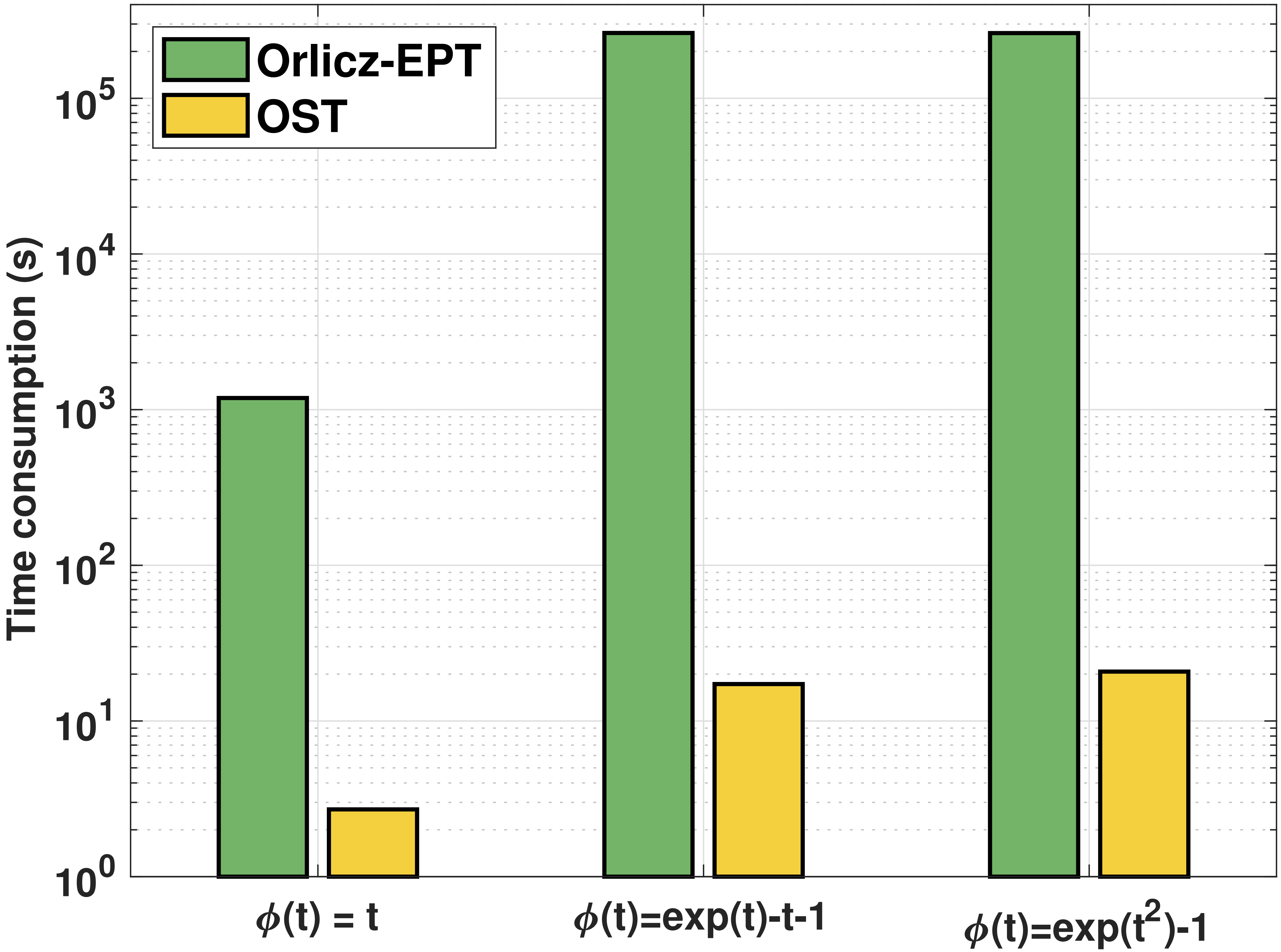
Empirical Results (G_Sqrt)



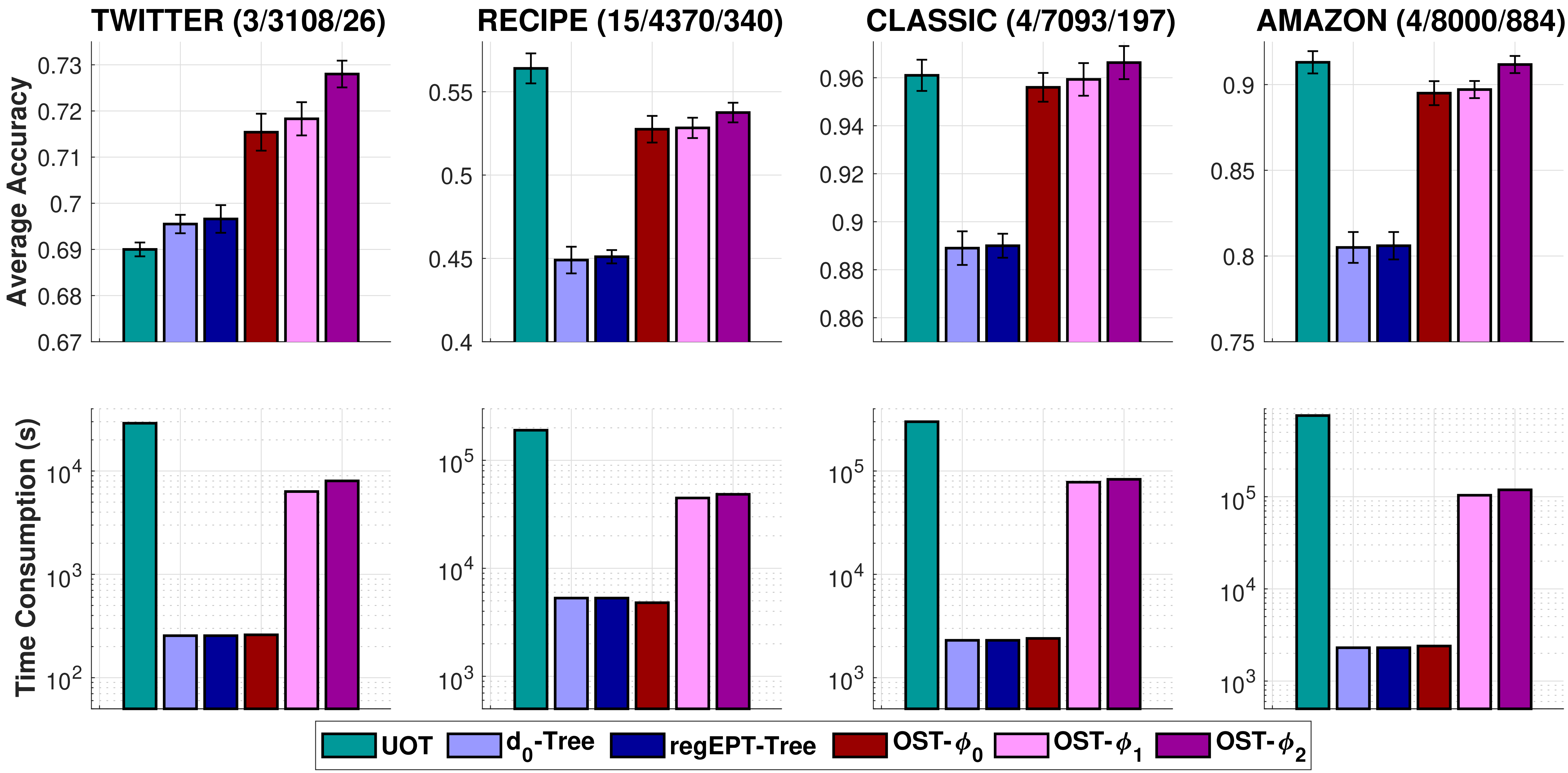
Empirical Results (G_Sqrt)



Empirical Results (G_Log)



Empirical Results (G_Log)



Empirical Results (G_Log)