On Transferring Transferability: Towards a Theory for Size Generalization

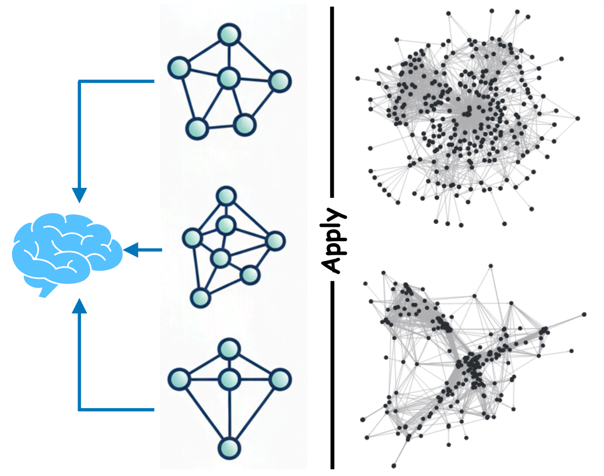
Eitan Levin (Caltech), Yuxin Ma (JHU), Mateo Díaz (JHU), Soledad Villar (JHU)

NeurIPS arXiv

Motivation

Motivation: Size Generalization Problem

- Any-dimensional models: neural networks that accept inputs of arbitrary size (e.g. GNN, DeepSet, PointNet)
- Goal: train on small-sized instances and generalize to large-sized ones

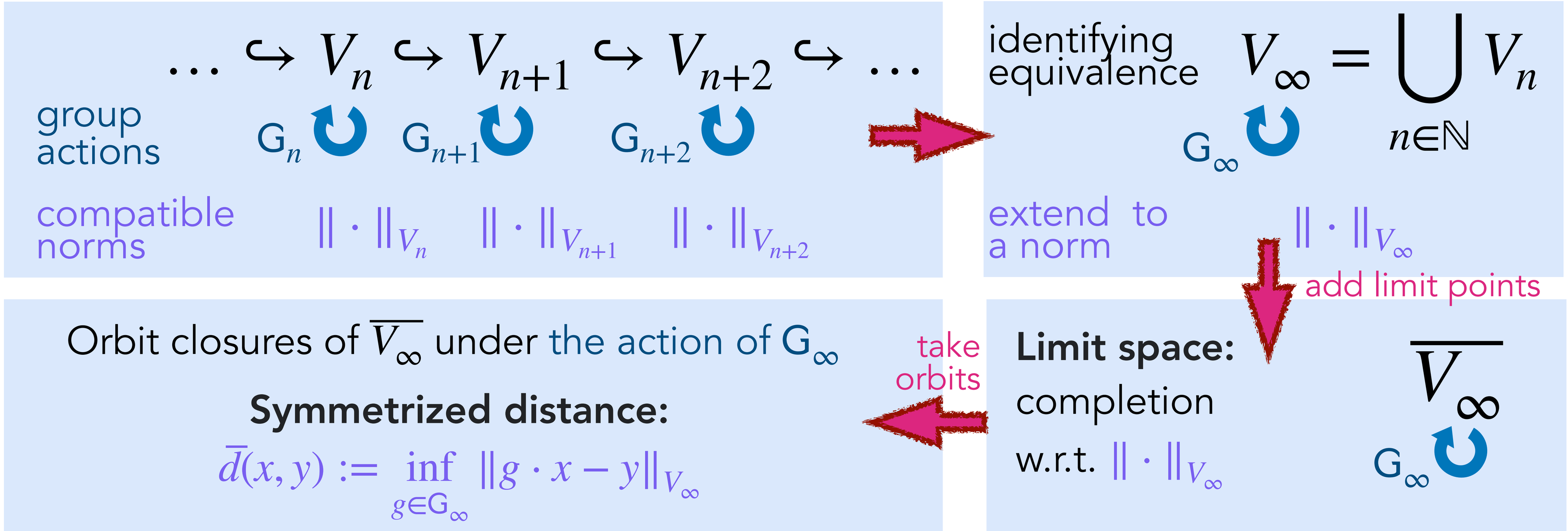


What properties of the model, data, and learning task ensure that learning performance transfers well across dimensions?

Prior Work: GNN Transferability under the Graphon Framework [1]

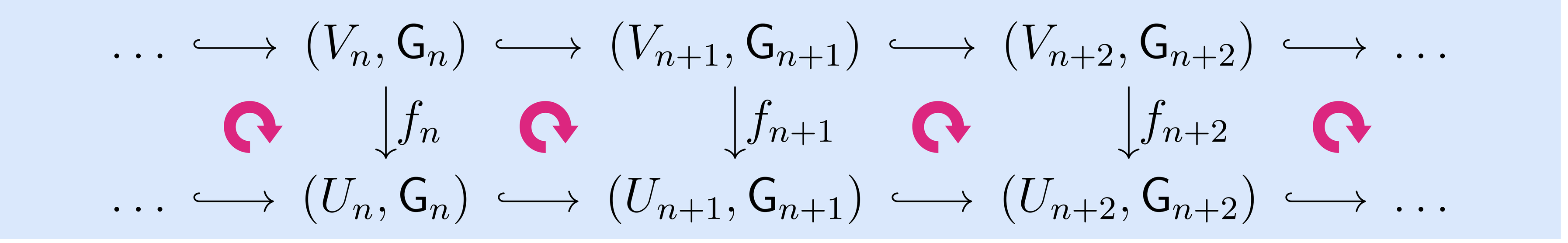
General Framework for Transferability

Equivalence of Differently-Sized Objects [2]



Compatibility and Transferability of $(f_n : V_n \to U_n)$

- **Compatibility = extension to the limit**



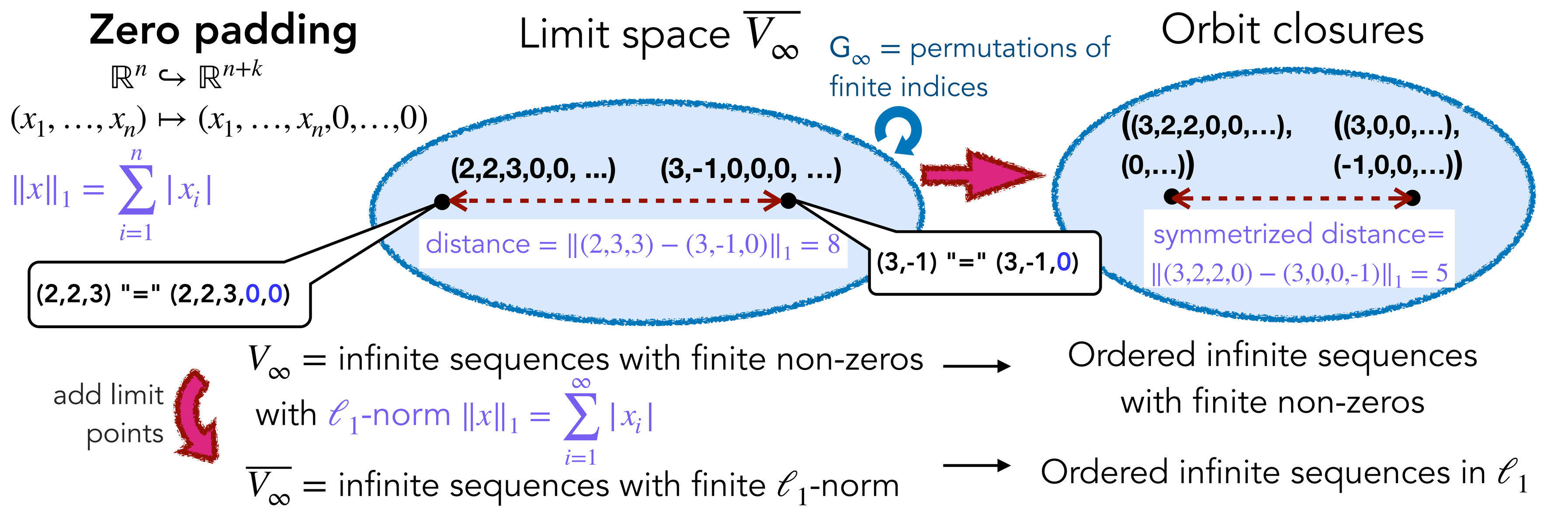
$\Updownarrow$

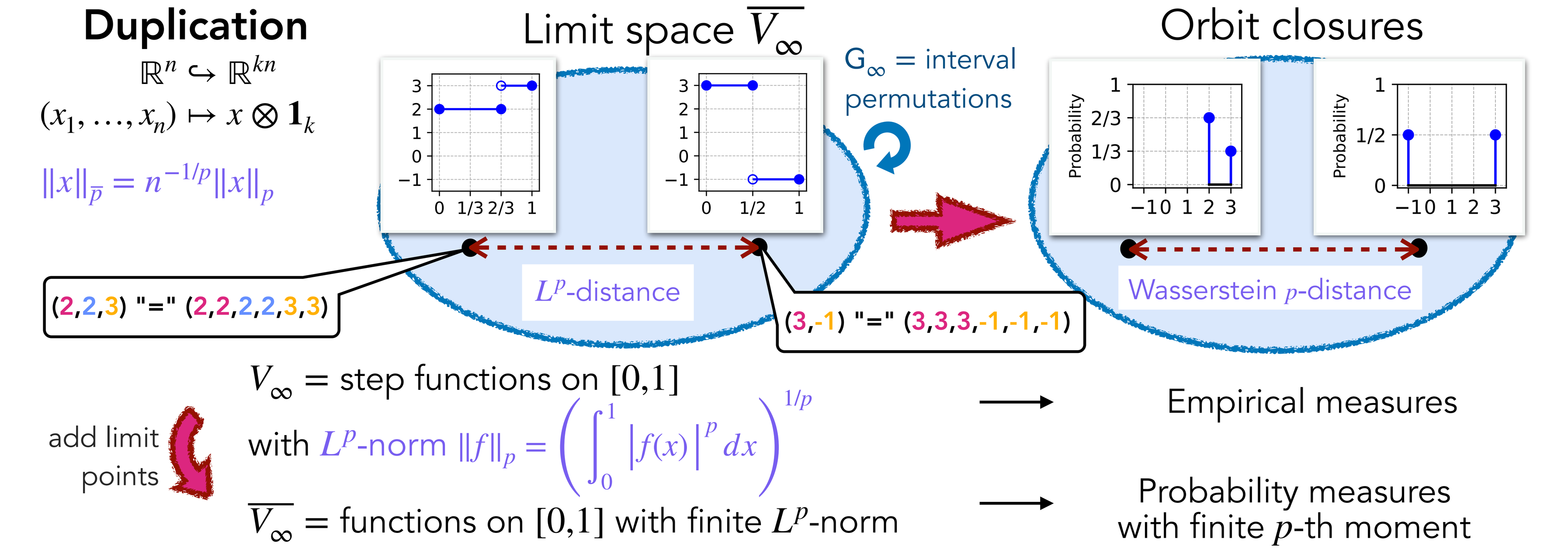
Exists $f_\infty : V_\infty \to U_\infty$ s.t. $f_\infty|_{V_n} = f_n$ for all n

- **Transferability = continuity in the limit**

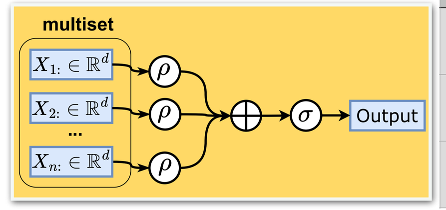
$f_n(x_n) = f_\infty(x_n) \approx f_\infty(x_m) = f_m(x_m)$ if $x_n \approx x_m$

Example: Sets

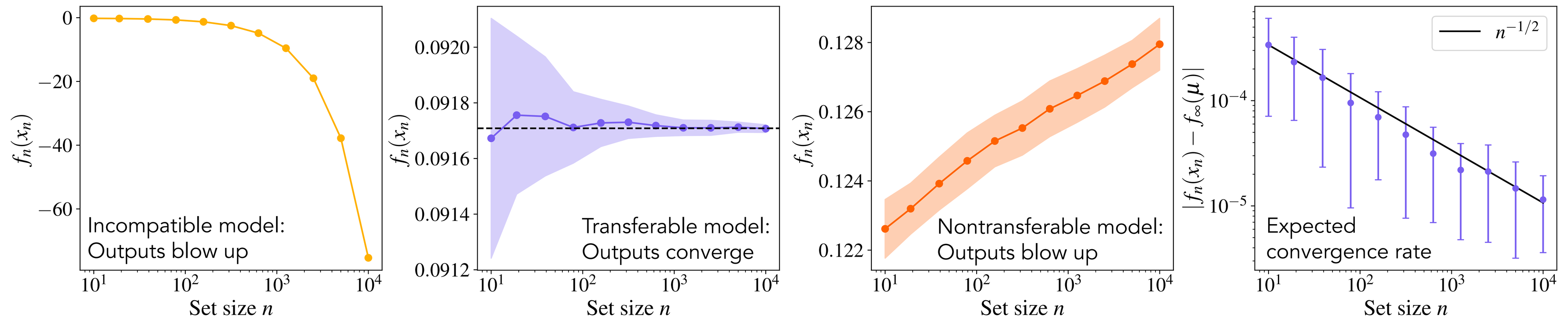




Neural Networks on Sets



Networks on Sets	Zero-padding with $\|\cdot\|_1$ norm	Duplication with $\|\cdot\|_{\bar{p}}$ norm
DeepSet (sum)	**Transferable** if $\rho(0) = 0$	Incompatible
$\overline{\text{DeepSet}}$ (mean)	Incompatible	**Transferable**
PointNet (max)	Incompatible	**Compatible**; **Transferable** if $p = \infty$



(a) DeepSet outputs (b) $\overline{\text{DeepSet}}$ outputs (c) PointNet outputs (d) $\overline{\text{DeepSet}}$ errors

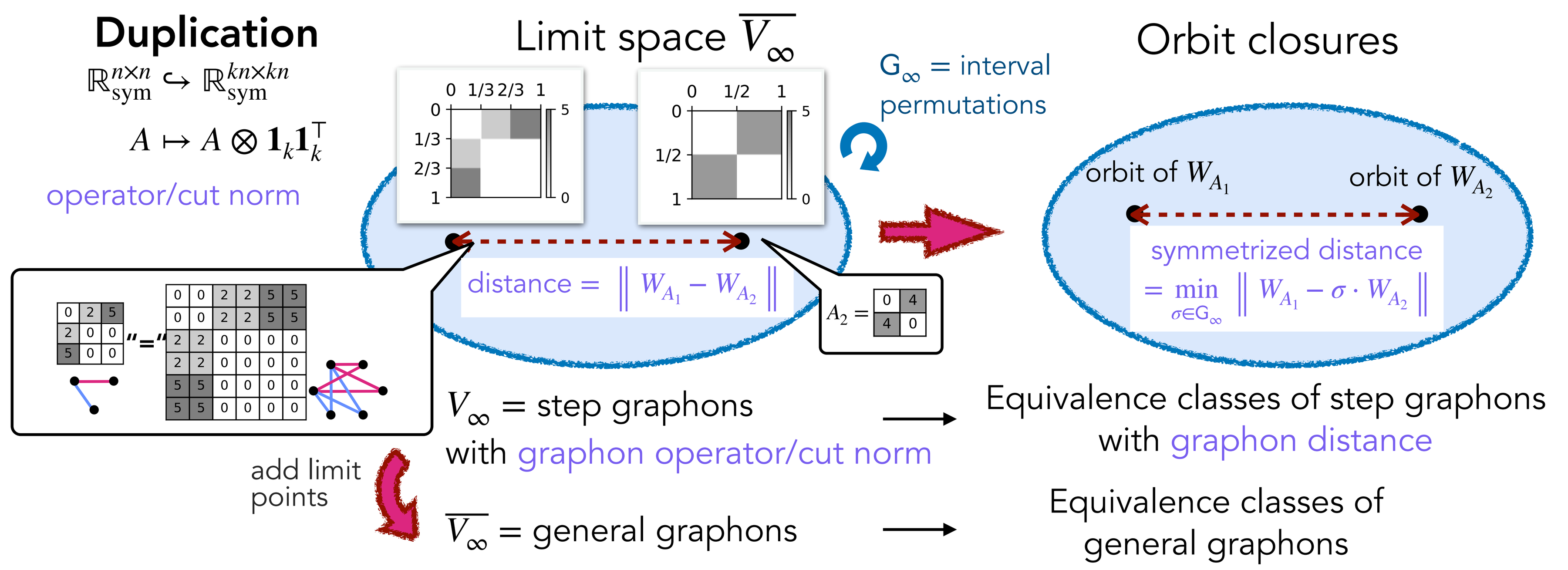
Bibliography

[1] Ruiz, Luana, Luiz Chamon, and Alejandro Ribeiro. **"Graphon neural networks and the transferability of graph neural networks."** *NeurIPS 2020.*

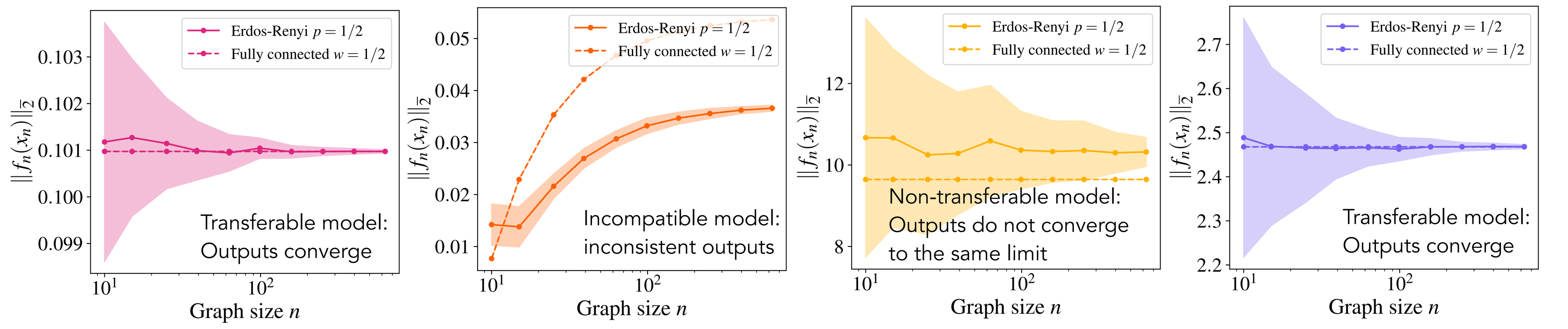
[2] Levin, Eitan, and Mateo Díaz. **"Any-dimensional equivariant neural networks."** *AISTATS 2024.*

[3] Maron, Haggai, et al. **"Invariant and equivariant graph networks."** *ICLR 2019.*

Example: Graphs



- **MPNNs**: **transferable** under constraints
- Modifying **Invariant Graph Network (IGN)** [3]
  - **IGN**: alternating equivariant linear layers and pointwise nonlinearities
  - **Generalizable GNN (GGNN)**: enforce compatible linear layers; use message-passing-like nonlinearity ⇒ **compatible** [2]
  - **Continuous GGNN**: restrict linear layers with bounded operator norm ⇒ **transferable**

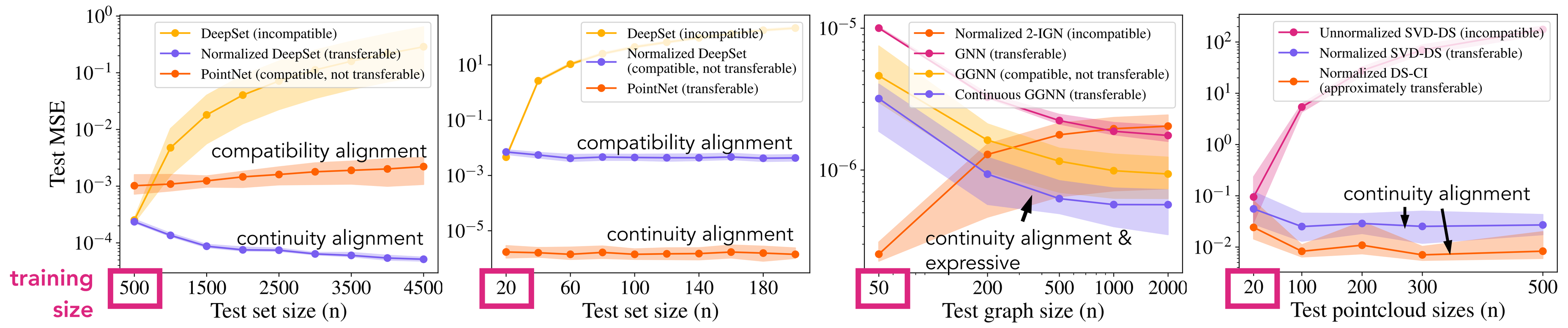


(a) GNN [1] outputs (b) 2-IGN [3] outputs (c) GGNN outputs (d) Continuous GGNN outputs

Transferability Implies Size Generalization

- Size generalization depends on **task-model alignment**
- Asymptotic guarantee
  - **Test on "infinite size":** $(x,y) \sim \mu$ in the limit space $\overline{V_\infty} \times \overline{U_\infty}$
  - **Train on small size:** $s = (x_i, y_i)_{i=1}^N \sim \mu_n$ in the size-n space $V_n \times U_n$, μ_n is induced from μ

$$\underbrace{\left|\frac{1}{N}\sum_{i=1}^N \ell(\mathscr{A}_s(x_i), y_i)\right.}_{\text{training error}} - \underbrace{\left.\mathbb{E}_{(x,y)\sim\mu}\ell(\mathscr{A}_s(x), y)\right|}_{\text{test error}} \le \underbrace{\left|\frac{1}{N}\sum_{i=1}^N \ell(\mathscr{A}_s(x_i), y_i) - \mathbb{E}_{(x,y)\sim\mu_n}\ell(\mathscr{A}_s(x), y)\right|}_{\substack{\text{in-distribution generalization} \\ \le g(N) \to 0 \text{ as \#data } N \to \infty}} + \underbrace{\left|\mathbb{E}_{(x,y)\sim\mu_n}\ell(\mathscr{A}_s(x), y) - \mathbb{E}_{(x,y)\sim\mu}\ell(\mathscr{A}_s(x), y)\right|}_{\substack{\text{distribution shift} \\ \le C\cdot W_1(\mu,\mu_n) \to 0 \\ \text{as training size } n \to \infty}}$$



Task 1: (Set) Population Statistics; Task 2: (Set) Maximal distance from origin; Task 3: (Graph) Weighted triangle density; Task 4: (Point Cloud) Gromov-Wasserstein lower bound