On Transferring Transferability: Towards a Theory for Size Generalization

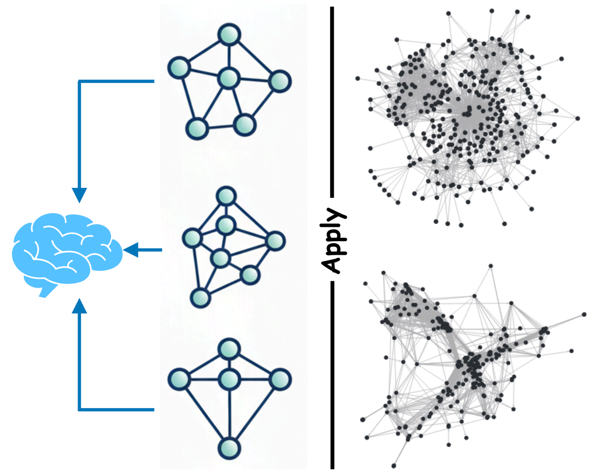
Eitan Levin (Caltech), Yuxin Ma (JHU), Mateo Díaz (JHU), Soledad Villar (JHU)

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]

$$\cdots \hookrightarrow V_n \hookrightarrow V_{n+1} \hookrightarrow V_{n+2} \hookrightarrow \cdots$$

group actions: G_n, G_{n+1}, G_{n+2}

compatible norms: $\|\cdot\|_{V_n}$, $\|\cdot\|_{V_{n+1}}$, $\|\cdot\|_{V_{n+2}}$

identifying equivalence: $V_\infty = \bigcup_{n\in\mathbb{N}} V_n$, G_∞

extend to a norm: $\|\cdot\|_{V_\infty}$

add limit points

Limit space: completion w.r.t. $\|\cdot\|_{V_\infty}$: $\overline{V_\infty}$, G_∞

take orbits

Orbit closures of $\overline{V_\infty}$ under the action of G_∞

Symmetrized distance:

$$\bar{d}(x,y) := \inf_{g\in G_\infty} \|g\cdot x - y\|_{V_\infty}$$

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

- **Compatibility** = extension to the limit

$$\begin{tikzcd}
\cdots \arrow[r] & (V_n, G_n) \arrow[r] \arrow[d, "f_n"] & (V_{n+1}, G_{n+1}) \arrow[r] \arrow[d, "f_{n+1}"] & (V_{n+2}, G_{n+2}) \arrow[r] \arrow[d, "f_{n+2}"] & \cdots \\
\cdots \arrow[r] & (U_n, G_n) \arrow[r] & (U_{n+1}, G_{n+1}) \arrow[r] & (U_{n+2}, G_{n+2}) \arrow[r] & \cdots
\end{tikzcd}$$

$$\Updownarrow$$

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

- **Transferability** = continuity in the limit

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

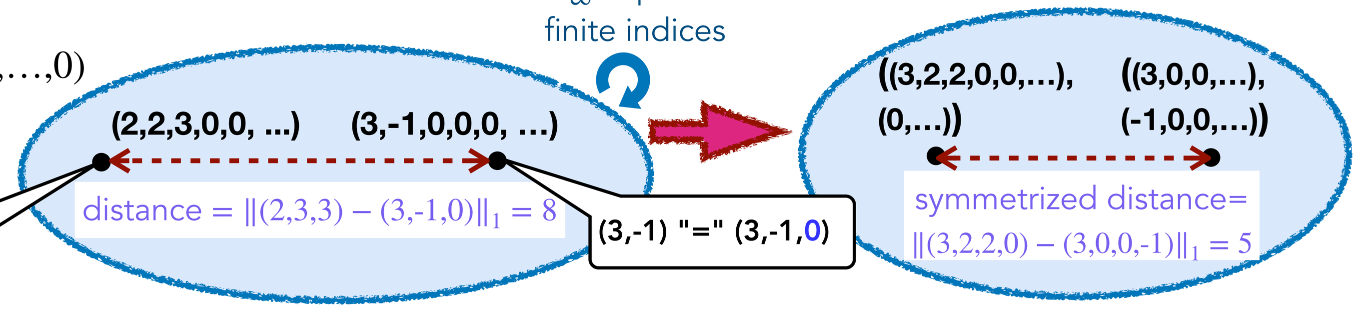
Example: Sets

Zero padding

$\mathbb{R}^n \hookrightarrow \mathbb{R}^{n+k}$

$(x_1,\ldots,x_n) \mapsto (x_1,\ldots,x_n,0,\ldots,0)$

$\|x\|_1 = \sum_{i=1}^n |x_i|$



V_∞ = infinite sequences with finite non-zeros with ℓ_1-norm $\|x\|_1 = \sum_{i=1}^\infty |x_i|$ ⟶ Ordered infinite sequences with finite non-zeros

add limit points

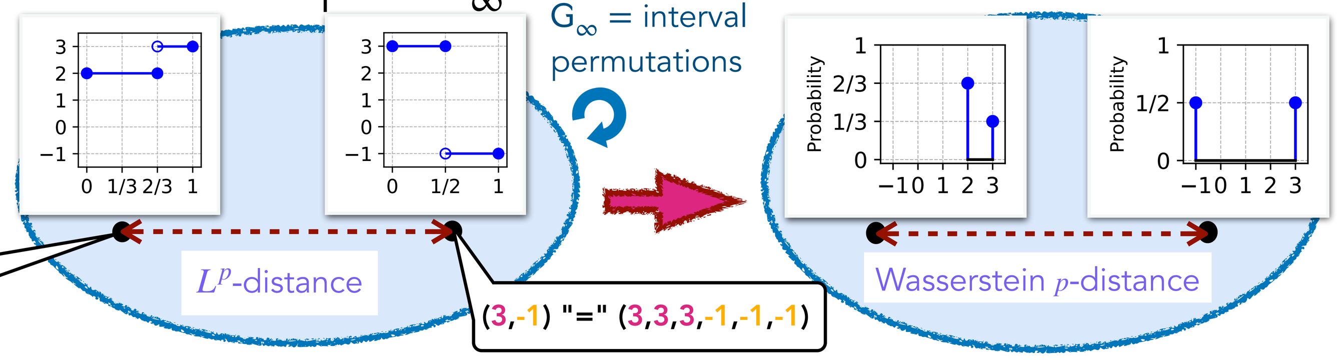
$\overline{V_\infty}$ = infinite sequences with finite ℓ_1-norm ⟶ Ordered infinite sequences in ℓ_1

Duplication

$\mathbb{R}^n \hookrightarrow \mathbb{R}^{kn}$

$(x_1,\ldots,x_n) \mapsto x \otimes \mathbf{1}_k$

$\|x\|_{\bar{p}} = n^{-1/p}\|x\|_p$

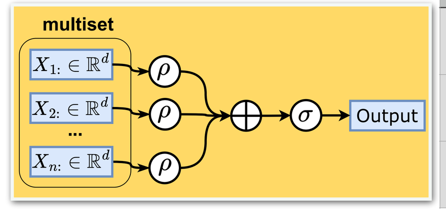


V_∞ = step functions on $[0,1]$ with L^p-norm $\|f\|_p = \left(\int_0^1 |f(x)|^p\,dx\right)^{1/p}$ ⟶ Empirical measures

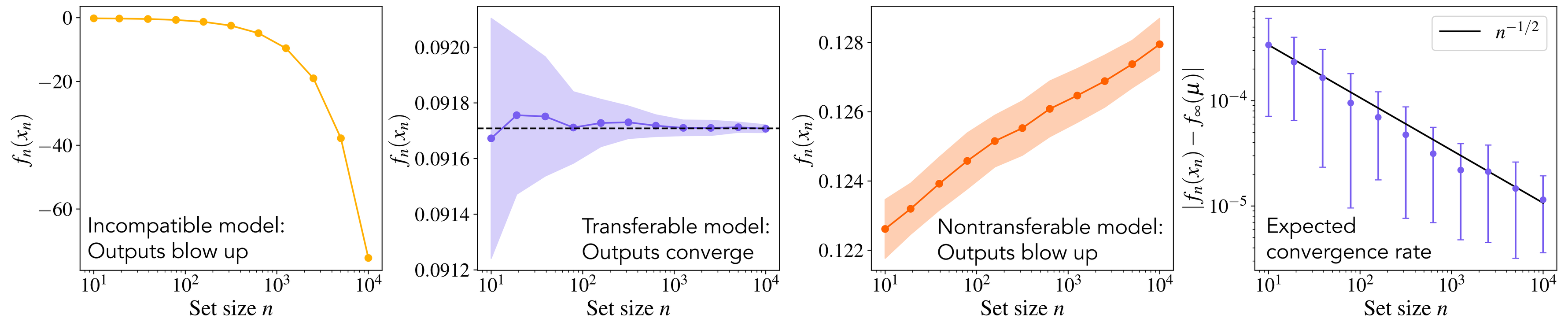
add limit points

$\overline{V_\infty}$ = functions on $[0,1]$ with finite L^p-norm ⟶ Probability measures with finite p-th moment

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.*

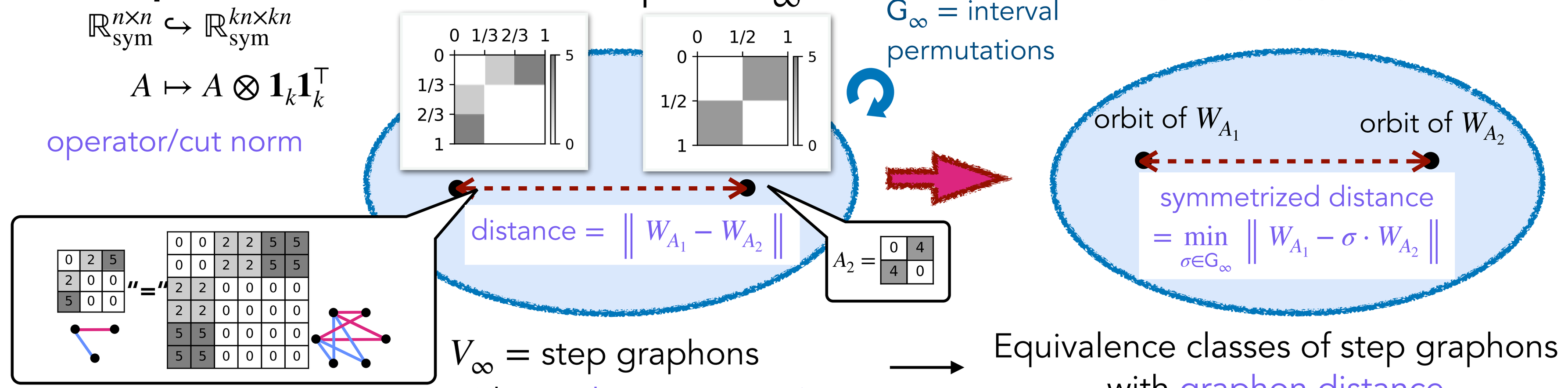
Example: Graphs

Duplication

$\mathbb{R}^{n\times n}_{\mathrm{sym}} \hookrightarrow \mathbb{R}^{kn\times kn}_{\mathrm{sym}}$

$A \mapsto A \otimes \mathbf{1}_k\mathbf{1}_k^\top$

operator/cut norm

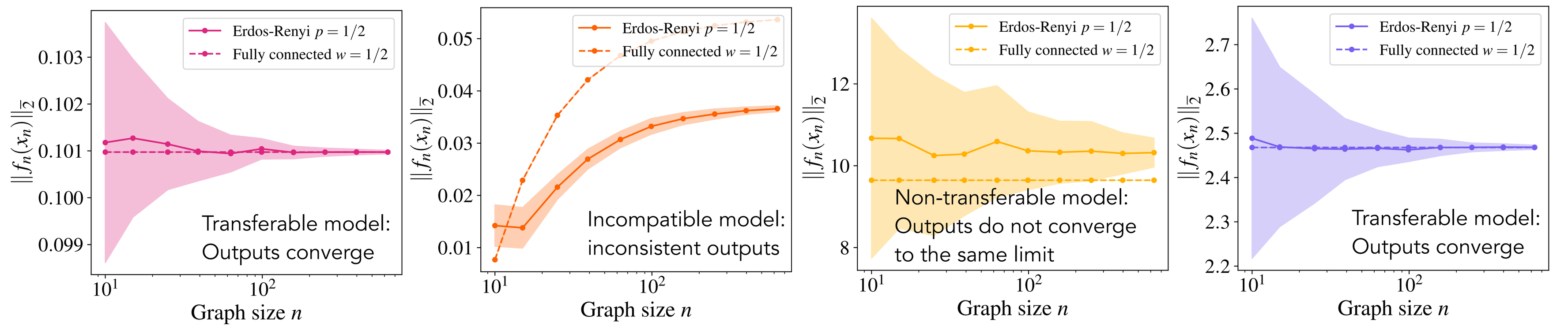


V_∞ = step graphons with graphon operator/cut norm ⟶ Equivalence classes of step graphons with graphon distance

add limit points

$\overline{V_\infty}$ = general graphons ⟶ Equivalence classes of general graphons

- **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

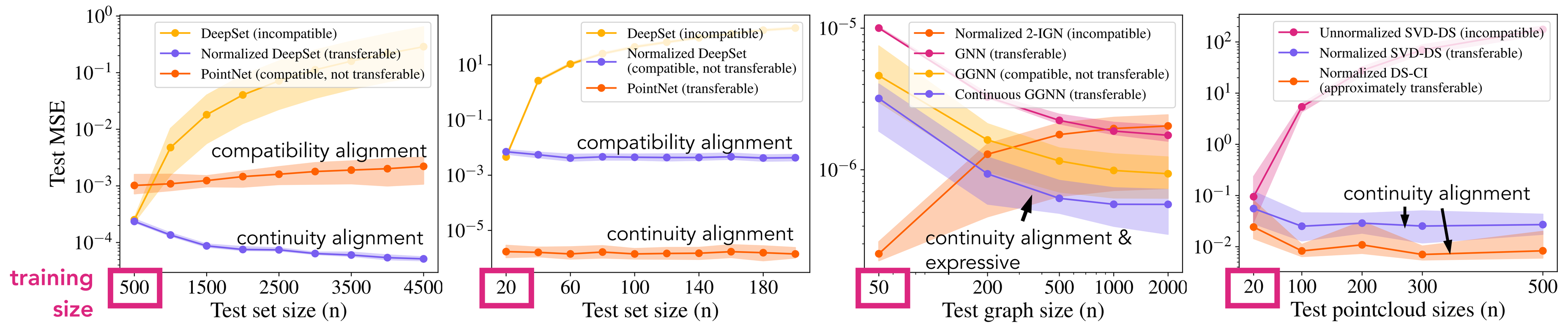
Transferability Implies Size Generalization

- 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

$$\left|\underbrace{\frac{1}{N}\sum_{i=1}^N \ell(\mathcal{A}_s(x_i),y_i)}_{\text{training error}} - \underbrace{\mathbb{E}_{(x,y)\sim\mu}\ell(\mathcal{A}_s(x),y)}_{\text{test error}}\right| \le \underbrace{\left|\frac{1}{N}\sum_{i=1}^N \ell(\mathcal{A}_s(x_i),y_i) - \mathbb{E}_{(x,y)\sim\mu_n}\ell(\mathcal{A}_s(x),y)\right|}_{\text{in-distribution generalization}} + \underbrace{\left|\mathbb{E}_{(x,y)\sim\mu_n}\ell(\mathcal{A}_s(x),y) - \mathbb{E}_{(x,y)\sim\mu}\ell(\mathcal{A}_s(x),y)\right|}_{\text{distribution shift}}$$

in-distribution generalization $\le g(N)\to 0$ as #data $N\to\infty$

distribution shift $\le C\cdot W_1(\mu,\mu_n)\to 0$ 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