# Approximately Equivariant Graph Networks

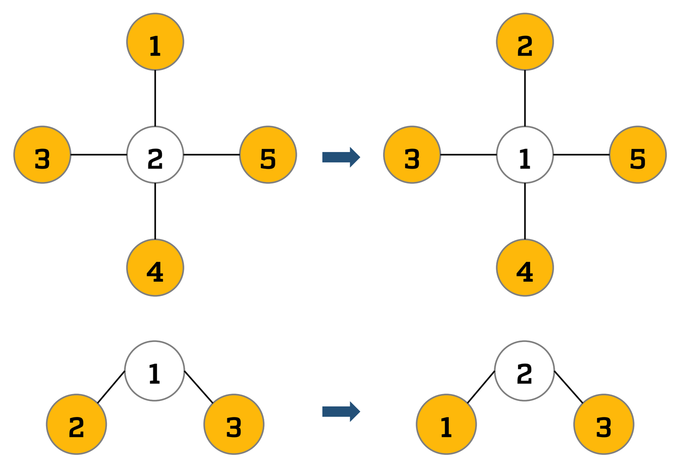
Ningyuan (Teresa) Huang [1], Ron Levie [2], Soledad Villar [1]

[1] Johns Hopkins University, [2] Technion – Israel Institute of Technology

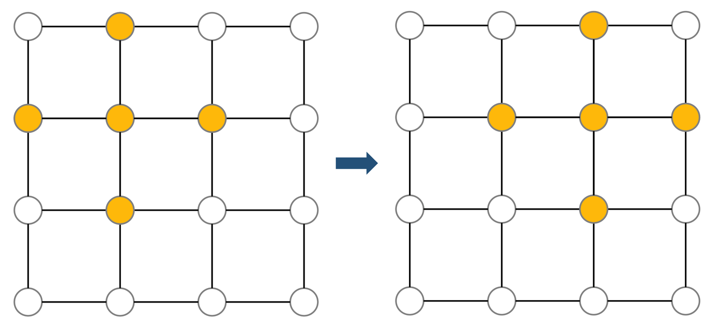
## Symmetry in Graph Learning: Passive Symmetry and Active Symmetry

### Graph: Input or Domain?

**Permutation invariance**



**Translation invariance**



**GNN Symmetry**

- Domain: different graphs
- Signal: X: $\Omega \to \mathbb{R}^k$
- The group acts on signals and domain together

**CNN Symmetry**

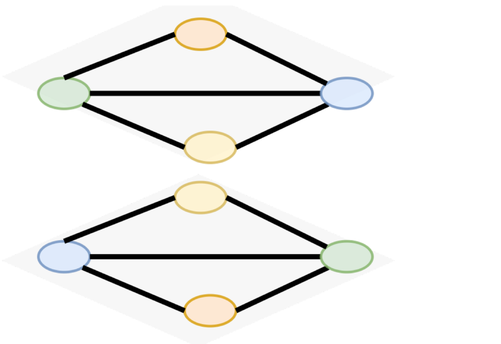
- Domain: grid $\Omega = \mathbb{Z}/n\mathbb{Z} \times \mathbb{Z}/n\mathbb{Z}$
- Signal: X: $\Omega \to \mathbb{R}^3$
- The group shifts the signals, keeping the domain fixed

### Passive Sym.

$$f(\Pi A \Pi^\top, \Pi X) = \Pi f(A, X)$$

for all permutations $\Pi \in \mathcal{S}_N$

Example: Node Relabeling

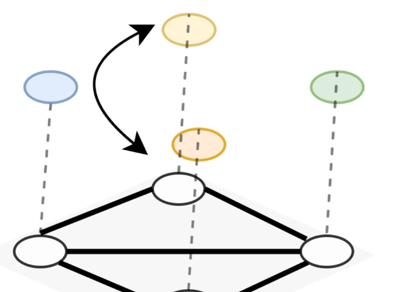


### Active Sym.

$$f(\Pi X) \approx \Pi f(X)$$

for permutations $\Pi \in \mathcal{G} \subseteq \mathcal{S}_N$

Example: Signal Swapping



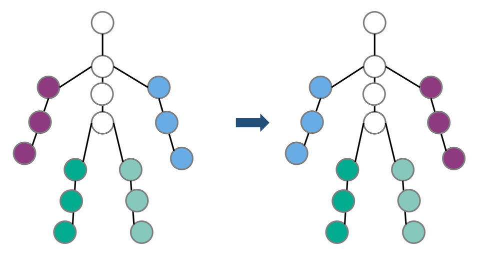
**Our contribution: enforce active symmetries in GNN**

## Learning on a Fixed Domain with Approximately Equivariant Graph Networks

### Equivariant Graph Networks using Graph Automorphisms

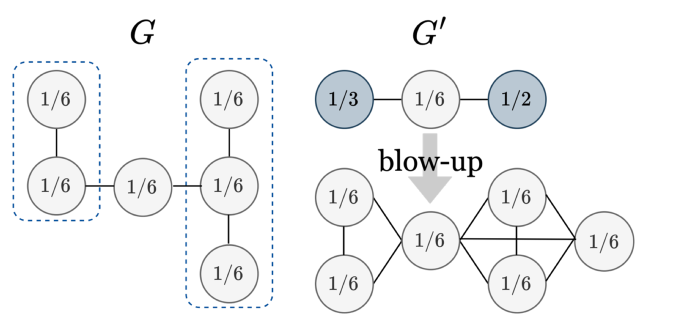
$$\mathcal{A}_G = \{\Pi : \Pi A = A\Pi\}$$

- Domain: **fixed graph**
- Signal: X: $\Omega \to \mathbb{R}^k$
- The group acts on signals only
- Network Architecture:
  - Equivariant linear map
  - Pointwise nonlinearity



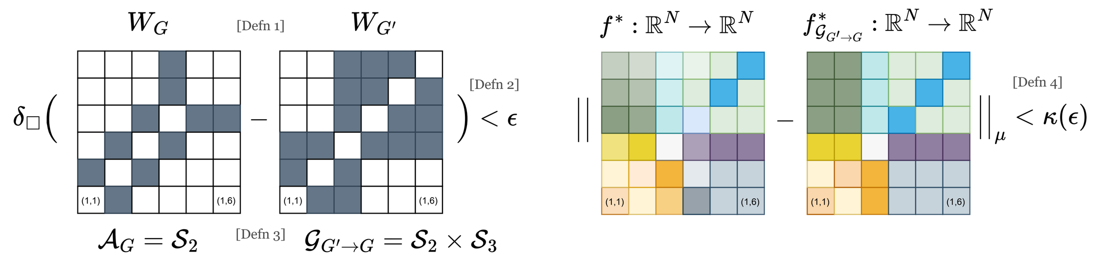
**Large real-world graphs tend to be asymmetric... How to introduce meaningful symmetries?**

### Approximate Symmetries: Graphon Analysis




Graph coarsening and blow-up to induce more symmetries:

- local permutation
- global automorphism induced from $G'$




Define approximate equivariant mapping using notions of induced graphons and coarsening error

Example: Random Geometric Graph

- Approximate symmetries: local deformations
- Approximate equivariant mapping: functions that are stable to local deformations.

## Theoretical Results: Bias-Variance Tradeoff in Symmetry Model Selection

Let $\phi, \psi$ be representations of $\mathcal{G}$ on $\mathcal{X}, \mathcal{Y}$ respectively. We say a function $f : \mathcal{X} \to \mathcal{Y}$ is equivariant w.r.t group $\mathcal{G}$ if

$$f(\phi(g)\, x) = \psi(g)\, f(x).$$

Define the projection operator by averaging over group orbits

$$(\mathcal{Q}_{\mathcal{G}} f)(x) = \int_{\mathcal{G}} \psi(g^{-1})\, f\,(\phi(g)x)\; \mathrm{d}\lambda(g),$$

Let $V$ be the space of all (measurable) functions $f : \mathcal{X} \to \mathcal{Y}$ such that $\|f\|_\mu = \sqrt{\langle f, f\rangle_\mu} < \infty$.

### Theorems

**Lemma (Risk Gap)**

*Let $X \sim \mu$ where $\mu$ is a $\mathcal{S}_N$-invariant distribution on $\mathcal{X} = \mathbb{R}^N$. Let $Y = f^*(X) + \xi \in \mathbb{R}^N$, where $\xi \in \mathbb{R}^N$ that is random, independent of $X$ with zero mean and finite variance and $f^* : \mathcal{X} \to \mathbb{R}^N$ is $\mathcal{G}$-equivariant. Then, for any $f \in V$ and for any compact groups $\mathcal{G}_L, \mathcal{G} \subseteq \mathcal{S}_N$, we can decompose it into*

$$f = \bar{f}_{\mathcal{G}_L} + f^{\perp}_{\mathcal{G}_L},$$

*where $\bar{f}_{\mathcal{G}_L} = \mathcal{Q}_{\mathcal{G}_L} f$, $f^{\perp}_{\mathcal{G}_L} = f - \bar{f}_{\mathcal{G}_L}$. Moreover, the risk gap satisfies*

$$\Delta(f, \bar{f}_{\mathcal{G}_L}) := \mathbb{E}\left[\|Y - f(X)\|_2^2\right] - \mathbb{E}\left[\|Y - \bar{f}_{\mathcal{G}_L}(X)\|_2^2\right] = \underbrace{-2\langle f^*, f^{\perp}_{\mathcal{G}_L}\rangle_\mu}_{\text{mismatch}} + \underbrace{\left\|f^{\perp}_{\mathcal{G}_L}\right\|_\mu^2}_{\text{constraint}}.$$

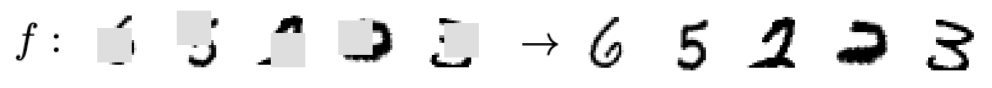
**Theorem (Bias-Variance Tradeoff)**

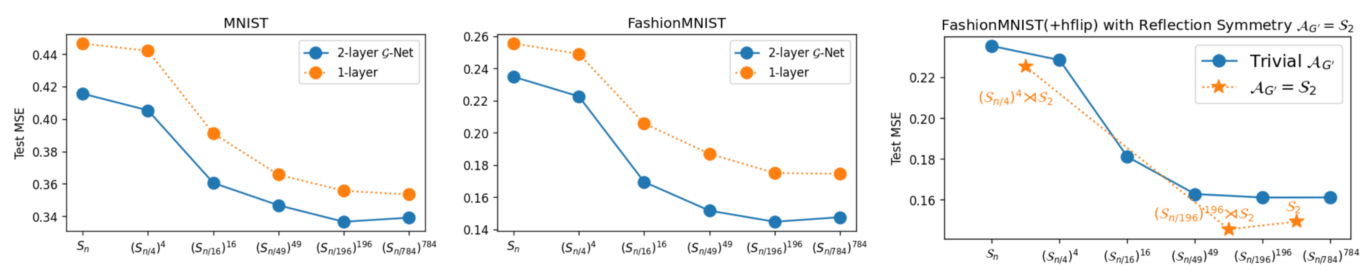
*Let $X \sim \mathcal{N}\left(0, \sigma_X^2 I_N\right), Y = f^*(X) + \xi$ where $f^*(x) = \Theta^\top x$ is $\mathcal{G}$-equivariant, $\Theta \in \mathbb{R}^{N \times N}$, and $\xi$ is white noise. Given $n$ i.i.d. examples $\{(X_i, Y_i) : i = 1, \ldots, n\}$, let $W$ be the least-squares estimate of $\Theta$, $\bar{W}_L = \mathcal{Q}_{\mathcal{G}_L}(W)$ be its equivariant version with respect to $\mathcal{G}_L$. Let $(\chi_{\psi_L} \mid \chi_{\phi_L}) = \int_{\mathcal{G}} \chi_{\psi_L}(g)\chi_{\phi_L}(g)\mathrm{d}\lambda(g)$ denote the inner product of group characters. If $n > N + 1$ the generalisation gain is*

$$\mathbb{E}\left[\Delta\left(f_W, f_{\bar{W}_L}\right)\right] = \underbrace{-\,\sigma_X^2 \left\|\mathcal{Q}^{\perp}_{\mathcal{G}_L}(\Theta)\right\|_F^2}_{\text{bias}} + \underbrace{\sigma_\xi^2 \frac{N^2 - (\chi_{\psi_L} \mid \chi_{\phi_L})}{n - N - 1}}_{\text{variance}}.$$

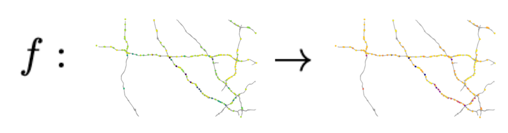
## Empirical Evidence: Using Active Symmetries in GNN Outperforms Passive Symmetries

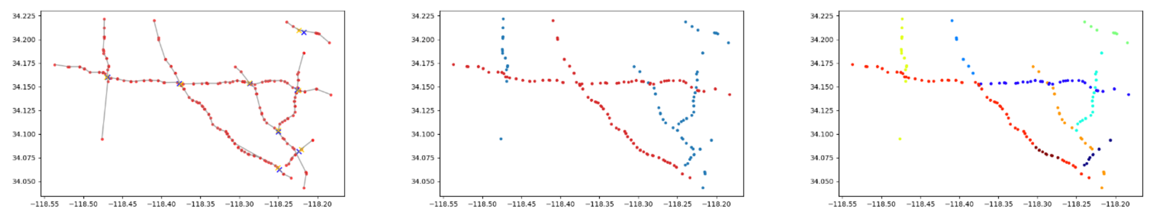
### Image Inpainting




### Traffic Forecasting

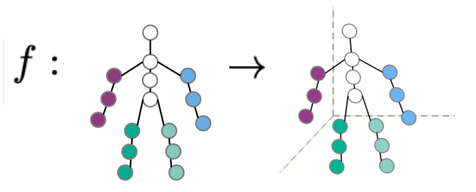


(a) Our faithful traffic graph (b) Graph clustering (2 clusters) (c) Graph clustering (9 clusters)

| $\mathcal{G}$-Net (gc) | $\mathcal{S}_N$ | $\mathcal{S}_{c_1} \times \mathcal{S}_{c_2}$ | $\mathcal{S}_{c_1} \times \ldots \times \mathcal{S}_{c_9}$ |
|---|---|---|---|
| Graph $G_s$ | $3.173 \pm 0.013$ | $\mathbf{3.150 \pm 0.008}$ | $3.204 \pm 0.006$ |
| Graph $G$ | $3.106 \pm 0.013$ | $\mathbf{3.092 \pm 0.008}$ | $3.174 \pm 0.013$ |

### Pose Estimation




| $\mathcal{G}$-Net (gc + ew) | $\mathcal{S}_{16}$ | Relax- $\mathcal{S}_{16}$ | $\mathcal{A}_G = (\mathcal{S}_2)^2$ | Trivial |
|---|---|---|---|---|
| MPJPE ↓ | $42.55 \pm 0.88$ | $\mathbf{39.87 \pm 0.46}$ | $42.18 \pm 0.49$ | $41.60 \pm 0.32$ |
| P-MPJPE ↓ | $34.48 \pm 0.44$ | $\mathbf{31.38 \pm 0.14}$ | $32.08 \pm 0.20$ | $31.69 \pm 0.17$ |