Charlotte Cambier van Nooten (charlotte.cambiervannooten@ru.nl)
Faculty of Science, Institute for Computing and Information Sciences (iCIS) – Radboud University Nijmegen

ICML 2026 Workshop on Weight-Space Symmetries

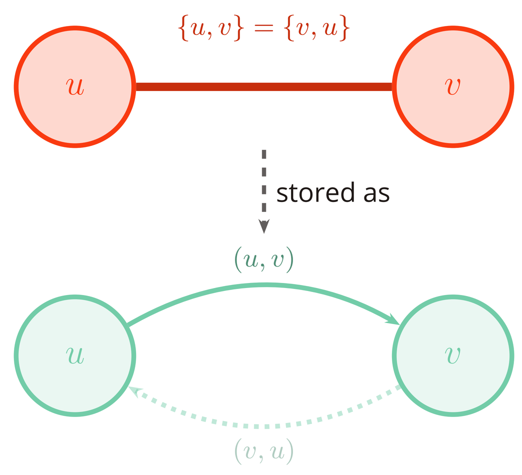
Endpoint Symmetry for Edge Updates: Weight-Space Redundancy in GNNs on Undirected Graphs

Charlotte Cambier van Nooten, Stijn van den Beemt, Yuliya Shapovalova, Tom Heskes

1. The Problem

Edge-centric GNNs update embeddings by concatenating endpoint features:

$$g^{(k)}_{u,v} = \xi\left(g^{(k-1)},\ h_u,\ h_v\right)$$



For an undirected edge $\{u,v\} = \{v,u\}$, the ordering of h_u, h_v carries no information, it is determined only by memory storage.

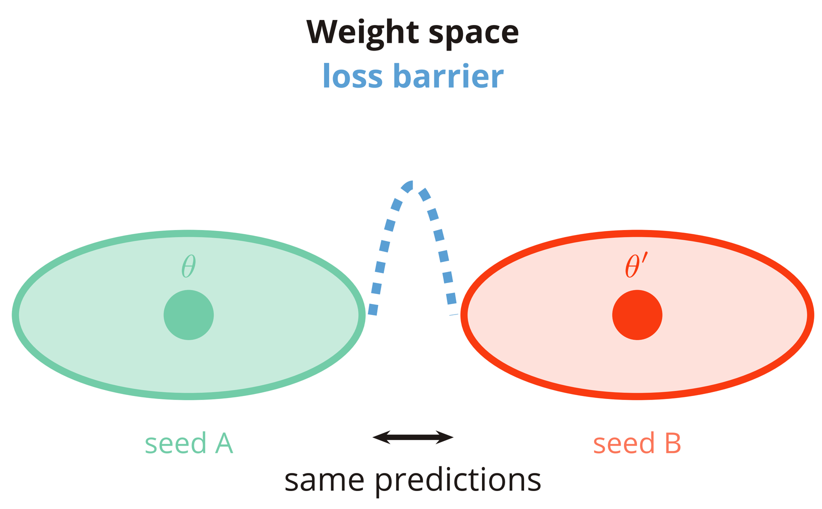
Endpoint symmetry requires $\xi(g, h_u, h_v) = \xi(g, h_v, h_u)$. Standard MLP violates this by construction.

For **edge classification**, $g^{(k)}$ is used directly as output, the asymmetry propagates to the prediction.

2. S_2 Redundant Orbit

For any $\theta = \{W_1, \ldots, W_L\}$, the block-swap P (swaps $h_u \leftrightarrow h_v$) constructs a partner:

$$\theta' = \{W_1P,\ W_2, \ldots\} \implies \xi_{\theta'}(g, h_u, h_v) = \xi_\theta(g, h_v, h_u)$$



$\{\theta, \theta'\}$ is a redundant S_2 orbit: two distinct weight vectors producing identical predictions. Seeds may land in either half.

THEOREM 3.2 — SYMMETRIC DEEP SETS APPROXIMATION

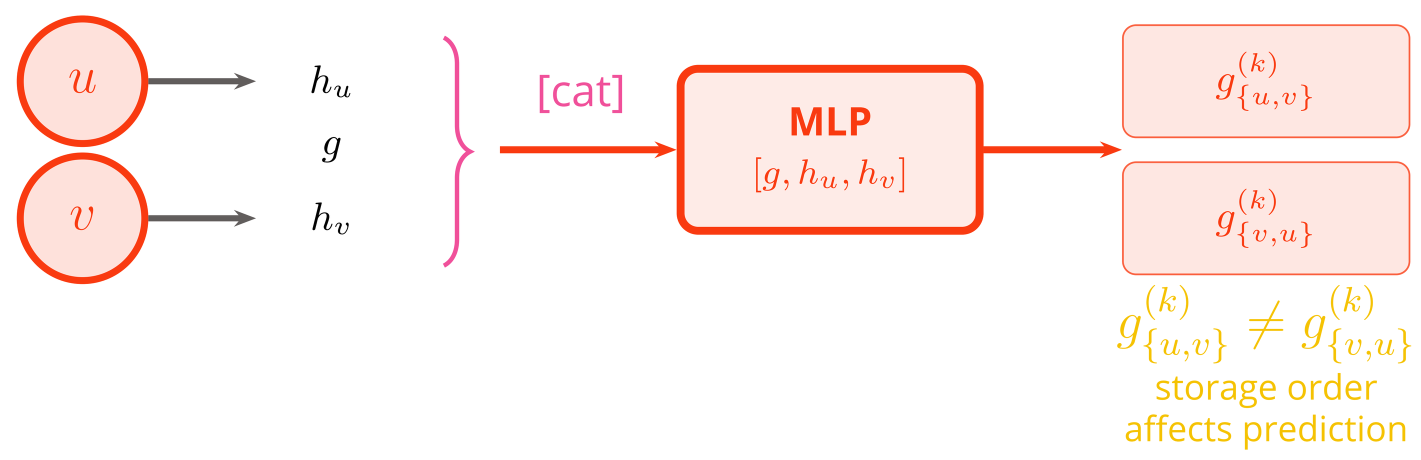
$$g^{(k)}_{\{u,v\}} = \rho\left(g^{(k-1)},\ \underbrace{\varphi(h_u) + \varphi(h_v)}_{\text{commutative: } a+b=b+a}\right)$$

endpoint-symmetric by construction · universally approximating on compact domains · no expressivity loss

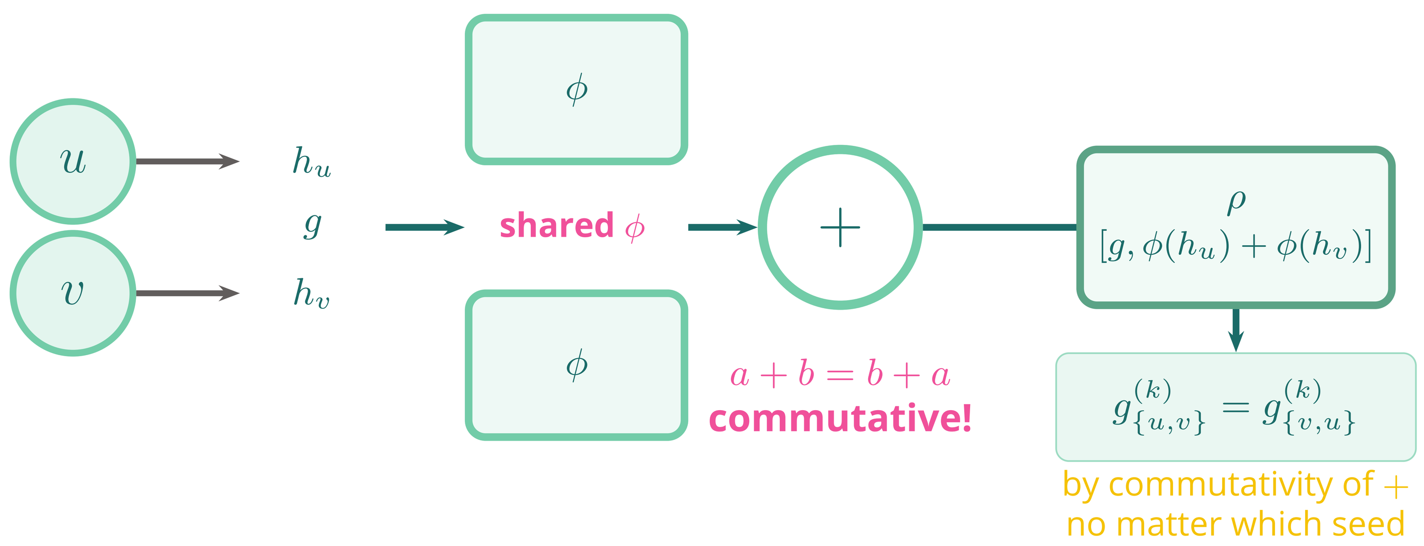
Why this eliminates the orbit: Any (φ, ρ) has a trivial S_2 orbit, no distinct (φ', ρ') can be obtained by swapping endpoints, because $\varphi(h_u) + \varphi(h_v) = \varphi(h_v) + \varphi(h_u)$ already. Contrast with concatenation, where $W_1P \neq W_1$ but $\{W_1P, W_2, \ldots\}$ computes the same function: a redundant orbit.

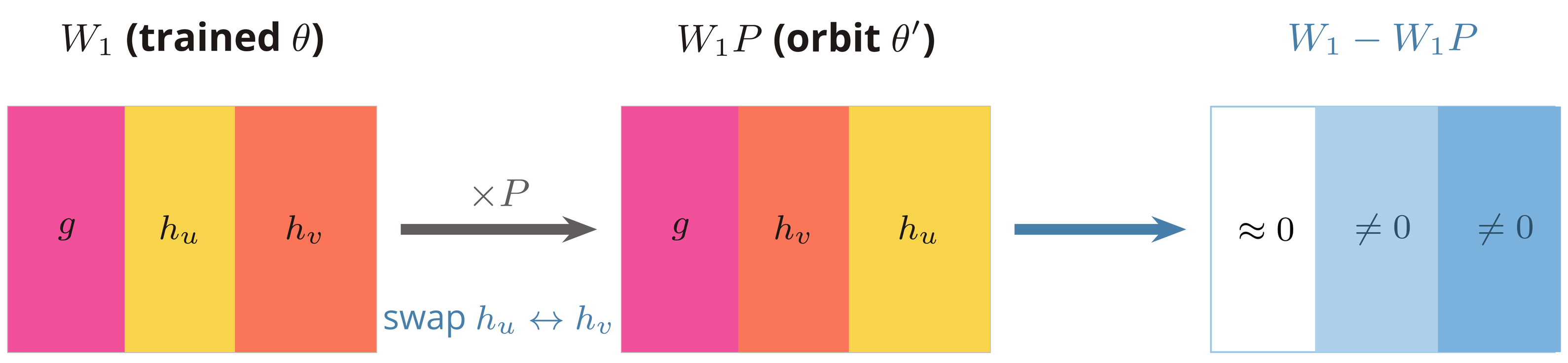
3. Deep Sets Edge Update

Asymmetric (broken):



Symmetric (ours):





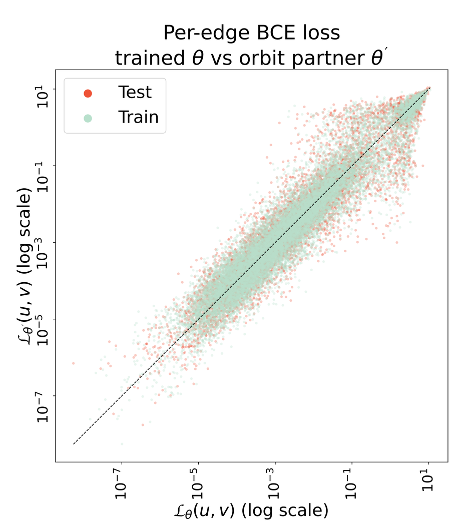
g block unchanged · h_u, h_v columns swapped · $|W_1 - W_1P| = 0.066$

Theorem 3.2. $\rho(g, \phi(h_u) + \phi(h_v))$ uniformly approximates any continuous endpoint-symmetric function on compact domains. Width $m \geq 4d+1$ ensures sum-injectivity (Amir et al., 2023). No expressivity loss for admissible functions.

Why + not concat? Addition is commutative, the output is the same regardless of endpoint ordering. φ is a shared encoder; the sum $\varphi(h_u) + \varphi(h_v)$ loses no information when φ is injective (width $m \geq 4d+1$ suffices).

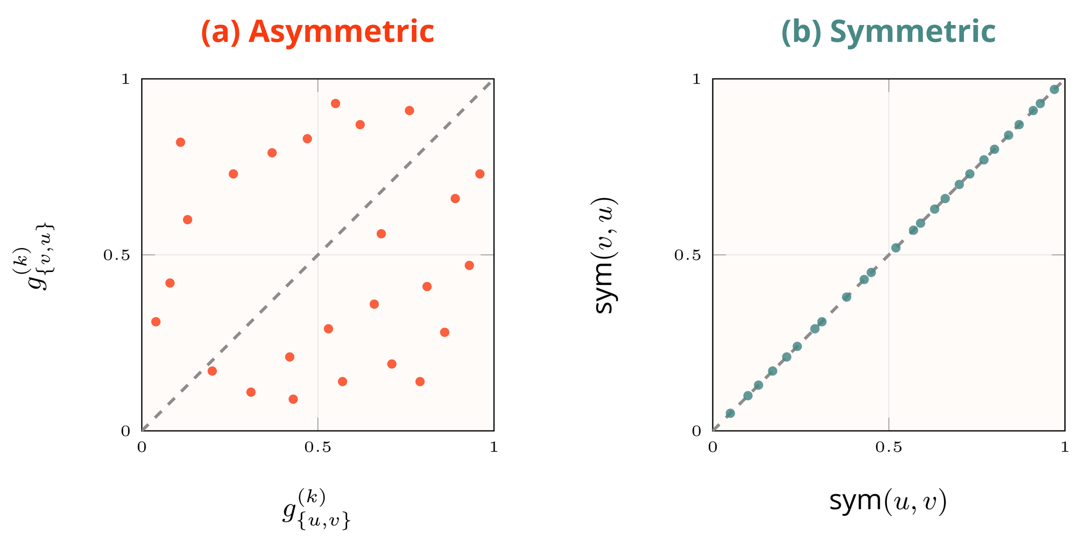
$2\times$ variance reduced	-33% fewer parameters	0 loss barrier (symmetric)	$p<.05$ Wilcoxon, power grid

4. Empirical Results



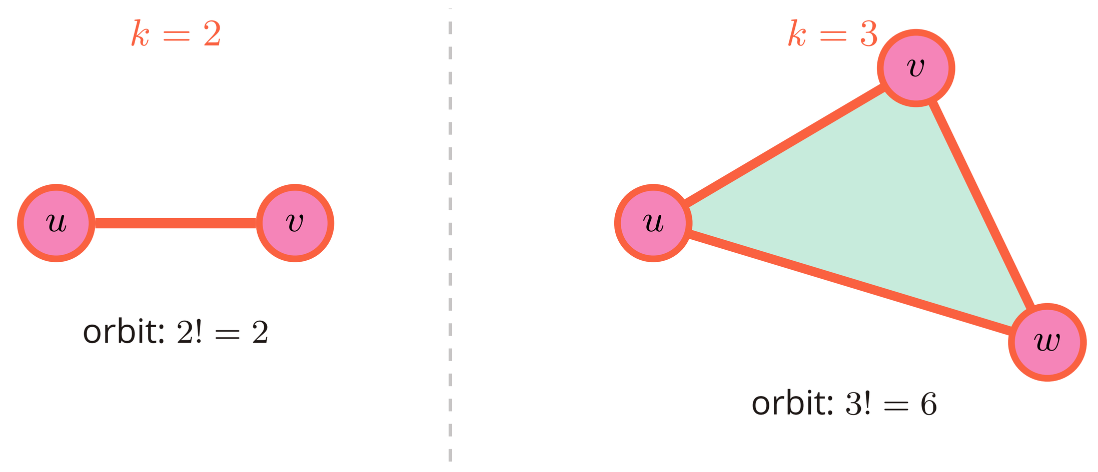
Synthetic task: symmetric ground-truth labels, 250 undirected graphs ($n = 15$ nodes, $d = 10$ features), 15 seeds, edge label $y_{\{u,v\}} = \sigma(0.8(\cos(h_u^\top w) + \cos(h_v^\top w)))$.

5. Symmetry Violation Test



Violation: 0.057 mean, 0.962 max vs 0 by construction

6. Generalisation



Use $\rho\left(\sum_{i\in e} \phi(h_i)\right)$, orbit $k!$, $m \geq 2kd+1$. Match architecture symmetry to data. Eliminate orbits, no cost to expressivity.

[1] Zaheer, M., Kottur, S., Ravanbakhsh, S., Poczos, B., Salakhutdinov, R.R., & Smola, A.J. (2017). Deep Sets. *Advances in Neural Information Processing Systems*, 30, 3391–3401.
[2] Amir, T., Gortler, S., Avni, I., Ravina, R., & Dym, N. (2023). Neural Injective Functions for Multisets, Measures and Graphs via a Finite Witness Theorem. *Advances in Neural Information Processing Systems.*

ICML International Conference On Machine Learning

Radboud Universiteit