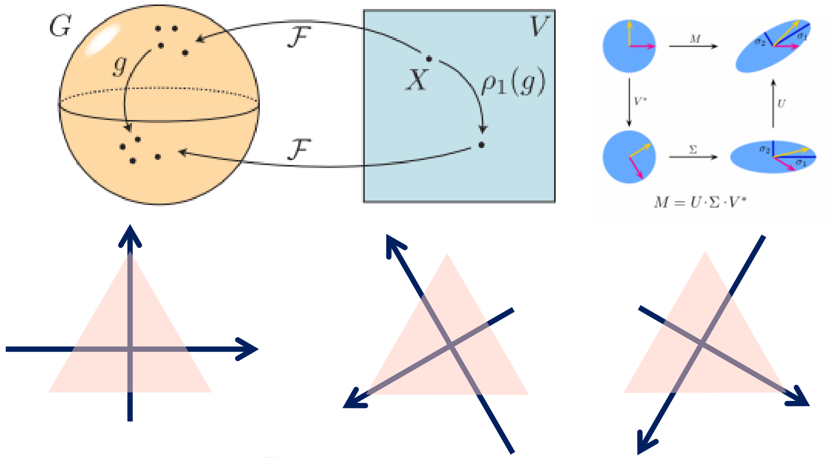
Are High-Degree Representations Really Unnecessary in Equivariant Graph Neural Networks?

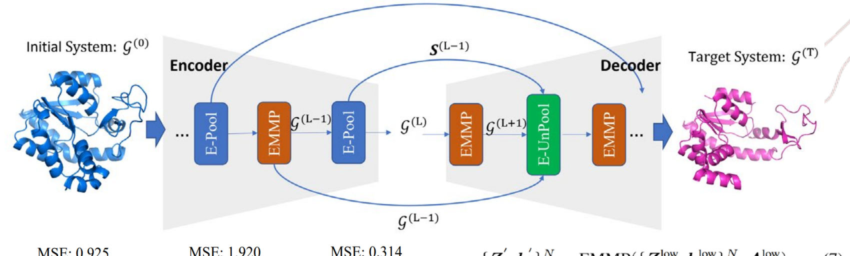
Jiacheng Cen, Anyi Li, Ning Lin, Yuxiang Ren, Zihe Wang, Wenbing Huang

2024.10.30

Global Features in Geometric GNNs

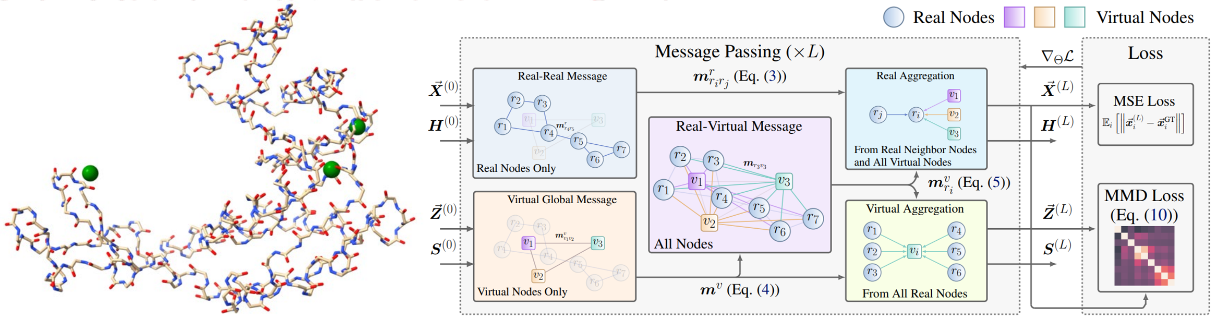


(a) Frames
in Frame Averaging (ICLR’22)

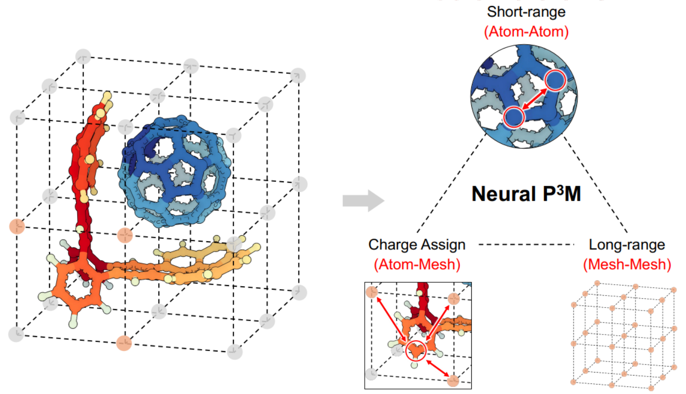


$$\{Z_i', h_i'\}_i^N = \mathrm{EMMP}(\{Z_i^{\mathrm{low}}, h_i^{\mathrm{low}}\}_i^N, A^{\mathrm{low}}), \quad (7)$$
$$s_i = \mathrm{Softmax}(\mathrm{MLP}(h_i')), \quad (8)$$
$$Z_j^{\mathrm{high}} = \frac{1}{\sum_{i=1}^N s_{ij}} \sum_{i=1}^N s_{ij} Z_i', \quad (9)$$
$$h_j^{\mathrm{high}} = \frac{1}{\sum_{j=1}^N s_{ij}} \sum_{i=1}^N s_{ij} h_i^{\mathrm{low}}, \quad (10)$$
$$A^{\mathrm{high}} = S^\top A^{\mathrm{low}} S, \quad (11)$$

(b) Equivariant Pooling
in EGHN (NeurIPS’22)

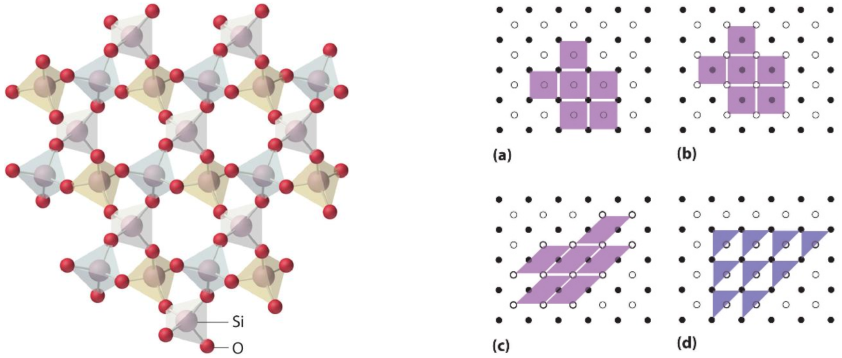


(c) Virtual Nodes
in FastEGNN (ICML’24)



(d) Mesh
in Neural P³M (NeurIPS’24)

Symmetric Graph



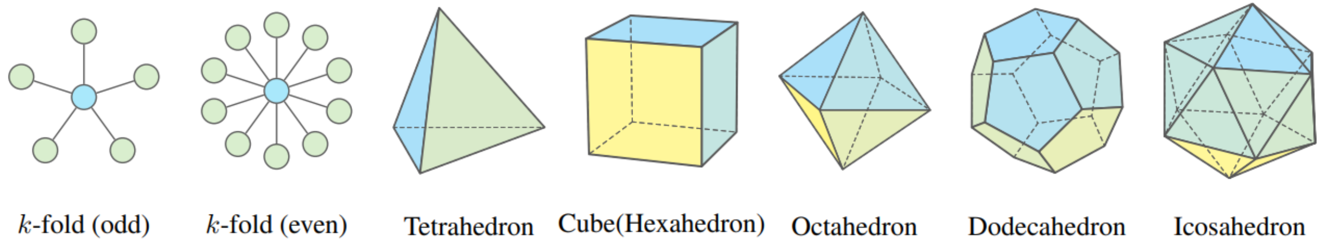
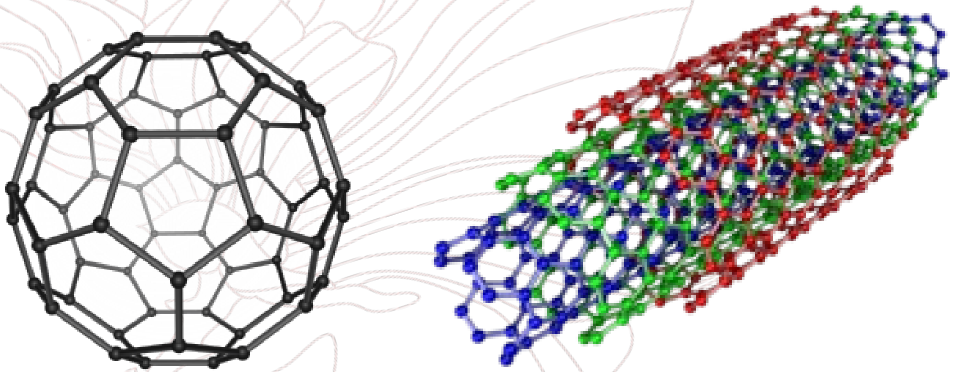


Figure 1: Common symmetric graphs. Equivariant GNNs on symmetric graphs will degenerate to a zero function if the degree of their representations is fixed as 1.

Crystal Structures



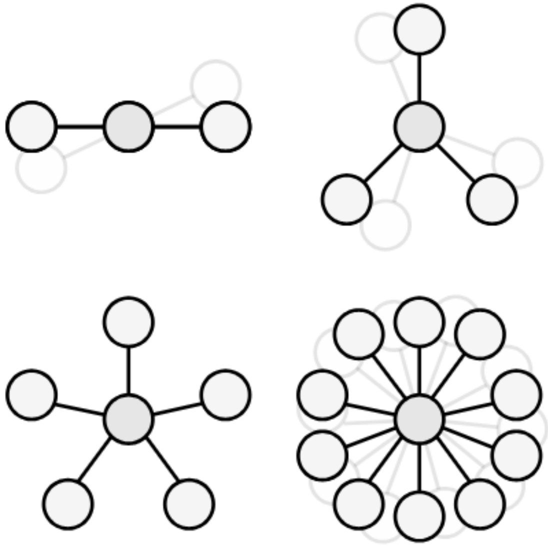
C_{60} & Carbon Nanotube

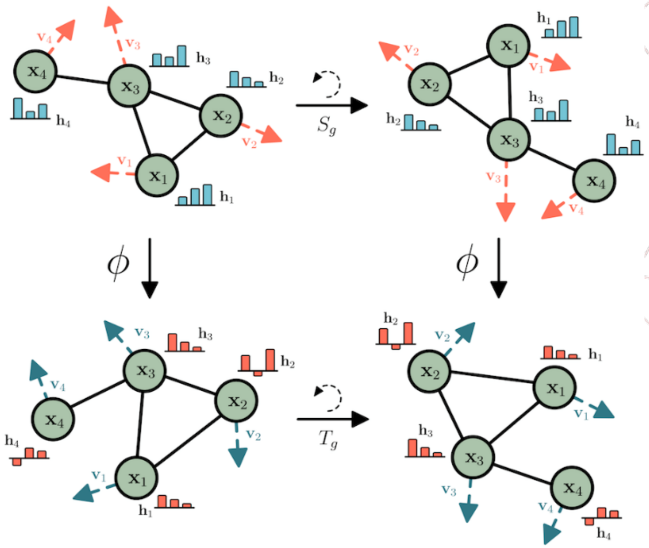
Symmetrical Structure

Coincides with itself under certain transformations

$$\forall \mathfrak{h} \in \mathfrak{H}, \mathfrak{h} \cdot \mathcal{G} = \mathcal{G}$$

The Degeneration Phenomenon





Symmetrical Structure + **Equivariant GNNs** = **Degeneracy at a specific order l**

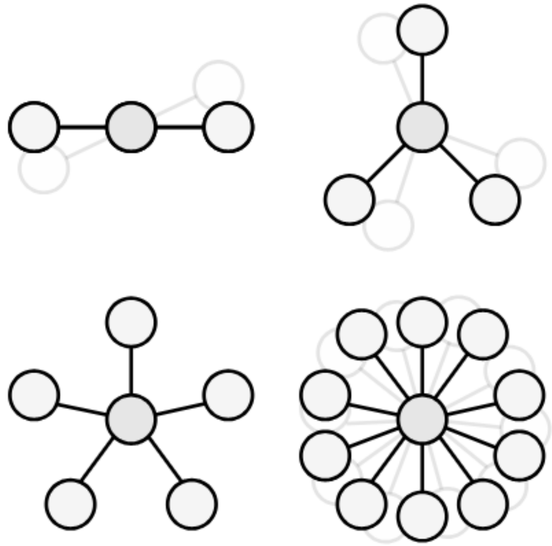
Coincides with itself under certain transformations

$$\forall \mathfrak{h} \in \mathfrak{H}, \mathfrak{h} \cdot \mathcal{G} = \mathcal{G}$$

$$\rho^{(l)}(\mathfrak{g}) f^{(l)}(\mathcal{G}) = f^{(l)}(\mathfrak{g} \cdot \mathcal{G})$$

$$f^{(l)}(\mathcal{G}) \equiv \mathbf{0}$$

Theoretical Results



Symmetrical Structure

Coincides with itself under certain transformations

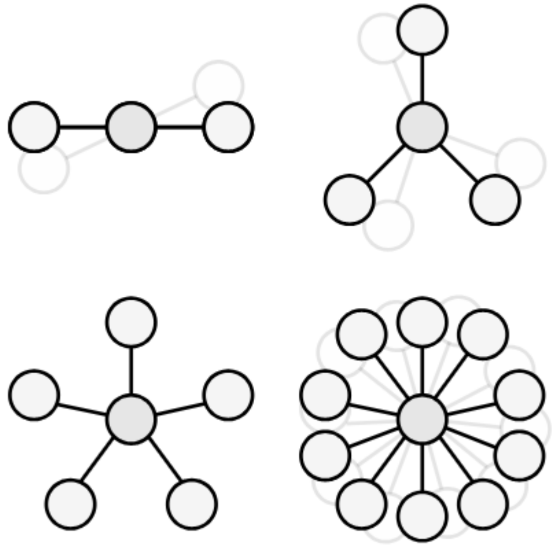
$$\forall \mathfrak{h} \in \mathfrak{H}, \mathfrak{h} \cdot \mathcal{G} = \mathcal{G}$$

Using the definition of **symmetric structure** and **equivariant function**, we can get the equation

$$\begin{aligned} f^{(l)}(\mathcal{G}) &= f^{(l)}(\mathfrak{h} \cdot \mathcal{G}) \\ &= \rho^{(l)}(\mathfrak{h}) \cdot f^{(l)}(\mathcal{G}) \\ &= \left(\frac{1}{|\mathfrak{H}|} \sum_{\mathfrak{h} \in \mathfrak{H}} \rho^{(l)}(\mathfrak{h}) \right) \cdot f^{(l)}(\mathcal{G}) \\ &\triangleq \rho^{(l)}(\mathfrak{H}) f^{(l)}(\mathcal{G}) \end{aligned}$$

Note that the types of point groups are **finite**, so we only need to **enumerate all the groups** to represent the average.

Theoretical Results



Symmetrical Structure

Coincides with itself under certain transformations

$$\forall \mathfrak{h} \in \mathfrak{H}, \mathfrak{h} \cdot \mathcal{G} = \mathcal{G}$$

Using the definition of **symmetric structure** and **equivariant function**, we can get the equation

$$\left(I_{2l+1} - \rho^{(l)}(\mathfrak{H})\right) f^{(l)}(\mathcal{G}) = 0$$

Left Matrix is full-rank ⟺ **The Degeneration Phenomenon**

$$\det\left(I_{2l+1} - \rho^{(l)}(\mathfrak{H})\right) \neq 0$$

$$f^{(l)}(\mathcal{G}) \equiv 0$$

Note that the types of point groups are **finite**, so we only need to **enumerate all the groups** to represent the average.

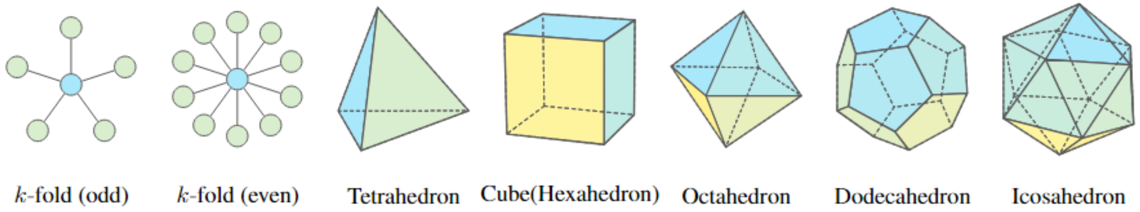
Motivation for HEGNN

Trace of point group average representation

Group	Notation	Data for Wigner-D matrix traces $D^{(l)}(H)$	
Reflection group	C_i	$(2l+1)\cdot\delta_{l \bmod 2, 0}$	
Cyclic group	C_n	$2\lfloor l/n \rfloor + 1$	
Dihedral group	D_n	$\lfloor l/n \rfloor + \delta_{l \bmod 2, 0}$	
Tetrahedral group	T	$r = 6$	$b = 100110$
Octahedral group	O	$r = 12$	$b = 100010101110$
Icosahedral group	I	$r = 30$	$b = 1000001000101001101011101111110$

Prediction of degenerate results for various symmetric graphs

Symmetric Graph $\mathcal{G}$	Symmetry Group $\mathfrak{H} \in \mathbb{H}(\mathcal{G})$	l leading to $f^{(l)}(\mathcal{G}) \equiv \mathbf{0}$
$2k$-fold	C_i, D_{2k}	l is odd
$(2k+1)$-fold	D_{2k+1}	$l < 2k+1$ and l is odd
Tetrahedron	T	$l \in \{1, 2, 5\}$
Cube/Octahedron	C_i, O	$l = 2$ or l is odd
Dodecahedron/Icosahedron	C_i, I	$l \in \{2, 4, 8, 14\}$ or l is odd



Difficulties

- Previous models generate **all representations of $|l_1 - l_2| \sim l_1 + l_2$** through CG tensor product, and cannot extract representations of special orders for verification

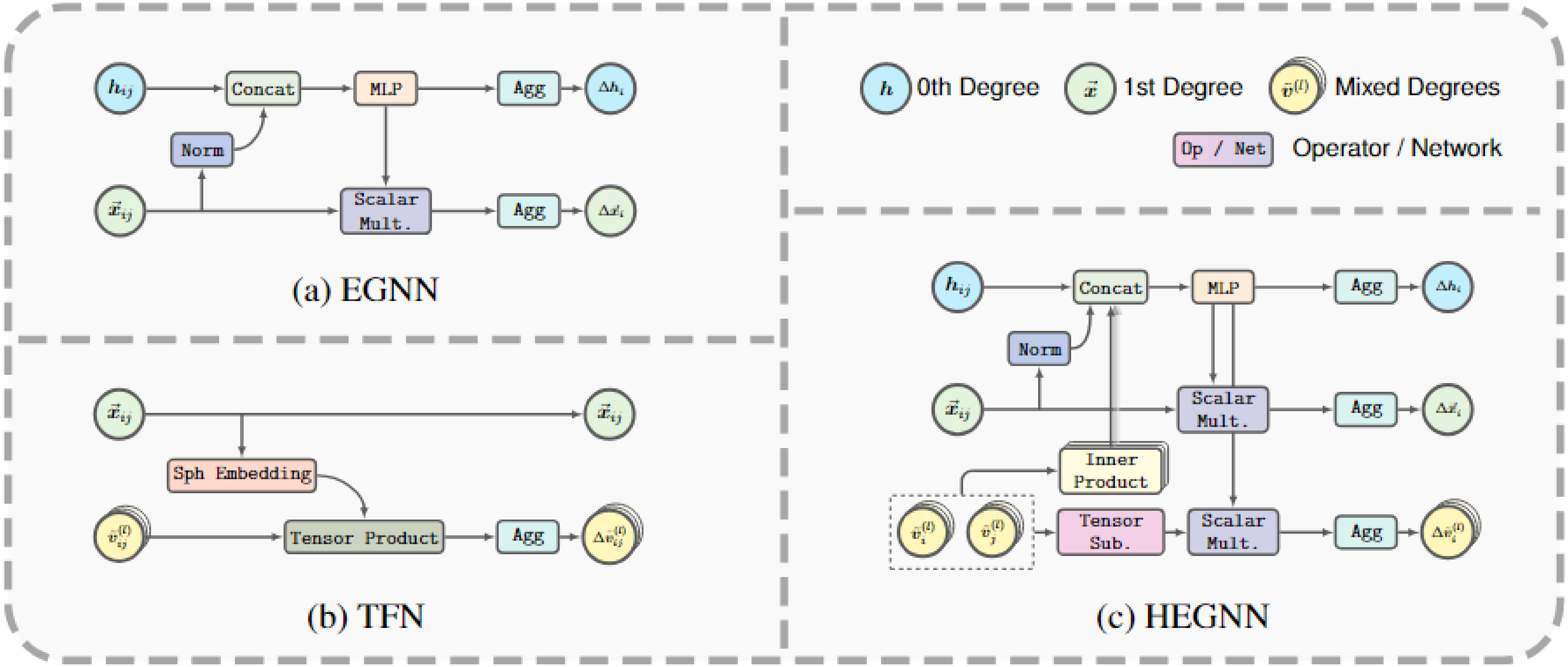
Additional requirements

- Can the model used for verification have **good application value**? For example, use it on **actual datasets**?
- Traditional high-order models use CG tensor products, with a complexity of up to $O(L^6)$, Can we design a model with **lower complexity**?
- Can you explain the **theoretical basis** for using high-order representations other than distinguishing symmetric structures?

Architecture of HEGNN

HEGNN: Use the scalarization-trick to introduce high-order representations, which reduce the time complexity to $O(L^2)$ from $O(L^6)$ of CG tensor-product

- Initialization: Use spherical harmonics and calculate coefficients for different orders
- Expression ability: Use the relationship between spherical harmonics and Legendre polynomials to prove that **HEGNN can fully express all inner product information of geometric graphs**



(a) EGNN (b) TFN (c) HEGNN

Architecture of HEGNN

- Initialization of high-degree steerable feature

$$\tilde{\boldsymbol{v}}_{i,\texttt{init}}^{(l)} = \frac{1}{|\mathcal{N}(i)|} \sum_{j\in\mathcal{N}(i)} \varphi_{\tilde{\boldsymbol{v}},\texttt{init}}^{(l)}(\boldsymbol{m}_{ij,\texttt{init}}) \cdot Y^{(l)}\left(\frac{\vec{\boldsymbol{x}}_i - \vec{\boldsymbol{x}}_j}{\|\vec{\boldsymbol{x}}_i - \vec{\boldsymbol{x}}_j\|}\right)$$

- Calculation of cross-degree invariant messages

$$d_{ij} = \|\vec{\boldsymbol{x}}_i - \vec{\boldsymbol{x}}_j\|, \quad z_{ij}^{(l)} = \left\langle \tilde{\boldsymbol{v}}_i^{(l)}, \tilde{\boldsymbol{v}}_j^{(l)} \right\rangle, \quad \boldsymbol{m}_{ij} = \varphi_{\boldsymbol{m}}\left(\boldsymbol{h}_i, \boldsymbol{h}_j, \boldsymbol{e}_{ij}, d_{ij}^2, \oplus_{l=0}^{L} z_{ij}^{(l)}\right)$$

- Aggregation of neighbor messages

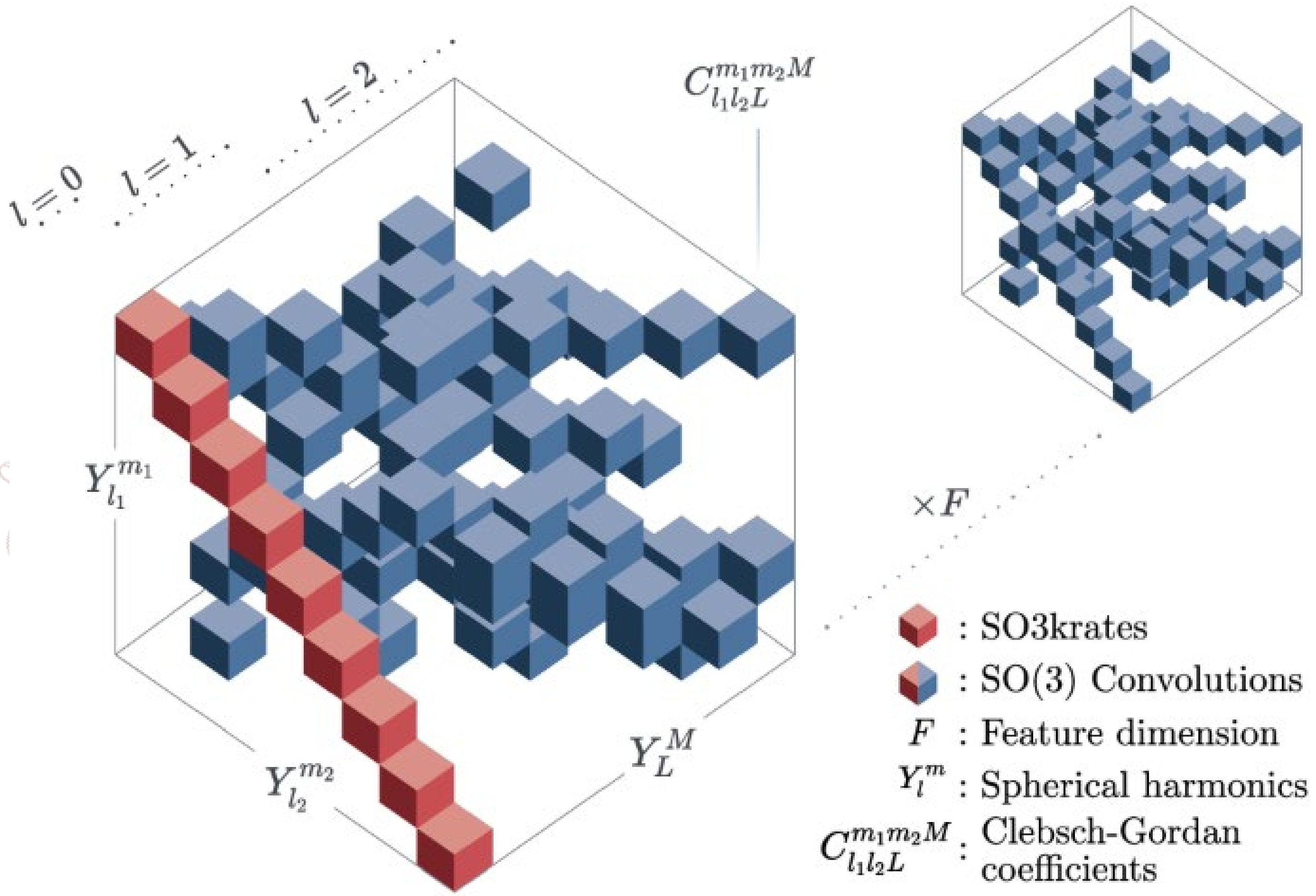
$$\Delta \boldsymbol{h}_i = \varphi_{\boldsymbol{h}}\left(\boldsymbol{h}_i, \frac{1}{|\mathcal{N}(i)|}\sum_{j\in\mathcal{N}(i)} \boldsymbol{m}_{ij}\right), \quad \Delta\vec{\boldsymbol{x}}_i = \frac{1}{|\mathcal{N}(i)|}\sum_{j\in\mathcal{N}(i)} \varphi_{\vec{\boldsymbol{x}}}(\boldsymbol{m}_{ij}) \cdot (\vec{\boldsymbol{x}}_i - \vec{\boldsymbol{x}}_j),$$

$$\Delta\tilde{\boldsymbol{v}}_i^{(l)} = \frac{1}{|\mathcal{N}(i)|}\sum_{j\in\mathcal{N}(i)} \varphi_{\tilde{\boldsymbol{v}}}^{(l)}(\boldsymbol{m}_{ij}) \cdot \left(\tilde{\boldsymbol{v}}_i^{(l)} - \tilde{\boldsymbol{v}}_j^{(l)}\right)$$

- Aggregation of neighbor messages

$$\oplus_{l=0}^{L} \Delta\tilde{\boldsymbol{v}}_i^{(l)} = \frac{1}{|\mathcal{N}(i)|}\sum_{j\in\mathcal{N}(i)} 1 \otimes_{\text{cg}}^{\varphi_{\tilde{\boldsymbol{v}}}(\boldsymbol{m}_{ij})} \left(\oplus_{l=0}^{L}\left(\tilde{\boldsymbol{v}}_i^{(l)} - \tilde{\boldsymbol{v}}_j^{(l)}\right)\right)$$

An equivalent Concurrent Work: SO3krates



$$x^{00}_{ij} = \sum_{l_1} \underbrace{\sum_{m_1} C^{m_1 -m_1 0}_{l_1 l_1 0} x^{l1m1}_{ij} x^{l1-m1}_{ij}}_{\equiv \bigoplus_{l=0}^{l_{\max}} \boldsymbol{x}_{ij,l\to 0}},$$

$\Updownarrow$

$$z^{(l)}_{ij} = \left\langle \tilde{\boldsymbol{v}}^{(l)}_i, \tilde{\boldsymbol{v}}^{(l)}_j \right\rangle$$

Architecture of HEGNN

The message passing formulas of our HEGNN, EGNN and TFN

	EGNN [1]	TFN [12]	HEGNN (Ours)
Msg	$\boldsymbol{m}_{ij}=\phi_{\boldsymbol{m}}(\boldsymbol{h}_i,\boldsymbol{h}_j,\boldsymbol{e}_{ij},d_{ij}^2)$ $\vec{\boldsymbol{m}}_{ij}=\varphi_{\vec{\boldsymbol{x}}}(\boldsymbol{m}_{ij})\cdot(\vec{\boldsymbol{x}}_i-\vec{\boldsymbol{x}}_j)$	$\tilde{\boldsymbol{m}}_{ij}^{(\mathbb{L})}=\tilde{\boldsymbol{v}}_i^{(\mathbb{L})}\otimes_{\text{cg}}^{\boldsymbol{W}(d_{ij})}Y^{(\mathbb{L})}\left(\frac{\vec{\boldsymbol{x}}_{ij}}{\Vert\vec{\boldsymbol{x}}_{ij}\Vert}\right)$	$\boldsymbol{m}_{ij}=\varphi_{\boldsymbol{m}}(\boldsymbol{h}_i,\boldsymbol{h}_j,\boldsymbol{e}_{ij},d_{ij}^2,\bigoplus_{l\in\mathbb{L}}z_{ij}^{(l)})$ $\vec{\boldsymbol{m}}_{ij}=\varphi_{\vec{\boldsymbol{x}}}(\boldsymbol{m}_{ij})\cdot(\vec{\boldsymbol{x}}_i-\vec{\boldsymbol{x}}_j)$ $\tilde{\boldsymbol{v}}_{ij}^{(l)}=\varphi_{\tilde{\boldsymbol{v}}}^{(l)}(\boldsymbol{m}_{ij})\cdot(\tilde{\boldsymbol{v}}_i^{(l)}-\tilde{\boldsymbol{v}}_j^{(l)})$
Agg	$\boldsymbol{m}_i=\alpha_i\sum_{j\in\mathcal{N}(i)}\boldsymbol{m}_{ij}$ $\vec{\boldsymbol{m}}_i=\alpha_i\sum_{j\in\mathcal{N}(i)}\vec{\boldsymbol{m}}_{ij}$	$\tilde{\boldsymbol{m}}_i^{(\mathbb{L})}=\alpha_i\sum_{j\in\mathcal{N}(i)}\tilde{\boldsymbol{m}}_{ij}^{(\mathbb{L})}$	$\boldsymbol{m}_i=\alpha_i\sum_{j\in\mathcal{N}(i)}\boldsymbol{m}_{ij}$ $\vec{\boldsymbol{m}}_i=\alpha_i\sum_{j\in\mathcal{N}(i)}\vec{\boldsymbol{m}}_{ij}$ $\tilde{\boldsymbol{m}}_i^{(l)}=\alpha_i\sum_{j\in\mathcal{N}(i)}\tilde{\boldsymbol{m}}_{ij}^{(l)}$
Upd	$\Delta\boldsymbol{h}_i=\varphi_{\boldsymbol{h}}(\boldsymbol{h}_i,\boldsymbol{m}_i)$ $\Delta\vec{\boldsymbol{x}}_i=\vec{\boldsymbol{m}}_i$	$\Delta\boldsymbol{v}_i^{(\mathbb{L})}=\boldsymbol{m}_i^{(\mathbb{L})}$	$\Delta\boldsymbol{h}_i=\varphi_{\boldsymbol{h}}(\boldsymbol{h}_i,\boldsymbol{m}_i)$ $\Delta\vec{\boldsymbol{x}}_i=\vec{\boldsymbol{m}}_i$ $\Delta\tilde{\boldsymbol{v}}_i^{(l)}=\tilde{\boldsymbol{m}}_i^{(l)}$

Theorem 4.1. *For any geometric graph, there exists a bijection between the set of inner products* $\{z_{ij}^{(l)}\}_{l=1}^{|\mathbb{A}_{ij}|}$ *given by Eq. (10) and the set of edge angles* $\mathbb{A}_{ij}=\{\theta_{is,jt}:=\langle\vec{\boldsymbol{x}}_{is},\vec{\boldsymbol{x}}_{jt}\rangle\}_{s\in\mathcal{N}(i),t\in\mathcal{N}(j)}$.

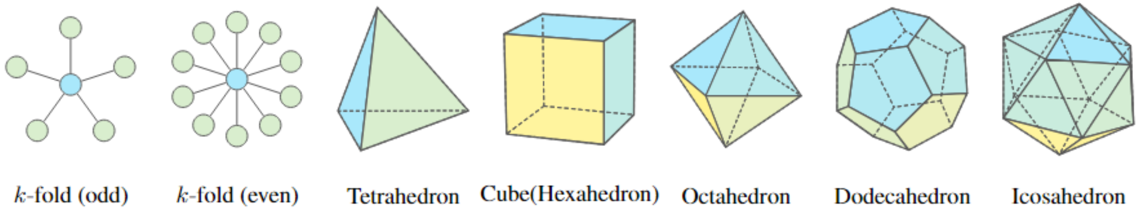
$$\left\langle\sum_{s\in\mathcal{N}(i)}Y^{(l)}(\vec{\boldsymbol{x}}_{is}),\sum_{t\in\mathcal{N}(j)}Y^{(l)}(\vec{\boldsymbol{x}}_{jt})\right\rangle=\frac{4\pi}{2l+1}\sum_{s\in\mathcal{N}(i)}\sum_{t\in\mathcal{N}(j)}P^{(l)}(\langle\vec{\boldsymbol{x}}_{is},\vec{\boldsymbol{x}}_{jt}\rangle),$$

Experiments

Symmetric polyhedron experiment: Theoretical and experimental results are completely consistent

N-body (N=5, 20, 50, 100): consistently outperforms other models

MD-17: outperforms most molecules (6/8)



	GNN Layer	Rotational symmetry: Tetrahedron	Cube	Octahedron	Dodecahedron	Icosahedron
Cart.	E-GNN$_{l=1}$	50.0 ± 0.0	50.0 ± 0.0	50.0 ± 0.0	50.0 ± 0.0	50.0 ± 0.0
	GVP-GNN$_{l=1}$	50.0 ± 0.0	50.0 ± 0.0	50.0 ± 0.0	50.0 ± 0.0	50.0 ± 0.0
Single Type Spherical	HEGNN$_{l=1}$	50.0 ± 0.0	50.0 ± 0.0	50.0 ± 0.0	50.0 ± 0.0	50.0 ± 0.0
	HEGNN$_{l=2}$	50.0 ± 0.0	50.0 ± 0.0	50.0 ± 0.0	50.0 ± 0.0	50.0 ± 0.0
	HEGNN$_{l=3}$	**100.0 ± 0.0**	50.0 ± 0.0	50.0 ± 0.0	50.0 ± 0.0	50.0 ± 0.0
	HEGNN$_{l=4}$	**100.0 ± 0.0**	90.0 ±30.0	90.0 ±30.0	50.0 ± 0.0	50.0 ± 0.0
	HEGNN$_{l=5}$	50.0 ± 0.0	50.0 ± 0.0	50.0 ± 0.0	50.0 ± 0.0	50.0 ± 0.0
	HEGNN$_{l=6}$	**100.0 ± 0.0**	**100.0 ± 0.0**	**100.0 ± 0.0**	**100.0 ± 0.0**	**100.0 ± 0.0**
	HEGNN$_{l=7}$	**100.0 ± 0.0**	50.0 ± 0.0	50.0 ± 0.0	50.0 ± 0.0	50.0 ± 0.0
	HEGNN$_{l=8}$	**100.0 ± 0.0**	90.0 ±30.0	90.0 ±30.0	50.0 ± 0.0	50.0 ± 0.0
	HEGNN$_{l=9}$	**100.0 ± 0.0**	50.0 ± 0.0	50.0 ± 0.0	50.0 ± 0.0	50.0 ± 0.0
	HEGNN$_{l=10}$	**100.0 ± 0.0**	**100.0 ± 0.0**	95.0 ±15.0	**100.0 ± 0.0**	**100.0 ± 0.0**
	HEGNN$_{l=11}$	**100.0 ± 0.0**	50.0 ± 0.0	50.0 ± 0.0	50.0 ± 0.0	50.0 ± 0.0
Sph.	HEGNN/TFN/MACE$_{l\le2}$	50.0 ± 0.0	50.0 ± 0.0	50.0 ± 0.0	50.0 ± 0.0	50.0 ± 0.0
	HEGNN/TFN/MACE$_{l\le3}$	**100.0 ± 0.0**	50.0 ± 0.0	50.0 ± 0.0	50.0 ± 0.0	50.0 ± 0.0
	HEGNN/TFN/MACE$_{l\le4}$	**100.0 ± 0.0**	**100.0 ± 0.0**	**100.0 ± 0.0**	50.0 ± 0.0	50.0 ± 0.0
	HEGNN/TFN/MACE$_{l\le6}$	**100.0 ± 0.0**	**100.0 ± 0.0**	**100.0 ± 0.0**	**100.0 ± 0.0**	**100.0 ± 0.0**

	5-body MSE ($\times10^{-2}$)	5-body Relative Time	20-body MSE ($\times10^{-2}$)	20-body Relative Time	50-body MSE ($\times10^{-2}$)	50-body Relative Time	100-body MSE ($\times10^{-2}$)	100-body Relative Time
Linear	7.72	0.01	10.12	0.02	11.81	0.02	12.69	0.01
MPNN [35]	1.80	0.49	2.50	0.51	2.96	0.50	3.55	0.45
SchNet [36]	11.31	2.93	17.72	6.24	22.14	31.63	22.14	27.04
RF [34]	1.51	0.54	3.41	0.65	4.75	0.67	5.72	0.49
GVP-GNN [37]	7.26	2.36	5.76	2.38	7.07	2.42	7.55	2.33
EGNN [1]	0.65	1.00	1.01	1.00	1.00	1.00	1.36	1.00
TFN$_{l\le2}$	1.49	2.69	1.86	3.19	2.20	2.87	3.42	6.58
TFN$_{l\le3}$	1.76	3.91	1.87	4.54	1.94	4.89	OOM	-
SE(3)-Tr.$_{l\le2}$	3.24	4.94	3.19	5.88	2.54	5.97	2.33	5.15
HEGNN$_{l\le1}$	0.52	1.77	0.79	1.84	0.88	1.60	1.13	1.45
HEGNN$_{l\le2}$	**0.47**	1.88	**0.78**	1.94	0.90	1.71	0.97	1.55
HEGNN$_{l\le3}$	0.48	2.11	0.80	2.23	**0.84**	1.84	0.94	1.61
HEGNN$_{l\le6}$	0.69	2.14	0.86	2.43	0.96	2.18	**0.86**	1.90

	Aspirin	Benzene	Ethanol	Malonaldehyde	Naphthalene	Salicylic	Toluene	Uracil
RF	10.94±0.01	103.72±1.29	4.64±0.01	13.93±0.03	0.50±0.01	1.23±0.01	10.93±0.04	0.64±0.01
EGNN	14.41±0.15	62.40±0.53	4.64±0.01	13.64±0.01	0.47±0.02	1.02±0.02	11.78±0.07	0.64±0.01
EGNNReg	13.82±0.19	61.68±0.37	6.06±0.01	13.49±0.06	0.63±0.01	1.68±0.01	11.05±0.01	0.66±0.01
GMN	10.14±0.03	**48.12**±0.40	4.83±0.01	13.11±0.03	0.40±0.01	0.91±0.01	**10.22**±0.08	0.59±0.01
TFN$_{l\le2}$	12.37±0.18	58.48±1.98	4.81±0.04	13.62±0.08	0.49±0.01	1.03±0.02	10.89±0.01	0.84±0.02
SE(3)-Tr.$_{l\le2}$	11.12±0.06	68.11±0.67	4.74±0.13	13.89±0.02	0.52±0.01	1.13±0.02	10.88±0.06	0.79±0.02
HEGNN$_{l\le1}$	10.32±0.58	62.53±7.62	4.63±0.01	12.85±0.01	0.38±0.01	0.90±0.05	10.56±0.10	0.56±0.02
HEGNN$_{l\le2}$	10.04±0.45	61.80±5.92	4.63±0.01	12.85±0.01	0.39±0.01	0.91±0.06	10.56±0.05	0.55±0.01
HEGNN$_{l\le3}$	10.20±0.23	62.82±4.25	4.63±0.01	12.85±0.02	**0.37**±0.01	0.94±0.10	10.55±0.16	**0.52**±0.01
HEGNN$_{l\le6}$	**9.94**±0.07	59.93±5.21	**4.62**±0.01	12.85±0.01	**0.37**±0.02	**0.88**±0.02	10.56±0.33	0.54±0.01

Perturbation Experiment

Most molecules may **not be symmetrical**, and even affected by molecular vibration, the structural changes are enough to **eliminate the original symmetry**.

So what are the **advantages of HEGNN** at this time? The answer is **better robustness**!

Table 5: Take the tetrahedron as an example and compare the cases of EGNN, HEGNN$_{l=3}$, and HEGNN$_{l\leq 3}$ when adding noise perturbations. Here, ε represents the ratio of noise, and the modulus of the noise obeys $\mathcal{N}(0, \varepsilon \cdot \mathbb{E}[||\vec{x} - \vec{x}_c||] \cdot I)$. It can be observed that the performance of EGNN is slightly improved in the presence of noise (from 50% when $\varepsilon = 0.01$ to 60% when $\varepsilon = 0.5$), while HEGNN demonstrates better robustness.

	$\varepsilon = 0.01$	$\varepsilon = 0.05$	$\varepsilon = 0.10$	$\varepsilon = 0.50$
EGNN	50.0 ± 0.0	45.0 ± 15.0	65.0 ± 22.9	60.0 ± 20.0
HEGNN$_{l=3}$	**100.0 ± 0.0**	**100.0 ± 0.0**	**100.0 ± 0.0**	**100.0 ± 0.0**
HEGNN$_{l\leq 3}$	**100.0 ± 0.0**	**100.0 ± 0.0**	**100.0 ± 0.0**	**100.0 ± 0.0**

Reference

[1] Puny O, Atzmon M, Smith E J, et al. Frame Averaging for Invariant and Equivariant Network Design. ICLR'22.

[2] Han J, Huang W, Xu T, et al. Equivariant graph hierarchy-based neural networks[C]. NeurIPS'22.

[3] Zhang Y, Cen J, Han J, et al. Improving Equivariant Graph Neural Networks on Large Geometric Graphs via Virtual Nodes Learning. ICML'24.

[4] Wang Y, Cheng C, Li S, et al. Neural P3M: A Long-Range Interaction Modeling Enhancer for Geometric GNNs. NeurIPS'24.

[5] Joshi C K, Bodnar C, Mathis S V, et al. On the expressive power of geometric graph neural networks. ICML'23.

[6] Satorras V G, Hoogeboom E, Welling M. E (n) equivariant graph neural networks. ICML'21.

[7] Frank J T, Unke O T, Müller K R, et al. A Euclidean transformer for fast and stable machine learned force fields. Nature Communications'24.