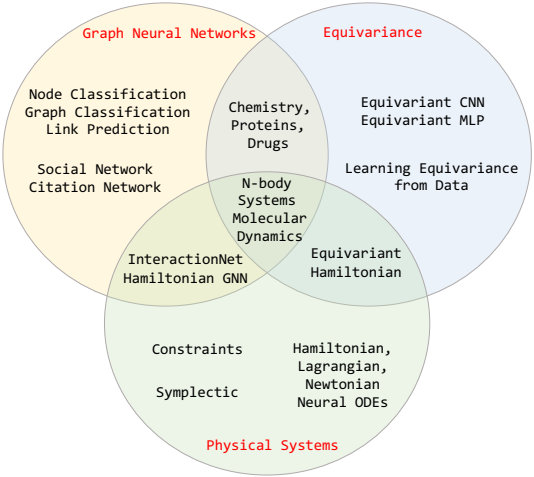
Equivariant Graph Hierarchy-Based Neural Networks

Jiaqi Han
hanjq21@mails.tsinghua.edu.cn

February 17, 2022

Backgrounds

Reasoning about the relations and dynamics of interacting objects and physical systems is a vital topic in machine learning.

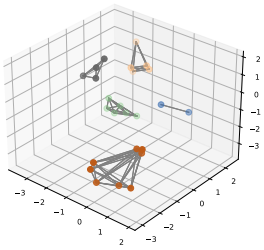


Challenges

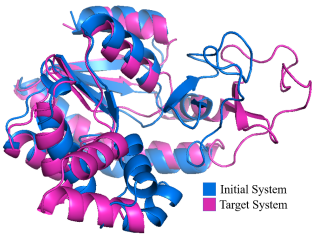
In this work, we focus on the systems that exhibit **hierarchical** structure, *e.g.*, the constrained N-body systems (or dubbed M-complex), proteins.

Challenges:

- Satisfying the equivariance constraints.
- Identifying the hierarchical structure.
- Incorporating the above two recipes in one network.



M-complex



Protein

Notations

Each input multi-body system is modeled as a graph $\mathcal{G}$ consisting of N particles (nodes) $\mathcal{V}$ and the interactions (edges) $\mathcal{E}$ among them. For each node i, it is assigned with a feature tuple $(\boldsymbol{Z}_i^{(0)}, \boldsymbol{h}_i^{(0)})$, where the directional matrix $\boldsymbol{Z}_i^{(0)} \in \mathbb{R}^{n \times m}$ is composed of m n-dimension vectors, such as the concatenation of position $\boldsymbol{x}_i \in \mathbb{R}^3$ and velocity $\boldsymbol{v}_i \in \mathbb{R}^3$; $\boldsymbol{h}_i \in \mathbb{R}^c$ is the non-directional feature, such as the atom number in molecules. The edges are represented by an adjacency matrix $\boldsymbol{A} \in \mathbb{R}^{N \times N}$. We henceforth abbreviate the entire information of a system, *i.e.*, $(\{\boldsymbol{Z}_i^{(0)}, \boldsymbol{h}_i^{(0)}\}_{i=1}^N, \boldsymbol{A})$ as the notation $\mathcal{G}^{\text{in}}$ if necessary.

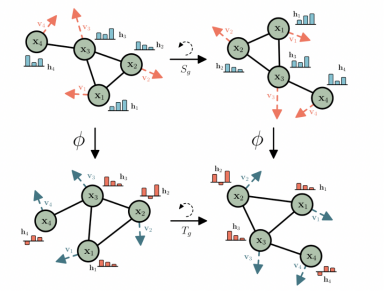
Equivariance and Invariance

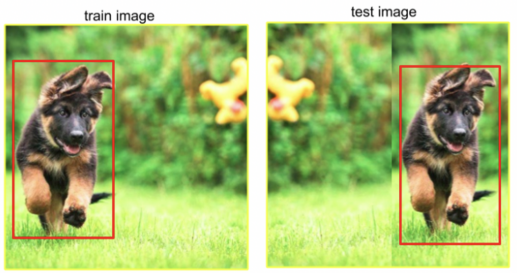
Group equivariant and invariant

Given a set of transformations $T_g : \mathcal{V} \to \mathcal{V}$ for $g \in G$, a function ϕ is called $G-$equivariant if for every g there exists a transformation $S_g : \mathcal{Y} \to \mathcal{Y}$ such that for all $g \in G, v \in \mathcal{V}$,

$$S_g[\phi(v)] = \phi(T_g[v])$$

Specifically, if $S_g = id$, then ϕ is $G-$invariant, where id is the identity transformation.





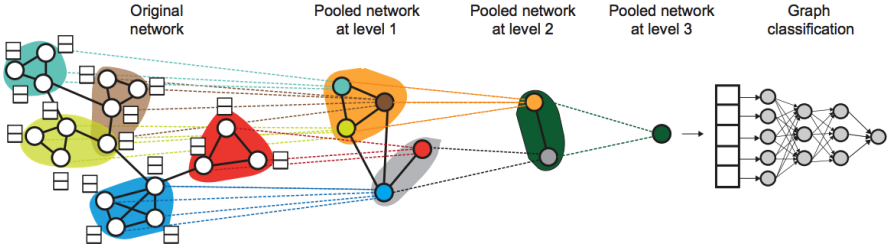
Equivariance

ϕ should be equivariant to any translation/reflection/rotation of the input states. By saying equivariance, we imply

$$\phi(\{g \cdot \boldsymbol{Z}_i^{(0)}\}_{i=1}^N, \cdots) = g \cdot \phi(\{\boldsymbol{Z}_i^{(0)}\}_{i=1}^N, \cdots),$$

where $g \cdot \boldsymbol{Z}_i^{(0)}$ conducts the orthogonal transformation as $\boldsymbol{R}\boldsymbol{Z}_i^{(0)}$ for both the position and velocity vectors and is additionally implemented as the translation $\boldsymbol{x}_i + \boldsymbol{b}$ for the position vector; the ellipsis denotes the input variables uninfluenced by g, including $\boldsymbol{h}_i^{(0)}$ and $\boldsymbol{A}$.

Hierarchical Learning in GNNs: DiffPool



$$\boldsymbol{A}^{(l+1)}, \boldsymbol{Z}^{(l+1)} = \text{DiffPool}(\boldsymbol{A}^{(l)}, \boldsymbol{Z}^{(l)})$$

$$\boldsymbol{Z}^{(l+1)} = \boldsymbol{S}^{(l)\top} \boldsymbol{Z}^{(l)},$$
$$\boldsymbol{A}^{(l+1)} = \boldsymbol{S}^{(l)\top} \boldsymbol{A}^{(l)} \boldsymbol{S}^{(l)}.$$

$\boldsymbol{S}^{(l)}$ is a GNN-parameterized assignment score matrix for the l-th layer.

Equivariant Matrix Message Passing (EMMP)

Given input features $\{(\boldsymbol{Z}_i, \boldsymbol{h}_i)\}_{i=1}^N$, EMMP performs information aggregation on the same graph to obtain the new features $\{(\boldsymbol{Z}'_i, \boldsymbol{h}'_i)\}_{i=1}^N$. The EMMP layer is undated by

$$\boldsymbol{H}_{ij} = \text{MLP}\left(\hat{\boldsymbol{Z}}_{ij}^\top \hat{\boldsymbol{Z}}_{ij}, \boldsymbol{h}_i, \boldsymbol{h}_j\right), \tag{1}$$

$$\boldsymbol{M}_{ij} = \hat{\boldsymbol{Z}}_{ij} \boldsymbol{H}_{ij}, \tag{2}$$

$$\boldsymbol{h}'_i = \text{MLP}(\boldsymbol{h}_i, \sum_{j \in \mathcal{N}(i)} \boldsymbol{H}_{ij}), \tag{3}$$

$$\boldsymbol{Z}'_i = \boldsymbol{Z}_i + \sum_{j \in \mathcal{N}(i)} \boldsymbol{M}_{ij}, \tag{4}$$

where $\hat{\boldsymbol{Z}}_{ij} = (\boldsymbol{Z}_i - \bar{\boldsymbol{Z}}, \boldsymbol{Z}_j - \bar{\boldsymbol{Z}})$ is a concatenation of the translated matrices on the edge ij. $\bar{\boldsymbol{Z}}$ is the mean of all nodes for the position vectors and zero for other vectors.
EMMP is equivariant *w.r.t.* E(n).

E-Pool

The role of E-Pool is to coarsen the low-level system $\mathcal{G}^{\text{low}} = (\{(\boldsymbol{Z}_i^{\text{low}}, \boldsymbol{h}_i^{\text{low}})\}_{i=1}^N, \boldsymbol{A}^{\text{low}})$ into an abstract and high-level system $\mathcal{G}^{\text{high}} = (\{(\boldsymbol{Z}_i^{\text{high}}, \boldsymbol{h}_i^{\text{high}})\}_{i=1}^K, \boldsymbol{A}^{\text{high}})$ with fewer particles, $K < N$. We proceed the following equations:

$$\{\boldsymbol{Z}_i^{'}, \boldsymbol{h}_i^{'}\}_i^N = \text{EMMP}(\{\boldsymbol{Z}_i^{\text{low}}, \boldsymbol{h}_i^{\text{low}}\}_i^N, \boldsymbol{A}^{\text{low}}), \tag{5}$$

$$\boldsymbol{s}_i = \text{SoftMax}(\text{MLP}(\boldsymbol{h}_i^{'})), \tag{6}$$

$$\boldsymbol{Z}_j^{\text{high}} = \frac{1}{\sum_{i=1}^N s_{ij}} \sum_{i=1}^N s_{ij} \boldsymbol{Z}_i^{'}, \tag{7}$$

$$\boldsymbol{h}_j^{\text{high}} = \frac{1}{\sum_{j=1}^N s_{ij}} \sum_{i=1}^N s_{ij} \boldsymbol{h}_i^{\text{low}}, \tag{8}$$

$$\boldsymbol{A}^{\text{high}} = \boldsymbol{S}^\top \boldsymbol{A}^{\text{low}} \boldsymbol{S}, \tag{9}$$

where the score matrix is given by $\boldsymbol{S} = [s_{ij}]_{N \times K}$, and $\boldsymbol{s}_i$ is its i-th row.

E-UpPool

E-UpPool maps the information of the high-level system $\mathcal{G}^{\text{high}}$ back to the original system space $\mathcal{G}^{\text{low}}$, leading to an output system $\mathcal{G}^{\text{out}}$. Particularly,

$$\boldsymbol{Z}_i^{\text{agg}} = \sum_{j=1}^{K} s_{ij} \boldsymbol{Z}_j^{\text{high}}, \tag{10}$$

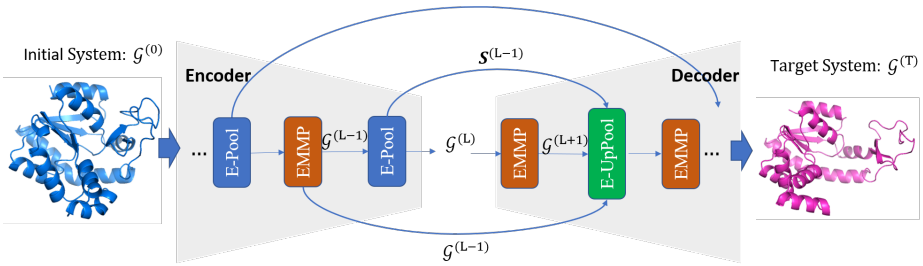
$$\boldsymbol{h}_i^{\text{agg}} = \sum_{j=1}^{K} s_{ij} \boldsymbol{h}_j^{\text{high}}, \tag{11}$$

$$\boldsymbol{h}_i^{\text{out}} = \text{MLP}\left(\hat{\boldsymbol{Z}}_i^{\top} \hat{\boldsymbol{Z}}_i, \boldsymbol{h}_i^{\text{low}}, \boldsymbol{h}_i^{\text{agg}}\right), \tag{12}$$

$$\boldsymbol{Z}_i^{\text{out}} = \hat{\boldsymbol{Z}}_i \boldsymbol{h}_i^{\text{out}}, \tag{13}$$

where $\hat{\boldsymbol{Z}}_i = [\boldsymbol{Z}_i^{\text{low}} - \bar{\boldsymbol{Z}}^{\text{low}}; \boldsymbol{Z}_i^{\text{agg}} - \bar{\boldsymbol{Z}}^{\text{agg}}]$ is the column-wise concatenation of the mean-translated low-level matrix $\boldsymbol{Z}_i^{\text{low}}$ and the high-level matrix $\boldsymbol{Z}_i^{\text{agg}}$.

Overall Architecture



The training objective of EGHN is given by:

$$\mathcal{L} = \sum_{i=1}^{N} \|\boldsymbol{Z}_i^{\text{out}} - \boldsymbol{Z}_i^{\text{gt}}\|_F^2 + \lambda \sum_{l=1}^{L} \|(\boldsymbol{S}^{(l)})^\top \boldsymbol{A}^{(l-2)} \boldsymbol{S}^{(l)} - \boldsymbol{I}\|_F^2,$$

Experiments: M-complex

Table: Prediction error ($\times 10^{-2}$) on various types of simulated datasets. The "Multiple System" contains $J = 5$ different systems. For each column, $(M, N/M)$ indicates that each system contains M complexes of average size N/M. Results averaged across 3 runs. "OOM" denotes out of memory.

	Single System				Multiple Systems			
	(3, 3)	(5, 5)	(5, 10)	(10, 10)	(3, 3)	(5, 5)	(5, 10)	(10, 10)
Linear	35.15±0.01	35.22±0.00	30.14±0.00	31.44±0.01	35.91±0.01	35.29±0.01	30.88±0.01	32.49±0.01
TFN	25.11±0.15	29.35±0.17	26.01±0.22	OOM	27.33±0.21	29.01±0.13	25.57±0.14	OOM
SE(3)-Tr.	27.12±0.26	28.87±0.09	24.48±0.35	OOM	28.14±0.16	28.66±0.10	25.00±0.28	OOM
GNN	16.00±0.11	17.55±0.19	16.15±0.08	15.91±0.15	16.76±0.13	17.58±0.11	16.55±0.21	16.05±0.16
RF	14.20±0.09	18.37±0.12	17.08±0.03	18.57±0.30	15.17±0.10	18.55±0.12	17.24±0.11	19.34±0.25
EGNN	12.69±0.19	15.37±0.13	15.12±0.11	14.64±0.27	13.33±0.12	15.48±0.16	15.29±0.12	15.02±0.18
EGHN	**11.58**±0.01	**14.42**±0.08	**14.29**±0.40	**13.09**±0.66	**12.80**±0.56	**14.85**±0.03	**14.50**±0.08	**13.11**±0.92

Experiments: M-complex

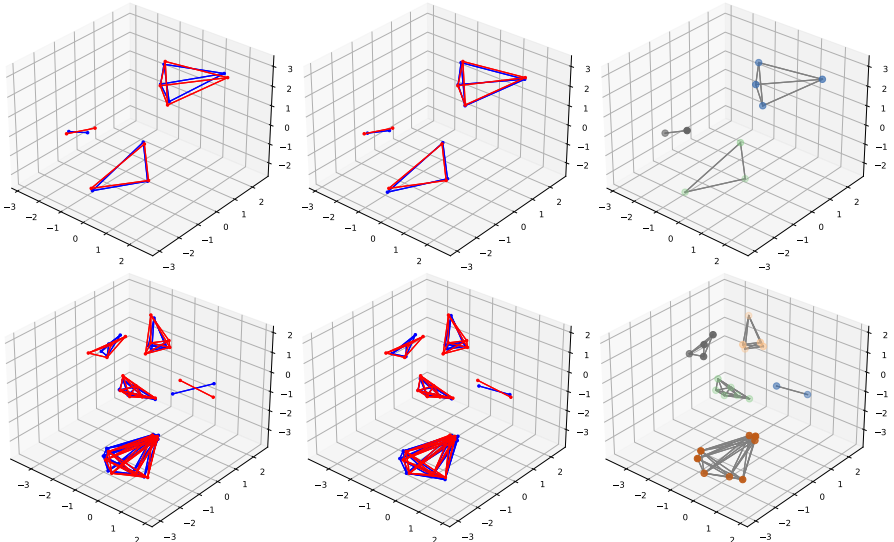


Figure: Visualization on M-complex systems. Left: EGNN; Middle: EGHN; Right: clusters.

Experiments: Motion Capture

Table: Prediction error ($\times 10^{-2}$) on the motion capture dataset. Results averaged across 3 runs.

	Subject #35 Walk	Subject #9 Run
GNN	36.1 $\pm$1.5	66.4 $\pm$2.2
RF	188.0 $\pm$1.9	521.3$\pm$2.3
TFN	32.0 $\pm$1.8	56.6 $\pm$1.7
SE(3)-Tr.	31.5 $\pm$2.1	61.2 $\pm$2.3
EGNN	28.7 $\pm$1.6	50.9 $\pm$0.9
GMN	21.6 $\pm$1.5	44.1 $\pm$2.3
EGHN	**8.5** $\pm$2.2	**25.9** $\pm$0.3

Experiments: Motion Capture

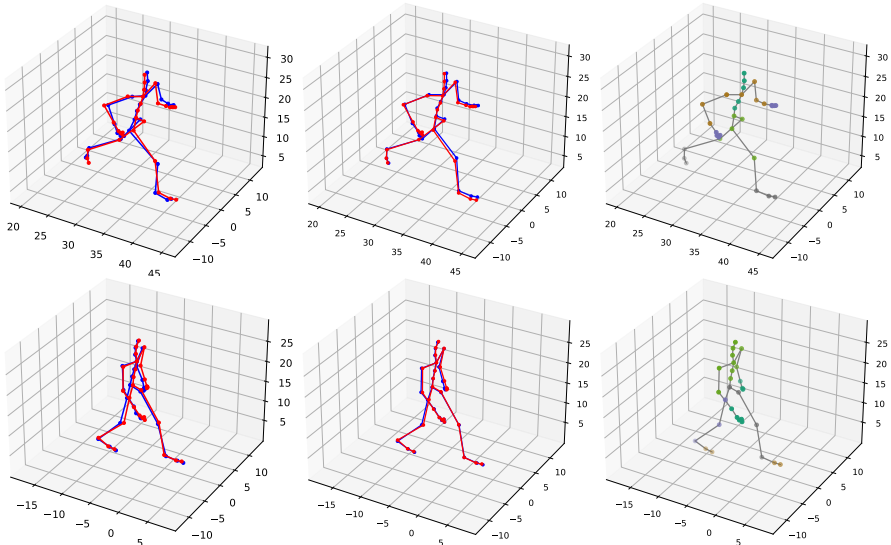


Figure: Visualization on Motion Capture. Left: EGNN; Middle: EGHN; Right: clusters.

Experiments: Protein MD

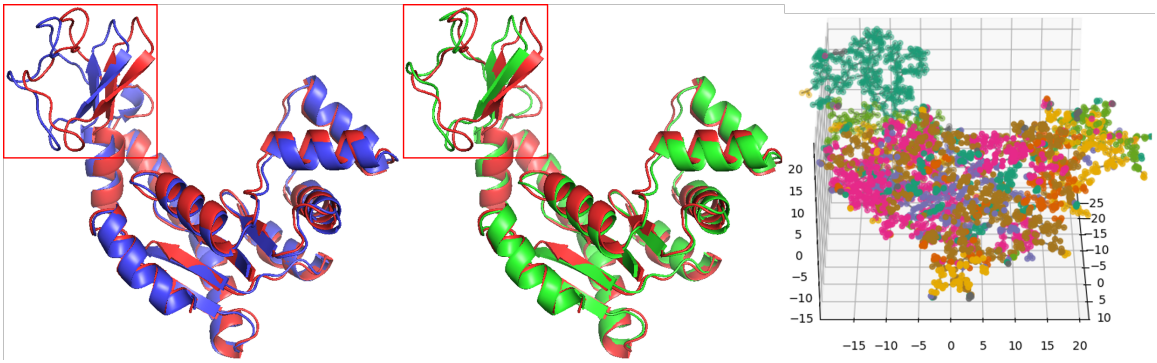


Figure: Visualization on the MDAnalysis dataset. *Left*: the prediction of EGNN. *Middle*: the prediction of EGHN. *Right*: the pooling results of EGHN with each color indicating a cluster. In the left and middle figure, ground truth in red, prediction for EGNN in blue, and prediction for EGHN in green.

Experiments: Protein MD

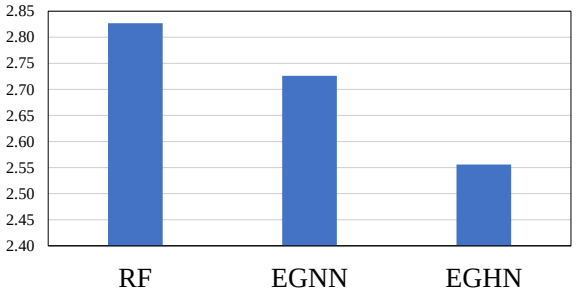


Figure: The prediction error of equivariant models on the protein molecular dynamics dataset.

The prediction error is computed as the MSE.