Schur Net

Effectively exploiting local structure for equivariance in higher order graph neural networks

Qingqi Zhang[1*], Ruize Xu[2*], Risi Kondor[1,2]

[1]Computational and Applied Mathematics
[2]Department of Computer Science
The University of Chicago

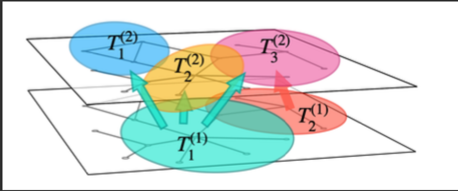
Nov. 10, 2024

Main contribution

- We developed *Schur* Net, a simple algorithm based on spectral graph theory for constructing a basis of automorphism equivariant operations on any possible subgraph.
- *Schur* Net is fast and easy to use on any user-defined subgraph template, bypassing the complex step of the group theoretical approach.
- On standard molecule benchmark datasets, we showed *Schur* Net achieves state-of-the-art performance among higher-order MPNNs.

Higher order MPNN

In higher-order MPNN, a **kth order tensor** is used to represent a **subgraph**, and the tensors communicate with each other by means of higher-order message passing (typically by intersection).



Permutation acting on a tensor

Definition ($\sigma \in S_m$ act on kth order tensor)

The action of a permutation $\sigma \in S_m$ on a kth order tensor $T \in R^{m^k}$ is a linear map R^{m^k} onto itself, denoted by $T^\sigma = \sigma \circ T$, which permutes the indexes of T simultaneously.

$$(T^\sigma)_{i_1,i_2,\ldots,i_k} = T_{\sigma^{-1}(i_1),\sigma^{-1}(i_2),\ldots,\sigma^{-1}(i_k)} \tag{1}$$

For example, when $k = 2$ and $T = A$ is the adjacency matrix, we have:

$$(A^\sigma)_{i,j} = A_{\sigma^{-1}(i),\sigma^{-1}(j)}$$

Equivariance

Definition (Equivariance)

We say a map $\phi : T^{\text{in}} \mapsto T^{\text{out}}$ is equivariant if

$$\phi(\sigma \circ T^{\text{in}}) = \sigma \circ \phi(T^{\text{in}}) = \sigma \circ T^{\text{out}} \quad \text{for all } \sigma \in S_m \tag{2}$$

The key concept of *equivariance* is the output is permuted as the manner of the input's permutation, or the map ϕ and action σ commutes, i.e., $\phi \circ \sigma = \sigma \circ \phi$.

The space of equivariant maps

Theorem (Homothety map on each stable subspaces are equivariant)

*Let G be a finite group acting on a vector space U by the linear action $\{g : U \to U\}_{g \in G}$ (In other words, $\rho : G \to \textbf{GL}(U)$ is a linear representation of G and write in short $g = \rho(g)$), assume that we have a decomposition of U into **stable** subspaces:*

$$U = U_1 \oplus \ldots \oplus U_p$$

Let $\phi : U \to U$ be a linear map that is a homothety on each U_i, i.e., $\phi(w) = \alpha_i w$ for some fixed scalar α_i and for $W \in U_i$. Then ϕ is equivariant to the action of G on U in the sense that $\phi(g(u)) = g(\phi(u))$ for any $u \in U$ and any $g \in G$.

The space of equivariant maps Cont.

Theorem (Necessary and sufficient condition for equivariant map)

Let G be a finite group acting on a vector space U by the linear action $\{g : U \to U\}_{g \in G}$ and assume the action can be decomposed into irreps $\rho_1, \cdots, \rho_p$ with multiplicities $\kappa_1, \cdots, \kappa_p$ and degree of ρ_i is d_i:

$$U = U_1 \oplus U_2 \oplus \cdots \oplus U_p, \quad U_i = \bigoplus_{j}^{\kappa_i} V_i^j$$

Let $\tau^i_{j \to j'} : V^i_j \to V^i_{j'}$ be the isomorphism between the two spaces. Then $\phi : U \to U$ is a equivariant map w.r.t. this group action if and only if ϕ is of the form:

$$\phi(v) = \sum_{j'} \alpha^i_{j,j'} \tau^i_{j \to j'}(v) \quad \textit{for } v \in V^i_j \tag{3}$$

for some fixed set of coefficient $\{\alpha^i_{j,j'}\}$.

Example on S_m

Example

(i) First order case: R^m decompose to two *irreps*: type (m) and $(m-1,1)$. There're two equivariant maps: the first is identity map and the second is $T \mapsto \sum_i T_i$.

(ii) For second order tensor $T \in R^{m^2}$,R^{m^2} decompose to: 2 isomorphic copies of type (m), 3 copies of $(m-1,1)$, 1 copy of $(m-2,2)$ and $(m-2,1,1)$, thus in total 15 equivariant maps in the basis.

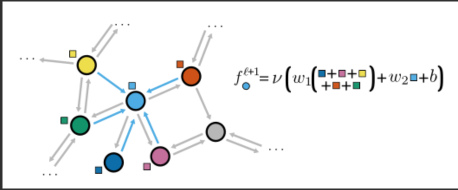
Rethinking MPNN: Equivariance with side information

MPNN is : $\phi(H) = AHW$ and

$$\phi(\sigma \circ H) = A^{\sigma} H^{\sigma} W = P_{\sigma} A P_{\sigma}^{T} P_{\sigma} H W = P_{\sigma} A H W = \sigma \circ \phi(H)$$

The key to equivariance is to use an graph object (namely adjacency matrix A) that permutes with $\sigma \in S_m$. The equivariance condition becomes:

$$\phi_{\sigma(A)}(\sigma \circ T^{\text{in}}) = \sigma \circ \phi_{A}(T^{\text{in}})$$



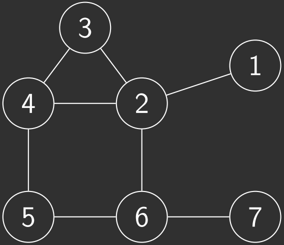
Automorphism group

Definition (Automorphism group of a graph)

Let $G = (V, E)$ with $|V| = m$ and adjacency matrix $A \in R^{m*m}$, the *automorphism* group of G is a subgroup of S_m with all $\sigma \in S_m$ such that:

$$A^\sigma = A$$

That is, after renumbering the vertices with σ, the edge set is still the same.



We only need equivariance w.r.t. the automorphism group, not S_m!

Eigenspaces of graph Laplacian are stable subspaces

We can build equivariant maps w.r.t. automorphism group by leveraging graph Laplacian as side information.

Lemma

Let S be an undirected graph with m vectices, Aut_S its automorphism group, and $L = D - A$ its combinatorial graph Laplacian. Assume L has t distinct eigenvalues $\lambda_1, \ldots, \lambda_t$ and corresponding subspaces $U_1, \ldots, U_t$. Then each U_i is invariant under the permutation (or first order) action of Aut_S on R^m and $R^m = U_1 \oplus \ldots \oplus U_t$.

Proof: by $\sigma \circ L = P_\sigma L P_\sigma^T = L$.

Schur Layer

Corollary

Let S, L be as in the lemma, and let $\phi : R^m \to R^m$ is defined by:

$$\phi(v) = \alpha_i v \quad \text{for } v \in U_i$$

Then ϕ is equivariant w.r.t. Aut_S. In matrix form, ϕ is given by:

$$\phi(T) = \sum_i M_i M_i^T T W_i$$

where $M = (M_1, \ldots, M_t)$ is orthonormal basis of $R^m = U_1 \oplus \ldots \oplus U_t$ (M_i corresponds to U_i and is given by eigenvalue decomposition), $T \in R^m$ and W_i is a scalar coefficient.

When we apply the same transformation to all occurrences of subgraph S in the graph, we call it *Schur* layer.

Analysis of *Schur* Layer

Graph	Aut_S	**# of distinct λ_i (*Schur* Layer)**	$\sum_i (\kappa_i)^2$ Group theoretical approach
6-cycle	D_6	4	4
5-cycle	D_5	3	3
4-cycle	D_4	3	3
3-cycle	D_3	2	2
5-star	S_4	3	5
4-star	S_3	3	5
5-cycle with one branch	S_2	6	20
6-cycle with one branch	S_2	7	29

The gap is because this approach can't take into account the isomorphic subspaces corresponding to the same type of *irrep*, thus ignoring the possible equivariant maps between V_j^i and $V_{j'}^i$.

Implementation

A few details:

1. Used *P*-tensor [1] framework to do message passing between subgraphs. *P*-tensor is a publicly available software designed for higher-order message passing, see https://github.com/risi-kondor/ptens.
2. Experiments used cycles of length $\{3-8, 11\}$ and branched cycles.
3. Added 0th order representation on nodes and edges in each layer.
4. Code is available at https://github.com/risilab/SchurNet.

[1] Andrew R. Hands, Tianyi Sun, Risi Kondor Proceedings of The 27th International Conference on Artificial Intelligence and Statistics, 2024.

Schur layer improves over *Linmaps*

We perform controlled experiments on various datasets and architectural settings, where we only replace *Linmaps* with *Schur* layer, but keep everything the same in each setting. *Linmaps* refers to equivariant maps w.r.t. full S_m, which is used in *P*-tensor paper.

Dataset	*Linmaps*	*Schur* Layer
Proteins	74.7 ± 3.8	$\mathbf{75.4 \pm 4.8}$
MUTAG	89.9 ± 5.5	$\mathbf{90.94 \pm 4.7}$
PTC_MR	61.1 ± 6.9	$\mathbf{64.6 \pm 5.9}$
NCI1	82.1 ± 1.8	$\mathbf{82.7 \pm 1.9}$

Table 8: Comparison of *Linmaps* and *Schur* Layer performance on TUdatasets. Numbers are binary classification accuracy.

Schur layer improves over *Linmaps* Cont.

Layer	Pass message when overlap ≥ 1	Pass message when overlap ≥ 2
Linmaps (baseline)	0.074 ± 0.008	0.074 ± 0.005
Schur layer	$\mathbf{0.070 \pm 0.006}$	$\mathbf{0.071 \pm 0.003}$

Table 1: Comparison between *Schur* Layer and *Linmaps* with different message passing schemes. The message passing scheme is a design choice in *P*-tensor framework, where the user can set when two subgraph's representations communicate. The mostly common use case is to require at least k vertices in the intersection of two subgraphs for them to communicate. Experiments on ZINC-12k dataset and all scores are test MAE. Cycle sizes of {3,4,5,6,7,8,11} are used.

Model	Test MAE
Linmaps	0.071 ± 0.004
Simple *Schur*-Net	0.070 ± 0.005
Linmap *Schur*-Net	$\mathbf{0.068 \pm 0.002}$
Complete *Schur*-Net	$\mathbf{0.064 \pm 0.002}$

Table 2: An experiment demonstrating different ways of using *Schur* layer. "Complete *Schur* Layer" means that we apply *Schur* Layer on the incoming messages together with the original cycle representation. "Linmap SchuLayer" means that we just apply the *Schur* Layer on the aggregated subgraph representation feature. "Simple *Schur* Layer" means we directly apply *Schur* Layer on the subgraph features without any preprocessing. We can observe that as the subgraph information diversifies, *Schur* layer tends to decouple the dense information better and results in better performance. The test MAE of *Linmaps* in this table is taken from [22].

Schur layer improves over *Linmaps* Cont.

Runtime comparison. Shows *Schur* layer didn't add significant extra computation cost while achieving better performance.

Dataset	*Linmaps*	*Schur* Layer
Zinc-12k	25.4s	27.6s
NCI1	9.5s	11.5s

Table: Runtime per epoch.

Benchmark results

Model	ZINC-12K MAE(↓)	OGB-HIV ROC-AUC(% ↑)
GCN	0.321 ± 0.009	76.07 ± 0.97
GIN	0.408 ± 0.008	75.58 ± 1.40
GINE	0.252 ± 0.014	75.58 ± 1.40
PNA	0.133 ± 0.011	79.05 ± 1.32
HIMP	0.151 ± 0.002	78.80 ± 0.82
N^2-GNN	$\mathbf{0.059 \pm 0.002}$	-
CIN	0.079 ± 0.006	80.94 ± 0.57
P-tensors	0.071 ± 0.004	80.76 ± 0.82
DS-GNN (EGO+)	0.105 ± 0.003	77.40 ± 2.19
DSS-GNN (EGO+)	0.097 ± 0.006	76.78 ± 1.66
GNN-AK+	0.091 ± 0.011	79.61 ± 1.19
SUN (EGO+)	0.084 ± 0.002	80.03 ± 0.55
Autobahn	0.106 ± 0.004	78.0 ± 0.30
Schur-Net	0.064 ± 0.002	$\mathbf{81.6 \pm 0.295}$

Table 3: Comparison of different models on the ZINC-12K and OGBG-MOLHIV datasets.

Conclusion

We proposed *Schur* Net, a higher-order GNN that:

- Utilizes the equivariance w.r.t. automorphism group to get more possible equivariant maps.
- Uses EVD of graph Laplacian to directly get equivariant maps, bypassing group theoretical steps.
- Achieves strong performance on benchmark datasets.

The limitations and future directions are:

- To theoretically explore what are the gaps in all scenarios and to find a way to incorporate the maps enabled by isomorphic eigenspaces under group action.
- To experiment on higher order tensor representation and more diverse subgraphs.

The End