Invariant and Equivariant Graph Networks

Haggai Maron, Heli Ben-Hamu, Nadav Shamir & Yaron Lipman

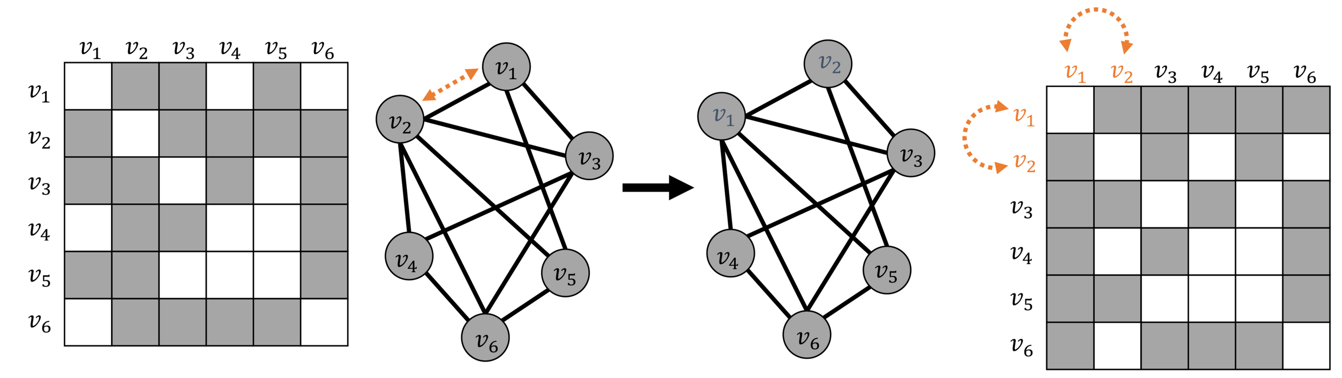
Weizmann Institute of Science

Introduction

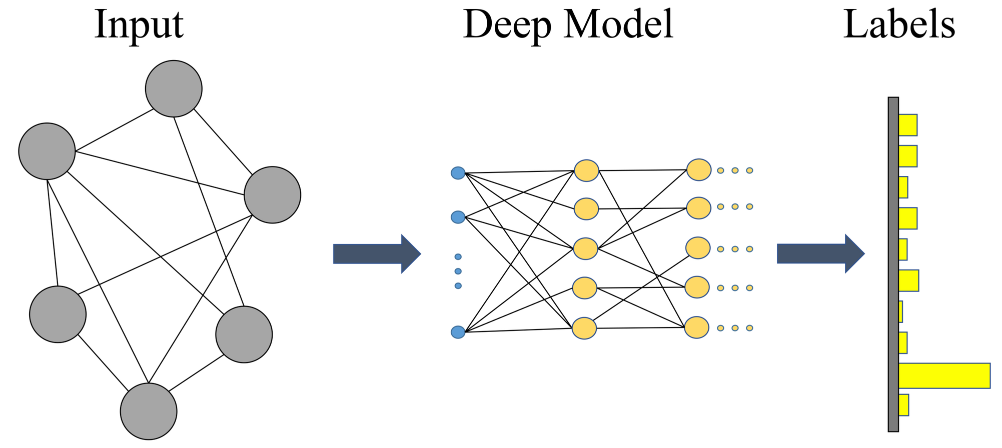
We consider the problem of graph learning, namely finding a functional relation between input graphs (more generally, hyper-graphs) $\mathcal{G}^\ell$ and corresponding targets T^ℓ.

Data. A (hyper-)graph data point $\mathcal{G} = (\mathbb{V}, \mathsf{A})$ consists of a set of n nodes $\mathbb{V}$, and values A attached to its *hyper-edges*. These values are encoded in a tensor A. For example, it is customary to represent a graph using a binary adjacency matrix A, where A_{ij} equals one if vertex i is connected to vertex j and zero otherwise. We denote the set of order-k tensors by $\mathbb{R}^{n^k}$.

Data symmetry. Relabeling the nodes does not change the graph.



Task. The task at hand is constructing a functional relation $f(\mathsf{A}^\ell) \approx T^\ell$, where f is a neural network.



invariance. If $T^\ell = t^\ell$ is a single output response then it is natural to ask that f is *order invariant*. For graph data, f is order invariant if it satisfies $f(\boldsymbol{P}^T\boldsymbol{A}\boldsymbol{P}) = f(\boldsymbol{A})$ for any permutation matrix $\boldsymbol{P}$.

equivariance. If the targets T^ℓ specify output response in a form of a tensor, $T^\ell = \mathsf{T}^\ell$, then it is natural to ask that f is *order equivariant*, that is, f commutes with the renumbering of nodes operator acting on tensors. Using the above adjacency matrix example f is equivariant if it satisfies $f(\boldsymbol{P}^T\boldsymbol{A}\boldsymbol{P}) = \boldsymbol{P}^T f(\boldsymbol{A})\boldsymbol{P}$ for every permutation matrix $\boldsymbol{P}$.

Goal

Following the standard paradigm of neural-networks where a network f is defined by alternating compositions of linear layers and non-linear activations, we set as a goal to characterize all *linear* invariant and equivariant layers.

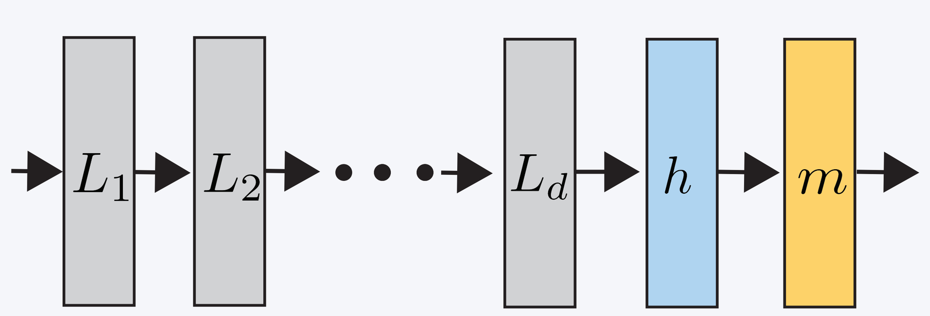


Figure 1. Invariant network architecture. L_i denotes linear equivariant layers, h denotes an invariant layer and m is a fully connected network

Contributions

In this paper, we provide a full characterization of permutation invariant and equivariant linear layers for general tensor input and output data. Our main contributions are :

1. We show that the space of equivariant linear layers $L : \mathbb{R}^{n^k} \to \mathbb{R}^{n^l}$ is of dimension $\mathrm{b}(k+l)$
2. We present a formula for an orthogonal basis for this space.
3. We show that our model can represent message passing layers [1] to an arbitrary precision on compact sets.

$\mathrm{b}(k)$ is the k-th *Bell number*: the number of possible partitions of a set of size k.

Fixed-point equations

Let $\boldsymbol{L} \in \mathbb{R}^{1\times n^2}$ denote the matrix representing a general linear operator $L : \mathbb{R}^{n\times n} \to \mathbb{R}$ in the standard basis, then L is order invariant iff $\boldsymbol{L}\mathrm{vec}(\boldsymbol{P}^T\boldsymbol{A}\boldsymbol{P}) = \boldsymbol{L}\mathrm{vec}(\boldsymbol{A})$, which is equivalent to:

$$\boldsymbol{P} \otimes \boldsymbol{P}\, \mathrm{vec}(\boldsymbol{L}) = \mathrm{vec}(\boldsymbol{L}) \tag{1}$$

for every permutation matrix $\boldsymbol{P}$.

In the general case for order-k tensor data $\mathsf{A} \in \mathbb{R}^{n^k}$ over one node set, $\mathbb{V}$, we have:

$$\text{invariant } \boldsymbol{L}: \quad \boldsymbol{P}^{\otimes k}\mathrm{vec}(\boldsymbol{L}) = \mathrm{vec}(\boldsymbol{L}) \tag{2}$$

$$\text{equivariant } \boldsymbol{L}: \quad \boldsymbol{P}^{\otimes 2k}\mathrm{vec}(\boldsymbol{L}) = \mathrm{vec}(\boldsymbol{L}) \tag{3}$$

for every permutation matrix $\boldsymbol{P}$, where $\boldsymbol{P}^{\otimes k} = \overbrace{\boldsymbol{P} \otimes \cdots \otimes \boldsymbol{P}}^{k}$.

The fixed-point equations can be equivalently formulated as $p \star \boldsymbol{L} = \boldsymbol{L}$, for any permutation p where $\star$ denotes the relabeling operator.

Solving the fixed point equations

Equality patterns. For multi-indices $\boldsymbol{a}, \boldsymbol{b} \in [n]^\ell$ we set $\boldsymbol{a} \sim \boldsymbol{b}$ iff $\boldsymbol{a}, \boldsymbol{b}$ have the same *equality pattern*, that is $\boldsymbol{a}_i = \boldsymbol{a}_j \Leftrightarrow \boldsymbol{b}_i = \boldsymbol{b}_j$ for all $i, j \in [\ell]$.

Each equivalence class can be represented by a unique partition of the set $[\ell]$ where each set in the partition indicates maximal set of identical values. Let us exemplify. For $\ell = 2$ we have two equivalence classes $\gamma_1 = \{\{1\}, \{2\}\}$ and $\gamma_2 = \{\{1, 2\}\}$; γ_1 represents all multi-indices (i, j) where $i \neq j$, while γ_2 represents all multi-indices (i, j) where $i = j$.

Indicator tensors. For each equivalence class $\gamma \in [n]^\ell/_\sim$ we define an order-ℓ tensor $\mathsf{B}^\gamma \in \mathbb{R}^{n^\ell}$ by setting:

$$\mathsf{B}^\gamma_{\boldsymbol{a}} = \begin{cases} 1 & \boldsymbol{a} \in \gamma \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

Since we have a tensor B^γ for every equivalence class γ, and the equivalence classes are in one-to-one correspondence with partitions of the set $[\ell]$, we have $\mathrm{b}(\ell)$ tensors.

Proposition. The tensors B^γ form an orthogonal basis to the solution set of fixed point equations. The dimension of the solution set is therefore $\mathrm{b}(\ell)$.

Characterizing Equivariant layers

Theorem. The space of equivariant linear layers $\mathbb{R}^{n^k} \to \mathbb{R}^{n^l}$ is of dimension $\mathrm{b}(k+l)$ with basis elements B^γ, where γ are equivalence classes in $([n]^{k+l}/_\sim)$.

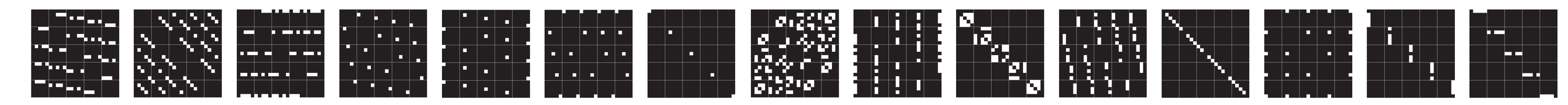


Figure 2. The full basis for equivariant linear layers for edge-value data $\mathsf{A} \in \mathbb{R}^{n\times n}$, for $n = 5$. The purely linear 15 basis elements, B^μ, are represented by $n^2 \times n^2$ matrices.

Generalization to multiple sets

Theorem The linear space of invariant linear layers $L : \mathbb{R}^{n_1^{k_1}\times n_2^{k_2}\times\cdots\times n_m^{k_m}} \to \mathbb{R}$ is of dimension $\Pi_{i=1}^m \mathrm{b}(k_i)$. The equivariant linear layers $L : \mathbb{R}^{n_1^{k_1}\times n_2^{k_2}\times\cdots\times n_m^{k_m}} \to \mathbb{R}^{n_1^{l_1}\times n_2^{l_2}\times\cdots\times n_m^{l_m}}$ has dimension $\Pi_{i=1}^m \mathrm{b}(k_i + l_i)$. Orthogonal bases for these layers are listed in the paper.

Previous work

The case of node-value input was treated in the pioneering works of [6, 4]. The general equivariant tensor case was partially treated in [3]. In [2] the authors provide an impressive generalization of the case of node-value data to several node sets.

Experiments

Table 1. Graph Classification Results on the datasets from [5]

dataset	MUTAG	PTC	PROTEINS	NCI1	NCI109	COLLAB	IMDB-B	IMDB-M
size	188	344	1113	4110	4127	5000	1000	1500
classes	2	2	2	2	2	3	2	3
avg node #	17.9	25.5	39.1	29.8	29.6	74.4	19.7	13
Results								
DGCNN	85.83±1.7	58.59±2.5	75.54±0.9	74.44±0.5	NA	73.76±0.5	70.03±0.9	47.83±0.9
PSCN (k=10)	88.95±4.4	62.29±5.7	75±2.5	76.34±1.7	NA	72.6±2.2	71±2.3	45.23±2.8
DCNN	NA	NA	61.29±1.6	56.61± 1.0	NA	52.11±0.7	49.06±1.4	33.49±1.4
ECC	76.11	NA	NA	76.82	75.03	NA	NA	NA
DGK	87.44±2.7	60.08±2.6	75.68±0.5	80.31±0.5	80.32±0.3	73.09±0.3	66.96±0.6	44.55±0.5
DiffPool	NA	NA	78.1	NA	NA	75.5	NA	NA
CCN	91.64±7.2	70.62±7.0	NA	76.27±4.1	75.54±3.4	NA	NA	NA
GK	81.39±1.7	55.65±0.5	71.39±0.3	62.49±0.3	62.35±0.3	NA	NA	NA
RW	79.17±2.1	55.91±0.3	59.57±0.1	> 3 days	NA	NA	NA	NA
PK	76±2.7	59.5±2.4	73.68±0.7	82.54±0.5	NA	NA	NA	NA
WL	84.11±1.9	57.97±2.5	74.68±0.5	84.46±0.5	85.12±0.3	NA	NA	NA
FGSD	92.12	62.80	73.42	79.80	78.84	80.02	73.62	52.41
AWE-DD	NA	NA	NA	NA	NA	73.93±1.9	74.45 ±5.8	51.54 ±3.6
AWE-FB	87.87±9.7	NA	NA	NA	NA	70.99 ± 1.4	73.13 ±3.2	51.58 ± 4.6
ours	84.61±10	59.47±7.3	75.19±4.3	73.71±2.6	72.48±2.5	77.92±1.7	71.27±4.5	48.55±3.9

References

[1] Justin Gilmer, Samuel S Schoenholz, Patrick F Riley, Oriol Vinyals, and George E Dahl. Neural message passing for quantum chemistry. In *Proceedings of the 34th International Conference on Machine Learning-Volume 70*, pages 1263–1272. JMLR. org, 2017.

[2] Jason S. Hartford, Devon R. Graham, Kevin Leyton-Brown, and Siamak Ravanbakhsh. Deep models of interactions across sets. In *ICML*, 2018.

[3] Risi Kondor, Hy Truong Son, Horace Pan, Brandon Anderson, and Shubhendu Trivedi. Covariant compositional networks for learning graphs. *arXiv preprint arXiv:1801.02144*, 2018.

[4] Charles R Qi, Hao Su, Kaichun Mo, and Leonidas J Guibas. Pointnet: Deep learning on point sets for 3d classification and segmentation. *Proc. Computer Vision and Pattern Recognition (CVPR), IEEE*, 1(2):4, 2017.

[5] Pinar Yanardag and S.V.N. Vishwanathan. Deep Graph Kernels. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '15*, 2015.

[6] Manzil Zaheer, Satwik Kottur, Siamak Ravanbakhsh, Barnabas Poczos, Ruslan R Salakhutdinov, and Alexander J Smola. Deep sets. In *Advances in Neural Information Processing Systems*, pages 3391–3401, 2017.