Universal Invariant and Equivariant Graph Neural Networks

Nicolas Keriven[1], Gabriel Peyré[2]

[1]CNRS, GIPSA-lab, Grenoble [2]CNRS, École Normale Supérieure, Paris

Summary

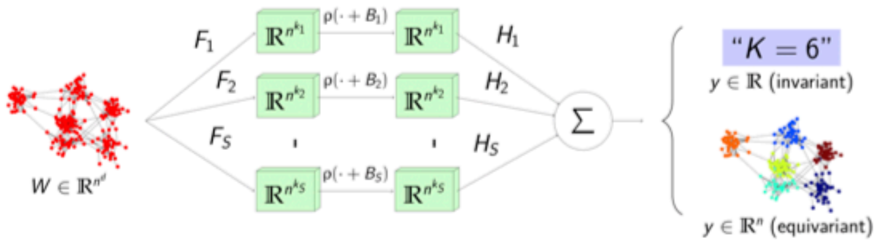
We prove **universal approximation theorems** for 1-layer **invariant and equivariant Graph Neural Networks**.

- The studied (theoretical) GNNs have **unbounded width** and **tensorization order**.
- Results are uniformly valid for (hyper)graphs of **varying number of nodes**, for a single set of parameters.
- The **equivariant case** is much more involved and requires a **new Stone-Weierstrass theorem**.

Notations

- **Graph**: $W \in \mathbb{R}^{n\times n}$ d-**Hypergraph**: $W \in \mathbb{R}^{n^d}$
- **Permutation**: bijection $\sigma : [n] \to [n]$
- Permuted (hyper)graph: $\sigma \star W \in \mathbb{R}^{n^d}$
- **Invariant function**: $f(\sigma \star W) = f(W)$
- **Equivariant function**: $f(\sigma \star W) = \sigma \star f(W)$

Studied architecture



1-layer GNNs:
$$\mathcal{F}_\ell = \left\{ f(\cdot) = \sum_{s=1}^{S} H_s\left[\rho\left(F_s[\cdot] + B_s\right)\right] + b \right\} \tag{1}$$

Param.
$$\begin{cases} S, k_s \in \mathbb{N} & \text{width of network, tensor orders (unbounded)} \\ F_s : \mathbb{R}^{n^d} \to \mathbb{R}^{n^{k_s}} & \textbf{linear, equivariant, can increase tensor order} \\ H_s : \mathbb{R}^{n^{k_s}} \to \mathbb{R}^{n^\ell} & \textbf{linear, invariant } (\ell = 0) \textbf{ or equivariant } (\ell = 1) \\ B_s \in \mathbb{R}^{n^{k_s}},\ b \in \mathbb{R}^{n^\ell} & \text{equivariant bias s.t. } \sigma \star B_s = B_s,\ \sigma \star b = b \end{cases}$$

ρ : non-linearity, any for which the MLP universality theorem applies.

Invariant and equivariant linear layers

Theorem (Maron et al. [1])

There is a basis of $b(k+p)$ equivariant linear operators $\mathbb{R}^{n^k} \to \mathbb{R}^{n^p}$, where $b(k)$ is the k^{th} Bell number.
(invariant case: just take $p = 0$)

- Does not depend on n. Ex: there are exactly 15 equivariant linear operators $\mathbb{R}^{n^2} \to \mathbb{R}^{n^2}$.
- the number of trainable parameters of f in (1) is
$$\sum_{s=1}^{S} \left(b(d + k_s) + b(k_s + \ell) + b(k_s) \right) + 1$$
- a GNN (1) with a fixed set of parameters can be applied to graphs of any size

Main results: universality of GNNs

Compact set of graphs: $\mathcal{G} = \left\{ W \in \mathbb{R}^{n^d} \ ;\ n \leqslant n_{\max},\ \|W\| \leqslant R \right\}$.

Theorem (Maron et al. [2]; Keriven and Peyré [3])

The set $\mathcal{F}_0$ of **invariant GNNs** is dense in the set of invariant continuous functions on $\mathcal{G}$ (for the sup norm).

Theorem (Keriven and Peyré [3])

The set $\mathcal{F}_1$ of **equivariant GNNs** is dense in the set of equivariant continuous functions on $\mathcal{G}$ (for the sup norm).

- A single set of parameters approximate functions on graphs of varying size uniformly well
- Equivariant case: much more difficult to prove (see below). Valid only for full group of permutations, and order-1 output $y \in \mathbb{R}^n$.

Sketch of proof

Invariant case

Apply Stone-Weierstrass theorem (like in Hornik et al. [4]), quotienting $\mathcal{G}$ by graph isomorphisms.

Theorem (Stone-Weierstrass)

An algebra of continuous functions that separates points is dense in the set of continuous functions.

- Algebra of GNNs *(aka "the cos trick")* EASY
  1. Authorize product of GNNs to obtain an algebra
  2. Prove universality for $\rho = \cos$
  3. A product of cos is also a sum!
  4. Approximate cos with any ρ using MLP universality theorem
- Separation of points HARD
  - "For any two distinct points, there is a function that distinguishes them".
  - Here: **"For two non-isomorphic graphs, there is a GNN that distinguishes them."**
    - We prove: **"Two graphs that coincide for every GNNs are isomorphic."**

Equivariant case

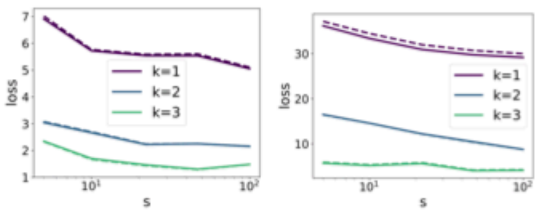
Impossible to use invariant polynomials [2] or regular Stone-Weierstrass theorem.

Theorem (Stone-Weierstrass for equivariant functions; Keriven and Peyré [3])

An algebra of equivariant continuous functions that separates points and separates coordinates is dense in the set of continuous functions.

- Separation of coordinates: "for a given graph W, and any two coordinates $1 \leqslant i, j \leqslant n$ that are not related by an automorphism of W (i.e. $\sigma \star W = W$), there is an equivariant GNN that distinguishes them."
- Proof: non-trivial adaptation of [5]

Numerics (toy)



Approximation results on synthetic data for invariant (left) or equivariant (right) GNNs. The tensorization order k plays a greater role than the width S.

Outlooks

- Convolutional GNNs
- Approximation power/stability with respect to weaker metrics on graphs (e.g. cut-metric)
- Behavior in the large-graph limit (see [6])

[1] Maron, Ben-Hamu, Shamir and Lipman. **Invariant and Equivariant Graph Neural Networks**. *ICLR*, 2019.

[2] Maron, Fetaya, Segol and Lipman. **On the Universality of Invariant Networks**. *ICML*, 2019.

[3] Keriven and Peyré. **Universal Invariant and Equivariant Graph Neural Networks**. *NeurIPS*, 2019.

[4] Hornik, Stinchcombe and White. **Multilayer Feedforward Networks are Universal Approximators**. *Neural Networks*, 1989.

[5] Brosowski and Deutsch. **An elementary proof of the Stone-Weierstrass Theorem**. *Proceedings of the American Mathematical Society*, 1981.

[6] Keriven, Bietti and Vaiter. **Convergence and Stability of Graph Convolutional Networks on Large Random Graphs**. *Preprint*, 2020.

nicolas.keriven@gipsa-lab.grenoble-inp.fr

This work was supported in part by the CFM-ENS chair *"Modèles et Sciences des données"* and the European Research Council, NORIA project.