Universal Invariant and Equivariant GNNs [1]

Nicolas Keriven, Gabriel Peyré

Presented by: Rozhan Akhound-Sadegh

UNIVERSITY OF WATERLOO | FACULTY OF MATHEMATICS

1
Introduction

Universal Approximation

Universal Approximation Theorem

MLPs with a single hidden layer are universal approximators

Universal Approximation

Universal Approximation Theorem

MLPs with a single hidden layer are universal approximators

- Maron et al. [2] showed certain invariant GNNs are universal approximators of invariant continuous functions on graphs

[2] Haggai Maron et al. "On the Universality of Invariant Networks". In: *Proceedings of the 36th International Conference on Machine Learning*. Ed. by Kamalika Chaudhuri and Ruslan Salakhutdinov. Vol. 97. Proceedings of Machine Learning Research. PMLR, Sept. 2019, pp. 4363–4371. URL: https://proceedings.mlr.press/v97/maron19a.html.

Universal Approximation

Universal Approximation Theorem

MLPs with a single hidden layer are universal approximators

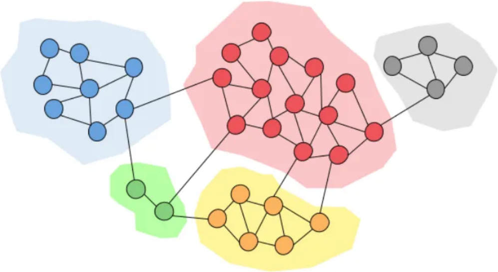
- Maron et al. [2] showed certain invariant GNNs are universal approximators of invariant continuous functions on graphs
- **Are *equivarient* GNNs universal approximators of *equivariant* continuous functions on graphs?**

[2] Haggai Maron et al. "On the Universality of Invariant Networks". In: *Proceedings of the 36th International Conference on Machine Learning*. Ed. by Kamalika Chaudhuri and Ruslan Salakhutdinov. Vol. 97. Proceedings of Machine Learning Research. PMLR, Sept. 2019, pp. 4363–4371. URL: https://proceedings.mlr.press/v97/maron19a.html.

Equivalance vs Invariance

- Community detection

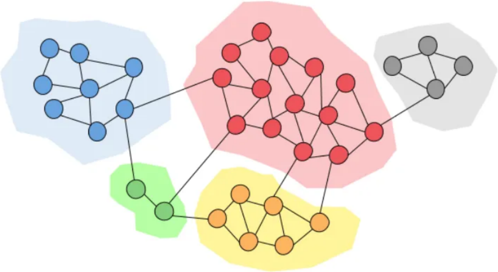
$$G = (V, E) \quad y : V \to \{1, \cdots, C\}$$



Equivariance vs Invariance

- Community detection

$$G = (V, E) \quad y : V \to \{1, \cdots, C\}$$



- Recommender Systems

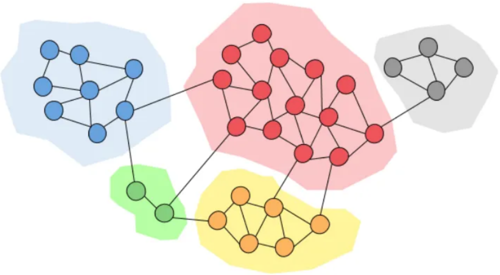
Equivalence vs Invariance

- Community detection

$$G = (V, E) \quad y : V \to \{1, \cdots, C\}$$

- Recommender Systems

- State Prediction



One-Layer GNN

- $f(G) = \sum_{s=1}^{S} H_s[\rho(F_s[G] + B_s)] + b$

One-Layer GNN

- $f(G) = \sum_{s=1}^{\mathcal{S}} H_s[\rho(F_s[G] + B_s)] + b$
- Hyper-graph: $G \in \mathbb{R}^{n^d}$ ($d = 2$ for classical graphs)

One-Layer GNN

- $f(G) = \sum_{s=1}^{\mathcal{S}} H_s[\rho(F_s[G] + B_s)] + b$
- Hyper-graph: $G \in \mathbb{R}^{n^d}$ ($d = 2$ for classical graphs)
- $F_s \colon \mathbb{R}^{n^d} \to \mathbb{R}^{n^{k_s}}$: linear equivariant functions

One-Layer GNN

- $f(G) = \sum_{s=1}^{\mathcal{S}} H_s[\rho(F_s[G] + B_s)] + b$
- Hyper-graph: $G \in \mathbb{R}^{n^d}$ ($d = 2$ for classical graphs)
- $F_s \colon \mathbb{R}^{n^d} \to \mathbb{R}^{n^{k_s}}$: linear equivariant functions
- $\rho \in \mathcal{F}_{\mathsf{MLP}}$: locally Lipschitz pointwise non-linearity where Universal Approximation Theorem for MLP applies

UNIVERSITY OF WATERLOO | FACULTY OF MATHEMATICS

One-Layer GNN

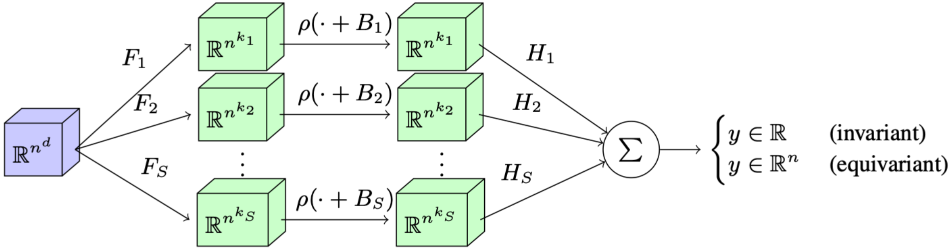
- $f(G) = \sum_{s=1}^{\mathcal{S}} H_s[\rho(F_s[G] + B_s)] + b$
- Hyper-graph: $G \in \mathbb{R}^{n^d}$ ($d = 2$ for classical graphs)
- $F_s \colon \mathbb{R}^{n^d} \to \mathbb{R}^{n^{k_s}}$: linear equivariant functions
- $\rho \in \mathcal{F}_{\text{MLP}}$: locally Lipschitz pointwise non-linearity where Universal Approximation Theorem for MLP applies
- $H_s \colon \mathbb{R}^{n^{k_s}} \to \mathbb{R}/\mathbb{R}^n$: invariant/equivariant linear operators

UNIVERSITY OF WATERLOO | FACULTY OF MATHEMATICS

One-Layer GNN

- $f(G) = \sum_{s=1}^{\mathcal{S}} H_s[\rho(F_s[G] + B_s)] + b$
- Hyper-graph: $G \in \mathbb{R}^{n^d}$ ($d = 2$ for classical graphs)
- $F_s \colon \mathbb{R}^{n^d} \to \mathbb{R}^{n^{k_s}}$: linear equivariant functions
- $\rho \in \mathcal{F}_{\text{MLP}}$: locally Lipschitz pointwise non-linearity where Universal Approximation Theorem for MLP applies
- $H_s \colon \mathbb{R}^{n^{k_s}} \to \mathbb{R}/\mathbb{R}^n$: invariant/equivariant linear operators
- $B_s = \sigma B_s$: equivariant Bias

Model of GNN


$\mathbb{R}^{n^d}$
F_1
F_2
F_S
$\mathbb{R}^{n^{k_1}}$
$\mathbb{R}^{n^{k_2}}$
$\mathbb{R}^{n^{k_S}}$
$\rho(\cdot + B_1)$
$\rho(\cdot + B_2)$
$\rho(\cdot + B_S)$
$\mathbb{R}^{n^{k_1}}$
$\mathbb{R}^{n^{k_2}}$
$\mathbb{R}^{n^{k_S}}$
H_1
H_2
H_S
Σ
$\begin{cases} y \in \mathbb{R} & \text{(invariant)} \\ y \in \mathbb{R}^n & \text{(equivariant)} \end{cases}$

Number of parameters

- Maron et al. [3] showed $f\colon \mathbb{R}^{n^k} \to \mathbb{R}^{n^\ell}$ lives in vector space of dimension $b(k+\ell)$

[3] Haggai Maron et al. "Invariant and Equivariant Graph Networks". In: *International Conference on Learning Representations*. 2018.

Number of parameters

- Maron et al. [3] showed $f\colon \mathbb{R}^{n^k} \to \mathbb{R}^{n^\ell}$ lives in vector space of dimension $b(k+\ell)$
- Notably, number of parameters doesn't depend on n

[3] Haggai Maron et al. "Invariant and Equivariant Graph Networks". In: *International Conference on Learning Representations*. 2018.

Number of parameters

- Maron et al. [3] showed $f: \mathbb{R}^{n^k} \to \mathbb{R}^{n^\ell}$ lives in vector space of dimension $b(k+\ell)$
- Notably, number of parameters doesn't depend on n
- GNN described by a single set of parameters can be applied to graphs of any size

[3] Haggai Maron et al. "Invariant and Equivariant Graph Networks". In: *International Conference on Learning Representations*. 2018.

Number of parameters

- Maron et al. [3] showed $f\colon \mathbb{R}^{n^k} \to \mathbb{R}^{n^\ell}$ lives in vector space of dimension $b(k+\ell)$
- Notably, number of parameters doesn't depend on n
- GNN described by a single set of parameters can be applied to graphs of any size
- We want to show that a GNN approximates a continuous function uniformly well for several n at once

[3] Haggai Maron et al. "Invariant and Equivariant Graph Networks". In: *International Conference on Learning Representations*. 2018.

8
Proof for Invariant Case

Proof Idea: Dense subsets

- If $\mathcal{A} \subset \mathcal{B}$ is dense in $\mathcal{B}$, then $\forall b \in \mathcal{B}, \forall \epsilon > 0, \exists a \in \mathcal{A}$ such that $\|a - b\| < \epsilon$

Proof Idea: Dense subsets

- If $\mathcal{A} \subset \mathcal{B}$ is dense in $\mathcal{B}$, then $\forall b \in \mathcal{B}, \forall \epsilon > 0, \exists a \in \mathcal{A}$ such that $\|a - b\| < \epsilon$

- We want to show for any continuous invariant function, there's a 1-layer GNN that approximates it

Proof Idea: Dense subsets

- If $\mathcal{A} \subset \mathcal{B}$ is dense in $\mathcal{B}$, then $\forall b \in \mathcal{B}, \forall \epsilon > 0, \exists a \in \mathcal{A}$ such that $\|a - b\| < \epsilon$
- We want to show for any continuous invariant function, there's a 1-layer GNN that approximates it

Proof Idea: Dense subsets

- If $\mathcal{A} \subset \mathcal{B}$ is dense in $\mathcal{B}$, then $\forall b \in \mathcal{B}, \forall \epsilon > 0, \exists a \in \mathcal{A}$ such that $\|a - b\| < \epsilon$
- We want to show for any continuous invariant function, there's a 1-layer GNN that approximates it

Proof Idea: Dense subsets

- If $\mathcal{A} \subset \mathcal{B}$ is dense in $\mathcal{B}$, then $\forall b \in \mathcal{B}, \forall \epsilon > 0, \exists a \in \mathcal{A}$ such that $\|a - b\| < \epsilon$

- We want to show for any continuous invariant function,

  there's a 1-layer GNN that approximates it

- Show class of invariant 1-layer GNNs is dense in space of real-valued continous invariant functions on graphs

Alternate Proof for [2]

Show class of invariant 1-layer GNNs is dense in space of real-valued continous invariant functions on graphs

- Equivalence class of isomorphic graphs
  $\mathcal{G}_{\mathsf{inv}} := \left\{ \mathcal{O}(G);\, G \in \mathbb{R}^{n^d} \text{ with } n \leq n_{\max},\, \|G\| \leq R \right\}$

Alternate Proof for [2]

Show class of invariant 1-layer GNNs is dense in space of real-valued continous invariant functions on graphs

- Equivalence class of isomorphic graphs
  $\mathcal{G}_{\text{inv}} := \left\{ \mathcal{O}(G);\, G \in \mathbb{R}^{n^d} \text{ with } n \leq n_{\max}, \|G\| \leq R \right\}$

Isomorphism

G_1, G_2 isomorphic if $G_1 = \sigma \star G_2$

Invariance

$f(G_1) = f(\sigma \star G_2) = f(G_2)$

Alternate Proof for [2]

Show class of invariant 1-layer GNNs is dense in space of real-valued continous invariant functions on graphs

- Equivalence class of isomorphic graphs
  $\mathcal{G}_{\text{inv}} := \left\{ \mathcal{O}(G); G \in \mathbb{R}^{n^d} \text{ with } n \leq n_{\max}, \|G\| \leq R \right\}$

- $\mathcal{O}$: set of permutations $\sigma : [n] \to [n]$

Alternate Proof for [2]

Show class of invariant 1-layer GNNs is dense in space of real-valued continous invariant functions on graphs

- Equivalence class of isomorphic graphs
  $\mathcal{G}_{\text{inv}} := \left\{ \mathcal{O}(G); G \in \mathbb{R}^{n^d} \text{ with } n \leq n_{\max}, \|G\| \leq R \right\}$
- $\mathcal{O}$: set of permutations $\sigma : [n] \to [n]$
- $\mathcal{O}(G) := \{\sigma \star G; \sigma \in \mathcal{O}\}$

Edit Distance

- Graph edit distance

$$d_{\textsf{edit}}(\mathcal{O}(G_1), \mathcal{O}(G_2)) := \min_{(o_1, o_2, \cdots, o_k) \in \mathcal{P}(G_1,\, G_2)} \sum_{i=1}^{k} c(o_i)$$

Edit Distance

- Graph edit distance

$$d_{\text{edit}}(\mathcal{O}(G_1), \mathcal{O}(G_2)) := \min_{(o_1, o_2, \cdots, o_k) \in \mathcal{P}(G_1, G_2)} \sum_{i=1}^{k} c(o_i)$$

- $\mathcal{P}(G_1, G_2)$: every sequence of operations to transform G_1 to G_2

Edit Distance

- Graph edit distance

$$d_{\textsf{edit}}(\mathcal{O}(G_1), \mathcal{O}(G_2)) := \min_{(o_1, o_2, \cdots, o_k) \in \mathcal{P}(G_1,\, G_2)} \sum_{i=1}^{k} c(o_i)$$

- $\mathcal{P}(G_1,\, G_2)$: every sequence of operations to transform G_1 to G_2
- $c(o)$: cost associated with each set of elementary operations o

Edit Distance

- Graph edit distance

$$d_{\mathsf{edit}}(\mathcal{O}(G_1), \mathcal{O}(G_2)) := \min_{(o_1, o_2, \cdots, o_k) \in \mathcal{P}(G_1, G_2)} \sum_{i=1}^{k} c(o_i)$$

- $\mathcal{P}(G_1, G_2)$: every sequence of operations to transform G_1 to G_2
- $c(o)$: cost associated with each set of elementary operations o
  - $c(\texttt{node\_addition}) = c$

Edit Distance

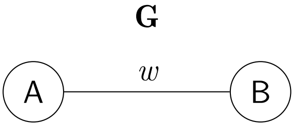
- Graph edit distance

$$d_{\mathsf{edit}}(\mathcal{O}(G_1), \mathcal{O}(G_2)) := \min_{(o_1, o_2, \cdots, o_k) \in \mathcal{P}(G_1, G_2)} \sum_{i=1}^{k} c(o_i)$$

- $\mathcal{P}(G_1, G_2)$: every sequence of operations to transform G_1 to G_2
- $c(o)$: cost associated with each set of elementary operations o
  - $c(\texttt{node\_addition}) = c$
  - $c(\texttt{edge\_weight\_modification}) = |w - w'|$

Edit Distance Example

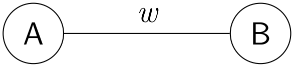
Edit Distance Example



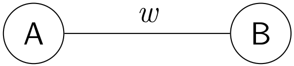
cost accumulated: c

Edit Distance Example

$\mathbf{G}$

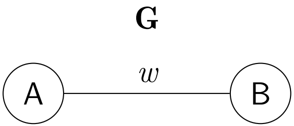


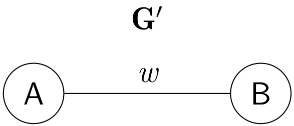
$\mathbf{G'}$



cost accumulated: $c + w$

Edit Distance Example





cost accumulated: $c + w$

$$d_{\text{edit}}(G, G') = c + w$$

Edit Distance Observations

Assume $d_{\text{edit}}(\mathcal{O}(G_1), \mathcal{O}(G_2)) < c$. Then the following hold:

- G_1 and G_2 have the same number of nodes

Edit Distance Observations

Assume $d_{\text{edit}}(\mathcal{O}(G_1), \mathcal{O}(G_2)) < c$. Then the following hold:

- G_1 and G_2 have the same number of nodes
- $d_{\text{edit}}(\mathcal{O}(G_1), \mathcal{O}(G_2)) = \min_{\sigma \in \mathcal{O}_n} \|G_1 - \sigma \star G_2\|_1$ since each edge must be transformed into another

Theorem 1

- For any $\rho \in \mathcal{F}_{\mathsf{MLP}}, \mathcal{N}_{\mathsf{inv}}(\rho)$ is dense in $\mathcal{C}(\mathcal{G}_{\mathsf{inv}}, d_{\mathsf{edit}})$

Theorem 1

- For any $\rho \in \mathcal{F}_{\mathsf{MLP}}$, $\mathcal{N}_{\mathsf{inv}}(\rho)$ is dense in $\mathcal{C}(\mathcal{G}_{\mathsf{inv}}, d_{\mathsf{edit}})$
- $\mathcal{N}_{\mathsf{inv}}(\rho)$: class of invariant 1-layer networks $f(G)$

Theorem 1

- For any $\rho \in \mathcal{F}_{\text{MLP}}, \mathcal{N}_{\text{inv}}(\rho)$ is dense in $\mathcal{C}(\mathcal{G}_{\text{inv}}, d_{\text{edit}})$
- $\mathcal{N}_{\text{inv}}(\rho)$: class of invariant 1-layer networks $f(G)$
- $\mathcal{C}(\mathcal{G}_{\text{inv}}, d_{\text{edit}})$: the space of real-valued functions on $\mathcal{G}_{\text{inv}}$ that are continuous w.r.t d_{edit}

Theorem 1

- For any $\rho \in \mathcal{F}_{\mathsf{MLP}}, \mathcal{N}_{\mathsf{inv}}(\rho)$ is dense in $\mathcal{C}(\mathcal{G}_{\mathsf{inv}}, d_{\mathsf{edit}})$
- $\mathcal{N}_{\mathsf{inv}}(\rho)$: class of invariant 1-layer networks $f(G)$
- $\mathcal{C}(\mathcal{G}_{\mathsf{inv}}, d_{\mathsf{edit}})$: the space of real-valued functions on $\mathcal{G}_{\mathsf{inv}}$ that are continuous w.r.t d_{edit}
- This shows any continuous invariant function is *uniformly* well-approximated by a GNN on $\mathcal{G}_{\mathsf{inv}}$

Stone-Weierstrass Theorem

- Conditions:

Stone-Weierstrass Theorem

- Conditions:
  - X compact, Hausdorff

Stone-Weierstrass Theorem

- Conditions:
  - X compact, Hausdorff
  - A subalgebra of $\mathcal{C}(X)$

Stone-Weierstrass Theorem

- Conditions:
  - X compact, Hausdorff
  - A subalgebra of $\mathcal{C}(X)$
  - A separates points in X

Stone-Weierstrass Theorem

- Conditions:
  - X compact, Hausdorff
  - A subalgebra of $\mathcal{C}(X)$
  - A separates points in X
- Then A is dense in $\mathcal{C}(X)$

Stone-Weierstrass Theorem

- Conditions:
  - $(\mathcal{G}_{\text{inv}}, d_{\text{edit}})$ compact, Hausdorff
  - $\mathcal{N}_{\text{inv}}$ subalgebra of $\mathcal{C}(\mathcal{G}_{\text{inv}}, d_{\text{edit}})$
  - $\mathcal{N}_{\text{inv}}$ separates points in $\mathcal{C}(\mathcal{G}_{\text{inv}}, d_{\text{edit}})$
- Then $\mathcal{N}_{\text{inv}}$ is dense in $\mathcal{C}(\mathcal{G}_{\text{inv}}, d_{\text{edit}})$

Stone-Weierstrass Theorem

- Conditions:
  - $(\mathcal{G}_{\text{inv}}, d_{\text{edit}})$ compact, Hausdorff
  - $\mathcal{N}_{\text{inv}}$ subalgebra of $\mathcal{C}(\mathcal{G}_{\text{inv}}, d_{\text{edit}})$
  - $\mathcal{N}_{\text{inv}}$ separates points in $\mathcal{C}(\mathcal{G}_{\text{inv}}, d_{\text{edit}})$
- Then $\mathcal{N}_{\text{inv}}$ is dense in $\mathcal{C}(\mathcal{G}_{\text{inv}}, d_{\text{edit}})$

Stone-Weierstrass Theorem

- Conditions:
  - $(\mathcal{G}_{\text{inv}}, d_{\text{edit}})$ compact, Hausdorff

Hasudorff

Separate points have separate neighborhoods
✓ all metric spaces are Hausdorff

UNIVERSITY OF WATERLOO | FACULTY OF MATHEMATICS

Compactness

Compactness Property

A continuous image of a compact space is compact

Compactness

Compactness Property

A continuous image of a compact space is compact

- $\mathcal{G}_{\mathsf{inv}}$ is the continuous image of $\bigcup_{n=1}^{n_{\max}} \left\{ G \in \mathbb{R}^{n^d} ; \|G\| \leq R \right\}$

Compactness

Compactness Property

A continuous image of a compact space is compact

- $\mathcal{G}_{\mathsf{inv}}$ is the continuous image of $\bigcup_{n=1}^{n_{\max}} \left\{ G \in \mathbb{R}^{n^d}; \|G\| \leq R \right\}$
- Why continuous?

Compactness

Compactness Property

A continuous image of a compact space is compact

- $\mathcal{G}_{\mathsf{inv}}$ is the continuous image of $\bigcup_{n=1}^{n_{\max}} \left\{ G \in \mathbb{R}^{n^d}; \|G\| \leq R \right\}$
- Why continuous?
  - Showed before that $d_{\mathsf{edit}}(\mathcal{O}(G_1), \mathcal{O}(G_2)) < c \implies G_1$ and G_2 have the same number of nodes

Compactness

Compactness Property

A continuous image of a compact space is compact

- $\mathcal{G}_{\text{inv}}$ is the continuous image of $\bigcup_{n=1}^{n_{\max}} \left\{ G \in \mathbb{R}^{n^d}; \|G\| \leq R \right\}$
- Why continuous?
  - Showed before that $d_{\text{edit}}(\mathcal{O}(G_1), \mathcal{O}(G_2)) < c \implies G_1$ and G_2 have the same number of nodes
  - $d_{\text{edit}}(\mathcal{O}(G_1), \mathcal{O}(G_2)) \leq \min_\sigma \|G_1 - \sigma \star G_2\|_1 \leq \|G_1 - G_2\|_1$

Compactness

Compactness Property

A continuous image of a compact space is compact

- $\mathcal{G}_{\text{inv}}$ is the continuous image of $\bigcup_{n=1}^{n_{\max}} \left\{ G \in \mathbb{R}^{n^d}; \|G\| \leq R \right\}$
- Why continuous?
  - Showed before that $d_{\text{edit}}(\mathcal{O}(G_1), \mathcal{O}(G_2)) < c \implies G_1$ and G_2 have the same number of nodes
  - $d_{\text{edit}}(\mathcal{O}(G_1), \mathcal{O}(G_2)) \leq \min_\sigma \|G_1 - \sigma \star G_2\|_1 \leq \|G_1 - G_2\|_1$
  - Therefore $G \mapsto \mathcal{O}(G)$ is continuous

Stone-Weierstrass Theorem

- Conditions:
  - $(\mathcal{G}_{\text{inv}}, d_{\text{edit}})$ compact, Hausdorff
  - $\mathcal{N}_{\text{inv}}$ subalgebra of $\mathcal{C}(\mathcal{G}_{\text{inv}}, d_{\text{edit}})$
  - $\mathcal{N}_{\text{inv}}$ separates points in $\mathcal{C}(\mathcal{G}_{\text{inv}}, d_{\text{edit}})$
- Then $\mathcal{N}_{\text{inv}}$ is dense in $\mathcal{C}(\mathcal{G}_{\text{inv}}, d_{\text{edit}})$

Stone-Weierstrass Theorem

- Conditions:
  - ✓ $(\mathcal{G}_{\text{inv}}, d_{\text{edit}})$ compact, Hausdorff
  - $\mathcal{N}_{\text{inv}}$ subalgebra of $\mathcal{C}(\mathcal{G}_{\text{inv}}, d_{\text{edit}})$
  - $\mathcal{N}_{\text{inv}}$ separates points in $\mathcal{C}(\mathcal{G}_{\text{inv}}, d_{\text{edit}})$
- Then $\mathcal{N}_{\text{inv}}$ is dense in $\mathcal{C}(\mathcal{G}_{\text{inv}}, d_{\text{edit}})$

Stone-Weierstrass Theorem

- Conditions:
  - ✓ $(\mathcal{G}_{\text{inv}}, d_{\text{edit}})$ compact, Hausdorff
  - $\mathcal{N}_{\text{inv}}$ subalgebra of $\mathcal{C}(\mathcal{G}_{\text{inv}}, d_{\text{edit}})$

Stone-Weierstrass Theorem

- Conditions:
  - ✓ $(\mathcal{G}_{\text{inv}}, d_{\text{edit}})$ compact, Hausdorff
  - $\mathcal{N}_{\text{inv}}$ subalgebra of $\mathcal{C}(\mathcal{G}_{\text{inv}}, d_{\text{edit}})$

Subalgebra

a vector subspace closed under the multiplication of vectors

Subalgebra

- Problem: $\mathcal{N}_{\text{inv}}(\rho)$ is not a subalgebra

Subalgebra

- Problem: $\mathcal{N}_{\mathsf{inv}}(\rho)$ is not a subalgebra
- Solution: Define $\mathcal{N}_{\mathsf{inv}}^{\otimes}(\rho)$ that is.

Subalgebra

- Problem: $\mathcal{N}_{\text{inv}}(\rho)$ is not a subalgebra
- Solution: Define $\mathcal{N}_{\text{inv}}^{\otimes}(\rho)$ that is.
- $f(G) = \sum_{s=1}^{S} H_s \left[\rho(F_{s1}[G] + B_{s1}) \otimes \cdots \otimes \rho(F_{sT_s}[G] + G_{sT_s})\right] + b$

Stone-Weierstrass Theorem

- Conditions:
  - ✓ $(\mathcal{G}_{\text{inv}}, d_{\text{edit}})$ compact, Hausdorff
  - $\mathcal{N}_{\text{inv}}^{\otimes}$ subalgebra of $\mathcal{C}(\mathcal{G}_{\text{inv}}, d_{\text{edit}})$
  - $\mathcal{N}_{\text{inv}}$ separates points in $\mathcal{C}(\mathcal{G}_{\text{inv}}, d_{\text{edit}})$
- Then $\mathcal{N}_{\text{inv}}$ is dense in $\mathcal{C}(\mathcal{G}_{\text{inv}}, d_{\text{edit}})$

Stone-Weierstrass Theorem

- Conditions:
  - ✓ $(\mathcal{G}_{\text{inv}}, d_{\text{edit}})$ compact, Hausdorff
  - ✓ $\mathcal{N}_{\text{inv}}^{\otimes}$ subalgebra of $\mathcal{C}(\mathcal{G}_{\text{inv}}, d_{\text{edit}})$
  - $\mathcal{N}_{\text{inv}}$ separates points in $\mathcal{C}(\mathcal{G}_{\text{inv}}, d_{\text{edit}})$
- Then $\mathcal{N}_{\text{inv}}$ is dense in $\mathcal{C}(\mathcal{G}_{\text{inv}}, d_{\text{edit}})$

Stone-Weierstrass Theorem

- Conditions:
  - ✓ $(\mathcal{G}_{\text{inv}}, d_{\text{edit}})$ compact, Hausdorff
  - ✓ $\mathcal{N}_{\text{inv}}^{\otimes}$ subalgebra of $\mathcal{C}(\mathcal{G}_{\text{inv}}, d_{\text{edit}})$
  - $\mathcal{N}_{\text{inv}}$ separates points in $\mathcal{C}(\mathcal{G}_{\text{inv}}, d_{\text{edit}})$
- Then $\mathcal{N}_{\text{inv}}$ is dense in $\mathcal{C}(\mathcal{G}_{\text{inv}}, d_{\text{edit}})$

Stone-Weierstrass Theorem

- Conditions:
  - ✓ $(\mathcal{G}_{\text{inv}}, d_{\text{edit}})$ compact, Hausdorff
  - ✓ $\mathcal{N}_{\text{inv}}^{\otimes}$ subalgebra of $\mathcal{C}(\mathcal{G}_{\text{inv}}, d_{\text{edit}})$
  - $\mathcal{N}_{\text{inv}}$ separates points in $\mathcal{C}(\mathcal{G}_{\text{inv}}, d_{\text{edit}})$

Separability

For all $x \neq y$ in X, there exists $f \in A$ such that $f(x) \neq f(y)$

Separability

- $\mathcal{N}_{\text{inv}}^{\otimes}(\rho)$ separates points

Separability

- $\mathcal{N}_{\text{inv}}^{\otimes}(\rho)$ separates points

Separability

For two distinct classes of isomorphic graphs in $\mathcal{G}_{\text{inv}}$ $(\mathcal{O}(G_1) \neq \mathcal{O}(G_2))$, there is a class of 1-layer GNNs $f \in \mathcal{N}^{\otimes}{}_{\text{inv}}$ with $f(\mathcal{O}(G_1)) \neq f(\mathcal{O}(G_2))$.

Separability

- $\mathcal{N}_{\text{inv}}^{\otimes}(\rho)$ separates points

Separability

For two distinct classes of isomorphic graphs in $\mathcal{G}_{\text{inv}}$ $(\mathcal{O}(G_1) \neq \mathcal{O}(G_2))$, there is a class of 1-layer GNNs $f \in \mathcal{N}^{\otimes}{}_{\text{inv}}$ with $f(\mathcal{O}(G_1)) \neq f(\mathcal{O}(G_2))$.

- Proof by contradiction and graphs G_1, G_2 are permutations of each other if $f(\mathcal{O}(G_1)) = f(\mathcal{O}(G_2))$ for $f \in \mathcal{N}_{\text{inv}}^{\otimes}(\rho_{\text{sig}})$

Stone-Weierstrass Theorem

- Conditions:
  - ✓ $(\mathcal{G}_{\mathsf{inv}}, d_{\mathsf{edit}})$ compact, Hausdorff
  - ✓ $\mathcal{N}_{\mathsf{inv}}^{\otimes}$ subalgebra of $\mathcal{C}(\mathcal{G}_{\mathsf{inv}}, d_{\mathsf{edit}})$
  - $\mathcal{N}_{\mathsf{inv}}^{\otimes}(\rho_{\mathsf{sig}})$ separates points in $\mathcal{C}(\mathcal{G}_{\mathsf{inv}}, d_{\mathsf{edit}})$
- Then $\mathcal{N}_{\mathsf{inv}}^{\otimes}(\rho_{\mathsf{sig}})$ is dense in $\mathcal{C}(\mathcal{G}_{\mathsf{inv}}, d_{\mathsf{edit}})$

Stone-Weierstrass Theorem

- Conditions:
  - ✓ $(\mathcal{G}_{\text{inv}}, d_{\text{edit}})$ compact, Hausdorff
  - ✓ $\mathcal{N}_{\text{inv}}^{\otimes}$ subalgebra of $\mathcal{C}(\mathcal{G}_{\text{inv}}, d_{\text{edit}})$
  - ✓ $\mathcal{N}_{\text{inv}}^{\otimes}(\rho_{\text{sig}})$ separates points in $\mathcal{C}(\mathcal{G}_{\text{inv}}, d_{\text{edit}})$
- Then $\mathcal{N}_{\text{inv}}^{\otimes}(\rho_{\text{sig}})$ is dense in $\mathcal{C}(\mathcal{G}_{\text{inv}}, d_{\text{edit}})$

Wrapping up Theorem 1

- So far proved $\mathcal{N}^{\otimes}_{\text{inv}}(\rho_{\text{sig}})$ is dense in $\mathcal{C}(\mathcal{G}_{\text{inv}}, d_{\text{edit}})$
- Can go from $\mathcal{N}^{\otimes}_{\text{inv}}(\rho_{\text{sig}})$ to $\mathcal{N}_{\text{inv}}(\rho)$
- Show $\mathcal{N}_{\text{inv}}(\rho)$ dense in $\mathcal{N}^{\otimes}_{\text{inv}}(\rho_{\text{sig}})$
- Have re-proved results from [2] but without fixing n

9
Proof for Equivariant Case

Equivariant Case

- Isomorphic graphs aren't equivalent anymore

Equivariant Case

- Isomorphic graphs aren't equivalent anymore

Equivariance

For isomorphic G_1, G_2, $(G_1 = \sigma \star G_2)$ we have

$$f(G_1) = f(\sigma \star G_2) = \sigma \star f(G_2) \neq f(G_2)$$

Equivariant Case

- Isomorphic graphs aren't equivalent anymore
- So define

$$\mathcal{G}_{\text{eq}} := \left\{ G \in \mathbb{R}^{n^d} ; n \leq n_{\max}, \|G\| \leq R \right\}$$

Equivariant Case

- Isomorphic graphs aren't equivalent anymore
- So define

$$\mathcal{G}_{\text{eq}} := \left\{ G \in \mathbb{R}^{n^d}; n \leq n_{\max}, \|G\| \leq R \right\}$$

- Metric:

$$d(G, G') = \begin{cases} \|G - G'\| & \text{if } G \text{ and } G' \text{ have the sane number of nodes,} \\ \infty & \text{otherwise.} \end{cases}$$

Equivariant Case

- Isomorphic graphs aren't equivalent anymore
- So define

$$\mathcal{G}_{\text{eq}} := \left\{ G \in \mathbb{R}^{n^d}; n \leq n_{\max}, \|G\| \leq R \right\}$$

- Metric:

$$d(G, G') = \begin{cases} \|G - G'\| & \text{if } G \text{ and } G' \text{ have the sane number of nodes,} \\ \infty & \text{otherwise.} \end{cases}$$

- H_s equivariant output operators

Theorem 2

For any $\rho \in \mathcal{F}_{\mathsf{MLP}}, \mathcal{N}_{\mathsf{eq}}(\rho)$ is dense in $\mathcal{C}(\mathcal{G}_{\mathsf{eq}}, d)$

Theorem 2

For any $\rho \in \mathcal{F}_{\mathsf{MLP}}, \mathcal{N}_{\mathsf{eq}}(\rho)$ is dense in $\mathcal{C}(\mathcal{G}_{\mathsf{eq}}, d)$

Problem

$\mathcal{C}(\mathcal{G}_{\mathsf{eq}}, d)$ is a particular subset of continuous functions

Classical Stone-Weierstrass theorem doesn't apply!

Theorem 2

For any $\rho \in \mathcal{F}_{\mathsf{MLP}}, \mathcal{N}_{\mathsf{eq}}(\rho)$ is dense in $\mathcal{C}(\mathcal{G}_{\mathsf{eq}}, d)$

Problem

$\mathcal{C}(\mathcal{G}_{\mathsf{eq}}, d)$ is a particular subset of continuous functions

Classical Stone-Weierstrass theorem doesn't apply!

Solution

Define specialized version of Stone-Weierstrass theorem for equivariant functions

Stone-Weierstrass for equivariant functions

- Additional condition: self-separability
- $f(G)$ can have different values on different coordinates
- Rest of the proof similar to invariant case

Conclusion

- Using higher-order tensorization inside GNN, we can achieve both invariance and equivariance

Conclusion

- Using higher-order tensorization inside GNN, we can achieve both invariance and equivariance
- Invariance (re)-proved using the Stone-Weierstrass theorem

Conclusion

- Using higher-order tensorization inside GNN, we can achieve both invariance and equivariance
- Invariance (re)-proved using the Stone-Weierstrass theorem
- Equivariance proved using a new generalized Stone-Weierstrass theorem

Conclusion

- Using higher-order tensorization inside GNN, we can achieve both invariance and equivariance
- Invariance (re)-proved using the Stone-Weierstrass theorem
- Equivariance proved using a new generalized Stone-Weierstrass theorem
- Works on graphs of varying size (don't need to fix the number of nodes)

Conclusion

- Using higher-order tensorization inside GNN, we can achieve both invariance and equivariance
- Invariance (re)-proved using the Stone-Weierstrass theorem
- Equivariance proved using a new generalized Stone-Weierstrass theorem
- Works on graphs of varying size (don't need to fix the number of nodes)
- Can't derive an bounds for order of tensorization $k_s \leq n^d(n^d - 1)/2$ and $k_s \geq n^d$

Conclusion

- Using higher-order tensorization inside GNN, we can achieve both invariance and equivariance
- Invariance (re)-proved using the Stone-Weierstrass theorem
- Equivariance proved using a new generalized Stone-Weierstrass theorem
- Works on graphs of varying size (don't need to fix the number of nodes)
- Can't derive an bounds for order of tensorization $k_s \leq n^d(n^d - 1)/2$ and $k_s \geq n^d$
- Equivariant proof doesn't exted to higher order *output* $\mathbb{R}^{n^\ell}$ (graph to graph mappings)

Conclusion

- Using higher-order tensorization inside GNN, we can achieve both invariance and equivariance
- Invariance (re)-proved using the Stone-Weierstrass theorem
- Equivariance proved using a new generalized Stone-Weierstrass theorem
- Works on graphs of varying size (don't need to fix the number of nodes)
- Can't derive an bounds for order of tensorization $k_s \leq n^d(n^d - 1)/2$ and $k_s \geq n^d$
- Equivariant proof doesn't exted to higher order *output* $\mathbb{R}^{n^\ell}$ (graph to graph mappings)
- Equivarant proof doesn't extend to subgroups of $\mathcal{O}_n$

References I

[1] Nicolas Keriven and Gabriel Peyré. "Universal Invariant and Equivariant Graph Neural Networks". In: *Advances in Neural Information Processing Systems*. Ed. by H. Wallach et al. Vol. 32. Curran Associates, Inc., 2019. URL: https://proceedings.neurips.cc/paper_files/paper/2019/file/ea9268cb43f55d1d12380fb6ea5bf572-Paper.pdf.

[2] Haggai Maron et al. "On the Universality of Invariant Networks". In: *Proceedings of the 36th International Conference on Machine Learning*. Ed. by Kamalika Chaudhuri and Ruslan Salakhutdinov. Vol. 97. Proceedings of Machine Learning Research. PMLR, Sept. 2019, pp. 4363–4371. URL: https://proceedings.mlr.press/v97/maron19a.html.

[3] Haggai Maron et al. "Invariant and Equivariant Graph Networks". In: *International Conference on Learning Representations*. 2018.