Expressive Power of Invariant and Equivariant Graph Neural Networks

Waïss Azizian & Marc Lelarge, ENS & INRIA & LJK

Presented at ICLR 2021

GNNs, Invariance and Equivariance

Neural Networks on graphs

Goal: take graphs as inputs of neural network models

Dataset: $(G_1, y_1), \ldots, (G_M, y_M)$ with G_i graphs

Tasks:

- Classification/regression: one label per graph $y_i \in \mathbb{R}$

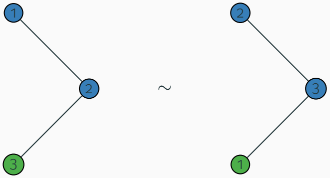
$$G \text{ graph} \xrightarrow{\text{GNN}} y \in \mathbb{R}$$

  Example: protein/molecule clasification.
- Node embedding: one label per nodes of the graph, $y_i \in \mathbb{R}^n$ if G_i has n nodes.

$$G \text{ graph with } n \text{ nodes} \xrightarrow{\text{GNN}} y \in \mathbb{R}^n$$

  Example: community detection.

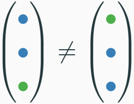
No canonical representation of graphs



Adjacency matrix:

$$\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \neq \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}$$

Node features:



Invariant and equivariant functions

For a permutation $\sigma \in \mathcal{S}_n$, we define,

- for $X \in \mathbb{R}^n$, $(\sigma \cdot X)_i = X_{\sigma^{-1}(i)}$
- for $G \in \mathbb{R}^{n \times n}$, $(\sigma \cdot G)_{i_1, i_2} = G_{\sigma^{-1}(i_1), \sigma^{-1}(i_2)}$

Two graphs with adjacency matrices G_1, G_2 are isomorphic iff $\exists \sigma$, $G_1 = \sigma \cdot G_2$.

Invariant and equivariant functions

For a permutation $\sigma \in \mathcal{S}_n$, we define,

- for $X \in \mathbb{R}^n$, $(\sigma \cdot X)_i = X_{\sigma^{-1}(i)}$
- for $G \in \mathbb{R}^{n \times n}$, $(\sigma \cdot G)_{i_1, i_2} = G_{\sigma^{-1}(i_1), \sigma^{-1}(i_2)}$

Two graphs with adjacency matrices G_1, G_2 are isomorphic iff $\exists \sigma$, $G_1 = \sigma \cdot G_2$.

Definition

($k = 1$ or $k = 2$)

A function $f : \mathbb{R}^{n^k} \to \mathbb{R}$ is invariant if $f(\sigma \cdot G) = f(G)$.

A function $f : \mathbb{R}^{n^k} \to \mathbb{R}^n$ is equivariant if $f(\sigma \cdot G) = \sigma \cdot f(G)$.

Practical GNNs and their limitations

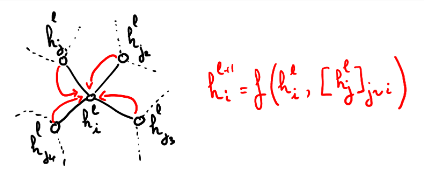
A first example: Message passing GNN (MPGNN)

- MPGNN take as input a discrete graph $G = (V, E)$ with n nodes and node features $h^0 \in \mathbb{R}^n$
- Defined inductively: given h^ℓ node features at layer ℓ, $h^{\ell+1}$ is obtained by:

$$\text{for node } i, \quad h_i^{\ell+1} = f_0\left(h_i^\ell, \sum_{(j,i)\in E} f_1\left(h_j^\ell\right)\right),$$

where f_0 and f_1 are learnable functions.

Prop: The message passing layer is equivariant.



MPGNN cannot separate all non-isormophic graphs

An example of a problematic pair for MPGNN:

For any F MPGNN, $F(G_1) = F(G_2)$.

MPGNN cannot separate all non-isormophic graphs

An example of a problematic pair for MPGNN:

For any F MPGNN, $F(G_1) = F(G_2)$.

Consequence for approximation: if $(F_k)_k$ sequence of MPGNN s.t.

$$F_k \to F\,,$$

then $F(G_1) = F(G_2)$. $\Rightarrow$ MPGNN cannot approximate all invariant/equivariant functions.

Weisfeiler-Lehman test

We can precisely characterize which graphs can be separated by MPGNN.

- **Weisfeiler-Lehman test:** approximate isomorphism test for graphs,

  $$WL(G_1) \neq WL(G_2) \implies G_1 \nsim G_2\ .$$

  But there are $G_1 \nsim G_2$ such that $WL(G_1) = WL(G_2)$.
- **Prop (Xu et al., 2018):** There exists F MPGNN such that $F(G_1) \neq F(G_2)$ iff $WL(G_1) \neq WL(G_2)$.
- **Consequence:** If F approximated by MPGNNs, then $WL(G_1) = WL(G_2)$ implies $F(G_1) = F(G_2)$.

Holy grail of GNN: A class of GNN which is universal, i.e. which can approximate all invariant/equivariant continuous functions.

Existing approximation results for GNN: only universality results

- Relational pooling (Murphy et al., 2019) / Group averaging (Ravanbakhsh, 2020) but has complexity $O(n!)$.
- Tensor based architectures (Maron et al., 2018, 2019a) but tensors of size $O(n^n)$ needed (Maron et al., 2019b; Keriven and Peyré, 2019; Chen et al., 2019).

Universality

Holy grail of GNN: A class of GNN which is universal, i.e. which can approximate all invariant/equivariant continuous functions.

Existing approximation results for GNN: only universality results

- Relational pooling (Murphy et al., 2019) / Group averaging (Ravanbakhsh, 2020) but has complexity $O(n!)$.
- Tensor based architectures (Maron et al., 2018, 2019a) but tensors of size $O(n^n)$ needed (Maron et al., 2019b; Keriven and Peyré, 2019; Chen et al., 2019).

Question: what happens for tractable classes of GNN, both equivariant and invariant, even though their separation power is limited?

Revisiting the Stone-Weierstrass theorem for invariant and equivariant functions

Theorem(Stone-Weierstrass): X compact space, $\mathcal{F} \subset \mathcal{C}(X, \mathbb{R})$ subalgebra, i.e. linear space, stable by pointwise multiplication s.t. $x \mapsto \mathbf{1} \in \mathcal{F}$.

Assume that $\mathcal{F}$ separates points,

$$\forall x, x' \in X, \quad x \neq x' \implies \exists f \in \mathcal{F}, f(x) \neq f(x')\,.$$

Then, for the uniform norm,

$$\overline{\mathcal{F}} = \mathcal{C}(X, \mathbb{R})$$

But our classes of GNN do not separate points...

Separating power

Separation (Timofte, 2005): Let $\mathcal{F}$ be a set of functions defined on X. Define the equivalence relation $\rho(\mathcal{F})$ on X is: for any $x, x' \in X$,

$$(x, x') \in \rho(\mathcal{F}) \iff \forall f \in \mathcal{F},\ f(x) = f(x')\,.$$

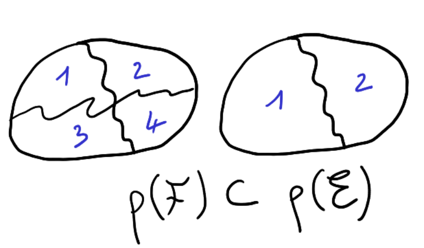
Given two sets of functions $\mathcal{F}$ and $\mathcal{E}$, $\rho(\mathcal{F}) \subset \rho(\mathcal{E})$ means that

Separating power

Separation (Timofte, 2005): Let $\mathcal{F}$ be a set of functions defined on X. Define the equivalence relation $\rho(\mathcal{F})$ on X is: for any $x, x' \in X$,

$$(x, x') \in \rho(\mathcal{F}) \iff \forall f \in \mathcal{F},\ f(x) = f(x')\,.$$

Given two sets of functions $\mathcal{F}$ and $\mathcal{E}$, $\rho(\mathcal{F}) \subset \rho(\mathcal{E})$ means that $\mathcal{F}$ is "more separating" than $\mathcal{E}$.

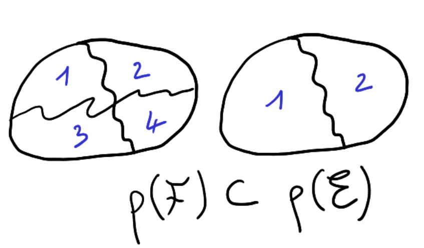


Separating power

Separation (Timofte, 2005): Let $\mathcal{F}$ be a set of functions defined on X. Define the equivalence relation $\rho(\mathcal{F})$ on X is: for any $x, x' \in X$,

$$(x, x') \in \rho(\mathcal{F}) \iff \forall f \in \mathcal{F},\ f(x) = f(x')\,.$$

Given two sets of functions $\mathcal{F}$ and $\mathcal{E}$, $\rho(\mathcal{F}) \subset \rho(\mathcal{E})$ means that $\mathcal{F}$ is "more separating" than $\mathcal{E}$.



Example

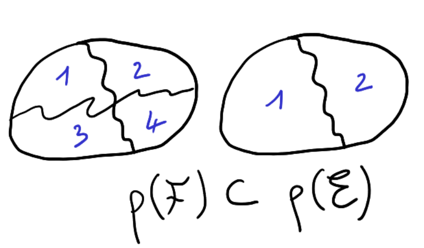
If $\mathcal{F}$ separates points, then $\rho(\mathcal{F}) =$

Separating power

Separation (Timofte, 2005): Let $\mathcal{F}$ be a set of functions defined on X. Define the equivalence relation $\rho(\mathcal{F})$ on X is: for any $x, x' \in X$,

$$(x, x') \in \rho(\mathcal{F}) \iff \forall f \in \mathcal{F},\ f(x) = f(x')\,.$$

Given two sets of functions $\mathcal{F}$ and $\mathcal{E}$, $\rho(\mathcal{F}) \subset \rho(\mathcal{E})$ means that $\mathcal{F}$ is "more separating" than $\mathcal{E}$.



Example

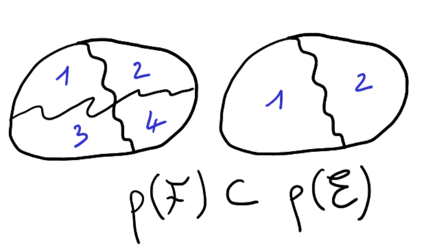
If $\mathcal{F}$ separates points, then $\rho\,(\mathcal{F}) = \{(x, x) : x \in X\}$

Separating power

Separation (Timofte, 2005): Let $\mathcal{F}$ be a set of functions defined on X. Define the equivalence relation $\rho(\mathcal{F})$ on X is: for any $x, x' \in X$,

$$(x, x') \in \rho(\mathcal{F}) \iff \forall f \in \mathcal{F},\ f(x) = f(x')\,.$$

Given two sets of functions $\mathcal{F}$ and $\mathcal{E}$, $\rho(\mathcal{F}) \subset \rho(\mathcal{E})$ means that $\mathcal{F}$ is "more separating" than $\mathcal{E}$.



Proposition (Xu et al. (2018), revisited)

$$\rho(MPGNN) = \rho(WL)$$

Theorem (Timofte (2005), simplified)

X compact space, $\mathcal{F} \subset \mathcal{C}(X, \mathbb{R}^p)$ subalgebra which contains $x \mapsto (1, \ldots, 1)$. Then,

$$\overline{\mathcal{F}} = \left\{ f \in \mathcal{C}(X, \mathbb{R}^p) : \rho\left(\mathcal{F}\right) \subseteq \rho\left(f\right), \ \forall x \in X, \ f(x) \in \overline{\mathcal{F}(x)} \right\} ,$$

where $\mathcal{F}(x) = \{f(x), \, f \in \mathcal{F}\}$,

under the assumption that

Theorem (Timofte (2005), simplified)

X compact space, $\mathcal{F} \subset \mathcal{C}(X, \mathbb{R}^p)$ subalgebra which contains $x \mapsto (1, \ldots, 1)$. Then,

$$\overline{\mathcal{F}} = \left\{ f \in \mathcal{C}(X, \mathbb{R}^p) : \rho(\mathcal{F}) \subseteq \rho(f)\,,\ \forall x \in X,\ f(x) \in \overline{\mathcal{F}(x)} \right\}\ ,$$

where $\mathcal{F}(x) = \{f(x),\ f \in \mathcal{F}\}$,

under the assumption that there exists $\mathcal{S} \subset \mathcal{C}(X, \mathbb{R})$ a class of real-valued functions s.t.

$$\mathcal{S}\mathcal{F} \subset \mathcal{F}$$

which is more separating that $\mathcal{F}$: $\rho(\mathcal{S}) \subset \rho(\mathcal{F})$.

Theorem (Timofte (2005), simplified)

X compact space, $\mathcal{F} \subset \mathcal{C}(X, \mathbb{R}^p)$ subalgebra which contains $x \mapsto (1, \dots, 1)$. Then,

$$\overline{\mathcal{F}} = \left\{ f \in \mathcal{C}(X, \mathbb{R}^p) : \rho(\mathcal{F}) \subseteq \rho(f),\ \forall x \in X,\ f(x) \in \overline{\mathcal{F}(x)} \right\},$$

where $\mathcal{F}(x) = \{f(x),\ f \in \mathcal{F}\}$,

under the assumption that there exists $\mathcal{S} \subset \mathcal{C}(X, \mathbb{R})$ a class of real-valued functions s.t.

$$\mathcal{S}\mathcal{F} \subset \mathcal{F}$$

which is more separating that $\mathcal{F}$: $\rho(\mathcal{S}) \subset \rho(\mathcal{F})$.

As a consequence, for most classes of real-valued invariant GNN,

$$\overline{\text{GNN}} = \{f \in \mathcal{C}(X, \mathbb{R}) :\ \rho(\text{GNN}) \subset \rho(f)\}.$$

Vector-valued functions

Theorem (Timofte (2005), simplified)

X compact space, $\mathcal{F} \subset \mathcal{C}(X, \mathbb{R}^p)$ subalgebra which contains $x \mapsto (1, \ldots, 1)$. Then,

$$\overline{\mathcal{F}} = \left\{ f \in \mathcal{C}(X, \mathbb{R}^p) : \rho(\mathcal{F}) \subseteq \rho(f), \; \forall x \in X, \; f(x) \in \overline{\mathcal{F}(x)} \right\},$$

where $\mathcal{F}(x) = \{f(x), \; f \in \mathcal{F}\}$,

under the assumption that there exists $\mathcal{S} \subset \mathcal{C}(X, \mathbb{R})$ a class of real-valued functions s.t.

$$\mathcal{S}\mathcal{F} \subset \mathcal{F}$$

which is more separating that $\mathcal{F}$: $\rho(\mathcal{S}) \subset \rho(\mathcal{F})$.

As a consequence, for most classes of real-valued invariant GNN,

$$\overline{\text{GNN}} = \{f \in \mathcal{C}(X, \mathbb{R}) : \; \rho(\text{GNN}) \subset \rho(f)\}.$$

In particular, $\overline{\text{MPGNN}} = \{f \in \mathcal{C}(X, \mathbb{F}) : \; \rho(WL) \subset \rho(f)\}$ (Scarselli et al., 2009)

Can it be applied to equivariant functions?

Take $\mathcal{F} \subset \mathcal{C}_E(X, \mathbb{R}^n)$ set of equivariant functions, and $\mathcal{S} \subset \mathcal{C}(X, \mathbb{R})$ a set of scalar functions satisfying the assumptions of the theorem.

- $\mathcal{S}\mathcal{F} \subset \mathcal{F}$ implies that $\{x \mapsto (f(x), \ldots, f(x)) : f \in \mathcal{S}\} \subset \mathcal{F}$.
- Then, since the functions in $\mathcal{F}$ are equivariant, this means that the functions in $\mathcal{S}$ are invariant.
- But $\rho(\mathcal{S}) \subset \rho(\mathcal{F})$ would imply that the functions in $\mathcal{F}$ are invariant too!

Can it be applied to equivariant functions?

Take $\mathcal{F} \subset \mathcal{C}_E(X, \mathbb{R}^n)$ set of equivariant functions, and $\mathcal{S} \subset \mathcal{C}(X, \mathbb{R})$ a set of scalar functions satisfying the assumptions of the theorem.

- $\mathcal{S}\mathcal{F} \subset \mathcal{F}$ implies that $\{x \mapsto (f(x), \ldots, f(x)) : f \in \mathcal{S}\} \subset \mathcal{F}$.
- Then, since the functions in $\mathcal{F}$ are equivariant, this means that the functions in $\mathcal{S}$ are invariant.
- But $\rho(\mathcal{S}) \subset \rho(\mathcal{F})$ would imply that the functions in $\mathcal{F}$ are invariant too!

$\Rightarrow$ The previous theorem cannot be applied to classes of equivariant functions!

Can it be applied to equivariant functions?

Take $\mathcal{F} \subset \mathcal{C}_E(X, \mathbb{R}^n)$ set of equivariant functions, and $\mathcal{S} \subset \mathcal{C}(X, \mathbb{R})$ a set of scalar functions satisfying the assumptions of the theorem.

- $\mathcal{S}\mathcal{F} \subset \mathcal{F}$ implies that $\{x \mapsto (f(x), \ldots, f(x)) : f \in \mathcal{S}\} \subset \mathcal{F}$.
- Then, since the functions in $\mathcal{F}$ are equivariant, this means that the functions in $\mathcal{S}$ are invariant.
- But $\rho(\mathcal{S}) \subset \rho(\mathcal{F})$ would imply that the functions in $\mathcal{F}$ are invariant too!

$\Rightarrow$ The previous theorem cannot be applied to classes of equivariant functions!

Solution: Using equivariance, relax the original assumption $\rho(\mathcal{S}) \subset \rho(\mathcal{F})$, i.e.,

$$\forall x, x', \; (\exists f \in \mathcal{F}, f(x) \neq f(x')) \implies \exists g \in \mathcal{S}, \, g(x) \neq g(x'),$$

to $\rho(\mathcal{S}) \subset \rho(\pi \circ \mathcal{F})$, i.e.,

$$\forall x, x', \; (\exists f \in \mathcal{F}, Orb(f(x)) \neq Orb(f(x'))) \implies \exists g \in \mathcal{S}, \, g(x) \neq g(x'),$$

where $\pi : \mathbb{R}^n \to \mathbb{R}^n / \mathcal{S}_n$ projection, and $Orb(y) = \{\sigma \cdot y : \sigma \in \mathcal{S}_n\}$.

Equivariant approximation theorem

Theorem (A. & Lelarge, 2021)

X compact space, $\mathcal{F} \subset \mathcal{C}_E(X, \mathbb{R}^p)$ subalgebra of equivariant functions which contains $x \mapsto (1, \dots, 1)$. Then,

$$\overline{\mathcal{F}} = \left\{ f \in \mathcal{C}_E(X, \mathbb{R}^p) : \rho(\mathcal{F}) \subseteq \rho(f),\ \forall x \in X,\ f(x) \in \overline{\mathcal{F}(x)} \right\},$$

where $\mathcal{F}(x) = \{f(x),\ f \in \mathcal{F}\}$,

under the assumption that there exists $\mathcal{S} \subset \mathcal{C}(X, \mathbb{R})$ a class of real-valued functions s.t.

$$\mathcal{S}\mathcal{F} \subset \mathcal{F}$$

which is more separating that $\mathcal{F}$: $\rho(\mathcal{S}) \subset \rho(\pi \circ \mathcal{F})$.

Equivariant approximation theorem

Theorem (A. & Lelarge, 2021)

X compact space, $\mathcal{F} \subset \mathcal{C}_E(X, \mathbb{R}^p)$ subalgebra of equivariant functions which contains $x \mapsto (1, \dots, 1)$. Then,

$$\overline{\mathcal{F}} = \left\{ f \in \mathcal{C}_E(X, \mathbb{R}^p) : \rho\left(\mathcal{F}\right) \subseteq \rho\left(f\right),\ \forall x \in X,\ f(x) \in \overline{\mathcal{F}(x)} \right\},$$

where $\mathcal{F}(x) = \{f(x), f \in \mathcal{F}\}$,

under the assumption that there exists $\mathcal{S} \subset \mathcal{C}(X, \mathbb{R})$ a class of real-valued functions s.t.

$$\mathcal{S}\mathcal{F} \subset \mathcal{F}$$

which is more separating that $\mathcal{F}$: $\rho\left(\mathcal{S}\right) \subset \rho\left(\pi \circ \mathcal{F}\right)$.

Corollary (A. & Lelarge, 2021)

Let X be a compact space, Let $\mathcal{F}_0 \subseteq \bigcup_{h=1}^{\infty} \mathcal{C}_E\left(X, (\mathbb{R}^h)^n\right)$ be a set of equivariant functions, stable by concatenation, and consider,

$$\mathcal{F} = \{x \mapsto (m(f(x)_1), \dots, m(f(x)_n)) : f \in \mathcal{F}_0 \cap \mathcal{C}\left(X, (\mathbb{R}^h)^n\right), \, m : \mathbb{R}^h \to \mathbb{R} \text{ MLP}, \, h \geq 1\}$$

Then the closure of $\mathcal{F}$ is,

$$\overline{\mathcal{F}} = \{f \in \mathcal{C}_E(X, \mathbb{R}^n) : \rho(\mathcal{F}_0) \subseteq \rho(f)\} \, .$$

under the assumption that,

$$f \in \mathcal{F}_0 \implies x \mapsto \left(\sum_{i=1}^n f(x)_i, \sum_{i=1}^n f(x)_i, \dots, \sum_{i=1}^n f(x)_i\right) \in \mathcal{F}_0 \, .$$

As a consequence, for most classes of invariant and equivariant GNN,

$$\overline{\mathrm{GNN}} \quad = \quad \{f :\ \rho(\mathrm{GNN}) \subset \rho(f)\}\,.$$

In particular, we apply our results to tensor-based architectures from Maron et al. (2018, 2019a).

Applications to GNNs

As a consequence, for most classes of invariant and equivariant GNN,

$$\overline{\mathrm{GNN}} = \{f : \rho(\mathrm{GNN}) \subset \rho(f)\}.$$

In particular, we apply our results to tensor-based architectures from Maron et al. (2018, 2019a).

Bonus: combining these approximation results with graph theory,

- The 2-tensor architecture of Maron et al. (2019a) can approximate functions of the spectrum.
- Tensor architectures of order $O(\sqrt{n})$ are universal on planar graphs.

Conclusion

Summary:

- A general framework to characterize the approximation capabilities of practical classes of GNNs.
- See our paper for the details of the application to tensor architectures.

Perspectives:

- Extension to approximate non-continuous functions (for example, angles between eigenvectors)
- Control the number of layers and the dimensions of the intermediate features.

Thank You!

Bibliography (1)

Z. Chen, S. Villar, L. Chen, and J. Bruna. On the equivalence between graph isomorphism testing and function approximation with gnns. In H. M. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. B. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, 8-14 December 2019, Vancouver, BC, Canada*, pages 15868–15876, 2019.

N. Keriven and G. Peyré. Universal invariant and equivariant graph neural networks. In H. M. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. B. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, 8-14 December 2019, Vancouver, BC, Canada*, pages 7090–7099, 2019.

H. Maron, H. Ben-Hamu, N. Shamir, and Y. Lipman. Invariant and equivariant graph networks. *arXiv preprint arXiv:1812.09902*, 2018.

H. Maron, H. Ben-Hamu, H. Serviansky, and Y. Lipman. Provably powerful graph networks. In *Advances in Neural Information Processing Systems*, pages 2153–2164, 2019a.

Bibliography (2)

H. Maron, E. Fetaya, N. Segol, and Y. Lipman. On the universality of invariant networks. *arXiv preprint arXiv:1901.09342*, 2019b.

R. Murphy, B. Srinivasan, V. Rao, and B. Ribeiro. Relational pooling for graph representations. In *International Conference on Machine Learning*, pages 4663–4673. PMLR, 2019.

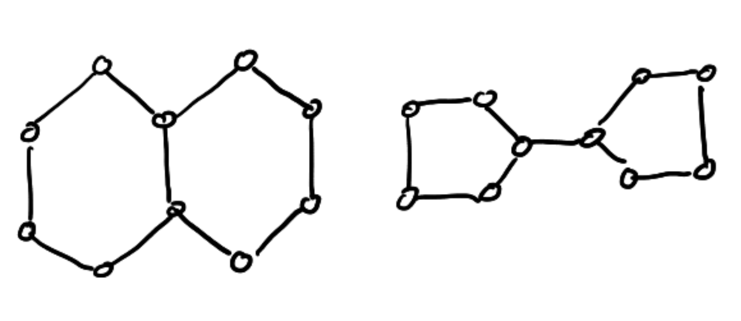
S. Ravanbakhsh. Universal equivariant multilayer perceptrons. *arXiv preprint arXiv:2002.02912*, 2020.

F. Scarselli, M. Gori, A. C. Tsoi, M. Hagenbuchner, and G. Monfardini. Computational capabilities of graph neural networks. *IEEE Trans. Neural Networks*, 20(1):81–102, 2009. doi: 10.1109/TNN.2008.2005141.

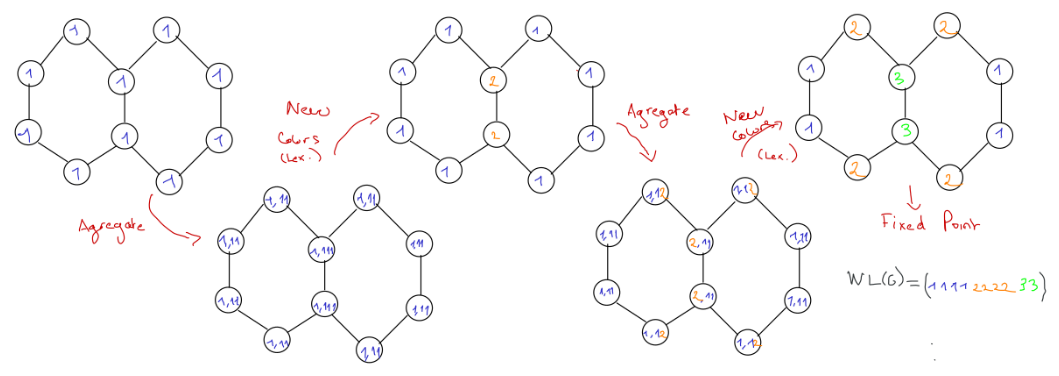
V. Timofte. Stone–weierstrass theorems revisited. *Journal of Approximation Theory*, 136(1):45 – 59, 2005. ISSN 0021-9045. doi: https://doi.org/10.1016/j.jat.2005.05.004.

K. Xu, W. Hu, J. Leskovec, and S. Jegelka. How powerful are graph neural networks? *arXiv preprint arXiv:1810.00826*, 2018.

Another problematic pair



WL test



New
Colors
(Lex.)
Agregate
Agregate
New
Colors
(Lex.)
Fixed Point
WL(G) = (1 1 1 1 2 2 2 2 3 3)

Application to GNNs

We show that:

$$\overline{\text{GNN}} \quad = \quad \{f : \ \rho(\text{GNN}) \subset \rho(f)\} \,.$$

In particular, we obtain the expressive power of Linear GNN (k-LGNN) and Folklore GNN (k-FGNN) with tensors of order k:

- $\rho((k+1)\text{-WL}) \subsetneq \rho(k\text{-WL})$
- $\rho(2\text{-LGNN}) = \rho(2\text{-WL})$
- $\rho(k\text{-LGNN}) \subset \rho(k\text{-WL})$

$$\begin{aligned} \overline{k\text{-LGNN}} \quad &= \quad \{f \in \mathcal{C}(X, \mathbb{F}) : \ \rho(k\text{-WL}) \subset \rho(f)\} \\ \overline{k\text{-FGNN}} \quad &= \quad \{f \in \mathcal{C}(X, \mathbb{F}) : \ \rho((k+1)\text{-WL}) \subset \rho(f)\} \end{aligned}$$

Better expressive power with FGNN

(Maron et al., 2019a) adapted the Folklore version of the Weisfeiler-Lehman test to propose the folklore graph layer (FGL):

$$h_{i\to j}^{\ell+1} = f_0\left(h_{i\to j}^{\ell}, \sum_{k\in V} f_1\left(h_{i\to k}^{\ell}\right) f_2\left(h_{k\to j}^{\ell}\right)\right),$$

where f_0, f_1 and f_2 are learnable functions.

For FGNNs, messages are associated with pairs of vertices as opposed to MPGNN where messages are associated with vertices.

FGNN: a FGNN is the composition of FGLs and a final invariant/equivariant reduction layer from $\mathbb{F}^{n^2}$ to $\mathbb{F}$ or $\mathbb{F}^n$.