Expressive Power of Invariant and Equivariant Graph Neural Networks

Waïss Azizian & Marc Lelarge
ENS & INRIA
waiss.azizian@ens.fr & marc.lelarge@ens.fr

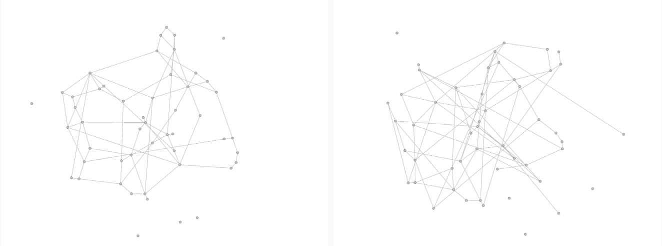
ICLR 2021

Learning with graph symmetries

An example : alignment of graphs

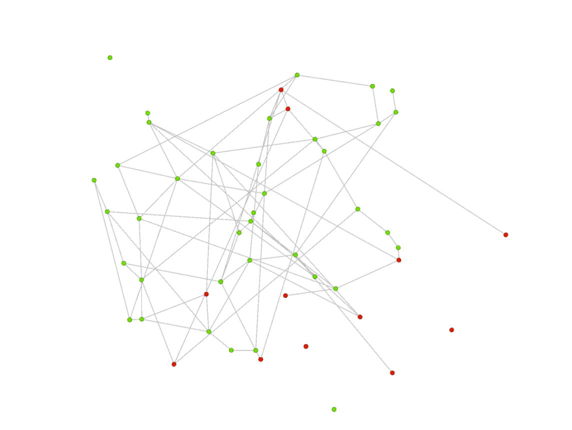
From graph 1 (on the left), put indices on its vertices, perturb the graph by adding and removing a few edges and remove indices to obtain graph 2 (on the right).

Task : recover the indices on vertices of graph 2.



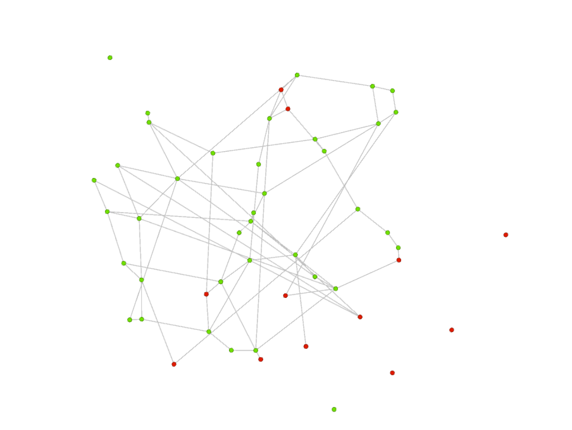
Result with FGNN

Green vertices are good predictions. Red vertices are errors (graph 2).



Result with FGNN

Green vertices are good predictions. Red vertices are errors (graph 1).



Invariant and Equivariant GNNs

Invariant and equivariant functions

For a permutation $\sigma \in \mathcal{S}_n$, we define ($\mathbb{F} = \mathbb{R}^p$ feature space) :

- for $X \in \mathbb{F}^n$, $(\sigma \star X)_{\sigma(i)} = X_i$
- for $G \in \mathbb{F}^{n \times n}$, $(\sigma \star G)_{\sigma(i_1),\sigma(i_2)} = G_{i_1,i_2}$

Invariant and equivariant functions

For a permutation $\sigma \in \mathcal{S}_n$, we define ($\mathbb{F} = \mathbb{R}^p$ feature space) :

- for $X \in \mathbb{F}^n$, $(\sigma \star X)_{\sigma(i)} = X_i$
- for $G \in \mathbb{F}^{n \times n}$, $(\sigma \star G)_{\sigma(i_1),\sigma(i_2)} = G_{i_1,i_2}$

G_1, G_2 are isomorphic iff $G_1 = \sigma \star G_2$.

Invariant and equivariant functions

For a permutation $\sigma \in \mathcal{S}_n$, we define ($\mathbb{F} = \mathbb{R}^p$ feature space) :

- for $X \in \mathbb{F}^n$, $(\sigma \star X)_{\sigma(i)} = X_i$
- for $G \in \mathbb{F}^{n \times n}$, $(\sigma \star G)_{\sigma(i_1),\sigma(i_2)} = G_{i_1,i_2}$

G_1, G_2 are isomorphic iff $G_1 = \sigma \star G_2$.

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

A function $f : \mathbb{F}^{n^k} \to \mathbb{F}$ is said to be invariant if $f(\sigma \star G) = f(G)$.

A function $f : \mathbb{F}^{n^k} \to \mathbb{F}^n$ is said to be equivariant if $f(\sigma \star G) = \sigma \star f(G)$.

Invariant and equivariant functions

For a permutation $\sigma \in \mathcal{S}_n$, we define ($\mathbb{F} = \mathbb{R}^p$ feature space) :

- for $X \in \mathbb{F}^n$, $(\sigma \star X)_{\sigma(i)} = X_i$
- for $G \in \mathbb{F}^{n \times n}$, $(\sigma \star G)_{\sigma(i_1),\sigma(i_2)} = G_{i_1,i_2}$

G_1, G_2 are isomorphic iff $G_1 = \sigma \star G_2$.

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

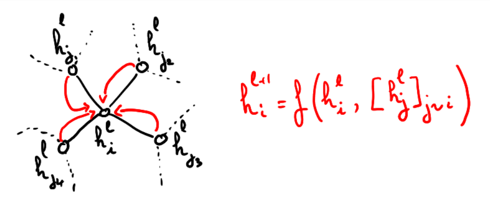
A function $f : \mathbb{F}^{n^k} \to \mathbb{F}$ is said to be invariant if $f(\sigma \star G) = f(G)$.

A function $f : \mathbb{F}^{n^k} \to \mathbb{F}^n$ is said to be equivariant if $f(\sigma \star G) = \sigma \star f(G)$.

For the graph alignment problem, we used an equivariant GNN from $\{0, 1\}^{n \times n}$ to $\mathbb{F}^n$.

Practical GNNs are not universal

A first example : Message passing GNN (MGNN)

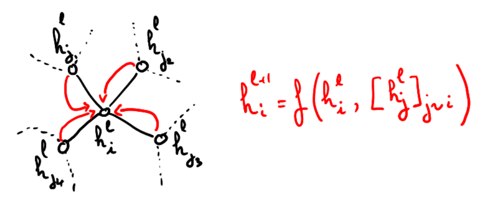


MGNN takes as input a discrete graph $G = (V, E)$ with n nodes and are defined inductively as : $h_i^\ell \in \mathbb{F}$ being the features at layer ℓ associated with node i, then

$$h_i^{\ell+1} = f\left(h_i^\ell, \left\{\left\{h_j^\ell\right\}\right\}_{j\sim i}\right) = f_0\left(h_i^\ell, \sum_{j\sim i} f_1\left(h_i^\ell, h_j^\ell\right)\right),$$

where f or f_0 and f_1 are learnable functions.

A first example : Message passing GNN (MGNN)



MGNN takes as input a discrete graph $G = (V, E)$ with n nodes and are defined inductively as : $h_i^\ell \in \mathbb{F}$ being the features at layer ℓ associated with node i, then

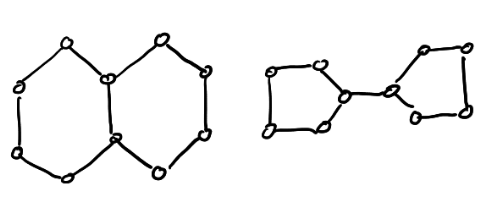
$$h_i^{\ell+1} = f\left(h_i^\ell, \left\{\left\{h_j^\ell\right\}\right\}_{j\sim i}\right) = f_0\left(h_i^\ell, \sum_{j\sim i} f_1\left(h_i^\ell, h_j^\ell\right)\right),$$

where f or f_0 and f_1 are learnable functions.

Prop : The message passing layer is equivariant and both expressions above are equivalent (i.e. for each f, there exists f_0 and f_1).

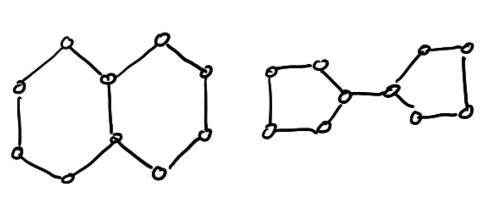
MGNN are not universal

An example of a problematic pair for MGNN :



MGNN are not universal

An example of a problematic pair for MGNN :



Another example :

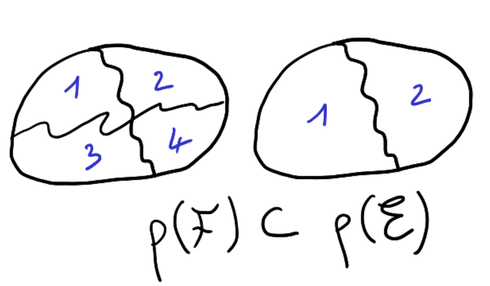
Prop : MGNN are useless on d-regular graphs (without features).

Separating power of MGNN

Separation : Let $\mathcal{F}$ be a set of functions f defined on a set X. The equivalence relation $\rho(\mathcal{F})$ defined by $\mathcal{F}$ 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 $\mathcal{E}$, we say that $\mathcal{F}$ is more separating than $\mathcal{E}$ if $\rho(\mathcal{F}) \subset \rho(\mathcal{E})$.

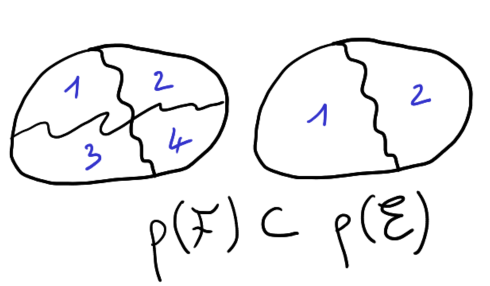


Separating power of MGNN

Separation : Let $\mathcal{F}$ be a set of functions f defined on a set X. The equivalence relation $\rho(\mathcal{F})$ defined by $\mathcal{F}$ 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 $\mathcal{E}$, we say that $\mathcal{F}$ is more separating than $\mathcal{E}$ if $\rho(\mathcal{F}) \subset \rho(\mathcal{E})$.



Xu et al. (2019) Prop : $\rho(\mathrm{MGNN}) = \rho(2\text{-WL})$

Our contribution :
from Separation to Approximation

An easy general fact :

If there exists $x \neq x'$ with $(x, x') \in \rho(\mathcal{F})$, all functions in $\mathcal{F}$ take the same values at x and x' and $\mathcal{F}$ cannot be dense.

Approximation $\Rightarrow$ **Separation**

An easy general fact :

If there exists $x \neq x'$ with $(x, x') \in \rho(\mathcal{F})$, all functions in $\mathcal{F}$ take the same values at x and x' and $\mathcal{F}$ cannot be dense.

Approximation $\Rightarrow$ **Separation**

If $\mathcal{F}$ is an algebra containing the constant function $\mathbf{1}$, i.e. vector space closed under pointwise multiplication then : **Separation** $\Leftrightarrow$ **Approximation**.

Stone-Weierstrass theorem

An easy general fact :

If there exists $x \neq x'$ with $(x, x') \in \rho(\mathcal{F})$, all functions in $\mathcal{F}$ take the same values at x and x' and $\mathcal{F}$ cannot be dense.

Approximation $\Rightarrow$ Separation

If $\mathcal{F}$ is an algebra containing the constant function $\mathbf{1}$, i.e. vector space closed under pointwise multiplication then : **Separation $\Leftrightarrow$ Approximation**.

Pb : we know MGNNs do not separate all graphs !

Stone-Weierstrass theorem

An easy general fact :

If there exists $x \neq x'$ with $(x, x') \in \rho(\mathcal{F})$, all functions in $\mathcal{F}$ take the same values at x and x' and $\mathcal{F}$ cannot be dense.

Approximation $\Rightarrow$ **Separation**

If $\mathcal{F}$ is an algebra containing the constant function $\mathbf{1}$, i.e. vector space closed under pointwise multiplication then : **Separation** $\Leftrightarrow$ **Approximation**.

Pb : we know MGNNs do not separate all graphs !

Sol : we need to relax the separation assumption... and consider vector-valued functions

Stone-Weierstrass theorem for vector-valued functions with symmetries

Building on the work of Timofte (2005), and we proved :

Theorem
Let $\mathcal{F} \subset \mathcal{C}_I(X, \mathbb{R}^p)$ be a sub-algebra of continuous invariant functions, (...).

If the set of functions $\mathcal{F}_{scal} \subset \mathcal{C}(X, \mathbb{R})$ defined by,

$$\mathcal{F}_{scal} = \{f \in \mathcal{C}(X, \mathbb{R}) : f\mathbf{1} \in \mathcal{F}\}$$

is more separating than $\mathcal{F}$, i.e. satisfies,

$$\rho(\mathcal{F}_{scal}) \subset \rho(\mathcal{F}) \, .$$

Then any function less separating than $\mathcal{F}$ can be approximated, i.e.

$$\overline{\mathcal{F}} = \left\{f \in \mathcal{C}_I(X, \mathbb{R}^p) : \rho(\mathcal{F}) \subset \rho(f)\right\} \, .$$

Stone-Weierstrass theorem for vector-valued functions with symmetries

Building on the work of Timofte (2005), and we proved :

Theorem
Let $\mathcal{F} \subset \mathcal{C}_I(X, \mathbb{R}^p)$ be a sub-algebra of continuous invariant functions, (...).

If the set of functions $\mathcal{F}_{scal} \subset \mathcal{C}(X, \mathbb{R})$ defined by,

$$\mathcal{F}_{scal} = \{f \in \mathcal{C}(X, \mathbb{R}) : f\mathbf{1} \in \mathcal{F}\}$$

is more separating than $\mathcal{F}$, i.e. satisfies,

$$\rho(\mathcal{F}_{scal}) \subset \rho(\mathcal{F}) .$$

Then any function less separating than $\mathcal{F}$ can be approximated, i.e.

$$\overline{\mathcal{F}} = \{f \in \mathcal{C}_I(X, \mathbb{R}^p) : \rho(\mathcal{F}) \subset \rho(f)\} .$$

See our paper for the equivariant version.

Application to GNNs

For all GNNs studied, the technical condition on $\mathcal{F}_{scal}$ is satisfied!

As a consequence, we show that :

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

Application to GNNs

For all GNNs studied, the technical condition on $\mathcal{F}_{scal}$ is satisfied !

As a consequence, we show that :

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

Recall : $\rho(\text{MGNN}) = \rho(2\text{-WL})$

so that : $\overline{\text{MGNN}} = \{f \in \mathcal{C}(X, \mathbb{F}) : \ \rho(2\text{-WL}) \subset \rho(f)\}$

Application to GNNs

For all GNNs studied, the technical condition on $\mathcal{F}_{scal}$ is satisfied!

As a consequence, we show that :

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

Recall : $\rho(\text{MGNN}) = \rho(2\text{-WL})$

so that : $\overline{\text{MGNN}} = \{f \in \mathcal{C}(X, \mathbb{F}) : \rho(2\text{-WL}) \subset \rho(f)\}$

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

$$\overline{k\text{-LGNN}} = \{f \in \mathcal{C}(X, \mathbb{F}) : \rho(k\text{-WL}) \subset \rho(f)\}$$

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

Learning with (practical i.e. $k = 2$) FGNN

Better expressive power with FGNN

(Maron et al., 2019) 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 MGNN 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 $\mathbb{F}/\mathbb{F}^n$.

Properties of Folklore GNN (FGNN)

(Maron et al., 2019) Prop : FGL is equivariant and $\rho(\text{FGNN}) = \rho(\text{3-WL})$.

Properties of Folklore GNN (FGNN)

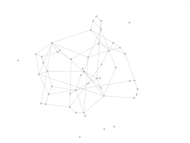
(Maron et al., 2019) Prop : FGL is equivariant and $\rho(\text{FGNN}) = \rho(\text{3-WL})$.

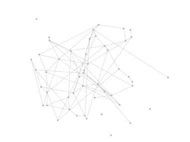
Approximation for FGNN :

$$\overline{\text{FGNN}} = \{f \in \mathcal{C}(X, \mathbb{F}) : \ \rho(\text{3-WL}) \subset \rho(f)\}$$

FGNN has the best power of approximation among all architectures working with tensors of order **2** presented so far.

Learning the graph alignment problem with Siamese FGNNs





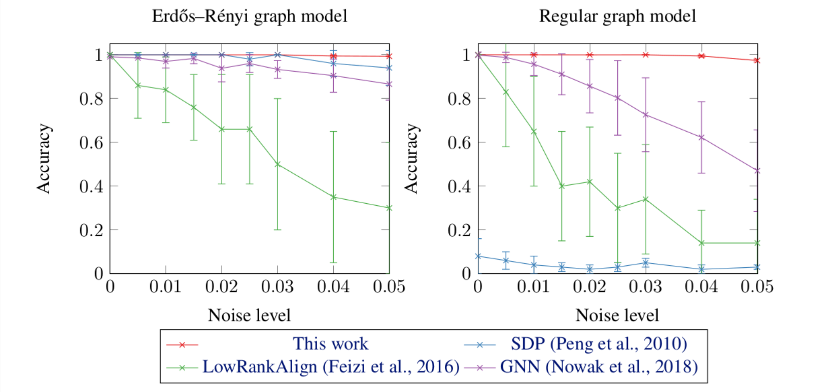
$$G_1 \in \{0,1\}^{n^2} \xrightarrow{\text{FGNN}} E_1 \in \mathbb{R}^{n\times b}$$

$$E_1 E_2^T \in \mathbb{R}^{n^2}$$

$$G_2 \in \{0,1\}^{n^2} \xrightarrow{\text{FGNN}} E_2 \in \mathbb{R}^{n\times b}$$

From the node similarity matrix $E_1E_2^T$, we extract a mapping from nodes of G_1 to nodes of G_2.

Results on synthetic data



- Graphs : $n = 50$, density = 0.2
- Training set : 20000 samples
- Validation and Test sets : 1000 samples

Conclusion

- For various GNNs, we characterized their separating power in term of the k-WL test in the invariant and equivariant cases.
- For GNNs : **Power of Separation** $\Leftrightarrow$ **Power of Approximation**.
- FGNN has the best power of approximation among all GNNs dealing with tensors of order 2.
- FGNN shows the best empirical results in the equivariant setting of the graph alignment problem : `https://github.com/mlelarge/graph_neural_net`

Thank You!

Références

H. Maron, H. Ben-Hamu, H. Serviansky, and Y. Lipman. Provably powerful graph 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 2153–2164, 2019. URL `http://papers.nips.cc/paper/8488-provably-powerful-graph-networks`.

A. Nowak, S. Villar, A. S. Bandeira, and J. Bruna. Revised note on learning quadratic assignment with graph neural networks. In *2018 IEEE Data Science Workshop, DSW 2018, Lausanne, Switzerland, June 4-6, 2018*, pages 229–233. IEEE, 2018. doi : 10.1109/DSW.2018.8439919. URL `https://doi.org/10.1109/DSW.2018.8439919`.

J. Peng, H. D. Mittelmann, and X. Li. A new relaxation framework for quadratic assignment problems based on matrix splitting. *Math. Program. Comput.*, 2 (1) :59–77, 2010. doi : 10.1007/s12532-010-0012-6. URL `https://doi.org/10.1007/s12532-010-0012-6`.

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. URL `http:`