GNNs based on **local features** capture a logic heavily used in **Knowledge Representation**.

GNNs mixing **local + global** features can express First Order logic **with 2 variables and counting**.

Logical Expressiveness of Graph Neural Networks

Aggregate-Combine GNNs (AC-GNNs)

$$\boldsymbol{x}_v^{(i)} = \mathrm{COM}^{(i)}\left(\boldsymbol{x}_v^{(i-1)}, \mathrm{AGG}^{(i)}\left(\{\!\{\boldsymbol{x}_u^{(i-1)} \mid u \in \mathcal{N}_G(v)\}\!\}\right)\right)$$

- $\boldsymbol{x}_v^{(i)}$: features of node *v* at layer *i*
- $\boldsymbol{x}_v^{(i-1)}$: features of node *v* at layer *i - 1*
- $\{\!\{\boldsymbol{x}_u^{(i-1)} \mid u \in \mathcal{N}_G(v)\}\!\}$: multiset of features of *v* neighbors
- $\mathrm{COM}^{(i)}$: combine node and neighbor features
- $\mathrm{AGG}^{(i)}$: aggregate *v*-neighbors (local) features

First Order logic with counting (*FOC*)

Example: A node ***v*** satisfies the *FOC* formula

$$\mathrm{Red}(x) \ \wedge \ \exists y\left(\neg E(x,y) \ \wedge \ \exists^{\geq 2} x\left[\, E(y,x) \wedge \mathrm{Blue}(x)\,\right]\right)$$

If and only if:

- ***v*** is red, and
- there is a node
- not connected with ***v***
- that has at least two
- blue neighbors

FOC-k: *FOC* with only ***k*** variables in every formula.

AC-GNNs cannot always implement *FOC-2* formulas

Proposition

There are (infinite) ***FOC-2*** formulas that cannot be captured by ***AC-GNNs***

ALCQ: a Description Logic to define *concepts*

Example: The following is an *ALCQ* formula

$$\mathrm{Man} \ \sqcap \ \mathrm{married.Doctor} \ \sqcap \ \exists^{\geq 3}\mathrm{child.Happy}$$

defining “men married with a doctor that have at least three happy children”.

Theorem 1

A formula is captured by an ***AC-GNN*** **if and only if** it is expressible in ***ALCQ***

ACR-GNNs: adding global features (Readouts)

$$\boldsymbol{x}_v^{(i)} = \mathrm{COM}^{(i)}\Big(\boldsymbol{x}_v^{(i-1)}, \mathrm{AGG}^{(i)}\left(\{\!\{\boldsymbol{x}_u^{(i-1)} \mid u \in \mathcal{N}_G(v)\}\!\}\right), \mathrm{READ}^{(i)}\left(\{\!\{\boldsymbol{x}_u^{(i-1)} \mid u \in G\}\!\}\right)\Big)$$

$\mathrm{READ}^{(i)}$: considers the features of all nodes in the graph

Theorem 2

Every formula expressible in ***FOC2*** can be captured by an ***ACR-GNN***

Results on synthetic data and *FOC-2* formulas

	α_1 Train	α_1 Test		α_2 Train	α_2 Test	
		same-size	bigger		same-size	bigger
AC 10-layers	0.839	0.826	0.671	0.694	0.695	0.667
GIN 10-layers	0.567	0.566	0.536	0.689	0.693	0.672
ACR 1-layer	1.000	1.000	1.000	0.827	0.834	0.726
ACR 2-layers	1.000	1.000	1.000	0.895	0.897	0.770
ACR 3-layers	1.000	1.000	1.000	0.903	0.902	0.836

PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE

CIENCIAS DE LA COMPUTACIÓN UNIVERSIDAD DE CHILE

Instituto Milenio Fundamentos de los datos

UNIVERSITY OF OXFORD

Pablo Barceló
Mikaël Monet
Juan L. Reutter
PUC & IMFD

Egor V. Kostylev
Oxford

Jorge Pérez
Juan-Pablo Silva
UChile & IMFD

Contact: <jperez@dcc.uchile.cl>