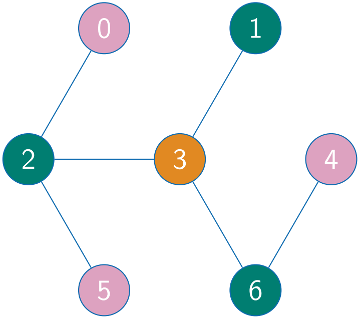
Logical Distillation of Graph Neural Networks

Alexander Pluska, Pascal Welke, Thomas Gärtner, Sagar Malhotra

November 7th, 2024

TECHNISCHE
UNIVERSITÄT
WIEN



A GNN comprises for each layer $k \in [l]$

- an aggregation function $\mathrm{agg}_k : \text{Finite Multiset}\,(\mathbb{R}^n) \to \mathbb{R}^n$
- a combination function $\mathrm{comb}_k : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$

Example 1

$$x_{3}^{(1)} = \mathrm{comb}_0\left(x_{3}^{(0)}, \mathrm{aggr}_0\left(\left\{\left\{x_{1}^{(0)}, x_{2}^{(0)}, x_{6}^{(0)}\right\}\right\}\right)\right)$$

Motivation: GNNs and Logic

THE LOGICAL EXPRESSIVENESS OF GRAPH NEURAL NETWORKS

Pablo Barceló
IMC, PUC & IMFD Chile

Egor V. Kostylev
University of Oxford

Mikaël Monet
IMFD Chile

Jorge Pérez
DCC, UChile & IMFD Chile

Juan Reutter
DCC, PUC & IMFD Chile

Juan-Pablo Silva
DCC, UChile

The Logic of Graph Neural Networks

Martin Grohe
RWTH Aachen University, Germany

Theorem 2 (Barceló et al. (2020), Theorem 4.2)

Any first-order classifier can be computed by a GNN without readout if and only if it is expressible in the guarded fragment of C^2.

Definition 3

A *modal parameter* S is one of the following

$$1, I, A, 1-I, 1-A, I+A, 1-I-A.$$

An $\mathcal{EMLC}$ formula is then built by the grammar

$$\varphi ::= U \mid \top \mid \varphi \wedge \varphi \mid \varphi \vee \varphi \mid \neg\varphi \mid S\varphi > n$$

- U ranges over node attributes,
- S ranges over modal parameters,
- n ranges over $\mathbb{N}$.

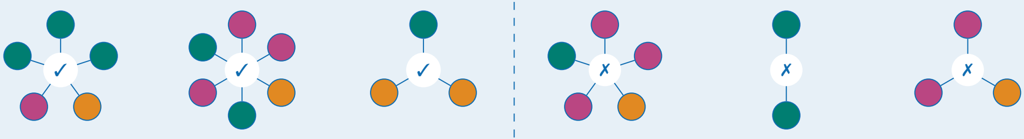
EMLC: Examples

Example 4

Each $\mathcal{EMLC}$ formula expresses a node property.

$$A(\text{green} \vee \text{orange}) > 2$$

expresses that a node has more than 2 neighbors that are green or orange.



EMLC: Examples

Example 5

- Modal parameters can also be nested.

$$A(A\top = 10) > 5$$

  expresses that a node has more than 5 neighbors of degree 10.

- Utilizing the model parameter 1, it is possible to express global properties.

$$1 \bullet > 3$$

  classifies graphs with more than 3 green nodes.

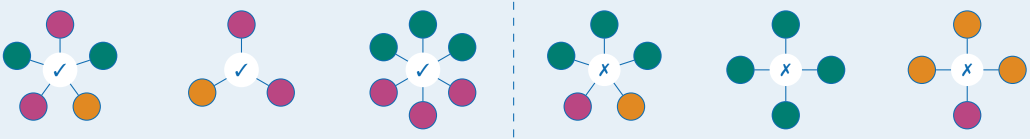
- There are first-order node properties not expressible by C^2 and therefore also EMLC, such as being part of a triangle.

EMLC%

Example 6

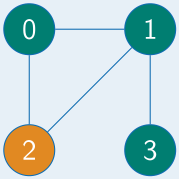
$$A \bullet > 1/3$$

expresses that among a node's neighbors, more than a third are red.



Example 7

G

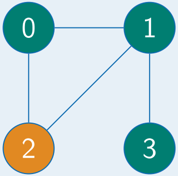


$$A = \begin{pmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} \qquad \textcolor{teal}{\bullet} = \begin{pmatrix} 1 \\ 1 \\ 0 \\ 1 \end{pmatrix} \qquad \textcolor{orange}{\bullet} = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}$$

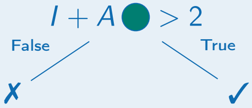
$$\begin{aligned} A\textcolor{teal}{\bullet} > 1 \wedge A\textcolor{orange}{\bullet} > 0 &= \begin{pmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 0 \\ 1 \end{pmatrix} > 1 \wedge \begin{pmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} > 0 \\ &= \begin{pmatrix} 1 \\ 2 \\ 2 \\ 1 \end{pmatrix} > 1 \wedge \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \end{pmatrix} > 0 = \begin{pmatrix} 0 \\ 1 \\ 1 \\ 0 \end{pmatrix} \wedge \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} \end{aligned}$$

Example 8

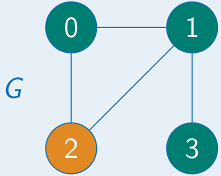
G



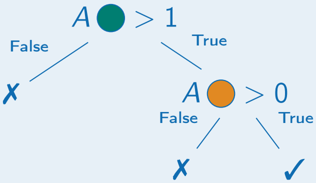
	$\color{teal}\bullet$	$\color{orange}\bullet$	$A\color{teal}\bullet$	$A\color{orange}\bullet$	$\cdots$	$I + A\color{teal}\bullet$	$\cdots$	y
v_0	1	0	1	1		2		0
v_1	1	0	2	1		3		1
v_2	0	1	2	0		2		0
v_3	1	0	1	0		2		0



Example 9 (Assuming only modal parameter A)

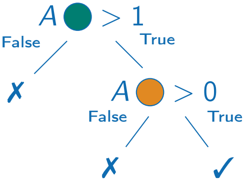


	●	●	A ●	A ●	y
v_0	1	0	1	1	0
v_1	1	0	2	1	1
v_2	0	1	2	0	0
v_3	1	0	1	0	0



Learning deeper EMLC formulas: IDTs

	🟢	🟠	A 🟢	A 🟠
v_0	1	0	1	1
v_1	1	0	2	1
v_2	0	1	2	0
v_3	1	0	1	0



Potential new features:

$U_0 := \neg A$ 🟢 > 1

$U_1 := A$ 🟢 $> 1 \wedge \neg A$ 🟠 > 1

$U_2 := A$ 🟢 $> 1 \wedge A$ 🟠 > 1

$U_3 := U_0 \vee U_1, \ldots$

	🟢	🟠	U_2	U_3	A 🟢	A 🟠	AU_2	AU_3
v_0	1	0	0	1	1	1	1	1
v_1	1	0	1	0	2	1	0	3
v_2	0	1	0	1	2	0	1	1
v_3	1	0	0	1	1	0	1	0

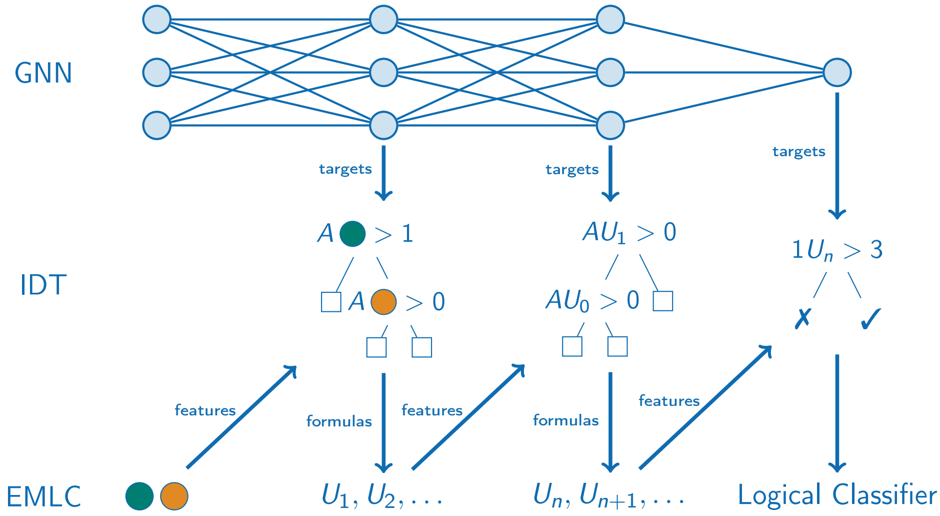
- Train Decision Tree
- Choose new features
- Iterate

Learning deeper EMLC formulas: IDTs

Q: What are the target values for the Decision Trees?

A: GNN Activations!

Learning deeper EMLC formulas: IDTs

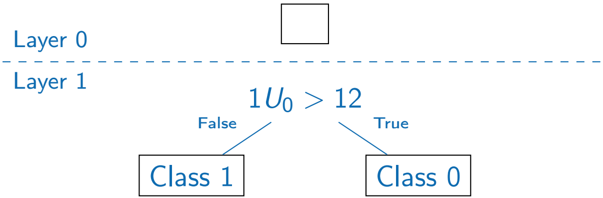


Experiments

- Models
  - GCN+GraphNorm, GIN+Graphnorm (Cai et al., 2021) as baselines.
  - IDT(GCN), IDT(GIN), IDT(GCN+True), IDT(GIN+True), IDT(True).
- Datasets
  - Real-world datasets
    - AIDS (Riesen and Bunke, 2008)
    - BZR (Sutherland et al., 2003)
    - PROTEINS (Borgwardt et al., 2005)
  - Synthetic datasets based on EMLC% formulas of increasing complexity.
  - BAMultiShapes (Azzolin et al., 2023) based on sub-graph motives.
- Metrics
  - Accuracy, F1-Score, Fidelity.

Quantitative Results

Test Accuracy	GCN	IDT (GCN)	IDT (GCN+True)	IDT (True)
AIDS	0.92 ± 0.02	0.99 ± 0.01	1.00 ± 0.00	1.00 ± 0.00
BZR	0.81 ± 0.06	0.79 ± 0.08	0.83 ± 0.06	0.81 ± 0.04
PROTEINS	0.72 ± 0.04	0.73 ± 0.04	0.74 ± 0.03	0.71 ± 0.03
ψ_0	1.00 ± 0.00	1.00 ± 0.00	1.00 ± 0.00	1.00 ± 0.00
ψ_1	0.88 ± 0.02	0.93 ± 0.02	0.97 ± 0.07	0.96 ± 0.04
ψ_2	0.81 ± 0.02	0.94 ± 0.01	0.96 ± 0.01	0.99 ± 0.03
BAMulti	0.99 ± 0.02	1.00 ± 0.01	1.00 ± 0.00	1.00 ± 0.00



Distilled IDT for AIDS. The rule derived for class 0 is $1\top > 12$.
It expresses that the graph has more than 12 nodes and achieves 99% accuracy.

Conclusions and Future Work

- We present EMLC%, an extension of EMLC.
- Leveraging the deep connection between EMLC% and GNNs, we introduce Iterated Decision Trees, a logical distillation model for graph neural networks.
- While highly interpretable, we demonstrate that IDTs are also perform well, in contrast to many other distillation methods.
- Surprisingly, IDTs perform well as a stand-alone model.

References

Azzolin, S., Longa, A., Barbiero, P., Liò, P., and Passerini, A. (2023). Global explainability of gnns via logic combination of learned concepts. In *International Conference on Learning Representations, (ICLR)*.

Barceló, P., Kostylev, E. V., Monet, M., Pérez, J., Reutter, J. L., and Silva, J. P. (2020). The logical expressiveness of graph neural networks. In *International Conference on Learning Representations, (ICLR)*.

Borgwardt, K. M., Ong, C. S., Schönauer, S., Vishwanathan, S., Smola, A. J., and Kriegel, H.-P. (2005). Protein function prediction via graph kernels. *Bioinformatics*, 21(1):47–56.

References

Cai, T., Luo, S., Xu, K., He, D., Liu, T.-y., and Wang, L. (2021). Graphnorm: A principled approach to accelerating graph neural network training. In *International Conference on Machine Learning (ICML)*.

Riesen, K. and Bunke, H. (2008). IAM graph database repository for graph based pattern recognition and machine learning. In *International Workshop on Structural, Syntactic, and Statistical Pattern Recognition*.

Sutherland, J. J., O'brien, L. A., and Weaver, D. F. (2003). Spline-fitting with a genetic algorithm: A method for developing classification structure- activity relationships. *Journal of chemical information and computer sciences*, 43(6):1906–1915.

Definition 10

Assume that each node has an associated initial Boolean feature vector $x_v = x_v^{(0)}$. A GNN computes a vector $x_v^{(k)}$ for every node v via the following recursive formula

$$x_v^{(k+1)} = \mathrm{comb}_k(x_v^{(k)}, \mathrm{agg}_k(\{\!\{x_w^{(k)} : w \in N(v)\}\!\})), \tag{1}$$

where $k \in [l]$. In graph classification, the vectors $x_v^{(l)}$ are then *pooled*

$$\hat{y} = \mathrm{pool}(\{\!\{x_v^{(l)} : v \in V\}\!\}) \tag{2}$$

to give a single graph vector $\hat{y}$, the output of the GNN.

Experiments: EMLC% formulas

The following formulas of increasing complexity are considered:

- $\psi_0 := 1U_1 > 0.5.$
  "More than half of the nodes satisfy U_1."
- $\psi_1 := 1((AU_0 < 4) \vee (AU_0 > 9)) > 0.$
  "There is a node v such that $d_v < 4$ or $d_v > 9$."
- $\psi_2 := 1(A(AU_0 > 6) > 0.5) > 0.5$
  "For at least half the nodes at least half of their neighbors have degree greater than 6"

TECHNISCHE UNIVERSITÄT WIEN

Quantitative Results

F1-Score (macro)	GCN	IDT (GCN)	IDT (GCN+True)	IDT (True)
AIDS	0.88 ± 0.04	0.98 ± 0.02	1.00 ± 0.00	1.00 ± 0.00
BZR	0.73 ± 0.07	0.65 ± 0.12	0.63 ± 0.08	0.68 ± 0.05
PROTEINS	0.71 ± 0.04	0.72 ± 0.04	0.73 ± 0.03	0.69 ± 0.04
ψ_0	1.00 ± 0.00	1.00 ± 0.00	1.00 ± 0.00	1.00 ± 0.00
ψ_1	0.86 ± 0.03	0.92 ± 0.02	0.96 ± 0.09	0.95 ± 0.05
ψ_2	0.80 ± 0.02	0.94 ± 0.01	0.95 ± 0.01	0.99 ± 0.03
BAMulti	0.99 ± 0.02	1.00 ± 0.01	1.00 ± 0.00	1.00 ± 0.01

Quantitative Results

Fidelity (GCN)	GCN	IDT (GCN)	IDT (True)
AIDS	0.92 ± 0.02	0.92 ± 0.02	0.92 ± 0.02
BZR	0.90 ± 0.05	0.80 ± 0.06	0.79 ± 0.05
PROTEINS	0.90 ± 0.05	0.84 ± 0.04	0.80 ± 0.06
ψ_0	1.00 ± 0.00	1.00 ± 0.00	1.00 ± 0.00
ψ_1	0.94 ± 0.01	0.92 ± 0.01	0.85 ± 0.02
ψ_2	0.86 ± 0.01	0.83 ± 0.02	0.81 ± 0.02
BAMulti	0.97 ± 0.02	0.99 ± 0.02	0.98 ± 0.02