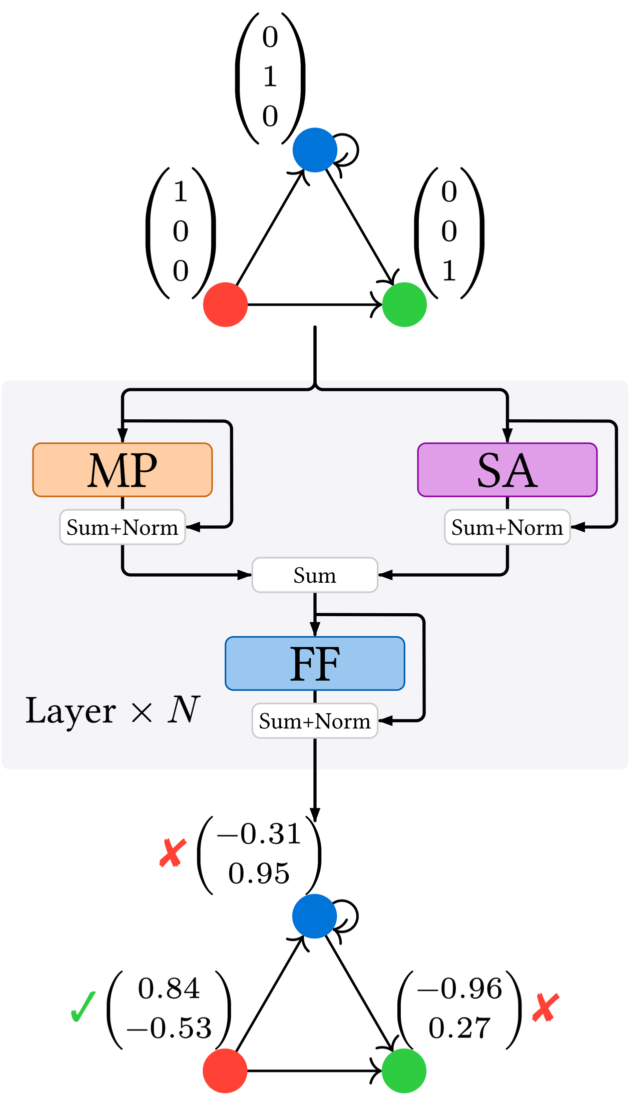
# Expressive Power of Graph Transformers via Logic

Veeti Ahvonen[1], Maurice Funk[2, 3], Damian Heiman[1], Antti Kuusisto[1], and Carsten Lutz[2, 3]
[1]Tampere University, [2]Leipzig University, [3]ScaDS.AI Center Dresden/Leipzig

## Graph Transformers and GPS Networks

**GPS networks** [1] combine message passing of graph neural networks (GNNs) with the attention mechanisms of graph transformers (GTs) [2]. GPS networks compute isomorphism invariant vertex embeddings.




**MP** – **Message Passing**

For every $v \in V$, update feature vector $X_v$:

$$X'_v = \text{COM}(X_v, \text{AGG}(\{\!\{X_u \mid (v,u) \in E\}\!\})).$$

Typically, COM is a FFNN and AGG is +

**SA** – **Self Attention**

Multiple attention heads on vertex features $X$:

$$X' = \text{softmax}\left(\frac{(XW_Q)(XW_K)^T}{\sqrt{d_h}}\right)XW_V.$$

Self attention ignores edges.

**FF** – **Feed-Forward Neural Network**

## Expressive Power via Logic

We view GPS networks $N$ as a binary vertex classifiers: $G, v \models N$ if $X_{v,1} > 0$.

Which vertex properties can be expressed by GPS networks?

Logical characterizations of expressiveness help understanding of theoretical limits and selection of suitable architectures.

GPS networks have the **same expressive power** as a logic $\mathcal{L}$ if for every GPS network $N$ there is a formula $\varphi \in \mathcal{L}$ such that for all labeled graphs $G$ and vertices $v$,

$$G, v \models N \text{ if and only if } G, v \models \varphi$$

and vice versa.

**Theorem (Barceló et al. 2020) [3]**

GNNs and graded modal logic (GML) have the same expressive power relative to first order logic (FO).

Message passing is purely local.
In contrast, self attention can express global properties, like

there is a blue vertex somewhere in the graph.

## Graded Modal Logic

**GML formulas** $\varphi$ are defined by

$$\varphi := p \mid \neg\varphi \mid \varphi \wedge \varphi \mid \varphi \vee \varphi \mid \Diamond_{\geq k}\varphi.$$

$\Diamond_{\geq 2}$Blue is satisified at vertices that have at least two red neighbors.

**GML with global counting** (GML + GC) can additionally use

$$\langle G \rangle_{\geq k}\varphi.$$

$\langle G \rangle_{\geq 2}$Blue is satisfied if there are at least two red vertices in the graph

**GML with global modality** (GML + G) can only use $\langle G \rangle_{\geq 1}\varphi$.

GML + G cannot count globally

**Propositional logic with global counting** (PL + GC) or with global modality (PL +G) cannot use $\Diamond_{\geq k}$.

PL+GC cannot speak about neighbors

## Results on Real Numbers

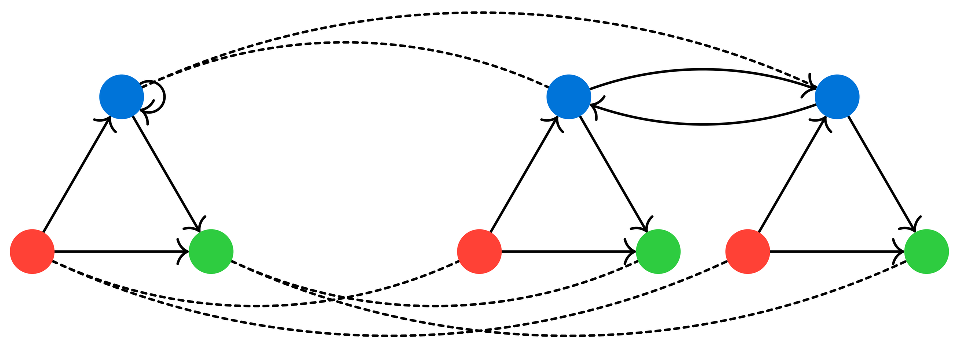
A theoretical model of computations with infinite precision.

**Theorem**

Relative to FO, the following pairs have the same expressive power:

GPS networks and GML + G
GTs and PL + G

1. GML+G formulas can be inductively translated into GPS networks:
   - use message passing for $\Diamond_{\geq k}\varphi$,
   - use self attention for $\langle G \rangle_{\geq 1}\varphi$.
2. GPS networks are invariant under a certain kind of bisimulation.



We show that every FO formula that is invariant under this relation is equivalent to a GML+G formula.

Hence, GPS networks on real numbers cannot express $\langle G \rangle_{\geq 2}$Blue. However, GPS networks on real numbers can express properties not expressible in FO, e.g.:

$\frac{1}{3}$ of the vertices are blue.

## Results on Floating Point Numbers

Floating point numbers are restricted to a finite number of bits and operations are not associative. This breaks bisimulation invariance.

**Theorem**

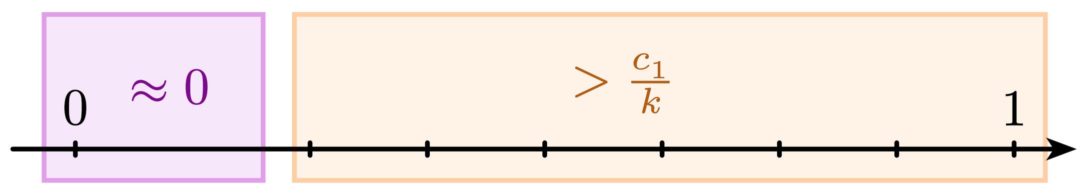
The following pairs have the same expressive power:

floating point GPS networks and GML + GC
floating point GTs and PL + GC

This is not relative to FO but holds absolutly!

Hence, GPS networks on floating point numbers can express $\langle G \rangle_{\geq 2}$Blue and they cannot express that $\frac{1}{3}$ of the vertices are blue.

$\langle G \rangle_{\geq k}\varphi$ can be expressed in floating point numbers, as there always exist $c_1, c_2$ such that $\left(\frac{1}{x}c_1c_2\right) = 0$ if $x \geq k$ and $\left(\frac{1}{x}c_1c_2\right) \geq 1$ if $x < k$:




## Future Work

We want to:
- transfer our results to graph classification,
- consider models that include common positional encodings like homomorphism counts and graph Laplacians,
- better understand the expressive power of GPS Networks and ACR-GNNs beyond FO.

## References

1. Rampásek L, Galkin M, Dwivedi VP, et al (2022) Recipe for a General, Powerful, Scalable Graph Transformer. In: Proceedings of NeurIPS
2. Dwivedi VP, Bresson X (2020) A Generalization of Transformer Networks to Graphs
3. Barceló P, Kostylev EV, Monet M, et al (2020) The Logical Expressiveness of Graph Neural Networks. In: Proceedings of ICLR

Paper ID: 12185
A long version with proofs is available at https://arxiv.org/abs/2508.01067