# Extended Abstract: Aggregate-Combine-Readout GNNs Can Express Logical Classifiers Beyond the Logic C2

**Stan P Hauke**[1] , **Przemysław Andrzej Wałęga**[2]
[1]Department of Informatics, King's College London, UK
[2]School of Electronic Engineering and Computer Science, Queen Mary University of London, UK
stanislaw.hauke@kcl.ac.uk, p.walega@qmul.ac.uk

**Abstract**

In recent years, there has been a growing interest in studying the expressive power of graph neural networks (GNNs) by linking them to logical languages. This line of study was started by a key finding by Barceló et al. (2020), who proved that graded modal logic (or the guarded part of the logic $C^2$) characterises the logical expressiveness of aggregate-combine GNNs. They left a "challenging open problem" asking if $C^2$ characterises the logical expressiveness of aggregate-combine-readout GNNs. This question has stayed open for five years. In this paper, we solve this open problem by showing that aggregate-combine-readout GNNs can express logical classifiers beyond $C^2$.

## 1 Introduction

*Graph Neural Networks* (GNNs) (Gilmer et al., 2017) are specialised machine learning models customized to process graph-structured data. A key result (Morris et al., 2019; Xu et al., 2019) shows that GNNs have the same *distinguishing power* as the Weisfeiler–Leman (WL) algorithm, which also has the same distinguishing power as the fragment $C^2$ of first-order logic (FO). This creates a close link between GNNs, the WL algorithm, and the logic $C^2$.

These findings explain the distinguishing capabilities of GNNs, but they do not explain which functions GNNs can express. The latter is known as the *uniform expressive power*, and has been recently intensively studied from the logical perspective (Benedikt et al., 2024; Tena Cucala et al., 2025; Cuenca Grau, Feng, and Walega, 2026). In particular, Barceló et al. (2020) have established the seminal result showing that the FO node properties expressible by aggregate-combine GNNs (AC-GNNs) are exactly those definable in graded modal logic. Moreover, the FO node properties expressible by aggregate-combine-readout GNNs (ACR-GNNs) contain all properties definable in $C^2$. Note that the latter is a one way inclusion. The reverse inclusion was left by the authors as an open problem.

The problem of Barceló et al. (2020) is, therefore, whether the FO node properties expressible by ACR-GNNs are exactly those definable in $C^2$. Their paper clearly states that this is "a challenging open problem." Later papers also pointed out this question; for example, Grohe (2021) listed it as Question 4 among "interesting theoretical questions that remain open." This question has stayed unresolved for the last five years.

## 2 Contributions

In this paper, we solve the open problem mentioned above by proving that ACR-GNNs can express FO node classifiers beyond $C^2$. We define a node property, prove that it is expressible in both FO and by an ACR-GNN, and demonstrate that this property is not expressible in $C^2$. We obtain such results in two settings, when graphs given as inputs to GNNs are directed and when they are undirected. Furthermore, we use our results to study the expressive power of infinitary logics. As we show, the infinitary version of $C^2$ can express strictly more FO properties than the standard, finitary $C^2$.

In what follows we will highlight the main ideas and techniques used in the paper. First, we introduce a bounded version, $\mathrm{WL}_c$, of the one-dimensional WL algorithm for node-labelled graphs. As we show, $\mathrm{WL}_c$ characterises the distinguishing power of $C^2_{\ell,c}$, which is the fragment of $C^2$, allowing for formulas of quantifier depth bounded by $\ell$ and whose counting quantifiers $\exists_k$ satisfy $k \leqslant c$. Similarly to the standard WL, our $\mathrm{WL}_c$ is an algorithm iteratively colouring nodes of a (node-coloured) graph. Initially, each vertex $v$ has the colour $W_c^0(v)$, as in the input graph. Every next iteration $\ell + 1$ of the algorithm computes $W_c^{\ell+1}(v)$. In the case of undirected graphs, we let $W_c^{\ell+1}(v) = \Big(W_c^\ell(v), \{\!\{W_c^\ell(w)\}\!\}^c_{w\in N_G(v)}, \{\!\{W_c^\ell(w)\}\!\}^c_{w\in V\setminus(N_G(v)\cup\{v\})}\Big)$, where $N_G(v)$ is the set of direct neighbours of a node $v$, and $\{\!\{\cdot\}\!\}^c$ is a multiset bounded to $c$ multiplicities, for example, $\{\!\{7,7,7,3\}\!\}^2 = \{\!\{7,7,3\}\!\}$. In the case of directed graphs, the definition is more complex.

We show that $\ell$ rounds of applying $\mathrm{WL}_c$ allows us to characterise the expressiveness of $C^2_{\ell,c}$.

**Theorem 1.** *Let $\ell, c \in \mathbb{N}$. For any graphs $G$ and $H$ with nodes $u$ and $v$, $W_c^\ell(u) = W_c^\ell(v)$ if and only if $G, u$ and $H, v$ agree on all $C^2_{\ell,c}$ formulas with one free variable.*

Next, we examine the expressiveness of ACR-GNNs, starting from the case of directed graphs. We prove that checking if the edges of a graph form a strict linear order (irreflexive, transitive, and total relation) can be expressed in FO and by ACR-GNNs, but it cannot be defined in $C^2$. While this is a graph property, we can also frame it as a node classifier as follows.

**Definition 2.** *We let $\varphi_{Lin}(x)$ be a node classifier accepting a node of a graph $G$ if and only if $G$ is a strict linear order.*

Clearly, $\varphi_{Lin}(x)$ is expressible in FO, but showing that $\varphi_{Lin}(x)$ can be expressed as an ACR-GNN is more challenging, since ACR-GNNs cannot detect transitivity. We construct the required ACR-GNN, whose application to a linear order of length four is presented in Figure 1. The idea is as follows. The first layer maps the initial vector of a node $v$ into the number $10^n$, where $n$ is the in-degree of $v$. The second layer maps a vector of $v$ into a vector in $\mathbb{R}^2$ of the form $(10^n, 10^{k_1} + \cdots + 10^{k_n})$ where $10^n$ is as in the first layer, whereas each $k_i$ is the in-degree of the $i$th among the $n$ in-neighbours of $v$. The third layer maps each vector to $(1)$ if both of the following conditions hold:

(i) $x[1] \neq y[1]$, for every pair $x, y \in M$.

(ii) if $x[1] = 10^n$, then $x[2] = \underbrace{1 \dots 1}_{n \text{ times}}$, for each $x \in M$,

and otherwise it maps each vector to $(0)$.

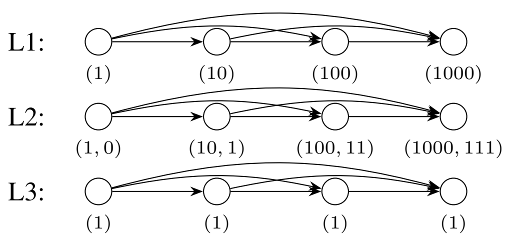


Figure 1: Application of layers 1–3 of the ACR-GNN to the strict linear order with four nodes

Condition (i) guarantees that each node has a different in-degree. If this is the case, then Condition (ii) verifies if the graph is total. This suffices as we can prove that a finite binary relation $E$ is a strict linear order if and only if $E$ is irreflexive, total, and each element has a different number of $E$-successors.

We can also show that $\varphi_{Lin}(x)$ cannot be expressed in $\mathrm{C}^2$. The proof is by contradiction exploiting Theorem 1. If $\varphi_{Lin}(x)$ is expressible in $\mathrm{C}^2$, it is definable in $\mathrm{C}^2_{\ell,c}$ for some $\ell, c \in \mathbb{N}$. To obtain a contradiction, we let $n = \ell \cdot c + 1$ and $G$ be a strict linear order of size $2n + 1$. Let $G'$ be the same as $G$ but with the edge between the $(n-1)$-th and $(n+1)$-th elements pointing the other direction. We obtain that $G \models \varphi_{Lin}(v)$ and $G' \not\models \varphi_{Lin}(v')$, but applying applying Theorem 1, we get $G, v \equiv_{\mathrm{C}^2_{\ell,c}} G', v'$ for any node $v$ and its corresponding node $v'$.

As a result we obtain the following result.

**Theorem 3.** *Over directed graphs, there are FO classifiers expressible by ACR-GNNs which are not expressible in $\mathrm{C}^2$.*

For undirected graphs, we cannot use the $\varphi_{Lin}(x)$. We use a more complex $\varphi_{GadLin}(x)$, which accepts nodes which are a part of a *gadgetised linear order*, namely an undirected graph $\mathsf{gad}(G)$ formed by encoding (gadgetising) a strict linear order $G$. In particular, for a directed graph $G = (V, E)$, we let $\mathsf{gad}(G)$ be an undirected graph on $|V| + 2|E|$ vertices, where each directed edge $(u, w)$ in $G$ is replaced with a new path of three undirected edges, referred to as *gadgetised edges*, and coloured in a way that encodes its direction.

**Definition 4.** *We let $\varphi_{GadLin}(x)$ be a node classifier accepting a node of a graph $G$ if and only if $G$ is isomorphic to $\mathsf{gad}(G')$, for some strict linear order $G'$.*

Following the approach from the directed case, we first show that $\varphi_{GadLin}(x)$ can be expressed in FO. We then prove that it can be expressed by an ACR-GNN, and finally that it cannot be expressed in $\mathrm{C}^2$.

**Theorem 5.** *Over undirected graphs, there are FO node classifiers expressible by ACR-GNNs which are not expressible in $\mathrm{C}^2$. In particular, $\varphi_{GadLin}(x)$ is such a classifier.*

Our results can be also used to show a relation between the expressive power of finitary and infinitary logics. Let $\mathsf{inf}$-$\mathrm{C}^2$ denote the extension of $\mathrm{C}^2$ which allows for infinitary conjunctions and disjunctions. This naturally leads us to the question: what are the FO properties expressible in $\mathsf{inf}$-$\mathrm{C}^2$? It may be tempting to assume that those are exactly the properties expressible in $\mathrm{C}^2$. As we can show, however, this is not the case.

**Theorem 6.** *There are strictly more FO properties expressible in $\mathsf{inf}$-$C^2$ than the properties expressible in $C^2$. This result holds both over directed and undirected graphs.*

## References

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 *Proc. of ICLR*.

Benedikt, M.; Lu, C.; Motik, B.; and Tan, T. 2024. Decidability of graph neural networks via logical characterizations. In *Proc. of ICALP*.

Cuenca Grau, B.; Feng, E.; and Walega, P. 2026. The correspondence between bounded graph neural networks and fragments of first-order logic. In *Proc. of AAAI*.

Gilmer, J.; Schoenholz, S. S.; Riley, P. F.; Vinyals, O.; and Dahl, G. E. 2017. Neural message passing for quantum chemistry. In *Proc. of ICML*, 1263–1272.

Grohe, M. 2021. The logic of graph neural networks. In *Proc. of LICS*, 1–17.

Morris, C.; Ritzert, M.; Fey, M.; Hamilton, W. L.; Lenssen, J. E.; Rattan, G.; and Grohe, M. 2019. Weisfeiler and leman go neural: Higher-order graph neural networks. In *Proc. of AAAI*, 4602–4609.

Tena Cucala, D. J.; Cuenca Grau, B.; Motik, B.; and Kostylev, E. V. 2025. Expressive power of monotonic graph neural networks via datalog. In *Handbook on Neurosymbolic AI and Knowledge Graphs*. IOS Press. 780–807.

Xu, K.; Hu, W.; Leskovec, J.; and Jegelka, S. 2019. How powerful are graph neural networks? In *Proc. of ICLR*.