The Correspondence Between Bounded Graph Neural Networks and Fragments of First-Order Logic

Bernardo Cuenca Grau[1], Eva Feng[1], Przemysław Andrzej Wałęga[2]
[1] Department of Computer Science, University of Oxford
[2] School of Electronic Engineering and Computer Science, Queen Mary University of London
bernardo.grau@cs.ox.ac.uk, eva.feng@cs.ox.ac.uk, p.walega@qmul.ac.uk

Abstract

Graph Neural Networks (GNNs) address two key challenges in applying deep learning to graph-structured data: they handle varying size input graphs and ensure invariance under graph isomorphism. While GNNs have demonstrated broad applicability, understanding their expressive power remains an important question. In this paper, we propose GNN architectures that correspond precisely to prominent fragments of first-order logic (FO), including various modal logics as well as more expressive two-variable fragments. To establish these results, we apply methods from finite model theory of first-order and modal logics to the domain of graph representation learning. Our results provide a unifying framework for understanding the logical expressiveness of GNNs within FO.

Extended version — https://arxiv.org/abs/2505.08021

1 Introduction

Learning on graphs or relational structures presents two fundamental challenges. First, neural networks require fixed size inputs, making them ill-suited for graphs of varying size. Second, predictions about graphs should not depend on how the graph is represented, i.e., they should be invariant under isomorphism (Hamilton 2020).

Graph Neural Networks (GNNs) (Gilmer et al. 2017) overcome these limitations by operating natively on graph-structured data, inherently handling variable sizes and ensuring representation invariance. The flagship *aggregate-combine* (AC) architecture can be viewed as a layered network operating over an input graph. Each node maintains a state (a real-valued vector) and, in each layer, a node's state is updated based on its current state and that of its neighbours. This update mechanism is specified by an *aggregation function* that takes the current states of the neighbours and aggregates them into a vector, and a *combination function* that takes the aggregate value from the neighbours and the current state of the node and computes the updated state. Their *aggregate-combine-readout* (ACR) extension includes an additional *readout function* which aggregates states across all nodes in the graph, rather than just local neighbours (Barceló et al. 2020). GNNs have been widely applied. They drive recommendation systems (Ying et al. 2018), predict molecular properties (Besharatifard and Vafaee 2024), enhance traffic navigation (Derrow-Pinion et al. 2021), interpret scenes in computer vision (Chen et al. 2022), and enable reasoning over incomplete knowledge graphs (Tena Cucala et al. 2022; Zhang and Chen 2018; Huang et al. 2024).

GNNs encompass many architectures and a central question is understanding their *expressive power*—i.e., the classes of functions they can compute. This has been addressed from multiple angles. Early works studied the *discriminative power* of GNNs: given two graphs, can a GNN from a given family yield distinct outputs for them? By design, no GNN can separate isomorphic graphs, but more subtly, certain non-isomorphic graphs may remain indistinguishable. In particular, if two graphs cannot be distinguished by the 1-dimensional Weisfeiler-Leman (WL) graph isomorphism test, then no GNN can differentiate them either (Morris et al. 2019; Xu et al. 2019). Generalised k-dimensional GNNs, which handle higher-order graph structures, have also been connected to the WL hierarchy of increasingly powerful isomorphism tests (Morris et al. 2019). Through the correspondence between WL and finite-variable logics (Cai, Fürer, and Immerman 1992), the limitation extends to logical distinguishability.

The expressiveness of GNNs has also been studied through the lenses of database query languages. As node classifiers, GNNs compute a *unary query*—an isomorphism-invariant function mapping each graph and node to a truth value. For a family of GNN classifiers, what is the logic expressing these unary queries? This is the *logical expressiveness* (or *uniform expressiveness*) of GNNs. The expressiveness of GNNs goes beyond first-order logic (FO) since aggregation can only be captured using extensions such as counting terms (Grohe 2024; Huang et al. 2023), Presburger quantifiers (Benedikt et al. 2024), or linear programming (Nunn et al. 2024). Other GNN variants, such as recursive GNNs (Ahvonen et al. 2024; Pflueger, Cucala, and Kostylev 2024) require fixpoint operators.

A connection between GNNs and FO fragments has also been established (Barceló et al. 2020). Graded modal logic ($\mathcal{GML}$) formulas, a.k.a. concepts in the description logic $\mathcal{ALCQ}$ (Baader et al. 2003), can be captured by GNNs without readout functions, just as FO formulas with two variables and counting quantifiers (C^2) can be realised by a

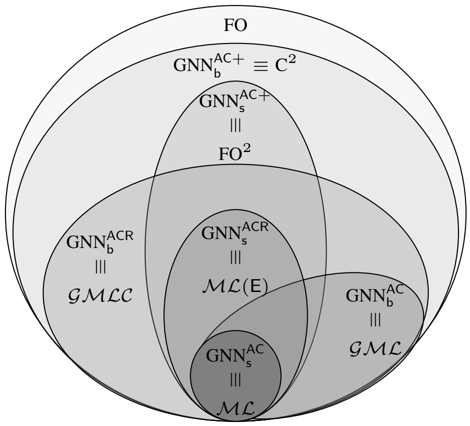

Figure 1: The landscape of our expressive power results

GNNs with readouts. This relationship is, however, asymmetric: while any $\mathcal{GML}$ classifier can be expressed by an AC GNN, the converse requires an assumption of FO expressibility, and in the case of case of ACR GNNs and C^2 the converse does not hold even with this assumption (Hauke and Wałęga 2025). The conditions ensuring FO expressibility of a GNN remain largely unexplored: the only sufficient condition known to us is that monotonic GNNs with max aggregation precisely match unions of tree-shaped conjunctive queries (Tena Cucala et al. 2023).

Contributions We introduce *bounded GNNs* with *k-bounded aggregation*, where multiplicities greater than k in a multiset are capped at k. If $k = 1$, the multiplicities do not matter, and we speak of *set-based aggregation*. As we show, bounded GNNs correspond to modal and two-variable FO fragments as depicted in Figure 1.

We first establish that AC GNNs with set-based aggregation ($\text{GNN}_{\text{s}}^{\text{AC}}$) correspond to basic modal logic ($\mathcal{ML}$) and thus to concepts in the description logic $\mathcal{ALC}$. This extends to GNNs using bounded aggregation ($\text{GNN}_{\text{b}}^{\text{AC}}$), which correspond to graded modal logic ($\mathcal{GML}$), that is, concepts of $\mathcal{ALCQ}$. Readouts enable global quantification: GNNs with set-based aggregation and readout ($\text{GNN}_{\text{s}}^{\text{ACR}}$) capture modal logic with the global modality ($\mathcal{ML}(\mathsf{E})$), which corresponds to $\mathcal{ALC}$ with the universal role (Baader et al. 2003), while those with bounded aggregation and readouts ($\text{GNN}_{\text{b}}^{\text{ACR}}$) match graded modal logic with counting ($\mathcal{GMLC}$), which corresponds to $\mathcal{ALCQ}$ equipped with the universal role. While bounded readouts enable global quantification, they cannot express certain first-order properties like "nodes with exactly k non-neighbours". To overcome this limitation, we introduce $\text{GNN}_{\text{b}}^{\text{AC+}}$: a family of bounded GNNs augmented with an aggregation function over non-neighbours. We prove that $\text{GNN}_{\text{b}}^{\text{AC+}}$ captures C^2 (two-variable FO with counting), while its set-based variant $\text{GNN}_{\text{s}}^{\text{AC+}}$ corresponds to the two-variable FO fragment FO^2.

2 Graphs and Classifiers

Graphs We consider *(finite, undirected, simple, and node-labelled) graphs* $G = (V, E, \lambda)$, where V is a finite set of nodes, E a set of undirected edges with no self-loops, and $\lambda : V \to \{0, 1\}^d$ assigns to each node a binary vector[1] of dimension d. The dimension d of all vectors in G is the same, and we refer to it as the *dimension* of G. A *pointed graph* is a pair (G, v) of a graph and one of its nodes.

Node Classifiers A *node classifier* is a function mapping pointed graphs to true or false. The classifier accepts the input if it returns true and it rejects if it returns false. Two classifiers are *equivalent* if they compute the same function. A family $\mathcal{F}$ of classifiers is *at most as expressive* as $\mathcal{F}'$, written $\mathcal{F} \leq \mathcal{F}'$, if each classifier in $\mathcal{F}$ has an equivalent one in $\mathcal{F}'$. If $\mathcal{F} \leq \mathcal{F}'$ and $\mathcal{F}' \leq \mathcal{F}$, we write $\mathcal{F} \equiv \mathcal{F}'$, and say that $\mathcal{F}$ and $\mathcal{F}'$ have the *same expressiveness*. Such *uniform expressiveness* contrasts with other notions such as discriminative power (Morris et al. 2019; Wang and Zhang 2022) and non-uniform expressiveness (Grohe 2024).

GNN Classifiers We consider standard GNNs with *aggregate-combine* (AC) and *aggregate-combine-readout* (ACR) layers (Benedikt et al. 2024; Barceló et al. 2020), and propose also *extended aggregate-combine* (AC+) layers equipped with one aggregation over neighbours and another over non-neighbours. An AC layer is a pair $(\mathsf{agg}, \mathsf{comb})$, an ACR layer is a triple $(\mathsf{agg}, \mathsf{comb}, \mathsf{read})$ and an AC+ layer is a triple $(\mathsf{agg}, \overline{\mathsf{agg}}, \mathsf{comb})$, where agg and $\overline{\mathsf{agg}}$ are *aggregation functions* and read is a *readout function*, all mapping multisets of vectors into single vectors, whereas comb is a *combination function* mapping vectors to vectors. An application of a layer to a graph $G = (V, E, \lambda)$ yields a graph $G' = (V, E, \lambda')$ with the same nodes and edges, but with an updated labelling function λ'. For an AC layer, vector $\lambda'(v)$ is defined as follows for each node v, where $N_G(v) = \{w \mid \{u, w\} \in E\}$ is the set of neighbours of v and $\overline{N}_G(v) = \{w \mid \{u, w\} \notin E\} \setminus \{v\}$ is the set of non-neighbours of v excluding v itself (note that the graphs we consider have no self loops).

$$\mathsf{comb}\Big(\lambda(v), \mathsf{agg}(\{\!\!\{\lambda(w)\}\!\!\}_{w \in N_G(v)})\Big). \tag{1}$$

For an ACR layer, vector $\lambda'(v)$ is defined as

$$\begin{aligned} \mathsf{comb}\Big(\lambda(v),&\mathsf{agg}(\{\!\!\{\lambda(w)\}\!\!\}_{w \in N_G(v)}),\\ &\mathsf{read}(\{\!\!\{\lambda(w)\}\!\!\}_{w \in V})\Big). \end{aligned} \tag{2}$$

In turn, for an AC+ layer, vector $\lambda'(v)$ is defined as

$$\begin{aligned} \mathsf{comb}\Big(\lambda(v),&\mathsf{agg}(\{\!\!\{\lambda(w)\}\!\!\}_{w \in N_G(v)}),\\ &\overline{\mathsf{agg}}(\{\!\!\{\lambda(w)\}\!\!\}_{w \in \overline{N}_G(v)})\Big). \end{aligned} \tag{3}$$

Each agg, $\overline{\mathsf{agg}}$, comb, and read has domain and range of some fixed dimension (but each can have a different dimension), which we refer to as input and output dimensions. For

[1] The assumption that node labels are binary, i.e., nodes are coloured, is standard when studying logical characterisation of GNNs (Barceló et al. 2020; Benedikt et al. 2024; Nunn et al. 2024)

layer application to be meaningful, these dimensions need to match: if in an AC layer agg has input dimension d and output dimension d', then the input dimension of comb is $d + d'$; in an ACR layer, if agg has dimensions d and d', and read has dimensions d and d'' (note that input dimensions of agg and read need to match), the input dimension of comb is $d + d' + d''$; if in an AC+ layer, agg has input dimension d and output dimension d', and $\overline{\mathsf{agg}}$ has dimensions d and d'' , then the input dimension of comb is $d + d' + d''$. A *GNN classifier* $\mathcal{N}$ of dimension d consists of L layers and a classification function cls from vectors to truth values. The input dimension of the first layer is d and consecutive layers have matching dimensions: the output dimension of layer i matches the input dimension of layer $i+1$. We write $\lambda(v)^{(\ell)}$ for the vector of node v upon application of layer ℓ; $\lambda(v)^{(0)}$ is the initial label of v, and $\lambda(v)^{(L)}$ is its final label. The application of $\mathcal{N}$ to (G, v) is the truth value $\mathcal{N}(G, v) = \mathsf{cls}(\lambda(v)^{(L)})$.

Logic Classifiers We consider formulas over finite signatures consisting of a set PROP of propositions (unary predicates) $p_1, p_2, \ldots$ for node colours. Formulas of the *graded modal logic of counting* ($\mathcal{GMLC}$) are defined as follows:

$$\varphi := p \mid \neg\varphi \mid \varphi \wedge \varphi \mid \Diamond_k\varphi \mid \exists_k\varphi,$$

where $p \in$ PROP and, for each $k \in \mathbb{N}$, $\Diamond_k$ and $\exists_k$ are the k-graded modality and the k-counting modality, respectively. The formula $\Diamond_k\varphi$ expresses that at least k accessible worlds satisfy φ, while $\exists_k\varphi$ states that at least k worlds in total satisfy φ. Graded modal logic ($\mathcal{GML}$) is obtained from $\mathcal{GMLC}$ by disallowing counting modalities. Modal logic with the global modality ($\mathcal{ML}(\mathsf{E})$) is obtained from $\mathcal{GMLC}$ by restricting both counting and graded modalities to $k = 1$. Basic modal logic ($\mathcal{ML}$) further restricts $\mathcal{ML}(\mathsf{E})$ by disallowing counting modalities entirely. We also consider the two-variable fragment of first-order logic with counting quantifiers (C^2), where $\exists_k$ denotes counting quantifiers.[2] The classical two-variable fragment (FO^2) is obtained from C^2 by restricting counting quantifiers to $k = 1$.

The *depth* of formula φ the maximum nesting of modal operators ($\Diamond_k$ and $\exists_k$) or quantifiers in it. The *counting rank*, $\mathsf{rk}_\#(\varphi)$, is the maximal among numbers k occurring in its graded and counting modalities ($\Diamond_k$ and $\exists_k$) or in counting quantifiers, or 0 if the formula does not mention any modalities or quantifiers. For $\mathcal{L}$ any of the logics defined above, we denote as $\mathcal{L}_{\ell,c}$ the set of all $\mathcal{L}$ formulas of depth at most ℓ and counting rank at most c.

Formulas are evaluated over pointed models $(\mathfrak{M}_G, v)$, each corresponding to a pair of a (coloured) graph $G = (V, E, \lambda)$ of some dimension d and a node $v \in V$. In the case of modal logics, model $\mathfrak{M}_G = (V, E, \nu)$ has V as the set of modal worlds, E as the symmetric accessibility relation, and the valuation function ν maps each $p_i \in$ PROP to the subset of nodes in V whose vectors have 1 on the i-th position. Valuation ν extends to all modal formulas:

$$\nu(\neg\varphi) := V \setminus \nu(\varphi), \qquad \nu(\varphi_1 \wedge \varphi_2) := \nu(\varphi_1) \cap \nu(\varphi_2),$$
$$\nu(\Diamond_k\varphi) := \{v \mid k \le |\{w \mid \{v, w\} \in E \text{ and } w \in \nu(\varphi)\}|\},$$
$$\nu(\exists_k\varphi) := V \text{ if } k \le |\nu(\varphi)|, \text{ and } \emptyset \text{ otherwise.}$$

A pointed model $(\mathfrak{M}, v)$, *satisfies* a formula φ, denoted $(\mathfrak{M}, v) \models \varphi$, if $v \in \nu(\varphi)$. In the case of C^2 formulas, we treat $\mathfrak{M}_G$ as the corresponding FO structure, and evaluate formulas using the standard FO semantics.

For a logic $\mathcal{L}$, we write $(\mathfrak{M}, w) \equiv_{\mathcal{L}} (\mathfrak{M}', w')$ if $(\mathfrak{M}, w)$ and $(\mathfrak{M}', w')$ satisfy the same formulas of $\mathcal{L}$. A *logic classifier* of dimension d is a formula φ with at most d propositions (unary predicates) $p_1, \ldots, p_d$. The application of φ to (G, v) is true if $(\mathfrak{M}_G, v) \models \varphi$ and false otherwise. By convention, we use the same symbol for a logic and its associated classifier family.

3 Bounded GNN Classifiers

We next introduce *bounded GNNs*, which generalise max GNNs (Tena Cucala and Cuenca Grau 2024) and max-sum GNNs (Tena Cucala et al. 2023). Bounded GNNs restrict aggregation and readout by requiring existence of a bound k such that all multiplicities $k' > k$ in an input multiset are replaced with k. Thus, multiplicities greater than k do not affect the output of a k-bounded function. Set-based functions ignore multiplicities altogether.

Definition 1. *An aggregation (or readout) function f is k-*bounded*, for $k \in \mathbb{N}$, if $f(M) = f(M_k)$ for each multiset M in the domain of f, where M_k is the multiset obtained from M by replacing all multiplicities greater than k with k. Function f is* set-based *if it is 1-bounded, and it is* bounded *if it is k-bounded for some $k \in \mathbb{N}$.*

Example 2. *Consider the example functions below.*

- *The aggregation in max GNNs (Tena Cucala and Cuenca Grau 2024) is set-based. It maps a multiset of vectors to a vector being their componentwise maximum, for instance $\{\!\{(3, 2), (2, 4), (2, 4)\}\!\} \mapsto (3, 4)$.*
- *The aggregation in max-k-sum GNNs (Tena Cucala et al. 2023) is k-bounded. It maps M to a vector whose ith component is the sum of the k largest ith components in M; if $k = 2$, $\{\!\{(3, 2), (2, 4), (2, 4)\}\!\} \mapsto (5, 8)$.*
- *Examples of unbounded functions include componentwise sum $\{\!\{(3, 2), (2, 4), (2, 4)\}\!\} \mapsto (7, 10)$ and the average mapping $\{\!\{(3, 2), (2, 4), (2, 4)\}\!\} \mapsto (\frac{7}{3}, \frac{10}{3})$.*

Equipped with the notion of bounded aggregation and readout functions, we are ready to define bounded GNNs.

Definition 3. *We consider families of GNN classifiers, GNN_X^Y, where $X \in \{\mathsf{s}, \mathsf{b}, \mathsf{m}\}$ indicates the type of aggregation and readout: set-based (s), bounded (b), or arbitrary—also called multiset—(m), whereas $Y \in \{\mathsf{AC}, \mathsf{ACR}, \mathsf{AC+}\}$ indicates whether the GNN uses only* AC, ACR, *or* AC+ *layers.* Bounded GNN classifiers *are those with bounded aggregation and readout functions.*

All GNN_X^Y classifiers, with $X \in \{\mathsf{s}, \mathsf{b}\}$, are bounded. Family $\mathrm{GNN}_{\mathsf{m}}^{\mathsf{AC}}$ corresponds to aggregate-combine GNNs,

[2] We use the symbols $\exists_k$ for both counting quantifiers and counting modalities.

$\text{GNN}_{\text{m}}^{\text{ACR}}$ to aggregate-combine-readout GNNs (Barceló et al. 2020), $\text{GNN}_{\text{b}}^{\text{AC}}$ contains monotonic max-sum GNNs (Tena Cucala et al. 2023), and $\text{GNN}_{\text{s}}^{\text{AC}}$ contains max GNNs (Tena Cucala and Cuenca Grau 2024).

As shown later, the expressiveness of bounded GNN classifiers falls within FO. This is intuitively so, because bounded GNNs have finite spectra, as defined below.

Definition 4. *(Benedikt et al. 2024) The* spectrum, $\mathsf{sp}(\mathcal{N})$*, of a GNN classifier $\mathcal{N}$ (of dimension d), is the set of all vectors that can occur as node labels in any layer of $\mathcal{N}$ application (to graphs of dimension d). For L the number of layers of $\mathcal{N}$ and $\ell \leq L$, we let $\mathsf{sp}(\mathcal{N}, \ell)$ be the subset of the spectrum consisting of the vectors that can occur upon application of layer ℓ. By convention, we let $\mathsf{sp}(\mathcal{N}, 0)$ be the set of Boolean vectors of the classifier's dimension.*

Since $\mathsf{sp}(\mathcal{N}, 0)$ is always finite and bounded functions applied to multisets with a bounded number of vectors yield finitely many possible outcomes, the spectra of bounded GNN classifiers are bounded. Using combinatorial arguments we can obtain explicit bounds as below.

Proposition 5. *Each bounded GNN classifier $\mathcal{N}$ has a finite spectrum. In particular, if $\mathcal{N}$ has dimension d, L layers, and k is the largest bound of its aggregation and readout functions, then $|\mathsf{sp}(\mathcal{N}, 0)| = 2^d$ and $|\mathsf{sp}(\mathcal{N}, \ell + 1)|$, for each $0 \leq \ell \leq L - 1$, is bounded by the following values:*

$$|\mathsf{sp}(\mathcal{N}, \ell)| \cdot (k+1)^{|\mathsf{sp}(\mathcal{N}, \ell)|}, \quad \text{if } \mathcal{N} \in \mathit{GNN}_{\mathsf{b}}^{\mathsf{AC}},$$
$$|\mathsf{sp}(\mathcal{N}, \ell)| \cdot (k+1)^{2|\mathsf{sp}(\mathcal{N}, \ell)|}, \quad \text{if } \mathcal{N} \in \{\mathit{GNN}_{\mathsf{b}}^{\mathsf{ACR}}, \mathit{GNN}_{\mathsf{b}}^{\mathsf{AC}+}\}.$$

4 Overview and Technical Approach

In what follows, we systematically establish the correspondences between bounded GNNs and logic classifiers depicted in Figure 1. In Section 5, we show that bounded aggregate-combine GNNs correspond precisely to the modal logics without global counting ($\mathcal{ML}$ and $\mathcal{GML}$). Next, in Section 6, we show that bounded aggregate-combine-readout GNNs capture the expressive power of modal logics with global counting ($\mathcal{ML}(\mathsf{E})$ and $\mathcal{GMLC}$). Finally, in Section 7, we prove that extended aggregate-combine GNNs are equivalent in expressive power to two-variable logics (C^2 and FO^2).

To characterise a family of GNN classifiers $\mathcal{F}$ via a logic $\mathcal{L}$ we establish a bidirectional correspondence. We first show that every formula of $\mathcal{L}$ can be simulated by a GNN in $\mathcal{F}$. We then show the converse: every GNN classifier in $\mathcal{F}$ admits an equivalent $\mathcal{L}$-classifier.

The second step builds on finite model theory characterisations of $\mathcal{L}$-equivalence through model comparison games (independent of GNNs). The existence of a winning strategy induces an equivalence relation $\sim$ on pointed models, which extends to pointed coloured graphs: $(G, v) \sim (G', v')$ holds precisely when $(\mathfrak{M}_G, v) \sim (\mathfrak{M}_{G'}, v')$. By limiting games to a fixed number of rounds and bounded grading, and by considering finite signatures, we ensure that $\sim$-invariant classes of models can be represented as a finite disjunction of *characteristic formulas of* $\mathcal{L}$ (Otto 2019; Libkin 2004). Specifically, for modal logics and two-variable fragments we use suitable variants of bisimulation and 2-pebble games, respectively.

The final requirement to establish the connection to GNNs is to show that each GNN classifier $\mathcal{N}$ in $\mathcal{F}$ is invariant under $\sim$, i.e., $(G_1, v_1) \sim (G_2, v_2)$ implies $\mathcal{N}(G_1, v_1) = \mathcal{N}(G_2, v_2)$, for all pointed graphs (G_1, v_1) and (G_2, v_2).

5 Modal Logics Without Global Counting

We first study bounded aggregate-combine GNNs, and start by showing that $\mathcal{GML}$ and $\mathcal{ML}$ formulas can be captured by $\text{GNN}_{\text{b}}^{\text{AC}}$ and $\text{GNN}_{\text{s}}^{\text{AC}}$ classifiers, respectively. For this, we adapt the construction simulating $\mathcal{GML}$ formulas with GNNs which uses unbounded summation (Barceló et al. 2020). For $\mathcal{GML}$, our GNN construction uses max-k-sum aggregation, and for $\mathcal{ML}$ it uses max aggregation (which coincides with max-k-sum, for $k = 1$).

Theorem 6. $\mathcal{GML} \leq \mathit{GNN}_{\mathsf{b}}^{\mathsf{AC}}$ *and* $\mathcal{ML} \leq \mathit{GNN}_{\mathsf{s}}^{\mathsf{AC}}$.

Proof Sketch. Let $\varphi \in \mathcal{GML}$ be a logic classifier of dimension d with subformulas $\varphi_1, \ldots, \varphi_L$, such that $k \leq \ell$ if φ_k is a subformula of φ_ℓ. We construct $\mathcal{N}_\varphi$ with layers $0, \ldots, L$ and a classification function that maps a vector to $\mathtt{true}$ iff its last element is 1. Layer 0 multiplies input vectors by a matrix $\mathbf{D} \in \mathbb{R}^{d \times L}$, namely $\lambda(v)^{(1)} = \lambda(v)^{(0)}\mathbf{D}$, where $D_{k\ell} = 1$ if the kth position of the input vectors corresponds to a proposition φ_ℓ; other entries of $\mathbf{D}$ are 0. All other layers are AC layers of dimension L using max-n-sum, for n the counting rank of φ. The combination function is $\mathsf{comb}(\mathbf{x}, \mathbf{y}) = \sigma(\mathbf{xC} + \mathbf{yA} + \mathbf{b})$, where $\sigma(x) = min(max(0, x), 1)$ is the truncated ReLU and where entries of matrices $\mathbf{A}, \mathbf{C} \in \mathbb{R}^{L \times L}$ and bias vector $\mathbf{b} \in \mathbb{R}^L$ depend on the subformulas of φ as follows: *(i)* if φ_ℓ is a proposition, $C_{\ell\ell} = 1$, *(ii)* if $\varphi_\ell = \varphi_j \wedge \varphi_k$, then $C_{j\ell} = C_{k\ell} = 1$ and $b_\ell = -1$; *(iii)* if $\varphi_\ell = \neg\varphi_k$, then $C_{k\ell} = -1$ and $b_\ell = 1$, and *(iv)* if $\varphi_\ell = \Diamond_c \varphi_k$, then $A_{k\ell} = 1$ and $b_\ell = -c + 1$. All other entries are zero. Note that if $\varphi \in \mathcal{ML}$, then $\mathcal{N}_\varphi \in \text{GNN}_{\text{s}}^{\text{AC}}$, as required. ☐

We next show that every classifier in $\text{GNN}_{b}^{\text{AC}}$ admits an equivalent $\mathcal{GML}$ classifier whereas each $\text{GNN}_{s}^{\text{AC}}$ classifier admits an equivalent $\mathcal{ML}$ classifier. To this end, we first discuss the game-theoretic characterisations of logical indistinguishability for $\mathcal{GML}$ and $\mathcal{ML}$.

The *ℓ-round c-graded bisimulation game* (Otto 2019) is played by Spoiler (him) and Duplicator (her) on finite pointed models $(\mathfrak{M}, v)$ and $(\mathfrak{M}', v')$. A *configuration* is a tuple $(\mathfrak{M}, w, \mathfrak{M}', w')$, stating that one pebble is placed on world w in $\mathfrak{M}$, and the other on w' in $\mathfrak{M}'$. The initial configuration is $(\mathfrak{M}, v, \mathfrak{M}', v')$. Each round proceeds in the following two steps, leading to the next configuration. (1) Spoiler selects a pebble and a set $U_1 \neq \emptyset$ of at most c neighbours of the world marked by this pebble. Duplicator responds with a set U_2 of neighbours of the world marked by the other pebble, such that $|U_1| = |U_2|$. (2) Spoiler selects a world in U_2 and Duplicator responds with a world in U_1. If a player cannot pick an appropriate set (U_1 or U_2), they lose. If in some configuration worlds marked by pebbles do not satisfy the same propositions, Duplicator loses. Hence, Duplicator wins if she has responses for all ℓ moves of Spoiler, or if Spoiler cannot make a move in some round. Spoiler wins if Duplicator loses. We write $(\mathfrak{M}, v) \sim_{\ell,c} (\mathfrak{M}', v')$ if Duplicator has a winning strategy starting from configuration $(\mathfrak{M}, v, \mathfrak{M}', v')$. Importantly, $\sim_{\ell,c}$ determines indistinguishability of pointed models in $\mathcal{GML}_{\ell,c}$ and any class of pointed models closed under $\sim_{\ell,c}$ is definable by a $\mathcal{GML}_{\ell,c}$ formula.

Theorem 7. *(Otto 2019) For any pointed models $(\mathfrak{M}, v)$ and $(\mathfrak{M}', v')$, and any $\ell, c \in \mathbb{N}$: $(\mathfrak{M}, v) \sim_{\ell,c} (\mathfrak{M}', v')$ iff $(\mathfrak{M}, v) \equiv_{\mathcal{GML}_{\ell,c}} (\mathfrak{M}', v')$ iff $(\mathfrak{M}', v') \models \varphi^{\ell,c}_{[\mathfrak{M},v]}$.*

Here, the characteristic formula $\varphi^{\ell,c}_{[\mathfrak{M},v]}$ *is defined inductively on $n \geq 0$ as follows, for $p \in \mathsf{PROP}$ and $U^n_{v,w}$ the set of worlds u such that $\{v, u\} \in E$ and $(\mathfrak{M}, u) \sim_{n,c} (\mathfrak{M}, w)$.*

$$\varphi^{0,c}_{[\mathfrak{M},v]} := \bigwedge\{p : (\mathfrak{M}, v) \models p\} \wedge \bigwedge\{\neg p : (\mathfrak{M}, v) \not\models p\},$$
$$\varphi^{n+1,c}_{[\mathfrak{M},v]} := \varphi^{0,c}_{[\mathfrak{M},v]} \wedge$$
$$\bigwedge \left\{\Diamond_k \varphi^{n,c}_{[\mathfrak{M},w]} : \{v, w\} \in E, k \leq min(|U^n_{v,w}|, c)\right\} \wedge$$
$$\bigwedge \left\{\neg\Diamond_k \varphi^{n,c}_{[\mathfrak{M},w]} : \{v, w\} \in E \text{ and } |U^n_{v,w}| < k \leq c\right\} \wedge$$
$$\neg\Diamond_1 \bigwedge_{\{v,w\}\in E} \neg\varphi^{n,c}_{[\mathfrak{M},w]}.$$

Moreover, any $\sim_{\ell,c}$-closed class $\mathcal{C}$ of pointed models is definable by the formula $\bigvee_{(\mathfrak{M},v)\in\mathcal{C}} \varphi^{\ell,c}_{[\mathfrak{M},v]}$.

We note two important observations regarding Theorem 7. First, our construction of characteristic formulas corrects an error of Otto (2019), by including a final conjunct that is necessary for the theorem to hold. Second, we observe that the ℓ-round 1-graded bisimulation games coincide with ℓ-round bisimulation games for $\mathcal{ML}$ (Goranko and Otto 2007), and that the characteristic formulas $\varphi^{\ell,1}_{[\mathfrak{M},v]}$ are characteristic formulas of $\mathcal{ML}$.

We can now shift our attention to GNNs and show that classifiers in $\text{GNN}^{\mathsf{AC}}_{\mathsf{b}}$ and $\text{GNN}^{\mathsf{AC}}_{\mathsf{s}}$ are invariant under the bisimulation games for $\mathcal{GML}$ and $\mathcal{ML}$, respectively.

Theorem 8. *The following hold:*

1. *$\text{GNN}^{\mathsf{AC}}_{\mathsf{b}}$ classifiers with L layers and k-bounded aggregation are $\sim_{L,k}$-invariant;*
2. *$\text{GNN}^{\mathsf{AC}}_{\mathsf{s}}$ classifiers with L layers are $\sim_{L,1}$-invariant.*

Proof Sketch. We show by induction on $\ell \leq L$ that, for any pointed graphs satisfying $(G_1, v_1) \sim_{\ell,k} (G_2, v_2)$, the execution of a k-bounded GNN $\mathcal{N} \in \text{GNN}^{\mathsf{AC}}_{\mathsf{b}}$ satisfies $\lambda_1(v_1)^{(\ell)} = \lambda_2(v_2)^{(\ell)}$, and thus $\mathcal{N}(G_1, v_1) = \mathcal{N}(G_2, v_2)$.

If $\ell = 0$, $(G_1, v_1) \sim_{0,k} (G_2, v_2)$ implies that v_1 and v_2 satisfy the same propositions, so $\lambda_1(v_1)^{(0)} = \lambda_2(v_2)^{(0)}$. For $\ell \geq 1$, $(G_1, v_1) \sim_{\ell,k} (G_2, v_2)$ implies $(G_1, v_1) \sim_{\ell-1,k} (G_2, v_2)$, so $\lambda_1(v_1)^{(\ell-1)} = \lambda_2(v_2)^{(\ell-1)}$ by induction. It remains to show that $\mathsf{agg}(\{\!\{\lambda_1(w)^{(\ell-1)}\}\!\}_{w\in N_{G_1}(v_1)})$ equals $\mathsf{agg}(\{\!\{\lambda_2(w)^{(\ell-1)}\}\!\}_{w\in N_{G_2}(v_2)})$, which subsequently implies $\lambda_1(v_1)^{(\ell)} = \lambda_2(v_2)^{(\ell)}$ by Equation (1). Suppose for the sake of contradiction, and without loss of generality, that there is a neighbour w_2 of v_2 such that $\lambda_2(w_2)^{(\ell-1)}$ occurs $k_2 < k$ times in $\{\!\{\lambda_2(w)^{(\ell-1)}\}\!\}_{w\in N_{G_2}(v_2)}$ and $k_1 > k_2$ times in $\{\!\{\lambda_1^{(\ell-1)}(w)\}\!\}_{w\in N_{G_1}(v_1)}$. The strategy for Spoiler is to select a set U_1 of $min(k, k_1)$ elements of $w \in N_{G_1}(v_1)$ satisfying $\lambda_1(w)^{(\ell-1)} = \lambda_1(w_1)^{(\ell-1)}$. Duplicator must respond with a subset U_2 of neighbours of v_2 in G_2 of the same cardinality. Any such U_2 must contain w_2 such that $\lambda_2(w_2)^{(\ell-1)} \neq \lambda_1(w_1)^{(\ell-1)}$. In the second part of the round, Spoiler chooses w_2; then, whichever element w_1' in U_1 Duplicator chooses, we have $\lambda_1(w_1')^{(\ell-1)} \neq \lambda_2^{(\ell-1)}(w_2)$. Hence, by the inductive hypothesis, $(G_1, w_1') \not\sim_{\ell-1,k} (G_2, w_2)$ and hence $(G_1, v_1) \not\sim_{\ell,k} (G_2, v_2)$, raising a contradiction. The proof for $\mathcal{N} \in \text{GNN}^{\mathsf{AC}}_{\mathsf{s}}$ is a particular case, where $k = 1$. □

By Theorem 7, invariance under bisimulation games implies that graphs accepted by a GNN can be characterised by a disjunction of characteristic formulas.

Corollary 9. *Let $\mathcal{N}$ be a GNN with L layers and $\mathcal{C}$ the pointed models $(\mathfrak{M}, v)$ accepted by $\mathcal{N}$. If $\mathcal{N} \in \text{GNN}^{\mathsf{AC}}_{\mathsf{b}}$ with k-bounded aggregations, it is equivalent to the $\mathcal{GML}$ formula $\bigvee_{(\mathfrak{M},v)\in\mathcal{C}} \varphi^{L,k}_{[\mathfrak{M},v]}$. If $\mathcal{N} \in \text{GNN}^{\mathsf{AC}}_{\mathsf{s}}$, it is equivalent to the $\mathcal{ML}$ formula $\bigvee_{(\mathfrak{M},v)\in\mathcal{C}} \varphi^{L,1}_{[\mathfrak{M},v]}$.*

Therefore, $\mathcal{GML} \geq \text{GNN}^{\mathsf{AC}}_{\mathsf{b}}$ and $\mathcal{ML} \geq \text{GNN}^{\mathsf{AC}}_{\mathsf{s}}$. By combining these results with Theorem 6 we obtain the following exact correspondence.

Corollary 10. *$\mathcal{GML} \equiv \text{GNN}^{\mathsf{AC}}_{\mathsf{b}}$ and $\mathcal{ML} \equiv \text{GNN}^{\mathsf{AC}}_{\mathsf{s}}$.*

6 Modal Logics with Global Counting

We now consider bounded GNNs with readouts. We first show that $\mathcal{GMLC}$ is captured by $\text{GNN}^{\mathsf{ACR}}_{\mathsf{b}}$ using max-k-sum as aggregation, whereas $\mathcal{ML}(\mathsf{E})$ is captured by $\text{GNN}^{\mathsf{ACR}}_{\mathsf{s}}$ using componentwise maximum aggregation.

Theorem 11. *$\mathcal{GMLC} \leq \text{GNN}^{\mathsf{ACR}}_{\mathsf{b}}$ and $\mathcal{ML}(\mathsf{E}) \leq \text{GNN}^{\mathsf{ACR}}_{\mathsf{s}}$.*

Proof Sketch. Let $\varphi \in \mathcal{GMLC}$; we will construct a GNN by modifying the construction from the proof of Theorem 6. We now use ACR layers with $\mathsf{comb}(\mathbf{x}, \mathbf{y}, \mathbf{z}) = \sigma(\mathbf{xC} + \mathbf{yA} + \mathbf{zR} + \mathbf{b})$, where $\mathbf{R}$ simulates global modalities. In particular, for subformulas $\varphi_\ell = \exists_c \varphi_k$, we set $R_{k\ell} = 1$ and $b_\ell = -c + 1$, whereas other entries are set to zeros. Moreover, we let read (and agg) be the max-n-sum function where n is the counting rank of φ. The remaining components are as in the proof of Theorem 6. The construction for $\mathcal{ML}(\mathsf{E})$ is obtained as a particular case by taking $c = 1$ and noting the max-1-sum function corresponds to max aggregation. □

We next show that $\text{GNN}^{\mathsf{ACR}}_{\mathsf{b}}$ and $\text{GNN}^{\mathsf{ACR}}_{\mathsf{s}}$ classifiers admit equivalent $\mathcal{GMLC}$ and $\mathcal{ML}(\mathsf{E})$ classifiers, respectively. As a first step, our approach involves developing a new game-theoretic characterisation of $\mathcal{GMLC}$ that naturally covers $\mathcal{ML}(\mathsf{E})$ as a special case.

To this end, we introduce *ℓ-round c-graded global bisimulation games*, by extending the games from Section 5. In each round, Spoiler can now choose to play either a standard (local) round as before, or a *global round*. Each global round proceeds by Spoiler selecting a non-empty set U_1 of worlds of size bounded by c in one of the models, and Duplicator subsequently picking a set U_2 of worlds of the same size in the other model; Spoiler then places a pebble in a world u in U_2 and Duplicator responds by placing a pebble on a world u' in U_1, leading to a new configuration $(\mathfrak{M}, u, \mathfrak{M}', u')$. We write $(\mathfrak{M}, v) \sim^{\exists}_{\ell,c} (\mathfrak{M}', v')$ if Duplicator has a winning strategy in the ℓ-round c-graded global bisimulation game starting at $(\mathfrak{M}, v, \mathfrak{M}', v')$. Similarly as in the case of non-global games (Theorem 7), we can show the following characterisation result.

Theorem 12. *For any pointed models $(\mathfrak{M}, v)$ and $(\mathfrak{M}', v')$ and $\ell, c \in \mathbb{N}$, $(\mathfrak{M}, v) \sim^{\exists}_{\ell,c} (\mathfrak{M}', v')$ iff $(\mathfrak{M}, v) \equiv_{\mathcal{GMLC}_{\ell,c}} (\mathfrak{M}', v')$ iff $(\mathfrak{M}', v') \models \varphi^{\ell,c}_{\exists[\mathfrak{M},v]}$.*

Here, the characteristic formulas are defined inductively on $n \geq 0$ as follow, where $U^n_{v,w}$ is the set of all u with $\{v, u\} \in E$ and $(\mathfrak{M}, u) \sim^{\exists}_{n,c} (\mathfrak{M}, w)$, and J^n_w is the set of all $u \in V$ with $(\mathfrak{M}, u) \sim^{\exists}_{n,c} (\mathfrak{M}, w)$.

$$\varphi^{0,c}_{\exists[\mathfrak{M},v]} := \bigwedge\{p : (\mathfrak{M}, v) \models p\} \wedge \bigwedge\{\neg p : (\mathfrak{M}, v) \not\models p\},$$

$$\varphi^{n+1,c}_{\exists[\mathfrak{M},v]} := \varphi^{0,c}_{\exists[\mathfrak{M},v]} \wedge$$

$$\bigwedge \left\{ \Diamond_k \varphi^{n,c}_{\exists[\mathfrak{M},w]} : \{v, w\} \in E, k \leq min(|U^n_{v,w}|, c) \right\} \wedge$$

$$\bigwedge \left\{ \neg\Diamond_k \varphi^{n,c}_{\exists[\mathfrak{M},w]} : \{v, w\} \in E \textit{ and } |U^n_{v,w}| < k \leq c \right\} \wedge$$

$$\bigwedge \left\{ \exists_k \varphi^{n,c}_{\exists[\mathfrak{M},w]} : w \in V, k \leq |J^n_w|, \textit{ and } k \leq c \right\} \wedge$$

$$\bigwedge \{ \neg\exists_k \varphi^{n,c}_{\exists[\mathfrak{M},w]} : w \in V \textit{ and } |J^n_w| < k \leq c \} \wedge$$

$$\neg\Diamond_1 \bigwedge_{\{v,w\} \in E} \neg\varphi^{n,c}_{\exists[\mathfrak{M},w]} \wedge \neg\exists_1 \bigwedge_{w \in V} \neg\varphi^{n,c}_{\exists[\mathfrak{M},w]}.$$

Moreover, any class $\mathcal{C}$ of pointed models closed under $\sim^{\exists}_{\ell,c}$ is definable by $\bigvee_{(\mathfrak{M},v)\in\mathcal{C}} \varphi^{\ell,c}_{\exists[\mathfrak{M},v]}$.

We can observe that if $c = 1$, then characteristic formulas are in $\mathcal{ML}(\mathsf{E})$, and our games correspond to global bisimulation games developed for $\mathcal{ML}(\mathsf{E})$ (Goranko and Otto 2007, Section 5.1).

We are now ready to show that classifiers in $\text{GNN}^{\mathsf{ACR}}_{\mathsf{b}}$ and $\text{GNN}^{\mathsf{ACR}}_{\mathsf{s}}$ are invariant under the bisimulation games for $\mathcal{GMLC}$ and $\mathcal{ML}(\mathsf{E})$, respectively.

Theorem 13. *The following hold:*

1. *$GNN^{\mathsf{ACR}}_{\mathsf{b}}$ classifiers with L layers and k-bounded aggregations and readouts are $\sim^{\exists}_{L,k}$-invariant;*
2. *$GNN^{\mathsf{ACR}}_{\mathsf{s}}$ classifiers with L layers are $\sim^{\exists}_{L,1}$-invariant.*

Proof Sketch. The proof has the same structure as in Theorem 8. The inductive hypothesis implies $\lambda_1(v_1)^{(\ell-1)} = \lambda_2(v_2)^{(\ell-1)}$ and $min(k, \{\!\{\lambda_1(w)^{(\ell-1)}\}\!\}_{w\in N_{G_1}(v_1)})$ equals $min(k, \{\!\{\lambda_2(w)^{(\ell-1)}\}\!\}_{w\in N_{G_2}(v_2)})$. Additionally we can show now that $\mathsf{read}(\{\!\{\lambda_1(w)^{(\ell-1)}\}\!\}_{w\in V_1})$ equals $\mathsf{read}(\{\!\{\lambda_2(w)^{(\ell-1)}\}\!\}_{w\in V_2})$, which then implies that $\lambda_1(v_1)^{(\ell)} = \lambda_2(v_2)^{(\ell)}$ by Equation (2). Indeed, if these multisets were not equal, Spoiler could find (w.l.o.g.) a node $w_2 \in V_2$ such that $\lambda_2(w_2)^{(\ell-1)}$ occurs $k_2 < k$ times in $\{\!\{\lambda_2(w)^{(\ell-1)}\}\!\}_{w\in V_2}$ and $k_1 > k_2$ times in $\{\!\{\lambda_1^{(\ell-1)}(w)\}\!\}_{w\in V_1}$. This would allow Spoiler to play a global round that wins the game.

The proof for the set-based case can again be obtained as a particular case, with $c = 1$. □

As before, Theorems 12 and 13 imply the following.

Corollary 14. *Let $\mathcal{N}$ be a GNN with L layers and $\mathcal{C}$ the pointed models $(\mathfrak{M}, v)$ accepted by $\mathcal{N}$. If $\mathcal{N} \in GNN^{\mathsf{ACR}}_{\mathsf{b}}$ with k-bounded aggregations, it is equivalent to the $\mathcal{GMLC}$ formula $\bigvee_{(\mathfrak{M},v)\in\mathcal{C}} \varphi^{L,k}_{\exists[\mathfrak{M},v]}$. If $\mathcal{N} \in GNN^{\mathsf{ACR}}_{\mathsf{s}}$, it is equivalent to the $\mathcal{ML}(\mathsf{E})$ formula $\bigvee_{(\mathfrak{M},v)\in\mathcal{C}} \varphi^{L,1}_{\exists[\mathfrak{M},v]}$.*

Therefore, $\mathcal{GMLC} \geq \text{GNN}^{\mathsf{ACR}}_{\mathsf{b}}$ and $\mathcal{ML}(\mathsf{E}) \geq \text{GNN}^{\mathsf{ACR}}_{\mathsf{s}}$. Combining these with Theorem 11 we have the following.

Corollary 15. *The following equivalences hold: $\mathcal{GMLC} \equiv GNN^{\mathsf{ACR}}_{\mathsf{b}}$ and $\mathcal{ML}(\mathsf{E}) \equiv GNN^{\mathsf{ACR}}_{\mathsf{s}}$.*

7 Two-Variable Fragments

Finally, we consider bounded GNNs with non-neighbour aggregation. We show that $\text{GNN}^{\mathsf{AC+}}_{\mathsf{b}}$ and $\text{GNN}^{\mathsf{AC+}}_{\mathsf{s}}$ exactly correspond to C^2 and FO^2, respectively.

We first show that each C^2 classifier can be expressed in $\text{GNN}^{\mathsf{AC+}}_{\mathsf{b}}$ using max-k-sum for neighbour and non-neighbour aggregations, whereas each FO^2 classifier can be expressed in $\text{GNN}^{\mathsf{AC+}}_{\mathsf{s}}$ using componentwise max.

Theorem 16. *$C^2 \leq GNN^{\mathsf{AC+}}_{\mathsf{b}}$ and $FO^2 \leq GNN^{\mathsf{AC+}}_{\mathsf{s}}$.*

Proof Sketch. Similarly to Barceló et al. (2020), we exploit the fact that C^2 has the same expressive power as the modal logic $\mathcal{EMLC}$ with complex modalities (Lutz, Sattler, and Wolter 2001, Theorem 1). We can show that $\mathcal{EMLC}$ classifiers in normal form (Barceló et al. 2020, Lemma D.4) can be captured by $\text{GNN}^{\mathsf{AC+}}_{\mathsf{b}}$. The construction is similar to that of Theorem 6, but using $\mathsf{AC+}$ layers. We use max-n-sum aggregations where n is the counting rank of the $\mathcal{EMLC}$ formula. Then, $\mathsf{comb}(\mathbf{x}, \mathbf{y}, \mathbf{z}) = \sigma(\mathbf{xC} + \mathbf{yA} + \mathbf{z}\overline{\mathbf{A}} + \mathbf{b})$, with matrix and vector entries depending on the subformulas φ_ℓ. If φ_ℓ is a proposition, conjunction, or negation, the ℓth columns of $\mathbf{A}$, $\mathbf{C}$, and $\mathbf{b}$ are as in Theorem 6, and the ℓth column of $\overline{\mathbf{A}}$ has only 0s. For the remaining cases, the ℓth columns of $\mathbf{A}$, $\mathbf{C}$, $\overline{\mathbf{A}}$, and $\mathbf{b}$ are defined as in the construction of (Barceló et al. 2020, Theorem 5.1), except that we use combinations of bounded neighbour and non-neighbour aggregation instead of combinations of unbounded aggregation and global readouts. The proof for FO^2 is a particular case, where operators in $\mathcal{EMLC}$ can count up to 1. □

Games for two-variable logics (Libkin 2004) are similar to bisimulation games, but they are now played with two pairs of pebbles. In what follows we define a variant of the game for C^2 (Grädel and Otto 1999), obtained by imposing restrictions on both the number of rounds and on the possible counting. We let the *2-pebble ℓ-round c-graded game* be played on two models $\mathfrak{M}$ and $\mathfrak{M}'$ by Spoiler and Duplicator with two pairs of pebbles: $(p^1_{\mathfrak{M}}, p^1_{\mathfrak{M}'})$ and $(p^2_{\mathfrak{M}}, p^2_{\mathfrak{M}'})$. After each round, the pebble positions define a mapping π of two elements in $\mathfrak{M}$ into two elements of $\mathfrak{M}'$. Duplicator has a winning strategy if she can ensure that, after each round, π is a partial isomorphism between the models. For node classification, we consider games in which the starting configuration has $p^1_{\mathfrak{M}}$ and $p^1_{\mathfrak{M}'}$ placed on some elements of $\mathfrak{M}$ and $\mathfrak{M}'$, respectively. Each round is played as follows: (1) Spoiler chooses a model (say, $\mathfrak{M}$), one of pebble pairs $i \in \{1, 2\}$, and a non-empty subset $U \subseteq V$ with $|U| \leq c$. Duplicator responds with a subset $U' \subseteq V'$ such that $|U'| = |U|$. (2) Spoiler places pebble $p^i_{\mathfrak{M}'}$ on some $u' \in U'$. Duplicator responds by placing $p^i_{\mathfrak{M}}$ on some $u \in U$. We write $(\mathfrak{M}, a) \sim^2_{\ell,c} (\mathfrak{M}', a')$ if Duplicator has a winning strategy when $p^1_{\mathfrak{M}}$ and $p^1_{\mathfrak{M}'}$ are initially placed on elements a and a', respectively.

As we establish next, this game variant characterises indistinguishability in the logic ${C^2}_{\ell,c}$. What is crucial is that we consider both bounded depth and counting rank. As a result, formulas of ${C^2}_{\ell,c}$ have finitely many equivalence classes (with respect to the logical equivalence), and so, any class of models closed under the equivalence in ${C^2}_{\ell,c}$ is definable by a (finite) ${C^2}_{\ell,c}$ formula.

Theorem 17. *For any pointed models $(\mathfrak{M}, a)$ and $(\mathfrak{M}', a')$ and any $\ell, c \in \mathbb{N}$: $(\mathfrak{M}, a) \sim^2_{\ell,c} (\mathfrak{M}', a')$ iff a in $\mathfrak{M}$ and a' in $\mathfrak{M}'$ satisfy the same $C^2_{\ell,c}$ formulas with one free variable. Furthermore, any class $\mathcal{C}$ of pointed models closed under $\sim^2_{\ell,c}$ is definable by a $C^2_{\ell,c}$ formula.*

Proof Sketch. To show the equivalence, we prove a stronger result, where a and a' are vectors of length at most 2. We show each implication by induction on ℓ. For the forward implication, we show the contrapositive, namely that $\mathfrak{M} \models \varphi(a)$ and $\mathfrak{M}' \not\models \varphi(a')$, imply the existence of a winning strategy for Spoiler. For the opposite direction, we show that whenever $\mathfrak{M} \models \varphi(\mathbf{a})$ iff $\mathfrak{M}' \models \varphi(\mathbf{a}')$, there is a winning strategy for Duplicator.

The above shows that $(\mathfrak{M}, a) \sim^2_{\ell,c} (\mathfrak{M}', a')$ iff a in $\mathfrak{M}$ and a' in $\mathfrak{M}'$ satisfy the same $C^2_{\ell,c}$ formulas with one free variable. Since, up to logical equivalence, there are finitely many $C^2_{\ell,c}$ formulas (Cai, Fürer, and Immerman 1992, Lemma 4.4), each $\sim^2_{\ell,c}$ equivalence class can be expressed as a (finite) disjunction of (finite) $C^2_{\ell,c}$ formulas. □

Next, we show that bounded GNNs with non-neighbour aggregation are invariant under our variant of the 2-pebble games.

Theorem 18. *The following hold:*

1. *$GNN^{\mathsf{AC+}}_{\mathsf{b}}$ classifiers with L layers and k-bounded aggregations and readout are $\sim^2_{L,k}$-invariant;*
2. *$GNN^{\mathsf{AC+}}_{\mathsf{s}}$ classifiers with L layers are $\sim^2_{L,1}$-invariant.*

Proof Sketch. The structure of the proof follows that of Theorem 8, but now games use two pairs of pebbles and GNNs have AC+ layers—with two types of aggregation. The important part of the proof is in the inductive step, where we show that $(G_1, v_1) \sim^2_{\ell,k} (G_2, v_2)$ implies that $min(k, \{\!\{\lambda_1(w)^{(\ell-1)}\}\!\}_{w \in X_{G_1}(v_1)})$ equals $min(k, \{\!\{\lambda_2(w)^{(\ell-1)}\}\!\}_{w \in X_{G_2}(v_2)})$, for both $X \in \{\mathcal{N}, \overline{N}\}$. Towards a contradiction suppose that there is $u \in X_{G_1}(v_1)$ such that $\lambda_1(u)^{(\ell-1)}$ appears k_1 times in $\{\!\{\lambda_1(w)^{(\ell-1)}\}\!\}_{w \in X_{G_1}(v_1)}$ and $k_2 < k_1$ times in $\{\!\{\lambda_2(w)^{(\ell-1)}\}\!\}_{w \in X_{G_2}(v_2)}$, with $k_2 < k$. The winning strategy for Spoiler is to pick a set U_1 of $\min(k, k_1)$ elements from $\{\!\{\lambda_1(w)^{(\ell-1)}\}\!\}_{w \in X_{G_1}(v_1)}$ with label $\lambda_1(u)^{(\ell-1)}$. This allows Spoiler to get to a configuration $(\mathfrak{M}_G, v_1, u_1, \mathfrak{M}_{G'}, v_2, u_2)$ with $\lambda_1(u)^{(\ell-1)} \neq \lambda_2(u_2)^{(\ell-1)}$ which, by the inductive hypothesis, implies $(G_1, v_1) \not\sim^2_{\ell,k} (G_2, v_2)$. The proof for the set-based case is a particular case, with $k = 1$. □

As before, Theorem 17 implies the following.

Corollary 19. *Let $\mathcal{N}$ be a GNN with L layers and $\mathcal{C}$ the pointed models $(\mathfrak{M}, v)$ accepted by $\mathcal{N}$. If $\mathcal{N} \in GNN^{\mathsf{AC+}}_{\mathsf{b}}$ it is equivalent to a C^2 formula. If $\mathcal{N} \in GNN^{\mathsf{AC+}}_{\mathsf{s}}$, it is equivalent to an FO^2 formula.*

Therefore, $C^2 \geq GNN^{\mathsf{AC+}}_{\mathsf{b}}$ and $FO^2 \geq GNN^{\mathsf{AC+}}_{\mathsf{s}}$. Combining these with Theorem 16 we have the following.

Corollary 20. *The following equivalences hold: $C^2 \equiv GNN^{\mathsf{AC+}}_{\mathsf{b}}$ and $FO^2 \equiv GNN^{\mathsf{AC+}}_{\mathsf{s}}$.*

8 Conclusion and Future Work

We have introduced families of bounded GNNs, whose expressive power corresponds exactly to well-known modal logics and 2-variable first-order logics. Among others, we have showed that standard aggregate-combine GNNs with bounded aggregation have the same expressive power as the graded modal logic. This, together with the result of Barceló et al. (2020), implies that an aggregate-combine GNN classifier is FO-expressible if and only if it is equivalent to a bounded aggregate-combine GNN classifier. The correspondence between FO-expressibility and bounding aggregation (and readout) occurs as an interesting phenomenon to study. In particular, we find it interesting to determine for which classes of GNNs classifiers, FO-expressibility is equivalent to expressibility by bounded GNNs. Future work directions we consider include also establishing tight bounds on the size of logical formulas capturing GNNs and practical extraction of logical formulas from GNNs.

Acknowledgements

Eva Feng is generously supported by a Google DeepMind Scholarship (CS2324_DeepMind_1594092).

References

Ahvonen, V.; Heiman, D.; Kuusisto, A.; and Lutz, C. 2024. Logical characterizations of recurrent graph neural networks with reals and floats. In Globersons, A.; Mackey, L.; Belgrave, D.; Fan, A.; Paquet, U.; Tomczak, J. M.; and Zhang, C., eds., *Proc. of NeurIPS*.

Baader, F.; Calvanese, D.; McGuinness, D. L.; Nardi, D.; and Patel-Schneider, P. F., eds. 2003. *The Description*

Logic Handbook: Theory, Implementation, and Applications. Cambridge University Press. ISBN 0-521-78176-0.

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.-H.; Motik, B.; and Tan, T. 2024. Decidability of Graph Neural Networks via Logical Characterizations. In *Proc. of ICALP*, volume 297, 127:1–127:20. Schloss Dagstuhl.

Besharatifard, M.; and Vafaee, F. 2024. A review on graph neural networks for predicting synergistic drug combinations. *Artif. Intell. Rev.*, 57(3): 49.

Cai, J.-Y.; Fürer, M.; and Immerman, N. 1992. An optimal lower bound on the number of variables for graph identification. *Combinatorica*, 12(4): 389–410.

Chen, C.; Wu, Y.; Dai, Q.; Zhou, H.; Xu, M.; Yang, S.; Han, X.; and Yu, Y. 2022. A Survey on Graph Neural Networks and Graph Transformers in Computer Vision: A Task-Oriented Perspective. *CoRR*, abs/2209.13232.

Derrow-Pinion, A.; She, J.; Wong, D.; Lange, O.; Hester, T.; Perez, L.; Nunkesser, M.; Lee, S.; Guo, X.; Wiltshire, B.; Battaglia, P. W.; Gupta, V.; Li, A.; Xu, Z.; Sanchez-Gonzalez, A.; Li, Y.; and Velickovic, P. 2021. ETA Prediction with Graph Neural Networks in Google Maps. In *Proc. of CIKM*, CIKM '21, 3767–3776. ACM.

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*, volume 70, 1263–1272.

Goranko, V.; and Otto, M. 2007. Model theory of modal logic. In Blackburn, P.; van Benthem, J. F. A. K.; and Wolter, F., eds., *Handbook of Modal Logic*, volume 3 of *Studies in logic and practical reasoning*, 249–329. North-Holland.

Grädel, E.; and Otto, M. 1999. On Logics with Two Variables. *Theor. Comput. Sci.*, 224(1-2): 73–113.

Grohe, M. 2024. The descriptive complexity of graph neural networks. *TheoretiCS*, 3.

Hamilton, W. L. 2020. *Graph Representation Learning*. Synthesis Lectures on Artificial Intelligence and Machine Learning. Morgan & Claypool Publishers. ISBN 978-3-031-00460-5.

Hauke, S. P.; and Wałęga, P. A. 2025. Aggregate-Combine-Readout GNNs Are More Expressive Than Logic C2. *arXiv preprint arXiv:2508.06091*.

Huang, X.; Orth, M. A. R.; Barceló, P.; Bronstein, M. M.; and Ceylan, İ. İ. 2024. Link Prediction with Relational Hypergraphs. *CoRR*, abs/2402.04062.

Huang, X.; Romero, M.; Ceylan, İ. İ.; and Barceló, P. 2023. A Theory of Link Prediction via Relational Weisfeiler-Leman on Knowledge Graphs. In Oh, A.; Naumann, T.; Globerson, A.; Saenko, K.; Hardt, M.; and Levine, S., eds., *Proc. of NeurIPS*.

Libkin, L. 2004. *Elements of Finite Model Theory*. Texts in Theoretical Computer Science. An EATCS Series. Springer.

Lutz, C.; Sattler, U.; and Wolter, F. 2001. Modal Logic and the Two-Variable Fragment. In Fribourg, L., ed., *Proc. of CSL*, volume 2142 of *Lecture Notes in Computer Science*, 247–261. Springer.

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*, volume 33, 4602–4609.

Nunn, P.; Sälzer, M.; Schwarzentruber, F.; and Troquard, N. 2024. A Logic for Reasoning about Aggregate-Combine Graph Neural Networks. In *Proc. of IJCAI*, 3532–3540. ijcai.org.

Otto, M. 2019. Graded modal logic and counting bisimulation. *CoRR*, abs/1910.00039.

Pflueger, M.; Cucala, D. T.; and Kostylev, E. V. 2024. Recurrent Graph Neural Networks and Their Connections to Bisimulation and Logic. In Wooldridge, M. J.; Dy, J. G.; and Natarajan, S., eds., *Proc. of AAAI*, 14608–14616. AAAI Press.

Tena Cucala, D.; Cuenca Grau, B.; Motik, B.; and Kostylev, E. V. 2023. On the Correspondence Between Monotonic Max-Sum GNNs and Datalog. In Marquis, P.; Son, T. C.; and Kern-Isberner, G., eds., *Proc. of KR*, 658–667.

Tena Cucala, D. J.; and Cuenca Grau, B. 2024. Bridging Max Graph Neural Networks and Datalog with Negation. In Marquis, P.; Ortiz, M.; and Pagnucco, M., eds., *Proc. of KR*.

Tena Cucala, D. J.; Cuenca Grau, B.; Kostylev, E. V.; and Motik, B. 2022. Explainable GNN-Based Models over Knowledge Graphs. In *Proc. of ICLR*.

Wang, X.; and Zhang, M. 2022. How Powerful are Spectral Graph Neural Networks. In *Proc. of ICML*.

Xu, K.; Hu, W.; Leskovec, J.; and Jegelka, S. 2019. How Powerful are Graph Neural Networks? In *Proc. of ICLR*.

Ying, R.; He, R.; Chen, K.; Eksombatchai, P.; Hamilton, W. L.; and Leskovec, J. 2018. Graph Convolutional Neural Networks for Web-Scale Recommender Systems. In *Proc. of the 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '18, 974–983. ACM.

Zhang, M.; and Chen, Y. 2018. Link Prediction Based on Graph Neural Networks. In Bengio, S.; Wallach, H. M.; Larochelle, H.; Grauman, K.; Cesa-Bianchi, N.; and Garnett, R., eds., *Proc. of NeurIPS*, 5171–5181.