Logical Characterizations of GNNs with Mean Aggregation

Moritz Schönherr, Carsten Lutz

Department of Computer Science, Leipzig University, Germany
Center for Scalable Data Analytics and Artificial Intelligence (ScaDS.AI), Dresden/Leipzig, Germany
{schoenherr, clu}@informatik.uni-leipzig.de

Abstract

We study the expressive power of graph neural networks (GNNs) with mean as the aggregation function, with the following results. In the non-uniform setting, such GNNs have exactly the same expressive power as ratio modal logic, which has modal operators expressing that at least a certain ratio of the successors of a vertex satisfies a specified property. In the uniform setting, the expressive power relative to MSO is exactly that of modal logic, and thus identical to the (absolute) expressive power of GNNs with max aggregation. The proof, however, depends on constructions that are not satisfactory from a practical perspective. This leads us to making the natural assumptions that combination functions are continuous and classification functions are thresholds. The resulting class of GNNs with mean aggregation turns out to be much less expressive: relative to MSO and in the uniform setting, it has the same expressive power as alternation-free modal logic. This is in contrast to the expressive power of GNNs with max and sum aggregation, which is not affected by these assumptions.

	Non-Uniform	Uniform wrt. MSO	Uniform absolute
Mean^{c,t}	RML Th. 2, 3, 5	AFML Th. 13, 14, 15	$>$ AFML Th. 16
Mean	RML Th. 2, 3, 5	ML Cor. 3, Th. 10, 11	$>$ ML Th. 16
Sum	GML Th. 3, 5	GML†	$>$ GML‡
Max	ML Th. 4, Th. 6	ML Th. 4, Th. 6	ML Th. 4, Th. 6

Table 1: Overview of results, with $\cdot^{\dagger}$ from (Barceló et al. 2020) and $\cdot^{\ddagger}$ from (Benedikt et al. 2024).

1 Introduction

Graph neural networks (GNNs) are a family of deep-learning architectures that act directly on graphs, removing the need for prior serialization or encoding, in this way ensuring isomorphism invariance (Scarselli et al. 2009; Wu et al. 2021; Zhou et al. 2020). GNNs have been applied successfully in domains ranging from molecular property prediction and drug discovery (Bongini, Bianchini, and Scarselli 2021) to fraud detection (Deng and Hooi 2021) and e-commerce recommendation (Wu et al. 2022), traffic forecasting (Jiang and Luo 2022), and physics simulation (Shlomi, Battaglia, and Vlimant 2020). Numerous GNN variants exist, all sharing central features such as message passing and the iterative update of vertex embeddings, and an important foundational question is which properties a given GNN architecture can represent. Beginning with Barceló et al. (2020) and Grohe (2021), an expanding body of work has addressed this question by associating the expressive power of GNNs with that of various logical formalisms. Such logical characterizations may be used to guide the choice of a GNN model for a specific task at hand and to reveal potential gaps between a model's representational capacity and the task's requirements.

Barceló et al. (2020) focus on GNNs with constant iteration depth, sum as the aggregation function, and truncated ReLU as the activation function. One main result is that, relative to first-order logic (FO), the expressive power of GNNs as a vertex classifier is exactly that of graded modal logic (GML). In other articles, such as (Tena Cucala et al. 2023; Tena Cucala and Cuenca Grau 2024), sum aggregation is replaced with max aggregation, and GML is replaced with certain restrictions of non-recursive datalog. The main purpose of the current paper is to characterize the expressive power of GNNs as a vertex classifier when the aggregation function is arithmetic mean. This is in fact a very natural and important case. Influential graph learning systems such as GraphSAGE (Hamilton, Ying, and Leskovec 2017) and Pinterest's web-scale recommender PinSAGE (Ying et al. 2018) use mean aggregation, and the popular Graph Convolutional Networks (GCNs) use a weighted version of mean (Kipf and Welling 2017). Moreover, mean aggregation is amenable to random sampling and thus a good choice for graphs in which vertices may have very large degree. Our characterizations are in terms of suitable versions of modal logic. We also provide several new observations regarding GNNs with max and sum aggregation.

We consider the expressive power of GNNs in two different settings that have both received attention in the literature. In the uniform setting, a GNN model $\mathcal{M}$ has the same expressive power as a logic $\mathcal{L}$ if $\mathcal{M}$ and $\mathcal{L}$ can express exactly the same vertex classifiers, across all graphs. In the non-uniform setting, we only require that for every $n \geq 1$, $\mathcal{M}$ and $\mathcal{L}$ can express exactly the same vertex classifiers across all graphs with n vertices. In the uniform case, we follow

Barceló et al. (2020) in studying the expressive power relative to FO, that is, we restrict our attention to GNNs that express a vertex property definable by an FO formula. In fact we use monadic second-order logic (MSO) in place of FO, based on the observation from (Ahvonen et al. 2024) that every GNN that expresses an MSO-definable property actually expresses an FO-definable property. In the non-uniform case, we seek absolute characterizations that are independent of any background logic.

In the non-uniform case, we prove that GNNs with mean aggregation have the same expressive power as ratio modal logic (RML) which provides modal operators $\Diamond^{\geq r}\varphi$ and $\Diamond^{>r}\varphi$ expressing that the fraction of successors that satisfy φ is at least r (resp. exceeds r). This should be contrasted with GNNs based on sum aggregation and max aggregation, which have the same expressive power as GML and modal logic (ML), respectively. While these latter results are not surprising given existing work, we are not aware that they have been proved anywhere in this form, and we provide proof details here. All results are summarized in Table 1. The expressive equivalence between Max-GNNs and ML even holds in the uniform case. In the non-uniform setting, GML is strictly more expressive than RML, which in turn is strictly more expressive than ML. Our results thus reflect the known fact that Sum-GNNs are (non-uniformly) strictly more expressive than Mean-GNNs, which are in turn strictly more expressive than Max-GNNs (Xu et al. 2019).

The uniform setting turns out to be significantly more subtle and interesting. The fragment of RML that is expressible in MSO is exactly ML. Given our results from the non-uniform case, one may thus expect that in the uniform case and relative to MSO, Mean-GNNs have the same expressive power as ML. We prove that this is indeed the case. It follows that (uniformly and relative to MSO) Mean-GNNs have the same expressive power as Max-GNNs, but are strictly less expressive than Sum-GNNs. However, the translation of ML formulas into Mean-GNNs is unsatisfactory from a practical standpoint: one may either use a combination function that is not differentiable, and in fact not even continuous, or a rather unnatural classification function (in our translation, the rational numbers are classified as 0 and the irrational numbers as 1). This leads us to study Mean-GNNs under the natural assumptions that (i) the combination functions are continuous and (ii) the classification function is a threshold function. This case is denoted Mean^{c,t} in Table 1.

We prove that assumptions (i) and (ii) result in a significant drop of expressive power: relative to MSO, GNNs with mean aggregation now have the same expressive power as *alternation-free* modal logic (AFML) in which modal diamonds and boxes cannot be mixed. We believe that from a practical perspective, AFML provides a more realistic characterization of the expressive power of Mean-GNNs than ML (relative to MSO). It is also interesting to note that assumptions (i) and (ii) have no impact on expressive power in the cases of sum aggregation and max aggregation.

Throughout the paper, we also consider a 'simple' version of GNNs in which the combination function is a feedforward neural network without hidden layers. We then pay special attention to the activation function. Notably, we prove that the non-uniform results for mean and sum aggregation stated above hold for all continuous non-polynomial activation functions. We also prove that our result relating Mean-GNNs and AFML in the uniform case also holds for ReLU, truncated ReLU, and sigmoid activation. This is in contrast to Sum-GNNs where transitioning from (truncated) ReLU to sigmoid reduces the expressive power in the uniform case (Khalife and Tonelli-Cueto 2025).

This is the extended version of (Schönherr and Lutz 2026).

Related Work. We start with the uniform setting. The link between GNNs and modal logic was established in (Barceló et al. 2020), relative to FO. The expressive power of Sum-GNNs beyond FO was studied in (Benedikt et al. 2024), and linked to modal logic with Presburger quantifiers. In (Tena Cucala and Cuenca Grau 2024), Max-GNNs are translated into non-recursive datalog programs with negation-as-failure that adhere to a certain tree-shape. This formalism is closely related to modal logic. Recently, GNNs with transformer layers have been characterized in terms of modal logic (Ahvonen et al. 2026). Such layers are closely related to mean aggregation. Sound logical explanations for a certain kind of monotone Mean-GNN have been studied in (Morris and Horrocks 2025). In the non-uniform setting, Grohe (2024) shows that GNNs have the same expressive power as an extension of GFO+C, the guarded fragment of FO with counting capabilities, with built-in relations. Some related results are in (Grohe and Rosenbluth 2024). Without reference to logic, the expressive power of different aggregation functions is studied in (Xu et al. 2019; Rosenbluth, Toenshoff, and Grohe 2023).

2 Preliminaries

We use $\mathbb{N}$ and $\mathbb{N}^+$ to denote the set of all non-negative and positive integers, respectively. For $n \geq 1$, we write $[n]$ for the set $\{1, \ldots, n\}$. With $\mathcal{M}(X)$ we mean the set of all finite multisets over the set X, that is, the set of functions $X \to \mathbb{N}$ where all but finitely many elements of X are mapped to 0. For a vector $\bar{x} \in \mathbb{R}^\delta$ we use $x_1, \ldots, x_\delta$ or, when more readable, $(\bar{x})_1, \ldots, (\bar{x})_\delta$ to refer to its components.

Graph Neural Networks. Let $\Pi = \{P_1, \ldots, P_n\}$ be a finite set of *vertex labels*. A *(Π-labeled directed) graph* is a tuple $G = (V, E, \pi)$ that consists of a set of vertices V, a set of edges $E \subseteq V \times V$, and a vertex labeling function $\pi : V \to 2^\Pi$. Unless noted otherwise, all graphs in this article are finite. For easier reference, we may write V^G for V, E^G for E, and π^G for π. The *neighborhood* of a vertex v in a graph G is the set of its successors, formally $\mathcal{N}(G, v) = \{u \mid (v, u) \in E\}$. A *pointed graph* is a pair (G, v) with G a graph and $v \in V^G$ a distinguished vertex. A *(vertex) property* is a class of pointed graphs, over a common set of labels Π, that is closed under isomorphism.

A *graph neural network (GNN)* on Π-labeled graphs is a tuple

$$\mathcal{G} = (L, \{\mathsf{AGG}^\ell\}_{\ell\in[L]}, \{\mathsf{COM}^\ell\}_{\ell\in[L]}, \mathsf{CLS})$$

where $L \geq 1$ is the number of layers and for each $\ell \in [L]$, $\mathsf{AGG}^\ell : \mathcal{M}(\mathbb{R}^{\delta^{\ell-1}}) \to \mathbb{R}^{\delta^{\ell-1}}$ is an *aggregation function*, $\mathsf{COM}^\ell : \mathbb{R}^{\delta^{\ell-1}} \times \mathbb{R}^{\delta^{\ell-1}} \to \mathbb{R}^{\delta^\ell}$ is a *combination function*, and $\mathsf{CLS} : \mathbb{R}^{\delta^L} \to \{0,1\}$ is a *classification function*. We call $\delta^{\ell-1}$ the *input dimension* of layer ℓ and δ^ℓ the *output dimension*, with the input dimension δ^0 of Layer 1 being at least $|\Pi|$. Typical aggregation functions are sum, max, and mean, applied component-wise. We only consider GNNs in which every layer uses the same aggregation function. If we want to highlight the aggregation function, we may speak of a Sum-GNN, a Max-GNN, or a Mean-GNN.

We now make precise the semantics of GNNs. For $1 \leq \ell \leq L$, the ℓ-th layer assigns to each Π-labeled graph G and vertex $v \in V^G$ a feature vector $\bar{x}^\ell_{G,v} \in \mathbb{R}^{\delta^\ell}$. The *initial feature vector* $\bar{x}^0_{G,v} \in \mathbb{R}^{\delta^0}$ of vertex v in graph G is defined as follows: for all $i \in [|\Pi|]$, the i-th value of $\bar{x}^0_{G,v}$ is 1 if $P_i \in \pi^G(v)$, and all other values are 0. The feature vector $\bar{x}^\ell_{G,v}$ assigned by the ℓ-th layer is

$$\bar{x}^\ell_{G,v} = \mathsf{COM}^\ell(\bar{x}^{\ell-1}_{G,v}, \mathsf{AGG}^\ell(\{\{\bar{x}^{\ell-1}_{G,u} \mid u \in \mathcal{N}(v)\}\})) \quad (*)$$

where $\{\{\cdots\}\}$ denotes multiset; we define the mean of the empty multiset to be 0. We write $\bar{x}^\ell_v$ instead of $\bar{x}^\ell_{G,v}$ if the graph G is clear from the context. The output of the last layer is then passed to the classification function. The property *defined* by a GNN $\mathcal{G}$ is the set of pointed Π-labeled graphs

$$\{(G,v) \mid \mathsf{CLS}(\bar{x}^L_{G,v}) = 1\}.$$

We also say that $\mathcal{G}$ *accepts* the graphs in this set.

In a *simple* GNN, as considered for example in (Barceló et al. 2020; Ahvonen et al. 2024; Tena Cucala et al. 2023), the combination function is restricted to the form

$$\mathsf{COM}(\bar{x}_v, \bar{x}_a) = f(\bar{x}_v \cdot C + \bar{x}_a \cdot A + \bar{b}),$$

where $f : \mathbb{R} \to \mathbb{R}$ is a (typically non-linear) *activation function* applied component-wise, $A, C \in \mathbb{R}^{\delta^{\ell-1} \times \delta^\ell}$ are matrices, $\bar{b} \in \mathbb{R}^{\delta^\ell}$ is a bias vector, and $\bar{x}_a$ is the result of the aggregation function, applied as in $(*)$. Relevant choices for the activation function f include the *truncated ReLU* $\mathsf{ReLU}^*(x) = \min(\max(0,x),1)$, the *ReLU* $\mathsf{ReLU}(x) = \max(0,x)$ and the *sigmoid function* $\sigma(x) = \frac{1}{1+e^{-x}}$. The following is an easy consequence of the observation that $\mathsf{ReLU}^*(x) = \mathsf{ReLU}(x) - \mathsf{ReLU}(x-1)$ for all $x \in \mathbb{R}$.

Lemma 1. *Let* $\mathsf{AGG} \in \{\mathsf{Max}, \mathsf{Sum}, \mathsf{Mean}\}$. *If a property is definable by a* simple AGG-*GNN with* ReLU^* *activation, then it is definable by a* simple AGG-*GNN with* ReLU *activation.*

Modal Logic Formulas of *modal logic (ML)* over a set of vertex labels Π are defined by the grammar rule

$$\varphi ::= P \mid \neg\varphi \mid \varphi \vee \varphi \mid \Diamond\varphi,$$

where P ranges over Π (Blackburn, de Rijke, and Venema 2001). As usual, we use $\varphi \wedge \psi$ as abbreviation for $\neg(\neg\varphi \vee \neg\psi)$, $\Box\varphi$ as abbreviation for $\neg\Diamond\neg\varphi$, $\top$ as abbreviation for $P \vee \neg P$ with $P \in \Pi$ chosen arbitrarily, and $\bot$ as abbreviation for $\neg\top$. Satisfaction of a formula φ by a vertex v in a Π-labeled graph G is defined inductively as follows:

$$\begin{array}{lcll}
G,v & \models & P & \text{if } P \in \pi(v) \\
G,v & \models & \neg\varphi & \text{if not } G,v \models \varphi \\
G,v & \models & \varphi \vee \psi & \text{if } G,v \models \varphi \text{ or } G,v \models \psi \\
G,v & \models & \Diamond\varphi & \text{if } G,u \models \varphi \text{ for some } u \in \mathcal{N}(G,v).
\end{array}$$

Graded modal logic (GML) is a well-known extension of ML in which the diamond is replaced with a counting version $\Diamond^{\geq n}$, $n \in \mathbb{N}$, where

$$G,v \models \Diamond^{\geq n}\varphi \text{ if } |\{u \in \mathcal{N}(G,v) \mid G,u \models \varphi\}| \geq n.$$

As a shortcut, we may use $\Diamond^{=k}\psi$ to mean $\Diamond^{\geq k}\psi \wedge \neg\Diamond^{\geq k+1}\psi$. For more details on GML, see for instance (Goble 1970; De Rijke 2000).

We next introduce *ratio modal logic (RML)* in which the standard ML diamond is also replaced with a counting version, but here the counting is relative rather than absolute. Diamonds take the form $\Diamond^{\geq r}$ and $\Diamond^{> r}$ with $r \in [0,1]$. We have $G,v \models \Diamond^{\geq r}\varphi$ (resp. $G,v \models \Diamond^{>r}$) if the fraction of successors of v that satisfy φ is at least r (resp. exceeds r). If a vertex v has no successors, then we define $G,v \models \Diamond^{\geq t}\varphi$ and $G,v \not\models \Diamond^{>t}\varphi$ for all φ. As a useful shortcut, we may write $\Diamond^{=f_i}$ to mean $\Diamond^{\geq f_i}\varphi \wedge \neg \Diamond^{>f_i} \varphi$. Modal operators of this kind have occasionally been considered in the literature, see for instance (Pacuit and Salame 2004). RML is a fragment of modal logic with Presburger constraints (Demri and Lugiez 2010). We remark that while ML and GML are fragments of first-order logic (FO), the diamond operators of RML cannot even be expressed in MSO.

The *modal depth* of a modal formula φ, no matter whether GML, RML, or ML, is the nesting depth of diamonds in φ.

Notions of Expressive Power. Both GNNs and modal logic formulas may be viewed as vertex classifiers. We say that a property P is *(uniformly) expressible* by a class of vertex classifiers $\mathcal{C}$, such as $\mathcal{C} =$ GML, if there is a $C \in \mathcal{C}$ such that the pointed graphs accepted by C are exactly those in P. For classes of classifiers $\mathcal{C}_1, \mathcal{C}_2$, we write $\mathcal{C}_1 \subseteq \mathcal{C}_2$ to mean that every property expressible by a classifier from $\mathcal{C}_1$ is also expressible by a classifier from $\mathcal{C}_2$. We then also say that $\mathcal{C}_2$ is *at least as expressive* as $\mathcal{C}_1$. This notion of expressive power is commonly referred to as the *uniform setting*.

This is in contrast to the *non-uniform setting* where a property P is *(non-uniformly) expressible* by a class of classifiers $\mathcal{C}$ if for every graph size $n \geq 0$, there is a $C \in \mathcal{C}$ such that the pointed graphs of size n accepted by C are exactly those in P. Also in this setting, we may write $\mathcal{C}_1 \subseteq \mathcal{C}_2$ to mean that every property (non-uniformly) expressible by a classifier from $\mathcal{C}_1$ is also (non-uniformly) expressible by a classifier from $\mathcal{C}_2$. We shall always make clear whether we refer to the uniform or non-uniform setting.

In both the uniform and the non-uniform setting, we may write $\mathcal{C}_1 = \mathcal{C}_2$ as an abbreviation for $\mathcal{C}_1 \subseteq \mathcal{C}_2 \subseteq \mathcal{C}_1$. The following is clear from the definitions.

Lemma 2. *Let $\mathcal{C}_1$ and $\mathcal{C}_2$ be two classes of classifiers. If $\mathcal{C}_1 \subseteq \mathcal{C}_2$ in the uniform setting, then $\mathcal{C}_1 \subseteq \mathcal{C}_2$ in the non-uniform setting.*

We now clarify the relative expressive power of the modal logics introduced above, both in the uniform and in the non-uniform setting.

Lemma 3. *In the uniform setting,*

1. ML $\subsetneq$ RML *and* ML $\subsetneq$ GML*;*
2. RML $\not\subseteq$ GML *and* GML $\not\subseteq$ RML.

In the non-uniform setting, ML $\subsetneq$ RML $\subsetneq$ GML.

3 Non-Uniform Setting

We give logical characterizations of GNNs in the non-uniform setting. The characterizations are absolute, that is, they are not relative to FO, MSO, or any other background logic. We start with a summary of all results in this section.

Theorem 1. *In the non-uniform setting,*

1. Mean-GNN $\subseteq$ RML $\subseteq$ simple Mean-GNN
2. Sum-GNN $\subseteq$ GML $\subseteq$ simple Sum-GNN
3. Max-GNN $\subseteq$ ML $\subseteq$ simple Max-GNN.

This holds for truncated ReLU and ReLU, and for Points 1 and 2 for every continuous non-polynomial activation function.

Point 3 of Theorem 1 even holds in the uniform setting, also there without relativization to a background logic. The same is true for the second inclusion in Point 2 as per (Barceló et al. 2020), but as we shall see, not for any of the other inclusions. Together with Lemma 3, Theorem 1 also reproves the following result from (Xu et al. 2019).

Corollary 1. *In the non-uniform setting,*

$$\text{Max-GNN} \subsetneq \text{Mean-GNN} \subsetneq \text{Sum-GNN}.$$

To prove Theorem 1, we first give the translations from logic to GNNs, starting with Point 1. The general strategy is as in (Barceló et al. 2020), that is, we translate a ratio modal logic (RML) formula φ with L subformulas into a simple GNN with L layers, each of output dimension L. We first observe that we may assume w.l.o.g. that φ contains no RML diamonds of the form $\Diamond^{\geq t}$. This is because on graphs with at most n vertices such a diamond can be replaced by $\Diamond^{>t'}$ where t' is the largest rational number that is smaller than t and can occur as a fraction in a graph where every vertex has at most n successors.

For the GNN translation, we choose a suitable enumeration $\varphi_1, \ldots, \varphi_L$ of the subformulas of φ and design the GNN to compute the truth value (that is, 0 or 1) of each subformula φ_i at every vertex v and store it in the i-th component of the feature vector for v. The only interesting case are subformulas of the form $\Diamond^{>t}\varphi_i$. If such a diamond is satisfied at a node v in a graph G with at most n vertices, then at least a fraction of

$$t' = \min\left\{\frac{\ell}{m} \mid 0 \leq \ell \leq m \leq n, \frac{\ell}{m} > t\right\}$$

successors of v must satisfy φ_i. Note that t' is the smallest rational number that is larger than t and can occur as a fraction in a graph where every vertex has at most n successors. If $\Diamond^{>t}\varphi_i$ is violated, then at most a fraction of t'' successors of v can satisfy φ_i where t'' is defined like t' except that $\min$ is now $\max$ and '$>$' is '$\leq$'. The gap between these fractions is at least $\frac{1}{n^2}$ and can be amplified using the matrix A from simple combination functions, which through the bias vector allows us to compute the desired truth value.

Theorem 2. RML $\subseteq$ simple Mean-GNN *in the non-uniform setting. This holds for truncated ReLU and ReLU activation.*

We next strengthen Theorem 2 to all continuous non-polynomial activation functions. This is based on universal approximation theorems from machine learning (Pinkus 1999). Intuitively, approximation suffices because the constant bound on the size of graphs imposed in the non-uniform setting ensures that a GNN can generate only a constant number of different feature vectors, across all input graphs, and we are good as long as we can distinguish these. A variation of the proof also works for sum aggregation, delivering the second inclusion in Point 2 of Theorem 1.

Theorem 3. *In the non-uniform setting and for all continuous non-polynomial activation functions:*

1. RML $\subseteq$ simple Mean-GNN*;*
2. GML $\subseteq$ simple Sum-GNN.

Point 2 was proved in (Barceló et al. 2020) in the uniform setting, but only for the special case of truncated ReLU activation.

We next treat the second inclusion in Point 3 of Theorem 1, using a minor variation of the proof in (Barceló et al. 2020). Like that proof, our proof even works in the uniform setting.

Theorem 4. ML $\subseteq$ simple Max-GNN *in the uniform setting. This holds for truncated ReLU and ReLU activation.*

We do not know whether this can be strengthened to all continuous non-polynomial activation functions.

We now turn to the translations from GNNs to logic. These also rely on the fact that in the non-uniform setting, a GNN can generate only a constant number of different feature vectors, across all input graphs. For every feature vector $\bar{x}$ and every layer ℓ of the GNN, we can construct a modal logic formula $\varphi^{\ell}_{\bar{x}}$ such that the GNN assigns $\bar{x}$ to a vertex v in layer ℓ if and only if $G, v \models \varphi^{\ell}_{\bar{x}}$. Depending on the aggregation function of the GNN, the formula $\varphi^{\ell}_{\bar{x}}$ needs to describe the distribution of feature vectors at the successors of v computed by level $\ell - 1$ in varying degrees of detail. For mean, we only need to know the fraction of successors at which each feature vector was computed, and thus the formula can be formulated in RML. For sum, we need exact multiplicities and thus require a GML formula.

Theorem 5. *In the non-uniform setting,*

1. Mean-GNN $\subseteq$ RML*;*
2. Sum-GNN $\subseteq$ GML.

As observed already in (Tena Cucala and Cuenca Grau 2024), with max aggregation the set of feature vectors ever generated by a GNN, across all input graphs, is finite even without bounding the graph size. Using similar arguments as for Theorem 5, we can thus show the following.

Theorem 6. *In the uniform setting,* Max-GNN $\subseteq$ ML.

We remark that all our results obtained in the non-uniform setting also hold in the uniform setting when a constant bound is imposed on the outdegree of vertices.

4 Uniform Setting

We again start with a summary of our results. Note that these are now relative to MSO, except for the case of Max-GNNs.

Theorem 7. *In the uniform setting,*

1. Mean-GNN $\cap$ MSO $\subseteq$ ML $\subseteq$ simple Mean-GNN
2. Sum-GNN $\cap$ MSO $\subseteq$ GML $\subseteq$ simple Sum-GNN

3. Max-GNN $\subseteq$ ML $\subseteq$ simple Max-GNN.

This holds for truncated ReLU and ReLU activation.

Also note that, compared to Theorem 1, the logic associated with Mean-GNNs has changed from RML to ML, because RML diamonds are not expressible in MSO. With Lemma 3, we obtain the following corollary.

Corollary 2. *In the uniform setting,*

$$\text{Max-GNN} = \text{Mean-GNN} \cap \text{MSO} \subsetneq \text{Sum-GNN} \cap \text{MSO}.$$

We now prove Theorem 7. We have already shown Point 3 as Theorems 4 and 6. The result in Point 2 is from (Barceló et al. 2020), stated there for truncated ReLU and for FO in place of MSO. We may invoke Lemma 1 for ReLU. Regarding the replacement of FO with MSO, it was observed in (Ahvonen et al. 2024) that the following is a consequence of results by (Elberfeld, Grohe, and Tantau 2016).

Lemma 4. *Any property expressible in MSO and by a GNN is also FO-expressible. This only depends on invariance under unraveling and thus holds for all choices of aggregation, activation, and classification function.*

In the remainder of this section, we prove Point 1 of Theorem 7. For the first inclusion, we need suitable versions of Ehrenfeucht-Fraïssé (EF) games. Such games are played by two players, *Spoiler (S)* and *Duplicator (D)*, who play on two potentially infinite pointed graphs $(G_1, v_1), (G_2, v_2)$. Spoiler's aim is to show that the graphs are dissimilar while D wishes to show that they are similar. The game is played in rounds. In the *GML game*, which in addition is parameterized by a *number of rounds* $\ell \in \mathbb{N}$ and a *counting bound* $c \in \mathbb{N}^+$, each round consists of the following steps (Otto 2019):

1. S chooses $i \in \{1, 2\}$ and a set $U_i \subseteq \mathcal{N}(G_i, v_i)$ with $0 < |U_i| \leq c$;
2. D selects a set $U_{3-i} \subseteq \mathcal{N}(G_{3-i}, v_{3-i})$ with $|U_1| = |U_2|$;
3. S selects a vertex $u_{3-i} \in U_{3-i}$;
4. D selects a vertex $u_i \in U_i$.

The game proceeds on the graphs (G_1, u_1) and (G_2, u_2). Spoiler wins as soon as one of the following conditions hold, possibly at the very beginning of the game:

- $\pi^{G_1}(v_1) \neq \pi^{G_2}(v_2)$;
- D fails in Step 2 because $|U_i| > |\mathcal{N}(G_{3-i}, v_{3-i})|$.

Duplicator wins if one of the following conditions hold:

- S cannot choose a non-empty set U_i in Step 1,
- after ℓ rounds, S has not won.

We write $\mathcal{E}_\ell^{\mathrm{GML}[c]}(G_1, v_1, G_2, v_2)$ to denote the ℓ-round GML game with counting bound c on pointed graphs (G_1, v_1) and (G_2, v_2). We may vary GML games to obtain games for ML. An *ML game* is a GML game with counting bound 1. Thus Spoiler selects a singleton set U_i in Step 1 and consequently Steps 3 and 4 are trivialized. In other words, a round consists of first S choosing $i \in \{1, 2\}$ and a vertex $u_i \in \mathcal{N}(G_i, v_i)$, and D replying with a vertex $u_{3-i} \in \mathcal{N}(G_{3-i}, v_{3-i})$. We denote these games with $\mathcal{E}_\ell^{\mathrm{ML}}(G_1, v_1, G_2, v_2)$.

We use GML$[c]$ to denote the fragment of GML in which in all diamonds $\Diamond^{\geq n}$ we have $n \leq c$.

Theorem 8. *Let $\mathcal{L} \in \{\mathrm{ML}\} \cup \{\mathrm{GML}[c] \mid c \geq 0\}$, and let P be a vertex property. The following are equivalent for all $\ell \geq 0$:*

1. *there exists an $\mathcal{L}$ formula φ of modal depth at most ℓ such that for all pointed graphs (G, v): $G, v \models \varphi$ if and only if $(G, v) \in P$.*
2. *Spoiler has a winning strategy in $\mathcal{E}_\ell^{\mathcal{L}}(G_1, v_1, G_2, v_2)$ for all pointed graphs $(G_1, v_1), (G_2, v_2)$ with $(G_1, v_1) \in P$ and $(G_2, v_2) \notin P$.*

Proofs can be found in the literature, see for instance (Otto 2019) and Chapter 3.2 in (Goranko and Otto 2007). We next recall that the following was proved in (Barceló et al. 2020), not specifically for Mean-GNNs, but in fact independently of the aggregation function used.

Theorem 9. Mean-GNN $\cap$ MSO $\subseteq$ GML *in the uniform setting.*

We improve this from GML to ML, exploiting the fact that properties definable by Mean-GNNs are invariant under scaling the graph, that is, choosing a $c \geq 1$ and multiplying each vertex in the graph exactly c times. To make this formal, let $G = (V, E, \pi)$ be a Π-labeled graph and $c \geq 1$. The *c-scaling of G* is the Π-labeled graph $c \cdot G := (V', E', \pi')$ where $V' = \{(v, i) \mid v \in V,\ 1 \leq i \leq c\}$, $E' = \{((v, i), (u, j)) \mid (v, u) \in E\}$, and $\pi'((v, i)) = \pi(v)$ for all $v \in V$ and $i \in [c]$. The following is immediate from the fact that the mean of a multiset is invariant under multiplying all multiplicities by a constant $c \geq 1$.

Lemma 5. *Let $\mathcal{G}$ be a* Mean-GNN *on Π-labeled graphs, (G, v) a Π-labeled pointed graph, and $c \geq 1$. Then for all $i \in [c]$: $\bar{x}_{G,v}^L = \bar{x}_{c \cdot G, (v,i)}^L$.*

The following relates EF-games for ML to EF-games for GML$[c]$ on the corresponding c-scaled graphs.

Lemma 6. *Let $(G_1, v_1),\ (G_2, v_2)$ be pointed graphs and $\ell \geq 0$. If D has a winning strategy in $\mathcal{E}_\ell^{\mathrm{ML}}(G_1, v_1, G_2, v_2)$, then D also has a winning strategy in*

$$\mathcal{E}_\ell^{\mathrm{GML}[c]}(c \cdot G_1, (v_1, k_1), c \cdot G_2, (v_2, k_2)),$$

for all $c \geq 1$ and $k_1, k_2 \in [c]$.

With Lemma 6 as the main ingredient, we can now show the following.

Corollary 3. Mean-GNN $\cap$ MSO $\subseteq$ ML *in the uniform setting.*

Proof. Let P be a vertex property that is expressible by a Mean-GNN and by an MSO formula. By Theorem 9, P is definable by a GML formula φ. Let c be maximal such that φ contains a diamond $\Diamond^{\geq c}$.

Assume to the contrary of what we have to show that P cannot be expressed in modal logic (ML). Then by Theorem 8 for each $\ell \geq 0$ there exist pointed graphs $(G_1, v_1) \in P$ and $(G_2, v_2) \notin P$ such that Duplicator wins $\mathcal{E}_\ell^{\mathrm{ML}}(G_1, v_1, G_2, v_2)$. By Lemma 6, Duplicator also wins $\mathcal{E}_\ell^{\mathrm{GML}[c]}(c \cdot G_1, (v_1, 1), c \cdot G_2, (v_2, 1))$. Since P is definable by a Mean-GNN, Lemma 5 yields $(c \cdot G_1, (v_1, 1)) \in P$ and $(c \cdot G_2, (v_2, 1)) \notin P$. Therefore, again by Theorem 8, P cannot be defined by a GML$[c]$ formula; a contradiction. ❑

Regarding the second inclusion of Point 1 of Theorem 7, we actually prove something stronger.

Theorem 10. RML $\subseteq$ simple Mean-GNN *in the uniform setting.*

The proof is similar to that of Theorem 2. In particular, truth and falsity of formulas is represented by the values 1 and 0 in feature vectors. Importantly, we use a step function as the activation function. The reason why we cannot use a continuous activation function such as ReLU in our translation is the modal diamond $\Diamond\varphi$: if a vertex has a successor that satisfies φ, then the mean over all successors may still be arbitrarily close to 0, inducing a discontinuity.

5 Uniform Setting, Reloaded

From a practical perspective, the use of a non-continuous activation function (resulting in a non-continuous combination function) in the proof of Theorem 10 is unsatisfactory. The combination function of a GNN is often represented as a feed-forward neural network (FNN) with a continuous activation function such as truncated ReLU, ReLU, or sigmoid, and is then guaranteed to be continuous. Importantly, the use of a non-continuous and thus non-differentiable combination function precludes a direct use of backpropagation, gradient descent, and related methods. It is thus natural to ask whether the translation of ML to Mean-GNNs can also be realized using a continuous combination function. The answer turns out to be positive. We use Mean^c-GNN to denote the class of GNNs in which the combination functions COM^ℓ, viewed as functions $\mathsf{COM}^\ell : \mathbb{R}^{2\delta^{\ell-1}} \to \mathbb{R}^{\delta^\ell}$, are continuous.

Theorem 11. ML $\subseteq$ Mean^c-GNN *in the uniform setting.*

From a practical perspective, however, the construction in the proof of Theorem 11 is even more unsatisfactory. It uses a combination function that is continuous, but very artificial: the function is inspired by and derived from the proof of Cantor's isomorphism theorem. Moreover, there is a price to pay in terms of a rather unnatural classification function. We represent truth and falsity of logical formulas as irrational and rational numbers, and consequently the classification function has to return 1 for all irrational numbers and 0 for all rational ones. We conjecture that our proof can be improved to yield simple Mean^c-GNNs, at the expense of making it more technical.

In practice, classification functions are often threshold functions.[1] It is therefore relevant to ask about the expressive power of Mean-GNNs that use a continuous combination function and a threshold classification function. We denote this class with Mean^{c,t}-GNN. To be more precise, the classification function is required to be of the form

$$\mathsf{CLS}(\bar{x}) = \begin{cases} 1 & \text{if } x_i \sim c, \\ 0 & \text{otherwise} \end{cases}$$

[1] Or they are represented by a feed-forward neural network (FNN) to which a threshold is applied. Our model captures this case because we can include the FNN in the COM function of the final GNN layer.

where $i \in [\delta^L]$, $\sim \in \{\geq, >\}$, and $c \in \mathbb{R}$. Note that adding the options $\sim \in \{\leq, <\}$ is syntactic sugar because we can replace the last combination function COM^L with $(-1)\cdot\mathsf{COM}^L$ and compare x_i against $-c$ in the classification function. We remark that all translations from logic to GNN given in this paper use threshold classification, except Theorem 11.

We characterize the expressive power of Mean^{c,t}-GNNs, relative to MSO, in terms of an alternation-free fragment of ML. Formally, *alternation-free modal logic (AFML)* is defined by the grammar rule

$$\begin{array}{lll} \varphi & ::= & \psi \mid \vartheta \\ \psi & ::= & P \mid \neg P \mid \Box\bot \mid \psi \wedge \psi \mid \psi \vee \psi \mid \Diamond\psi \\ \vartheta & ::= & P \mid \neg P \mid \Diamond\top \mid \vartheta \wedge \vartheta \mid \vartheta \vee \vartheta \mid \Box\vartheta. \end{array}$$

We use AFML[1] to denote all AFML formulas formed according to the grammar rule for ψ in the definition of AFML, and likewise for AFML[2] and the grammar rule for ϑ. Our main result is as follows.

Theorem 12. *In the uniform setting,*

$$\text{Mean}^{c,t}\text{-GNN} \cap \text{MSO} \subseteq \text{AFML} \subseteq \text{simple Mean}^{c,t}\text{-GNN}.$$

This holds for truncated ReLU, ReLU, and sigmoid activation.

The proof of the first inclusion in Theorem 12 relies on EF games for AFML. For $k \in \{1, 2\}$, an AFML$[k]$ *game* is an ML game subject to the modification that Spoiler chooses the same value $i = k$ in the first step of each round, except that S may choose $i = 3 - k$ in case that v_k has no successors. Note that in the latter case, Spoiler immediately wins because Duplicator cannot respond with a successor of v_k (we in fact have $G_k, v_k \models \Box\bot$ and $G_{3-k}, v_{3-k} \not\models \Box\bot$). We denote these games with $\mathcal{E}_\ell^{\mathsf{AFML}[k]}(G_1, v_1, G_2, v_2)$. A version of Theorem 8 for AFML[1] and AFML[2] is proved in the appendix.

Using AFML games, it is easy to prove that basic ML properties such as $\varphi = \Diamond P \wedge \Box Q$ are not expressible in AFML, that is, AFML $\subsetneq$ ML (both in the uniform and non-uniform setting). Details are in the appendix.

To prove Theorem 12, we start from Corollary 3. We need some preliminaries. Let $G = (V, E, \pi)$ be a graph and $v \in V$. A *path* in G is a sequence $p = v_0, \ldots, v_n$ of vertices from V such that $(v_i, v_{i+1}) \in E$ for all $i < n$. The path *starts at* v_0 and is of *length* n, and we use $\mathsf{tail}(p)$ to denote v_n. The *unraveling of G at v* is the potentially infinite tree-shaped graph $\mathsf{Unr}(G, v) = (V', E', \pi')$ defined as follows:

- V' is the set of all paths in G that start at v;
- E' contains an edge (p, pu) if $(\mathsf{tail}(p), u) \in E$;
- $\pi'(p) = \pi(\mathsf{tail}(p))$.

For $L \geq 0$, the *unraveling of G at v up to depth L*, denoted $\mathsf{Unr}^L(G, v)$, is the (finite) subgraph of $\mathsf{Unr}(G, v)$ induced by all paths of length at most L.

It is well-known that modal formulas are invariant under unraveling up to their modal depth. A similar statement holds for GNNs.

Lemma 7 (Barceló et al. (2020))**.** *Let G be a graph, $v \in V^G$, $\mathcal{G} = (L, \{\mathsf{AGG}^\ell\}_{\ell\in[L]}, \{\mathsf{COM}^\ell\}_{\ell\in[L]}, \mathsf{CLS})$, and $1 \leq \ell \leq L$. Then $\bar{x}^\ell_{G,v} = \bar{x}^\ell_{\mathsf{Unr}^L(G,v),v}$.*

We next show that slightly changing a highly scaled input graph to a Mean^{c,t}-GNN does not change the computed value in an unbounded way. We first formalize what we mean by 'slight change'.

Definition 1. *Let $G = (V, E, \pi)$ be a graph and $n \geq 0$. A graph $G' = (V', E', \pi')$ is an n-*extension *of G if it satisfies the following conditions for all $v \in V$:*

1. *$V \subseteq V'$, $E \subseteq E'$,*
2. *for all $v \in V$: $\pi(v) = \pi'(v)$*
3. *if $\mathcal{N}(G, v) \neq \emptyset$, then $|\mathcal{N}(G', v) \setminus \mathcal{N}(G, v)| \leq n$, and*
4. *if $\mathcal{N}(G, v) = \emptyset$, then $\mathcal{N}(G', v) = \emptyset$.*

Note that n-extensions can add up to n fresh successors to any vertex that already has at least one successor. Intuitively, the value of n will be small compared to the scaling of the graph G of which the n-extension is taken.

In what follows, we use the maximum metric and define the distance of two vectors $\bar{x}$ and $\bar{y}$ of dimension δ to be $||\bar{x} - \bar{y}||_\infty = \max_{1 \leq i \leq \delta} |x_i - y_i|$. The following makes precise why we are interested in n-extensions.

Lemma 8. *Let $\mathcal{G}$ be a* Mean^c-GNN *with L layers. Then for all $\varepsilon > 0$, $n \geq 1$, and $\ell \in [L]$, there exists a constant c such that for all $c' \geq c$, graphs $H = c' \cdot G$, vertices (v, i) in H, and n-extensions H' of H: $||\bar{x}^\ell_{H,(v,i)} - \bar{x}^\ell_{H',(v,i)}||_\infty < \varepsilon$.*

One ingredient to the proof of Lemma 8 is the observation that for every Mean^c-GNN, there are constant upper and lower bounds on the values that may occur in feature vectors, across all input graphs. The reader might want to compare this with Max-GNNs where even the number of such values is bounded by a constant, see the discussion before Theorem 6.

Theorem 13. Mean^{c,t}-GNN$\cap$MSO $\subseteq$ AFML *in the uniform setting.*

Proof. (sketch) Assume to the contrary that there exists a Mean^{c,t}-GNN $\mathcal{G}$ with L layers that is equivalent to an MSO formula, but not to an AFML formula. By Corollary 3, $\mathcal{G}$ is equivalent to an ML formula φ. Assume first that the classification function of $\mathcal{G}$ uses '$>$' rather than '$\geq$'.

Because φ is not expressible in AFML, for each $\ell \in \mathbb{N}$ there exist pointed graphs (G, v) and (G', v') with $G, v \models \varphi$ and $G', v' \not\models \varphi$ such that D has a winning strategy in $\mathcal{E}^{\mathsf{AFML}[1]}_\ell(G, v, G', v')$. Our aim is to transform (G, v) into a pointed graph (H, u) such that

(i) $H, u \models \varphi$ and

(ii) D has a winning strategy in $\mathcal{E}^{\mathsf{ML}}_\ell(H, u, G', v')$.

Then, by Theorem 8, φ is not expressible in ML. A contradiction.

It can be shown that since D has a winning strategy in $\mathcal{E}^{\mathsf{AFML}[1]}_\ell(G, v, G', v')$, they also have one in $\mathcal{E}^{\mathsf{AFML}[1]}_\ell(\mathsf{Unr}^K(G, v), v, G', v')$ for all $K \geq \ell$. Moreover, in AFML[1]-games in which the first graph is tree-shaped, the existence of a winning strategy for D implies the existence of a memoryless winning strategy. We may view such a strategy as a function $\mathsf{ws} : V^{\mathsf{Unr}^K(G,v)} \to V^{G'}$ such that if S plays vertex u in $\mathsf{Unr}^K(G, v)$, then D always answers with $\mathsf{ws}(u)$. We also set $\mathsf{ws}(v) = v'$.

Let $m = |V^{G'}|$ and $K = \max(\ell, L)$. Since classification is based on '$>$', we find an $\varepsilon > 0$ such that all pointed graphs (H, u) with $||\bar{x}^L_{G,v} - \bar{x}^L_{H,u}||_\infty < \varepsilon$ are accepted by $\mathcal{G}$. By Lemma 8, there exists a c such that $||\bar{x}^L_{G'',v} - \bar{x}^L_{X,v}||_\infty < \varepsilon$ in each m-extension X of $G'' = c \cdot \mathsf{Unr}^K(G, v)$. Now, (H, u) is defined as follows:

1. start with $G'' = c \cdot \mathsf{Unr}^K(G, v)$;
2. take the disjoint union with all $\mathsf{Unr}^K(G', v')$, $v' \in V^{G'}$;
3. for each vertex $(u, i) \in V^{G''}$ that has at least one successor, let $\mathcal{N}(G', \mathsf{ws}(u)) = \{u'_1, \ldots, u'_m\}$. Add to (u, i) the fresh successors $u'_1, \ldots, u'_m$;
4. $u = (v, 1)$.

We show in the appendix that Conditions (i) and (ii) are satisfied. We also deal there with the case where the classification function is based on '$\geq$', using AFML[2]. ❑

Next, we show that each formula in AFML can be realized by a simple Mean^{c,t}-GNN. As in the proof of Theorem 10, we face the challenge that the mean over all successors may diminish values. Here, we address this by representing the truth of subformulas by values from the range $(0, 1]$ and falsity by value 0. With this encoding, we can realize the modal diamonds of AFML[1] using truncated ReLU activation, but we cannot realize modal boxes. Still, it is easy to also treat AFML[2] using duality arguments.

Theorem 14. AFML $\subseteq$ simple Mean^{c,t}-GNN *in the uniform setting. This holds for truncated ReLU and ReLU as activation functions.*

The proof of Theorem 14 can be adapted to sigmoid activation. This once more requires non-trivial modifications. Truth is now encoded by values from the range $(\frac{1}{2}, 1)$ and falsity by value $\frac{1}{2}$.

Theorem 15. AFML $\subseteq$ simple Mean^{c,t}-GNN *in the uniform setting, with sigmoid as the activation function.*

6 Conclusion

We have identified logical characterizations of graph neural networks with mean aggregation, in several different settings. Some interesting questions remain open. For instance, we would like to know whether Theorem 4 can be strengthened to all continuous non-polynomial activation functions, in the spirit of Theorem 3. It would also be interesting to consider broader classes of activation functions in the uniform setting. Finally, it would be interesting to find absolute logical characterization of Mean-GNNs and Mean^{c,t}-GNNs, that is, characterizations that are not relative to any background logic such as FO or MSO. In fact, we can already provide two first observations on this subject.

Theorem 16.

1. *The property 'there exist more successors that satisfy P_1 than successors that satisfy P_2' is not expressible in RML, but by a simple* Mean^{c,t}-GNN.
2. *The RML formula $\Diamond^{>\frac{1}{2}} \Diamond^{>\frac{1}{2}} P$ is not expressible by a* Mean^{c,t}-GNN.

Acknowledgements

Carsten Lutz was supported by DFG project LU 1417/4-1.

References

Ahvonen, V.; Funk, M.; Heimann, D.; Kuusisto, A.; and Lutz, C. 2026. Expressive Power of Graph Transformers via Logic. In *Proc. of AAAI*. AAAI Press.

Ahvonen, V.; Heiman, D.; Kuusisto, A.; and Lutz, C. 2024. Logical characterizations of recurrent graph neural networks with reals and floats. In *Proc. of NeurIPS*, volume 37, 104205–104249. Curran Associates, Inc.

Barceló, P.; Kostylev, E. V.; Monet, M.; Pérez, J.; Reutter, J.; and Silva, J.-P. 2020. The Logical Expressiveness of Graph Neural Networks. In *Proc. of ICLR*.

Barceló, P.; Kostylev, E. V.; Monet, M.; Pérez, J.; Reutter, J. L.; and Silva, J. P. 2020. The Expressive Power of Graph Neural Networks as a Query Language. *SIGMOD Rec.*, 49(2): 6–17.

Benedikt, M.; Lu, C.; Motik, B.; and Tan, T. 2024. Decidability of Graph Neural Networks via Logical Characterizations. In *Proc. of ICALP*, volume 297 of *LIPIcs*, 127:1–127:20. Schloss Dagstuhl - Leibniz-Zentrum für Informatik.

Blackburn, P.; de Rijke, M.; and Venema, Y. 2001. *Modal Logic*, volume 53 of *Cambridge Tracts in Theoretical Computer Science*. Cambridge University Press. ISBN 978-1-10705088-4.

Bongini, P.; Bianchini, M.; and Scarselli, F. 2021. Molecular generative Graph Neural Networks for Drug Discovery. *Neurocomputing*, 450: 242–252.

Cantor, G. 1895. Beiträge zur Begründung der transfiniten Mengenlehre. *Mathematische Annalen*, 46(4): 481–512.

De Rijke, M. 2000. A note on graded modal logic. *Studia Logica*, 64(2): 271–283.

Demri, S.; and Lugiez, D. 2010. Complexity of modal logics with Presburger constraints. *J. of Applied Logic*, 8(3): 233–252.

Deng, A.; and Hooi, B. 2021. Graph Neural Network-Based Anomaly Detection in Multivariate Time Series. *Proc. of the AAAI*, 35(5): 4027–4035.

Elberfeld, M.; Grohe, M.; and Tantau, T. 2016. Where First-Order and Monadic Second-Order Logic Coincide. *ACM Transactions on Computational Logic*, 17(4).

Goble, L. F. 1970. Grades of modality. *Logique et Analyse*, 13(51): 323–334.

Goranko, V.; and Otto, M. 2007. 5 Model theory of modal logic. In Blackburn, P.; Van Benthem, J.; and Wolter, F., eds., *Handbook of Modal Logic*, volume 3 of *Studies in Logic and Practical Reasoning*, 249–329. Elsevier.

Grohe, M. 2021. The logic of graph neural networks. In *Proc. of LICS*, 5:1–5:17. ACM.

Grohe, M. 2024. The Descriptive Complexity of Graph Neural Networks. *TheoretiCS*, 3.

Grohe, M.; and Rosenbluth, E. 2024. Are Targeted Messages More Effective? In *Proc. of LICS*, 40:1–40:14. ACM.

Hamilton, W.; Ying, Z.; and Leskovec, J. 2017. Inductive Representation Learning on Large Graphs. In *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc.

Hornik, K.; Stinchcombe, M.; and White, H. 1989. Multilayer feedforward networks are universal approximators. *Neural Networks*, 2(5): 359–366.

Jiang, W.; and Luo, J. 2022. Graph neural network for traffic forecasting: A survey. *Expert Systems with Applications*, 207: 117921.

Khalife, S.; and Tonelli-Cueto, J. 2025. Is uniform expressivity too restrictive? Towards efficient expressivity of GNNs. In *Proc. of ICLR*. OpenReview.net.

Kipf, T. N.; and Welling, M. 2017. Semi-Supervised Classification with Graph Convolutional Networks. In *Proc. of ICLR*.

Leshno, M.; Lin, V. Y.; Pinkus, A.; and Schocken, S. 1993. Multilayer feedforward networks with a nonpolynomial activation function can approximate any function. *Neural Networks*, 6(6): 861–867.

Morris, M.; and Horrocks, I. 2025. Sound Logical Explanations for Mean Aggregation Graph Neural Networks. In *Proc. of NeurIPS*.

Otto, M. 2019. Graded modal logic and counting bisimulation. *arXiv preprint arXiv:1910.00039*.

Pacuit, E.; and Salame, S. 2004. Majority Logic. In *Proc. of KR*, 598–605. AAAI Press.

Pinkus, A. 1999. Approximation theory of the MLP model in neural networks. *Acta Numerica*, 8: 143–195.

Rockafellar, R. T. 1970. *Convex Analysis*. Princeton: Princeton University Press. ISBN 9781400873173.

Rosenbluth, E.; Toenshoff, J.; and Grohe, M. 2023. Some might say all you need is sum. In *Proc. of IJCAI*.

Scarselli, F.; Gori, M.; Tsoi, A. C.; Hagenbuchner, M.; and Monfardini, G. 2009. The Graph Neural Network Model. *IEEE Transactions on Neural Networks*, 20(1): 61–80.

Schönherr, M.; and Lutz, C. 2026. Logical Characterizations of GNNs with Mean Aggregation. In *Proc. of AAAI*. AAAI Press.

Shlomi, J.; Battaglia, P.; and Vlimant, J.-R. 2020. Graph neural networks in particle physics. *Machine Learning: Science and Technology*, 2(2): 021001.

Tena Cucala, D.; Cuenca Grau, B.; Motik, B.; and Kostylev, E. V. 2023. On the Correspondence Between Monotonic Max-Sum GNNs and Datalog. In *Proc. of KR*, 658–667.

Tena Cucala, D. J.; and Cuenca Grau, B. 2024. Bridging Max Graph Neural Networks and Datalog with Negation. In *Proc. of KR*, 950–961.

Wu, S.; Sun, F.; Zhang, W.; Xie, X.; and Cui, B. 2022. Graph Neural Networks in Recommender Systems: A Survey. *ACM Computing Surveys*, 55(5).

Wu, Z.; Pan, S.; Chen, F.; Long, G.; Zhang, C.; and Yu, P. S. 2021. A Comprehensive Survey on Graph Neural Networks. *IEEE Transactions on Neural Networks and Learning Systems*, 32(1): 4–24.

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

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 SIGKDD*, KDD '18, 974–983. ACM.

Zhou, J.; Cui, G.; Hu, S.; Zhang, Z.; Yang, C.; Liu, Z.; Wang, L.; Li, C.; and Sun, M. 2020. Graph neural networks: A review of methods and applications. *AI Open*, 1: 57–81.

A Proofs for Section 2

Lemma 1. *Let* $\mathsf{AGG} \in \{\mathsf{Max}, \mathsf{Sum}, \mathsf{Mean}\}$. *If a property is definable by a* simple AGG-*GNN with* ReLU^* *activation, then it is definable by a* simple AGG-*GNN with* ReLU *activation.*

Proof. To convert a simple AGG-GNN $\mathcal{G}$ with ReLU^* activation into a simple AGG-GNN $\mathcal{G}'$ with ReLU activation, we can thus proceed as follows. We double the number of layers. Each layer ℓ of $\mathcal{G}$ with output dimension δ^ℓ is simulated by layers $(2\ell - 1, 2\ell)$ of $\mathcal{G}'$ with output dimensions $2\delta^\ell$ and δ^ℓ, respectively. If the result of layer ℓ of $\mathcal{G}$ is $\mathsf{ReLU}^*(\bar{x})$, then the $2\ell - 1$-st layer of $\mathcal{G}'$ computes both $\mathsf{ReLU}(\bar{x})$ and $\mathsf{ReLU}(\bar{x} - \bar{1})$, where $\bar{1}$ is the all-1 vector of appropriate dimension, and stores the result in the now twice as large output feature vectors. The 2ℓ-th layer of $\mathcal{G}'$ then computes $\mathsf{ReLU}(\mathsf{ReLU}(\bar{x}) - \mathsf{ReLU}(\bar{x} - \bar{1}))$, which is equal to $\mathsf{ReLU}(\bar{x}) - \mathsf{ReLU}(\bar{x} - \bar{1})$, in a straightforward way. ❏

Lemma 3. *In the uniform setting,*

1. ML $\subsetneq$ RML *and* ML $\subsetneq$ GML*;*
2. RML $\not\subseteq$ GML *and* GML $\not\subseteq$ RML.

In the non-uniform setting, ML $\subsetneq$ RML $\subsetneq$ GML.

Proof. We start with the uniform setting. It is clear that ML $\subseteq$ GML and ML $\subseteq$ RML:

- in GML, we have $\Diamond\varphi \equiv \Diamond^{\geq 1}\varphi$;
- in RML, we have $\Diamond\varphi \equiv \Diamond^{>0}\varphi$.

We prove below in the non-uniform setting that GML $\not\subseteq$ ML, RML $\not\subseteq$ ML, and GML $\not\subseteq$ RML. By Lemma 2, these results carry over to the uniform setting.

It thus remains to argue that RML $\not\subseteq$ GML. Consider the RML formula $\varphi = \Diamond^{\geq \frac{1}{2}} P$. We argue that it is not expressible in GML. To prove this, notice that a GML formula φ with maximal counting constant c cannot distinguish between a vertex with c successors that satisfy P and c successors that do not satisfy P, and a vertex with c successors that satisfy P and $c+1$ successors that do not satisfy P. This can in fact be proved by induction on the structure of φ. But these two graphs can be distinguished by $\Diamond^{\geq \frac{1}{2}} P$.

In the non-uniform setting, we obtain ML $\subseteq$ RML from the uniform setting and Lemma 2. To show RML $\subseteq$ GML, we note that on graphs of size at most n, the RML diamond $\Diamond^{\geq r}\varphi$ can be expressed as

$$\Diamond^{=0}\top \vee \bigvee_{\substack{0 \leq p \leq q \leq n \\ q \neq 0,\, p/q \geq r}} \Diamond^{=q}\top \wedge \Diamond^{=p}\varphi.$$

The RML diamond $\Diamond^{>r}\varphi$ can be expressed similarly.

To prove RML $\not\subseteq$ ML, consider the RML formula $\Diamond^{>\frac{1}{2}} P$. We claim that it is not nonuniformly expressible in ML. We construct two graphs that can be distinguished by $\Diamond^{>\frac{1}{2}} P$ but cannot be distinguished by any formula in ML. In fact, ML cannot count and thus cannot distinguish between a vertex with two successors, exactly one of which satisfies P, and a vertex with one successor that satisfies P and two successors that do not. Formally, this can be proved using bisimulations, see for instance (Blackburn, de Rijke, and Venema 2001). The proof of GML $\not\subseteq$ ML is very similar, using the GML formula $\Diamond^{\geq 2} P$.

It remains to prove GML $\not\subseteq$ RML. Consider the GML formula $\varphi = \Diamond^{\geq 2} P$. We claim that it is not expressible in RML. In fact, it is easy to see that there does not exist an RML formula ψ that can distinguish between a vertex that has one successor that satisfies P (and no other successors) and a vertex that has two successors that satisfy P. Formally, this can be proved by induction on the structure of ψ. ❏

B Proofs for Section 3

Proof of Theorem 2

Theorem 2. RML $\subseteq$ simple Mean-GNN *in the non-uniform setting. This holds for truncated ReLU and ReLU activation.*

Proof. We start with truncated ReLU. Let φ be a formula in RML over a finite set $\Pi = \{P_1, \ldots, P_r\}$ of vertex labels and fix a maximum graph size $n \geq 1$. As observed in the main body of the paper, we may assume w.l.o.g. that φ contains no RML diamonds of the form $\Diamond^{\geq t}$, since they can be replaced by $\Diamond^{>t'}$, where

$$t' = \max\left\{\frac{\ell}{m} \mid 0 \leq \ell \leq m \leq n, \frac{\ell}{m} < t\right\}.$$

Note that t' is the largest rational number that is smaller than t and that can occur as a fraction in a graph where every vertex has at most n successors.

Let $\varphi_1, \ldots, \varphi_L$ be an enumeration of the subformulas of φ such that (i) $\varphi_i = P_i$ for $1 \leq i \leq r$ and (ii) if φ_ℓ is a subformula of φ_k, then $\ell < k$. As in (Barceló et al. 2020), we construct a simple GNN $\mathcal{G}$ with L layers, all of input and output dimension L. To encode satisfaction of a subformula φ_i at a vertex v, we store value 1 in the i-th component of the feature vector of v; likewise, falsification of φ_i is encoded by value 0. The GNN evaluates one subformula in each layer so that for every $k \in [L]$, from layer k on all subformulas $\varphi_1, \ldots, \varphi_k$ are encoded correctly. All layers use the same combination function.[2]

We define the matrices $A, C \in \mathbb{R}^L \times \mathbb{R}^L$ and the bias vector $\bar{b} \in \mathbb{R}^L$ that define the combination function of the (simple) GNN in the following way, where all entries that are not mentioned explicitly have value 0. Let $k \in [L]$. We make a case distinction to set certain entries:

Case 1: $\varphi_k = P_k$. Set $C_{k,k} = 1$.
Case 2: $\varphi_k = \neg\varphi_i$. Set $C_{i,k} = -1$ and $b_k = 1$.
Case 3: $\varphi_k = \varphi_i \vee \varphi_j$. Set $C_{i,k} = C_{j,k} = 1$.
Case 4: $\varphi_k = \Diamond^{>t}\varphi_i$. Set $A_{i,k} = n^2$ and

$$b_k = -n^2 \cdot \max\left\{\frac{\ell}{m} \mid 0 \leq \ell \leq m \leq n, \frac{\ell}{m} \leq t\right\}.$$

As the classification function we use

$$\mathsf{CLS}(\bar{x}) = \begin{cases} 1 & \text{if } x_L > 0 \\ 0 & \text{otherwise.} \end{cases}$$

[2] Such GNNs in are called 'homogeneous' in (Barceló et al. 2020).

We next show correctness of the translation.

Claim 1. For all φ_k, $1 \leq k \leq L$, the following holds: if $v \in V^G$ and $k \leq k' \leq L$, then

$$(\bar{x}^{k'}_{G,v})_k = \begin{cases} 1 & \text{if } G, v \models \varphi_k, \\ 0 & \text{otherwise.} \end{cases}$$

The proof is by induction on k, distinguishing Cases 1 to 4. For Cases 1, 2, 3, the (straightforward) arguments can be found in (Barceló et al. 2020). For Case 4, we first observe that different fractions with denominators bounded by n differ at least by a certain amount.

Claim 2. For all $n, n_1, n_2 \in \mathbb{N}^+$, $m_1, m_2 \in \mathbb{N}$ with $n_1, n_2 \leq n$ and $\frac{m_1}{n_1} \neq \frac{m_2}{n_2}$:

$$\left| \frac{m_1}{n_1} - \frac{m_2}{n_2} \right| \geq \frac{1}{n^2}.$$

The claim holds because the least common multiple between the denominators n_1 and n_2 is at most n^2 and the smallest non-zero difference between any two such fractions is clearly at least $\frac{1}{n^2}$.

Now consider Case 4 with $\varphi_k = \Diamond^{>t} \varphi_i$, assume that Claim 1 has already been shown for $\varphi_1, \dots, \varphi_{k-1}$, and let $k \leq k' \leq L$. Consider any vertex $v \in V^G$. Since $|V^G| \leq n$, the fraction of successors of v that satisfy φ_i can clearly be represented by $\frac{\ell}{m}$ for some ℓ, m with $\ell \leq m \leq n$. Since $i < k$, Claim 1 has already been shown for φ_i. Consequently, the i-th component of the vector computed by mean aggregation is $\frac{\ell}{m}$ and the i-th component of the new feature vector stored by level k' of the GNN at node v is

$$(\bar{x}^{k'}_v)_k = \mathsf{ReLU}^*(n^2 \frac{\ell}{m} + b_k).$$

First assume that $G, v \models \varphi_k$. Then $\frac{\ell}{m} > t$ and by definition of $\bar{b}_k$ and by Claim 2, this implies $n^2 \frac{\ell}{m} + b_k \geq 1$. Thus, by definition of the truncated ReLU the above value is 1, as required.

Now assume that $G, v \not\models \varphi_k$. Then $\frac{\ell}{m} \leq t$ and by definition of $\bar{b}_k$, this implies $n^2 \cdot \frac{\ell}{m} + b_k \leq 0$. Thus, by definition of the truncated ReLU the above value is 0, as required. This finishes the proof of Claim 1.

For the non-truncated ReLU, it suffices to invoke Lemma 1. ❑

Proof of Theorem 3

The proof uses universal approximation theorems which we introduce next.

Definition 2. *A function $f : \mathbb{R} \to \mathbb{R}$ has the universal approximation property if for every compact $K \subseteq \mathbb{R}^n$, every continuous $g : K \to \mathbb{R}^m$, and every $\varepsilon > 0$ there exist $d \in \mathbb{N}$, matrices $W_1 \in \mathbb{R}^{n \times d}$ and $W_2 \in \mathbb{R}^{d \times m}$ and a vector $\bar{b} \in \mathbb{R}^d$ such that*

$$\sup_{\bar{x} \in K} ||g(\bar{x}) - f(\bar{x}W_1 + \bar{b})W_2||_\infty < \varepsilon.$$

There are several theorems characterizing functions which have the universal approximation property. For example, (Leshno et al. 1993) show that

Lemma 9. *A continuous function $f : \mathbb{R} \to \mathbb{R}$ has the universal approximation property if and only if f is not polynomial.*

Corollary 4. *Truncated ReLU, ReLU, and sigmoid have the universal approximation property.*

(Hornik, Stinchcombe, and White 1989) show the universal approximation property for all monotone and bounded, but possibly noncontinuous, functions. See the survey (Pinkus 1999) for more information.

For the proof of Theorem 3 we will use GNNs which satisfy some continuity conditions for the combination and aggregation function.

We remind the reader that for $X \subseteq \mathbb{R}^\gamma$, a function $f : X \to \mathbb{R}^{\gamma'}$ is *continuous* if for all $\bar{x} \in X$ and all $\varepsilon > 0$ there exists a $\delta > 0$ such that for all $\bar{y} \in X$, $||\bar{x} - \bar{y}||_\infty < \delta$ implies $||f(\bar{x}) - f(\bar{y})||_\infty < \varepsilon$. The function f is *uniformly continuous* if δ can be chosen independently of $\bar{x}$. That is, for all $\varepsilon > 0$ there exists a $\delta > 0$ such that for all $\bar{x}, \bar{y} \in X$, $||\bar{x} - \bar{y}||_\infty < \delta$ implies $||f(\bar{x}) - f(\bar{y})||_\infty < \varepsilon$.

Definition 3. *We call an aggregation function* AGG *with input dimension γ* bounded continuous *if for all $\varepsilon > 0$ and $\bar{x}_1, \dots, \bar{x}_n \in \mathbb{R}^\gamma$, there exists a $\delta > 0$ such that for all $\bar{y}_1, \dots, \bar{y}_n \in \mathbb{R}^\gamma$ with $||\bar{x}_i - \bar{y}_i||_\infty < \delta$ for all $i \in [n]$, we have $||\mathsf{AGG}(\{\{\bar{x}_1, \dots, \bar{x}_n\}\}) - \mathsf{AGG}(\{\{\bar{y}_1, \dots, \bar{y}_n\}\})||_\infty < \varepsilon$.*

As we will see, the notion of bounded continuity guarantees that in the non-uniform setting the aggregation function behaves like a continuous function, since in graphs up to size n, the aggregation function is applied to multisets of size at most n.

Lemma 10. MEAN*,* MAX *and* SUM *are bounded continuous.*

Proof. In case $\mathsf{AGG} \in \{\mathsf{MEAN}, \mathsf{MAX}\}$ it can be verified easily that if $||\bar{x}_i - \bar{y}_i||_\infty < \varepsilon$ for all $i \in [n]$, then $||\mathsf{AGG}(\{\{\bar{x}_1, \dots, \bar{x}_n\}\}) - \mathsf{AGG}(\{\{\bar{y}_1, \dots, \bar{y}_n\}\})||_\infty < \varepsilon$. In these cases, the choice of δ thus only depends on ε, and therefore MEAN and MAX actually satisfy a very strong version of bounded continuity. In particular, it is bounded continuous in the sense of Definition 3

For $\mathsf{AGG} = \mathsf{SUM}$, δ additionally depends on the size of the multisets. If $||\bar{x}_i - \bar{y}_i||_\infty < \frac{\varepsilon}{n}$ for all $i \in [n]$, then $||\mathsf{AGG}(\{\{\bar{x}_1, \dots, \bar{x}_n\}\}) - \mathsf{AGG}(\{\{\bar{y}_1, \dots, \bar{y}_n\}\})||_\infty < \varepsilon$. It is still bounded continuous in the sense of Definition 3, even in a stronger sense since δ does not depend on the feature vectors but only on the size of the multisets. The reader should think of Definition 3 as a minimum requirement for bounded continuity. ❑

Lemma 11. *Let* $\mathsf{AGG}, \mathsf{COM}$ *be a GNN layer with input dimension γ where* AGG *is bounded continuous and* COM *is continuous. Then for all finite sets $\chi \subseteq \mathbb{R}^\gamma$, $n \in \mathbb{N}$, $\varepsilon > 0$ there exists a $\delta > 0$ such that for all $m \in [n]$, $\bar{x}_0, \dots, \bar{x}_m \in \chi$ and $\bar{y}_0, \dots, \bar{y}_m \in \mathbb{R}^\gamma$ with $||\bar{x}_i - \bar{y}_i||_\infty < \delta$,* *we have*

$$||\mathsf{COM}(\bar{x}_0, \mathsf{AGG}(\{\{\bar{x}_1, \dots, \bar{x}_m\}\})) \\ - \mathsf{COM}(\bar{y}_0, \mathsf{AGG}(\{\{\bar{y}_1, \dots, \bar{y}_m\}\}))||_\infty < \varepsilon.$$

Proof. It suffices to show that for each size of the multisets m and each χ and ε there exists such δ_m. For a given n we then can choose $\delta = \min_{0 \le i \le n} \delta_i$.

Since m is fixed, we can view AGG as a continuous function $\mathsf{AGG} : \mathbb{R}^{m\gamma} \to \mathbb{R}^\gamma$. Thus, we can view the layer as a function $L : \mathbb{R}^{(m+1)\gamma} \to \mathbb{R}^\gamma$. L is continuous, since continuous functions are closed under composition.

We define for each $\bar{x} \in \chi$ the set $I_{\bar{x}} = \{\bar{y} \in \mathbb{R}^\gamma \mid ||\bar{x} - \bar{y}||_\infty \le 1\}$, which is closed and bounded. We define $I = \bigcup_{\bar{x} \in \chi} I_{\bar{x}}$, which is a finite union of closed and bounded sets and thus also closed and bounded. By the Heine-Borel theorem from real analysis, I is compact and by the Heine-Cantor theorem, L restricted to the domain I^{m+1} is uniformly continuous. Hence, we can find a δ' such that for all $\bar{x}, \bar{y} \in I^{m+1}$ with $||\bar{x} - \bar{y}||_\infty < \delta'$ we have $||L(\bar{x}) - L(\bar{y})||_\infty < \varepsilon$. We can now choose $\delta_m = \min(1, \delta')$. ❑

We also observe that if the size of the input graphs is bounded by a constant, a GNN can generate only a constant number of different feature vectors, across all input graphs. In fact, this is an immediate consequence of the fact that, for every finite set of vertex labels Π and every $n \ge 1$, there are only finitely many Π-labeled graphs of size at most n.

Lemma 12. *Let $\mathcal{G}$ be a GNN with L layers over some finite set of vertex labels Π. For each layer ℓ of output dimension δ^ℓ and for all $n \ge 1$, let $\chi_n^\ell \subseteq \mathbb{R}^{\delta^\ell}$ denote the set of feature vectors $\bar{x}$ such that for some input graph G of size at most n and some vertex v in G, $\mathcal{G}$ generates $\bar{x}$ at v in layer ℓ. Then χ_n^ℓ is finite.*

We consider aggregation functions that commute with matrix multiplication.

Lemma 13. *For all $\mathsf{AGG} \in \{\mathsf{SUM}, \mathsf{MEAN}\}$, $x_1, \dots, x_n \in \mathbb{R}^\delta$ and matrices $A \in \mathbb{R}^{\delta \times \gamma}$*

$$\mathsf{AGG}(\{\{\bar{x}_1, \dots, \bar{x}_n\}\})A = \mathsf{AGG}(\{\{\bar{x}_1 A, \dots, \bar{x}_n A\}\}).$$

Proof. MEAN commutes with matrix multiplication, since for all $\bar{x}_1, \dots, \bar{x}_n \in \mathbb{R}^\delta$ and $A \in \mathbb{R}^{\delta \times \gamma}$

$$\mathsf{MEAN}(\{\{\bar{x}_1, \dots, \bar{x}_n\}\})A = \left(\frac{1}{n}\sum_{i=1}^n \bar{x}_i\right) A \\ = \frac{1}{n}\sum_{i=1}^n (\bar{x}_i A) = \mathsf{MEAN}(\{\{\bar{x}_1 A, \dots, \bar{x}_n A\}\}).$$

The proof that SUM also commutes with matrix multiplication is analogous. ❑

We remark that MAX does not commute with matrix multiplication. A counterexample is $\bar{x}_1 = 0, \bar{x}_2 = 1$ and $A = (-1)$, since $\mathsf{MAX}(\{\{0,1\}\})A = -1 \ne 0 = \mathsf{MAX}(\{\{0,-1\}\}) = \mathsf{MAX}(\{\{0A, 1A\}\})$.

The following lemma is the main ingredient to the proof of Theorem 3.

Lemma 14. *Let $\mathcal{G} = (L, \{\mathsf{AGG}^\ell\}_{\ell \in [L]}, \{\mathsf{COM}^\ell\}_{\ell \in [L]}, \mathsf{CLS})$ be a GNN where each AGG^ℓ is bounded continuous and each COM^ℓ is continuous. Let each AGG^ℓ commute with matrix multiplication. Let $f : \mathbb{R} \to \mathbb{R}$ be a function with the following properties.*

1. *f has the universal approximation property.*
2. *There exist $a, b \in \mathbb{R}$ and an $\varepsilon > 0$ such that all $x \in (a - \varepsilon, a + \varepsilon)$ and $y \in (b - \varepsilon, b + \varepsilon)$ satisfy $f(x) < f(y)$.*

Then for all $n \in \mathbb{N}^+$ there exists a simple GNN $\mathcal{G}'$ with activation function f and $L + 2$ layers such that $\mathcal{G}$ and $\mathcal{G}'$ are equivalent on graphs up to size n. $\mathcal{G}'$ uses for each layer $1 \le \ell \le L$ the aggregation function AGG^ℓ and AGG^{L+1} and AGG^{L+2} can be chosen arbitrarily.

Proof. Let $\mathcal{G}$ and f be as in the lemma, and let $n \in \mathbb{N}^+$ be a bound on the graph size. Let $\delta^0, \dots, \delta^L$ be the dimensions of the layers of $\mathcal{G}$. We will use the values a and b from Point 2 to define the classification function of $\mathcal{G}'$, and the $\pm\varepsilon$-balls around a and b are needed because the layers of $\mathcal{G}'$ will only approximate those of $\mathcal{G}$. To prepare for this, we first introduce a modest extension $\mathcal{H}$ of $\mathcal{G}$ with one additional layer of output dimension 1 that already introduces the values a and b. For each $\bar{x}_v \in \chi_n^L$ and $\bar{x}_a \in \mathbb{R}^{\delta^L}$, define

$$\mathsf{COM}^{L+1}(\bar{x}_v, \bar{x}_a) = \begin{cases} b & \text{if } \mathsf{CLS}(\bar{x}_v) = 1 \\ a & \text{otherwise.} \end{cases}$$

Since χ_n^L is a finite set and $\bar{x}_a$ is ignored by COM^{L+1}, we can find a polynomial (which is continuous) that extends COM^{L+1} to the domain $\mathbb{R}^{\delta^L + \delta^L}$. Let AGG^{L+1} be any bounded continuous [3] aggregation function. For the classification function, we choose a threshold between a and b. Let $\sim$ be $>$ if $b > a$ and let $\sim$ be $<$ otherwise. Then

$$\mathsf{CLS}'(x) = \begin{cases} 1 & \text{if } x \sim \frac{a+b}{2} \\ 0 & \text{otherwise.} \end{cases}$$

The GNN

$$\mathcal{H} = (L + 1, \{\mathsf{COM}^\ell\}_{\ell \in [L+1]}, \{\mathsf{AGG}^\ell\}_{\ell \in [L+1]}, \mathsf{CLS}')$$

is equivalent to $\mathcal{G}$ on graphs of size at most n, since $\mathsf{CLS}(\bar{x}) = \mathsf{CLS}'(\mathsf{COM}^{L+1}(\bar{x}, \bar{y}))$ for all $\bar{x} \in \chi_n^L$ and $\bar{y} \in \mathbb{R}^{\delta^L}$. In addition, each aggregation function is bounded continuous and each combination function is continuous. For all $\ell \in [L+1]$, let χ_n^ℓ be the set of feature vectors computed by $\mathcal{H}$ in layer ℓ on some pointed graph of size at most n. By Lemma 12, all these sets are finite.

We now show how to approximate each layer in $\mathcal{H}$ using f, in a way that shall allow us to construct the desired GNN $\mathcal{G}'$. We define the approximations of the layers inductively, starting with the last layer. The desired error bound of the last layer is $\varepsilon_{L+1} = \varepsilon$.

Knowing the desired error bound ε_ℓ we define the approximation of the ℓ-th layer in the following way. Since AGG^ℓ is

[3] In principle, there is no need for AGG^{L+1} to be bounded continuous. We only assume this here so that the preconditions of Lemma 11 are satisfied.

bounded continuous and COM^ℓ is continuous, by Lemma 11 there exists an $\varepsilon_{\ell-1}$ such that for all $m \in [n]$, $\bar{x}_0, \ldots, \bar{x}_m \in \chi_n^{\ell-1}$ and $\bar{y}_0, \ldots, \bar{y}_m \in \mathbb{R}^{\delta^\ell}$ with $||\bar{x}_i - \bar{y}_i||_\infty < \varepsilon_{\ell-1}$, we have

$$||\mathsf{COM}^\ell(\bar{x}_0, \mathsf{AGG}^\ell(\{\{\bar{x}_1, \ldots, \bar{x}_m\}\})) \\ - \mathsf{COM}^\ell(\bar{y}_0, \mathsf{AGG}^\ell(\{\{\bar{y}_1, \ldots, \bar{y}_m\}\}))||_\infty < \frac{\varepsilon_\ell}{2}.$$

We define $I_{\ell-1} = \bigcup_{\bar{x} \in \chi_n^{\ell-1}} \{\bar{y} \in \mathbb{R}^{\delta^{\ell-1}} \mid ||\bar{x} - \bar{y}||_\infty \leq \varepsilon_{\ell-1}\}$. It contains all vectors that have a distance of at most $\varepsilon_{\ell-1}$ to some vector in $\chi_n^{\ell-1}$. By the Heine-Borel theorem and since $I_{\ell-1}$ is bounded and closed, $I_{\ell-1}$ is compact. We define $J_{\ell-1}$ similarly, but with the aggregation function in mind. Let

$$J_{\ell-1} = \bigcup_{0 \leq i \leq n} \mathsf{AGG}^\ell\{\{\bar{y}_1, \ldots, \bar{y}_i \mid \forall k \in [i] \colon \bar{y}_k \in I_{\ell-1}\}\}.$$

$J_{\ell-1}$ is a compact set, because $(I_{\ell-1})^i$ is compact for all $i \in [n]$, AGG^ℓ is a continuous function when restricted to multisets of size i and because continuous functions preserve compact sets.

Because f has the universal approximation property, there exist a dimension γ^ℓ and matrices $W_1^\ell \in \mathbb{R}^{2\delta^{\ell-1} \times \gamma^\ell}$ and $W_2^\ell \in \mathbb{R}^{\gamma^\ell \times \delta^\ell}$ and a vector $\bar{b}^\ell \in \mathbb{R}^{\gamma^\ell}$ such that for all $\bar{x} \in I_{\ell-1}$ and $\bar{y} \in J_{\ell-1}$ we have

$$||\mathsf{COM}^\ell(\bar{x}, \bar{y}) - f((\bar{x} \circ \bar{y})W_1^\ell + \bar{b}^\ell)W_2^\ell||_\infty < \frac{\varepsilon_\ell}{2}, \qquad (*)$$

where $\circ$ denotes vector concatenation. We choose $\varepsilon_{\ell-1}$ as the desired error bound of layer $\ell - 1$.

We now use the approximations identified above to define a GNN $\mathcal{G}'$. To obtain a simple GNN, we split each $W_1^\ell \in \mathbb{R}^{2\delta^{\ell-1} \times \gamma^\ell}$ into two halves $C^\ell, A^\ell \in \mathbb{R}^{\delta^{\ell-1} \times \gamma^\ell}$ such that $\bar{x}C^\ell + \bar{y}A^\ell = (\bar{x} \circ \bar{y})W_1^\ell$ for all $\bar{x}, \bar{y} \in \mathbb{R}^{\delta^{\ell-1}}$. Note that in $(*)$, W_2^ℓ is multiplied outside of f. We thus have to defer this multiplication to the subsequent layer. For $1 \leq \ell \leq L+1$, the ℓ-th layer computes

$$\mathsf{COM'}^\ell(\bar{x}_v, \bar{x}_a) = f(\bar{x}_v \cdot W_2^{\ell-1} \cdot C^\ell + \bar{x}_a \cdot W_2^{\ell-1} \cdot A^\ell + \bar{b}^\ell),$$

where W_2^0 is the identity matrix of appropriate size. The $L + 2$-nd layer computes

$$\mathsf{COM'}^{L+2}(\bar{x}_v, \bar{x}_a) = f(\bar{x}_v W_2^{L+1}).$$

Let $c = \inf f((b - \varepsilon, b + \varepsilon))$ and let $\sim$ be $>$ if $c \in f((a - \varepsilon, a + \varepsilon))$ and $\geq$ otherwise. As the classification function, we use

$$\mathsf{CLS}_2(\bar{x}) = \begin{cases} 1 & \text{if } x_L \sim c \\ 0 & \text{otherwise.} \end{cases}$$

We define

$$\mathcal{G}' = (L + 2, \{\mathsf{COM'}^\ell\}_{\ell \in [L+2]}, \{\mathsf{AGG}^\ell\}_{\ell \in [L+2]}, \mathsf{CLS}_2).$$

Since the last layer ignores the aggregation vector, we can choose AGG^{L+2} arbitrarily.

Claim. $\mathcal{G}'$ is equivalent to $\mathcal{H}$ (and thus also to $\mathcal{G}$) on graphs of size at most n.

Let G be a graph of size at most n. For each vertex $v \in V^G$, we use $\bar{x}_v^\ell$ and $\bar{y}_v^\ell$ to denote the feature vectors computed by $\mathcal{H}$ and $\mathcal{G}'$ at v in layer ℓ, respectively. By definition, $\bar{x}_v^\ell \in \chi_n^\ell$ for all $v \in V(G)$. Let

$$\mathsf{APR}^\ell(\bar{x}_v, \bar{x}_a) = f((\bar{x}_v \circ \bar{x}_a)W_1^\ell + \bar{b}^\ell)W_2^\ell$$

be the approximation of $\mathsf{COM}^\ell(\bar{x}_v, \bar{x}_a)$. We have to show that for each $0 \leq \ell \leq L+1$,

$$||\bar{x}_v^\ell - \bar{y}_v^\ell W_2^\ell||_\infty < \varepsilon_\ell.$$

For $\ell = 0$ this statement holds, since $\bar{y}_v^\ell = \bar{x}_v^\ell$ and W_2^0 is the identity matrix. For the induction step, we observe that

$$\begin{aligned} &\bar{y}_v^\ell W_2^\ell \\ &= f(\bar{y}_v^{\ell-1} W_2^{\ell-1} C^\ell + \mathsf{AGG}\{\{\bar{y}_u^{\ell-1}\}\} W_2^{\ell-1} A^\ell + \bar{b}^\ell) W_2^\ell \\ &= f(\bar{y}_v^{\ell-1} W_2^{\ell-1} C^\ell + \mathsf{AGG}\{\{\bar{y}_u^{\ell-1} W_2^{\ell-1}\}\} A^\ell + \bar{b}^\ell) W_2^\ell \\ &= f(((\bar{y}_v^{\ell-1} W_2^{\ell-1}) \circ \mathsf{AGG}\{\{\bar{y}_u^{\ell-1} W_2^{\ell-1}\}\}) W_1^\ell + \bar{b}^\ell) W_2^\ell \\ &= \mathsf{APR}^\ell(\bar{y}_v^{\ell-1} W_2^{\ell-1}, \mathsf{AGG}\{\{\bar{y}_u^{\ell-1} W_2^{\ell-1}\}\}) \end{aligned}$$

where the second equation exploits that AGG commutes with matrix multiplication.

By induction hypothesis, $||\bar{x}_w^{\ell-1} - \bar{y}_w^{\ell-1} W_2^{\ell-1}||_\infty < \varepsilon_{\ell-1}$ for all vertices $w \in V^G$, and therefore

$$||\mathsf{COM}^\ell(\bar{x}_v^{\ell-1}, \mathsf{AGG}^\ell(\{\{\bar{x}_u^{\ell-1}\}\})) \\ - \mathsf{COM}^\ell(\bar{y}_v^{\ell-1} W_2^{\ell-1}, \mathsf{AGG}^\ell(\{\{\bar{y}_u^{\ell-1} W_2^{\ell-1}\}\}))||_\infty < \frac{\varepsilon_\ell}{2}.$$

Using APR^ℓ instead of COM^ℓ adds a second error which is bounded by $\frac{\varepsilon_\ell}{2}$, since all $\bar{y}_w W_2^{\ell-1}$ are in $I_{\ell-1}$ and $\mathsf{AGG}^\ell(\{\{\bar{y}_u^{\ell-1} W_2^{\ell-1}\}\}) \in J_{\ell-1}$. Therefore,

$$\begin{aligned} &||\bar{x}_v^\ell - \bar{y}_v^\ell W_2^\ell||_\infty \\ =&||\bar{x}_v^\ell - \mathsf{COM}^\ell(\bar{y}_v^{\ell-1} W_2^{\ell-1}, \mathsf{AGG}^\ell(\{\{\bar{y}_u^{\ell-1} W_2^{\ell-1}\}\})) \\ +&\mathsf{COM}^\ell(\bar{y}_v^{\ell-1} W_2^{\ell-1}, \mathsf{AGG}^\ell(\{\{\bar{y}_u^{\ell-1} W_2^{\ell-1}\}\})) - \bar{y}_v^\ell W_2^\ell||_\infty \\ \leq&||\bar{x}_v^\ell - \mathsf{COM}^\ell(\bar{y}_v^{\ell-1} W_2^{\ell-1}, \mathsf{AGG}^\ell(\{\{\bar{y}_u^{\ell-1} W_2^{\ell-1}\}\}))||_\infty \\ +&||\mathsf{COM}^\ell(\bar{y}_v^{\ell-1} W_2^{\ell-1}, \mathsf{AGG}^\ell(\{\{\bar{y}_u^{\ell-1} W_2^{\ell-1}\}\})) - \bar{y}_v^\ell W_2^\ell||_\infty \\ \leq& \varepsilon_\ell \end{aligned}$$

where the topmost inequality is due to the triangle inequality.

The last layer applies the last missing W_2^{L+1} to $\bar{y}_v^{L+1}$. If a pointed graph (G, v) with size at most n is accepted by $\mathcal{H}$, that is $x_v^{L+1} = b$, then $\bar{y}_v^{L+1} W_2^{L+1} \in (b - \varepsilon, b + \varepsilon)$, since $\varepsilon_{L+1} = \varepsilon$. Therefore, $\mathsf{CLS}_2(\bar{y}_v^{L+2}) = 1$ and $\mathcal{G}'$ accepts (G, v). Likewise, if (G, v) is not accepted by $\mathcal{H}$, then $x_v^{L+1} = a$ and $\bar{y}_v^{L+1} W_2^{L+1} \in (a - \varepsilon, a + \varepsilon)$. Thus, $\mathsf{CLS}_2(\bar{y}_v^{L+2}) = 0$ and $\mathcal{G}'$ does not accept (G, v). Thus, the GNN $\mathcal{G}'$ accepts the same pointed graphs as $\mathcal{H}$ and $\mathcal{G}$ up to size n. ❏

Theorem 3 is now a straightforward consequence.

Theorem 3. *In the non-uniform setting and for all continuous non-polynomial activation functions:*

1. RML $\subseteq$ simple Mean-GNN*;*

2. GML $\subseteq$ simple Sum-GNN.

Proof. Let n be a maximum graph size. Further let f be a continuous, non-polynomial activation function. By Lemma 9, f has the universal approximation property. Since f is not a polynomial, it is not constant, and since f is continuous, there exist $a, b \in \mathbb{R}$ and $\varepsilon > 0$ such that for all $x \in (a-\varepsilon, a+\varepsilon)$ and $y \in (b-\varepsilon, b+\varepsilon)$ we have $f(x) < f(y)$.

We now show Point 1 of Theorem 3. By Theorem 2, for each formula φ in RML there exists a simple Mean-GNN $\mathcal{G}$ with truncated ReLU that is equivalent to φ on graphs of size at most n. Truncated ReLU is continuous, thus the combination functions in $\mathcal{G}$ are continuous. Since MEAN is bounded continuous and commutes with matrix multiplication, by Lemma 14 there is a Mean-GNN that uses f as its activation function and is equivalent to $\mathcal{G}$ on graphs up to size n.

Now for Point 2 of Theorem 3. It is shown in (Barceló et al. 2020) that for each formula ψ in GML there exists a simple Sum-GNN $\mathcal{G}$ with truncated ReLU that is equivalent to ψ (on all graphs). Since SUM is bounded continuous and commutes with matrix multiplication, Lemma 14 yields a simple Sum-GNN that uses f as its activation function and is equivalent to $\mathcal{G}$ on graphs up to size n. ❑

Proof of Theorems 4, 5, and 6

Theorem 4. ML $\subseteq$ simple Max-GNN *in the uniform setting. This holds for truncated ReLU and ReLU activation.*

Proof. The proof is a minor variation of that of Theorem 2, but simpler. In particular, in the case $\varphi_k = \Diamond\varphi_i$, we can simply set $A_{i,k} = 1$ and $\bar{b}_k = 0$. Note that in contrast to Case 4 in the proof of Theorem 2 and in analogy with the proof in (Barceló et al. 2020), the graph size is not used. One can then easily prove an analogue of Claim 1 in the proof of Theorem 2, without using the fact that the graph size is bounded. ❑

Theorem 5. *In the non-uniform setting,*

1. Mean-GNN $\subseteq$ RML*;*
2. Sum-GNN $\subseteq$ GML.

Proof. We prove Points 1 and 2 of Theorem 5 simultaneously. Let

$$\mathcal{G} = (L, \{\mathsf{AGG}^\ell\}_{\ell\in[L]}, \{\mathsf{COM}^\ell\}_{\ell\in[L]}, \mathsf{CLS})$$

be a GNN on Π-labeled graphs, with $\Pi = \{P_1, \dots, P_r\}$, and let $n \geq 1$ be an upper bound on the size of input graphs. Let χ_n^ℓ be the set of feature vectors $\bar{x}$ such that for some input graph G of size at most n and some vertex $v \in V^G$, $\mathcal{G}$ generates $\bar{x}$ at v in layer ℓ. By Lemma 12, each χ_n^ℓ is finite. We construct one modal logic formula $\varphi_{\bar{x}}^\ell$ for every layer ℓ and every $\bar{x} \in \chi_n^\ell$, proceeding by induction on ℓ.

For $\ell = 0$, we can derive the desired formulas from the initial feature vector. For all $\bar{x} \in \chi_n^\ell$, set

$$\varphi_{\bar{x}}^0 = \bigwedge_{\substack{1\leq i\leq r\\ x_i=0}} \neg P_i \wedge \bigwedge_{\substack{1\leq i\leq r\\ x_i=1}} P_i.$$

In the induction step, we enumerate all possibilities for $\mathcal{G}$ to generate $\bar{x}$. This is different depending on the aggregation function and logic we work with. We start with the case of RML. Let $\chi_n^{\ell-1} = \{\bar{y}_1, \dots, \bar{y}_p\}$ and

$$F = \left\{ \frac{\ell}{m} \mid 1 \leq \ell \leq m \leq n,\ \ell, m \in \mathbb{N}^+ \right\} \cup \{0\}.$$

For all $\bar{y} \in \chi_n^{\ell-1}$ and all $f_1, \dots, f_p \in F$ with $\sum_{i=1}^p f_i = 1$, define

$$\varphi_{\bar{y},f_1,\dots,f_p}^\ell = \varphi_{\bar{y}}^{\ell-1} \wedge \bigwedge_{i=1}^p \Diamond^{=f_i} \varphi_{\bar{y}_i}^{\ell-1}.$$

For each $\bar{x} \in \chi_n^\ell$, we can now define the formula $\varphi_{\bar{x}}^\ell$ to be the disjunction of all formulas $\varphi_{\bar{y},f_1,\dots,f_p}^\ell$ such that $\bar{x} = \mathsf{COM}^\ell(\bar{y}, \mathsf{MEAN}\, M)$ where M is any finite multiset over $\chi_n^{\ell-1}$ that realizes the fractions $f_1, \dots, f_p$, that is, $\frac{M(\bar{y}_i)}{|M|} = f_i$ for $1 \leq i \leq p$.

We can now show the following:

Claim 1. For all Π-labeled graphs of size at most n, all $v \in V^G$, all layers ℓ of $\mathcal{G}$, and each feature vector $\bar{x} \in \chi_n^\ell$: $\mathcal{G}$ assigns $\bar{x}$ to v in G in layer ℓ if and only if $G, v \models \varphi_{\bar{x}}^\ell$.

This is in fact easy to prove by induction on ℓ, using the fact that the result of mean aggregation only depends on the fraction of successors at which each possible feature vector was computed by the previous layer (and not on the exact number of such successors).

The case of Sum-GNNs is very similar, except that for sum aggregation we need to know the exact number of times that each feature vector has been computed at some successor by the previous layer. Since the number of successor is bounded by the constant n, we can express all possibilities in GML. For all $\bar{y} \in \chi_n^{\ell-1}$ and all $m_1, \dots, m_p \in [n] \cup \{0\}$, define

$$\varphi_{\bar{y},m_1,\dots,m_p}^\ell = \varphi_{\bar{y}}^{\ell-1} \wedge \bigwedge_{i=1}^p \Diamond^{=m_i} \varphi_{\bar{y}_i}^{\ell-1}.$$

For each $\bar{x} \in \chi_n^\ell$, we then define the formula $\varphi_{\bar{x}}^\ell$ to be the disjunction of all formulas $\varphi_{\bar{y},m_1,\dots,m_p}^\ell$ such that $\bar{x} = \mathsf{COM}^\ell(\bar{y}, \mathsf{SUM}\, M)$ where M is the multiset over $\chi_n^{\ell-1}$ defined by $M(\bar{y}_i) = m_i$ for $1 \leq i \leq p$. The rest of the proof remains unchanged. ❑

Theorem 6. *In the uniform setting,* Max-GNN $\subseteq$ ML.

Proof. The proof is analogous to that of Lemma 5, with two differences. First, finiteness of the relevant sets of feature vectors holds already without imposing a constant bound on the graph size. And second, we once more have to adapt the formulas $\varphi_{\bar{y},f_1,\dots,f_p}^\ell$.

Claim. Let $\mathcal{G}$ be a Max-GNN with L layers over some finite set of vertex labels Π. For each layer ℓ of output dimension δ^ℓ let $\chi^\ell \subseteq \mathbb{R}^{\delta^\ell}$ denote the set of feature vectors $\bar{x}$ such that for some input graph G and some vertex v in G, $\mathcal{G}$ generates $\bar{x}$ at v in layer ℓ. Then χ^ℓ is finite.

Proof of claim. The proof is straightforward by induction on ℓ. χ^0 is the set of all vectors over the set $\{0,1\}$. Then the set χ^ℓ consists of all vectors $\bar{x}$ that can be obtained by choosing a vector $\bar{y} \in \chi^{\ell-1}$ and a set $S \subseteq \chi^{\ell-1}$ and setting $\bar{x} = \mathsf{COM}^\ell(\bar{y}, \mathsf{MAX}\, S)$. Clearly, this implies that χ^ℓ is finite. The crucial difference to the sum and mean aggregation cases is that we can work with a set here rather than with multisets since max aggregation is oblivious to multiplicities.

Now for the construction of the ML formulas $\varphi^\ell_{\bar{x}}$. In the case of Max-GNNs, the result of aggregation only depends on which feature vectors have been computed at some successor by the previous layer, and which have not. We can thus use the exact same arguments as in the proof of Lemma 5, except that the formulas $\varphi^\ell_{\bar{y},f_1,\ldots,f_p}$ are replaced as follows. For all $\bar{y} \in \chi^{\ell-1}$ and all subsets $S \subseteq \chi^{\ell-1}$, define

$$\varphi_{\bar{y},S} = \varphi^{\ell-1}_{\bar{y}} \wedge \bigwedge_{\bar{z}\in S} \Diamond\varphi^{\ell-1}_{\bar{z}} \wedge \bigwedge_{\bar{z}\in\chi^{\ell-1}\setminus S} \Box\neg\varphi^{\ell-1}_{\bar{z}}.$$

For each $\bar{x} \in \chi^\ell$, we then define the formula $\varphi^\ell_{\bar{x}}$ to be the disjunction of all formulas $\varphi_{\bar{y},S}$ such that $\bar{x} = \mathsf{COM}^\ell(\bar{y}, \mathsf{MAX}\, S)$. The rest of the proof remains unchanged. ❑

C Proofs for Section 4

Theorem 9. Mean-GNN $\cap$ MSO $\subseteq$ GML *in the uniform setting.*

Proof. It is well-known that any property P definable by a GNN, independently of the aggregation, combination, and classification function used, is invariant under graded bisimulation, that is, if $(G_1, v_1) \in P$ and (G_1, v_1) and (G_2, v_2) are graded bisimilar, then $(G_2, v_2) \in P$. We refer to (Barceló et al. 2020) for a proof and also for a detailed definition of graded bisimulations. It was further proved in (Otto 2019) that every FO formula in one free variable that is invariant under graded bisimulation (on finite models) is equivalent to a GML formula. Together with Lemma 4, this clearly implies the statement in the theorem. ❑

Lemma 6. *Let (G_1, v_1), (G_2, v_2) be pointed graphs and $\ell \geq 0$. If D has a winning strategy in $\mathcal{E}^{\mathrm{ML}}_\ell(G_1, v_1, G_2, v_2)$, then D also has a winning strategy in*

$$\mathcal{E}^{\mathrm{GML}[c]}_\ell(c\cdot G_1, (v_1, k_1), c\cdot G_2, (v_2, k_2)),$$

for all $c \geq 1$ and $k_1, k_2 \in [c]$.

Proof. The proof is by induction on ℓ. The statement is obviously true for $\ell = 0$ because (v_i, k_i) satisfies the same propositional variables as v_i. Therefore, because the winning condition for Spoiler is not satisfied in $\mathcal{E}^{\mathrm{ML}}_0(G_1, v_1, G_2, v_2)$, it is also not satisfied in

$$\mathcal{E}^{\mathrm{GML}[c]}_0(c\cdot G_1, (v_1, k_1), c\cdot G_2, (v_2, k_2)).$$

For the induction step, let the statement hold for $\ell - 1$ and assume that D has a winning strategy in $\mathcal{E}^{\mathrm{ML}}_\ell(G_1, v_1, G_2, v_2)$. Then we find a winning strategy for D in

$$\mathcal{E}^{\mathrm{GML}[c]}_\ell(c\cdot G_1, (v_1, k_1), c\cdot G_2, (v_2, k_2)),$$

as follows. In the first round:

- Assume that in Step 1 of the first round S selects $i \in \{1,2\}$ and $U_i = \{(u^1_i, a_1), \ldots, (u^n_i, a_n)\} \subseteq \mathcal{N}_{G_i}(v_i, k_i)$, with $n \leq c$. In the game $\mathcal{E}^{\mathrm{ML}}_\ell(G_1, v_1, G_2, v_2)$, S may choose the same i and any of the vertices $u^1_i, \ldots, u^n_i$ (or other elements of $\mathcal{N}_{G_i}(v_i)$) and for each choice u^j_i the winning strategy for D provides an answer $u^j_{3-i} \in \mathcal{N}_{G_{3-i}}(v_{3-i})$. In Step 2, we let D choose $U_{3-i} = \{(u^1_{3-i}, 1), \ldots (u^n_{3-i}, n)\} \subseteq \mathcal{N}_{G_{3-i}}(v_{3-i}, k_{3-i})$, which satisfies $|U_1| = |U_2|$.
- Assume that S chooses (u^j_{3-i}, j) in Step 3. Then D chooses (u^j_i, a_j).

Since D has a winning strategy in $\mathcal{E}^{\mathrm{ML}}_\ell(G_1, v_1, G_2, v_2)$, D also has a winning strategy in $\mathcal{E}^{\mathrm{ML}}_{\ell-1}(G_1, u^j_1, G_2, u^j_2)$. By the induction hypothesis, D thus has a winning strategy in

$$\mathcal{E}^{\mathrm{GML}[c]}_{\ell-1}(c\cdot G_1, (u^j_1, a^j_1), c\cdot G_2, (u^j_2, a^j_2)),$$

where $a^j_i = a_j$ and $a^j_{3-i} = j$. The remaining rounds are played according to that strategy. ❑

Theorem 10. RML $\subseteq$ simple Mean-GNN *in the uniform setting.*

Proof. We use the following step function as the activation function:

$$f(x) = \begin{cases} 1 & \text{if } x > 0 \\ 0 & \text{otherwise.} \end{cases}$$

The proof is then a minor variation of that of Theorem 2, but simpler. We can assume that each diamond uses $>$ since $\Diamond^{\geq t}\varphi \equiv \neg\Diamond^{>(1-t)}\neg\varphi$. We only give the relevant weights:

Case 1: $\varphi_k = P_k$. Set $C_{k,k} = 1$.
Case 2: $\varphi_k = \neg\varphi_i$. Set $C_{i,k} = -1$ and let $b_k = 1$.
Case 3: $\varphi_k = \varphi_i \vee \varphi_j$. Set $C_{i,k} = C_{j,k} = 1$.
Case 4: $\varphi_k = \Diamond^{>t}\varphi_i$. Set $A_{i,k} = 1$ and let $b_k = -t$.

As the classification function (as in Theorem 2) we use

$$\mathsf{CLS}(\bar{x}) = \begin{cases} 1 & \text{if } x_L > 0 \\ 0 & \text{otherwise.} \end{cases}$$

It is not difficult to show correctness of the translation:

Claim. For all φ_k, $1 \leq k \leq L$, the following holds: if $v \in V^G$ and $k \leq k' \leq L$, then

$$(\bar{x}^{k'}_{G,v})_k = \begin{cases} 1 & \text{if } G, v \models \varphi_k, \\ 0 & \text{otherwise.} \end{cases}$$

We leave details to the reader. ❑

D Proofs for Section 5

Proof of Theorem 11

We show that for each ML formula there exists an equivalent Mean-GNN with continuous combination functions and classification function.

$$\mathsf{CLS}(x) = \begin{cases} 1 & \text{if } x \in \mathbb{R}\setminus\mathbb{Q} \\ 0 & \text{otherwise.} \end{cases}$$

In this translation, we represent truth of subformulas by irrational numbers and falsity by rational numbers. We choose $a \in \mathbb{Q}$ and $b \in \mathbb{R} \setminus \mathbb{Q}$ arbitrarily to represent the falsity and truth of atomic formulas. For non-atomic formulas, we are not able to maintain exactly this representation. Instead we represent falsity by rational numbers from the interval I between a and b and truth by values from the set

$$\{p + qb \mid p, q \in \mathbb{Q}, q > 0\} \cap I.$$

We will see that this encoding supports the implementation of disjunction and modal diamonds in a natural way. Implementing negation, however, is less straightforward. We will rely on the following lemma.

Lemma 15. *Let $a \in \mathbb{Q}$ and $b \in \mathbb{R} \setminus \mathbb{Q}$ with $a < b$.*[4] *Let $I = [a, b]$ be the interval between a and b. Let $F = I \cap \mathbb{Q}$ and $T = \{p + qb \mid p, q \in \mathbb{Q},\ q > 0\} \cap I$. Then there exists a continuous function $f : \mathbb{R} \to \mathbb{R}$ such that $f(F) = T$ and $f(T) = F$.*

Proof. We first construct a function with domain and range $F \cup T$ that swaps F and T. After defining the image of a and b, this proof extends Cantor's isomorphism theorem, stating that all countable, dense, linearly ordered sets without maximum and minimum are isomorphic (Cantor 1895). We then show that this function is continuous on its domain and that it can be extended to a continuous function on $\mathbb{R}$.

Since $F \setminus \{a\}$ and $T \setminus \{b\}$ are countable, there exist enumerations $a_1, a_2, \dots$ and $b_1, b_2, \dots$ of $F \setminus \{a\}$ and $T \setminus \{b\}$, respectively. Consider the enumeration $x_1 = a_1, x_2 = b_1, x_3 = a_2, x_4 = b_2, \dots$ of $F \cup T \setminus \{a, b\}$. We define a sequence $f_0, f_1, \dots$ of partial functions from $F \cup T$ and obtain the desired function f in the limit. Let $f_0 = \{(a, b), (b, a)\}$. For f_{i+1}, let k be the smallest number such that x_k is not in the domain of f_i. If $x_k = a_{k'}$, then choose the smallest j such that

- b_j is not in the domain of f_i, and
- for all x in the domain of f_i:
  - if $a_{k'} < x$, then $b_j > f_i(x)$ and
  - if $a_{k'} > x$, then $b_j < f_i(x)$.

Since T is dense, such b_j exists. We then define

$$f_{i+1} = f_i \cup \{(a_{k'}, b_j), (b_j, a_{k'})\}$$

and proceed analogously if $x_k = b_{k'}$. Finally define

$$f = \bigcup_{i \in \mathbb{N}} f_i.$$

We observe the following. Note that Point 2 is the crucial property that we want to attain.

Claim 1.

1. f is a bijection from $F \cup T$ to itself;
2. $f(F) = T$ and $f(T) = F$.

[4] The condition $a < b$ is not necessary, but it simplifies our exposition.

We start with Point 1. It is clear from the definition that f is a function. Also by definition, the domain of each f_i is identical to its range. It follows that each f_i is injective, and thus so is f. It is also surjective: we eventually choose any element of $F \cup T$, and if this happens during the definition of f_i, then the element is in the range of f_i, thus of f.

Point 2 easily follows from Point 1 and the definition of the functions f_i, which only map elements of F to elements of T, and vice versa. This finishes the proof of Claim 1.

To define the extension of f to domain $\mathbb{R}$, we first analyze the images of convergent sequences.

Claim 2.

1. f reverses order: if $x, y \in F \cup T$ with $x < y$, then $f(y) < f(x)$;
2. if $(x_i)_{i \in \mathbb{N}^+}$ and $(y_i)_{i \in \mathbb{N}^+}$ are sequences over $F \cup T$ that converge in $\mathbb{R}$ to the same limit and such that $(f(x_i))_{i \in \mathbb{N}^+}$ and $(f(y_i))_{i \in \mathbb{N}^+}$ converge in $\mathbb{R}$, then $(f(x_i))_{i \in \mathbb{N}^+}$ and $(f(y_i))_{i \in \mathbb{N}^+}$ converge to the same limit.
3. if $(x_i)_{i \in \mathbb{N}^+}$ is a sequence over $F \cup T$ that converges in $\mathbb{R}$, then $(f(x_i))_{i \in \mathbb{N}^+}$ also converges in $\mathbb{R}$;

We prove Point 1 by induction on i. It holds for f_0, since $f_0(a) = b$ and $f_0(b) = a$. Assume that Point 1 holds for f_{i-1} and let $f_i = f_{i-1} \cup \{(c, d), (d, c)\}$, where c the first element in $x_1, x_2, \dots$ that is not in the domain of f_{i-1}. For all x in the domain of f_{i-1} with $c < x$, we have $f_i(x) = f_{i-1}(x) < d = f_i(c)$ by choice of d. The analogous statement hold if $c > x$. We also have to show this point for d. So let x be any element in the domain of f_{i-1}. Since the domain of f_{i-1} is equal to its range, there exists a y such that $x = f_{i-1}(y)$. If $d < x = f_{i-1}(y)$, then $c \geq y$ by definition of f_i. Moreover, $c > y$ since c is a fresh element. Therefore, $f_i(d) = c > y = f_i(x)$ The analogous statement hold if $d > x$. This point also holds if $x, y \in \{c, d\}$ and if x, y are both in the domain of f_{i-1}.

For Point 2, let $(x_i)_{i \in \mathbb{N}^+}$ and $(y_i)_{i \in \mathbb{N}^+}$, $x_i, y_i \in F \cup T$, be two sequences that converge to the same limit. Let $x = \lim_{i \to \infty} f(x_i)$ and $y = \lim_{i \to \infty} f(y_i)$. Assume to the contrary that $x \neq y$. W.l.o.g. let $x < y$. Let $d_1, d_2 \in F \cup T$ such that $x < d_1 < d_2 < y$. Such d_1 and d_2 exists since $F \cup T$ is dense in the interval between a and b. Let c_1 and c_2 such that $f(c_i) = d_i$ for $i \in \{1, 2\}$. Then there exists an index N such that for all $n \geq N$ we have $f(x_n) < d_1 < d_2 < f(y_n)$. By Point 1, $y_n < c_1 < c_2 < x_n$ for each $n \geq N$. Hence, $\lim_{i \to \infty} y_i \leq c_1 < c_2 \leq \lim_{i \to \infty} x_i$. This is a contradiction to $\lim_{i \to \infty} y_i = \lim_{i \to \infty} x_i$. Thus, $f(x_i)$ and $f(y_i)$ converge to the same limit.

To prove Point 3, let $(x_i)_{i \in \mathbb{N}^+}$, $x_i \in F \cup T$, be a sequence that converges to some value $x \in \mathbb{R}$. We split $(x_i)_{i \in \mathbb{N}^+}$ into the subsequence $(z_i^>)$ containing all elements greater than x, the subsequence $(z_i^<)$ containing all elements less than x and the subsequence $(z_i^=)$ containing all elements equal to x. At least one of these is infinite and the limit of the infinite ones is x. We now show for all $\sim \in \{>, <, =\}$ that if $(z_i^\sim)$ is infinite, then $(f(z_i^\sim))_{i \in \mathbb{N}^+}$ converges.

If $(z_i^=)$ is infinite, then $f(z_i^=)$ converges to $f(x)$ since it is constant. We now show the statement for $(z_i^<)$. The proof for $(z_i^>)$ is analogous. Since $(z_i^<)_{i \in \mathbb{N}^+}$ is bounded by a and

b and by the Bolzano-Weierstrass theorem, there exists a monotone subsequence $(z^<_{\ell_i})_{i\in\mathbb{N}}$ which is also convergent to $x \in \mathbb{R}$. Since all $z^<_{\ell_i} < x$, this sequence has to be monotone increasing. By Point 1, $(f(z^<_{\ell_i}))_{i\in\mathbb{N}^+}$ is a monotone decreasing subsequence of $(f(z^<_i))_{i\in\mathbb{N}^+}$ which is bounded by a and b. Therefore, this subsequence is convergent to some $y \in \mathbb{R}$. We now show that $(f(z^<_i))_{i\in\mathbb{N}^+}$ also converges to y. Let $\varepsilon > 0$. Then there exists an N such that $|f(z^<_{\ell_n}) - y| < \varepsilon$ for all $n \geq N$. Since $(z^<_i)_{i\in\mathbb{N}^+}$ converges to x there has to be an M such that $z^<_{\ell_N} \leq z^<_m < x$ for all $m \geq M$. For each $m \geq M$ we can find an index m' such that $z^<_{m'}$ is part of $(z^<_{\ell_i})_{i\in\mathbb{N}^+}$ and is closer to x, that is $z^<_{\ell_N} \leq z^<_m \leq z^<_{m'} < x$. By Point 1, $f(z^<_{m'}) < f(z^<_m) \leq f(z_{\ell_N})$. And since $(f(z^<_i))_{i\in\mathbb{N}^+}$ is monotone decreasing, we have $y \leq f(z^<_{m'})$ and thus $|f(z^<_m) - y| \leq |f(z^<_{\ell_N}) - y| < \varepsilon$. Thus, $(f(z^<_i))_{i\in\mathbb{N}^+}$ converges to y.

It follows from Point 2 that if two or all three of $(f(z^<_i))$, $(f(z^>_i))$ and $(f(z^=_i))$ are infinite, then they converge to the same limit y. We show that if all three are infinite, then $(f(x_i))_{i\in\mathbb{N}^+}$ also converges. If only two of them converge, we can argue analogously. If all three are infinite, then for each $\varepsilon > 0$ we find indices $N^<, N^>, N^=$ and $N = \max(N^<, N^>, N^=)$ such that for all $n \geq N$ and $\sim \in \{<, >, =\}$, $|f(z^\sim_n) - y| < \varepsilon$. Hence, we can find an N' such that for all $n \geq N'$ we have $|f(x_n) - y| < \varepsilon$. This finishes the proof of Claim 2.

We now define a continuous extension $\hat{f}$ of f to domain $\mathbb{R}$. We remind the reader that $a < b$.

Start with setting

- $\hat{f}(x) = b$ for all $x < a$;
- $\hat{f}(x) = a$ for all $x > b$;

Since $f(a) = b$ and $f(b) = a$, it immediately follows that $\hat{f}$ is continuous at all those x.

It remains to define $\hat{f}$ for all $x \in I = (a, b)$. Since $(F \cup T) \setminus \{a, b\}$ is dense in I, there exists a sequence $(x_i)_{i\in\mathbb{N}^+}$ with limit x such that $x_i \in (F \cup T) \setminus \{a, b\}$ for all $i \in \mathbb{N}^+$. We define

$$\hat{f}(x) = \lim_{i\to\infty} f(x_i).$$

This limit exists by Claim 2.3 and is well-defined by Claim 2.2. $\hat{f}$ is an extension of f, that is $\hat{f}(x) = f(x)$ for all $x \in F \cup T$, because $f(x) = \lim_{i\to\infty} f(x) = \hat{f}(x)$.

We have to show that $\hat{f}$ is continuous at all elements of $[a, b]$.

We first observe that the following is an easy consequence of Claim 2.1 and the definition of $\hat{f}$.

Claim 3. $\hat{f}$ reverses order: if $x < y$ then $\hat{f}(y) \leq \hat{f}(x)$.

Claim 4. $\hat{f}$ is continuous on $[a, b]$.

We show the claim for all $x \in I$ first. Let $(x_i)_{i\in\mathbb{N}^+}$ be a sequence with limit x. We have to show that $\hat{f}(x) = \lim_{i\to\infty} \hat{f}(x_i)$.

Since $x \in I$ and I is an open interval, there exists an index N such that $x_n \in I$ for all $n \geq N$. Thus, we can assume w.l.o.g that every x_i is in I. Since I is an open interval and $(F \cup T) \setminus \{a, b\}$ is dense in I, there exist upper and lower bounds of $(x_i)_{i\in\mathbb{N}^+}$ in $(F \cup T) \setminus \{a, b\}$. That is, there exist sequences $(y_i)_{i\in\mathbb{N}^+}$ and $(z_i)_{i\in\mathbb{N}^+}$ such that $y_i, z_i \in (F \cup T) \setminus \{a, b\}$, both converge to x, and $y_i \leq x_i \leq z_i$ for all $i \in \mathbb{N}^+$. By Claim 3, the image of $(x_i)_{i\in\mathbb{N}^+}$ is also bounded by the images of $(y_i)_{i\in\mathbb{N}^+}$ and $(z_i)_{i\in\mathbb{N}^+}$. Since the order is reversed, $f(z_i) = \hat{f}(z_i) \leq \hat{f}(x_i) \leq \hat{f}(y_i) = f(y_i)$. By definition of $\hat{f}$, we have $\hat{f}(x) = \lim_{i\to\infty} f(z_i) = \lim_{i\to\infty} f(y_i)$, and by the squeeze theorem we also have $\hat{f}(x) = \lim_{i\to\infty} \hat{f}(x_i)$. It remains to argue that $\hat{f}$ is continuous at a and b. We show that $\hat{f}$ is continuous at a; the argument that $\hat{f}$ is continuous at b is analogous. Let $(x_i)_{i\in\mathbb{N}^+}$ be a sequence with limit a. We have to show that $\lim_{i\to\infty} f(x_i) = b$. There can only be finitely many i such that $x_i > b$, hence these indices can be ignored without changing convergence. We can also safely ignore all $x_i < a$, since they satisfy $\hat{f}(x_i) = b$. If there are infinitely many i left such that $x_i \in I$, we can argue analogously to the previous case $x \in I$ to prove $\lim_{i\to\infty} \hat{f}(x_i) = b$. If there are only finitely many i such that $x_i \in I$, then all $x_i < a$ define the limit of $(f(x_i))_{i\in\mathbb{N}^+}$ which then has to be b.

❑

Theorem 17. *In the uniform setting when we allow arbitrary classification functions:*

$$\text{ML} \subseteq \text{Mean}^c\text{-GNN}.$$

Proof. Let φ be an ML formula over a finite set $\Pi = \{P_1, \ldots, P_r\}$ of vertex labels. Let $\varphi_1, \ldots, \varphi_L$ be an enumeration of the subformulas of φ such that (i) $\varphi_i = P_i$ for $1 \leq i \leq r$ and (ii) if φ_ℓ is a subformula of φ_k then $\ell < k$.

We construct a GNN $\mathcal{G}$ with $L + 1$ layers, all of output dimension L. Choose a rational number a and an irrational number b such that $a < b$, and let f be the function constructed for these choices of a and b in Lemma 15. The purpose of the first layer is to convert the initial feature vector into the intended format where truth is represented by irrational numbers and falsity by rational numbers, as explained at the beginning of this section. For the first layer, the k-th entry of $\mathsf{COM}^1(\bar{x}, \bar{y})$ is thus defined as follows:

Case 1: $\varphi_k = P_k$. Return $(b - a)x_k + a$

Case 2: Otherwise return a.

Since $(b - a)x_k + a$ is continuous, COM^1 is continuous.

For all layers $\ell \geq 2$, we use a different combination function. The k-th entry of $\mathsf{COM}^\ell(\bar{x}, \bar{y})$ is defined as follows:

Case 1: $\varphi_k = P_k$. Return x_k.

Case 2: $\varphi_k = \neg\varphi_i$. Return $f(x_i)$.

Case 3: $\varphi_k = \varphi_i \vee \varphi_j$. Return $\frac{x_i + x_j}{2}$.

Case 4: $\varphi_k = \Diamond\varphi_i$. Return y_i.

Since each of these functions is continuous, COM^ℓ is continuous. $\mathcal{G}$ uses the following classification function:

$$\mathsf{CLS}(x) = \begin{cases} 1 & \text{if } x \in \mathbb{R} \setminus \mathbb{Q}, \\ 0 & \text{otherwise.} \end{cases}$$

Let $I = [a, b]$, $F = \mathbb{Q} \cap I$ and $T = \{p + qb \mid p, q \in \mathbb{Q}, q > 0\} \cap I$. We now show the correctness of our construction.

Claim. For all φ_k, $1 \leq k \leq L$, the following holds: if $v \in V^G$ and $k \leq k' \leq L$, then

$$(\bar{x}^{k'}_{G,v})_k \in T \text{ if } G, v \models \varphi_k, \text{ and}$$
$$(\bar{x}^{k'}_{G,v})_k \in F \text{ otherwise.}$$

The proof of the claim is by induction on k.

Case 1. For $1 \leq k \leq r$, $\varphi_k = P_k$ is an atomic formula. If $G, v \models \varphi_k$, then $(\bar{x}^0_{G,v})_k = 1$ and the first layer returns

$$(\bar{x}^1_{G,v})_k = (b-a) \cdot 1 + a = b \in T.$$

All subsequent layers ℓ return $(\bar{x}^\ell_{G,v})_k = b \in T$. If $G, v \not\models \varphi_k$, then $(\bar{x}^0_{G,v})_k = 0$, the first layer returns

$$(\bar{x}^1_{G,v})_k = (b-a) \cdot 0 + a = a \in F,$$

and for all subsequent layers ℓ, we have $(\bar{x}^\ell_{G,v})_k = a \in F$.

Case 2. Let $\varphi_k = \neg\varphi_i$. If $G, v \models \neg\varphi_i$, then since $i < k$ the induction hypothesis yields $(\bar{x}^{k'-1}_{G,v})_i \in F$ and thus $(\bar{x}^{k'}_{G,v})_k \in T$ by Lemma 15. If $G, v \not\models \neg\varphi_i$, then $(\bar{x}^{k'-1}_{G,v})_i \in T$ and thus $(\bar{x}^{k'}_{G,v})_k \in F$.

Case 3. Let $\varphi_k = \varphi_i \vee \varphi_j$. If $G, v \models \varphi_k$, then at least one of $(\bar{x}^{k'-1}_{G,v})_i$ and $(\bar{x}^{k'-1}_{G,v})_j$ is in T. Since the definition of T requires $q > 0$,

$$\frac{(\bar{x}^{k'-1}_{G,v})_i + (\bar{x}^{k'-1}_{G,v})_j}{2} \qquad (*)$$

can be written as $p' + q'b$ where $p', q' \in \mathbb{Q}$ and $q' > 0$. If $G, v \not\models \varphi_k$, then both $(\bar{x}^{k'-1}_{G,v})_i$ and $(\bar{x}^{k'-1}_{G,v})_j$ are in F, thus both are rational numbers. Therefore, $(*)$ is also rational, and it is easy to verify that it is in F.

Case 4. Let $\varphi_k = \Diamond\varphi_i$. First assume that $G, v \models \varphi_k$. Let $\mathcal{N}(G, v) = \{u_1, \ldots, u_n\}$. For $1 \leq j \leq n$, there exist $p_j, q_j \in \mathbb{Q}$ such that $(\bar{x}^{k'-1}_{G,u_j})_i = p_j + q_j b$. Then

$$\begin{aligned} &\mathsf{MEAN}\{\{(\bar{x}^{k'-1}_{G,u_j})_i \mid 1 \leq j \leq n\}\} \\ &= \mathsf{MEAN}\{\{p_j \mid 1 \leq j \leq n\}\} + \mathsf{MEAN}\{\{q_j \mid 1 \leq j \leq n\}\} b. \end{aligned}$$

By induction hypothesis, $q_j \geq 0$ for all j and $q_j > 0$ for at least one j. Thus, the above can be written as $p + qb$ where $p, q \in \mathbb{Q}$ and $q > 0$. Thus $(\bar{x}^{k'}_{G,v})_k \in T$, as required. If $G, v \not\models \varphi_k$, then $q_j = 0$ for all j, and thus $(\bar{x}^{k'-1}_{G,u_j})_i$ is a rational number. It is easy to verify that it is also in F. ❏

Relations of AFML to EF games and ML

Theorem 18. *Let P be a vertex property and let $k \in \{1, 2\}$. The following are equivalent for all $\ell \geq 0$:*

1. *there exists an* AFML$[k]$ *formula φ of modal depth at most ℓ such that for all pointed graphs (G, v): $G, v \models \varphi$ if and only if $(G, v) \in P$.*
2. *Spoiler has a winning strategy in $\mathcal{E}^{\mathrm{AFML}[k]}_\ell(G_1, v_1, G_2, v_2)$ for all pointed graphs $(G_1, v_1), (G_2, v_2)$ with $(G_1, v_1) \in P$ and $(G_2, v_2) \notin P$.*

We start with proving a basic connection between AFML logic and AFML games. Note that the subsequent theorem implies Theorem 18 because up to equivalence, there are only finitely many AFML formulas of any fixed modal depth ℓ.

Theorem 19. *Let $k \in \{1, 2\}$. The following are equivalent for all possibly infinite pointed graphs (G_1, v_1) and (G_2, v_2) and all $\ell \geq 0$:*

1. *there is an* AFML$[k]$ *formula φ of modal depth at most ℓ such that $G_1, v_1 \models \varphi$ and $G_2, v_2 \not\models \varphi$,*
2. *Spoiler has a winning strategy in $\mathcal{E}^{\mathrm{AFML}[k]}_\ell(G_1, v_1, G_2, v_2)$.*

Proof. We start with the proof for AFML$[1]$.

"$1 \Rightarrow 2$". The proof is by induction on φ. For the induction start, there are three cases. First let $\varphi = P$ or $\varphi = \neg P$ and assume that $G_1, v_1 \models \varphi$ and $G_2, v_2 \not\models \varphi$. Thus, $\pi_1(v_1) \neq \pi_2(v_2)$ and Spoiler wins $\mathcal{E}^{\mathrm{AFML}[1]}_0(G_1, v_1, G_2, v_2)$. The third case is $\varphi = \Box\bot$. Assume that $G_1, v_1 \models \varphi$ and $G_2, v_2 \not\models \varphi$. Then v_1 has no successors and by the definition of AFML$[1]$-games Spoiler is allowed to choose a $u_2 \in \mathcal{N}(v_2)$, which exists because v_2 does not satisfy φ. Duplicator has no response in $\mathcal{N}(v_1) = \emptyset$, thus S has a winning strategy for $\mathcal{E}^{\mathrm{AFML}[1]}_1(G_1, v_1, G_2, v_2)$—note that the modal depth of $\Box\bot$ is 1.

For the induction step, assume that the statement holds for ψ_1 and ψ_2. We have to show that the statement also holds for $\psi_1 \wedge \psi_2$, $\psi_1 \vee \psi_2$ and $\Diamond\psi_1$. Let $\varphi = \psi_1 \wedge \psi_2$ and assume that $G_1, v_1 \models \varphi$ and $G_2, v_2 \not\models \varphi$. Then $G_2, v_2 \not\models \psi_k$ for some $k \in \{1, 2\}$ and by induction hypothesis, Spoiler has a winning strategy for $\mathcal{E}^{\mathrm{AFML}[1]}_{\ell_k}(G_1, v_1, G_2, v_2)$ where ℓ_k is the modal depth of ψ_k. Since the modal depth ℓ of φ satisfies $\ell \geq \ell_k$, S also has a winning strategy for $\mathcal{E}^{\mathrm{AFML}[1]}_\ell(G_1, v_1, G_2, v_2)$. The case $\varphi = \psi_1 \vee \psi_2$ is analogous.

Now let $\varphi = \Diamond\psi$ with ψ of modal depth ℓ and assume that $G_1, v_1 \models \varphi$ and $G_2, v_2 \not\models \varphi$. Then there exists a $u_1 \in \mathcal{N}(v_1)$ such that $G_1, u_1 \models \psi$, and for all $u_2 \in \mathcal{N}(v_2)$ we have $G_2, u_2 \not\models \psi$. By induction hypothesis, Spoiler has a winning strategy for $\mathcal{E}^{\mathrm{AFML}[1]}_\ell(G_1, u_1, G_2, u_2)$ for all $u_2 \in \mathcal{N}(v_2)$. We can extend this to a winning strategy for $\mathcal{E}^{\mathrm{AFML}[1]}_{\ell+1}(G_1, v_1, G_2, v_2)$, by letting S choose u_1 in the first round.

"$2 \Rightarrow 1$". The proof is by induction on ℓ. For the induction start, assume that Spoiler wins $\mathcal{E}^{\mathrm{AFML}[1]}_0(G_1, v_1, G_2, v_2)$. Then the vertex labels of v_1 and v_2 differ. We thus find an AFML$[1]$ formula φ of the form P or $\neg P$ such that $G_1, v_1 \models \varphi$ and $G_2, v_2 \not\models \varphi$.

For the induction step, let the statement hold for $\ell - 1 \geq 0$ and assume that Spoiler wins $\mathcal{E}^{\mathrm{AFML}[1]}_\ell(G_1, v_1, G_2, v_2)$. There are two ways in which this may happen. In case S wins because they choose $u_2 \in \mathcal{N}(v_2)$ in G_2 and D cannot respond with a $u_1 \in \mathcal{N}(v_1)$, we have $G_1, v_1 \models \Box\bot$ while $G_2, v_2 \not\models \Box\bot$. Therefore, $\Box\bot$ is an AFML$[1]$ formula that distinguishes v_1 and v_2. It has modal depth $1 \leq \ell$.

The second case is that S chooses $u_1 \in \mathcal{N}(v_1)$ in G_1 and for every response $u_2 \in \mathcal{N}(v_2)$ that D may choose, S has a winning strategy for the game $\mathcal{E}_{\ell-1}^{\text{AFML}[1]}(G_1, u_1, G_2, u_2)$. By induction hypothesis, for each $u_2 \in \mathcal{N}(v_2)$ there exists a formula ψ_{u_2} of modal depth at most $\ell - 1$ such that $G_1, u_1 \models \psi_{u_2}$ and $G_2, u_2 \not\models \psi_{u_2}$. But then $\varphi = \Diamond(\bigwedge_{u_2 \in \mathcal{N}(v_2)} \psi_{u_2})$ is an AFML[1] formula of modal depth ℓ such that $G_1, v_1 \models \varphi$ and $G_2, v_2 \not\models \varphi$ (we take the empty conjunction to be $\top$).

For AFML[2], the result follows from the AFML[1] case and the observations that (i) an EF-game for AFML[2] is exactly an EF-game for AFML[1] with the roles of (G_1, v_1) and (G_2, v_2) swapped and (ii) every AFML[1] formula is equivalent to the complement of an AFML[2] formula and vice versa. ❏

Lemma 16. AFML $\subsetneq$ ML.

Proof. We show that $\varphi = \Diamond P \wedge \Box Q$ is not equivalent to an AFML formula. We use the following three graphs:

- $V^A = \{a, b, c\}$,
  $E^A = \{(a, b), (a, c)\}$,
  $\pi^A(a) = \emptyset$, $\pi^A(b) = \{P, Q\}$ and $\pi^A(c) = \{Q\}$.
- $V^{B_1} = \{a, b, c, d\}$,
  $E^{B_1} = \{(a, b), (a, c), (a, d)\}$,
  $\pi^{B_1}(a) = \pi^{B_1}(d) = \emptyset$, $\pi^{B_1}(b) = \{P, Q\}$ and $\pi^{B_1}(c) = \{Q\}$.
- $V^{B_2} = \{a, c\}$,
  $E^{B_2} = \{(a, c)\}$,
  $\pi^{B_2}(a) = \emptyset$ and $\pi^{B_2}(c) = \{Q\}$.

We have $A, a \models \varphi$, $B_1, a \not\models \varphi$ and $B_2, a \not\models \varphi$.

We now show that for each $\ell \in \mathbb{N}^+$ Duplicator wins $\mathcal{E}_\ell^{\text{AFML}[1]}(A, a, B_1, a)$ as well as $\mathcal{E}_\ell^{\text{AFML}[2]}(A, a, B_2, a)$.

In $\mathcal{E}_\ell^{\text{AFML}[1]}(A, a, B_1, a)$, Spoiler does not win with zero rounds, because both vertices a satisfy $\pi^A(a) = \pi^{B_1}(a) = \emptyset$. In the first round S can choose $b \in V^A$ or $c \in V^A$. Duplicator can answer with $b \in V^{B_1}$ and $c \in V^{B_1}$ respectively. Both b satisfy $\pi(b) = \{P, Q\}$, and both c satisfy $\pi(c) = \{Q\}$, so Spoiler does not win in the first round. Additionally, each of these vertices does not have successors, hence S has no vertex they can play in the second round. Hence, D will win in the next round. This implies, that D also has a winning strategy for all $\ell \geq 2$.

In $\mathcal{E}_\ell^{\text{AFML}[2]}(A, a, B_2, a)$, Spoiler does not win with zero rounds, because both vertices a satisfy $\pi^A(a) = \pi^{B_2}(a) = \emptyset$. In the first round, S has to choose $c \in V^{B_2}$ and Duplicator can respond with c in V^A, which has the same labeling. Since both vertices c do not have successors, D wins in the second round. This also means that D will win the EF-game for all $\ell \geq 2$. ❏

Proof of Theorem 13

Lemma 17. *For each* Mean^c-GNN $\mathcal{G}$ *with L layers, there exist m_-, $m_+ \in \mathbb{R}$ such that for all input graph G, vertices $v \in V^G$, and $\ell \in [L]$, the vector $\bar{x}_{G,v}^\ell$ only contains values from the range $[m_-, m_+]$.*

Proof. We identify sequences $m_-^0, m_-^1, \ldots, m_-^L$ and $m_+^0, m_+^1, \ldots, m_+^L$ such that for all $\ell \in [L]$, the vector $\bar{x}_{G,v}^\ell$ only contains values from the range $[m_-^\ell, m_+^\ell]$. We may then set m_- to the minimum of all m_-^ℓ and m_+ to the maximum of all m_+^ℓ.

We proceed by induction on ℓ. To start, we may set $m_-^0 = 0$ and $m_+^0 = 1$ because the initial feature vectors only contain the values 0 and 1.

Now let $\ell > 0$ and let $\delta^{\ell-1}$ and δ^ℓ be the output dimensions of layers $\ell - 1$ and ℓ. Then the feature vectors that get computed by MEAN in layer ℓ are in $[m_-^{\ell-1}, m_+^{\ell-1}]^{\delta^{\ell-1}}$ and COM gets as input a vector in $[m_-^{\ell-1}, m_+^{\ell-1}]^{2\delta^{\ell-1}}$. By the Heine-Borel theorem, $[m_-^{\ell-1}, m_+^{\ell-1}]^{2\delta^{\ell-1}}$ is a compact subset of $\mathbb{R}^{\delta^\ell}$. Since COM is continuous and it is well-known that the image of a compact set under a continuous function is compact, $\mathsf{COM}([m_-^{\ell-1}, m_+^{\ell-1}]^{2\delta^{\ell-1}})$ is also compact. Once more applying Heine-Borel, we obtain m_-^ℓ and m_+^ℓ such that $\mathsf{COM}([m_-^{\ell-1}, m_+^{\ell-1}]^{2\delta^{\ell-1}}) \subseteq [m_-^\ell, m_+^\ell]^{\delta^\ell}$. ❏

Lemma 8. *Let $\mathcal{G}$ be a* Mean^c-GNN *with L layers. Then for all $\varepsilon > 0$, $n \geq 1$, and $\ell \in [L]$, there exists a constant c such that for all $c' \geq c$, graphs $H = c' \cdot G$, vertices (v, i) in H, and n-extensions H' of H: $||\bar{x}_{H,(v,i)}^\ell - \bar{x}_{H',(v,i)}^\ell||_\infty < \varepsilon$.*

Proof. The proof is by induction on ℓ. Let

$$\mathcal{G} = (L, \{\mathsf{AGG}^\ell\}_{\ell \in [L]}, \{\mathsf{COM}^\ell\}_{\ell \in [L]}, \mathsf{CLS})$$

be a Mean^c-GNN with L layers, let $\varepsilon > 0$, $n \geq 1$, and $\ell \in [L]$.

In the induction start, where $\ell = 0$, we can choose $c = 1$ no matter what ε is because each vertex (v, i) in a graph H has the same vertex labels in H and any n-extension H' of H, and thus $\bar{x}_{H,(v,i)}^0 = \bar{x}_{H',(v,i)}^0$.

For the induction step, where $\ell > 0$, let m_-, m_+ be the values from Lemma 17. Let γ be the output dimension of layer $\ell - 1$ which is the input dimension of layer ℓ. By the Heine-Cantor theorem and since $[m_-, m_+]^{2\gamma} \subseteq \mathbb{R}^{2\gamma}$ is compact and COM^ℓ is continuous, the restriction of COM^ℓ to domain $[m_-, m_+]^{2\gamma}$ is uniformly continuous; note that for convenience here we view COM^ℓ as a function with a single input vector of dimension 2γ rather than with two input vectors of dimension γ. There is thus a value δ such that for all $\bar{x}, \bar{y} \in [m_-, m_+]^{2\gamma}$ with $||\bar{x} - \bar{y}||_\infty < \delta$ we have $||\mathsf{COM}^\ell(\bar{x}) - \mathsf{COM}^\ell(\bar{y})||_\infty < \varepsilon$. Let $c^{\ell-1}$ be the constant obtained in the induction hypothesis for $\varepsilon = \delta/5$ and the same value of n.

Let $m = \max(|m_-|, |m_+|)$ and choose c^* such that

$$\frac{mn}{c^*} \leq \frac{\delta}{5} \quad \text{and} \quad c^* \geq n.$$

Set $c = \max(c^{\ell-1}, c^*)$.

Now consider any $c' \geq c$, graph $H = c' \cdot G$, and n-extension H' of H. Let (v, i) be a vertex in H. Then

$$\bar{x}_{H',(v,i)}^\ell = \mathsf{COM}^\ell(\bar{x}_{H',(v,i)}^{\ell-1}, \bar{z}_{H',(v,i)})$$

where $\bar{z}_{H',(v,i)}$ is the result of mean aggregation, and likewise for $\bar{x}^{\ell}_{H,(v,i)}$.

Since we are working with the maximum metric and by choice of δ, to prove that

$$||\bar{x}^{\ell}_{H',(v,i)} - \bar{x}^{\ell}_{H,(v,i)}||_\infty < \varepsilon,$$

we have to show that $||\bar{x}^{\ell-1}_{H',(v,i)} - \bar{x}^{\ell-1}_{H,(v,i)}||_\infty < \delta$ and $||\bar{z}_{H',(v,i)} - \bar{z}_{H,(v,i)}||_\infty < \delta$.

The first inequality holds by induction hypothesis since c' is larger or equals to $c^{\ell-1}$.

We now show the second inequality. First assume that (v,i) has at least one successor in H. Because H is a c' scaled graph, (v,i) has at least c' and thus at least c successors. Let $N = |\mathcal{N}(H,(v,i))| \geq c$. Moreover,

$$\begin{aligned}\bar{z}_{H',(v,i)} &= \mathsf{MEAN}(\{\{\bar{x}^{\ell-1}_{H',(u,j)} \mid (u,j) \in \mathcal{N}(H',(v,i))\}\}) \\ &= \frac{1}{|\mathcal{N}(H',(v,i))|}\left(\sum_{(u,j)\in\mathcal{N}(H',(v,i))} \bar{x}^{\ell-1}_{H',(u,j)}\right).\end{aligned}$$

Let

$$\begin{aligned}\Delta_1 &= \frac{1}{|\mathcal{N}(H',(v,i))|} - \frac{1}{|\mathcal{N}(H,(v,i))|} \\ &= \frac{1}{|\mathcal{N}(H',(v,i))|} - \frac{1}{N}\end{aligned}$$

and

$$\Delta_2 = \sum_{(u,j)\in\mathcal{N}(H',(v,i))} \bar{x}^{\ell-1}_{H',(u,j)} - \sum_{(u,j)\in\mathcal{N}(H,(v,i))} \bar{x}^{\ell-1}_{H,(u,j)}.$$

We thus can write $\bar{z}_{H',(v,i)}$ as

$$\left(\Delta_1 + \frac{1}{N}\right)\left(\Delta_2 + \sum_{(u,j)\in\mathcal{N}(H,(v,i))} \bar{x}^{\ell-1}_{H,(u,j)}\right).$$

The difference between $\bar{z}_{H',(v,i)}$ and $\bar{z}_{H,(v,i)}$ is thus

$$\Delta_1\left(\sum_{(u,j)\in\mathcal{N}(H,(v,i))} \bar{x}^{\ell-1}_{H,(u,j)}\right) + \frac{\Delta_2}{N} + \Delta_1\Delta_2.$$

By the triangle inequality follows that $||\bar{z}_{H',(v,i)} - \bar{z}_{H,(v,i)}||_\infty$ is bounded by

$$|\Delta_1|\left|\left|\sum_{(u,j)\in\mathcal{N}(H,(v,i))} \bar{x}^{\ell-1}_{H,(u,j)}\right|\right|_\infty + \frac{||\Delta_2||_\infty}{N} + |\Delta_1|\cdot||\Delta_2||_\infty.$$

We bound each term as follows:

- $|\Delta_1|$ is bounded by $\frac{1}{N} - \frac{1}{N+n}$ since H' is an n-extension of H. We have $\frac{1}{N} - \frac{1}{N+n} = \frac{n}{N(N+n)} \leq \frac{n}{N^2}$.
- $||\Delta_2||_\infty$ is bounded by $nm + N\frac{\delta}{5}$. In H' there can be up to n new neighbors which add a difference of up to m each. And for each of the N successors in H, $||\bar{x}^{\ell-1}_{H',(u,j)} - \bar{x}^{\ell-1}_{H,(u,j)}||_\infty$ is bounded by $\frac{\delta}{5}$ by choice of $c^{\ell-1}$.
- $\left|\left|\sum_{(u,j)\in\mathcal{N}(H,(v,i))} \bar{x}^{\ell-1}_{H,(u,j)}\right|\right|_\infty$ is bounded by Nm since each entry is a sum that adds N terms which are all bounded by m.

Thus, $||\bar{z}_{H',(v,i)} - \bar{z}_{H,(v,i)}||_\infty$ can be bounded by

$$\frac{Nmn}{N^2} + \frac{1}{N}(nm + N\frac{\delta}{5}) + \frac{n}{N^2}\cdot(nm + N\frac{\delta}{5}).$$

It can now be verified easily that this sum is bounded by δ since $N \geq c^* \geq n$ and

$$\frac{mn}{N} \leq \frac{mn}{c^*} \leq \frac{\delta}{5}.$$

Now assume that (v,i) has no successors in H. Then $\mathcal{N}(H',(v,i)) = \emptyset = \mathcal{N}(H,(v,i))$. Thus,

$$\bar{z}_{H',(v,i)} = \bar{z}_{H,(v,i)}.$$

Therefore, for all vertices (v,i) in H, we have $||\bar{x}^{\ell-1}_{H',(v,i)} - \bar{x}^{\ell-1}_{H,(v,i)}||_\infty < \delta$ and $||\bar{z}^{\ell-1}_{H',(v,i)} - \bar{z}^{\ell-1}_{H,(v,i)}||_\infty < \delta$. And by choice of δ we have

$$\begin{aligned}&||\bar{x}^{\ell}_{H',(v,i)} - \bar{x}^{\ell}_{H,(v,i)}|| = \\ &||\mathsf{COM}^{\ell}(\bar{x}^{\ell-1}_{H',(v,i)}, \bar{z}^{\ell-1}_{H',(v,i)}) - \mathsf{COM}^{\ell}(\bar{x}^{\ell-1}_{H,(v,i)}, \bar{z}^{\ell-1}_{H,(v,i)})||_\infty \\ &< \varepsilon.\end{aligned}$$

❑

At this point, we have everything we need to prove Theorem 13.

Theorem 13. Mean^{c,t}-GNN$\cap$MSO $\subseteq$ AFML *in the uniform setting.*

Proof. Assume to the contrary that there exists a Mean^{c,t}-GNN $\mathcal{G}$ with L layers that is equivalent to an MSO formula, but not to an AFML formula. By Corollary 3, $\mathcal{G}$ is equivalent to an ML formula φ. Let

$$\mathsf{CLS}(\bar{x}) = \begin{cases} 1 & \text{if } x_i \sim c_0, \\ 0 & \text{otherwise} \end{cases}$$

be the classification function of $\mathcal{G}$, where $\sim \in \{>, \geq\}$.

First assume that $\sim$ is $>$. Because φ is not expressible in AFML, for each $\ell \in \mathbb{N}$ there exist pointed graphs (G,v) and (G',v') with $G,v \models \varphi$ and $G',v' \not\models \varphi$ such that D has a winning strategy in $\mathcal{E}^{\mathrm{AFML}[1]}_{\ell}(G,v,G',v')$. We show below how to transform (G,v) into a pointed graph (H,u) such that $H,u \models \varphi$ and D has a winning strategy in $\mathcal{E}^{\mathrm{ML}}_{\ell}(H,u,G',v')$. Thus, by Theorem 8, φ is not expressible in ML. A contradiction.

The transformation of (G,v) into (H,v) crucially relies on the observation that, due to the equivalence of φ to $\mathcal{G}$ and since $\sim$ is $>$, the following holds:

($*_1$) for each pointed graph (G,v) accepted by $\mathcal{G}$, there is an $\varepsilon > 0$ such that all pointed graphs (H,u) with $||\bar{x}^L_v - \bar{x}^L_u||_\infty < \varepsilon$ are also accepted by $\mathcal{G}$.

In fact, we may simply choose $\varepsilon = (\bar{x}_v^L)_i - c_0$.

Assume that D has a winning strategy in $\mathcal{E}_\ell^{\mathrm{AFML}[1]}(G, v, G', v')$. Then for all $K \geq \ell$, D also has a winning strategy in $\mathcal{E}_\ell^{\mathrm{AFML}[1]}(\mathsf{Unr}^K(G, v), v, G', v')$, as a consequence of Theorem 18 and since AFML formulas of modal depth ℓ are invariant under unraveling up to depth $K \geq \ell$. Since $\mathsf{Unr}^K(G, v)$ is tree-shaped, the only way for Spoiler to play a vertex $u = v_1 \cdots v_n$ is to play exactly the vertices $v_1, v_1v_2, \ldots, u$ on the unique path from the root $v = v_1$ to u. Once this has happened, S can never play u again. As a consequence, the winning strategy of D in $\mathcal{E}_\ell^{\mathrm{AFML}[1]}(\mathsf{Unr}^K(G, v), v, G', v')$ may be viewed as a function $\mathsf{ws} : V^{\mathsf{Unr}^K(G,v)} \to V^{G'}$ such that if S plays vertex u in $\mathsf{Unr}^K(G, v)$, then D answers with $\mathsf{ws}(u)$. We also set $\mathsf{ws}(v) = v'$.

Let $m = |V^{G'}|$ and $K = \max(\ell, L)$. By Lemma 8, there exists a c such that in each m-extension X of $G'' = c \cdot \mathsf{Unr}^K(G, v)$, we have $||\bar{x}_{G'',v}^L - \bar{x}_{X,v}^L||_\infty < \varepsilon$ with ε the constant from $(*_1)$.

The pointed graph (H, u) is defined as follows:

1. start with $G'' = c \cdot \mathsf{Unr}^K(G, v)$;
2. take the disjoint union with all $\mathsf{Unr}^K(G', v')$, $v' \in V^{G'}$;
3. for each vertex $(u, i) \in V^{G''}$ that has at least one successor, let $\mathcal{N}(G', \mathsf{ws}(u)) = \{u'_1, \ldots, u'_m\}$. Add to (u, i) the fresh successors $u'_1, \ldots, u'_m$ (which are all roots of unravelings added in Step 2);
4. $u = (v, 1)$.

Claim 1. $H, u \models \varphi$.

Since $G, v \models \varphi$, (G, v) is accepted by $\mathcal{G}$. By Lemmas 7 and 5, $\bar{x}_{G,v}^L = \bar{x}_{G'',v}^L$. By choice of c and since H is an m-extension of G'', we have $||\bar{x}_{G,v}^L - \bar{x}_{H,u}^L||_\infty < \varepsilon$. Consequently, it follows from $(*_1)$ that (H, u) is also accepted by $\mathcal{G}$ and therefore $H, u \models \varphi$, as desired.

Claim 2. D has a winning strategy in $\mathcal{E}_\ell^{\mathrm{ML}}(H, u, G', v')$.

To prove Claim 2, we show that for each $k \leq \ell$, the following holds:

(i) D has a winning strategy in $\mathcal{E}_{K-k}^{\mathrm{ML}}(H, v'_1 \cdots v'_k, G', v'_k)$, for each $v'_1 \cdots v'_k \in V^{\mathsf{Unr}^K(G',v'_1)}$.

(ii) If D has a winning strategy in
$$\mathcal{E}_k^{\mathrm{AFML}[1]}(\mathsf{Unr}^K(G, v_1), v_1 \cdots v_n, G', \mathsf{ws}(v_1 \cdots v_n)),$$
then for all i, D has a winning strategy in
$$\mathcal{E}_k^{\mathrm{ML}}(H, (v_1 \cdots v_n, i), G', \mathsf{ws}(v_1 \cdots v_n)).$$

Claim 2 follows from Point (ii) since D has a winning strategy in $\mathcal{E}_\ell^{\mathrm{AFML}[1]}(\mathsf{Unr}^K(G, v), v, G', v')$.

Point (i) is almost immediate. By Theorem 8 and since modal formulas of depth $K - k$ are invariant under unraveling up to depth K, D has a winning strategy in $\mathcal{E}_{K-k}^{\mathrm{ML}}(\mathsf{Unr}^K(G', v'_1), v'_1 \cdots v'_k, G', v'_k)$. Point (i) follows since in H there is no edge that leaves the disconnected component $\mathsf{Unr}^K(G', v'_1)$.

It thus remains to prove Point (ii), which we do by induction on k. For $k = 0$ notice that $v_1 \cdots v_n$ and $\mathsf{ws}(v_1 \cdots v_n)$ have the same vertex labels because D wins the AFML[1]-game. By definition of H, $(v_1 \cdots v_n, i)$ also has the same labels and thus D wins the 0-round ML-game.

For the induction step, let (ii) hold for $k - 1 \geq 0$ and assume that D wins
$$\mathcal{E}_k^{\mathrm{AFML}[1]}(\mathsf{Unr}^K(G, v_1), v_1 \cdots v_n, G', \mathsf{ws}(v_1 \cdots v_n)).$$

Therefore, $\mathcal{N}(\mathsf{Unr}^K(G, v_1), v_1 \cdots v_n) = \emptyset$ if and only if $\mathcal{N}(G', \mathsf{ws}(v_1 \cdots v_n)) = \emptyset$, because otherwise Spoiler could win in the first round.

If $\mathcal{N}(\mathsf{Unr}^K(G, v_1), v_1 \cdots v_n) = \emptyset$, then by definition of H also $\mathcal{N}(H, (v_1 \cdots v_n, i)) = \emptyset$ Hence, D wins $\mathcal{E}_k^{\mathrm{ML}}(H, (v_1 \cdots v_n, i), G', \mathsf{ws}(v_1 \cdots v_n))$ since S cannot choose a successor. Otherwise, Spoiler has three choices in this game:

- S can choose a vertex $(v_1 \cdots v_n u, j) \in \mathcal{N}(H, v_1 \cdots v_n)$ such that $v_1 \cdots v_n u \in V^{\mathsf{Unr}^K(G,v_1)}$.
  Then D may choose
  $$\mathsf{ws}(v_1 \cdots v_n u) \in \mathcal{N}(G', \mathsf{ws}(v_1 \cdots v_n)).$$
  But ws represents a winning strategy for D in
  $$\mathcal{E}_{k-1}^{\mathrm{AFML}[1]}(\mathsf{Unr}^K(G, v_1), v_1 \cdots v_n u, G', \mathsf{ws}(v_1 \cdots v_n u)).$$
  By induction hypothesis, D thus has a winning strategy for the remaining game
  $$\mathcal{E}_{k-1}^{\mathrm{ML}}(H, (v_1 \cdots v_n u, j), G', \mathsf{ws}(v_1 \cdots v_n u)).$$
- S can choose a vertex $u' \in \mathcal{N}(H, v_1 \cdots v_n)$ such that $u' \in V^{\mathsf{Unr}^K(G',u')}$.
  Then by definition of H, $u' \in \mathcal{N}(G', \mathsf{ws}(v_1, \cdots v_n))$. Thus D can choose u'. By Point (i), D has a winning strategy in the remaining game $\mathcal{E}_{k-1}^{\mathrm{ML}}(H, u', G', u')$.
- S can choose a successor u' of $\mathsf{ws}(v_1, \ldots, v_n)$ in G'.
  Since $\mathsf{ws}(v_1 \cdots v_n)$ has a successor in G', $v_1 \cdots v_n$ also has at least one successor in $\mathsf{Unr}^K(G, v_1)$. By construction of H, this implies that $(v_1 \cdots v_n, i)$ has u' as a successor in H. Thus, D can choose u'. By Point (i), D has a winning strategy in $\mathcal{E}_{k-1}^{\mathrm{ML}}(H, u', G', u')$.

This finishes the case where $\sim$ is $>$.

The case where $\sim$ is $\geq$ is analogous. Because φ is not expressible in AFML, for each $\ell \in \mathbb{N}$ there exist pointed graphs $(G, v), (G', v')$ with $G, v \models \varphi, G', v' \not\models \varphi$ and D has a winning strategy for $\mathcal{E}_\ell^{\mathrm{AFML}[2]}(G, v, G', v')$. It is easy to see that any such winning strategy is also a winning strategy for $\mathcal{E}_\ell^{\mathrm{AFML}[1]}(G', v', G, v)$. In fact, these games are exactly identical. Moreover, since $\sim$ is $\geq$, the following holds:

$(*_2)$ for each pointed graph (G, v) rejected by $\mathcal{G}$, there is an $\varepsilon > 0$ such that all pointed graphs (H, u) with $||\bar{x}_v^L - \bar{x}_u^L||_\infty < \varepsilon$ are also rejected by $\mathcal{G}$.

We may use exactly the same arguments as in the case where $\sim$ is $>$, except that the roles of (G, v) and (G', v') are swapped. This includes the exact same construction of (H, u), now from (G', v'). Instead of showing that $H, u \models \varphi$, we then have to show that $H, u \not\models \varphi$. This is exactly analogous since $(*_2)$ is about rejection while $(*_1)$ is about acceptance. ❑

Proof of Theorems 14 and 15

Theorem 14. AFML $\subseteq$ simple $\text{Mean}^{c,t}$-GNN *in the uniform setting. This holds for truncated ReLU and ReLU as activation functions.*

Proof. Recall that AFML $=$ AFML$[1] \cup$ AFML$[2]$. We start with AFML[1] formulas. Let φ be an AFML[1] formula over a finite set $\Pi = \{P_1, \dots, P_r\}$ of vertex labels, and let $\varphi_1, \dots, \varphi_L$ be an enumeration of the subformulas of φ such that (i) $\varphi_i = P_i$ for $1 \leq i \leq r$ and (ii) if φ_ℓ is a subformula of φ_k, then $\ell < k$. We construct a simple GNN $\mathcal{G}$ with $2L$ layers, all of input and output dimension $3L+1$ and with all combination functions identical. Each subformula gets computed within two layers. The truth value of the i-th subformula is stored in the i-th component of feature vectors. Additionally, we use up to two positions per subformula for bookkeeping purposes, which explains why we need at least $3L$ components in feature vectors. The additional $3L+1$-st component is needed for technical purposes and will be constant 1.

We define the matrices $A,\ C \in \mathbb{R}^{3L+1 \times 3L+1}$ and bias vector $\bar{b} \in \mathbb{R}^{3L+1}$ that define the combination function. All entries that are not mentioned explicitly have value 0. Set $b_{3L+1} = 1$ to achieve that the $3L{+}1$-st component is constant 1.[5] Let $k \in [L]$. We make a case distinction to set certain entries:

Case 1: $\varphi_k = P_k$: Set $C_{k,k} = 1$.

Case 2: $\varphi_k = \neg P_i$: Set $C_{i,k} = -1$ and $b_k = 1$.

Case 3: $\varphi_k = \Box\bot$: Set $A_{3n+1,k} = -1$ and $b_k = 1$.

Case 4: $\varphi_k = \varphi_i \vee \varphi_j$: Set $C_{i,k} = C_{j,k} = 1$.

Case 5: $\varphi_k = \Diamond\varphi_i$: Set $A_{i,k} = 1$.

Case 6: $\varphi_k = \varphi_i \wedge \varphi_j$: Set

$$\begin{aligned} C_{i,k+n} &= 1 \\ C_{j,k+n} &= -1 \\ C_{i,k+2n} &= -1 \\ C_{j,k+2n} &= 1 \\ C_{i,k} &= C_{j,k} = 1 \\ C_{k+n,k} &= C_{k+2n,k} = -1. \end{aligned}$$

Note that the use of bookkeeping values in components $L+1, \dots, 3L$ of feature vectors happens (only) in Case 6. We use the classification function

$$\mathsf{CLS}(\bar{x}) = \begin{cases} 1 & \text{if } x_L > 0 \\ 0 & \text{otherwise.} \end{cases}$$

[5]With the exception of the initial feature vectors, for which this is irrelevant.

We next show correctness of the translation.

Claim 1. For all φ_k, $1 \leq k \leq L$, the following holds: if $v \in V^G$ and $2k \leq k' \leq L$, then $(\bar{x}^{k'}_{G,v})_k \in (0, 1]$ if $G, v \models \varphi_k$ and $(\bar{x}^{k'}_{G,v})_k = 0$ otherwise.

We prove Claim 1 by induction on k. The correctness for Cases 1, 2, and 4 is straightforward to verify.

For Case 3, where $\varphi_k = \Box\bot$, we exploit that the $3L+1$-st component of every feature vector is constant 1. This implies that if a vertex has at least one successor, then

$$\mathsf{MEAN}(\{\{\bar{x}^{k'-1}_{G,u} \mid u \in \mathcal{N}(v)\}\})$$

returns a vector with 1 in the $3L+1$-st component and thus the k-th component in $\bar{x}^{k'}_{G,v}$ will be set to 0. When a vertex has no successors, the above mean returns a vector with 0 in the $3L+1$-st component and the k-th component in $\bar{x}^{k'}_{G,v}$ will be set to 1.

For Case 5, where $\varphi_k = \Diamond\varphi_i$, we know from the induction hypothesis that for each $u \in V^G$, the value stored in the i-th component of feature vector $\bar{x}^{k'-1}_{G,u}$ is from $(0, 1]$ if $G, u \models \varphi_i$ and 0 otherwise. Therefore,

$$\mathsf{MEAN}(\{\{\bar{x}^{k'-1}_{G,u} \mid u \in \mathcal{N}(v)\}\})$$

returns a vector that has 0 in the i-th component if v has no successors that satisfies φ_i and some value from $(0, 1]$ otherwise. This value is then stored in the k-th entry of $\bar{x}^{k'}_{G,v}$. Note that we do not have much control of that value except that it is from the stated range. This is the intuitive reason why there is no obvious way to encode the truth of formulas using value 1.

For Case 6, we first observe the following.

Claim 2. Let $x, y \in [0, 1]$. Then the following are equivalent:

1. $\mathsf{ReLU}^*(x + y - \mathsf{ReLU}^*(x - y) - \mathsf{ReLU}^*(y - x)) > 0$;
2. $x > 0$ and $y > 0$.

Proof of claim. When $x = 0$ then $\mathsf{ReLU}^*(x - y) = 0$ and $\mathsf{ReLU}^*(y - x) = \mathsf{ReLU}^*(y) = y$, both because $y \in [0, 1]$. Hence, $\mathsf{ReLU}^*(x + y - \mathsf{ReLU}^*(x - y) - \mathsf{ReLU}^*(y - x))$ becomes $\mathsf{ReLU}^*(0 + y - 0 - y) = 0$. The case $y = 0$ is analogous.

For the other direction, let $0 < x \leq y$. Then $\mathsf{ReLU}^*(y - x) = y - x$ and $\mathsf{ReLU}^*(x - y) = 0$, both because $x, y \in [0, 1]$. Thus, $\mathsf{ReLU}^*(x + y - \mathsf{ReLU}^*(x - y) - \mathsf{ReLU}^*(y - x))$ becomes $\mathsf{ReLU}^*(x + y - (y - x)) = \mathsf{ReLU}^*(2x) \in (0, 1]$. The case $0 < y \leq x$ is analogous. This finishes the proof of Claim 2.

Now let $\varphi_k = \varphi_i \wedge \varphi_j$. Analyzing the coefficient set in Case 6, it can be verified that after two layers, $(\bar{x}^{k'}_v)_k$ contains $\mathsf{ReLU}^*((\bar{x}^{k'-1}_v)_i + (\bar{x}^{k'-1}_v)_j - \mathsf{ReLU}^*((\bar{x}^{k'-2}_v)_i - (\bar{x}^{k'-2}_v)_j) - \mathsf{ReLU}^*((\bar{x}^{k'-2}_v)_j - (\bar{x}^{k'-2}_v)_i))$. By Claim 2, this is the desired value.

This finishes the proof for AFML[1] formulas. We now show how to realize AFML[2] formulas. Let $\varphi \in$ AFML[2]. Then there exists a $\psi \in$ AFML[1] such that $\varphi \equiv \neg\psi$. We have just seen that there if a GNN $\mathcal{G}$ that is equivalent to ψ. We add a layer to $\mathcal{G}$ that computes the function $f(\bar{x}) = 1 - x_L$, where L is the index used by the classification function in $\mathcal{G}$. This is easily realizable by a simple Mean^c-GNN. We also replace the classification function with

$$\mathsf{CLS}(x) = \begin{cases} 1 & \text{if } x \geq 1 \\ 0 & \text{otherwise.} \end{cases}$$

Now $\mathcal{G}$ accepts a pointed graph (G, v) if and only if the constructed GNN rejects (G, v). Thus, the latter is equivalent to $\neg\psi \equiv \varphi$. ❏

Theorem 15. AFML $\subseteq$ simple Mean^{c,t}-GNN *in the uniform setting, with sigmoid as the activation function.*

Proof. As in Lemma 14, we first concentrate on AFML[1] formulas. Let φ be such a formula and let $\varphi_1, \ldots, \varphi_L$ be an enumeration of the subformulas of φ such that (i) $\varphi_i = P_i$ for $1 \leq i \leq r$ and (ii) if φ_ℓ is a subformula of φ_k, then $\ell < k$. We construct a simple GNN $\mathcal{G}$ with $5L + 1$ layers, all of input and output dimension $13L + 1$. In contrast to previous reductions, the combination function of the first layer is different to that of the other layers, which share the same combination function. This helps us to deal with formulas of the form P_k and $\neg P_k$ in a technically simple way and could be avoided at the expense of a slightly more technical translation. Each subformula gets computed within five layers. As before, the truth value of the i-th subformula is stored in the i-th component of feature vectors. The additional positions are used for bookkeeping purposes, with the $13L + 1$-st component being constant $\frac{1}{2}$.

We first introduce a suitable helper function that will assist us to deal with conjunction.

Claim 1. Define

$$\tau(x) = -\sigma(\sigma(x)) - \sigma(\sigma(-x)).$$

Then the following properties hold:

1. τ has a strict global minimum at $x = 0$, with $e_{\min} = \tau(0) = -\frac{2}{1+\frac{1}{\sqrt{e}}}$,
2. τ is symmetric in the sense that $\tau(x) = \tau(-x)$ for all $x \in \mathbb{R}$, and
3. τ is injective on the interval $[0, \infty)$.

The function

$$\alpha(x, y) = \tau(\tau(x - y) - \tau(x + y - 1)) - e_{\min}$$

is 0 if $x = \frac{1}{2}$ or $y = \frac{1}{2}$ and is greater than 0 otherwise.

Properties 1-3 of τ are straightforward to verify. By definition and Property 2, the function α has a strict global minimum if $\tau(x - y) = \tau(x + y - 1)$ which by Properties 2 and 3 is the case if and only if $x - y = x + y - 1$ or $x - y = -x - y + 1$. The first case holds if $y = \frac{1}{2}$ and the second case holds if $x = \frac{1}{2}$. This finishes the proof of Claim 1.

We define the matrices $A,\ C \in \mathbb{R}^{13L+1\times 13L+1}$ and bias vector $\bar{b} \in \mathbb{R}^{13L+1}$ that define the combination function for the first layer. All entries that are not mentioned explicitly have value 0. Set $b_{13L+1} = 0$ to achieve that the $13L + 1$-st component is $\sigma(0) = \frac{1}{2}$. Let $k \in [L]$. We make a case distinction as follows:

Case 1: $\varphi_k = P_k$. Set $C_{k,k} = 1$ and $b_k = 0$.

Case 2: $\varphi_k = \neg P_i$. Set $C_{i,k} = -1$ and $b_k = 1$.

What this achieves is that, for formulas φ_k of the form P_k or $\neg P_k$, $(\bar{x}_v^1)_k \in (\frac{1}{2}, 1)$ if $G \models \varphi_k$ and $(\bar{x}_v^1)_k = \frac{1}{2}$ if $G \not\models \varphi_k$. The first layer thus treats all these formulas simultaneously.[6]

The remaining $5L$ layers all use the same combination function. We next give the defining matrices $A,\ C \in \mathbb{R}^{13L+1\times 13L+1}$ and bias vector $\bar{b} \in \mathbb{R}^{13L+1}$, again only explicitly giving entries that are not 0. We set $b_{13L+1} = 0$ to achieve that the $13L + 1$-st component is $\sigma(0) = \frac{1}{2}$. Let $k \in [L]$. We make a case distinction as follows:

Case 1: $\varphi_k = P_k$. Set $C_{k,k} = 1$ and $b_k = -\frac{1}{2}$.

Case 2: $\varphi_k = \neg P_i$. Set $C_{k,k} = 1$ and $b_k = -\frac{1}{2}$.

Case 3: $\varphi_k = \Box\bot$. Set $A_{13L+1,k} = -1$ and $b_k = \frac{1}{2}$.

Case 4: $\varphi_k = \varphi_i \vee \varphi_j$. Set $C_{i,k} = C_{j,k} = 1$ and $b_k = -1$.

Case 5: $\varphi_k = \Diamond\varphi_i$. Set $A_{i,k} = 1$ and $A_{13L+1,k} = -1$.

Case 6: $\varphi_k = \varphi_i \wedge \varphi_j$. Using Claim 1, it is easy to see that

$$\sigma(\alpha((\bar{x}_v^{k'-1})_i, (\bar{x}_v^{k'-1})_j))$$

gives the desired result. Computing α requires three applications of τ, each in turn requiring four temporary values, namely for $\sigma(x), \sigma(\sigma(x)), \sigma(-x)$ and $\sigma(\sigma(-x))$, and 2 layers. In α, the two inner computations of τ can be computed in parallel, thus α needs 12 temporary values and 4 layers. It is straightforward to work out the corresponding entries in C and $\bar{b}$, we omit details.

Note that the use of bookkeeping values in components $L + 1, \ldots, 13L$ of feature vectors happens (only) in Case 6. We use the classification function

$$\mathsf{CLS}(\bar{x}) = \begin{cases} 1 & \text{if } x_L > \frac{1}{2} \\ 0 & \text{otherwise.} \end{cases}$$

We next show correctness of the translation.

Claim 2. For all φ_k, $1 \leq k \leq L$, the following holds: if $v \in V^G$ and $5k + 1 \leq k' \leq L$, then $(\bar{x}_v^{k'})_k \in (\frac{1}{2}, 1)$ if $G \models \varphi_k$ and $(\bar{x}_v^{k'})_k = \frac{1}{2}$ if $G \not\models \varphi_k$.

We prove Claim 2 by induction on k.

Cases 1 and 2 rely on the fact that formulas of the form P_k and $\neg P_k$ have already been treated in the first layer. They simply implement the 'identity function' in the sense that $(\bar{x}_v^{k'})_k \in (\frac{1}{2}, 1)$ if $(\bar{x}_v^{k'-1})_k \in (\frac{1}{2}, 1)$ and $(\bar{x}_v^{k'})_k = \frac{1}{2}$ if $(\bar{x}_v^{k'-1})_k = \frac{1}{2}$ for all $k' \geq 1$.

For Case 3, where $\varphi_k = \Box\bot$, we exploit that the $13L+1$-st component of every feature vector is constant $\frac{1}{2}$. This implies that if a vertex has at least one successor, then

$$\mathsf{MEAN}(\{\{\bar{x}_{G,u}^{k'-1} \mid u \in \mathcal{N}(v)\}\})$$

returns a vector with $\frac{1}{2}$ in the $13L + 1$-st component and thus the k-th component in $\bar{x}_{G,v}^{k'}$ will be set to $\sigma(0) = \frac{1}{2}$. When

[6]Which actually means that $5L - 5m$ of the remaining layers suffice, where m is the number of subformulas of φ that are of the form P_k or $\neg P_k$; however, additional layers are not harmful.

a vertex has no successors, the above mean returns a vector with 0 in the $13L+1$-st component and the k-th component in $\bar{x}^{k'}_{G,v}$ will be set to $\sigma(\frac{1}{2}) \in (\frac{1}{2}, 1)$.

Case 4 is straightforward to verify.

For Case 5, first assume that a vertex v has a successor u that satisfies φ_i. Then $(\bar{x}^{k'-1}_u)_i > \frac{1}{2}$ and therefore $\mathsf{MEAN}(\{\{(\bar{x}^{k'-1}_u)_i \mid u \in \mathcal{N}(v)\}\}) > \frac{1}{2}$. Additionally, $\mathsf{MEAN}(\{\{(\bar{x}^{k'-1}_u)_{13L+1} \mid u \in \mathcal{N}(v)\}\}) = \frac{1}{2}$, thus $(\bar{x}^{k'}_v)_k > \frac{1}{2}$. Now assume that v has at least one successor, but none of its successors satisfies φ_i. Then $\mathsf{MEAN}(\{\{(\bar{x}^{k'-1}_u)_i \mid u \in \mathcal{N}(v)\}\}) = \frac{1}{2}$, implying $(\bar{x}^{k'}_v)_k = \frac{1}{2}$. Finally assume that v has no successors. Then $\mathsf{MEAN}(\{\{(\bar{x}^{k'-1}_u)_i \mid u \in \mathcal{N}(v)\}\}) = \mathsf{MEAN}(\{\{(\bar{x}^{k'-1}_u)_{13L+1} \mid u \in \mathcal{N}(v)\}\}) = 0$ and thus $(\bar{x}^{k'}_v)_k = \sigma(0) = \frac{1}{2}$.

For Case 6, it suffices to invoke Claim 1 and observe that $\sigma(0) = \frac{1}{2}$ and $\sigma(x) \in (\frac{1}{2}, 1)$ for all $x > 0$.

If φ is a formula in AFML[2] we again use that there exists a formula $\psi \in$ AFML[1] such that $\varphi \equiv \neg\psi$. Let $\mathcal{G}$ be the GNN that is equivalent to ψ. We add a layer to $\mathcal{G}$ that computes $\sigma(\frac{1}{2} - x_L)$. The classification function is replaced with

$$\mathsf{CLS}(x) = \begin{cases} 1 & \text{if } x \geq \frac{1}{2} \\ 0 & \text{otherwise.} \end{cases}$$

Now $\mathcal{G}$ accepts a vertex v if and only if the extended GNN rejects v. Thus, the latter is equivalent to $\neg\psi \equiv \varphi$. ❑

E Proof of Theorem 16

To prove Theorem 16, we state as separate theorem the items stated there. These are Theorems 20 and 22.

Theorem 20. *The property $\mathcal{P}$ 'there exist more successors that satisfy P_1 than successors that satisfy P_2' is not expressible in RML, but by a simple* Mean^{c,t}-GNN.

Proof. Assume that φ is a formula in RML that expresses this property. Let R be the set of all rational numbers r such that $\Diamond^{>r}$ or $\Diamond^{\geq r}$ occurs in φ. Because R is finite, there exists $x = \min(R \cup (1-R)) \setminus \{0\}$. Clearly, $(0,x) \cap R = (1-x, 1) \cap R = \emptyset$. It is easy to see that φ cannot distinguish between different fractions in $(0, x)$.

Let $y \in \mathbb{Q}$ such that $0 < 3y < x$. We now consider a graph G_1 where a vertex v_1 has y successors that satisfy P_1, $2y$ successors that satisfy P_2, and $1-3y$ successors that satisfy neither P_1 nor P_2. Graph G_2 with vertex v_2 is defined analogously, but with the roles of P_1 and P_2 swapped.

In both graphs all fractions are in the intervals $(0, 3y] \subseteq (0, x)$ and $[(1-3y), 1) \subseteq (1-x, 1)$, thus φ will classify v_1 and v_2 identically. But v_2 satisfies property $\mathcal{P}$ while v_1 does not. Thus, there exists no RML formula that expresses $\mathcal{P}$.

The one-layer, simple GNN with ReLU^* activation, C the all zero matrix and

$$A = \begin{pmatrix} 1 \\ -1 \end{pmatrix} \text{ and } \mathsf{CLS}(x) = \begin{cases} 1 & \text{if } x > 0 \\ 0 & \text{otherwise} \end{cases}$$

computes the fraction of successors satisfying P_1 and subtracts the fraction of successors satisfying P_2. This difference is greater than 0 if and only if there are more successors satisfying P_1 than P_2. ❑

We recall the notions of convex sets and affine spaces, as well as some lemmas regarding the interior and closure of convex sets. See (Rockafellar 1970) for more in-depth information.

Definition 4. *The* affine hull $\mathsf{aff}(S)$ *of a set $S \subseteq \mathbb{R}^n$ is defined as*

$$\{\lambda_1 x_1 + \cdots + \lambda_m x_m \mid m \in \mathbb{N}, x_i \in S, \lambda_i \in \mathbb{R}, \sum_{i=1}^{m} \lambda_i = 1\}.$$

It is the smallest superset of S that contains all lines through two of its points. That is, for $x, y \in \mathsf{aff}(S)$,

$$\{(1-\lambda)x + \lambda y \mid \lambda \in \mathbb{R}\} \subseteq \mathsf{aff}(S).$$

Definition 5. *The* convex hull *of a set $S \subseteq \mathbb{R}^n$ is defined as*

$$\{\lambda_1 x_1 + \cdots + \lambda_m x_m \mid m \in \mathbb{N},\ x_i \in S,\ 0 \leq \lambda_i \leq 1,\ \sum_{i=1}^{m} \lambda_i = 1\}.$$

We denote the convex hull by $\mathsf{cvh}(S)$. *It is the smallest superset of S that contains all line segments between two of its points. That is, for x, y in the convex hull of S,*

$$\{\lambda x + (1-\lambda)y \mid 0 \leq \lambda \leq 1\}$$

is a subset of the convex hull. A set S is called convex if it is identical to its convex hull.

The convex hull of S and the means of multisets over S are tightly connected.

Lemma 18. *Let $S \subseteq \mathbb{R}^n$ be a set of vectors and let $\mathcal{M}(S)$ be the set of all finite multisets over S. Then*

$$\mathsf{MEAN}(\mathcal{M}(S)) = \{\mathsf{MEAN}(M) \mid M \in \mathcal{M}(S)\} \subseteq \mathsf{cvh}(S).$$

Moreover, the means of elements in $\mathcal{M}(S)$ define exactly the elements in the convex hull where the parameters λ are rational numbers. That is,

$$\{\mathsf{MEAN}(M) \mid M \in \mathcal{M}(S)\} = \{\lambda_1 x_1 + \cdots + \lambda_m x_m \mid m \in \mathbb{N}, x_i \in S, \lambda_i \in [0,1] \cap \mathbb{Q}, \sum_{i=1}^{m} \lambda_i = 1\}.$$

Lemma 19. *Let $T \subseteq S \subseteq \mathbb{R}^n$ and let T be dense in S. Then $\{\mathsf{MEAN}(M) \mid M \in \mathcal{M}(T)\}$ is dense in* $\mathsf{cvh}(S)$.

Proof. We can rewrite $\mathsf{MEAN}(\mathcal{M}(T))$ as

$$\mathsf{MEAN}(\mathcal{M}(T)) = \{\lambda_1 x_1 + \cdots + \lambda_m x_m \mid m \in \mathbb{N}, x_i \in T, \lambda_i \in [0,1] \cap \mathbb{Q}, \sum_{i=1}^{m} \lambda_i = 1\}.$$

Since $\mathbb{Q}$ is dense in $\mathbb{R}$ and T is dense in S, it is easy to verify that each vector in $\mathsf{cvh}(S)$ can be approximated by vectors in $\mathsf{MEAN}(\mathcal{M}(T))$. Thus, $\mathsf{MEAN}(\mathcal{M}(T))$ is dense in $\mathsf{cvh}(S)$. ❑

Definition 6. *The* relative interior *of a convex set $S \subseteq \mathbb{R}^n$ is the (topological) interior of S relative to its affine hull. That is*

$$\mathsf{ri}(S) = \{x \in S \mid \exists \varepsilon > 0\ \forall y \in \mathsf{aff}(S)\ ||x - y||_\infty < \varepsilon \implies y \in S\}.$$

Lemma 20. *For all convex sets S,*

- *the relative interior of S is empty if and only if S is the empty set, and*
- *the closure of $\mathsf{ri}(S)$ is the same as the closure of S. Furthermore, the closure is contained in the affine hull of S:*

$$\mathsf{cl}(\mathsf{ri}(S)) = \mathsf{cl}(S) \subseteq \mathsf{aff}(S).$$

Lemma 21 (Accessibility Lemma)**.** *Let $S \subseteq \mathbb{R}^n$ be a convex set. Let $x \in \mathsf{ri}(S)$ and $y \in \mathsf{cl}(S)$. Then*

$$\{\lambda x + (1-\lambda)y \mid 0 \leq \lambda < 1\} \subseteq \mathsf{ri}(S).$$

We also introduce some lemmas regarding dense sets in $\mathbb{R}^n$ with respect to the maximum metric $||\cdot||_\infty$.

Definition 7. *Let $X \subseteq Y \subseteq \mathbb{R}^\delta$. Then X is dense in Y if for all $y \in Y$ and $\varepsilon > 0$ we have*

$$B_\varepsilon(y) = \{y' \in Y \mid ||y' - y||_\infty < \varepsilon\} \cap X \neq \emptyset.$$

Lemma 22. *Let $S \subseteq X$ be dense in X and let f be a continuous function. Then $f(S)$ is dense in $f(X)$.*

Lemma 23. *Let $X, Y \subseteq Z \subseteq \mathbb{R}^n$ such that X is dense in Z and Y is open in Z. Then $X \cap Y$ is dense in Y.*

Proof. Let $y \in Y$. Further let $\varepsilon > 0$. Define the open ε-ball around y to be

$$B_\varepsilon(y) = \{z \in Y \mid ||z - y||_\infty < \varepsilon\}.$$

We have to show that $B_\varepsilon(y) \cap X \cap Y \neq \emptyset$.

As Y is open in Z, there is a δ with $0 < \delta < \varepsilon$ such that the open δ-ball around y is a subset of Y, that is

$$B_\delta(y) = \{z \in Z \mid ||z - y||_\infty < \delta\} \subseteq Y.$$

Therefore, $B_\delta(y) \subseteq B_\varepsilon(y)$. Since $y \in Z$ and X is dense in Z, we have $B_\delta(y) \cap X \neq \emptyset$. Thus $B_\varepsilon(y) \cap X \cap Y \supseteq B_\delta(y) \cap X \cap Y \neq \emptyset$. ❑

Definition 8. *A path in $\mathbb{R}^n$ is a set $X \subseteq \mathbb{R}^n$ such that there exists a continuous surjective function $f : [0,1] \to X$.*

In the rest of the proof, we will consider the class $\mathcal{T}$ of all graphs that take the form of a directed tree of depth two, with vertex labels from $\Pi = \{P\}$, and where P may only label leaves.

Theorem 21. *There is no two-layer* Mean^{c,t}-GNN *that classifies the roots of trees in $\mathcal{T}$ according to the RML formula $\varphi = \Diamond^{>\frac{1}{2}} \Diamond^{>\frac{1}{2}} P$.*

Proof. Let

$$\mathcal{G} = (L, \{\mathsf{AGG}^\ell\}_{\ell \in [2]}, \{\mathsf{COM}^\ell\}_{\ell \in [2]}, \mathsf{CLS})$$

be a two-layer Mean^{c,t}-GNN with input dimensions $\delta^0 = 1$ and δ^1. When $\mathcal{G}$ is run on a graph $G \in \mathcal{T}$, then the feature vectors computed at a leaf $v \in V(G)$, by layer 1 and 2, only depend on whether v is labeled with P or not, but not on any other aspects of G. Moreover, the feature vector computed by the first layer at the root is entirely independent of G. Let $\bar{x}^1$ be the feature vector computed by the first layer at the root, $\bar{y}^0$ the initial feature vector of the intermediate vertices, and $\bar{z}^\ell_+$ and $\bar{z}^\ell_-$ the feature vectors of the leaves that are and are not labeled with P, respectively.

For an intermediate vertex with a fraction of $r \in [0,1] \cap \mathbb{Q}$ successors labeled P, the vector computed in layer 1 is

$$\bar{y}^1_r = \mathsf{COM}^1(\bar{y}^0, r\bar{z}^0_+ + (1-r)\bar{z}^0_-).$$

We define the set of all such vectors by Y^1. This set is dense in $Y^1_{\mathbb{R}} = \{\mathsf{COM}^1(\bar{y}^0, r\bar{z}^0_+ + (1-r)\bar{z}^0_- \mid r \in [0,1])\}$. As an intuition, the set

$$\{(\bar{y}^0, r\bar{z}^0_+ + (1-r)\bar{z}^0_-) \mid r \in [0,1]\}$$

forms a line segment in $\mathbb{R}^{2\delta^0}$ and $Y^1_{\mathbb{R}}$ forms a path in $\mathbb{R}^{\delta^1}$. The set Y^1 can be partitioned into two sets $Y^1_+ = \{\bar{y}^1_r \mid r \in (\frac{1}{2}, 1] \cap \mathbb{Q}\}$ and $Y^1_- = \{\bar{y}^1_r \mid r \in [0, \frac{1}{2}] \cap \mathbb{Q}\}$. We can split $Y^1_{\mathbb{R}}$ analogously into $Y^1_{+,\mathbb{R}}$ and $Y^1_{-,\mathbb{R}}$.

For each root vertex v with

$$s = \frac{|\{u \in \mathcal{N}(v) \mid u \models \Diamond^{>\frac{1}{2}} P\}|}{|\mathcal{N}(v)|},$$

the MEAN computed in the second layer can be reformulated as follows:

$$\begin{aligned}
&\mathsf{MEAN}\{\{\bar{y}^1_u \mid u \in \mathcal{N}(v)\}\} \\
&= s \cdot \mathsf{MEAN}\{\{\bar{y}^1_u \mid u \in \mathcal{N}(v), u \models \Diamond^{>\frac{1}{2}} P\}\} \\
&\quad + (1-s) \cdot \mathsf{MEAN}\{\{\bar{y}^1_u \mid u \in \mathcal{N}(v), u \not\models \Diamond^{>\frac{1}{2}} P\}\}.
\end{aligned}$$

Recall that to classify the trees in $\mathcal{T}$ according to the RML formula $\Diamond^{>\frac{1}{2}} \Diamond^{>\frac{1}{2}} P$, the GNN has to decide whether $s > \frac{1}{2}$. Clearly, the first MEAN in the above sum processes a multiset in $\mathcal{M}(Y^1_+)$, while the second MEAN processes a multiset in $\mathcal{M}(Y^1_-)$. We define

$$\begin{aligned}
M_+ = \{s\bar{y}_+ + (1-s)\bar{y}_- \mid\ & s \in (\frac{1}{2}, 1] \cap \mathbb{Q}, \\
& \bar{y}_+ \in \mathsf{MEAN}(\mathcal{M}(Y^1_+)), \\
& \bar{y}_- \in \mathsf{MEAN}(\mathcal{M}(Y^1_-))\} \\
M_{+,\mathbb{R}} = \{s\bar{y}_+ + (1-s)\bar{y}_- \mid\ & s \in (\frac{1}{2}, 1], \\
& \bar{y}_+ \in \mathsf{cvh}(Y^1_{+,\mathbb{R}}),\ \bar{y}_- \in \mathsf{cvh}(Y^1_{-,\mathbb{R}})\}.
\end{aligned}$$

We define M_- and $M_{-,\mathbb{R}}$ analogously, but require that $s \in [0, \frac{1}{2}] \cap \mathbb{Q}$ and $s \in [0, \frac{1}{2}]$ respectively. Now for each $\bar{m}_+ \in M_+$, the vector $\mathsf{COM}^2(\bar{x}^1, \bar{m}_+)$ is a feature vector computed at the root of a tree in $\mathcal{T}$ that satisfies φ. And likewise, for each $\bar{m}_- \in M_-$, the vector $\mathsf{COM}^2(\bar{x}^1, \bar{m}_-)$ belongs to a tree in $\mathcal{T}$ that does not satisfy φ.

By Lemma 19, $\mathsf{MEAN}(\mathcal{M}(Y^1_+))$ is dense in $\mathsf{cvh}(Y^1_{+,\mathbb{R}})$ and $\mathsf{MEAN}(\mathcal{M}(Y^1_-))$ is dense in $\mathsf{cvh}(Y^1_{-,\mathbb{R}})$. Therefore, M_+ is dense in $M_{+,\mathbb{R}}$. Analogously M_- is dense in $M_{-,\mathbb{R}}$. We now show that $M_{+,\mathbb{R}}$ and $M_{-,\mathbb{R}}$ have a non-empty intersection. To be more precise, we show there exists a set in $M_{+,\mathbb{R}}$ that is open in $\mathsf{cvh}(Y^1_{+,\mathbb{R}})$ that is also contained in $M_{-,\mathbb{R}}$. Let $\bar{y} \in \mathsf{ri}(\mathsf{cvh}(Y^1_{+,\mathbb{R}}))$ and let $\varepsilon > 0$ such that

$$B_{\mathbb{R}} = \{y' \mid y' \in \mathsf{aff}(\mathsf{cvh}(Y^1_{+,\mathbb{R}})), ||y' - y||_\infty < \varepsilon\} \subseteq \mathsf{ri}(\mathsf{cvh}(Y^1_{+,\mathbb{R}})).$$

Consider the point

$$\bar{y}^1_{0.5} = \mathsf{COM}^1(\bar{y}^0, 0.5\bar{z}^0_+ + 0.5\bar{z}^0_-) \in Y^1_-.$$

Because COM^1 is continuous, we can approximate $\bar{y}^1_{0.5}$ by vectors in $Y^1_{+,\mathbb{R}}$, hence $\bar{y}^1_{0.5} \in \mathsf{cl}(\mathsf{cvh}(Y^1_{+,\mathbb{R}}))$. By the accessibility lemma,

$$B_{?,\mathbb{R}} = \frac{1}{4}B_{\mathbb{R}} + \frac{3}{4}\bar{y}^1_{0.5} \subseteq \mathsf{ri}(\mathsf{cvh}(Y^1_{+,\mathbb{R}})).$$

Since $\mathsf{ri}(\mathsf{cvh}(Y^1_{+,\mathbb{R}})) \subseteq M_{+,\mathbb{R}}$, it follows that $B_{?,\mathbb{R}} \subseteq M_{+,\mathbb{R}}$. But we also have $B_{?,\mathbb{R}} \subseteq M_{-,\mathbb{R}}$ since for each vector $\bar{v} \in B_{?,\mathbb{R}}$ we can find a $\bar{y}_+ \in \mathsf{cvh}(Y^1_{+,\mathbb{R}})$ and can choose $\bar{y}_- = \bar{y}^1_{0.5} \in \mathsf{cvh}(Y^1_{-,\mathbb{R}})$ and $s = \frac{1}{4} \in [0, \frac{1}{2}]$ such that $\bar{v} = s\bar{y}_+ + (1-s)\bar{y}_-$.

It remains to argue that $B_{?,+} = M_+ \cap B_{?,\mathbb{R}}$ and $B_{?,-} = M_- \cap B_{?,\mathbb{R}}$ are dense in $B_{?,\mathbb{R}}$.

We show both statements by using Lemma 23 by considering these sets relative to $\mathsf{cvh}(Y^1_{+,\mathbb{R}})$. For the first statement, consider $X = M_+ \cap \mathsf{cvh}(Y^1_{+,\mathbb{R}})$, $Y = B_{?,\mathbb{R}}$, $Z = \mathsf{cvh}(Y^1_{+,\mathbb{R}})$. The set $B_{?,\mathbb{R}}$ is open in $\mathsf{cvh}(Y_{+,\mathbb{R}})$ since $B_{\mathbb{R}}$ is open in $\mathsf{aff}(\mathsf{cvh}(Y^1_{+,\mathbb{R}}))$, $\bar{y}^1_{0.5} \in \mathsf{aff}(\mathsf{cvh}(Y^1_{+,\mathbb{R}}))$ and $B_{?,\mathbb{R}}$ is just a translation and scaling within $\mathsf{aff}(\mathsf{cvh}(Y^1_{+,\mathbb{R}}))$. Thus, $B_{?,\mathbb{R}}$ is open in $\mathsf{aff}(\mathsf{cvh}(Y^1_{+,\mathbb{R}}))$ and thus also in $\mathsf{cvh}(Y^1_{+,\mathbb{R}})$. Since $\mathsf{MEAN}(\mathcal{M}(Y^1_+)) \subseteq M_+ \cap \mathsf{cvh}(Y^1_{+,\mathbb{R}})$ and $\mathsf{MEAN}(\mathcal{M}(Y^1_+))$ is dense in $\mathsf{cvh}(Y^1_{+,\mathbb{R}})$, it follows that $M_+ \cap \mathsf{cvh}(Y^1_{+,\mathbb{R}})$ is dense in $\mathsf{cvh}(Y^1_{+,\mathbb{R}})$. We can rewrite $B_{?,+}$ as $(M_+ \cap \mathsf{cvh}(Y^1_{+,\mathbb{R}})) \cap B_{?,\mathbb{R}}$, since $B_{?,\mathbb{R}} \subseteq \mathsf{cvh}(Y^1_{+,\mathbb{R}})$.

Before showing that $B_{?,-}$ is dense in $B_{?,\mathbb{R}}$ we first consider the sets $X = \mathsf{MEAN}(\mathcal{M}(Y^1_+))$, $Y = B_{\mathbb{R}}$, as well as $Z = \mathsf{cvh}(Y_{+,\mathbb{R}})$. Using Lemma 23, we conclude that $B = B_{\mathbb{R}} \cap \mathsf{MEAN}(\mathcal{M}(Y^1_+))$ is dense in $B_{\mathbb{R}}$. We now transform these sets according to the definition of $B_{?,\mathbb{R}}$: $X = \frac{1}{4}\mathsf{MEAN}(\mathcal{M}(Y^1_+)) + \frac{3}{4}\bar{y}^1_{0.5}$, $Y = B_{?,\mathbb{R}}$, and $Z = \frac{1}{4}\mathsf{cvh}(Y_{+,\mathbb{R}}) + \frac{3}{4}\bar{y}^1_{0.5}$. X remains dense and Y remains open in Z, therefore $B_{?,\mathbb{R}} \cap (\frac{1}{4}\mathsf{MEAN}(\mathcal{M}(Y^1_+)) + \frac{3}{4}\bar{y}^1_{0.5})$ is dense in $B_{?,\mathbb{R}}$. The desired statement now follows from $\frac{1}{4}\mathsf{MEAN}(\mathcal{M}(Y^1_+)) + \frac{3}{4}\bar{y}^1_{0.5} \subseteq M_-$.

We can now finish our proof that $\mathcal{G}$ has to misclassify some graphs in $\mathcal{T}$. W.l.o.g we can assume that COM^2 has output dimension 1 since $\mathsf{CLS}(\bar{x})$ only depends on one entry in $\bar{x}$. Since projections are continuous, this does not affect the continuity of COM^2. By Lemma 22 and since COM^2 is continuous, $\mathsf{COM}^2(\bar{x}^1, B_{?,+})$ and $\mathsf{COM}^2(\bar{x}^1, B_{?,-})$ are dense in $\mathsf{COM}^2(\bar{x}^1, B_{?,\mathbb{R}})$ and the step function CLS cannot separate these two dense sets. ❑

We still have to show that the restriction to two-layer GNNs does not change the expressiveness on the class of graphs $\mathcal{T}$.

Theorem 22. *The RML formula $\Diamond^{>\frac{1}{2}} \Diamond^{>\frac{1}{2}} P$ is not expressible by a* Mean^{c,t}-GNN.

Proof. We show that for each Mean^{c,t}-GNN $\mathcal{G}$ with L layers there exists a GNN with two layers that classifies the root of each tree in $\mathcal{T}$ identically. The construction preserves the continuity of the combination functions.

Let $\bar{a}^0_+, \ldots, \bar{a}^L_+$ and $\bar{a}^0_-, \ldots, \bar{a}^L_-$ be the feature vectors assigned by $\mathcal{G}$ to a vertex without successors that does and does not satisfy P, respectively. Notice that all initial feature vectors are either $\bar{a}^0_+$ or $\bar{a}^0_-$.

The first layer of the two-layer GNN works as follows: $\mathsf{COM'}^1(x, y)$ is the concatenation of the vectors

$$\bar{v}_1 = \mathsf{COM}^1(\bar{a}^0_-, y\bar{a}^0_+ + (1-y)\bar{a}^0_-)$$
$$\bar{v}_i = \mathsf{COM}^i(\bar{v}_{i-1}, y\bar{a}^{i-1}_+ + (1-y)\bar{a}^{i-1}_-)$$

for all $1 < i \leq L$. Intuitively, y is the fraction of successors satisfying P and $\mathsf{COM'}^1$ computes all feature vectors of a vertex at height one that does not satisfy P. Since all COM^i are continuous, $\mathsf{COM'}^1$ is again continuous because continuous functions are closed under addition, function composition and vector concatenation.

The second layer is defined similarly: $\mathsf{COM'}^2((\bar{x}_1, \ldots, \bar{x}_L), (\bar{y}_1, \ldots, \bar{y}_L))$ returns the vector $\bar{u}_L$, where

$$\bar{u}_1 = \mathsf{COM}^1(\bar{a}^0_-, \bar{a}^0_-)$$
$$\bar{u}_i = \mathsf{COM}^i(\bar{u}_{i-1}, \bar{y}_{i-1})$$

Intuitively, these vectors are the feature vectors of a vertex of height two in a tree where the vertex itself and all immediate successors do not satisfy P. Therefore, this GNN computes the same feature vector as $\mathcal{G}$ at the root of each tree in $\mathcal{T}$. Analogously to $\mathsf{COM'}^1$, because all COM^i are continuous, $\mathsf{COM'}^2$ is continuous.

By using the same classification function as $\mathcal{G}$, the constructed two-layer GNN classifies trees in $\mathcal{T}$ in the same way as $\mathcal{G}$. It thus follows from Theorem 21 that $\mathcal{G}$ does not express $\Diamond^{>\frac{1}{2}} \Diamond^{>\frac{1}{2}} P$. ❑