CONSTANT-DEPTH RATIONAL-WEIGHT GNNS AND THEIR CONNECTION TO SCHANUEL'S CONJECTURE

Anonymous authors
Paper under double-blind review

ABSTRACT

A landmark result states that the expressive power of graph neural networks (GNNs) is bounded from above by the 1-dimensional Weisfeiler-Leman (1-WL) test. We ask whether a constant-depth GNN can match the expressive power of 1-WL while using only rational weights. We prove that this is the case when Schanuel's conjecture, a famous open problem from transcendental number theory, is taken to be true. Our GNN construction works even with integer weights, and more generally with a broad range of rational weights. As the activation function, it may use logistic sigmoid, hyperbolic tangent, or related functions. It is uniform across all graph sizes. We also prove a partial converse: if a GNN with exponentiation, hyperbolic sine or hyperbolic cosine as activation matches the expressive power of the 1-WL test in a slightly more expressive setting where graphs may contain certain auxiliary vertices, then this implies Schanuel's conjecture restricted to the numbers that arise in computations of the GNN. We also conduct experiments showing that the constructed GNNs indeed match the 1-WL test.

1 INTRODUCTION

A fundamental result in the theory of graph neural networks (GNNs) is that their expressive power is bounded from above by the 1-dimensional Weisfeiler–Leman (1-WL) test, in a version also known as color refinement; see Xu et al. (2019); Morris et al. (2019). In particular, if two vertices, or two graphs, cannot be distinguished by 1-WL, then they cannot be distinguished by a standard message-passing GNN either. Apart from motivating a lot of research on how to break the 1-WL barrier, this raises the converse question: to what extent can GNNs actually attain the 1-WL upper bound? This question has been answered in several different ways.

It was shown in Xu et al. (2019) that GNNs of constant depth L match the expressive power of L-round 1-WL when the aggregation and combination functions, as well as the graph-level readout, are injective. They instantiate this in an architecture called graph isomorphism networks (GINs) where combination function are represented as feedforward neural networks (FNNs) with ReLU activation. A more economic construction was proposed in Aamand et al. (2022). An important caveat to these results is that they are *non-uniform*, that is, the construction of the GNN depends on the maximum size of the input graphs. A uniform result was obtained by Bravo et al. (2024), refining a result by Amir et al. (2023). Their construction works with any non-polynomial analytic activation function, such as logistic sigmoid, and yields GNNs of internal dimension 1, that is, embeddings associate vertices with real numbers rather than with vectors of reals. While, remarkably, Bravo et al. show that in an L-layer GNN almost every choice of real-valued weights suffices to match L-round 1-WL, they do not establish that rational, or more generally finitely representable, weights suffice. It is interesting to contrast this with the result of Khalife & Basu (2026) which states that it is impossible for a constant-depth GNN with a piecewise polynomial activation function such as ReLU to match the expressive power of even 2-round 1-WL.

A complementary result was recently obtained by Rosenbluth & Grohe (2026). They show that recurrent GNNs with rational, finite-precision weights and ReLU activation can attain the 1-WL limit. That is, there is a recurrent GNN that distinguishes two vertices, or graphs, if and only if some number of 1-WL rounds is able to distinguish them. Note that recurrence significantly changes the game: allowing the number of GNN iterations to depend on the input makes it possible to simulate message-passing algorithms, potentially using a non-constant number of iterations for each round

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

of the 1-WL algorithm. When working with fixed-depth GNNs, in contrast, a single round of 1-WL must be simulated by only a constant number of GNN layers.

These results leave open a natural question, addressed in this article: *Can a constant-depth GNN uniformly match the expressive power of 1-WL while using only rational weights?* We answer this question conditionally on Schanuel's conjecture, a famous unsolved problem in transcendental number theory. Assuming the conjecture, we show that for a broad class of activation functions, including logistic sigmoid, hyperbolic tangent and the exponential function e^x, L-layer GNNs with rational weights can uniformly match the L-round 1-WL test, both at the vertex and at the graph level. Like all results discussed above, we use sum aggregation. Our construction yields GNNs of internal dimension 1 and works even with very simple weights: we may use self-weight 1, aggregation-weight 2, and bias 0 in every message-passing layer. This also resolves, under Schanuel's conjecture, an open question from Bravo et al. (2024) concerning the use of common weights in every layer. More generally, our construction applies to a broad range of rational weight choices, just as the result of Bravo et al. applies to almost every choice of real-valued weights. The appearance of Schanuel's conjecture can be understood from the arithmetic of sum aggregation. If the scalar embeddings at a given layer are linearly independent over $\mathbb{Q}$, then their sum uniquely encodes any finite multiset. To iterate this argument, the activation function must produce embeddings that are again linearly independent over $\mathbb{Q}$. Consider the basic case of exponential activation e^x. If $a_1, \ldots, a_n$ are embeddings and $m_1, \ldots, m_n$ are their multiplicities in a multiset, then $e^{m_1 a_1 + \cdots + m_n a_n} = (e^{a_1})^{m_1} \cdots (e^{a_n})^{m_n}$. Thus, different multisets naturally give rise to different monomials in the exponentials of the previous embeddings. Schanuel's conjecture provides the transcendence properties needed to ensure that these new values have the required independence.

Our connection to transcendental number theory also has a partial converse. In the slightly more expressive setting of graphs with a certain kind of auxiliary vertices akin to virtual nodes Gilmer et al. (2017); Hu et al. (2020), we prove that the existence of an L-layer GNN with exponentiation, sinh or cosh activation (and scalar embeddings on the final level) that attains the expressive power of L-round 1-WL implies the restriction of Schanuel's conjecture to the numbers that arise in its computation. This indicates that the connection between Schanuel's conjecture and GNN expressivity goes in both directions rather than being a mere artifact of our proof technique.

We complement our theoretical results with experiments. We evaluate GNNs of the form used in our construction, including simple configurations with integer weights, and observe that they indeed match the distinctions made by 1-WL.

Related Work. There has been a lot of work on extending the expressive power of GNNs beyond 1-WL, using e.g. homomorphism counts of pattern graphs Barceló et al. (2021); Maehara & NT (2024), random node initializations Abboud et al. (2021), positional distance encodings Li et al. (2020) and mechanisms inspired by higher-dimensional Weisfeiler-Leman tests Maron et al. (2019); Morris et al. (2019). Regarding Bravo et al. (2024) and Rosenbluth & Grohe (2026), we only add some minor clarifications over what was already said above. The results in Bravo et al. (2024) concern graph distinguishability, but analogous results for the vertex level can easily be derived from their proofs. The same is true for Xu et al. (2019). In contrast, Rosenbluth & Grohe (2026) have explicit results for the vertex level and for the graph level, on connected graphs. These are subject to the assumption that the graph size is provided as part of the initial embedding at every vertex. Additional results on the graph level neither need this assumption nor connectedness, but require GNNs that have additional global aggregation in every iteration. Our results also contrast those of Khalife & Tonelli-Cueto (2025) which imply that GNNs with ReLU activation can express first-order properties that cannot be expressed by any GNN activated by logistic sigmoid or hyperbolic tangent. Together with our results, assuming Schanuel's conjecture, this suggests that while logistic sigmoid can be more powerful than ReLU for exact distinguishability, ReLU can be more powerful than sigmoid for classification with a fixed margin. Indeed, the results in Khalife & Tonelli-Cueto (2025) require a uniform margin that separates the embeddings on positive and negative instances, whereas our notion of distinguishability only requires these outputs to be distinct. Schanuel's conjecture has also been used in previous work on machine learning. Assuming the conjecture, training Hankala et al. (2024) and verification Isac et al. (2023) have been shown to be decidable for FNNs with sigmoid activation. Note that these results concern computability rather than expressive power. We are not aware of previous results connecting Schanuel's conjecture to the expressive power of neural network architectures.

2 PRELIMINARIES

Abstract Algebra. We use $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$, $\bar{\mathbb{Q}}$, $\mathbb{R}$, $\mathbb{C}$ to denote natural, integer, rational, algebraic, real and complex numbers respectively. By $\mathbb{Q}_{\geq 0}$ we reference non-negative rationals. We recall that an algebraic number is a real number that is a root of a non-zero polynomial in one variable with rational coefficients. For a vector $\bar{v}$, we use $\bar{v}[i]$ to denote its i-th component. We also recall that a field consists of an underlying set F and two binary operations $+, \cdot : F \times F \to F$, subject to additional conditions. As customary, we denote a field by its underlying set, e.g. $\mathbb{Q}$ is used for the field of rational numbers with the usual addition and multiplication operations.

For two fields E and F, we say that E is a *field extension* of F, denoted E/F, if $F \subseteq E$ and the additive and multiplicative operations of E, when restricted to $F \times F$, are exactly the additive and multiplicative operations of F. For instance, $\mathbb{R}$ is a field extension of $\mathbb{Q}$. For a field extension E/F and $S \subseteq E$, we denote by $F(S)$ the field *generated* by F and S, that is, the smallest subfield G of E such that $S \subseteq G$ and G/F. For instance, $\mathbb{Q}(\sqrt{2}) = \{a + b\sqrt{2} \mid a, b \in \mathbb{Q}\}$ is the smallest subfield of $\mathbb{R}$ containing both $\mathbb{Q}$ and $\sqrt{2}$. Since $F(S)$ is a field, $\frac{P(s)}{Q(s)} \in F(S)$ for all polynomials $P(x)$ and $Q(x)$ over F and all $s \in S$ with $Q(s) \neq 0$. This relates to a family of generated fields used later in this article, namely, for any positive integer κ,

$$\mathbb{Q}(e^{1/\kappa}) = \left\{ \frac{P(e^{1/\kappa})}{Q(e^{1/\kappa})} \;\middle|\; P(x), Q(x) \text{ polynomials over } \mathbb{Q},\ Q(e^{1/\kappa}) \neq 0 \right\}.$$

A field extension E/F is *finitely generated* if there is a finite $S \subseteq E$ such that $E = F(S)$. Clearly, $\mathbb{Q}(\sqrt{2})$ and $\mathbb{Q}(e^{1/\kappa})$ are finitely generated.

Let E/F be a field extension and $S = \{s_1, \ldots, s_n\} \subseteq E$. Then S is *linearly independent* over F if $a_1 s_1 + \cdots + a_n s_n = 0$ in E only if $a_1 = \cdots = a_n = 0$, for all $a_1, \ldots, a_n \in F$. For instance, $\{1, \sqrt{2}\}$ is linearly independent over $\mathbb{Q}$. Similarly, S is *algebraically independent* over F if there is no non-zero polynomial $P(x_1, \ldots, x_n)$ with coefficients in F that satisfies $P(s_1, \ldots, s_n) = 0$. For instance, $\{\sqrt{2}\}$ is not algebraically independent over $\mathbb{Q}$ since $\sqrt{2}$ is a root of $x^2 - 2$. In contrast, $\{\pi\}$ is algebraically independent over $\mathbb{Q}$ since π is transcendental. A *transcendence basis* of a field extension E/F is a subset of E, maximal w.r.t. set inclusion, that is algebraically independent over F. For example, $\{\pi\}$ is a transcendence basis of $\mathbb{Q}(\pi)/\mathbb{Q}$. It is known that all transcendence bases share the same cardinality, see Clader & Ross (2025). It is thus meaningful to define the *transcendence degree* of E/F, denoted $\mathsf{trdeg}_F(E)$, as the cardinality of any transcendence basis of E/F. Thus $\mathsf{trdeg}_{\mathbb{Q}}(\mathbb{Q}(\pi)) = 1$ and $\mathsf{trdeg}_{\mathbb{Q}}(\mathbb{Q}(\sqrt{2})) = 0$. A field extension E/F is *algebraic* if $\mathsf{trdeg}_F(E) = 0$. Hence $\mathbb{Q}(\sqrt{2})/\mathbb{Q}$ is algebraic.

The following consequence of Corollary 6.31 of Clader & Ross (2025) will be useful.

Lemma 1. *Let E/F be a finitely generated field extension and let $S \subseteq E$. If $E/F(S)$ is algebraic and $\mathsf{trdeg}_F(E) \geq |S|$, then S is algebraically independent over F.*

Proof. Suppose that S is algebraically dependent over F, and let $T \subseteq S$ be maximal algebraically independent over F. Then $|T| < |S|$, and $F(S)/F(T)$ is algebraic. Since $E/F(S)$ is algebraic, so is $E/F(T)$. Point 1 of Corollary 6.31 of Clader & Ross (2025) states that if E/F is finitely generated and $E/F(T)$ is algebraic, then $\mathsf{trdeg}_F(E) \leq |T|$. Consequently, $\mathsf{trdeg}_F(E) \leq |T| < |S|$, contradicting $\mathsf{trdeg}_F(E) \geq |S|$. □

A classical result from transcendental number theory provides an important way of obtaining algebraically independent numbers, from Weierstrass (1885).

Theorem 2. (Lindemann–Weierstrass) *If $\{z_1, \ldots, z_n\} \subseteq \bar{\mathbb{Q}}$ is linearly independent over $\mathbb{Q}$, then $\{e^{z_1}, \ldots, e^{z_n}\}$ is algebraically independent over $\mathbb{Q}$.*

The following conjecture, due to Schanuel and first published by Lang (1966), generalizes the Lindemann–Weierstrass theorem by allowing the exponents to be arbitrary complex numbers, rather than only algebraic ones.

Conjecture 3. (Schanuel's conjecture) *If $\{z_1, \ldots, z_n\} \subseteq \mathbb{C}$ is linearly independent over $\mathbb{Q}$, then $\mathsf{trdeg}_{\mathbb{Q}}(\mathbb{Q}(z_1, \ldots, z_n, e^{z_1}, \ldots, e^{z_n})) \geq n$.*

Note that Theorem 2 is the special case of Schanuel's conjecture where $z_1, \ldots, z_n$ are algebraic. By definition, algebraic numbers contribute no transcendence degree and thus $\mathsf{trdeg}_{\mathbb{Q}}(\mathbb{Q}(z_1, \ldots, z_n, e^{z_1}, \ldots, e^{z_n})) \geq n$ implies $\mathsf{trdeg}_{\mathbb{Q}}(\mathbb{Q}(e^{z_1}, \ldots, e^{z_n})) \geq n$, which is the same as saying that $\{e^{z_1}, \ldots, e^{z_n}\}$ is algebraically independent over $\mathbb{Q}$.

Schanuel's conjecture naturally relates to activation functions defined in terms of exponentials such as sigmoid. We identify a concrete class of such functions that we shall analyze in this article. We say that a function $\sigma : \mathbb{R} \to \mathbb{R}$ is *Schanuel-ready* if it is of the form $\sigma(x) = \frac{P(e^x)}{Q(e^x)}$, with $P(y)$ and $Q(y)$ polynomials with rational coefficients, such that (i) $\frac{P(y)}{Q(y)}$ is non-constant, positive and strictly increasing on $y > 1$ and (ii) $Q(y) \neq 0$ for all $y > 0$. Notable Schanuel-ready functions include the exponentiation function $\exp = e^x$, logistic sigmoid $\frac{e^x}{e^x+1}$, hyperbolic tangent $\frac{e^{2x}-1}{e^{2x}+1}$, hyperbolic sine $\frac{e^{2x}-1}{2e^x}$, and hyperbolic cosine $\frac{e^{2x}+1}{2e^x}$.

Graphs. We work with directed graphs $G = (V, E)$ throughout this article. For each $v \in V$ we define the *neighborhood* $N(v) = \{u \in V \mid (v, u) \in E\}$ to be the set of vertices reachable from v in one step. A *pointed graph* is a pair (G, v) with $G = (V, E)$ a graph and $v \in V$ a distinguished vertex. An *embedded graph of dimension* $d > 0$ is a triple (V, E, λ) with $\lambda : V \to \mathbb{R}^d$ an *embedding*. Pointed embedded graphs are defined as expected, taking the form $((V, E, \lambda), v)$.

AC-GNNs and 1-WL. An *L-layer aggregate-combine graph neural network (AC-GNN)* is a tuple $\mathcal{N} = (\sigma, (C_i)_{1 \leq i \leq L}, (A_i)_{1 \leq i \leq L}, (b_i)_{1 \leq i \leq L}, I)$ where $\sigma : \mathbb{R} \to \mathbb{R}$ is an *activation function* applied component-wise, there are dimensions $d_0, \ldots, d_L \geq 1$ such that $A_i, C_i \in \mathbb{Q}^{d_i \times d_{i-1}}$ and $b_i \in \mathbb{Q}^{d_i}$ for $1 \leq i \leq L$, and I specifies an *initialization function*. More precisely, I is a tuple $(B_1, \ldots, B_k, c_1, \ldots, c_k)$ such that there are $d'_0, \ldots, d'_k \geq 1$ with $B_i \in \mathbb{Q}^{d'_i \times d'_{i-1}}$ and $c_i \in \mathbb{Q}^{d'_i}$ for $1 \leq i \leq k$. We require that $d'_k = d_0$ and call d'_0 the *input dimension* of $\mathcal{N}$, and d_L its *output dimension*. The *weights* of a GNN are the elements of the A_i, B_i, C_i, b_i and c_i.

As the input to GNNs, we use graphs endowed with non-negative rational vertex features. Formally, a *featured graph of dimension* n is an embedded graph (V, E, λ) of dimension n such that the image of λ contains only vectors of non-negative rationals. Let $\mathcal{N}$ be a GNN as above, and let $G = (V, E, \lambda)$ be a featured graph of dimension d'_0, the input dimension of $\mathcal{N}$. When applying $\mathcal{N}$ to G, in the first step I transforms G into an embedded graph $(V, E, \lambda^{(0)})$ of dimension d_0. Informally, I should be thought of as a feedforward neural network. Formally, we set $\widetilde{\lambda}^{(0)}(\bar{x}) = \sigma(B_k \sigma(\ldots \sigma(B_1 \bar{x} + c_1) \ldots) + c_k)$ for all $\bar{x} \in \mathbb{R}^{d'_0}$ and $\lambda^{(0)}(v) = \widetilde{\lambda}^{(0)}(\lambda(v))$ for all $v \in V$. Next, $\mathcal{N}$ constructs a sequence of L embeddings of (V, E), starting from $\lambda^{(0)}$. Let $1 \leq i \leq L$. Set $\widetilde{\lambda}^{(i)}(\bar{x}, M) = \sigma(C_i \bar{x} + A_i \sum M + b_i)$ for all $\bar{x} \in \mathbb{R}^{d_{i-1}}$ and all finite multisets M over $\mathbb{R}^{d_{i-1}}$. Then define $\lambda^{(i)}(v) = \widetilde{\lambda}^{(i)}(\lambda^{(i-1)}(v), \{\{\lambda^{(i-1)}(u) \mid u \in N(v)\}\})$ for all $v \in V$. For vertex classification, one may now work with the embedded graph $(V, E, \lambda^{(L)})$. For graph classification, a final *graph-level readout* is carried out to consolidate the graph embedding into a single vector $\lambda_G = \sum_{v \in V} \lambda^{(L)}(v)$. We refer to the functions $\widetilde{\lambda}^{(0)}, \ldots, \widetilde{\lambda}^{(L)}$ as *update functions*.

We remark that the initialization function is a slightly unusual feature of our GNN model. Its purpose is to synchronize the layers of the GNN with the rounds of the Weisfeiler–Leman algorithm. In principle, the initialization function could instead be implemented by additional GNN layers, at the cost of increasing the depth of the network.

We next introduce the 1-dimensional Weisfeiler–Leman algorithm (1-WL), in its color refinement variant; see also Grohe (2021). The algorithm takes a featured graph $G = (V, E, \lambda)$ as input and proceeds in rounds, producing increasingly refined colorings of the vertices of G. The initial coloring is defined by $\mu^{(0)}(v) = \mathsf{hash}(\lambda(v))$ where hash is an injective function whose range is a countable set of *colors*. In later rounds $i > 0$, the coloring is defined by

$$\mu^{(i)}(v) = \mathsf{hash}\Big(\mu^{(i-1)}(v), \{\{\mu^{(i-1)}(u) \mid u \in N(v)\}\}\Big).$$

When used as an approximative test for graph isomorphism, 1-WL is typically run until stabilization, that is, until $\mu^{(i)}(v_1) = \mu^{(i)}(v_2)$ implies $\mu^{(i+1)}(v_1) = \mu^{(i+1)}(v_2)$ for all vertices v_1, v_2. Here, we will mostly be interested in the coloring $\mu^{(L)}(v)$ for some fixed L. This coloring can be lifted to the graph level by taking the color of $G = (V, E)$ to be $\mu_G = \mathsf{hash}(\{\{\mu^L(v) \mid v \in V\}\})$.

Both the L-round 1-WL algorithm and every L-layer AC-GNN induce an equivalence relation on all graphs: G_1 and G_2 are equivalent if $\mu(G_1) = \mu(G_2)$ or $\lambda(G_1) = \lambda(G_2)$, respectively. Equivalence relations on pointed graphs (G, v) are induced in the same way, via $\mu^{(L)}(v)$ and $\lambda^{(L)}(v)$. It follows from Xu et al. (2019); Morris et al. (2019) that L-layer GNNs cannot classify graphs or vertices as being non-equivalent unless the L-round 1-WL does the same. The following, which can be inferred from Theorem 3 of Xu et al. (2019), states a converse. If the update functions of a GNN $\mathcal{N}$ are injective, then the equivalence relation on pointed graphs induced by $\mathcal{N}$ is exactly that of 1-WL. The same holds on the graph level if we additional demand that the image of the final update function is linearly independent over $\mathbb{Q}$.

Lemma 4. *Let $\mathcal{N}$ be an L-layer AC-GNN such that $\widetilde{\lambda}^{(0)}, \ldots, \widetilde{\lambda}^{(L)}$ are injective. Then for all pointed graphs (G_1, v_1) and (G_2, v_2), $\lambda^{(L)}(v_1) = \lambda^{(L)}(v_2)$ if and only if $\mu^{(L)}(v_1) = \mu^{(L)}(v_2)$. If, in addition, every finite subset of the image of $\widetilde{\lambda}^{(L)}$ is linearly independent over $\mathbb{Q}$, then for all graphs G_1, G_2, $\lambda_{G_1} = \lambda_{G_2}$ if and only if $\mu_{G_1} = \mu_{G_2}$.*

For our later use of Lemma 4, we remark that it suffices for the lemma to hold that the functions $\widetilde{\lambda}^{(0)}, \ldots, \widetilde{\lambda}^{(L)}$ are only *effectively injective*. For $1 \leq i \leq L$, this means that $\widetilde{\lambda}^{(i)}$ is injective on the restriction to inputs $(\bar{x}, M)$ such that there is a graph $G = (V, E)$ and a vertex $v \in V$ with $\bar{x} = \lambda^{(i-1)}(v)$ and $M = \{\{\lambda^{(i-1)}(u) \mid u \in N(v)\}\}$. Likewise, $\widetilde{\lambda}^{(0)}$ is effectively injective if it is injective on vertex labels, that is, on vectors of non-negative rationals.

3 Matching the upper bound

The following theorem states the main result of this article. When speaking of a GNN of *internal dimension 1*, we mean that all dimensions $d_0, \ldots, d_L$ from the definition of GNNs are 1 (while the dimensions $d'_1, \ldots, d'_k$ of the initialization function need not be 1).

Theorem 5. *Assume that Schanuel's conjecture is true. Then for every Schanuel-ready activation function σ and every $L \geq 1$, there is an L-layer GNN $\mathcal{N}$ of internal dimension 1 that uses σ activation and integer weights such that (i) for all pointed graphs, (G_1, v_1) and (G_2, v_2), $\lambda^{(L)}(v_1) = \lambda^{(L)}(v_2)$ if and only if $\mu^{(L)}(v_1) = \mu^{(L)}(v_2)$; and (ii) for all graphs G_1, G_2, $\lambda_{G_1} = \lambda_{G_2}$ if and only if $\mu_{G_1} = \mu_{G_2}$.*

In fact, our result is even stronger than what is stated in Theorem 5. Note that, for GNNs with internal dimension 1, the matrices A_i, C_i and bias vectors b_i are simply scalars. Our proof shows that it suffices to choose $C_i = 1$, $A_i = 2$, and $b_i = 0$ for every message-passing layer. For featured input graphs of dimension N, the matrices B_i in the initialization function use only the values $2^1, 2^2, \ldots, 2^N$, and the vectors c_i are either comprised entirely of zero or of one. Thus the weights in $\mathcal{N}$ are not only integers, but also reside in a finite set of natural numbers whose size only depends on N, but not on L. In the general setting, our construction allows all positive rational A_i and C_i satisfying $\frac{C_i}{A_i} \notin \mathbb{N}$. In this sense, our result applies to a broad range of rational weights.

We now give the technical details of our construction. Say that an L-layer GNN $\mathcal{N} = (\sigma, (C_i)_{1 \leq i \leq L}, (A_i)_{1 \leq i \leq L}, (b_i)_{1 \leq i \leq L}, I)$ of input dimension N and internal dimension 1 is *Schanuel-ready* if it uses a Schanuel-ready activation function and satisfies the following conditions:

1. the initialization function is $I = (B_1, B_2, c_1, c_2)$, where B_1 is the $N \times N$ identity matrix, $c_1 = \mathbf{1}_N$, $B_2 = (2^1, 2^2, \ldots, 2^N)$, and $c_2 = 0$. Then $\lambda^{(0)}(v) = \sigma\big(\sum_{j=1}^{N} 2^j \sigma(\lambda(v)[j]+1)\big)$ for all graphs $G = (V, E)$ and vertices $v \in V$;

2. $A_i, C_i > 0$ and $b_i \geq 0$ are rational numbers such that $\frac{C_i}{A_i} \notin \mathbb{N}$.

The purpose of the above initialization function I is to encode a vector of non-negative rational numbers, as given by the input features, into a single real number. One can show that, under Schanuel's conjecture, the values $\sigma(q + 1)$ are linearly independent over $\mathbb{Q}$, preventing different feature values from canceling out in the sum. The powers 2^j provide, for each rational value, a binary encoding of the positions in which it occurs. In the appendix, we prove the following.

Proposition 6. *Assuming Schanuel's conjecture is true, every L-layered Schanuel-ready GNN $\mathcal{N}$ satisfies $\lambda^{(L)}(v_1) = \lambda^{(L)}(v_2)$ if and only if $\mu^{(L)}(v_1) = \mu^{(L)}(v_2)$, for all pointed graphs (G_1, v_1) and (G_2, v_2).*

In what follows, we present a proof of a restricted version of Proposition 6 that is less technical, yet still reveals the underlying mechanisms. More concretely, we assume that the activation function is $\sigma = \exp$, $C_i = 1$, $A_i = 2$, and $b_i = 0$ for every layer i, and all feature vectors in input graphs are one-hot, that is, they have exactly one non-zero component. The assumption of one-hot feature vectors also allows us to simplify the initialisation to be $I = (B_1, c_1)$, where $B_1 = (1, 2, \ldots, N)$ and $c_1 = 0$. Then $\lambda^{(0)}(v) = \sigma\big(\sum_{j=1}^{n} j\, \lambda(v)[j]\big)$ for all graphs $G = (V, E)$ and vertices $v \in V$.

We now define, for every layer i of our GNN $\mathcal{N}$, a set of vectors Z_i that overapproximates the set of embedding vectors that can arise in layer i of $\mathcal{N}$, across all graphs. Since the sets Z_i can be infinite and Schanuel's conjecture applies only to finite sets, we shall also define, for each $r \geq 1$, a finite set $Z_{i,r}$. The reader should think of the set $Z_{1,2r+1}$ in the same way as Z_i, but on graphs of out-degree at most r. We proceed inductively:

$$\widehat{Z}_{0,r} = \{e\} \qquad Z_{0,r} = \{e^1, \ldots, e^r\},$$
$$\widehat{Z}_{i+1,r} = \{e^z \mid z \in Z_{i,r}\} \qquad Z_{i+1,r} = \{z_1 \cdots z_r \mid z_1, \ldots, z_r \in \widehat{Z}_{i+1,r} \cup \{1\} \text{ not all } 1\}.$$

Importantly, $z_1, \ldots, z_r$ do not need to be pairwise distinct. Set $Z_i = \bigcup_{r \geq 1} Z_{i,r}$ for $0 \leq i \leq L$.

Lemma 7. *For all featured graphs G, all vertices $v \in V$ and $0 \leq i \leq L$ we have $\lambda^{(i)}(v) \in Z_i$.*

Proof. We proceed by induction on i. The base case is immediate as the image of the initialisation function $\lambda^{(0)}$ is $\{e^1, \ldots, e^N\} \subseteq Z_0$. For the inductive step assume that $\lambda^{(i)}(v) \in Z_i$ for all $v \in V$. Then, aggregation and combination in $\mathcal{N}$ amount to $\sigma\big(z_0 + 2(z_1 + \cdots + z_n)\big)$ with $z_0, \ldots, z_n \in Z_i$. Applying σ yields $(e^{z_0}) \cdot (e^{z_1})^2 \cdot \ldots \cdot (e^{z_n})^2$, which is in Z_{i+1} as $e^{z_0}, \ldots, e^{z_n} \in \widehat{Z}_{i+1}$. □

All that is left to show at this point is that each finite subset of Z_i is linearly independent over $\mathbb{Q}$ for all $0 \leq i \leq L$. Indeed, then one need only apply Lemma 4 noting that $\widetilde{\lambda}^{(0)}(s) = \sigma((1, \ldots, N) \times z)$ and $\widetilde{\lambda}^{(i)}(s, M) = \sigma(s + \sum 2M)$ are injective when z is one-hot, $s \in Z_{i-1}$, and M is a finite multiset over Z_{i-1} (see appendix for details). To prove the required linear independence of finite subsets of Z_i, we first observe that it follows from algebraic independence of finite subsets of $\widehat{Z}_i = \bigcup_{r \geq 1} \widehat{Z}_{i,r}$ over $\mathbb{Q}$. For technical reasons, we establish a stronger result:

Lemma 8. *Let $0 \leq i \leq L$. If each finite subset of $\bigcup_{j=0}^{i} \widehat{Z}_j$ is algebraically independent over $\mathbb{Q}$, then each finite subset of $\bigcup_{j=0}^{i} Z_j \cup \{1\}$ is linearly independent over $\mathbb{Q}$.*

Proof. Suppose for contradiction that some finite $S \subseteq \bigcup_{j=0}^{i} Z_j \cup \{1\}$ is linearly dependent over $\mathbb{Q}$. Thus, for $S \setminus \{1\} = \{z_1 \ldots, z_n\}$ there are $c_1, \ldots, c_n, d \in \mathbb{Q}$ not all zero such that $\sum_{j=1}^{n} c_j z_j + d = 0$. Each value z_j is a finite product of elements in $\bigcup_{\ell=0}^{i} \widehat{Z}_\ell$. This follows directly from the definition for $z_j \in Z_\ell$ with $\ell \geq 1$, while every element $e^k \in Z_0$ is the product of k copies of the unique element $e \in \widehat{Z}_0$. We can thus write $z_j = s_1^{b_{1,j}} \cdot \ldots \cdot s_m^{b_{m,j}}$ where $b_{1,j}, \ldots, b_{m,j} \in \mathbb{N}$ not all zero and $s_1, \ldots, s_m \in \bigcup_{j=0}^{i} \widehat{Z}_j$. But then, $\sum_{j=1}^{n} c_j s_1^{b_{1,j}} \cdot \ldots \cdot s_m^{b_{m,j}} + d = 0$. Hence, by algebraic independence of $s_1, \ldots, s_m$ over $\mathbb{Q}$, the function $\sum_{j=1}^{n} c_j x_1^{b_{1,j}} \cdot \ldots \cdot x_m^{b_{m,j}} + d$ must be the zero polynomial. But since $z_1, \ldots, z_n$ are distinct, each vector of exponents $(b_{1,j}, \ldots, b_{m,j})$ must also be distinct. In fact, $(b_{1,j}, \ldots, b_{m,j}) \neq (0, \ldots, 0)$. Thus, no cancellations take place in the expression above, resulting in a non-zero polynomial which contradicts our initial assumption. □

At this point, it remains to show the following.

Lemma 9. *Assuming that Schanuel's conjecture is true, all finite subsets of $\bigcup_{j=0}^{L} \widehat{Z}_j$ are algebraically independent over $\mathbb{Q}$.*

Proof. Assuming Schanuel's conjecture, we prove by induction on i that, for $0 \leq i \leq L$ and $r \geq 1$,

1. $\bigcup_{j=0}^{i} \widehat{Z}_{j,r}$ is algebraically independent over $\mathbb{Q}$ and
2. the sets $\widehat{Z}_{0,r}, \ldots, \widehat{Z}_{i,r}$ are pairwise disjoint; and so are $\{1\}, Z_{0,r}, \ldots, Z_{i,r}$.

This clearly implies the intended result. For the base case note that $\bigcup_{j=0}^{0} \widehat{Z}_j = \{e\}$. Thus Condition 1 holds trivially as e is a transcendental number. Condition 2 follows from the above and the fact that $Z_{0,r}$ are positive powers of the transcendental number e, thus ensuring $1 \notin Z_{0,r}$.

Now suppose Conditions 1 and 2 hold for $i \geq 0$. We show that the same is true for $i+1$. By Condition 1 of our induction hypothesis, $\bigcup_{j=0}^{i} \widehat{Z}_{j,r}$ is algebraically independent over $\mathbb{Q}$. Thus by Lemma 8, the set $S = \bigcup_{j=0}^{i} Z_{j,r} \cup \{1\}$ is linearly independent over $\mathbb{Q}$. Applying Schanuel's conjecture to S yields $F = \mathbb{Q}(S \cup \bigcup_{j=0}^{i+1} \widehat{Z}_{j,r})$ such that $\mathsf{trdeg}_{\mathbb{Q}}(F) \geq |S|$. We next analyse the size of S. Notice, first, that $|Z_{j,r}| = |\widehat{Z}_{j+1,r}|$ for all j. Moreover, by Condition 2 of our induction hypothesis, $\{1\}, Z_{0,r}, \ldots, Z_{i,r}$ are all pairwise disjoint. Thus,

$$|S| = |\bigcup_{j=0}^{i} Z_{j,r} \cup \{1\}| = \sum_{j=0}^{i} |Z_{j,r}| + 1 = \sum_{j=1}^{i+1} |\widehat{Z}_{j,r}| + |\widehat{Z}_{0,r}| = \sum_{j=0}^{i+1} |\widehat{Z}_{j,r}|.$$

Write $E = \mathbb{Q}(\bigcup_{j=0}^{i+1} \widehat{Z}_{j,r})$. Notice that $S \subseteq E$ because elements of S are monomials over $\bigcup_{j=0}^{i} \widehat{Z}_{j,r}$ or 1. Hence, $E = F$ and $\mathsf{trdeg}_{\mathbb{Q}}(E) \geq |S| = \sum_{j=0}^{i+1} |\widehat{Z}_{j,r}|$. With Lemma 1, we obtain that $\bigcup_{j=0}^{i+1} \widehat{Z}_{j,r}$ is algebraically independent over $\mathbb{Q}$, thus establishing Condition 1.

The $\widehat{Z}_{0,r}, \ldots, \widehat{Z}_{i+1,r}$ part of Condition 2 follows from the facts that the sets $\{1\}, Z_{0,r}, \ldots, Z_{i,r}$ are pairwise disjoint, due to the induction hypothesis, and exponentiation is injective. Now, suppose that $Z_{i+1,r} \cap Z_{j,r} \neq \emptyset$ for $0 \leq j \leq i$, with z a common element. Clearly, z is then a monomial over $\widehat{Z}_{i+1,r}$ as well as over $\widehat{Z}_{j,r}$. But by our argument above, $\widehat{Z}_{i+1,r} \cap \widehat{Z}_{j,r} = \emptyset$. Additionally, by Condition 1, already established above for $i+1$, $\widehat{Z}_{i+1,r} \cup \widehat{Z}_{j,r}$ is algebraically independent over $\mathbb{Q}$. Hence, no monomial over $\widehat{Z}_{i+1,r}$ equates to a monomial over $\widehat{Z}_{j,r}$, a contradiction. Finally, $Z_{i+1,r} \cap \{1\} = \emptyset$ follows from the fact that 1 is algebraic whilst the elements of $Z_{i+1,r}$ are not. □

4 The Transcendence of GNN Computations

The aim of this section is to provide a partial converse to Theorem 5. In doing so we argue that the connection between GNN expressivity and Schanuel's conjecture goes both ways, and is not merely an artefact of our technique. To achieve this, we work in a slightly more expressive setting that admits virtual nodes. These are auxiliary vertices commonly used in GNNs to globally aggregate information, see for instance Gilmer et al. (2017); Hu et al. (2020). We use a variant tailored to our purposes in which virtual nodes need not be globally connected and, crucially, use a modified update function that does not depend on their own state. We also make the additional assumption that the embedding computed by the final GNN layer is of dimension 1.

More precisely, a *featured graph with virtual nodes* is a triple $G = (V, E, \lambda)$, where $\lambda : V \to \mathbb{Q}^N_{\geq 0} \cup \{\mathsf{virtual}\}$. Vertices v with $\lambda(v) = \mathsf{virtual}$ are called *virtual*, all other vertices are *ordinary*. We extend our definition of GNNs accordingly. An *L-layer GNN with virtual nodes* is a tuple

$$\mathcal{N} = \big(\sigma, (C_i)_{1 \leq i \leq L}, (A_i)_{1 \leq i \leq L}, (b_i)_{1 \leq i \leq L}, (A'_i)_{1 \leq i \leq L}, (b'_i)_{1 \leq i \leq L}, I\big),$$

where σ, C_i, A_i, b_i and I are as in the definition of GNNs with dimensions $d_0, \ldots, d_L$, $A'_i \in \mathbb{Q}^{d_i \times d_{i-1}}$ and $b'_i \in \mathbb{Q}^{d_i}$ for $1 \leq i \leq L$. For every node v with $\lambda(v) = \mathsf{virtual}$, the initial embedding $\lambda^{(0)}(v)$ is some vector shared by all virtual nodes, whilst $\lambda^{(i)}(v) = \sigma\big(A'_i \sum_{u \in N(v)} \lambda^{(i-1)}(u) + b'_i\big)$.

We also consider a *1-WL test over graphs with virtual nodes*: for vertices u with $\lambda(u) = \mathsf{virtual}$, we define $\mu^{(i+1)}(u) = \textsc{hash}(\mathsf{virtual}, \{\{\mu^{(i)}(v) \mid v \in N(u)\}\})$ where $\mathsf{virtual}$ is a distinguished value that is never assigned to an ordinary vertex. The following is routine to prove, see the appendix.

Proposition 10. *Suppose (G, v) and (H, u) are two featured pointed graphs with virtual nodes. Then, for all L-layered GNNs $\mathcal{N}$, $\mu^{(L)}(v) = \mu^{(L)}(u)$ implies $\lambda^{(L)}(v) = \lambda^{(L)}(u)$ and $\mu_G = \mu_H$ implies $\lambda_G = \lambda_H$.*

We are now in a position to establish the advertised converse.

Theorem 11. *Let $\mathcal{N}$ be an L-layer GNN that uses $\sigma \in \{\exp, \sinh, \cosh\}$ as activation and produces 1-dimensional vertex embeddings in its final layer. Further assume that $\lambda_{G_1} = \lambda_{G_2}$ if and only if*

$\mu_{G_1} = \mu_{G_2}$ for all featured graphs with virtual nodes G_1, G_2. Then for every featured graph $G = (V, E, \lambda)$ the values $\{\exp(A'_L \lambda^{(L-1)}(v)) \mid v \in V\}$ are algebraically independent over $\mathbb{Q}$.

Proof. We only treat $\sigma = \exp$ here. The cases $\sigma \in \{\sinh, \cosh\}$ are relegated to the appendix. Observe that, for every featured graph with virtual nodes $G = (V, E, \lambda)$ and every $u \in V$ with $\lambda(u) = \mathsf{virtual}$, $\lambda^{(L)}(u) = \exp\big(A'_L \sum_{v \in N(u)} \lambda^{(L-1)}(v) + b'_L\big) = \exp(b'_L) \prod_{v \in N(u)} \exp\big(A'_L \lambda^{(L-1)}(v)\big)$. Thus, the final embedding of a virtual node is $\exp(b'_L)$ times a monomial in the values $\mathcal{V} = \{\exp(A'_L \lambda^{(L-1)}(v)) \mid v \in V\}$ where the exponent of each value records how often it occurs among the neighbors of the virtual node. Towards a contradiction, assume that for some graph $G = (V, E, \lambda)$ this set is not algebraically independent over $\mathbb{Q}$. We construct two graphs G_Q and G_R that are distinguishable by L-round 1-WL but not by $\mathcal{N}$, contradicting our assumption.

Let $z_1, \ldots, z_n$ be an enumeration of $\mathcal{V}$. For $1 \leq i \leq n$ choose a vertex $v_i \in V$ such that $z_i = \exp(A'_L \lambda^{(L-1)}(v_i))$. Since the z_i are pairwise distinct, so are the embeddings $\lambda^{(L-1)}(v_i)$. By Proposition 10, the pointed graphs $(G, v_1), \ldots, (G, v_n)$ are all pairwise distinguishable by the $(L-1)$-round 1-WL algorithm. Now, since $\{z_1, \ldots, z_n\}$ is algebraically dependent over $\mathbb{Q}$, there is a non-zero polynomial $P(x_1, \ldots, x_n)$ such that $P(z_1, \ldots, z_n) = 0$. Multiplying by a common denominator, we may assume that P has integer coefficients. Moving all terms with negative coefficients to the right hand side, we obtain a polynomial equality $Q(x_1, \ldots, x_n) = R(x_1, \ldots, x_n)$ where Q and R have disjoint sets of monomials and non-negative integer coefficients. Exchanging Q and R if necessary, we may assume that R has no constant term. Multiplying by $\exp(b'_L)$ and writing out the monomials gives

$$\sum_{i=1}^{q} \alpha_i z_1^{\beta_{i,1}} \cdots z_n^{\beta_{i,n}} \exp(b'_L) + \alpha_{q+1} \exp(b'_L) = \sum_{j=1}^{r} \gamma_j z_1^{\delta_{j,1}} \cdots z_n^{\delta_{j,n}} \exp(b'_L),$$

where $\alpha_i, \gamma_j \in \mathbb{N} \setminus \{0\}$, with the exception that α_{q+1} may be zero, and the exponent vectors $(\beta_{i,1}, \ldots, \beta_{i,n})$ and $(\delta_{j,1}, \ldots, \delta_{j,n})$ are not all zero entries. The desired graphs G_Q and G_R realize the two sides of this equality. A monomial $z_1^{\beta_{i,1}} \cdots z_n^{\beta_{i,n}}$ is realized by a virtual node having, for every k, $\beta_{i,k}$ neighbors in fresh copies of (G, v_k). By the initial observation in this proof, its final embedding is $\exp(b'_L) z_1^{\beta_{i,1}} \cdots z_n^{\beta_{i,n}}$. An isolated virtual node, in turn, has final embedding $\exp(b'_L)$ and thus realizes the constant monomial. To ensure that all vertices belonging to the copies of G make the same contribution on both sides, we additionally put on each side copies corresponding to the monomials realized on the other side. Formally, G_Q is defined as follows:

1. for $1 \leq i \leq q$, add α_i disjoint copies of the graph constructed as follows:
   (a) for $1 \leq k \leq n$, add $\beta_{i,k}$ fresh copies of the pointed graph (G, v_k);
   (b) create a fresh virtual node u and add an edge from u to the distinguished vertex in each copy added in the previous step;
2. add α_{q+1} fresh isolated virtual vertices;
3. for $1 \leq j \leq r$ and $1 \leq k \leq n$, add $\gamma_j \delta_{j,k}$ disjoint copies of (G, v_k).

The graph G_R is defined symmetrically: interchange q and r as well as (α, β) and (γ, δ), and omit Step 2. Notice that the newly added edges point from virtual nodes into the copies of G and thus do not change the neighborhoods, and hence the GNN embeddings or 1-WL colors, of vertices inside these copies. We argue in the appendix that G_Q and G_R have the desired properties. □

Theorem 11 immediately yields the following restriction of Schanuel's conjecture to GNN outputs.

Corollary 12. *Let $\mathcal{N}$ be as in Theorem 11. Then, for every featured graph $G = (V, E, \lambda)$ and all distinct $x_1, \ldots, x_n \in \{A'_L \lambda^{(L-1)}(v) \mid v \in V\}$, we have $\mathsf{trdeg}_{\mathbb{Q}} \mathbb{Q}(x_1, \ldots, x_n, e^{x_1}, \ldots, e^{x_n}) \geq n$. Consequently, Schanuel's conjecture holds when restricted to $\mathbb{Q}$-linearly independent tuples arising in this way.*

Proof. Let $x_1, \ldots, x_n$ be as in the statement. By Theorem 11, $e^{x_1}, \ldots, e^{x_n}$ are algebraically independent over $\mathbb{Q}$. Hence $\mathrm{trdeg}_{\mathbb{Q}} \mathbb{Q}(e^{x_1}, \ldots, e^{x_n}) = n$. Since $\mathbb{Q}(e^{x_1}, \ldots, e^{x_n}) \subseteq \mathbb{Q}(x_1, \ldots, x_n, e^{x_1}, \ldots, e^{x_n})$, the claimed inequality follows, which is exactly the conclusion of Schanuel's conjecture. □

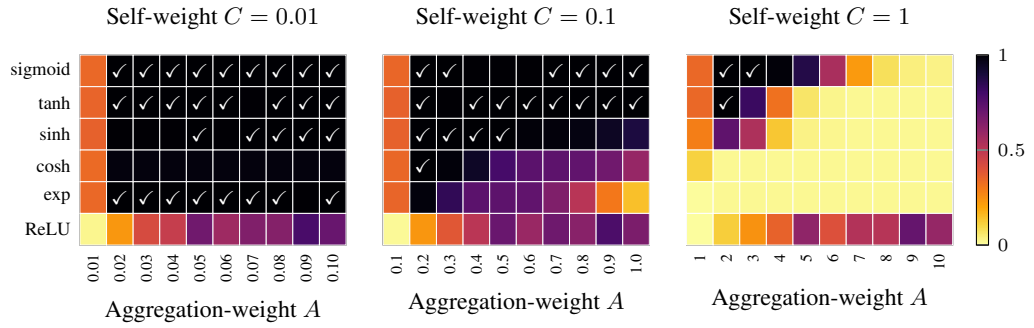

Figure 1: Success ratio in float64 arithmetic for GNNs with activation (row) and aggregation-weight (column) at fixed self-weights in three weight categories. Checkmarks denote success ratio 1.

5 Experiments

We empirically investigate the vertex distinguishing power of GNNs using sigmoid, tanh, sinh, cosh and exp as prominent Schanuel-ready activation functions. In addition, we evaluate ReLU as a popular activation function that is not Schanuel-ready. Since Schanuel-ready activation functions may compute transcendental numbers from rational ones, we vary the precision and use 32-, 64- and 128-bit floating point arithmetics. We randomly generate $N = 500$ Erdős–Rényi graphs with $n \in [20, 100]$ vertices using edge probability[1] $\frac{2}{n}$, with one-hot feature vectors from $\mathbb{Q}^4$. On each graph, we run 1-WL. If L rounds are needed for stabilization, we construct L-layer GNNs $(\sigma, (C_i)_{1\leq i\leq L}, (A_i)_{1\leq i\leq L}, (b_i)_{1\leq i\leq L}, I)$ where $b_1 = \cdots = b_L = 0$ and $C_1 = \cdots = C_L$ as well as $A_1 = \cdots = A_L$ range over $\{0.01, 0.02, \ldots, 0.1\}$, $\{0.1, 0.2, \ldots, 1\}$ and $\{1, 2, \ldots, 10\}$. We choose I so that the initialisation function is $\widetilde{\lambda}^{(0)}(x) = \sigma((2^{-1}, 2^{-2}, 2^{-3}, 2^{-4}) \times x)$. Say that $\mathcal{N}$ *succeeds* on a graph $G = (V, E, \lambda)$ if for all $u, v \in V$, $\mu^{(L)}(v) = \mu^{(L)}(u)$ if and only if $\lambda^{(L)}(v) = \lambda^{(L)}(u)$. The *success ratio* of $\mathcal{N}$ is the number of graphs that $\mathcal{N}$ succeeds on divided by N.

An excerpt of our results for float64 is displayed in Figure 1, see the appendix for 32- and 128-bit results. For many weight settings, GNNs with Schanuel-ready activation functions perfectly simulate the 1-WL algorithm. The exception of $\frac{C}{A} \in \mathbb{N}$ predicted by our theoretical analysis is observable in the leftmost columns. All Schanuel-ready activation functions encounter a performance degradation when weights get larger, which we believe is because computed values get exceedingly close to 1 and cannot be distinguished by a fixed-precision floating-point representation (for sigmoid and tanh), or overflow (for sinh, cosh, and exp). These different phenomena can also be used to explain why the performance of sinh, cosh, and exp degrades faster than that of sigmoid and tanh (see appendix). Regardless of which Schanuel-ready activation function is used, higher precision floats result in better performance. As expected from the results in Khalife & Basu (2026), ReLU GNNs exhibit weaker performance than GNNs with Schanuel-ready activation functions.

6 Conclusions

We have studied the capability of GNNs with constant depth and rational weights to match the expressive power of 1-WL, establishing an intimate connection to algebraic independence and, more specifically, Schanuel's conjecture. As future work, we would be interested in understanding whether the assumptions made in Section 4 can be weakened. Can the same result be proved for sigmoid or all Schanuel-ready activation functions? Is it necessary to assume the presence of virtual nodes and that the embedding computed by the final GNN layer is of dimension 1? Can the restricted form of Schanuel's conjecture obtained in Corollary 12 be strengthened, possibly even to Schanuel's conjecture without any restrictions? We conjecture that our methods can be extended to more expressive GNN architectures and their natural WL test counterparts; e.g. k-GNNs and the k-WL test, GNNs with global readout in each layer and 1-WL with global aggregation. Such directions are left for future research.

[1]We avoid graphs generated by a fixed edge probability as in this setting 1-WL is known to, with high probability, attribute distinct colours to all vertices in a graph; see Table 1 of Verbitsky & Zhukovskii (2025).

AI USE STATEMENT

The generative AI model ChatGPT-5.6 Sol (Max) was used as an aid when developing this paper. The extent of the model's involvement is as follows. All text (including the appendix) has been human written with occasional paraphrasing requests relegated to AI. All mathematical claims and statements have been formulated by humans without help of AI. All proofs excluding Claims 14, 15 and 21 were first obtained by humans without AI help. Claims 14, 15 and 21 have been realised by strengthening our initial human-generated proofs with AI guidance. Completed proofs were given to the AI agent in order to identify gaps and inaccuracies. Some suggestions arising from the above have been implemented. Figures have been generated by AI. Some references have been recommended by AI agents. Experimental code has been written with the help of an AI coding agent. The authors have verified the AI-generated figures, proof suggestions, code and references, and assume responsibility for all contents of this work.

REPRODUCIBILITY STATEMENT

Since our theoretical results rely on Schanuel's conjecture, we explicitly indicated this dependence in claim, lemma, proposition and theorem statements. The appendix houses full proofs and additional experimental results omitted from the main body of text. We provide code for running our experiments in the supplementary material with documentation.

ETHICS STATEMENT

Our work aims to increase understanding in the expressive power of GNNs. Our results are theoretic in nature. We therefore believe that this work is low risk in terms of possible misuse.

REFERENCES

Anders Aamand, Justin Chen, Piotr Indyk, Shyam Narayanan, Ronitt Rubinfeld, Nicholas Schiefer, Sandeep Silwal, and Tal Wagner. Exponentially improving the complexity of simulating the weisfeiler-lehman test with graph neural networks. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh (eds.), *Advances in Neural Information Processing Systems*, volume 35, pp. 27333–27346. Curran Associates, Inc., 2022. doi: 10.52202/068431-1982. URL https://proceedings.neurips.cc/paper_files/paper/2022/file/af0ad514b9cda46bd49e14ee11e2672f-Paper-Conference.pdf.

Ralph Abboud, İsmail İlkan Ceylan, Martin Grohe, and Thomas Lukasiewicz. The surprising power of graph neural networks with random node initialization. In Zhi-Hua Zhou (ed.), *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI-21*, pp. 2112–2118. International Joint Conferences on Artificial Intelligence Organization, 8 2021. doi: 10.24963/ijcai.2021/291. URL https://doi.org/10.24963/ijcai.2021/291. Main Track.

Tal Amir, Steven Gortler, Ilai Avni, Ravina Ravina, and Nadav Dym. Neural injective functions for multisets, measures and graphs via a finite witness theorem. In A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine (eds.), *Advances in Neural Information Processing Systems*, volume 36, pp. 42516–42551. Curran Associates, Inc., 2023. doi: 10.52202/075280-1845. URL https://proceedings.neurips.cc/paper_files/paper/2023/file/84b686f7cc7b7751e9aaac0da74f755a-Paper-Conference.pdf.

Pablo Barceló, Floris Geerts, Juan Reutter, and Maksimilian Ryschkov. Graph neural networks with local graph parameters. In M. Ranzato, A. Beygelzimer, Y. Dauphin, P.S. Liang, and J. Wortman Vaughan (eds.), *Advances in Neural Information Processing Systems*, volume 34, pp. 25280–25293. Curran Associates, Inc., 2021. URL https://proceedings.neurips.cc/paper_files/paper/2021/file/d4d8d1ac7e00e9105775a6b660dd3cbb-Paper.pdf.

César Bravo, Alexander Kozachinskiy, and Cristobal Rojas. On dimensionality of feature vectors in MPNNs. In Ruslan Salakhutdinov, Zico Kolter, Katherine Heller, Adrian Weller, Nuria

Oliver, Jonathan Scarlett, and Felix Berkenkamp (eds.), *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 4472–4481. PMLR, 21–27 Jul 2024. URL https://proceedings.mlr.press/v235/bravo24a.html.

Emily Clader and Dustin Ross. *Beginning in Algebraic Geometry*. Undergraduate Texts in Mathematics. Springer, 2025. ISBN 978-3-031-88819-9. doi: 10.1007/978-3-031-88819-9.

Justin Gilmer, Samuel S. Schoenholz, Patrick F. Riley, Oriol Vinyals, and George E. Dahl. Neural message passing for quantum chemistry. In *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pp. 1263–1272. PMLR, 2017.

Martin Grohe. The logic of graph neural networks. In *2021 36th Annual ACM/IEEE Symposium on Logic in Computer Science (LICS)*, pp. 1–17, 2021. doi: 10.1109/LICS52264.2021.9470677.

Teemu Hankala, Miika Hannula, Juha Kontinen, and Jonni Virtema. Complexity of neural network training and ETR: extensions with effectively continuous functions. In Michael J. Wooldridge, Jennifer G. Dy, and Sriraam Natarajan (eds.), *Thirty-Eighth AAAI Conference on Artificial Intelligence, AAAI 2024, Thirty-Sixth Conference on Innovative Applications of Artificial Intelligence, IAAI 2024, Fourteenth Symposium on Educational Advances in Artificial Intelligence, EAAI 2024, February 20-27, 2024, Vancouver, Canada*, pp. 12278–12285. AAAI Press, 2024. doi: 10.1609/AAAI.V38I11.29118. URL https://doi.org/10.1609/aaai.v38i11.29118.

Weihua Hu, Matthias Fey, Marinka Zitnik, Yuxiao Dong, Hongyu Ren, Bowen Liu, Michele Catasta, and Jure Leskovec. Open graph benchmark: Datasets for machine learning on graphs. In *Advances in Neural Information Processing Systems*, volume 33, pp. 22118–22133, 2020.

Omri Isac, Yoni Zohar, Clark Barrett, and Guy Katz. DNN Verification, Reachability, and the Exponential Function Problem. In Guillermo A. Pérez and Jean-François Raskin (eds.), *34th International Conference on Concurrency Theory (CONCUR 2023)*, volume 279 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pp. 26:1–26:18, Dagstuhl, Germany, 2023. Schloss Dagstuhl – Leibniz-Zentrum für Informatik. ISBN 978-3-95977-299-0. doi: 10.4230/LIPIcs.CONCUR.2023.26. URL https://drops.dagstuhl.de/entities/document/10.4230/LIPIcs.CONCUR.2023.26.

Sammy Khalife and Amitabh Basu. On the power of graph neural networks and the role of the activation function. *SIAM Journal on Applied Algebra and Geometry*, 10(2):350–365, 2026. doi: 10.1137/24M1659704. URL https://doi.org/10.1137/24M1659704.

Sammy Khalife and Josué Tonelli-Cueto. Is uniform expressivity too restrictive? towards efficient expressivity of GNNs. In *The Thirteenth International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=lsvGqR6OTf.

Serge Lang. *Introduction to Transcendental Numbers*. Addison-Wesley, Reading, MA, 1966.

Pan Li, Yanbang Wang, Hongwei Wang, and Jure Leskovec. Distance encoding: Design provably more powerful neural networks for graph representation learning. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin (eds.), *Advances in Neural Information Processing Systems*, volume 33, pp. 4465–4478. Curran Associates, Inc., 2020. URL https://proceedings.neurips.cc/paper_files/paper/2020/file/2f73168bf3656f697507752ec592c437-Paper.pdf.

Takanori Maehara and Hoang NT. Deep homomorphism networks. In A. Globerson, L. Mackey, D. Belgrave, A. Fan, U. Paquet, J. Tomczak, and C. Zhang (eds.), *Advances in Neural Information Processing Systems*, volume 37, pp. 56076–56107. Curran Associates, Inc., 2024. doi: 10.52202/079017-1784. URL https://proceedings.neurips.cc/paper_files/paper/2024/file/65f54fdf62cd5614dc5715ae7ece4ef6-Paper-Conference.pdf.

Haggai Maron, Heli Ben-Hamu, Hadar Serviansky, and Yaron Lipman. Provably powerful graph networks. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett (eds.), *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019. URL https://proceedings.neurips.cc/paper_files/paper/2019/file/bb04af0f7ecaee4aae62035497da1387-Paper.pdf.

Christopher Morris, Martin Ritzert, Matthias Fey, William L. Hamilton, Jan Eric Lenssen, Gaurav Rattan, and Martin Grohe. Weisfeiler and leman go neural: Higher-order graph neural networks. In *Proceedings of the Thirty-Third AAAI Conference on Artificial Intelligence, AAAI 2019*, pp. 4602–4609. AAAI Press, 2019. doi: 10.1609/AAAI.V33I01.33014602. URL https://doi.org/10.1609/aaai.v33i01.33014602.

Eran Rosenbluth and Martin Grohe. Repetition makes perfect: Recurrent graph neural networks match message passing limit. In Sven Koenig, Chad Jenkins, and Matthew E. Taylor (eds.), *Fortieth AAAI Conference on Artificial Intelligence, Thirty-Eighth Conference on Innovative Applications of Artificial Intelligence, Sixteenth Symposium on Educational Advances in Artificial Intelligence, AAAI 2026, Singapore, January 20-27, 2026*, pp. 25168–25175. AAAI Press, 2026. doi: 10.1609/AAAI.V40I30.39707. URL https://doi.org/10.1609/aaai.v40i30.39707.

Oleg Verbitsky and Maksim Zhukovskii. Canonical labeling of sparse random graphs. In Olaf Beyersdorff, Michał Pilipczuk, Elaine Pimentel, and Nguy˜ên Kim Thng (eds.), *42nd International Symposium on Theoretical Aspects of Computer Science (STACS 2025)*, volume 327 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pp. 75:1–75:20, Dagstuhl, Germany, 2025. Schloss Dagstuhl – Leibniz-Zentrum für Informatik. ISBN 978-3-95977-365-2. doi: 10.4230/LIPIcs.STACS.2025.75. URL https://drops.dagstuhl.de/entities/document/10.4230/LIPIcs.STACS.2025.75.

Karl Weierstrass. Zu Lindemann's Abhandlung: "über die Ludolph'sche Zahl". *Sitzungsberichte der Königlich Preussischen Akademie der Wissenschaften zu Berlin*, pp. 1067–1085, 1885.

Keyulu Xu, Weihua Hu, Jure Leskovec, and Stefanie Jegelka. How powerful are graph neural networks? In *International Conference on Learning Representations*, 2019. URL https://openreview.net/forum?id=ryGs6iA5Km.

APPENDIX

A FULL PROOF OF PROPOSITION 6

Let us recall the GNNs we are working with. Recall that an activation function σ is Schanuel-ready if there exist polynomials $P(y)$ and $Q(y)$ with rational coefficients such that

- $\sigma(x) = \frac{P(e^x)}{Q(e^x)}$,
- $\frac{P(y)}{Q(y)}$ is non-constant, positive and strictly increasing on $y > 1$,
- $Q(y) \neq 0$ for all $y > 0$.

When σ is clear from context we will often refer to $\frac{P(y)}{Q(y)}$ as $R(y)$.

Now, writing N for the dimensionality of initial feature vectors recall that, in our full setting, an L-layered GNN $\mathcal{N}$ is Schanuel-ready if

- each layer uses a common Schanuel-ready activation function σ,
- initialisation takes the form $\lambda^{(0)}(v) = \sigma((2^1, 2^2, \ldots, 2^N) \times \sigma(\lambda(v) + \mathbf{1}_N))$,
- each layer computes $\lambda^{(i)}(v) = \sigma(C_i\lambda^{(i-1)}(v) + A_i \sum_{u \in N(v)} \lambda^{(i-1)}(u) + b_i)$, where $A_i, C_i > 0$, $b_i \geq 0$ are rational numbers such that $\frac{C_i}{A_i} \notin \mathbb{N}$.

We again note that a GNN having $A_1 = \cdots = A_L = 2$, $C_1 = \cdots = C_L = 1$ and $b_1 = \cdots = b_L = 0$ is a valid choice for our purposes. Notice that each $\lambda^{(0)}, \ldots, \lambda^{(L)}$ produces 1-dimensional feature vectors, but λ – the feature map of any given graph – need not be so. We now prove the following in its full form:

Proposition 6. *Assuming Schanuel's conjecture is true, every L-layered Schanuel-ready GNN $\mathcal{N}$ satisfies $\lambda^{(L)}(v_1) = \lambda^{(L)}(v_2)$ if and only if $\mu^{(L)}(v_1) = \mu^{(L)}(v_2)$, for all pointed graphs (G_1, v_1) and (G_2, v_2).*

The overall proof structure follows the one in the main body. The setting, however, is more technical as we aim to accommodate a wider range of activation functions and the scope of graphs with non-negative rational features. Thus, recall that $\sigma(x) = \frac{P(e^x)}{Q(e^x)}$ is non-constant, where P, Q are polynomials over $\mathbb{Q}$ with $Q(x) \neq 0$ for all $x > 0$.

In the sequel we will make heavy use of a series of constants $\kappa_1, \kappa_2, \cdots \in \mathbb{N}$, which are defined as follows. Enumerate positive valued rationals in any way: $l_1, l_2 \ldots$ and recall that all weights of our GNN $A_1, C_1, b_1, \ldots, A_L, C_L, b_L$ are positive (non-negative in the case of $b_1, \ldots, b_L$) rationals. For a given $k \geq 1$ the constant κ_k is the least common denominator of $l_1, \ldots, l_k$ as well as $A_1, C_1, b_1, \ldots, A_L, C_L, b_L$. When k is clear from context, we simply write $\kappa = \kappa_k$.

With the above at hand we take any $k \geq 1$ and $r \geq 1$ and define the following sets inductively on $i < L$:

- $\widehat{Z}_{-1,k,r} = \{e^{\frac{1}{\kappa}}\}$,
- $Z_{-1,k,r} = \{\sigma(l_1), \ldots, \sigma(l_k)\} = \{R((e^{\frac{1}{\kappa}})^{l_j\kappa}) \mid 1 \leq j \leq k\}$,
- $\widehat{Z}_{i+1,k,r} = \{e^{\frac{z}{\kappa}} \mid z \in Z_{i,k,r}\}$, and
- $Z_{i+1,k,r} = \{R(z_1 \cdot \ldots \cdot z_r) \mid z_1, \ldots, z_r \in \widehat{Z}_{i+1,k,r} \cup \{1, e^{\frac{1}{\kappa}}\}$ not all in $\{1, e^{\frac{1}{\kappa}}\}\}$.

The definitions of $\widehat{Z}_{-1,k,r}$ and $Z_{-1,k,r}$ are technical and serve as an intermediary step for computing the initialisation $\lambda^{(0)}$. To aid intuition, we invite the reader to think of $Z_{0,k,r}$ as a k-sized sample of $\lambda^{(0)}$'s initialisation values; it being understood that r is "large enough". The set $Z_{1,k,r}$ is then again a sample of values that $\mathcal{N}$ can compute by, in a sense, aggregating and combining values of $Z_{0,k,r}$. Subsequent sets $Z_{i,k,r}$ then behave similarly. The intuition behind κ is that raising $e^{\frac{1}{\kappa}}$ to appropriate powers will ultimately recover (exponentiations of) $l_1, \ldots, l_k$ as well as $A_1, C_1, b_1, \ldots, A_L, C_L, b_L$. We show how this is achieved by writing $Z_{i,k} = \bigcup_{r \geq 1} Z_{i,k,r}$ and:

Claim 13. *For all featured graphs G, all vertices $v \in V$ and all $0 \leq i \leq L$ there is some $k' \geq 1$ such that for all $k \geq k'$ we have $\lambda^{(i)}(v) \in Z_{i,k}$.*

Proof. Take any graph G and any vertex $v \in V$. Recall that $\lambda(v)$ is a vector of non-negative rationals, say, $m_1, \ldots, m_N$, it being understood that $N = |\lambda(v)|$. Since each entry is non-negative, we may find $l_{k_1}, \ldots, l_{k_N}$ in our enumeration of the positive rational numbers, such that $l_{k_t} = m_t + 1$. Now, let $k'_v = \max(k_1, \ldots, k_N)$. Clearly, $\sigma(l_{k_t}) \in Z_{-1,k'_v}$ thus resulting in $\sigma(m_t + 1) \in Z_{-1,k'_v}$ for all $1 \leq t \leq N$. Now, taking $k' = \max\{k'_v \mid v \in V\}$ we have, for each $k \geq k'$, that every entry m in any feature vector of vertices in G satisfies $\sigma(m + 1) \in Z_{-1,k}$.

Taking any k as above and writing $\kappa = \kappa_k$ we proceed by induction on $0 \leq i < L$. For the base case take any v and let $m_1, \ldots, m_N$ be the entries of $\lambda(v)$ in order. As argued above, $\sigma(m_1 + 1), \ldots, \sigma(m_N + 1) \in Z_{-1,k}$ and thus $e^{\frac{\sigma(m_1+1)}{\kappa}}, \ldots, e^{\frac{\sigma(m_N+1)}{\kappa}} \in \widehat{Z}_{0,k}$. But then, again by definition, $R(\prod_{j=1}^{n} e^{2^j \kappa \cdot \frac{\sigma(m_j+1)}{\kappa}}) = \sigma((2^1, 2^2, \ldots, 2^N) \times \sigma(\lambda(v) + \mathbf{1}_N)) = \lambda^{(0)}(v)$ is in $Z_{0,k,r}$ for r large enough, and thus also in $Z_{0,k}$.

For the inductive step assume that $\lambda^{(i)}(v) \in Z_{i,k}$ for all $v \in V$. Then, aggregation and combination in $\mathcal{N}$ amount to $\sigma\big(C_{i+1} z_0 + A_{i+1}(z_1 + \cdots + z_n) + b_{i+1}\big)$, where $z_0, \ldots, z_n \in Z_{i,k}$. By the definition of σ we have that the above is $R(e^{C_{i+1} z_0} \cdot \prod_{j=1}^{n} e^{A_{i+1} z_j} \cdot e^{b_{i+1}})$. Noting that $C_{i+1} = \frac{C'_{i+1}}{\kappa}$, $A_{i+1} = \frac{A'_{i+1}}{\kappa}$, and $b_{i+1} = \frac{b'_{i+1}}{\kappa}$, for some $C'_{i+1}, A'_{i+1}, b'_{i+1} \in \mathbb{N}$ we have that $e^{C_{i+1} z_0} = (e^{\frac{z_0}{\kappa}})^{C'_{i+1}}$ with $e^{\frac{z_0}{\kappa}}$ being a member of $\widehat{Z}_{i+1,k}$. Similar observations hold for $e^{A_{i+1} z_j}$ with $1 \leq j \leq n$. Lastly, recalling that multiples of $e^{\frac{1}{\kappa}}$ are allowed to appear when forming $Z_{i+1,k}$, we have $e^{b_{i+1}} = (e^{\frac{1}{\kappa}})^{b'_{i+1}}$. Then, clearly, $R(e^{C_{i+1} z_0} \cdot \prod_{j=1}^{n} (e^{A_{i+1} z_j}) \cdot e^{b_{i+1}}) = R((e^{\frac{z_0}{\kappa}})^{C'_{i+1}} \cdot \prod_{j=1}^{n} (e^{\frac{z_j}{\kappa}})^{A'_{i+1}} \cdot (e^{\frac{1}{\kappa}})^{b'_{i+1}})$ which is then in $Z_{i+1,k,r}$ for large enough r, and thus also in $Z_{i+1,k}$. □

Our strategy again consists of establishing linear independence of $Z_{i,k}$ for all $0 \leq i \leq L$ and $k \geq 1$ in order to conclude that $\mathcal{N}$ matches the L-round 1-WL test. We fully prove this claim at the end of this section. The next few claims are a setup for the desired conclusion.

Claim 14. *There is some $t \in \mathbb{Q}$ and non-zero $r \in \mathbb{Z}$ such that $\alpha = \lim_{x \to 0^+} \frac{R(x) - t}{x^r}$ is non-zero.*

Proof. Since $P(x)$ and $Q(x)$ are non-zero polynomials we may find $p, q \in \mathbb{N}$ such that $P(x) = x^p P_0(x)$ and $Q(x) = x^q Q_0(x)$ for non-zero polynomials $P_0(x), Q_0(x)$ both having a non-zero constant term (and thus $P_0(0) \neq 0 \neq Q_0(0)$). Suppose first that $p \neq q$. Then, by taking $t = 0$ and $r = p - q$ we have

$$\alpha = \lim_{x \to 0^+} \frac{\frac{P(x)}{Q(x)} - t}{x^r} = \lim_{x \to 0^+} \frac{x^{p-q} \cdot \frac{P_0(x)}{Q_0(x)}}{x^{p-q}} = \lim_{x \to 0^+} \frac{P_0(x)}{Q_0(x)} \neq 0.$$

Now suppose that $p = q$. Then, taking P_0 and Q_0 as above we again have that $\frac{P_0(0)}{Q_0(0)} \neq 0$. Now, let us take a look at $S(x) = P_0(x)Q_0(0) - P_0(0)Q_0(x)$. Notice that $S(x)$ is not the zero polynomial as we would then have $\frac{P_0(x)}{Q_0(x)} = \frac{P_0(0)}{Q_0(0)}$ for all $x > 0$. Inspecting $S(x)$ we see that the constant terms of polynomials $P_0(x)Q_0(0)$ and $P_0(0)Q_0(x)$ cancel each other out. Thus we may assume that $S(x)$ takes the normal form $a_0 + a_1 x^1 + \cdots + a_n x^n$ with $a_0 = 0$. Now, let r be the smallest index such that $a_r \neq 0$. Clearly, x^r is then the smallest power of x present in $S(x)$. Thus, $S(x) = x^r S_0(x)$, where $S_0(x)$ is some non-zero polynomial. Note that $S_0(0) \neq 0$ as that would contradict $a_r \neq 0$. We now observe the following by setting $t = \frac{P_0(0)}{Q_0(0)}$:

$$\alpha = \lim_{x \to 0^+} \frac{\frac{P(x)}{Q(x)} - t}{x^r} = \lim_{x \to 0^+} \frac{\frac{x^p P_0(x)}{x^q Q_0(x)} - \frac{P_0(0)}{Q_0(0)}}{x^r} = \lim_{x \to 0^+} \frac{\frac{P_0(x)Q_0(0) - P_0(0)Q_0(x)}{Q_0(x)Q_0(0)}}{x^r} =$$

$$\lim_{x \to 0^+} \frac{\frac{x^r S_0(x)}{Q_0(x)Q_0(0)}}{x^r} = \lim_{x \to 0^+} \frac{S_0(x)}{Q_0(x)Q_0(0)} = \frac{S_0(0)}{Q_0(0)Q_0(0)} \neq 0$$

Thus concluding the current case and proof. □

A series of functions $f_1(x), \ldots, f_n(x)$ is said to be *linearly dependent* over $\mathbb{Q}$ if there are some $c_1, \ldots, c_n \in \mathbb{Q}$ not all zero such that $c_1 f_1(x) + \cdots + c_n f_n(x) = 0$ for all x over the common

domain of $f_1(x), \ldots, f_n(x)$. If no such $c_1, \ldots, c_n$ exist, then the functions are linearly independent.

Claim 15. *Let $B_1, \ldots, B_n \in \mathbb{N}$ be distinct and positive. Then, the functions $1, R(x^{B_1}), \ldots, R(x^{B_n})$ are linearly independent over $\mathbb{Q}$.*

Proof. Suppose for contradiction that $1, \frac{P(x^{B_1})}{Q(x^{B_1})}, \ldots, \frac{P(x^{B_n})}{Q(x^{B_n})}$ are linearly dependent over $\mathbb{Q}$. One may then find $c_1, \ldots, c_n, d \in \mathbb{Q}$ not all zero such that, for all positive $x \in \mathbb{Q}$,

$$\sum_{j=1}^{n} c_j \cdot \frac{P(x^{B_j})}{Q(x^{B_j})} + d = 0.$$

Now, fix t, r, α as the values promised by Claim 14 and write $T(x) = \frac{R(x)-t}{x^r}$. For all $1 \leq j \leq n$ we then have that $R(x^{B_j}) = x^{rB_j}T(x^{B_j}) + t$. Thus, the linear dependence identity can be written as:

$$\sum_{j=1}^{n} c_j \Big(x^{rB_j}T(x^{B_j}) + t \Big) + d = \sum_{j=1}^{n} c_j x^{rB_j}T(x^{B_j}) + D = 0,$$

where $D = d + t\sum_{j=1}^{n} c_j$.

Discarding terms with $c_j = 0$ suppose, first, that $D = 0$. Now, let j be such that $rB_j = \min\{rB_\ell \mid 1 \leq \ell \leq n, c_\ell \neq 0\}$. Without loss of generality, we assume that $j = 1$. Dividing the above by x^{rB_j} and taking the limit of the left-hand side we obtain the following:

$$\lim_{x\to 0^+} c_1 T(x^{B_1}) + \sum_{j=2}^{n} c_j x^{rB_j - rB_1} T(x^{B_j}) = c_1\alpha + \sum_{j=2}^{n} c_j \cdot 0 \cdot \alpha = c_1\alpha,$$

as for all $j \geq 2$, we have that $rB_j - rB_1 > 0$. Notice that since the original expression is zero, we have that the limit above is also zero. But neither $c_1 = 0$, nor $\alpha = 0$ thus resulting in a contradiction.

Now, suppose that $D \neq 0$. But then, if all rB_ℓ with $c_\ell \neq 0$ are positive we have

$$\lim_{x\to 0^+} \sum_{j=1}^{n} c_j x^{rB_j} T(x^{B_j}) + D = \sum_{j=1}^{n} c_j \cdot 0 \cdot \alpha + D = D.$$

But $D \neq 0$ thus resulting in a contradiction. Otherwise, if some rB_ℓ with $c_\ell \neq 0$ is negative, again take $rB_j = \min\{rB_\ell \mid 1 \leq \ell \leq n, c_\ell \neq 0\}$. Assuming w.l.o.g. that $j = 1$ we again have that

$$\lim_{x\to 0^+} c_1 T(x^{B_1}) + \sum_{j=2}^{n} c_j x^{rB_j - rB_1} T(x^{B_j}) + x^{-rB_1} D = c_1\alpha + \sum_{j=1}^{n} c_j \cdot 0 \cdot \alpha + 0D = c_1\alpha.$$

But again $c_1 \neq 0$ and $\alpha \neq 0$ thus contradicting our initial assumption and concluding the proof. □

Finally, writing $\widehat{Z}_{i,k} = \bigcup_{r\geq 1} \widehat{Z}_{i,k,r}$ we show:

Claim 16. *Suppose that each finite subset of $\bigcup_{j=-1}^{i} \widehat{Z}_{j,k}$ is algebraically independent over $\mathbb{Q}$. Then each finite subset of $\{1\} \cup \bigcup_{j=-1}^{i} Z_{j,k}$ is linearly independent over $\mathbb{Q}$.*

Proof. Suppose for contradiction that $\{1\} \cup \bigcup_{j=-1}^{i} Z_{j,k}$ is not linearly independent over $\mathbb{Q}$. Then, there are some elements $z_1, \ldots z_m \in \bigcup_{j=-1}^{i} Z_{j,k}$ and $c_1, \ldots, c_m, d \in \mathbb{Q}$ not all zero such that $\sum_{j=1}^{m} c_j z_j + d = 0$. Recall that $z_j = R(s_1^{a_{j,1}} \cdot \ldots \cdot s_n^{a_{j,n}})$ with $a_{j,1}, \ldots, a_{j,n} \in \mathbb{N}$ not all zero and $s_1, \ldots, s_n \in \bigcup_{j=-1}^{i} \widehat{Z}_{j,k}$. We can rewrite the relation above as

$$\sum_{j=1}^{m} c_j \cdot \frac{P(s_1^{a_{j,1}} \cdot \ldots \cdot s_n^{a_{j,n}})}{Q(s_1^{a_{j,1}} \cdot \ldots \cdot s_n^{a_{j,n}})} + d = 0.$$

Since Q is non zero on values above 0 we multiply by the denominators and obtain:

$$\sum_{j=1}^{m} c_j \cdot P(s_1^{a_{j,1}} \cdot \ldots \cdot s_n^{a_{j,n}}) \cdot \prod_{\ell=1}^{\ell\neq j} Q(s_1^{a_{\ell,1}} \cdot \ldots \cdot s_n^{a_{\ell,n}}) + d\prod_{j=1}^{m} Q(s_1^{a_{j,1}} \cdot \ldots \cdot s_n^{a_{j,n}}) = 0.$$

Clearly, when symbols $s_1, \ldots, s_n$ are replaced by variables, the expression above is a polynomial. Assuming it is non-zero, we have that $\bigcup_{j=-1}^{i} \widehat{Z}_{j,k}$ is not algebraically independent over $\mathbb{Q}$ thus contradicting our initial assumption. Thus the only thing left to show is that the above is non-zero.

Clearly, the polynomial mentioned above is non-zero if the following expression is non-zero:

$$\sum_{j=1}^{m} c_j \cdot \frac{P(y_1^{a_{j,1}} \cdot \ldots \cdot y_n^{a_{j,n}})}{Q(y_1^{a_{j,1}} \cdot \ldots \cdot y_n^{a_{j,n}})} + d.$$

Concentrating on the above let $\beta = \max\{a_{j,\ell} \mid 1 \leq j \leq m, 1 \leq \ell, \leq n\} + 1$ and substitute each y_ℓ with x^{β^ℓ}. Now, write $B_\ell = \beta^1 a_{\ell,1} + \cdots + \beta^n a_{\ell,n}$. Here B_ℓ is the β-ary encoding of the tuple $(a_{\ell,1}, \ldots, a_{\ell,n})$. Notice that each tuple is distinct as $z \in Z_{j,k}$ is formed by applying R to a non empty product of (possibly repeated) elements in $\widehat{Z}_{j,k}$ multiplied by some (possibly empty) product of (possibly repeated) elements in $\{e^{\frac{1}{\kappa}}\} = \widehat{Z}_{-1,k}$. Since each tuple is distinct, we have that all B_ℓ are as well. What is obtained is the following expression in a single variable:

$$\sum_{j=1}^{m} c_j \cdot \frac{P(x^{B_j})}{Q(x^{B_j})} + d.$$

But then, by Claim 15, the functions $1, \frac{P(x^{B_1})}{Q(x^{B_1})}, \ldots, \frac{P(x^{B_m})}{Q(x^{B_m})}$ are linearly independent. Thus, there is some value in $\mathbb{Q}$ under which the expression above is defined and non-zero as required. □

Once again, the only missing part of our proof is that of algebraic independence of (finite subsets of) $\widehat{Z}_{i,k}$ over $\mathbb{Q}$. This is again achieved by Schanuel's conjecture.

Lemma 17. *Under Schanuel's conjecture, finite subsets of $\bigcup_{j=-1}^{L} \widehat{Z}_{j,k}$ are algebraically independent over $\mathbb{Q}$.*

Proof. The proof is in similar spirit to the one in the main body of the work. Let us return to the setting of sets $\widehat{Z}_{i,k,r}$ and proceed by induction on $-1 \leq i \leq L$ maintaining

1. $\bigcup_{j=-1}^{i} \widehat{Z}_{j,k,r}$ is algebraically independent over $\mathbb{Q}$,
2. the sets $\widehat{Z}_{-1,k,r}, \ldots, \widehat{Z}_{i,k,r}$ are pairwise disjoint; and so are $\{1\}, Z_{-1,k,r}, \ldots, Z_{i,k,r}$,
3. each entry of $Z_{-1,k,r}, \ldots, Z_{i,k,r}$ is positive.

Again the required result will be achieved by proving the above for each $r \geq 1$, as $\widehat{Z}_{i,k}$ is the union of $\widehat{Z}_{i,k,r}$ for all $r \geq 1$ and $\widehat{Z}_{i,k,r} \subseteq \widehat{Z}_{i,k,r+1}$.

For the base case note that $\bigcup_{j=-1}^{-1} \widehat{Z}_{j,k,r} = \{e^{\frac{1}{\kappa}}\}$, where κ is a non-zero integer value that is the least common denominator of the positive rationals $l_1, \ldots, l_k$ as well as the weights of $\mathcal{N}$. Hence, condition 1 holds trivially as, by the Lindemann–Weierstrass Theorem, $e^{\frac{1}{\kappa}}$ is algebraically independent over $\mathbb{Q}$. Condition 2 is immediate in regards to $\widehat{Z}_{-1,k,r}$. Suppose that it does not hold in regards to $\{1\}, Z_{-1,k,r}$. Then, $1 \in Z_{-1,k,r}$ meaning that there is some positive rational l_j for which $\sigma(l_j) = 1$. But then $P(e^{l_j}) = Q(e^{l_j})$ and thus $P(e^{l_j}) - Q(e^{l_j}) = 0$. But since e^{l_j} is transcendental we have that $P(x) - Q(x) = 0$ for all x. Thus, $\frac{P(x)}{Q(x)} = 1$ for all x contradicting our requirement that the function is strictly increasing for $x > 1$. Condition 3 follows from the fact that $l_1, \ldots, l_k$ are positive rationals thus making $e^{\frac{z}{\kappa}} > 1$; it being understood that z is any sum (with possible repeats) of $l_1, \ldots, l_k$. By definition, $R(x)$ is positive on $x > 1$, thus $R(e^z)$ is positive as required.

Now, suppose the above conditions hold for $-1 \leq i < L$. We show that the same is true for $i + 1$ in place of i. By condition 1 of our inductive hypothesis, $\bigcup_{j=-1}^{i} \widehat{Z}_{j,k,r}$ is algebraically independent over $\mathbb{Q}$. Hence, by invoking Claim 16 on appropriate subsets, the set $\{1\} \cup \bigcup_{j=-1}^{i} Z_{j,k,r}$ is linearly independent over $\mathbb{Q}$, and thus so is $S = \{\frac{1}{\kappa}\} \cup \bigcup_{j=-1}^{i} \{\frac{z}{\kappa} \mid z \in Z_{j,k,r}\}$ as κ is a positive integer. With that, we may apply Schanuel's conjecture on S and, in doing so, obtain $F = \mathbb{Q}(S \cup \bigcup_{j=-1}^{i+1} \widehat{Z}_{j,k,r})$

such that $\mathsf{trdeg}_{\mathbb{Q}}(F) \geq |S|$. When turning to analyse the size of S. Notice, first, that $|Z_{i,k,r}| = |\widehat{Z}_{i+1,k,r}|$. Now, by condition 2 of our induction hypothesis, $\{1\}, Z_{-1,k,r}, \ldots, Z_{i,k,r}$ are all pairwise disjoint. Thus,

$$|S| = |\{1\} \cup \bigcup_{j=-1}^{i} Z_{j,k,r}| = 1 + \sum_{j=-1}^{i} |Z_{j,k,r}| = |\widehat{Z}_{-1,k,r}| + \sum_{j=0}^{i+1} |\widehat{Z}_{j,k,r}| = \sum_{j=-1}^{i+1} |\widehat{Z}_{j,k,r}|.$$

Now, write $E = \mathbb{Q}(\bigcup_{j=-1}^{i+1} \widehat{Z}_{j,k,r})$. Notice that $S \subseteq E$ as elements of S are either rational, already in the set $\bigcup_{j=-1}^{i+1} \widehat{Z}_{j,k,r}$, or take the form $\frac{1}{\kappa} \cdot R(z) = \frac{1}{\kappa} \cdot \frac{P(z)}{Q(z)}$, where z is a monomial over $\bigcup_{j=-1}^{i} \widehat{Z}_{j,k,r}$ and κ is a positive rational. By condition 3 of our induction hypothesis, $z > 1$ as it is the result of exponentiation by a positive power. Thus, by definition of $R(x)$, we have $Q(z) \neq 0$. Since fields are closed under multiplicative inverses, we obtain $E = F$ and thus $\mathsf{trdeg}_{\mathbb{Q}}(E) \geq |S| = \sum_{j=-1}^{i+1} |\widehat{Z}_{j,k,r}|$. The above combined with Lemma 1 gives us that $\bigcup_{j=-1}^{i+1} \widehat{Z}_{j,k,r}$ is algebraically independent over $\mathbb{Q}$. This is then exactly condition 1.

Condition 2 for $\widehat{Z}_{-1,k,r}, \ldots, \widehat{Z}_{i+1,k,r}$ is achieved noting that, by I.H., $\{1\}, Z_{-1,k,r}, \ldots, Z_{i,k,r}$ are pairwise disjoint sets and exponentiation (of $e^{\frac{1}{\kappa}}$) is injective. Now, suppose that $Z_{i+1,k,r} \cap Z_{j,k,r} \neq \emptyset$ for some $0 \leq j \leq i$. Let the common element be z. Clearly, z equals $R(s_1) = R(s_2)$, where s_1 is a monomial over $\widehat{Z}_{i+1,k,r}$ multiplied with some non-negative power of $e^{\frac{1}{\kappa}}$ and s_2 a monomial over $\widehat{Z}_{j,k,r}$ multiplied by some non-negative power of $e^{\frac{1}{\kappa}}$. But, by our argument above, $\widehat{Z}_{i+1,k,r} \cap \widehat{Z}_{j,k,r} = \emptyset$. Additionally, by condition 1 and 2 (already established above for $\widehat{Z}_{i+1,k,r}$), the sets $\widehat{Z}_{i+1,k,r}, \widehat{Z}_{j,k,r}, \{e^{\frac{1}{\kappa}}\}$ are distinct and their union is algebraically independent over $\mathbb{Q}$. Hence, $s_1 \neq s_2$. Notice that $s_1, s_2 > 1$ as a consequence of condition 3. Indeed, by I.H., all entries of $Z_{-1,k,r}, \ldots, Z_{i,k,r}$ are positive. By definition, s_1 and s_2 are then e raised to a positive power. Then, since $R(x)$ is strictly increasing on $x > 1$, we have that $R(s_1) \neq R(s_2)$ thus contradicting our initial assumption. The case where $Z_{i+1,k,r} \cap Z_{-1,k,r} \neq \emptyset$ is similar as $\widehat{Z}_{i+1,k,r}, \{e^{\frac{1}{\kappa}}\}$ are distinct and, when taken together, algebraically independent over $\mathbb{Q}$. Finally, $Z_{i+1,k,r} \cap \{1\} = \emptyset$ follows from the fact that 1 is algebraic over $\mathbb{Q}$ whilst elements of $Z_{i+1,k,r}$ are not.

Condition 3 holds following the same observations as for the base case. Indeed, by I.H. entries of $Z_{i,k,r}$ are all positive. Thus, $e^{\frac{z_1}{\kappa}} \cdot \ldots \cdot e^{\frac{z_n}{\kappa}} > 1$, for any choice of $z_1, \ldots, z_n \in Z_{i,k,r} \cup \{1\}$. Hence, again by definition, $R(e^{\frac{z_1}{\kappa}} \cdot \ldots \cdot e^{\frac{z_n}{\kappa}}) > 0$. □

Claim 18. *Under Schanuel's conjecture, the vertex embeddings produced by $\mathcal{N}$ partition the class of vertices into the same equivalence classes as the L-round 1-WL test. Additionally, the graph embeddings produced by $\mathcal{N}$ partition the class of graphs into the same equivalence classes as the L-round 1-WL test.*

Proof. For the first statement we show that each of the functions $\widetilde{\lambda}^{(0)}, \ldots, \widetilde{\lambda}^{(L)}$ are effectively injective on their respective domains and invoke Lemma 4. For the second statement, we show that each finite subset of vertex embeddings produced by $\mathcal{N}$ is linearly independent over $\mathbb{Q}$.

Let us first take a look at $\widetilde{\lambda}^{(0)}$. Take any featured graph $G = (V, E, \lambda)$ and any $v, u \in V$. Suppose for contradiction that $\lambda^{(0)}(v) = \lambda^{(0)}(u)$, but the initial features are such that $\lambda(v) \neq \lambda(u)$. Inspecting the initialisation function we see that its first layer gives vectors $\alpha = \sigma(\lambda(v) + \mathbf{1}_N)$ and $\beta = \sigma(\lambda(u) + \mathbf{1}_N)$. Since $\lambda(v) + \mathbf{1}_N$ and $\lambda(u) + \mathbf{1}_N$ contain only positive entries we have that $\alpha \neq \beta$ as $R(y)$ is strictly increasing on $y > 1$, thus ensuring that $\sigma(x)$ strictly increases on $x > 0$. It is in the second layer of the initialisation that we must have $\sigma\big((2^1, \ldots, 2^N) \times \alpha\big) = \sigma\big((2^1, \ldots, 2^N) \times \beta\big)$. Suppose first that $(2^1, \ldots, 2^N) \times \alpha = (2^1, \ldots, 2^N) \times \beta$. By Claim 13, the entries of α and β belong to $Z_{-1,k}$ for k large enough. By Lemma 17 and Claim 16, the values of $Z_{-1,k} = \{s_1, \ldots, s_k\}$ are linearly independent over $\mathbb{Q}$. Thus, it must be the case that $\sum_{1 \leq j \leq N}^{\alpha[j]=s_\ell} 2^j = \sum_{1 \leq j \leq N}^{\beta[j]=s_\ell} 2^j$ for all $1 \leq \ell \leq k$; it being understood that $\alpha[j]$ is the j-th entry of α (and similarly for β). But then, for a given j, the respective sum is a binary encoding of positions where s_j resides in α and β. Thus, $\alpha[j] = \beta[j]$ for all $1 \leq j \leq N$ contradicting $\alpha \neq \beta$. Lastly, writing $\alpha' = (2^1, \ldots, 2^N) \times \alpha$ and $\beta' = (2^1, \ldots, 2^N) \times \beta$ we must have that $\sigma(\alpha') = \sigma(\beta')$. But $\alpha', \beta' > 0$ thus resulting in $e^{\alpha}, e^{\beta'} > 1$ and $\sigma(\alpha') \neq \sigma(\beta')$, a contradiction to our initial assumption arising from the fact $R(x)$ is strictly increasing for $x > 1$.

For $\lambda^{(i)}$ with $1 \leq i \leq L$ we proceed as follows. Suppose for contradiction that $\lambda^{(i)}(v) = \lambda^{(i)}(u)$, but

$$\big(\lambda^{(i-1)}(v), \{\{\lambda^{(i-1)}(v') \mid v' \in N(v)\}\}\big) \neq \big(\lambda^{(i-1)}(u), \{\{\lambda^{(i-1)}(u') \mid u' \in N(u)\}\}\big).$$

But then,

$$\sigma(C_i\lambda^{(i-1)}(v) + A_i \sum_{v' \in N(v)} \lambda^{(i-1)}(v') + b_i) = \sigma(C_i\lambda^{(i-1)}(u) + A_i \sum_{u' \in N(u)} \lambda^{(i-1)}(u') + b_i).$$

By Claim 13 the embeddings given by λ^{i-1} are all in $Z_{i-1,k}$ for k large enough. It is easy to verify that the sums inside the activation function are positive. Since $\sigma(x)$ is strictly increasing when $x > 0$ we have that, in fact,

$$C_i\lambda^{(i-1)}(v) + A_i \sum_{v' \in N(v)} \lambda^{(i-1)}(v') + b_i = C_i\lambda^{(i-1)}(u) + A_i \sum_{u' \in N(u)} \lambda^{(i-1)}(u') + b_i.$$

Thus, supposing the relevant values of $Z_{i-1,k}$ are $s_1, \ldots, s_n$, we obtain the following relation:

$$C_i s_r + A_i \sum_{\ell=1}^{n} \alpha_\ell s_\ell = C_i s_t + A_i \sum_{\ell=1}^{n} \beta_\ell s_\ell,$$

where all α_ℓ, β_ℓ are in $\mathbb{N}$ and represent the number of neighbours (of v and u respectively) with the embedding s_ℓ. It is to be understood that $s_r = \lambda^{(i-1)}(v)$ and $s_t = \lambda^{(i-1)}(u)$. Supposing for the moment $s_r = s_t$ we have that $\sum_{\ell=1}^{n} \alpha_\ell s_\ell = \sum_{\ell=1}^{n} \beta_\ell s_\ell$. Since values of $Z_{i-1,k}$ are linearly independent over $\mathbb{Q}$ (Lemma 17 with Claim 16) we have that $\alpha_\ell = \beta_\ell$ for all $1 \leq \ell \leq n$. Thus, there is a bijection $f : N(v) \to N(u)$ such that $\lambda^{(i-1)}(v') = \lambda^{(i-1)}(f(v'))$. Thus, $\big(\lambda^{(i-1)}(v), \{\{\lambda^{(i-1)}(v') \mid v' \in N(v)\}\}\big) = \big(\lambda^{(i-1)}(u), \{\{\lambda^{(i-1)}(f(v')) \mid v' \in N(v)\}\}\big) = \big(\lambda^{(i-1)}(u), \{\{\lambda^{(i-1)}(u') \mid u' \in N(u)\}\}\big)$ thus contradicting our initial assumption.

Now suppose that $s_r \neq s_t$. If $\alpha_\ell = \beta_\ell$ for all $1 \leq \ell \leq n$, then $C_i s_r = C_i s_t$. Since $C_i > 0$ this results in a contradiction. Now, if $\alpha_\ell \neq \beta_\ell$ for some choice of ℓ, we obtain the following by grouping dependent terms together:

$$C_i s_r + A_i \alpha_r s_r = A_i \beta_r s_r \quad \text{and} \quad A_i \alpha_t s_t = C_i s_t + A_i \beta_t s_t.$$

Focusing on the first equation, recall that $\frac{C_i}{A_i} \notin \mathbb{N}$. In fact, since $A_i, C_i > 0$, we have that $\frac{C_i}{A_i} \notin \mathbb{Z}$. But then $\beta_r - \alpha_r = \frac{C_i}{A_i}$. Since $\alpha_r, \beta_r \in \mathbb{N}$ we have that $\frac{C_i}{A_i} \in \mathbb{Z}$, a contradiction.

For the second statement of the claim we proceed as follows. Take any finite subset $\{s_1, \ldots, s_n\}$ such that $s_j = \lambda^{(L)}(v_j)$ for some pointed graph (G_j, v_j). Now, let G be the disjoint union of $G_1, \ldots, G_n$. Claim 13 then establishes $s_1, \ldots, s_n \in Z_{L,k}$ for k large enough. Lemma 17 and Claim 16 then establish that these values are linearly independent over $\mathbb{Q}$. □

B Proof of Proposition 10

Proposition 10. *Suppose (G, v) and (H, u) are two featured pointed graphs with virtual nodes. Then, for all L-layered GNNs $\mathcal{N}$, $\mu^{(L)}(v) = \mu^{(L)}(u)$ implies $\lambda^{(L)}(v) = \lambda^{(L)}(u)$ and $\mu_G = \mu_H$ implies $\lambda_G = \lambda_H$.*

Proof. The proof is comprised of two inductive arguments. First, we show that $\mu^{(i+1)}(v) = \mu^{(i+1)}(u)$ implies $\mu^{(i)}(v) = \mu^{(i)}(u)$. For the base case suppose that $\mu^{(1)}(v) = \mu^{(1)}(u)$, but $\mu^{(0)}(v) \neq \mu^{(0)}(u)$. If v and u are ordinary vertices, then $\mu^{(1)}(v) = \mathsf{hash}(\mu^{(0)}(v), \ldots) \neq \mathsf{hash}(\mu^{(0)}(u), \ldots) = \mu^{(1)}(u)$, a contradiction. If v and u are virtual then, by definition, $\mu^{(0)}(v) = \mathsf{hash}(\mathsf{virtual}) = \mu^{(0)}(u)$. The case where one node is virtual and the other is ordinary is immediate as then $\mu^{(j)}(v) \neq \mu^{(j)}(u)$ for any $j \geq 0$.

Let us assume the required condition holds for $i > 0$. We extend this to the $i + 1$ setting as follows. Suppose $\mu^{(i+1)}(v) = \mu^{(i+1)}(u)$. If v and u are ordinary vertices, then $\mu^{(i)}(v) \neq \mu^{(i)}(u)$ would result in a similar contradiction as in the base case. On the other hand suppose v and u are virtual and that $\mu^{(i+1)}(v) = \mu^{(i+1)}(u)$, but $\mu^{(i)}(v) \neq \mu^{(i)}(u)$. Writing $S = \{\{\mu^{(i-1)}(v') \mid v' \in N(v)\}\}$

and $Z = \{\{\mu^{(i-1)}(u') \mid u' \in N(u)\}\}$ we observe that $S \neq Z$. But then, $S^+ = Z^+$, where $S^+ = \{\{\mu^{(i)}(v') \mid v' \in N(v)\}\}$ and $Z^+ = \{\{\mu^{(i)}(u') \mid u' \in N(u)\}\}$. Now, let f be a bijection between $N(v)$ and $N(u)$ such that $\mu^{(i)}(v') = \mu^{(i)}(f(v'))$ for all $v' \in N(v)$. But then, by I.H., the same bijection gives $\mu^{(i-1)}(v') = \mu^{(i-1)}(f(v'))$ and establishes $S = Z$, a contradiction. The case where one node is virtual and the other is ordinary is immediate.

We now prove that $\mu^{(L)}(v) = \mu^{(L)}(u)$ implies $\lambda^{(L)}(v) = \lambda^{(L)}(u)$ by induction on $0 \leq i \leq L$. The base case $i = 0$ is immediate as $\lambda(v) = \lambda(u)$ if and only if $\mu^{(0)}(v) = \mu^{(0)}(u)$. Since $\lambda^{(0)}$ is an FNN for ordinary vertices and produced the same vector for virtual ones, we have that $\mu^{(0)}(v) = \mu^{(0)}(u)$ implies $\lambda^{(0)}(v) = \lambda^{(0)}(u)$.

For the inductive step assume that our hypothesis holds for $0 \leq i < L$. We show that it holds for $i+1$ as well. For this purpose suppose $\mu^{(i+1)}(v) = \mu^{(i+1)}(u)$. Then, as argued above, $\mu^{(i)}(v) = \mu^{(i)}(u)$ and $\{\{\mu^{(i)}(v') \mid v' \in N(v)\}\} = \{\{\mu^{(i)}(u') \mid u' \in N(u)\}\}$. But then, by our inductive hypothesis, $\lambda^{(i)}(v) = \lambda^{(i)}(u)$ as well as $\{\{\lambda^{(i)}(v') \mid v' \in N(v)\}\} = \{\{\lambda^{(i)}(u') \mid u' \in N(u)\}\}$. Recall that, before activation, $\lambda^{(i+1)}(v)$ is a linear combination of $\sum\{\{\lambda^{(i)}(v') \mid v' \in N(v)\}\}$ and $\lambda^{(i)}(v)$; and, in fact, so is $\lambda^{(i+1)}(u)$. It is then easy to see that $\lambda^{(i+1)}(v) = \lambda^{(i+1)}(u)$ as required.

To see that $\mu_G = \mu_H$ implies $\lambda_G = \lambda_H$ one needs only note that $\{\{\mu^{(L)}(v) \mid v \in V\}\} = \{\{\mu^{(L)}(u) \mid u \in U\}\}$; it being understood that V is the set of vertices of G and U is the set of vertices of H. Then, $\sum\{\{\lambda^{(L)}(v) \mid v \in V\}\} = \sum\{\{\lambda^{(L)}(u) \mid u \in U\}\}$ as required. □

C Full proof of Theorem 11

Theorem 11. *Let $\mathcal{N}$ be an L-layer GNN that uses $\sigma \in \{\exp, \sinh, \cosh\}$ as activation and produces 1-dimensional vertex embeddings in its final layer. Further assume that $\lambda_{G_1} = \lambda_{G_2}$ if and only if $\mu_{G_1} = \mu_{G_2}$ for all featured graphs with virtual nodes G_1, G_2. Then for every featured graph $G = (V, E, \lambda)$ the values $\{\exp(A'_L \lambda^{(L-1)}(v)) \mid v \in V\}$ are algebraically independent over $\mathbb{Q}$.*

In this section we only fill in the parts of the above theorem. Namely we show that G_Q and G_R as constructed in the theorem are distinguishable by L-round 1-WL, that $\mathcal{N}$ is such that $\lambda_{G_Q} = \lambda_{G_R}$, and show that the above theorem holds with $\sigma \in \{\sinh, \cosh\}$.

We proceed by showing the first claim. For this purpose recall the pointed graphs $(G, v_1), \ldots, (G, v_n)$, values $z_i = \exp(A'_L \lambda^{(L-1)}(v_i))$, the identity

$$\sum_{i=1}^{q} \alpha_i z_1^{\beta_{i,1}} \cdots z_n^{\beta_{i,n}} \exp(b'_L) + \alpha_{q+1} \exp(b'_L) = \sum_{j=1}^{r} \gamma_j z_1^{\delta_{j,1}} \cdots z_n^{\delta_{j,n}} \exp(b'_L),$$

and graph construction:

1. for $1 \leq i \leq q$, add α_i disjoint copies of the graph constructed as follows:
   (a) for $1 \leq k \leq n$, add $\beta_{i,k}$ fresh copies of the pointed graph (G, v_k);
   (b) create a fresh virtual node u and add an edge from u to the distinguished vertex in each copy added in the previous step;
2. add α_{q+1} fresh isolated virtual vertices;
3. for $1 \leq j \leq r$ and $1 \leq k \leq n$, add $\gamma_j \delta_{j,k}$ disjoint copies of (G, v_k).

The graph G_R is defined symmetrically: interchange q and r as well as (α, β) and (γ, δ), and omit Step 2.

Claim 19. *For G_Q and G_R as in Theorem 11 we have $\mu_{G_Q} \neq \mu_{G_R}$.*

Proof. Recall that G_Q and G_R are built from pointed graphs $(G, v_1), \ldots, (G, v_n)$ that are pairwise distinguishable by the $(L-1)$-round 1-WL algorithm. Looking at steps 1 (a) and 3 of the construction in Theorem 11 we see that there were $\sum_{i=1}^{q} \alpha_i \beta_{i,k} + \sum_{j=1}^{r} \gamma_j \delta_{j,k}$ disjoint copies of (G, v_k) added to both G_Q and G_R. Notice that the only edges added were ones that originate from a fresh vertex introduced in step 1 (b). Thus, in both G_Q and G_R, the copies of (G, v_k) are coloured

identically by the L-round 1-WL and, as it were, cancel each-other out as far as 1-WL graph distinguishability is concerned.

We now inspect the virtual vertices added in steps 1 (b) and 2. Firstly, notice that vertices added in 2 have a unique colour when compared to those added in 1 (b). Indeed these vertices have no edges associated with them and thus, if $\alpha_{q+1} \neq 0$, it is immediate that $\mu_{G_Q} \neq \mu_{G_R}$.

On the other hand suppose that $\alpha_{q+1} = 0$. We claim that the vertices added in step 1 (b) make the two graphs distinguishable. For this purpose let $c_1, \ldots, c_n$ be the $(L-1)$-round 1-WL colours produced on the pointed vertices $(G, v_1), \ldots, (G, v_n)$. Recall that they are all distinct. Taking any $v \in V_Q$ added in step 1 (b) we see that it takes the colour $\mu^{(L)}(v) = \{(c_1 \mapsto \beta_{i,1}), \ldots, (c_n \mapsto \beta_{i,n})\}$ for some $1 \leq i \leq q$. By inspecting $Q(x_1, \ldots, x_n) = R(x_1, \ldots, x_n)$ in Theorem 11 we see that there is no $1 \leq j \leq r$ for which $(\beta_{i,1}, \ldots \beta_{i,n}) = (\delta_{j,1}, \ldots \delta_{j,n})$. Thus, by the construction in step 1, there is no $u \in V_R$ such that $\mu^{(L)}(u) = \mu^{(L)}(v)$. With that we conclude $\mu_{G_Q} \neq \mu_{G_R}$. □

Claim 20. *For G_Q and G_R as in Theorem 11 we have $\lambda_{G_Q} = \lambda_{G_R}$.*

Proof. For the final global readout notice that, by construction of G_Q and G_R, there is a bijection between the vertices of G_Q and G_R added in Steps 1(a) and 3 that preserves the colours attributed by the L-round 1-WL algorithm. Indeed, Step 1 adds $\alpha_i \cdot \beta_{i,k}$ copies of (G_k, v_k) to G_Q while Step 3 does the same for G_R, and vice versa for $\gamma_j, \delta_{j,k}$. Thus, when concerned with the final readout, we need only look at the contribution of virtual nodes added at Steps 1(b) and 2. But notice that the final readout on the graphs is as follows

$$\sum_{v \in V_Q} \lambda^{(L)}(v) = K + \exp(b'_L) \cdot Q(z_1, \ldots, z_n) = K + \exp(b'_L) \cdot R(z_1, \ldots, z_n) = \sum_{v \in V_R} \lambda^{(L)}(v),$$

where K is the contribution of vertices that were added in Steps 1(a) and 3. Since the readout over the two graphs is the same, we have that $\lambda_{G_Q} = \lambda_{G_R}$, as desired. □

A high level explanation of how the two claims above complete Theorem 11 is in order. Recall that we assumed (for contradiction) that the values $z_1, \ldots, z_n$ are algebraically dependent over $\mathbb{Q}$. This allowed us to construct two distinct polynomials Q and R with positive integer coefficients such that $Q(z_1, \ldots, z_n) = R(z_1, \ldots, z_n)$. They then gave rise to two graphs G_Q and G_R satisfying $\mu_{G_Q} \neq \mu_{G_R}$ (Claim 19) and $\lambda_Q = \lambda_R$ (Claim 20). But Theorem 11 requires $\lambda_{G_1} = \lambda_{G_2} \Leftrightarrow \mu_{G_1} = \mu_{G_2}$ for all graphs G_1, G_2 thus resulting in a contradictions. It must then be the case that if $\mathcal{N}$ matches the L-round 1-WL test on graphs with virtual nodes, the values $z_1, \ldots, z_n$ (where $z_i = \exp(A'_L \lambda^{(L-1)}(v_i))$) must be algebraically independent over $\mathbb{Q}$.

Lastly, we extend our reasoning to other advertised forms of activation:

Claim 21. *Theorem 11 extends to GNNs with $\sigma \in \{\sinh, \cosh\}$ activation.*

Proof. To prove the Theorem in full we must show that the values in $\{\exp(A'_L \lambda^{(L-1)}(v)) \mid v \in V\}$ are algebraically independent over $\mathbb{Q}$ for any given graph $G = (V, E, \lambda)$. For the purposes of contradiction suppose this is not the case. Enumerating the values as $z_1, \ldots, z_n$ we obtain a non-zero polynomial $P(x_1, \ldots, x_n)$ with rational coefficients for which $P(z_1, \ldots, z_n) = 0$. We again assume, by multiplying by the lowest common denominator if needed, that the coefficients in P are integers.

Let us now inspect the expression $R(x_1, \ldots, x_n) = P(x_1, \ldots, x_n) \cdot P(x_1^{-1}, \ldots, x_n^{-1})$. Clearly, $R(z_1, \ldots, z_n) = 0$. Now, peeking into $P(x_1, \ldots, x_n)$ we see that it takes the form $\sum_{i=1}^{m} a_i M_i + b$, where each M_i is a monomial in $x_1, \ldots, x_n$. Thus, $R(x_1, \ldots, x_n)$ takes the form:

$$\Big(\sum_{i=1}^{m} a_i M_i + b\Big) \cdot \Big(\sum_{i=1}^{m} a_i M_i^{-1} + b\Big).$$

Expanding the expression above we obtain:

$$b^2 + b\sum_{i=1}^{m} a_i M_i + b\sum_{i=1}^{m} a_i M_i^{-1} + \sum_{i=1}^{m}\sum_{j=1}^{m} a_i a_j M_i M_j^{-1}.$$

Having the above we wish to rid ourselves of monomials that that feature both negative and positive powers; i.e. $M_i M_j^{-1}$. For this purpose let N be any monomial over $x_1, \ldots, x_n$ such that, for all

$1 \leq j \leq m$, $NM_j^{-1} = x_1^{k_1} \cdot \ldots \cdot x_n^{k_n}$ for some $k_1, \ldots, k_n \in \mathbb{N} \setminus \{0\}$. Clearly $S(x_1, \ldots, x_n) = NR(x_1, \ldots, x_n)$ is a polynomial that takes the form:

$$b^2 N + b \sum_{i=1}^{m} a_i N M_i + b \sum_{i=1}^{m} a_i N M_i^{-1} + \sum_{i=1}^{m} \sum_{j=1}^{m} a_i a_j N M_i M_j^{-1}.$$

Then, $S(x_1, \ldots, x_n)$ can be rewritten into the compact form $\sum_{i=1}^{m'} a'_i N_i$, where each $a'_i \in \mathbb{Z}$ and N_i is a monomial of the form $x_1^{c_{i,1}}, \ldots, x_n^{c_{i,n}}$, where $c_{i,1}, \ldots, c_{i,n} \in \mathbb{N}$. Now, recalling that $S(x_1, \ldots, x_n) = NP(x_1, \ldots, x_n) \cdot P(x_1^{-1}, \ldots, x_n^{-1})$, we see that $S(z_1, \ldots, z_n) = 0$. Additionally, $S(z_1^{-1}, \ldots, z_n^{-1}) = 0$. Thus, the relations $T_+(x_1, \ldots, x_n)$ and $T_-(x_1, \ldots, x_n)$ defined as:

$$\frac{e^{b'_L} S(x_1, \ldots, x_n) \pm e^{-b'_L} S(x_1^{-1}, \ldots, x_n^{-1})}{2}$$

also evaluate to zero at $z_1, \ldots, z_n$; it being understood that $\pm \equiv +$ in T_+ with $\pm \equiv -$ in T_-, and with b'_L being the L-th layer bias vector used by virtual nodes. Peaking inside $T_\pm(z_1, \ldots, z_n)$ we see:

$$\sum_{i=1}^{m'} a'_i \cdot \frac{e^{b'_L} N_i(z_1, \ldots, z_n) \pm e^{-b'_L} N_i(z_1^{-1}, \ldots, z_n^{-1})}{2}.$$

Since N_i are monomials and $z_j = e^{s_j}$ we obtain:

$$\sum_{i=1}^{m'} a'_i \cdot \frac{(e^{s_1})^{c_{i,1}} \cdot \ldots \cdot (e^{s_n})^{c_{i,n}} \cdot e^{b'_L} \pm (e^{s_1})^{-c_{i,1}} \cdot \ldots \cdot (e^{s_n})^{-c_{i,n}} \cdot e^{-b'_L}}{2}.$$

Recalling that $\sinh(x) = \frac{e^x - e^{-x}}{2}$ and $\cosh(x) = \frac{e^x + e^{-x}}{2}$ we obtain that $T_-(z_1, \ldots, z_n)$ and $T_+(z_1, \ldots, z_n)$ are, respectively:

$$\sum_{i=1}^{m'} a'_i \sinh(s_1 c_{i,1} + \cdots + s_n c_{i,n} + b'_L) \quad \text{and} \quad \sum_{i=1}^{m'} a'_i \cosh(s_1 c_{i,1} + \cdots + s_n c_{i,n} + b'_L).$$

Noting that $s_1, \ldots, s_n \in \{A'_L \lambda^{(L-1)}(v) \mid v \in V\}$, each expression $\sinh(s_1 c_{i,1} + \cdots + s_n c_{i,n} + b'_L)$ is then computable by a virtual node using $\sinh$ as activation. Indeed, taking pointed graphs $(G_1, v_1), \ldots, (G_n, v_n)$ such that $A'_L \lambda^{(L-1)}(v_j) = s_j$ we have that a virtual node u_i connected to $c_{i,j}$ copies of (G_j, v_j) for all $1 \leq j \leq n$ has $\lambda^{(L)}(u_i) = \sinh(s_1 c_{i,1} + \cdots + s_n c_{i,n} + b'_L)$. Similar observations follow for $\cosh$. The outermost sum can then be treated as the global readout of $\mathcal{N}$ over virtual nodes; provided that terms with negative coefficients have been moved to the right hand side of $T\pm(z_1, \ldots, z_n) = 0$. Having the above one can proceed with the same graph construction as in Theorem 11.

In this setting one again needs to be careful so that $T_\pm(x_1, \ldots, x_n)$ does not turn out to be the zero function. Notice, first, that $R(x_1, \ldots, x_n)$ is non-zero. For this let us again look at $P(x_1, \ldots, x_n) = \sum_{i=1}^{m} a_i x_1^{d_{i,1}} \cdot \ldots \cdot x_n^{d_{i,n}} + b$, where $a_i \in \mathbb{Z} \setminus \{0\}$, $b \in \mathbb{Z}$ and $d_{i,1}, \ldots, d_{i,n} \in \mathbb{N}$ not all zero. Let $\beta = \max\{d_{i,j} \mid 1 \leq i \leq m, 1 \leq j \leq n\} + 1$ and $B_i = \beta^1 d_{i,1} + \cdots + \beta^n d_{i,n}$. Then $P(x^{\beta^1}, \ldots, x^{\beta^n}) = \sum_{i=1}^{m} a_i x^{B_i} + b$ is a polynomial, where B_i captures via a β-ary encoding the sequence $(d_{i,1}, \ldots, d_{i,n})$. Since all degree sequences were taken to be distinct, we have that all B_i are as well. Thus the polynomial is non-zero. Without loss of generality, let B_1 and B_m be, respectively, the minimum and maximum values amongst $B_1, \ldots, B_m$. Now, inspecting $R(x^{\beta^1}, \ldots, x^{\beta^n})$ we have:

$$b^2 + b \sum_{i=1}^{m} a_i x^{B_i} + b \sum_{i=1}^{m} a_i x^{-B_i} + \sum_{i=1}^{m} \sum_{j=1}^{m} a_i a_j x^{B_i - B_j}.$$

Let us first suppose that $b = 0$. Then, multiplying by $x^{B_m - B_1}$ and taking the limit we obtain:

$$\lim_{x \to 0} \sum_{i=1}^{m} \sum_{j=1}^{m} a_i a_j x^{B_i - B_j + B_m - B_1} = a_1 a_m.$$

Otherwise, when $b \neq 0$, we have the following by multiplying by x^{B_m} and taking the limit:

$$\lim_{x \to 0} \left(b^2 x^{B_m} + b \sum_{i=1}^{m} a_i x^{B_i + B_m} + b \sum_{i=1}^{m} a_i x^{-B_i + B_m} + \sum_{i=1}^{m} \sum_{j=1}^{m} a_i a_j x^{B_i - B_j + B_m} \right) = b a_m.$$

Having both cases we conclude that the expression $R(x_1, \ldots, x_n)$ is non-zero.

Almost immediately, because of the above, $S(x_1, \ldots, x_n) = NR(x_1, \ldots, x_n)$ is a non-zero polynomial. Indeed, begin by writing $R(x_1, \ldots, x_n)$ as $\sum_{i=1}^{m'} a'_i x_1^{d'_{i,1}} \cdot \ldots \cdot x_n^{d'_{i,n}} + b'$ where $a'_i, b' \in \mathbb{Z}$, $d_{i,1}, \ldots, d_{i,n} \in \mathbb{Z}$ not all zero and with sequences $(d_{i,1}, \ldots, d_{i,n}) \neq (d_{j,1}, \ldots, d_{j,n})$ for $i \neq j$. But then, multiplying by a common monomial $N = x_1^{d_{0,1}} \cdot \ldots \cdot x_n^{d_{0,n}}$ cannot result in terms cancelling out. Indeed, $(d_{i,1} + d_{0,1}, \ldots, d_{i,n} + d_{0,n}) = (d_{j,1} + d_{0,1}, \ldots, d_{j,n} + d_{0,n})$ implies $(d_{i,1}, \ldots, d_{i,n}) = (d_{j,1}, \ldots, d_{j,n})$, a contradiction. Thus the polynomial $S(x_1, \ldots, x_n)$ has at least a single term with a non-zero coefficient making the polynomial itself non-zero.

Lastly, we need to show that $e^{b'_L} S(x_1, \ldots, x_n) \pm e^{-b'_L} S(x_1^{-1}, \ldots, x_n^{-1})$ is non-zero. Clearly, if the above is the case, then the relevant expressions $T_{\pm}(x_1, \ldots, x_n)$ are non-zero as well. Inspecting the monomials (with non-zero coefficients) of $S(x_1, \ldots, x_n)$ and $S(x_1^{-1}, \ldots, x_n^{-1})$ we see that they are all distinct (and that there are no constant terms in either of the expressions). Since $e^{b'_L} \neq 0$ we have that $e^{b'_L} S(x_1, \ldots, x_n) \pm e^{-b'_L} S(x_1^{-1}, \ldots, x_n^{-1})$ does not cause cancellations leaving the expression non-zero. □

D EXPERIMENTS

We would first like to add some clarity regarding our experimental setup. When comparing embeddings produced by GNNs we compare without rounding. That is to say, two embeddings are equal if they have the same binary value. Graphs for which a given GNN produced values $\pm\infty$ or NaN are automatically taken to be unsuccessful for that GNN. In some systems 128-bit floating point numbers are unavailable. We thus emulated float128 (IEEE 754 binary128) in software.

We present the extended sample of Figure 1 in Figure 2. In particular we note that in the 128-bit setting success ratio for sigmoid and tanh is exactly 1 when $A, C < 1$, and exactly 1 for sinh, cosh and exp when $A, C < 0.1$ (and as long as $\frac{C}{A} \notin \mathbb{N}$). This indicates that GNN errors in the float64 setting are due to imprecisions in float64 arithmetic.

It is interesting to analyse the effects that increased precision has on GNNs with different activation functions. Firstly, we see no meaningful gains for ReLU activated GNNs. This is to be expected as the values computed by our ReLU GNNs are seemingly well represented even in the float32 setting. It is clear that increased precision correlates with better performance for sigmoid and tanh activated GNNs. Indeed, we see that in the float128 setting both sigmoid and tanh start to perform better for values above 1. It is then interesting to contrast this with the seemingly weaker effect precision increase has on sinh, cosh and exp. This phenomenon is consistent with 64- and 128-bit float structure. Note that float64 reserves 11 bits for the exponent and 53 bits of significand precision; float128 specifies 15 bits for the exponent and 113 bits of significand precision. Thus, float128 allows more accurate representation of values close to 1, which is advantageous for sigmoid and tanh. On the other hand, sinh, cosh and exp are unbounded from above. Thus, these activation functions coupled with larger weights exhaust the representable range of the float128 exponent. See Table 1 for how much of the failure rate is attributed to overflow.

Figure 3 presents the full runs for sigmoid, tanh and sinh activated GNNs with $A, C \in \{0.1, 0.2, \ldots, 1\}$ in the 128-bit float setting. Notice that in the case of sigmoid and tanh the success rate is 1 unless $\frac{C}{A} \in \mathbb{N}$.

	sigmoid	tanh	sinh	cosh	exp
Success ratio	0.982	0.998	0.974	0.634	0.116
Non-finite failure rate	0.000	0.000	0.022	0.366	0.884

Table 1: Float32 results at self-weight $C = 0.1$ and aggregation-weight $A = 0.7$. A non-finite failure means that at least one final node embedding is NaN, $+\infty$, or $-\infty$; these outcomes are combined into one category.



Figure 2: Success ratio for GNNs by floating-point precision (panel row), fixed self-weight C (panel column), activation (heatmap row), and aggregation-weight A (heatmap column). Checkmarks denote success ratio 1; all panels share the same color scale.

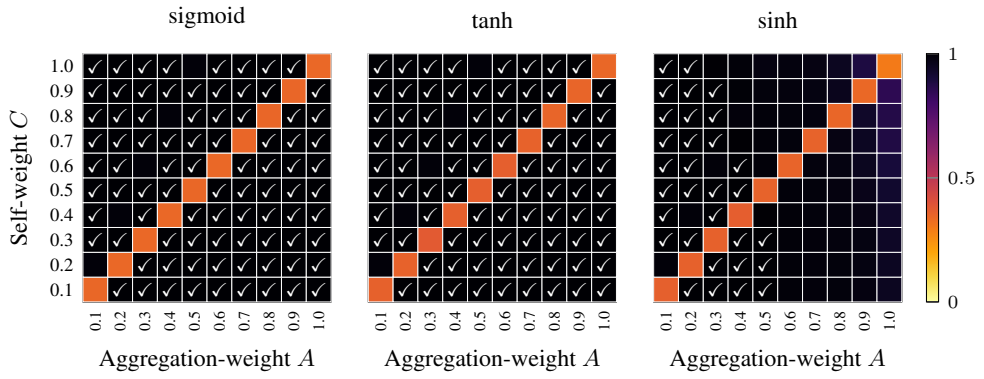

Figure 3: Success ratio in float128 arithmetic for GNNs with medium self-weight C (heatmap row), medium aggregation-weight A (heatmap column), and the sigmoid, tanh, and sinh activations (panels). Checkmarks denote success ratio 1; all panels share the same color scale.