ON THE EXPRESSIVE POWER OF SUBGRAPH GRAPH NEURAL NETWORKS FOR GRAPHS WITH BOUNDED CYCLES

ZIANG CHEN AND QIAO ZHANG

Abstract. Graph neural networks (GNNs) have been widely used in graph-related contexts. It is known that the separation power of GNNs is equivalent to that of the Weisfeiler-Lehman (WL) test; hence, GNNs are imperfect at identifying all non-isomorphic graphs, which severely limits their expressive power. This work investigates k-hop subgraph GNNs that aggregate information from neighbors with distances up to k and incorporate the subgraph structure. We prove that under appropriate assumptions, the k-hop subgraph GNNs can approximate any permutation-invariant/equivariant continuous function over graphs without cycles of length greater than $2k+1$ within any error tolerance. We also provide an extension to k-hop GNNs without incorporating the subgraph structure.

1. Introduction

Graph neural networks (GNNs) [14,20,22,25,32] have demonstrated remarkable effectiveness in modeling and analyzing graph-structured data. Their versatility has enabled impactful applications in diverse areas, including physics [21], bioinformatics [29], finance [23], electronic engineering [13, 15, 16], and operations research [8], just to name a few. From a theoretical perspective, GNNs are employed to learn or approximate functions on graphs, it is of essential importance to analyze and understand the expressiveness of GNNs, i.e., identify the function class on graphs that GNNs can well approximate, providing valuable insights to guide the design of more effective GNN architectures.

One of the cornerstone architectures in graph neural networks (GNNs) is the message-passing framework [12], which updates the features of each node layer-by-layer via incorporating information from its neighbors. Formally, consider a graph $G = (V, E)$, where $V = \{v_1, v_2, \ldots, v_n\}$ is the set of nodes and $E \subseteq V \times V$ denotes the edges. Each node is initially assigned features $h_1^{(0)}, h_2^{(0)}, \ldots, h_n^{(0)}$. Let AGGREGATE be a permutation-invariant operation (such as summation, averaging, or maximization), let $\mathcal{N}(v_i)$ represent the neighbors of v_i, and let $\{\{\cdot\}\}$ denote a multiset to handle duplicate elements. At the l-th layer, the feature of node v_i, namely $h_i^{(l)}$ is updated as follows, for learnable functions $f^{(l)}$ and $g^{(l)}$:

$$h_i^{(l)} = f^{(l)}\left(h_i^{(l-1)}, \text{AGGREGATE}\left(\left\{\left\{g^{(l)}(h_j^{(l-1)}) : v_j \in \mathcal{N}(v_i)\right\}\right\}\right)\right), \tag{1.1}$$

where $h_j^{(l-1)}$ is the vertex feature at the $(l-1)$-th layer.

Despite their empirical success, unfortunately, message-passing GNNs suffer from insufficient expressive power. Specifically, some non-isomorphic graphs cannot be distinguished by MP-GNNs. For example, Figure 1 shows two non-isomorphic graphs in which vertices of the same color have identical initial features and all edges have uniform weights. Although the graphs are non-isomorphic, vertices of the same color will always share identical features, regardless of the number of message-passing layers or the choice of functions $f^{(l)}, g^{(l)}$, AGGREGATE.

Date: January 11, 2025.

Key words and phrases. Graph neural networks, k-hop subgraph, expressive.

This occurs because each vertex receives the same aggregated information from its neighbors, rendering these graphs indistinguishable to MP-GNNs.

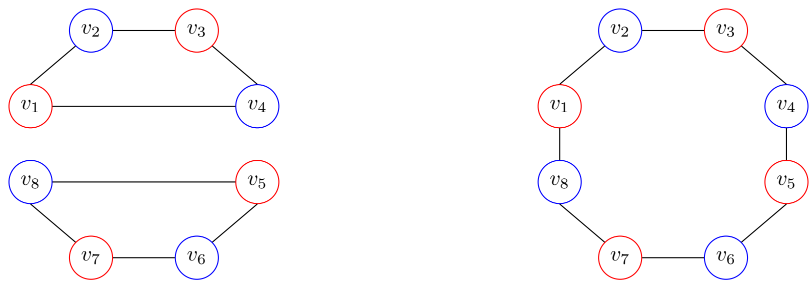

Figure 1. Two non-isomorphic graphs that cannot be distinguished by MP-GNNs or the WL test.

In general, the separation and expressive power of MP-GNNs are closely related to the Weisfeiler-Lehman (WL) test [24], a classical algorithm for addressing the graph isomorphism problem. The WL test is fundamentally a color refinement algorithm, where each vertex v_i is initially assigned a color $C^{(0)}(v_i)$ based on its initial features. The algorithm then iterates by applying the update:

$$C^{(l)}(v_i) = \text{HASH}\left(C^{(l-1)}(v_i), \left\{\left\{C^{(l-1)}(w) : w \in \mathcal{N}(v_i)\right\}\right\}\right), \tag{1.2}$$

which follows a similar structure to the update in (1.1). If the hash function is collision-free, two vertices share the same color at the l-th iteration if and only if they had the same color and identical multisets of neighbors' colors at the $(l-1)$-th iteration. The WL test terminates when the color partition stabilizes, typically within at most n iterations, and identifies two graphs as isomorphic if their final color multisets match.

It has been shown that MP-GNNs have the same separation power as the WL test [26], meaning two graphs are distinguished as non-isomorphic by the WL test if and only if they yield different outputs in some MP-GNN. It is further proven in [1, 11] that GNNs can universally approximate any continuous functions whose separation powers are upper bounded by the associated WL test. However, no polynomial-time algorithms are known to perfectly solve the graph isomorphism problem, so the WL test cannot distinguish certain pairs of non-isomorphic graphs, such as the example in Figure 1. Consequently, it is impossible for MP-GNNs to represent or approximate all permutation-invariant/equivariant functions.

In response, researchers have proposed alternative GNN architectures designed with enhanced separation capabilities. This project specifically focuses on unweighted graphs. One well-known approach in the literature is to employ higher-order GNNs [1, 9–11, 17–19, 31] that correspond to higher-order WL tests [4]. Roughly speaking, a k-th order GNN assigns a feature to each k-tuple of vertices and updates each tuple's feature based on information from its adjacent tuples.

In this work, we investigate another common technique in recent literature [3, 7, 28, 30] to enhance the expressive power of message-passing GNNs, that involves incorporating subgraph structures, rather than relying solely on vertex features from neighboring nodes. Such GNN architecture is termed subgraph GNNs and this paper rigorously characterizes their separation power by demonstrating that they can perfectly distinguish a large family of graphs with bounded cycles.

The rest of this paper will be organized as follows. We define subgraph GNNs and the associated WL test and introduce the motivation in Section 2. Our theory for the expressive power of subgraph GNNs are presented in Section 3 and Section 4, where we prove that any permutation-invariant/equivariant continuous function on graphs with bounded cycles can be approximated universally by subgraph GNNs. Our theory is extended in Section 5, and the whole paper is concluded in Section 6.

2. Subgraph graph neural networks and Weisfeiler-Lehman test

2.1. **Motivation.** One idea to enhance the expressive power of message-passing GNNs is to incorporate more information from neighboring vertices:

- The aggregation in (1.1) uses $\mathcal{N}(v_i)$, the set of neighbors of v_i. To incorporate additional information, one can define $d(u, v)$ as the distance between u and v in the graph G and
$$\mathcal{N}_k(v_i) := \{v \in G : d(v, v_i) \leqslant k\}, \quad k \geqslant 1.$$
- Beyond the features of vertices in $\mathcal{N}_k(v_i)$, one can also capture edge information, i.e., whether two vertices are connected. This means that the topology of $G|_{\mathcal{N}_k(v_i)}$, the subgraph of G restricted to $\mathcal{N}_k(v_i)$ (known as the k-hop subgraph rooted at v_i), can be used to update the feature of v_i.

Let $(G, h^{(l-1)})_{v_i,k}$ denote the subgraph $G|_{\mathcal{N}_k(v_i)}$ rooted at v_i, with each vertex having a feature from $h^{(l-1)}$. Accordingly, the vertex feature update rule is given by

$$h_i^{(l)} = f^{(l)}\left(h_i^{(l-1)}, g^{(l)}\left((G, h^{(l-1)})_{v_i,k}\right)\right). \tag{2.1}$$

The functions $f^{(l)}$ and $g^{(l)}$ are learnable, with $g^{(l)}$ taking constant value on isomorphic graphs. This scheme, termed the *k-hop subgraph GNN*, has various applications and adaptations in the existing literature [3, 7, 28, 30]. Notably, the permutation-invariant function $g^{(l)}$ is often parameterized as another GNN applied to the smaller subgraph $(G, h^{l-1})_{v_i,k}$.

Recall that the message-passing GNN (1.1) has limited separation power and fails to distinguish the graphs in Figure 1. However, the 2-hop subgraph GNN can successfully distinguish them. Specifically, the 2-hop subgraphs rooted at v_1 are shown in Figure 2 and are clearly non-isomorphic, indicating that v_1 in the two graphs in Figure 1 will have different feature after one layer of the 2-hop subgraph GNN.

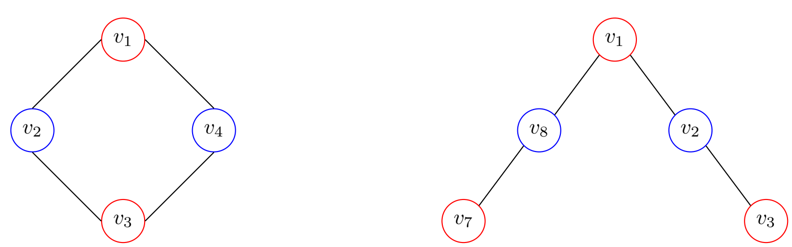

Figure 2. 2-hop subgraphs rooted at v_1 for graphs in Figure 1

Another observation is that the 2-hop subgraph GNN fails if we increase the cycle sizes in Figure 1—for instance, by changing one graph to have two 6-cycles and the other a single 12-cycle. In general, the k-hop subgraph GNN fails to distinguish between a graph with two $(2k + 2)$-cycles and one with a single $(4k + 4)$-cycle, though it succeeds when the cycle sizes are smaller. This suggests that larger cycles limit the separation power of the k-hop subgraph GNN.

This observation aligns with empirical findings in the literature. The ZINC dataset [6] consists of molecular graphs with no large cycles, and variants of subgraph GNNs have shown notable improvement over message-passing GNNs on this dataset [3, 7, 27, 28, 30].

2.2. **k-hop subgraph GNNs.** We rigorously define the k-hop subgraph GNNs in this subsection, for which we define the graph space first.

Definition 2.1 (Space of graphs with vertex features)**.** We use $\mathcal{G}_{n,m}$ to denote the space of all undirected unweighted graphs of n vertices with each vertex equipped with a feature in $\mathbb{R}^m$. The space $\mathcal{G}_{n,m}$ is equipped with the product topology of discrete topology (of graphs without vertex features) and Euclidean topology (of vertex features).

We use (G, H) to denote an element in $\mathcal{G}_{n,m}$ where $G = (E, V)$ is an undirected unweighted graph, and $H = (h_1, h_2, \ldots, h_n)$ is the collection of all vertex features. Given $(G, H) \in \mathcal{G}_{n,m}$ and $k \geqslant 1$, the k-hop subgraph GNN is defined as follows.

- The embedding layer maps each vertex feature $h_i \in \mathbb{R}^m$ as an embedding vector
$$h_i^{(0)} = f^{(0)}(h_i),$$
where $f^{(0)}$ is learnable.
- For $l = 1, 2, \ldots, L$, the information aggregation layer computes $h_i^{(l)}$ for $i = 1, 2, \ldots, n$.
- There are two types of outputs. The graph-level output computes a real number for the whole graph, namely
$$y = r\left(\text{AGGREGATE}\left(\left\{\left\{h_i^{(L)} : i \in \{1, 2, \ldots, n\}\right\}\right\}\right)\right),$$
where r is learnable. The vertex-level output assigns a real number for each vertex:
$$y_i = r(h_i^{(L)}), \quad i = 1, 2, \ldots, n.$$

In general, the intermediate vertex features $h_i^{(l)}$ can be defined in any topological space, while one usually uses Euclidean spaces in practice.

Definition 2.2 (Spaces of subgraph GNNs)**.** We use $\mathcal{F}_k$ to denote the collection of all k-hop subgraph GNNs with graph-level output, and use $\mathcal{F}_{k,v}$ to denote the collection of all k-hop subgraph GNNs with vertex-level output.

It is clear that a k-hop subgraph GNN with graph-level output is permutation-invariant, and a k-hop subgraph GNN with vertex-level output is permutation-equivariant, with respect to the following definition.

Definition 2.3 (Permutation-invariant and permutation-equivariant functions)**.** We say that a function $\Phi : \mathcal{G}_{n,m} \to \mathbb{R}$ is permutation-invariant if
$$\Phi(\sigma * (G, H)) = \Phi(G, H), \quad \forall\ \sigma \in S_n,$$
where $\sigma * (G, H)$ is the graph obtained by relabeling vertices in (G, H) according to the permutation σ, and that a function $\Phi : \mathcal{G}_{n,m} \to \mathbb{R}^n$ is permutation-equivariant if
$$\Phi(\sigma * (G, H)) = \sigma(\Phi(G, H)), \quad \forall\ \sigma \in S_n.$$

2.3. **Equivalent separation power of the k-hop subgraph WL test.** The WL test associated with the k-hop subgraph GNN is stated in Algorithm 1. The separation powers of k-hop subgraph GNN and the k-hop subgraph WL test are equivalent.

Definition 2.4. For $(G, H), (\hat{G}, \hat{H}) \in \mathcal{G}_{n,m}$, denote $\{\{C_i^{(L)} : i \in \{1, 2, \ldots, n\}\}\}$ and $\{\{\hat{C}_i^{(L)} : i \in \{1, 2, \ldots, n\}\}\}$ as their final color multisets output by the k-hop subgraph WL test.

(i) We say $(G, H) \overset{k}{\sim} (\hat{G}, \hat{H})$ if $\{\{C_i^{(L)} : i \in \{1, 2, \ldots, n\}\}\} = \{\{\hat{C}_i^{(L)} : i \in \{1, 2, \ldots, n\}\}\}$ for any $L > 0$ and any hash functions.

(ii) We say $(G, H) \overset{k,v}{\sim} (\hat{G}, \hat{H})$ if $C_i^{(L)} = \hat{C}_i^{(L)},\ i = 1, 2, \ldots, n$, for any $L > 0$ and any hash functions.

Algorithm 1 k-hop Subgraph Weisfeiler-Lehman test

Require: A graph $(G, H) \in \mathcal{G}_{n,m}$ and iteration limit $L > 0$.
Initialize the vertex color

$$C_i^{(0)} = \mathrm{HASH}(h_i), \quad i = 1, 2, \ldots, n$$

while $l = 1, 2, \ldots, L$ **do**
Refine the color

$$C_i^{(l)} = \mathrm{HASH}\left(C_i^{(l-1)}, (G, C_{v_i,k}^{(l-1)})\right). \tag{2.2}$$

end while
Output: Color multiset $\{\{C_i^{(L)} : i \in \{1, 2, \ldots, n\}\}\}$.

We remark that two multisets are identical if for any element, its multiplicities in two multisets are the same.

Theorem 2.5. *For any $(G, H), (\hat{G}, \hat{H}) \in \mathcal{G}_{n,m}$ and any $k > 0$, the following are equivalent:*

(i) $(G, H) \overset{k}{\sim} (\hat{G}, \hat{H})$.
(ii) $F(G, H) = F(\hat{G}, \hat{H})$ *for any* $F \in \mathcal{F}_k$.
(iii) *For any* $F_v \in \mathcal{F}_{k,v}$, *there exists* $\sigma \in S_n$ *such that* $F_v(G, H) = \sigma(F_v(\hat{G}, \hat{H}))$.

Moreover, $(G, H) \overset{k,v}{\sim} (\hat{G}, \hat{H})$ *if and only if* $F_v(G, H) = F_v(\hat{G}, \hat{H})$ *for any* $F_v \in \mathcal{F}_{k,v}$.

Proof. The proof follows similar lines as in the proof of [5, Theorem 4.2] that is inspired by [26]. □

Corollary 2.6. *For any $(G, H) \in \mathcal{G}_{n,m}$ and any $k > 0$. Let $\{\{C_i^{(L)} : i \in \{1, 2, \ldots, n\}\}\}$ be the color multiset output by the 1-hop subgraph WL test. For any $i, i' \in \{1, 2, \ldots, n\}$, the following are equivalent:*

(i) $C_i^{(L)} = C_{i'}^{(L)}$ *for any* $L > 0$ *and any hash function.*
(ii) $F_v(G, H)_i = F_v(G, H)_{i'}$ *for any* $F_v \in \mathcal{F}_{k,v}$.

Proof. Apply Theorem 2.5 to (G, H) and $\sigma * (G, H)$ where σ is the permutation that switches i, i' and keep all other indices unchanged. □

3. Expressive power of 1-hop subgraph GNNs

This section characterizes the expressive power of 1-hop subgraph GNNs. The main theorem is stated as follows, which proves that 1-hop subgraph GNNs can approximate any permutation-invariant/equivariant continuous functions on graphs without cycles of length greater than 3.

Theorem 3.1. *Let $\mathbb{P}$ be a Borel probability measure on $\mathcal{G}_{n,m}$. Suppose that for $\mathbb{P}$-almost surely (G, H), the graph G is connected and has no cycles of length greater than 3. Then, the following hold.*

(i) *For any $\epsilon, \delta > 0$ and any permutation-invariant continuous function $\Phi : \mathcal{G}_{n,m} \to \mathbb{R}$, there exists $F \in \mathcal{F}_1$ such that*

$$\mathbb{P}\left[|F(G, H) - \Phi(G, H)| > \delta\right] < \epsilon.$$

(ii) *For any $\epsilon, \delta > 0$ and any permutation-equivariant continuous function $\Phi_v : \mathcal{G}_{n,m} \to \mathbb{R}^n$, there exists $F_v \in \mathcal{F}_{1,v}$ such that*

$$\mathbb{P}\left[\|F_v(G, H) - \Phi_v(G, H)\| > \delta\right] < \epsilon.$$

Throughout this paper, we always denote $\|\cdot\|$ as the standard ℓ_2-norm on $\mathbb{R}^n$. We present the proof of Theorem 3.1 in the rest of this section, which is based on the following theorem and corollary.

Theorem 3.2. *Consider $(G,H),(\hat{G},\hat{H})\in\mathcal{G}_{n,m}$. Suppose that G and $\hat{G}$ are both connected and have no cycles of length greater than 3. If $(G,H)\overset{1}{\sim}(\hat{G},\hat{H})$, then (G,H) and $(\hat{G},\hat{H})$ must be isomorphic.*

Corollary 3.3. *Consider any $(G,H)\in\mathcal{G}_{n,m}$ where G is connected and has no cycles of length greater than 3. Let $\{\{C_i^{(L)}:i\in\{1,2,\ldots,n\}\}\}$ be the color multiset output by the 1-hop subgraph WL test. For any $i,i'\in\{1,2,\ldots,n\}$, if $C_i^{(L)}=C_{i'}^{(L)}$ holds for any $L>0$ and any hash function, then we have for any permutation-equivariant function $\Phi:\mathcal{G}_{n,m}\to\mathbb{R}^n$ that $\Phi(G,H)_i=\Phi(G,H)_{i'}$.*

In the acyclic graph setting, it is proven in [2] that two trees indistinguishable by the classic WL test (1.2) must be isomorphic. Theorem 3.2 can be viewed as a generalization of this result from [2]. We will postpone the proofs of Theorem 3.2 and Corollary 3.3 and first prove Theorem 3.1 (i) using Theorem 3.2 and the Stone-Weierstrass theorem.

Proof of Theorem 3.1 (i). There exists a compact and permutation-invariant subset $X\subseteq\mathcal{G}_{n,m}$ such that $\mathbb{P}[X]>1-\epsilon$ and that for any $(G,H)\in X$, G is connected and has no cycles of length greater than 3. Due to Theorem 3.2 and the permutation-invariant property of Φ, $\Phi|_X:X\to\mathbb{R}$ induces a continuous map on the quotient space $\widetilde{\Phi|_X}:X/\overset{1}{\sim}\to\mathbb{R}$ By the same reason, for $F\in\mathcal{F}_1$, $F|_X:X\to\mathbb{R}$ also induces a continuous map $\widetilde{F|_X}:X/\overset{1}{\sim}\to\mathbb{R}$. Consider any $(G,H),(\hat{G},\hat{H})\in X$ that represent different elements in $X/\overset{1}{\sim}$, Theorem 2.5 guarantees that there exists $F\in\mathcal{F}_1$ such that $F(G,H)\neq F(\hat{G},\hat{H})$, suggesting that $\{\widetilde{F|_X}:F\in\mathcal{F}_1\}$ separates points on $X/\overset{1}{\sim}$. Therefore, by the Stone-Weierstrass theorem, one can conclude that there exists $F\in\mathcal{F}_1$ such that

$$\left\|\widetilde{F|_X}-\widetilde{\Phi|_X}\right\|_{L^\infty(X/\overset{1}{\sim})}<\delta,$$

which implies that

$$|F(G,H)-\Phi(G,H)|<\delta,\quad\forall\ (G,H)\in X.$$

Thus, it holds that

$$\mathbb{P}\left[|F(G,H)-\Phi(G,H)|>\delta\right]\leqslant\mathbb{P}[\mathcal{G}_{n,m}\backslash X]<\epsilon,$$

which completes the proof. □

The proof of Theorem 3.1 (ii) requires a generalized Stone-Weierstrass theorem for equivariant functions.

Theorem 3.4 (Generalized Stone-Weierstrass theorem [1, Theorem 22])**.** *Let X be a compact topological space and let $\mathbf{G}$ be a finite group that acts continuously on X and $\mathbb{R}^n$. Define the collection of all equivariant continuous functions from X to $\mathbb{R}^n$ as follows:*

$$\mathcal{C}_e(X,\mathbb{R}^n)=\{F\in\mathcal{C}(X,\mathbb{R}^n):F(g*x)=g*F(x),\ \forall\ x\in X,g\in\mathbf{G}\}.$$

Consider any $\mathcal{F}\subset\mathcal{C}_e(X,\mathbb{R}^n)$ and any $\Phi\in\mathcal{C}_e(X,\mathbb{R}^n)$. Suppose the following conditions hold:

(i) *$\mathcal{F}$ is a subalgebra of $\mathcal{C}(X,\mathbb{R}^n)$ and $\mathbf{1}\in\mathcal{F}$.*
(ii) *For any $x,x'\in X$, if $f(x)=f(x')$ holds for any $f\in\mathcal{C}(X,\mathbb{R})$ with $f\mathbf{1}\in\mathcal{F}$, then for any $F\in\mathcal{F}$, there exists $g\in\mathbf{G}$ such that $F(x)=g*F(x')$.*
(iii) *For any $x,x'\in X$, if $F(x)=F(x')$ holds for any $F\in\mathcal{F}$, then $\Phi(x)=\Phi(x')$.*

(iv) *For any* $x \in X$*, it holds that* $\Phi(x)_i = \Phi(x)_{i'}, \ \forall \ (i,i') \in I(x)$*, where*

$$I(x) = \left\{(i,i') \in \{1,2,\ldots,n\}^2 : F(x)_i = F(x)_{i'}, \ \forall \ F \in \mathcal{F}\right\}.$$

Then for any $\epsilon > 0$*, there exists* $F \in \mathcal{F}$ *such that*

$$\sup_{x \in X} \|\Phi(x) - F(x)\| < \epsilon.$$

Proof of Theorem 3.1 (ii). There exists a compact and permutation-invariant subset $X \subseteq \mathcal{G}_{n,m}$ such that $\mathbb{P}[X] > 1-\epsilon$ and that for any $(G,H) \in X$, G is connected and has no cycles of length greater than 3. The rest is to apply Theorem 3.4 on X and $\mathcal{F} = \mathcal{F}_{1,v}$, for which one needs to verify the four condition in Theorem 3.4.

- *Verification of Condition (i).* By its construction, $\mathcal{F}_{1,v}$ is a subalgebra of $\mathcal{C}(X,\mathbb{R})$. In addition, $\mathbf{1} \in \mathcal{F}_{1,v}$ if the output function r always takes the constant value 1.
- *Verification of Condition (ii).* Notice that $\mathcal{F}_1\mathbf{1} \subset \mathcal{F}_{1,v}$. If $F(G,H) = F(\hat{G},\hat{H}), \ \forall \ F \in \mathcal{F}_1$, then Theorem 2.5 implies that for any $F_v \in \mathcal{F}_{1,v}$, one has $F_v(G,H) = \sigma(F_v(\hat{G},\hat{H}))$ for some permutation $\sigma \in S_n$.
- *Verification of Condition (iii).* Suppose that $F_v(G,H) = F_v(\hat{G},\hat{H}), \ \forall \ F_v \in \mathcal{F}_{1,v}$. By Theorem 2.5, it holds that $(G,H) \overset{1,v}{\sim} (\hat{G},\hat{H})$. Then apply Theorem 3.2 and Corollary 3.3, and one can conclude that $\Phi(G,H) = \Phi(\hat{G},\hat{H})$.
- *Verification of Condition (iv).* Condition (iv) is a direct corollary of Corollary 2.6 and Corollary 3.3.

□

Finally, we present the proof of Theorem 3.2 and Corollary 3.3.

Proof of Theorem 3.2. Let $\mathcal{A} = (A_1, A_2, \ldots, A_m)$ be an m-tuple of subgraphs of a graph A, and let $\mathcal{B} = (B_1, B_2, \ldots, B_m)$ be an m-tuple of subgraphs of a graph B. Let $\cup\mathcal{A}$ be the union of the vertices in $A_1, A_2, \ldots, A_m$, and let $\cup\mathcal{B}$ be the union of the vertices in $B_1, B_2, \ldots, B_m$. We say that $\mathcal{A}$ and $\mathcal{B}$ are isomorphic if there exists a bijective map of vertices of $\cup\mathcal{A}$ to vertices of $\cup\mathcal{B}$ such that for any $i \in \{1,2,\ldots,m\}$,

- all vertices of A_i are mapped to vertices of B_i with the same label and vice versa
- all edges of A_i are mapped to edges of B_i and vice versa.

Assume for the sake of contradiction that (G,H) and $(\hat{G},\hat{H})$ cannot be distinguished by the WL test, which means the multisets of labels of vertices in G and $\hat{G}$ are the same and any $v_1 \in G$ and $v_2 \in \hat{G}$ with the same label must have isomorphic 1-degree neighborhoods.

We abbreviate an induced subgraph of G or $\hat{G}$ as its set of vertices. For any set S of vertices, let $\mathcal{N}(S)$ be the set of all vertices in S or neighboring some vertex of S. We prove the following statement by induction: for any $m \in \{1,2,\ldots,|G|\}$, there exist connected isomorphic subsets $S_1 \subseteq G$ and $S_2 \subseteq \hat{G}$ of size m such that $(S_1, \mathcal{N}(S_1))$ and $(S_2, \mathcal{N}(S_2))$ are isomorphic. For the base case, choose any two vertices in G and $\hat{G}$ with the same label. For the inductive step, suppose that S_1 and S_2 are sets of size $m < |G|$, and we want to find two sets S_1' and S_2' with size $m+1$ that satisfy the inductive statement. Let v_1 be a vertex not in S_1 adjacent to a vertex in S_1, and let v_2 be the image of v_1 under any isomorphism from $(S_1, \mathcal{N}(S_1))$ to $(S_2, \mathcal{N}(S_2))$. Let $\mathcal{N}(v_1)$ and $\mathcal{N}(v_2)$ be the sets of vertices with distance at most 1 from v_1 and v_2, respectively. We claim that $(S_1 \cup \{v_1\}, \mathcal{N}(S_1) \cup \mathcal{N}(v_1))$ and $(S_2 \cup \{v_2\}, \mathcal{N}(S_2) \cup \mathcal{N}(v_2))$ are isomorphic. We know that there is an isomorphism from $\mathcal{N}(S_1)$ to $\mathcal{N}(S_2)$ taking v_1 to v_2. Consider $T_1 = \mathcal{N}(v_1) \backslash \mathcal{N}(S_1)$ and $T_2 = \mathcal{N}(v_2) \backslash \mathcal{N}(S_2)$. Notice that $(\mathcal{N}(S_1), v_1 \cup T_1)$ and $(\mathcal{N}(S_2), v_2 \cup T_2)$ are isomorphic. We claim that we cannot connect any vertex u_1 of T_1 to a vertex of $\mathcal{N}(S_1)$ other than v_1. If u_1 is connected to some vertex $u_2 \neq v_1$ in $\mathcal{N}(S_1)$, then both

v_1 and u_2 must have some neighbor in S_1: call these u_3 and u_4. If $u_3 = u_4$, then we have the cycle $u_1 \to v_1 \to u_3 \to u_2 \to u_1$. If $u_3 \neq u_4$, then there must be a path through edges of S_1 from u_3 to u_4, so we create a cycle containing $u_1 \to v_1 \to u_3 \to \cdots \to u_4 \to u_2 \to u_1$. Both of these cycles have length greater than 3, contradiction. Thus, u_1 is not connected to any vertex of $\mathcal{N}(S_1)$ other than v_1. Thus, we have proven the inductive step and the proof is completed. □

Proof of Corollary 3.3. By the proof of Theorem 3.2, there exists a permutation $\sigma \in S_n$ such that $\sigma(i) = i'$ and $\sigma * (G, H) = (G, H)$. Then the result holds immediately. □

4. Expressive power of k-hop subgraph GNNs

For the expressive power of k-hop subgraph GNNs, the theory is an extension of Theorem 3.1, in the sense that k-hop subgraph GNNs can approximate any permutation-invariant/equivariant continuous functions on graphs without cycles of length greater than $2k+1$, but an additional assumption is required.

Definition 4.1. A graph $(G, H) \in \mathcal{G}_{n,m}$ is said to be k-separable if the following condition holds when the k-hop subgraph WL test terminates without hash collisions: For any three vertices u, v_1, v_2 with $d(u, v_1) = d(u, v_2) = k$ and $v_1 \neq v_2$, the final colors of v_1 and v_2 output by the k-hop subgraph WL test are different.

Theorem 4.2. *Let $\mathbb{P}$ be a Borel probability measure on $\mathcal{G}_{n,m}$. Suppose that $\mathbb{P}$-almost surely, (G, H) is k-separated and G is connected with no cycles of length greater than $2k+1$. Then, the following hold.*

(i) *For any $\epsilon, \delta > 0$ and any permutation-invariant continuous function $\Phi : \mathcal{G}_{n,m} \to \mathbb{R}$, there exists $F \in \mathcal{F}_k$ such that*
$$\mathbb{P}\left[|F(G, H) - \Phi(G, H)| > \delta\right] < \epsilon.$$
(ii) *For any $\epsilon, \delta > 0$ and any permutation-equivariant continuous function $\Phi_v : \mathcal{G}_{n,m} \to \mathbb{R}$, there exists $F_v \in \mathcal{F}_{k,v}$ such that*
$$\mathbb{P}\left[\|F_v(G, H) - \Phi_v(G, H)\| > \delta\right] < \epsilon.$$

We need the following theorem to prove Theorem 4.2.

Theorem 4.3. *Consider $k \geqslant 2$ and $(G, H), (\hat{G}, \hat{H}) \in \mathcal{G}_{n,m}$ that are both k-separable. Suppose that G and $\hat{G}$ are both connected and have no cycles of length greater than $2k+1$. If $(G, H) \overset{k}{\sim} (\hat{G}, \hat{H})$, then (G, H) and $(\hat{G}, \hat{H})$ must be isomorphic.*

Proof of Theorem 4.2. Based on Theorem 4.3, the proof of Theorem 4.2 follows the same lines as the proof of Theorem 3.1. □

Next, we present the proof of Theorem 4.3. Let S be a subset of vertices of a graph. Define $\mathcal{N}_k(S)$ as the set of all vertices with distance at most k from any vertex in S and $\mathcal{N}_k(v)$ as the set of all vertices with distance at most k from v. If S is nonempty, define $d_S(v)$ as the minimum distance from v to any vertex in S.

To prove our Theorem 4.3, we need a lemma, which rules out the existence of undetected edges when we do our induction.

Lemma 4.4. *Let $k \geqslant 2$ and let S be a connected subset of vertices of a connected graph G with no cycles of length greater than $2k+1$. Let u_1 be a vertex not in S adjacent to a vertex in S. Then, no vertex in $T = \mathcal{N}_k(u_1) \backslash \mathcal{N}_k(S)$ can be connected to a vertex in $\mathcal{N}_k(S) \backslash \mathcal{N}_k(u_1)$.*

Proof. Assume for the sake of contradiction that there exists a vertex $u_{k+1} \in T$ connected to $v_k \in \mathcal{N}_k(S) \backslash \mathcal{N}_k(u_1)$. Notice that $d_S(v_k) \leqslant k$ because $v_k \in \mathcal{N}_k(S)$. If $d_S(v_k) < k$, then $d_S(u_{k+1}) \leqslant k$, which contradicts $u_{k+1} \notin \mathcal{N}_k(S)$. Thus, $d_S(v_k) = k$.

Therefore, there must exist vertices $u_2, u_3, \ldots, u_k$ and $v_0, v_1, \ldots, v_{k-1}$ such that u_i and u_{i+1} are connected for $i \in \{1, 2, \ldots, k\}$, v_i and v_{i+1} are connected for $i \in \{0, 1, \ldots, k-1\}$, and $v_0 \in S$.

We claim that $u_1, u_2, \ldots, u_{k+1}, v_0, \ldots, v_k$ are pairwise distinct. For any two connected vertices a and b, notice that $|d_S(a) - d_S(b)| \leqslant 1$ because any path of length n from a to a vertex of S can be extended to a path of length $n+1$ from b to a vertex of S and vice versa. Since $d_S(u_1) = 1$, $d_S(u_{k+1}) = k+1$, $d_S(v_0) = 0$, and $d_S(v_k) = k$, we must have $d_S(u_i) = i$ and $d_S(v_i) = i$ for all valid i. Thus, the only possible pairs of vertices of $u_1, u_2, \ldots, u_{k+1}, v_0, \ldots, v_k$ that can be equal are (u_i, v_i) for $i \in \{1, 2, \ldots, k\}$. Assume for the sake of contradiction that $u_i = v_i$ for some i. Then, there exists a path $u_1 \to u_2 \to \cdots \to u_i \to v_{i+1} \to \cdots \to v_k$ of length $k-1$ from u_1 to v_k, contradicting the fact that $v_k \notin \mathcal{N}_k(u_1)$. Thus, the vertices $u_1, u_2, \ldots, u_{k+1}, v_0, \ldots, v_k$ are pairwise distinct.

Since S is connected, there exists a path with edges in S from v_0 to a vertex in S adjacent to u_1. We can combine this path with $u_1 \to u_2 \to \cdots \to u_{k+1} \to v_k \to v_{k-1} \to \cdots \to v_0$ to create a cycle containing vertices $u_1, u_2, \ldots, u_{k+1}, v_0, \ldots, v_k$. This cycle contains at least $2k+2$ vertices, a contradiction. □

Proof of Theorem 4.3. Let $\mathcal{A} = (A_1, A_2, \ldots, A_m)$ be an m-tuple of subgraphs of a graph A, and let $\mathcal{B} = (B_1, B_2, \ldots, B_m)$ be an m-tuple of subgraphs of a graph B. Let $\cup\mathcal{A}$ be the union of the vertices in $A_1, A_2, \ldots, A_m$, and let $\cup\mathcal{B}$ be the union of the vertices in $B_1, B_2, \ldots, B_m$. We say that $\mathcal{A}$ and $\mathcal{B}$ are isomorphic if there exists a bijective map of vertices of $\cup\mathcal{A}$ to vertices of $\cup\mathcal{B}$ such that for any $i \in \{1, 2, \ldots, m\}$,

- all vertices of A_i are mapped to vertices of B_i with the same label and vice versa
- all edges of A_i are mapped to edges of B_i and vice versa.

Assume for the sake of contradiction that G and $\hat{G}$ cannot be distinguished by the WL test, which means the multisets of labels of vertices in G and $\hat{G}$ are the same and any $v_1 \in G$ and $v_2 \in \hat{G}$ with the same label must have isomorphic k-degree neighborhoods.

For any set S of vertices, let $\overline{S}$ be the set of all vertices with distance at most k from some vertex of S. We will use the following definition of graph isomorphism: two graphs are isomorphic if and only if there exists a bijective map between the vertices that preserves edges and labels.

We prove the following statement by induction: for any $k \in \{1, 2, \ldots, |G|\}$, there exist connected isomorphic subsets $S_1 \subseteq G$ and $S_2 \subseteq \hat{G}$ of size k such that $(S_1, \overline{S_1})$ and $(S_2, \overline{S_2})$ are isomorphic. For the base case, choose any two vertices in G and $\hat{G}$ with the same label. For the inductive step, suppose that S_1 and S_2 are valid sets of size $m < |G|$, and we want to find two sets S_1' and S_2' with size $m+1$. Let v_1 be a vertex not in S_1 adjacent to a vertex in S_1, and let v_2 be the image of v_1 under an isomorphism f from $(S_1, \overline{S_1})$ to $(S_2, \overline{S_2})$. Let N_1 be the set of vertices with distance at most k from v_1, and let N_2 be the set of vertices with distance at most k from v_2. Notice that f takes $\overline{S_1} \cap N_1$ to $\overline{S_2} \cap N_2$.

Notice that any vertex in $N_1 \backslash \overline{S_1}$ has distance k from v_1, so we know which vertices of N_1 it is connected to, since we can distinguish the vertices of $N_1 \backslash \overline{S_1}$ by label. We also know by the above lemma that no vertex in $N_1 \backslash \overline{S_1}$ is connected to any vertex in $\overline{S_1} \backslash N_1$, so we know whether any two vertices of $\overline{S_1} \cup N_1$ are connected. These edges must be the same edges as in $\overline{S_2} \cup N_2$, so $\overline{S_1} \cup N_1$ and $\overline{S_2} \cup N_2$ are isomorphic and the inductive step is complete.

If the graphs are infinite, the proof shows that the same isomorphism that sends S_1 to S_2 for one value of m also sends S_1 to S_2 for all values of m. Thus, by continuing this process to arbitrarily large m, we can construct an isomorphism between the two graphs. □

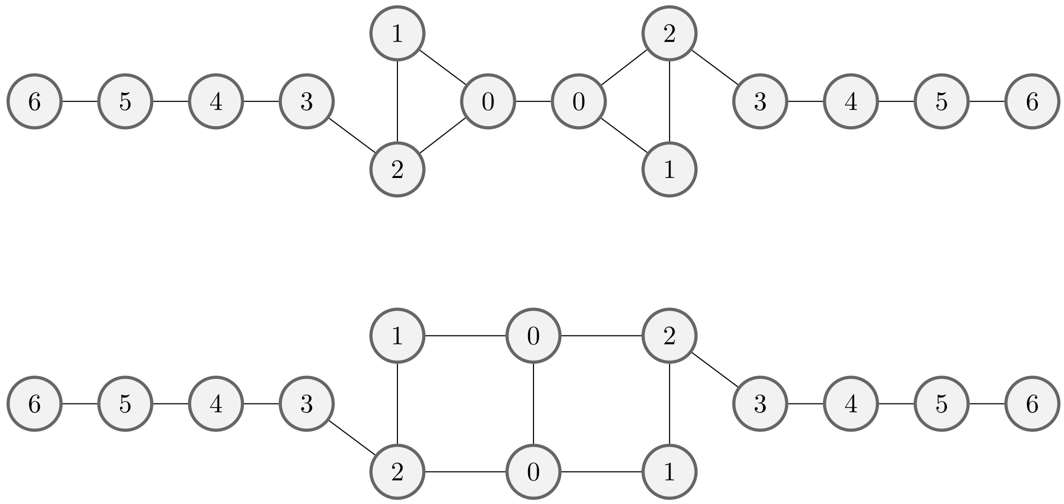

Figure 3

At the end of this section, we present two non-isomorphic k-separable graphs that can be distinguished by the k-hop subgraph WL test, but are, however, treated the same by the original WL test. Let $k = 3$, and consider the graphs in Figure 3 with initial node features as labeled. Notice that neither graph has a cycle with length more than $2k + 1 = 7$ vertices. Furthermore, for any vertex v in either graph, any distinct vertices v_1 and v_2 with distance exactly 3 from v are of different labels. Thus, our results imply that these two graphs can be distinguished by the 3-hop subgraph WL test. However, we can see that both graphs immediately stabilize when the regular WL test is applied, so the regular WL test cannot distinguish between the graphs.

5. An extension to k-hop GNNs

This section extends our theory to k-hop GNNs without incorporating the subgraph structure, for which the vertex feature is updated via

$$h_i^{(l)} = f^{(l)}\left(h_i^{(l-1)}, \text{AGGREGATE}\left(\left\{\left\{g^{(l)}(h_j^{(l-1)}, d(v_i, v_j)) : v_j \in \mathcal{N}_k(v_i)\right\}\right\}\right)\right).$$

We will use $\mathcal{F}'_k$ and $\mathcal{F}'_{k,v}$ to denote the collections of k-hop GNNs with graph-level and vertex-level outputs, respectively. The associated k-hop WL test implements the color refinement as follows:

$$C^{(l)}(v_i) = \text{HASH}\left(C^{(l-1)}(v_i), \left\{\left\{(C^{(l-1)}(v_j), d(v_i, v_j)) : v_j \in \mathcal{N}_k(v_i)\right\}\right\}\right).$$

The next theorem is our main result in this section.

Definition 5.1. A graph $(G, H) \in \mathcal{G}_{n,m}$ is said to be k-strongly separable if the following condition holds when the k-hop WL test terminates without hash collisions: For any two vertices v_1, v_2 with $d(v_1, v_2) \leqslant 2k$, the final colors of v_1 and v_2 output by the k-hop WL test are different.

Theorem 5.2. *Let $\mathbb{P}$ be a Borel probability measure on $\mathcal{G}_{n,m}$. Suppose that $\mathbb{P}$-almost surely, (G, H) is k-strongly separated and G is connected with no cycles of length greater than $2k - 1$. Then, the following hold.*

(i) *For any* $\epsilon, \delta > 0$ *and any permutation-invariant continuous function* $\Phi : \mathcal{G}_{n,m} \to \mathbb{R}$, *there exists* $F \in \mathcal{F}'_k$ *such that*

$$\mathbb{P}\left[|F(G,H) - \Phi(G,H)| > \delta\right] < \epsilon.$$

(ii) *For any* $\epsilon, \delta > 0$ *and any permutation-equivariant continuous function* $\Phi_v : \mathcal{G}_{n,m} \to \mathbb{R}$, *there exists* $F_v \in \mathcal{F}'_{k,v}$ *such that*

$$\mathbb{P}\left[\|F_v(G,H) - \Phi_v(G,H)\| > \delta\right] < \epsilon.$$

Proof. Based on the following theorem, one can prove Theorem 5.2 following the same lines in the proof of Theorem 3.1. □

Theorem 5.3. *Consider* $k \geqslant 2$ *and* $(G,H), (\hat{G},\hat{H}) \in \mathcal{G}_{n,m}$ *that are both* k*-stronly separable. Suppose that* G *and* $\hat{G}$ *are both connected and have no cycles of length greater than* $2k-1$*. If* (G,H) *and* $(\hat{G},\hat{H})$ *are indistinguishable by the* k*-hop WL test, then they must be isomorphic.*

Proof. We claim that using the k-hop WL test and the condition that any two vertices such that the distance between them is at most $2k$ are of different labels, we can obtain all the information we would otherwise obtain using the $(k-1)$-hop subgraph WL test and the same condition. For any vertex v, notice that any pair of vertices in $\mathcal{N}_k(v)$ have a distance at most $2k$ from each other, so they are of different colors. Suppose u_1 and u_2 are vertices in $\mathcal{N}_{k-1}(v)$. If the label of u_1 implies it has a neighbor with the same label as u_2, then this neighbor must be u_2, as the only neighbors of u_1 are in $\mathcal{N}_k(v)$ and all vertices in $\mathcal{N}_k(v)$ are of different colors. Otherwise, u_1 and u_2 cannot be connected by edges. Thus, for any u_1 and u_2 in $\mathcal{N}_{k-1}(v)$, we can find whether there is an edge between u_1 and u_2. Therefore, we can find the induced subgraph of vertices in $\mathcal{N}_{k-1}(v)$ for any vertex v. Then the result is a direct corollary of Theorem 4.3. □

Lastly, we show that Theorem 5.2 does not hold true if the k-strong separability is removed. Consider the two graphs in Figure 4, in which all vertices have the same initial feature. Each

Figure 4. The k-strong separability assumption is necessary in Theorem 5.2

vertex has three neighbors of distance 1, two neighbors of distance 2, and no neighbors of higher distance, so all vertices would have the same feature in any k-hop GNN for any positive integer k. Thus, these two graphs satisfy everything in Theorem 5.2 except the condition. However, the leftmost graph has no triangles but the rightmost one does, so the two graphs are not isomorphic.

6. Conclusion

This paper rigorously evaluates the efficiency of GNNs that leverage subgraph structures, particularly on graphs with bounded cycles, which represent many real-world datasets. In particular, we prove that k-hop subgraph GNNs can reliably predict properties of graphs without cycles of length greater than $2k+1$, which is unconditionally if $k=1$ and requires an additional assumption for $k \geqslant 2$. The theory is extended to k-hop GNNs without considering the subgraph structure for graphs without cycles of length greater than $2k-1$.

Let us also comment on the limitations of the current work. Firstly, it is unclear whether the k-separability in Theorem 4.2 can be removed or not. Secondly, though examples in Figure 4 illustrate that Theorem 5.2 cannot hold unconditionally, it remains unknown whether the k-strong separability can be weakened or what the weakest assumption is. Those directions deserve future research.

Acknowledgments

We thank the MIT PRIMES-USA program for making this paper possible. Z. Chen thanks Runzhong Wang for helpful discussions.

References

[1] Waiss Azizian and Marc Lelarge. Expressive power of invariant and equivariant graph neural networks. In *International Conference on Learning Representations*, 2021.

[2] Jacob Bamberger. A topological characterisation of weisfeiler-leman equivalence classes. In *Topological, Algebraic and Geometric Learning Workshops 2022*, pages 17–27. PMLR, 2022.

[3] Beatrice Bevilacqua, Fabrizio Frasca, Derek Lim, Balasubramaniam Srinivasan, Chen Cai, Gopinath Balamurugan, Michael M Bronstein, and Haggai Maron. Equivariant subgraph aggregation networks. *arXiv preprint arXiv:2110.02910*, 2021.

[4] Jin-Yi Cai, Martin Fürer, and Neil Immerman. An optimal lower bound on the number of variables for graph identification. *Combinatorica*, 12(4):389–410, 1992.

[5] Ziang Chen, Jialin Liu, Xinshang Wang, Jianfeng Lu, and Wotao Yin. On representing linear programs by graph neural networks. In *The Eleventh International Conference on Learning Representations*, 2023.

[6] Vijay Prakash Dwivedi, Chaitanya K Joshi, Anh Tuan Luu, Thomas Laurent, Yoshua Bengio, and Xavier Bresson. Benchmarking graph neural networks. *Journal of Machine Learning Research*, 24(43):1–48, 2023.

[7] Fabrizio Frasca, Beatrice Bevilacqua, Michael Bronstein, and Haggai Maron. Understanding and extending subgraph gnns by rethinking their symmetries. *Advances in Neural Information Processing Systems*, 35:31376–31390, 2022.

[8] Maxime Gasse, Didier Chételat, Nicola Ferroni, Laurent Charlin, and Andrea Lodi. Exact combinatorial optimization with graph convolutional neural networks. *Advances in neural information processing systems*, 32, 2019.

[9] Floris Geerts. The expressive power of kth-order invariant graph networks. *arXiv preprint arXiv:2007.12035*, 2020.

[10] Floris Geerts. Walk message passing neural networks and second-order graph neural networks. *arXiv preprint arXiv:2006.09499*, 2020.

[11] Floris Geerts and Juan L Reutter. Expressiveness and approximation properties of graph neural networks. In *International Conference on Learning Representations*, 2022.

[12] Justin Gilmer, Samuel S Schoenholz, Patrick F Riley, Oriol Vinyals, and George E Dahl. Neural message passing for quantum chemistry. In *International conference on machine learning*, pages 1263–1272. PMLR, 2017.

[13] Shiwen He, Shaowen Xiong, Yeyu Ou, Jian Zhang, Jiaheng Wang, Yongming Huang, and Yaoxue Zhang. An overview on the application of graph neural networks in wireless networks. *IEEE Open Journal of the Communications Society*, 2:2547–2565, 2021.

[14] Thomas N Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*, 2016.

[15] Mengyuan Lee, Guanding Yu, Huaiyu Dai, and Geoffrey Ye Li. Graph neural networks meet wireless communications: Motivation, applications, and future directions. *IEEE Wireless Communications*, 29(5):12–19, 2022.

[16] Wenlong Liao, Birgitte Bak-Jensen, Jayakrishnan Radhakrishna Pillai, Yuelong Wang, and Yusen Wang. A review of graph neural networks and their applications in power systems. *Journal of Modern Power Systems and Clean Energy*, 10(2):345–360, 2021.

[17] Haggai Maron, Heli Ben-Hamu, Hadar Serviansky, and Yaron Lipman. Provably powerful graph networks. *Advances in neural information processing systems*, 32, 2019.

[18] Christopher Morris, Gaurav Rattan, and Petra Mutzel. Weisfeiler and leman go sparse: Towards scalable higher-order graph embeddings. *Advances in Neural Information Processing Systems*, 33:21824–21840, 2020.

[19] 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 AAAI conference on artificial intelligence*, volume 33, pages 4602–4609, 2019.

[20] Franco Scarselli, Marco Gori, Ah Chung Tsoi, Markus Hagenbuchner, and Gabriele Monfardini. The graph neural network model. *IEEE transactions on neural networks*, 20(1):61–80, 2008.

[21] Jonathan Shlomi, Peter Battaglia, and Jean-Roch Vlimant. Graph neural networks in particle physics. *Machine Learning: Science and Technology*, 2(2):021001, 2020.

[22] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua Bengio. Graph attention networks. *arXiv preprint arXiv:1710.10903*, 2017.

[23] Jianian Wang, Sheng Zhang, Yanghua Xiao, and Rui Song. A review on graph neural network methods in financial applications. *arXiv preprint arXiv:2111.15367*, 2021.

[24] Boris Weisfeiler and Andrei Leman. The reduction of a graph to canonical form and the algebra which appears therein. *NTI, Series*, 2(9):12–16, 1968.

[25] Zonghan Wu, Shirui Pan, Fengwen Chen, Guodong Long, Chengqi Zhang, and S Yu Philip. A comprehensive survey on graph neural networks. *IEEE transactions on neural networks and learning systems*, 32(1):4–24, 2020.

[26] Keyulu Xu, Weihua Hu, Jure Leskovec, and Stefanie Jegelka. How powerful are graph neural networks? *arXiv preprint arXiv:1810.00826*, 2018.

[27] Bohang Zhang, Shengjie Luo, Liwei Wang, and Di He. Rethinking the expressive power of gnns via graph biconnectivity. *arXiv preprint arXiv:2301.09505*, 2023.

[28] Muhan Zhang and Pan Li. Nested graph neural networks. *Advances in Neural Information Processing Systems*, 34:15734–15747, 2021.

[29] Xiao-Meng Zhang, Li Liang, Lin Liu, and Ming-Jing Tang. Graph neural networks and their current applications in bioinformatics. *Frontiers in genetics*, 12:690049, 2021.

[30] Lingxiao Zhao, Wei Jin, Leman Akoglu, and Neil Shah. From stars to subgraphs: Uplifting any gnn with local structure awareness. *arXiv preprint arXiv:2110.03753*, 2021.

[31] Lingxiao Zhao, Neil Shah, and Leman Akoglu. A practical, progressively-expressive gnn. *Advances in Neural Information Processing Systems*, 35:34106–34120, 2022.

[32] Jie Zhou, Ganqu Cui, Shengding Hu, Zhengyan Zhang, Cheng Yang, Zhiyuan Liu, Lifeng Wang, Changcheng Li, and Maosong Sun. Graph neural networks: A review of methods and applications. *AI Open*, 1:57–81, 2020.

(ZC) Department of Mathematics, Massachusetts Institute of Technology, 77 Massachusetts Ave, Cambridge, MA 02139

Email address: ziang@mit.edu

(QZ) Sierra Canyon School, 20801 Rinaldi St, Chatsworth, CA 91311

Email address: tigerzhang343@gmail.com