Rethinking Graph Neural Networks for the Graph Coloring Problem

Wei Li*, Ruxuan Li†, Yuzhe Ma‡, Siu On Chan†, David Pan§, Bei Yu†
*Carnegie Mellon University †The Chinese University of Hong Kong
‡The Hong Kong University of Science and Technology (Guangzhou)
§The University of Texas at Austin

***Abstract*—Graph coloring, a classical NP-hard problem, is the problem of assigning connected nodes as different colors as possible. In this work, we aim to solve the coloring problem by graph neural networks (GNNs). However, we observe that state-of-the-art GNNs are less successful in the graph coloring problem. We analyze the reasons from two perspectives. First, most GNNs fail to generalize the task under homophily to heterophily, i.e., graphs where connected nodes are assigned different features. Second, GNNs are bounded by the network depth, making them possible to be a local method, which has been demonstrated to be non-optimal in Maximum Independent Set (MIS) problem. In this paper, we focus on the *aggregation-combine GNNs* (AC-GNNs), a popular class of GNNs. Instead of learning when AC-GNNs assign local equivalent node pairs to the same node embedding, which is designed for the task under homophily, we study the power of a GNN in the coloring problem by analyzing its ability to assign nodes different colors. Through the analysis, we find some specific settings/architectures that may harm the performance. Furthermore, we demonstrate the non-optimality of AC-GNNs due to their local property, and prove the positive correlation between model depth and its coloring power. Following the discussions above, we summarize a series of rules that make a GNN powerful in the coloring problem. Then, we propose a simple AC-GNN variation based on these rules. We empirically validate our theoretical findings and demonstrate that our simple model substantially outperforms state-of-the-art heuristic algorithms in both quality and runtime.**

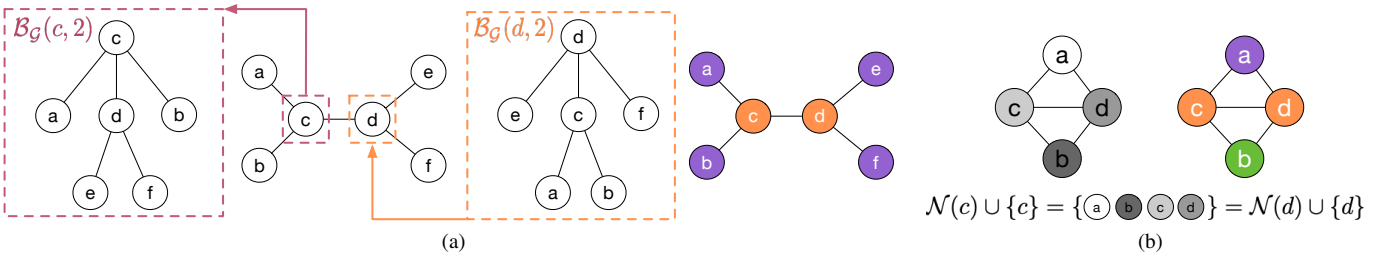
Introduction

Graph neural networks (GNNs) have shown overwhelming success in various fields, such as molecules, social networks, and web pages[1]. The main idea behind GNNs is a neighborhood aggregation scheme (or called message passing), where each node aggregates feature vectors from its neighbors and combines them with its own feature vector to produce a new one. GNNs following such a scheme are called aggregation-combination GNNs (AC-GNNs) [2]. After finite iterations of aggregation and combination, the corresponding feature vector of each node is called node embedding to represent the node.

In this work, we try to study the performance of GNNs for the coloring problem because 1) Graph coloring and its variations have a great demand in industry. while the target graph size is exploding nowadays. For example, the netlist graph in a commercial chip contains millions of gate nodes; the size of the user graph in the Internet company is also at a million level. These graphs are too large to be processed by transitional algorithms within an acceptable time, and therefore motivate us to apply the power of highly parallelable GNNs. 2) However, in our motivating experiment, most of existing GNNs even cannot beat the simplest heuristic algorithm. Therefore, we try to find the reason for the low performance so that we can provide some theoretical guidance on a powerful GNN for the coloring problem. 3) Graph problems under heterophily are the ones where connected nodes are expected to have *different* labels/features/colors instead of a similar one (homophily), and graph coloring problem is the most representative task under heterophily. Although some rules are shown to be able to enhance the expressive power of GNNs in previous work, whether these rules still hold under heterophily is still an open question. For example, in previous study under homophily, deeper GNNs often suffer from over-smooth problem (Zhao and Akoglu 2019) and it is believed that deepening GNNs do not improve (or sometimes worsen) their performance (Oono and Suzuki 2019). We will try to find the answer during our explorations in the coloring problem.

We study the problem by investigating the *power* of GNNs for the graph coloring problem. Some recent works [3], [2], [4], [5], [6], [7] study the power of a GNN by analyzing when a GNN maps two nodes to the same node embedding. In their study, a maximally powerful GNN with depth L should map two L-local equivalent nodes to the same node embedding [4], [3]. However, when applied in the graph coloring problem, such a definition raises some problems. First, the coloring task is not under homophily but *heterophily*. Therefore, two local equivalent nodes are not necessarily assigned the same node embedding. One example can be found in Figure 1(a), where the node pair $\{c, d\}$ is local equivalent but should be assigned different colors to avoid the conflict. Second, every AC-GNN is bounded by its depth L. Therefore, the maximally powerful GNN is identified by L-local equivalence instead of a *global* equivalence. This constraint makes an AC-GNN possible to be a *local method*, which has been demonstrated to be non-optimal in many NP-hard problems such as MIS [8], [9].

Motivated by these limitations, we define the power of AC-GNNs in the coloring problem as its ability to assign nodes different colors. We then observe and theoretically prove a set of conditions that may harm/contribute to the coloring performance. Based on these observations, we develop a series of rules to design powerful AC-GNNs *specifically for the graph coloring problem*. We make the following contributions: (1): We show that AC-GNNs cannot be optimal in the coloring problem and demonstrate the positive correlation between model depth and its power in the coloring problem. (2) We



(a) (b)

Fig. 1: Examples of graph structures in which some AC-GNNs fail to discriminate the equivalent node pair $\{c, d\}$. (a) left: the input graph with the same node attribute; right: the coloring results by the most powerful AC-GNN. (b) left: the input graph with different node attributes (represented by the gray scale); right: the coloring results by the most powerful *integrated* AC-GNN. The aggregation for any *integrated* AC-GNN in both c and d are the same since $\mathcal{N}(c) \cup \{c\} = \mathcal{N}(d) \cup \{d\}$. Here, the most powerful is an ideal object, and refers to a virtual integrated AC-GNN that has the most powerful ability to assign node pairs to different colors.

give simple but effective rules about the GNN architecture for the coloring problem, some of which are contrary to previous research under homophily. (3) Combining these rules, we develop a simple GNN-based approach by un-supervised learning. (4) We validate our findings by extensive empirical evaluation including three datasets from different subjects. Our method shows substantially superior performance compared with other existing AC-GNN variations and even outperforms state-of-the-art heuristic algorithms with a significant efficiency improvement.

PRELIMINARIES

Graph Terminology: Here we list the following graph theoretic terms encountered in our work. Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ and $\mathcal{G}' = (\mathcal{V}', \mathcal{E}')$ be graphs on vertex set $\mathcal{V}$ and $\mathcal{V}'$, we define

- *isomorphism*: we say that a bijection $\pi : \mathcal{V} \to \mathcal{V}'$ is an *isomorphism* if any two vertices $u, v \in \mathcal{V}$ are adjacent in $\mathcal{G}$ if and only if $\pi(u), \pi(v) \in \mathcal{V}'$ are adjacent in $\mathcal{G}'$, i.e., $\{u, v\} \in \mathcal{E}$ iff $\{\pi(u), \pi(v)\} \in \mathcal{E}'$.
- *isomorphic nodes*: If there exists the isomorphism between $\mathcal{G}$ and $\mathcal{G}'$, we say that $\mathcal{G}$ and $\mathcal{G}'$ are *isomorphic*.
- *automorphism*: When π is an isomorphism of a vertex set onto itself, i.e., $\mathcal{V} = \mathcal{V}'$, π is called an *automorphism* of $\mathcal{G}$.
- *topologically equivalent*: We say that the node pair $\{u, v\}$ is *topologically equivalent* if there is an automorphism mapping one to the other, i.e., $v = \pi(u)$.
- *equivalent*: $\{u, v\}$ is *equivalent* if it is topologically equivalent by π and $\boldsymbol{x}_w = \boldsymbol{x}_{\pi(w)}$ holds for every $w \in \mathcal{V}$, where $\boldsymbol{x}_w$ is the node attribute of node w.
- *r-local topologically equivalent*: The node pair $\{u, v\}$ is *r-local topologically equivalent* if π_r is an isomorphism from $\mathcal{B}_{\mathcal{G}}(u, r)$ to $\mathcal{B}_{\mathcal{G}}(v, r)$.
- *r-local equivalent*: $\{u, v\}$ is *r-local equivalent* if it is r-local topologically equivalent by π_r and $\boldsymbol{x}_w = \boldsymbol{x}_{\pi_r(w)}$ holds for every $w \in \mathcal{B}_{\mathcal{G}}(u, r)$.
- *r-local isomorphism*: A bijection π_r is an *r-local isomorphism* that maps u to v if π_r is an isomorphism that maps $\mathcal{B}_{\mathcal{G}}(u, r)$ to $\mathcal{B}_{\mathcal{G}}(v, r)$.
- *other local graph terminology*: For every positive integer r and every node $u \in \mathcal{V}$, we define $\mathcal{B}_{\mathcal{G}}(u, r)$ as the subgraph of $\mathcal{G}$ induced by node u with distance at most r from u.

One example is given in Figure 1(a).

Graph coloring.: Let k be the number of available colors, $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be the input graph and each vertex $v \in \mathcal{V}$ be associated with an attribute $\boldsymbol{x}_v$, a coloring function $f_k : (v, \mathcal{G}, \boldsymbol{x}_v) \to \{1, ..., k\}$ returns a color of v indexed by $col_v \in \{1, ..., k\}$. In the following pages, k follows the same definition if not specified and $f(\mathcal{G})$ represents the coloring solution on $\mathcal{G}$ by f for simplification. Given a graph $\mathcal{G}$ colored by f_k, a *conflict function* $c : (u, v, f_k) \to \{0, 1\}$ is used to measure the performance of f_k on $\mathcal{G}$. Specifically, $c(u, v, f_k) = 1$ when u and v are connected and assigned the same color:

$$c(u, v, f_k) = \begin{cases} 1, & \text{if } f_k(v, \mathcal{G}, \boldsymbol{x}_v) = f_k(u, \mathcal{G}, \boldsymbol{x}_u) \\ & \text{and } \{u, v\} \in \mathcal{E}; \\ 0, & \text{otherwise.} \end{cases} \tag{1}$$

The edge $e = \{u, v\}$ is called a *conflict* if $c(u, v, f) = 1$. The objective of the graph coloring problem is widely formulated in two ways: 1) *k-coloring problem*: Given k, minimize the number of conflicts as in Equation (2); 2) Given a conflict constraint c_{max}, minimize the number of used colors as in Equation (3).

$$\min \sum_{\{u,v\} \in \mathcal{E}} c(u, v, f_k). \tag{2}$$

$$\min k, \text{ s.t.} \sum_{\{u,v\} \in \mathcal{E}} c(u, v, f_k) \leq c_{max}. \tag{3}$$

When we set c_{max} as 0, i.e., no conflict is introduced by f_k, we refer the obtained minimum color number as the *chromatic number* of $\mathcal{G}$, which is often represented by $\mathcal{X}$. The corresponding coloring function $f_{\mathcal{X}}$ is called an *optimal* function.

Graph neural networks (GNNs).: GNNs are to learn the node embeddings or graph embedding based on the graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ and node features $\{\boldsymbol{x}_v : v \in \mathcal{V}\}$. We follow the same notations in [2] to formally define the basics for GNNs. Let $\{\text{AGG}^{(i)}\}_{i=1}^{L}$ and $\{\text{COM}^{(i)}\}_{i=1}^{L}$ be two sets of *aggregation* and *combination* functions. An *aggregation-combine GNN*

(AC-GNN) computes the feature vectors $\boldsymbol{h}_v^{(i)}$ for every node $v \in \mathcal{V}$ by:

$$\boldsymbol{h}_v^{(i)} = \text{COM}^{(i)}(\boldsymbol{h}_v^{(i-1)}, \text{AGG}^{(i)}(\{\boldsymbol{h}_u^{(i-1)} : u \in \mathcal{N}(v)\})), \quad (4)$$

where $\mathcal{N}(v)$ denotes the neighborhood of v, i.e., $\mathcal{N}(v) = \{u : \{u, v\} \in \mathcal{E}\}$ and $\boldsymbol{h}_v^{(0)}$ is the node attribute $\boldsymbol{x}_v$. Finally, each node v is classified by a node classification $CLS(\cdot)$ applied to the node embedding $\boldsymbol{h}_v^{(L)}$. When the AC-GNN is used for the graph coloring problem, $CLS(\cdot)$ returns a $col_v \in \{1, ..., k\}$. Then, an AC-GNN $\mathcal{A}$ with L layers is also called L-AC-GNN and defined as $\mathcal{A} = (\{\text{AGG}^{(i)}\}_{i=1}^{L}, \{\text{COM}^{(i)}\}_{i=1}^{L}, CLS(\cdot))$. Here, we define $\mathcal{A}(v, \mathcal{G}, \boldsymbol{x}_v)$ as the color of v assigned by $\mathcal{A}$.

simple AC-GNN: The properties of aggregation, combination and classification functions are widely studied and many variations of these functions are proposed. Among various function architectures, we say an AC-GNN is *simple* if the aggregation and combination functions are defined as follows:

$$\text{AGG}^{(i)}(\mathbb{X}) = \sum_{\boldsymbol{x} \in \mathbb{X}} \boldsymbol{x}, \quad (5)$$

$$\text{COM}^{(i)}(\boldsymbol{x}, \boldsymbol{y}) = \sigma(\boldsymbol{x}\boldsymbol{C}^{(i)} + \boldsymbol{y}\boldsymbol{A}^{(i)} + \boldsymbol{b}^{(i)}), \quad (6)$$

where $\boldsymbol{C}^{(i)}$, $\boldsymbol{A}^{(i)}$, and $\boldsymbol{b}^{(i)}$ are trainable parameters, σ is an activation function.

integrated AC-GNN: The aggregation and combination functions can also be integrated such as networks explored in [10], [3]. We say that such AC-GNN is *integrated* when aggregation and combination functions are integrated as follows:

$$\boldsymbol{h}_u^{(i)} = \text{COM}^{(i)}(\text{AGG}^{(i)}(\{\boldsymbol{h}_w^{(i-1)} : w \in \mathcal{N}(u) \cup \{u\}\})). \quad (7)$$

In integrated AC-GNNs, aggregation functions aggregate features from neighborhood and the node itself simultaneously, which means they treat the neighborhood information and ego-information (information from the node itself) equally.

Powerful GNNs for Graph Coloring

In this section, we focus on the question: *What kinds of designs make a GNN more/less powerful in the graph coloring problem?* Although GNNs demonstrate their power in various tasks, most of them even cannot beat the simplest heuristic algorithms in the coloring problem. One motivating experiment is shown in Table I, where the solved ratio is one minus the ratio between the number of conflicts and the number of edges. For example, given a graph with 100 edges and colored with 10 conflicts, then the solved ratio is calculated by $1 - 10/100 = 0.9$. The Greedy method colors nodes in the order of node IDs. We observe that all tested GNNs do not work in the coloring problem. We analyze the reason from the perspective of heterophily and homophily: That is, linked nodes should be assigned different colors rather than the same one. However, previous studies define the power of GNNs as the capability to map two equivalent nodes to the same embedding. Under the heterophily, it is critical to rethink the definition of a GNN's power *specifically for the coloring problem*. After the power is formalized, the next question is: *What factors may enhance or harm such a power?* Here, we discuss the power of GNNs for graph coloring by answering the questions raised above. We leave all proofs in Appendix due to the page limit.

TABLE I: Solved ratio of existing GNNs (GCN [10], SAGE [11], GIN [4], GAT [12]) and the simplest greedy algorithm on layout dataset over three runs. The node attributes are set to all-one vectors. d: depth.

	d = 2	d= 10
GCN	0.55 ±0.01	0 ±0
SAGE	0 ±0	0 ±0
GIN	0.59 ±0.01	0.58 ±0.01
GAT	0 ±0	0 ±0
Greedy	**0.962**	

Discrimination power under heterophily

Q: *How to determine whether a GNN is powerful in the coloring problem*?

In the graph coloring problem, connected nodes are assigned to different colors. Therefore, a powerful GNN should map the two connected nodes to node embeddings as differently as possible. Intuitively, we can study the power of a GNN in the coloring problem by analyzing ***its ability to assign nodes different colors.*** Here, we refer to the power as the *discrimination power* of GNNs to differ from previous expressive power under homophily. Formally, we define that a coloring method f discriminates a node pair (u, v) as follows:

Definition 1 (discriminate)**.** *A coloring method f discriminates a node pair (u, v) if f assigns u and v different colors, i.e., $f(u, \mathcal{G}, \boldsymbol{x}_u) \neq f(v, \mathcal{G}, \boldsymbol{x}_v)$.*

Following the definitions, we can answer the question above: *the more powerful a GNN is, the more node pairs it will discriminate*. In the following context, we consider the case where the number of colors k is the chromatic number $\mathcal{X}$. Ideally, an optimal GNN should be able to discriminate all node pairs.

Given the definition above, one may try to build an optimal AC-GNN which colors all graphs without conflict through discriminating all node pairs:

Q: *Can we design an AC-GNN which discriminates any node pair?*

The following Property 1 refutes the existence of such a "perfect" AC-GNN:

Property 1. *All AC-GNNs cannot discriminate any equivalent node pair.*

One example is given in Figure 1(a), where node pair $\{c, d\}$ is equivalent. Since the equivalent node pair have the same subgraph structure with the same node attribute distributions, the AC-GNN always return the same results in each layer. Hence, AC-GNNs are not optimal for any graph that contains these node pairs, i.e., connected and also equivalent pairs.

To avoid such a non-optimal case, we can break the equivalence between two nodes by assigning different node attributes such as random features [13] or one-hot vectors [2]. The solution also aligns with the conclusion in [3], which proves

that with different attributes GNNs become significantly more powerful. Indeed, making nodes different purposely strengthens the AC-GNN by eliminating the equivalent node pairs (although the topological equivalence is preserved). However, the superficial methods on the node attributes cannot influence and solve the underlying defects of some specific AC-GNNs, for example, integrated AC-GNN. In an integrated AC-GNN, the node and its neighbors are aggregated into the same multi-set, making it more difficult for integrated AC-GNNs to discriminate the node with its neighbors. Property 2 points out that an integrated AC-GNN cannot be optimal since there always exists a set of graphs in which at least one node pair is not discriminated by the integrated AC-GNN:

Property 2. *If nodes u and v in a graph $\mathcal{G}$ are connected and share the same neighborhood except each other, i.e., $\mathcal{N}(u)\backslash\{v\} = \mathcal{N}(v)\backslash\{u\}$, then an integrated AC-GNN cannot discriminate $\{u, v\}$.*

The property points to the deficiency of integrated AC-GNN even if nodes are differentiable by their attributes. One example is given in Figure 1(b), where the node pair $\{c, d\}$ cannot be discriminated by any integrated AC-GNN even the node attributes are different.

Locality

Local methods are widely used in the combinational optimization problems such as maximum independent set (MIS) and graph coloring. The formal definition of the local method for the coloring problem is described as follows, which is a direct rephrasing in [9]:

Definition 2 (local method [9])**.** *A coloring method f is r-local if it fails to discriminate any r-local equivalent node pair. A coloring method f is local if f is r-local for at least one positive integer r.*

Along with the study of the local methods, the upper bound of a local method for the MIS problem is investigated. David et al. [9] gives an upper bound $1/2 + 1/(2\sqrt{2})$ of an MIS produced by any local method in the random d-regular graph as $d \to \infty$ and [8] strengthens the bound to $1/2$. A random d-regular graph is a graph with n nodes and the nodes in each node pair are connected with a probability d/n. Starting from the upper bound of any local method for the MIS problem, we may try to figure out:

Q: *Whether a local method is also non-optimal in the graph coloring problem?*

The answer is given by following corollary:

Corollary 1. *A local coloring method is non-optimal in the random d-regular tree as $d \to \infty$.*

We finish the proof by making use of the upper bound studied in the MIS problem and bridging the connection between a local method for MIS problem and coloring problem. Due to the localized nature of the aggregation function in GNNs, an AC-GNN with a fixed number of layers, say L layers, cannot detect the structure or information of nodes at a distance further than L. Considering the non-optimality of a local coloring method stated in Corollary 1 and the localized nature of GNNs, we can reduce our analysis of whether there exists an optimal AC-GNN for graph coloring to whether an AC-GNN is a local coloring method? Corollary 2 answers the question as yes:

Corollary 2. *L-AC-GNN is an L-local coloring method and thus a local coloring method.*

Corollary 1 and Corollary 2 directly lead to the following theorem:

Theorem 1. *AC-GNN is not optimal, specifically for the random d-regular tree as $d \to \infty$.*

Based on the analysis above, we can see that the locality of AC-GNN makes it infeasible to to be an optimal coloring function. To solve the problems raised by locality of AC-GNNs, which inhibits AC-GNNs from detecting the global graph structure, many efforts have been made to devise a global scheme such as global readout functions [2], randomness [14], [15] and deeper networks [16], [17], [18]. Among all global techniques, a deep architecture is believed to be global as long as it covers the full graph. Given a graph with diameter R, a R-AC-GNN is able to detect the information from the whole graph. However, it is impossible to find an AC-GNN which is able to cover all graphs: any AC-GNN is always bounded by its depth. Then, if we cannot develop an optimal AC-GNN by simply stacking layers, does this method contribute to the discrimination power? Formally:

Q: *Is deeper AC-GNN more powerful in the coloring problem?*

We answer the question as yes, and give a more specific statement:

Property 3. *Let $\{u, v\}$ be a node pair in any graph $\mathcal{G}$, and L be any positive integer. If a L-AC-GNN discriminates $\{u, v\}$, a L^+-AC-GNN also discriminates it.*

A L^+-AC-GNN is an AC-GNN by stacking injective layers after L-AC-GNN (before $CLS(\cdot)$). An injective layer includes a pair of injective aggregation function and injective combination function.

Our Method

Based on the discussions above, we summarize a series of rules that make a GNN $\mathcal{A}$ powerful in the coloring problem as follows:

1) The input graph contains no equivalent node pair (Property 1);
2) $\mathcal{A}$ does not integrate the aggregation and combination function (Property 2);
3) $\mathcal{A}$ should be as deep as possible (Property 3);
4) Layers in $\mathcal{A}$ should be injective (Property 3);

With the guidance of the rules above, we propose a very simple architecture, *Graph Discrimination Network (GDN)* based on simple AC-GNN. Note that there is not solely one architecture that satisfies all the rules above. We select GDN as an example considering the balance between efficiency and performance. We describe GDN as follows:

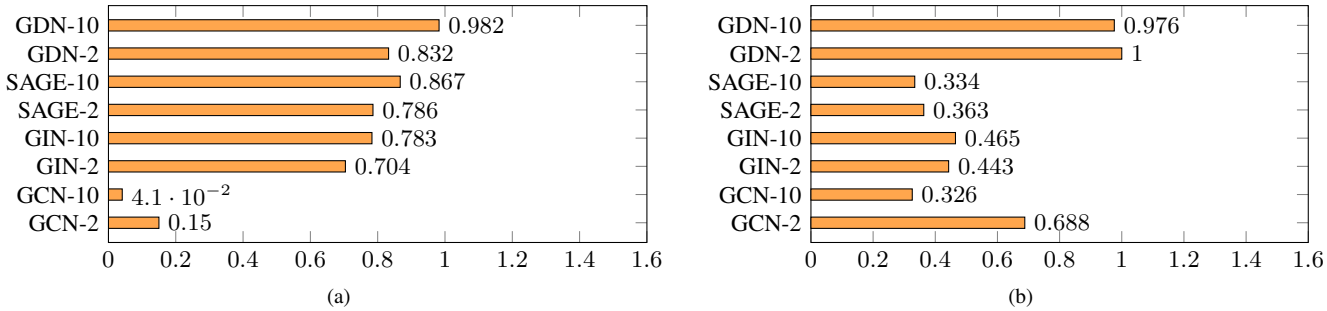


Fig. 2: (a) Solved ratio by different AC-GNN variations; (b) Fixed color ratio by different AC-GNN variations.

Forward Computation.: For a k-coloring problem, the node attribute is the centered probability distribution of k colors and is initialized randomly to eliminate the equivalent node pairs (rule 1).

The aggregation function is the same as Equation 5 (rule 2,4). Let $\boldsymbol{m}_v^{(i)} \in \mathbb{R}^k$ be the result returned by $\text{AGG}^{(i)}$ for the node v in the i-th layer, the aggregation layer is organized as follows:

$$\boldsymbol{m}_v^{(i)} = \sum_{u \in \mathcal{N}(v)} \boldsymbol{h}_u^{(i-1)}. \tag{8}$$

In the combination function, we define the $\text{COM}^{(i)}$ as follows to make GDN color equivariant. For the details of color equivariance, please refer to the Appendix.

$$\begin{aligned} \boldsymbol{h}_v^{(i)} = \boldsymbol{h}_v^{(i-1)}(\lambda_C^{(i)} + \gamma_C^{(i)}(\mathbf{1}\mathbf{1}^\top)) \\ + \boldsymbol{m}_v^{(i)}(\lambda_A^{(i)} + \gamma_A^{(i)}(\mathbf{1}\mathbf{1}^\top)) + \beta^{(i)}\mathbf{1}, \end{aligned} \tag{9}$$

where λ, γ, β are trainable scalars. Finally, the classification function $CLS(\cdot))$ in GDN is defined as an argmax function, since the final node embedding is still a probability distribution of colors.

Loss Function.: Considering that the permutation of colors cannot influence the result quality, it is not an easy job to develop a supervised training scheme since there are multiple optimal solutions. Here, we use an un-supervised margin loss, motivated by the fact that the final node embeddings of connected nodes should be as different as possible, and formulated by:

$$\min \sum_{\{u,v\} \in \mathcal{E}} \max\{m - d(\boldsymbol{h}_u, \boldsymbol{h}_v), 0\}, \tag{10}$$

where $\boldsymbol{h}_u \in \mathbb{R}^k$ is the probability distribution obtained by $\mathcal{A}$. d is the Euclidean distance between the node pair. m is the pre-defined margin.

Preprocess & Postprocess.: Our method also contains preprocess and postprocess procedures, which are widely used in other coloring methods [19], [20]. In the preprocess part, the node with a degree less than k is removed iteratively. In the postprocess part, we iteratively detect 1) whether a color change in a single node will decrease the cost or 2) whether a swap of colors between connected nodes will decrease the cost. We implement the two additional steps by tensor operations, which significantly boost the efficiency. The experiments on the two steps and detailed algorithms are shown in the Appendix D.

Combining with Other methods.: As stated in Theorem 1, any AC-GNN cannot be optimal in the coloring problem. Therefore, we may combine with other optimal methods to obtain better quality in a sacrifice of efficiency. Here, we propose a simple combination with ILP-based method. Specifically, after we obtain the color distribution of each node by GDN, if there is conflict(s) after $CLS(\cdot)$, we can set up a threshold for the final color distributions to get a partial coloring result. The partial result is then passed to a ILP solver to obtain the final result. The details can be found in the Appendix.

EXPERIMENTS

Experimental setup

Detailed settings, dataset introduction, more experiments and analysis are shown in the Appendix. We evaluate our models and baselines on three datasets here, the basic information on these datasets are shown in Table II, where column $\mathcal{X}(k)$ is the chromatic number except layout dataset, which is set to 3 in the real-world circuit design.

We mainly compare our models with three previous works (the results of other methods such as ILP and simple heuristics can be found in the Appendix): (1) GNN-GCP [21], combing GNN, RNN, and MLP to obtain the node embedding and using a k-means method to color the node. We obtain models from the author and directly obtain the results. (2) Tabucol [22], a well-known heuristic algorithm using Tabu search. We follow the original setting with an iteration limit of 1000 (or the time limit of 24 hours) and the number of uncolored node pairs is returned if the algorithm fails to find a perfect coloring assignment within the limit. (3) HybridEA [23], the state-of-the-art evolutionary algorithm for the coloring problem. We also compare different variants of AC-GNN in the previous works: GCN [10], GIN [4], and GraphSAGE [11]. All AC-GNN variations are only tested in the layout dataset since AC-GNN variations require a fixed number of colors to make output shape keep the same. To make the comparison fair, we use the same CPU to test all methods.

Comparison with other AC-GNN variations

The results are shown in Figure 2(a). "GDN-k" represents GDN with a depth of k. According to the results, we observe

TABLE II: Graph datasets information and results by different coloring methods.

Dataset	Graph	$\|\mathcal{V}\|$	$\|\mathcal{E}\|$	$d\%$	$\mathcal{X}(k)$	GNN-GCP		Tabucol		HybridEA		GDN	
						Cost	Time	Cost	Time	Cost	Time	Cost	Time
Layout	35158	641202	787242	-	3	386009	3896	2392	82301	1562	133285	**1557**$\pm$45	5.84 $\pm$0.23
ratio						247.9	667.1	1.54	14092	1.00	22822	**1.0**	**1.0**
	Cora	2708	5429	0.15	5	1291	3.90	31	15410	0	18921	**0** $\pm$0	0.81 $\pm$0.08
Citation	Citeseer	3327	4732	0.09	6	1733	2.74	6	44700	0	24230	**0** $\pm$0	1.42 $\pm$0.15
	Pubmed	19717	44338	0.03	8	4393	4.50	-	>24h	-	>24h	**21** $\pm$4	1.41 $\pm$0.12
ratio						353.2	3.06	-	-	-	-	**1.0**	**1.0**
	jean	80	254	8	10	76	0.06	**0**	0.95	**0**	0.01	**0** $\pm$1	0.13 $\pm$0.02
	anna	138	493	5	11	87	0.08	**0**	3.23	**0**	0.02	**0** $\pm$0	0.17 $\pm$0.03
	huck	74	301	11	11	117	0.05	**0**	0.15	**0**	0.01	**0** $\pm$0	0.06 $\pm$0.02
	david	87	406	11	11	-	-	**0**	4.83	**0**	0.01	**0** $\pm$0	0.19 $\pm$0.01
	homer	561	1628	1	13	1628	1.09	**0**	274	**0**	0.06	**0** $\pm$1	0.29 $\pm$0.02
	myciel5	47	236	22	6	35	0.04	**0**	0.20	**0**	0.01	**0** $\pm$0	0.12 $\pm$0.01
	myciel6	95	755	17	7	94	4.33	**0**	0.79	**0**	0.01	**0** $\pm$0	0.21 $\pm$0.02
	games120	120	638	9	9	301	0.07	**0**	0.93	**0**	0.01	**0** $\pm$1	0.08 $\pm$0.01
COLOR	Mug88_1	88	146	4	3	146	0.33	**0**	0.12	**0**	0.01	**0** $\pm$0	0.01 $\pm$0
	1-Insertions_4	67	232	10	2	42	0.05	**0**	0.16	**0**	0.01	**0** $\pm$0	0.07 $\pm$0
	2-Insertions_4	212	1621	7	4	360	0.09	**1**	255	**1**	101.1	**1** $\pm$0	60.08 $\pm$0.01
	Queen5_5	25	160	53	5	37	0.03	**0**	0.13	**0**	0.07	**0** $\pm$0	0.05 $\pm$0.01
	Queen6_6	36	290	46	6	290	0.38	**0**	4.93	**0**	1.89	**0** $\pm$0	0.08 $\pm$0
	Queen7_7	49	476	40	7	126	0.04	10	36.9	**9**	51.8	**9** $\pm$1	0.38 $\pm$0.08
	Queen8_8	64	728	36	8	188	0.05	8	61.3	5	74.1	**2** $\pm$0	0.13 $\pm$0.03
	Queen9_9	81	1056	33	9	296	0.07	**5**	97.8	6	126.9	6 $\pm$1	0.09 $\pm$0.01
	Queen8_12	96	1368	30	12	260	0.10	10	139	3	92.9	**0** $\pm$0	0.58 $\pm$0.09
	Queen11_11	121	3960	55	11	396	0.10	33	213	22	141.3	**21** $\pm$3	0.07 $\pm$0.01
	Queen13_13	169	6656	47	13	728	0.20	42	401	37	213	**33** $\pm$2	2.38 $\pm$0.32
ratio						-	-	1.51	22.63	1.15	12.10	**1.0**	**1.0**

the following: (1) GCN, the most representative integrated AC-GNN, is much worse than other AC-GNNs, which demonstrates our rule 1. (2) Most AC-GNNs benefit from a deeper network, which aligns with our rule 3. (3) Although some non-integrated AC-GNN achieve an acceptable solved ratio, our method is still far better than other AC-GNNs.

We also validate the color equivariance of models by simulating the pre-color constraint in the layout decomposition problem. For each instance, we randomly select one node and set its color by changing the node attribute, known as the color distribution. We measure the color equivariant capability by checking the fixed color ratio, defined as the ratio between the number of successfully fixed graphs and the number of total graphs. A successfully fixed graph is the graph whose selected node is colored as expected with the pre-assigned one . From the results shown in Figure 2(b), we can see that the fixed color ratio of our GDN is much higher than other variations, matching with our analysis.

Comparison with other graph coloring methods

The comparison with other graph coloring methods is conducted on all collected datasets. We combine with ILP for the complex dataset, i.e., the COLOR dataset. The results are shown in Table II, where k is the number of available colors and cost is the number of conflicts in the coloring result. GNN-GCP gives "-" if it fails to find a chromatic number prediction. Tabucol and HybridEA give "-" if they fails to color the graph within 24 hours. According to the results, we observe the following: (1) Our method outperforms the state-of-the-art algorithms with a better result quality and 10$\times$ speedup. (2) Our method is more advantageous for complex and large graphs (Citation and Queen), which are more beneficial for industrial demand.

CONCLUSION

In this paper, we established theoretical foundations for reasoning about the discrimination power of GNNs for the graph coloring problem. We identified the node pairs that a popular class of GNNs, AC-GNNs fail to discriminate and gave conditions on how an AC-GNN can be more discriminatively powerful. Moreover, we built the connection between the locality study in graph theory and the local property of AC-GNNs, and proved the non-optimality of AC-GNN due to the locality. Furthermore, we analyzed the color equivariance in the graph coloring problem and proposed a scheme to make AC-GNN color equivariant. Combining all the analysis above, we designed a simple variation of AC-GNN for the graph coloring problem, which proves to be discriminatively powerful and color equivariant. To complete the picture, it would be interesting to analyze the global terms for enhancing the discrimination power of GNNs.

REFERENCES

[1] W. L. Hamilton, R. Ying, and J. Leskovec, "Representation learning on graphs: Methods and applications," *arXiv preprint arXiv:1709.05584*, 2017.

[2] P. Barceló, E. V. Kostylev, M. Monet, J. Pérez, J. Reutter, and J. P. Silva, "The logical expressiveness of graph neural networks," in *International Conference on Learning Representations (ICLR)*, 2019.

[3] A. Loukas, "What graph neural networks cannot learn: depth vs width," *arXiv preprint arXiv:1907.03199*, 2019.

[4] K. Xu, W. Hu, J. Leskovec, and S. Jegelka, "How powerful are graph neural networks?" *arXiv preprint arXiv:1810.00826*, 2018.

[5] F. Gama, J. Bruna, and A. Ribeiro, "Stability properties of graph neural networks," *IEEE Transactions on Signal Processing*, vol. 68, pp. 5680–5695, 2020.

[6] A. Loukas, "How hard is to distinguish graphs with graph neural networks?" Tech. Rep., 2020.

[7] S. S. Du, K. Hou, B. Póczos, R. Salakhutdinov, R. Wang, and K. Xu, "Graph neural tangent kernel: Fusing graph neural networks with graph kernels," *arXiv preprint arXiv:1905.13192*, 2019.

[8] M. Rahman, B. Virag *et al.*, "Local algorithms for independent sets are half-optimal," *The Annals of Probability*, vol. 45, no. 3, pp. 1543–1577, 2017.

[9] D. Gamarnik and M. Sudan, "Limits of local algorithms over sparse random graphs," in *Proceedings on Innovations in theoretical computer science*, 2014, pp. 369–376.

[10] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," *arXiv preprint arXiv:1609.02907*, 2016.

[11] W. Hamilton, Z. Ying, and J. Leskovec, "Inductive representation learning on large graphs," in *Annual Conference on Neural Information Processing Systems (NIPS)*, 2017, pp. 1024–1034.

[12] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Lio, and Y. Bengio, "Graph attention networks," *arXiv preprint arXiv:1710.10903*, 2017.

[13] R. Sato, M. Yamada, and H. Kashima, "Random features strengthen graph neural networks," *arXiv preprint arXiv:2002.03155*, 2020.

[14] J. You, R. Ying, and J. Leskovec, "Position-aware graph neural networks," *arXiv preprint arXiv:1906.04817*, 2019.

[15] G. Dasoulas, L. D. Santos, K. Scaman, and A. Virmaux, "Coloring graph neural networks for node disambiguation," *arXiv preprint arXiv:1912.06058*, 2019.

[16] G. Li, M. Muller, A. Thabet, and B. Ghanem, "DeepGCNs: Can GCNs go as deep as CNNs?" in *IEEE International Conference on Computer Vision (ICCV)*, 2019, pp. 9267–9276.

[17] M. Chen, Z. Wei, Z. Huang, B. Ding, and Y. Li, "Simple and deep graph convolutional networks," in *International Conference on Machine Learning*. PMLR, 2020, pp. 1725–1735.

[18] K. Xu, C. Li, Y. Tian, T. Sonobe, K.-i. Kawarabayashi, and S. Jegelka, "Representation learning on graphs with jumping knowledge networks," in *International Conference on Machine Learning*. PMLR, 2018, pp. 5453–5462.

[19] Y. Zhou, J.-K. Hao, and B. Duval, "Reinforcement learning based local search for grouping problems: A case study on graph coloring," *Expert Systems with Applications*, vol. 64, pp. 412–422, 2016.

[20] W. Li, J. Xia, Y. Ma, J. Li, Y. Liny, and B. Yu, "Adaptive layout decomposition with graph embedding neural networks," in *2020 57th ACM/IEEE Design Automation Conference (DAC)*. IEEE, 2020, pp. 1–6.

[21] H. Lemos, M. Prates, P. Avelar, and L. Lamb, "Graph colouring meets deep learning: Effective graph neural network models for combinatorial problems," in *IEEE International Conference on Tools with Artificial Intelligence (ICTAI)*. IEEE, 2019, pp. 879–885.

[22] A. Hertz and D. de Werra, "Using tabu search techniques for graph coloring," *Computing*, vol. 39, no. 4, pp. 345–351, 1987.

[23] P. Galinier and J.-K. Hao, "Hybrid evolutionary algorithms for graph coloring," *Journal of combinatorial optimization*, vol. 3, no. 4, pp. 379–397, 1999.