

# FEATURE AUGMENTATION OF GNNs FOR ILPs: LOCAL UNIQUENESS SUFFICES

**Anonymous authors**

Paper under double-blind review

## ABSTRACT

Integer Linear Programs (ILPs) are central to real-world optimizations but notoriously difficult to solve. Learning to Optimize (L2O) has emerged as a promising paradigm, with Graph Neural Networks (GNNs) serving as the standard backbone. However, standard anonymous GNNs are limited in expressiveness for ILPs, and the common enhancement of augmenting nodes with globally unique identifiers (UIDs) typically introduces spurious correlations that severely harm generalization. To address this tradeoff, we propose a parsimonious Local-UID scheme based on  $d$ -hop uniqueness coloring, which ensures identifiers are unique only within each node’s  $d$ -hop neighborhood. Building on this scheme, we introduce ColorGNN, which incorporates color information via color-conditioned embeddings, and ColorUID, a lightweight feature-level variant. We prove that for  $d$ -layer networks, Local-UIDs achieve the expressive power of Global-UIDs while offering stronger generalization. Extensive experiments show that our approach (i) yields substantial gains on three ILP benchmarks, (ii) exhibits strong OOD generalization on linear programming datasets, and (iii) further improves a general graph-level task when paired with a state-of-the-art method.

## 1 INTRODUCTION

Integer linear programs (ILPs) are important optimization problems with linear objectives and linear constraints, where variables are constrained to be integers. ILPs arise widely in engineering and operations research—including vehicle routing (Ray et al., 2014; Schöbel, 2001), scheduling (Ku & Beck, 2016; Meng et al., 2020), and facility location (Holmberg et al., 1999; Melkote & Daskin, 2001)—yet remain computationally challenging. Traditional methods, most notably Branch and Bound (Boyd & Mattingley, 2007), exhibit exponential worst-case complexity.

Learning to Optimize (L2O) has recently gained attention as a data-driven paradigm for addressing challenging optimization tasks. L2O methods learn optimization strategies from existing problem instances to improve solution efficiency and quality (Kruber et al., 2017; Khalil et al., 2022; Paulus et al., 2022; Li et al., 2024). For ILPs, Graph Neural Networks (GNNs) have emerged as a natural choice of architecture (Gasse et al., 2019; Nair et al., 2020; Han et al., 2023; Huang et al., 2024; Liu et al., 2025), since an ILP can be naturally represented as a bipartite graph: variable nodes on one side, constraint nodes on the other, with edges indicating variable participation in constraints.

However, standard GNNs have limited expressive power to solve ILPs (Chen et al., 2023a; 2025). A common way to overcome this limitation is to enrich the graph with additional distinguishing features, typically in the form of unique identifiers (UIDs) for nodes. Several feature augmentation strategies have been explored for ILPs, including random features (Chen et al., 2023a) and positional embeddings (Han et al., 2023; Chen et al., 2024). These strategies can be viewed as instances of a Global-UID scheme, where each node is assigned a completely distinct identifier. In addition, to address variable symmetries, Chen et al. (2025) proposed an orbit-based augmentation method that assigns distinct UIDs to variables within the same symmetry orbit.

While feature augmentation can enhance expressiveness, it often comes at the cost of generalization, thereby requiring more training samples. Specifically, injecting extrinsic features or identifiers introduces irrelevant signals that increase the risk of overfitting: the model may rely on IDs as shortcut cues, memorizing training-specific patterns rather than learning structural regularities, thereby

spuriously associating IDs with the target labels (Keriven et al., 2021; Bechler-Speicher et al., 2024; 2025). To alleviate this issue, augmentation should remain as parsimonious as possible. This motivates the following design principle:

*Feature augmentation should be parsimonious while retaining sufficient expressive power.*

## 1.1 OUR CONTRIBUTION

In this paper, we introduce a simple yet effective feature augmentation scheme, termed *Local-UIDs*, designed to be parsimonious without compromising expressive power. Our approach is motivated by two key observations:

- **Observation 1:** Empirical GNNs for ILPs are typically *shallow* (no more than four layers) (Gasse et al., 2019; Han et al., 2023; Chen et al., 2024; Huang et al., 2024; Chen et al., 2025; Liu et al., 2025), partly due to over-smoothing and over-squashing issues.
- **Observation 2:** In a  $d$ -layer GNN, the receptive field of each node is restricted to its  $d$ -hop neighborhood.

Building on these insights, we introduce the notion of  *$d$ -hop unique coloring*, where all vertices within any node’s  $d$ -hop neighborhood receive distinct colors. The key intuition is that ensuring local uniqueness within receptive fields provides the same expressive power as Global-UIDs, but with fewer identifiers and better generalization. Our contributions can be summarized as follows:

- **Local-UID scheme.** We formalize  $d$ -hop unique coloring and propose a greedy algorithm for constructing such colorings efficiently. This provides a parsimonious alternative to Global-UIDs.
- **Architecture Design.** We propose ColorGNN, which injects color information via color-conditioned embedding mappings rather than as raw input features. ColorGNN is value-invariant: it automatically embeds colors without requiring predefined numeric encodings and can be seamlessly combined with other feature-augmentation techniques. In addition, following conventional UID design, we present a lightweight variant, ColorUID, which concatenates colors with node features.
- **Theoretical justification.** We establish both expressivity and generalization results, showing that Local-UIDs retain the expressive power of Global-UIDs, while significantly improving generalization and requiring far fewer training samples.
- **Empirical validation.** We conduct extensive experiments demonstrating that Local-UIDs achieve (i) state-of-the-art performance on three ILP benchmarks, (ii) strong transferability to Linear Programming (LP) tasks with robust out-of-distribution generalization, and (iii) competitive results on the real-world general graph ZINC dataset.

## 2 PRELIMINARIES

**Basic Notations** Let  $\mathbb{R}$  and  $\mathbb{Z}$  denote the sets of real and integer numbers respectively. We use bold lowercase letters (i.e.,  $\mathbf{x}$ ) to denote vectors, bold uppercase letters (i.e.,  $\mathbf{A}$ ) to denote matrices. Let  $G = (V, E)$  be a graph, Given a node  $v \in V$ , we define its  *$d$ -hop neighbor* as

$$N_d(v) := \{u \neq v \in V \mid \text{dist}(u, v) \leq d\},$$

where  $\text{dist}(u, v)$  is the shortest-path distance. And we let  $B_d(v)$  denote the induced subgraph on  $N_d(v) \cup \{v\}$ . For a vertex  $v \in V$ , we denote its degree as  $\text{deg}(v) = |N_1(v)|$ , and more generally its  $d$ -hop degree as  $\text{deg}_d(v) = |N_d(v)|$ . Let  $\Delta_d := \max_v \text{deg}_d(v)$ . The  $d$ -th power of  $G$ , denoted  $G^d = (V, E^d)$ , is defined as  $(u, v) \in E^d$  if and only if  $\text{dist}_G(u, v) \leq d$ .

**Integer Linear Programming** An integer linear program (ILP) takes the following form:

$$\min_{\mathbf{x}} \mathbf{c}^T \mathbf{x}, \quad \text{s.t. } \mathbf{A} \mathbf{x} \leq \mathbf{b}, \quad \mathbf{x} \in \mathbb{Z}^n$$

where  $\mathbf{x} \in \mathbb{Z}^n$  are integer variables,  $\mathbf{A} \in \mathbb{R}^{m \times n}$  is the constraint matrix,  $\mathbf{b} \in \mathbb{R}^m$  and  $\mathbf{c} \in \mathbb{R}^n$  are coefficients. An ILP instance  $(\mathbf{A}, \mathbf{b}, \mathbf{c})$  can be naturally encoded as a labeled bipartite graph  $G = (V \cup U, E, L, W)$ :

- A left node  $v_i$  represents a variable  $x_i$ , a right node  $u_j$  represents a constraint  $A_{j,:}x \leq b_j$ , and they are connected if the variable participates in the constraint;
- The label of a left node  $v_i$  is  $L(v_i) = c_i$ , and the label of a right node  $u_j$  is  $L(u_j) = b_j$ ;
- The weight of an edge  $e = (v_i, u_j)$  is  $W(e) = A_{ij}$ .

**Graph Neural Network** A  $d$ -depth Graph Neural Network (GNN) can be viewed as a parameterized function as follows: it takes a labeled graph  $G = (V, E, L, W)$  as input,

1.  $h_v^{(0)} := \text{emb}(L_v)$ , where  $\text{emb}(\cdot)$  is a learnable linear mapping.
2. For  $k = 1$  to  $d$ , compute

$$h_v^{(k)} := \text{Merge}^{(k)} \left( h_v^{(k-1)}, \text{AGGR}(\{h_u^{(k-1)}, W(u, v)\} \mid u \in N_1(v)) \right),$$

where  $\text{Merge}^{(k)}$  is a multi-layer perceptron (MLP), and AGGR is a fixed function which is permutation-invariant with respect to its arguments.

3. Finally, each node  $v$  outputs  $h_v^{(d)}$ .

### 3 METHODOLOGY

In this section, we present our Local-UID scheme, which is based on  $d$ -hop unique coloring, and introduce a novel color-aware architecture ColorGNN and its lightweight variant ColorUID. All proofs are provided in the appendix.

#### 3.1 LOCAL-UID SCHEME

Our Local-UID scheme is based on the notion of  $d$ -hop unique coloring of a graph  $G = (V, E)$ , defined as a mapping  $C: V \rightarrow \mathbb{N}$  satisfying that: for any  $v \in V$ ,  $C(u) \neq C(w)$  for any  $u \neq w \in N_d(v)$ . It turns out that this notion can be exactly characterized by the standard  $d$ -hop coloring. Recall that a  $d$ -hop coloring is a map  $C: V \rightarrow \mathbb{N}$  such that  $C(u) \neq C(v)$  for any two distinct  $u, v \in V$  with  $\text{dist}(u, v) \leq d$ . A  $d$ -hop coloring can equivalently defined as a proper coloring of the power graph  $G^d$ .

**Proposition 1.** *For any graph  $G$ , a coloring is  $d$ -hop unique if and only if it is a  $(2d)$ -hop coloring (equivalently, a proper coloring of  $G^{2d}$ ).*

Based on Proposition 1, we construct a  $d$ -hop unique coloring by performing a greedy algorithm to find a proper coloring of  $G^{2d}$ , depicted in Algorithm 1. Here,  $\perp$  denotes the uncolored state, and BFS denotes Breadth-First Search. We sort the vertices of  $G^{2d}$  in non-increasing degree order and apply greedy coloring accordingly, which empirically reduces the number of colors required. We have the following lemma.

**Lemma 1.** *For any graph  $G$ , Algorithm 1 constructs a  $d$ -hop unique coloring using at most  $\Delta_{2d} + 1$  colors in  $\tilde{O}(|G^{2d}|) = \tilde{O}(|V| \cdot \Delta_{2d})$  time.*

Lemma 1 implies that our Local-UID scheme can substantially reduce the number of identifiers: in contrast to Global-UIDs assigning  $|V|$  distinct identifiers, it requires only at most  $\Delta_{2d} + 1$  distinct colors. The savings are particularly significant when  $d$  is small and the graph is sparse. In Section 5.3 (see Table 5), we explicitly compare the number of identifiers required on ILP benchmarks.

#### 3.2 COLOR-AWARE ARCHITECTURAL DESIGN OF GNNs

Given a  $d$ -hop unique coloring produced by Algorithm 1, we introduce a novel GNN architecture, termed ColorGNN, that integrates the coloring into the network design.

**ColorGNN Architecture** The key idea is to incorporate colors via color-conditioned embedding mappings, thereby avoiding the need for predefined numeric encodings. Moreover, ColorGNN remains fully compatible with other feature-augmentation techniques. Formally, given a labeled graph  $G = (V, E, L, W)$  and a coloring  $C: V \rightarrow \mathbb{N}$ , ColorGNN proceeds as follows:

**Algorithm 1** Greedy  $d$ -hop unique coloring**Input:** A graph  $G = (V, E)$ , and a integer  $d \geq 1$ **Output:** Color assignment  $C : V \rightarrow \mathbb{N}$ 


---

```

1: for  $v \in V$  do
2:    $N_{2d}(v) \leftarrow \text{BFS}(G, 2d)$ 
3: end for
4: construct  $G^{2d}$  with  $\{N_{2d}(v)\}$ 
5:  $\tilde{V} \leftarrow$  vertices of  $V$  sorted in nonincreasing order of  $\text{deg}_{2d}(v)$ 
6:  $C(v) \leftarrow \perp$  for all  $v \in V$ 
7: Forbidden  $\leftarrow \emptyset$ 
8: for  $v \in \tilde{V}$  do
9:   Forbidden  $\leftarrow \{C(u) \mid u \in N_{2d}(v) \wedge C(u) \neq \perp\}$ 
10:   $c \leftarrow 0$ 
11:  while  $c \in \text{Forbidden}$  do
12:     $c \leftarrow c + 1$ 
13:  end while
14:   $C(v) \leftarrow c$ 
15: end for
16: return  $\{C(v) \mid v \in V\}$ 

```

---

1.  $\mathbf{h}_v^{(0)} := \text{emb}_{C(v)}(\mathbf{L}_v)$ , where  $\text{emb}_{C(v)}(\cdot)$  is a learnable linear transformation specific to color  $C(v)$ .

2. For  $k = 1$  to  $d$ , compute

$$\mathbf{h}_v^{(k)} := \text{Merge}^{(k)}\left(\mathbf{h}_v^{(k-1)}, \text{AGGR}(\{\mathbf{h}_u^{(k-1)}, W(u, v)\} \mid u \in N_1(v))\right),$$

3. Finally, each node  $v$  outputs  $\mathbf{h}_v^{(d)}$ .

*Design choice:* We condition the embedding functions rather than the aggregation AGGR or merge functions for two reasons. First, AGGR is typically required to remain permutation-invariant; conditioning it on colors risks violating this property or entangling UIDs with neighborhood multisets, which hinders transferability across architectures. Second, conditioning Merge offers no additional expressive power beyond embedding-level conditioning. In practice, the insight of color-conditioning improves performance; both emb-based ColorGNN and Merge-based variant achieve better performance (see Table 4).

**ColorUID Architecture** While ColorGNN shares embeddings among nodes of the same color, the parameter count still scales with the number of colors  $|C|$ ; for graphs with large  $|C|$ , this may cause parameter blow-up. To address this, we propose a lightweight variant, termed ColorUID. Architecturally, ColorUID encodes color identifiers  $C(v)$  directly as input features, using a numeric representation consistent with prior UID schemes (Chen et al., 2022; Han et al., 2023; Chen et al., 2025), i.e.,  $\mathbf{h}_v^{(0)} := \text{emb}(\mathbf{L}_v, C(v))$ . Empirical experiments show that ColorUID consistently improves accuracy across six benchmark datasets, demonstrating the effectiveness of the Local-UID framework.

## 4 THEORETICAL JUSTIFICATION

### 4.1 EXPRESSIVENESS ANALYSIS

In this subsection, we show that a  $d$ -hop unique coloring provides the same expressive power as global unique identifiers (Theorem 1).

Our proof leverages the equivalence between  $d$ -layer GNNs and  $d$ -round distributed algorithms (Loukas, 2020; Sato et al., 2019). Specifically, in the standard distributed computing model, the network is represented by a graph  $G = (V, E)$ , where (i) each node corresponds to a processor

and (ii) each edge  $(u, v) \in E$  corresponds to a bidirectional communication channel. The computation proceeds in synchronous rounds: in each round, all processors execute a message-passing procedure to exchange information with their neighbors, followed by an identical local computation. Unless otherwise stated, we consider *anonymous* distributed algorithms, where nodes do not have identifiers.

**Proposition 2.** *A  $d$ -layer ColorGNN, a  $d$ -layer ColorUID, and a  $d$ -round distributed algorithm equipped with the same coloring are equivalent in expressive power.*

*Proof.* Fix a coloring  $\chi$ . On the one hand, both ColorGNN and ColorUID with  $C$  can be simulated by a  $d$ -round distributed algorithm with the same coloring: one layer corresponds to one round, and the coloring  $C$  is provided as auxiliary input.

On the other hand, given any  $d$ -round distributed algorithm with coloring  $C$ , we can construct  $d$ -layer ColorGNN and ColorUID that faithfully simulate it. For ColorUID, by the universal approximation property of MLP, the merge function  $\text{Merge}^{(k)}$  can be parameterized to implement the local computation in round  $k$ . For ColorGNN, we use the embedding mapping  $\text{emb}_c(\mathbf{v}) = (c, \mathbf{v})$  to preserve the coloring, and similarly choose the merge function  $\text{Merge}^{(k)}$  to implement the local computations, with  $C$  explicitly incorporated as input features. Now, the proposition follows.  $\square$

**Theorem 1** (Local uniqueness suffices). *If a function  $f(G)$  can be computed by a  $d$ -layer GNN with global unique identifiers, then it can also be computed by a  $d$ -layer ColorGNN or ColorUID with a  $d$ -hop unique coloring.*

**Remark 1.** *Even 1-hop uniquely colored ColorGNN and ColorUID can solve several interesting ILP problems, including Maximal independent set (Alon et al., 1986; Luby, 1985) and Maximal Matching (Emek et al., 2014).*

## 4.2 GENERALIZATION ANALYSIS

We analyze the generalization error of ColorGNN for solving graph-level binary classification tasks, including the ILP feasibility problem, and provide a sample complexity estimation (Lemma 2). The generalization gap scales as  $O(\sqrt{|C|})$ , and the sample complexity scales as  $O(|C|)$ , which justifies the design principle that feature augmentation should be as parsimonious as possible.

We consider the standard supervised learning setting. Let  $\mathcal{D}$  an unknown distribution over labeled graphs  $(G, y)$  with  $y \in \{0, 1\}$ , and let  $Z = \{(G_i, y_i)\}_{i=1}^N \stackrel{\text{i.i.d.}}{\sim} \mathcal{D}$  be the training set. We consider the hypothesis class

$$\mathcal{H}^C := \{h : G \rightarrow \{0, 1\} \mid h \text{ is a ColorGNN using coloring } C\},$$

with fixed architecture (depth, widths, and aggregator), focusing only on the effect of the coloring  $C$ . For  $h \in \mathcal{H}^C$ , define population and empirical risks

$$L(h) := \Pr_{(G, y) \sim \mathcal{D}} [h(G) \neq y] \quad \text{and} \quad \hat{L}(h) := \frac{1}{N} \sum_{i=1}^N [h(G_i) \neq y_i].$$

and the generalization gap  $\text{gen}(h, \mathcal{D}, N) := L(h) - \hat{L}(h)$ .

As is the practice case, we assume that all real-valued parameters are stored with finite precision  $p$ . A ColorGNN with coloring  $C$  has two types of parameters: (i) the color-specific embedding matrices  $\text{emb}_c$ , one for each distinct color, contributing  $|C| \cdot |\boldsymbol{\theta}_{\text{emb}}|$  parameters in total and (ii) the remaining parameters of the network  $\{\text{Merge}^{(k)}\}_{k=1}^d$  contributing  $d \cdot |\boldsymbol{\theta}_{\text{Merge}}|$  parameters in total. Hence the overall parameter count of  $\mathcal{H}^C$  is  $|C| \cdot |\boldsymbol{\theta}_{\text{emb}}| + d \cdot |\boldsymbol{\theta}_{\text{Merge}}|$ , and the log size  $\log |\mathcal{H}^C|$  of  $\mathcal{H}^C$  is  $p \cdot (|C| \cdot |\boldsymbol{\theta}_{\text{emb}}| + d \cdot |\boldsymbol{\theta}_{\text{Merge}}|)$ . Then, we have the following lemma. For more details about its proof, we refer readers to Appendix B.4.

**Lemma 2.** *In the standard supervised learning setting described above, fix a  $d$ -hop unique coloring  $C$  and consider the hypothesis class  $\mathcal{H}_C$  realized by a depth- $d$  ColorGNN. Then, for any  $\delta \in (0, 1)$ , with probability at least  $1 - \delta$  over the draw of the training set  $Z \sim \mathcal{D}^N$ , every  $h \in \mathcal{H}_C$  satisfies*

$$|\text{gen}(h, \mathcal{D}, N)| \leq \sqrt{\frac{p \cdot (|C| \cdot |\boldsymbol{\theta}_{\text{emb}}| + d \cdot |\boldsymbol{\theta}_{\text{Merge}}|) + \log(2/\delta)}{2N}},$$

Thus, to guarantee a generalization gap of at most  $\varepsilon$  with probability at least  $1 - \delta$ , it suffices to choose the number of training samples  $N$  such that

$$N \gtrsim \frac{p \cdot (|C| \cdot |\theta_{\text{emb}}| + d \cdot |\theta_{\text{Merge}}|) + \log(2/\delta)}{2\varepsilon^2}.$$

Hence, for fixed depth  $d$ , widths, and precision  $p$ , the required number of training samples grows only linearly with the number of colors  $|C|$ . This quantitatively supports our design principle that feature augmentation should be as parsimonious as possible: Local-UIDs reduce  $|C|$  compared to Global-UIDs and thus enjoy a smaller sample complexity while retaining the same expressive power.

## 5 EXPERIMENTS

We organize the experiments into three parts: (1) setup, specifying the evaluation protocol and baseline methods; (2) main experiments, validating the proposed scheme and reporting further extended experiments; and (3) ablation studies, providing empirical analysis of the method and further extensions. The source code is available at [https://anonymous.4open.science/r/dhop\\_Coloring4GNN-B062/](https://anonymous.4open.science/r/dhop_Coloring4GNN-B062/).

### 5.1 SETUP

**Benchmark problems** We evaluate three standard ILP benchmarks and two LP benchmarks: (1) Bin Packing Problem (BPP; (Schwerin & Wäscher, 1997)), 500 instances, 420 variables and 40 constraints per instance; (2) Balanced Item Placement (BIP; (Gasse et al., 2022)), 300 instances, 1,083 variables and 195 constraints; (3) Steel Mill Slab Design Problem (SMSP; (Schaus et al., 2011)), 380 instances, 22,000–24,000 variables and  $\sim 10,000$  constraints. For LPs, we consider (4) PageRank (Applegate et al., 2021), 100 instances, 10,000 variables and 10,001 constraints, and (5) LP relaxation of Workload Appropriation (WA; (Gasse et al., 2022)) datasets, 100 instances, 5,200 variables and 6,334 constraints.

**Learning Tasks** For ILPs, the goal is to learn a  $f: G \rightarrow \mathbb{R}^n$  to predict the optimal integer solution vector. For LPs, we aim to learn a  $f: G \rightarrow \mathbb{R}^n$  to produce the optimal real solution vector.

**Model architectures** (1) ILPs: Following Chen et al. (2025), we use four “half-layer” graph-convolution blocks to extract hidden representations, followed by a two-layer MLP head for prediction; (2) LPs: Following Li et al. (2024), we use PDHG-Net as our GNN model.

**Baselines** (1) **No-Aug**: A baseline with no feature augmentation. (2) **Position** (Global-UID): A feature-augmentation scheme that assigns distinct numbers to variable nodes (Han et al., 2023; Chen et al., 2024). (3) **Uniform** (Global-UID): A feature-augmentation scheme that adds i.i.d. noise sampled from a uniform distribution to every node (Chen et al., 2023b). (4) **Orbit**: A feature-augmentation scheme that assigns UIDs based on symmetry orbits (Chen et al., 2025). (5) **Orbit+**: An extension of Orbit that additionally leverages inter-orbit relations (Chen et al., 2025).

**Metrics** For ILPs, following (Chen et al., 2025), we adopt *Top- $m\%$  error* as the metric. Specifically, given the ground-truth  $y$  and a prediction  $\hat{y}$ , define the closest symmetric equivalent of  $y$  as  $\tilde{y} = \pi^*(y)$  with  $\pi^* = \arg \min_{\pi} \|\hat{y} - \pi(y)\|_1$ . Let  $M$  be the index set of the top  $m\%$  variables with the smallest values of  $|\text{Round}(\hat{y}_j) - \tilde{y}_j|$ . The Top- $m\%$  error is defined as  $\mathcal{E}(m) = \sum_{i \in M} |\text{Round}(\hat{y}_i) - \tilde{y}_i|$ . For LPs, following (Li et al., 2024), we evaluate the mean squared error (MSE) between predicted and optimal solutions, averaged over the entire dataset.

**Training protocol** We use the unmodified training configurations from (Chen et al., 2025; Li et al., 2024) for the respective architectures, including learning rate, batch size, epochs, optimizer, weight decay, and scheduler (see the original papers for specifications); Given time-complexity considerations, we use greedy 1-hop unique coloring in the experiments. We conduct all experiments under the same configuration on a server equipped with two Intel Xeon Gold 5117 CPUs (2.0 GHz), 256 GB RAM, and two NVIDIA V100 GPUs.

## 5.2 MAIN EXPERIMENTS

This part comprises two components. First, we conduct experiments on the effectiveness of the parsimonious Local-UID scheme, validating our methods on BPP dataset that permits rapid iteration. Second, we present advanced extension experiments, where we extend the lightweight ColorUID to more complex and larger-scale datasets, and even transfer it to the LP domain to verify the method’s strong transferability.

**Validation of the Local-UID scheme** Table 1 empirically validates the effectiveness of the ColorGNN architecture in three respects: (1) Effectiveness of color-aware embedding. ColorGNN consistently outperforms Global-UID baselines (Non-Aug, Uniform, Position), surpasses the Orbit design, and is comparable to Orbit+. (2) Compatibility with prior UID schemes. Augmenting ColorGNN with Orbit or Orbit+ yields substantial additional gains; both variants exhibit strong improvements. (3) Lightweight variant. The special case ColorUID achieves performance comparable to ColorGNN, making it a practical lightweight substitute for large-scale graphs.

Table 1: Comparative results on the BPP dataset for our methods and baselines, evaluated across Top- $m$ % thresholds (lower is better).

Method	Top-30%	Top-50%	Top-70%	Top-90%	Top-100%
Non-Aug	5.81	9.74	14.43	18.28	27.94
Uniform	0.01	0.21	3.45	10.81	19.32
Position	0.00	0.44	3.45	10.58	18.92
Orbit	0.00	0.04	2.33	7.98	17.43
Orbit+	0.00	0.00	1.18	6.96	19.17
ColorGNN	0.00	0.01	0.98	7.80	18.29
ColorGNN,Orbit	0.00	0.03	1.23	6.44	17.42
ColorGNN,Orbit+	0.00	0.00	<b>0.92</b>	<b>5.87</b>	<b>16.55</b>
ColorUID	0.00	0.01	1.02	7.71	18.27

**Extended experiments** To assess scalability and transferability, we evaluate on larger, more complex datasets and extend the evaluation to other domains.

- **ILPs:** Our method scales well to larger and more complex datasets. Table 2 Shows that the ColorUID consistently outperforms the baselines across the board on BIP. On SMSP, it yields substantial gains on the Top-70%, Top-90%, and Top-100% metrics. However, on Top-30% and Top-50% for SMSP, Orbit and Orbit+ perform better. This likely stems from the fact that Orbit incorporates instance-specific information by computing solution orbits and leverages an evaluation metric that normalizes predictions under these symmetries, whereas our approach derives enhanced representational power solely from the underlying graph structure, without relying on such instance-level preprocessing.
- **LPs:** Our method exhibits remarkably strong Out-of-distribution (OOD) generalization on the two LP benchmarks, PageRank and WA (Table 3). We train on smaller scale instances, and test on larger instances (PageRank: 30,000 variables, 30,001 constraints; WA: 7,440 variables, 9,035 constraints). For in-distribution (In-D) evaluation, vanilla GNNs already achieve high accuracy without feature augmentation, consistent with the theoretical analysis (Chen et al., 2022; Li et al., 2024), and adding UIDs (either Uniform or ColorUID) offers little benefit. For the OOD evaluation, Uniform leads to a significant degradation in performance, consistent with our intuition that excessive UIDs harm generalization. However, ColorUID attains *unexpectedly* strong improvement, highlights its robustness across scales.

## 5.3 ABLATION STUDIES

This section has three aims: (1) provide evidence that the observed gains are consistent with our feature-augmentation principle; (2) probe architectural sensitivity by varying the color-embedding function, suggesting that benefits arise from color integration rather than a specific functional choice; (3) illustrate feasibility on a general-graph benchmark via integration with a state-of-the-art

Table 2: Comparative results on two ILP datasets for the lightweight variant (ColorUID) and baselines, evaluated across Top- $m\%$  thresholds (lower is better).

Methods	BIP					SMSP				
	30%	50%	70%	90%	100%	30%	50%	70%	90%	100%
No-Aug	30.34	51.29	72.94	93.77	104.13	32.79	56.31	81.32	109.03	215.38
Uniform	4.54	14.22	44.71	80.57	104.13	17.75	32.21	48.25	72.40	152.54
Position	4.12	12.23	44.41	81.49	103.17	0.13	3.13	26.02	64.09	146.33
Orbit	3.09	6.20	35.69	73.92	101.64	<b>0.00</b>	0.63	16.09	49.37	145.71
Orbit+	2.87	5.17	38.75	78.47	100.76	0.01	<b>0.49</b>	13.06	50.67	139.45
ColorUID	<b>2.74</b>	<b>4.91</b>	<b>32.64</b>	<b>72.23</b>	<b>98.72</b>	0.15	1.55	<b>6.60</b>	<b>16.49</b>	<b>115.57</b>

Table 3: Comparative results on two LP datasets for the lightweight variant (ColorUID) and baselines, using MSE (lower is better).

Methods	PageRank		WA	
	In-D Test MSE	OOD Test MSE	In-D Test MSE	OOD Test MSE
No-Aug	<b>4.51e-07</b>	1.18e-05	3.14e-05	6.34e+02
Uniform	4.66e-07	1.11e-03	2.38e-05	7.07e+02
ColorUID	4.56e-07	<b>6.53e-06</b>	<b>2.30e-05</b>	<b>1.32e-01</b>

positional-encoding model; (4) perform an ablation study on the parameter  $d$  to further substantiate our theoretical claims.

**Sources of Performance Gains** (1) The gains of our methods are not due to increased dimensionality: On BPP, we compare ColorGNN against Non-Aug and Non-Aug+ (which scales the baseline embedding dimension by a factor of  $|C|$ ). We find that merely increasing model capacity without color information does not improve performance and can even degrade it (e.g., Non-Aug vs. Non-Aug+, see the upper half of Table 4). (2) ColorGNN typically uses a smaller but sufficient number of colors: Table 5 shows that, our methods uses very few UIDs on BPP and SMSP. On BIP, our method uses more colors than the Orbit/Orbit+ schemes yet achieves better performance. In BIP, a 1-hop unique coloring requires at least 106 colors, as there exists a constraint involving 106 variables. Consequently, Orbit and Orbit+, which use only 51 colors, lack sufficient expressive power to fully distinguish all 1-hop neighbors.

Table 4: Comparative results on the BPP dataset for color-embedding variants, the Non-Aug baseline, and their combinations, evaluated across Top- $m\%$  thresholds (lower is better). ColorGNN(Merge) is the variant that applies coloring to Merge<sup>(1)</sup> instead of emb.

Method	Top-30%	Top-50%	Top-70%	Top-90%	Top-100%
Non-Aug	5.81	9.74	14.43	18.28	27.94
Non-Aug+	6.05	10.03	14.92	18.88	28.87
ColorGNN	0.00	<b>0.01</b>	<b>0.98</b>	7.80	18.29
ColorGNN(Merge)	0.00	0.02	1.04	<b>7.69</b>	<b>18.28</b>
Orbit	0.00	0.04	2.33	7.98	17.43
ColorGNN,Orbit	0.00	0.03	1.23	6.44	17.42
ColorGNN(Merge),Orbit	0.00	0.03	1.22	6.82	17.43
Orbit+	0.00	0.00	1.18	6.96	19.17
ColorGNN,Orbit+	0.00	0.00	<b>0.92</b>	<b>5.87</b>	<b>16.55</b>
ColorGNN(Merge),Orbit+	0.00	0.01	0.97	6.15	16.96

**Coloring Schemes with Alternative Functional Forms** Embedding structural UID information into the GNN architecture improves performance (see Table 4) : We observe that (1) the coloring function is largely immaterial—ColorGNN(Merge) and ColorGNN(emb) achieve comparable performance; and (2) both variants compose seamlessly with other UID schemes and yield additional gains, demonstrating compatibility.

Table 5: Number of unique identifiers (UIDs) used by each method across three ILP datasets.

Method	BPP	BIP	SMSP
Uniform	420	1082.41	23137.25
Position	420	1083	23417.97
Orbit	140	<b>51</b>	1000
Orbit+	140	<b>51</b>	6070
ColorGNN	<b>32</b>	114	<b>171.68</b>

**General-Graph Applications** We evaluate on the ZINC dataset ((Irwin et al., 2012; Dwivedi et al., 2023)), a graph-level task that predicts the penalized water–octanol partition coefficient (penalized logP). Baselines include GIN(Xu et al., 2018), a GIN augmented with random IDs, SignNet(Lim et al., 2022), an eigenvector-based positional-encoding method, and the state-of-the-art positional-encoding model PEARL (Kanatsoulis et al., 2025). Following PEARL, we adopt the PEARL positional-encoding framework for graph-level modeling and refer to its random-feature variant as R-PEARL. Performance is measured by mean absolute error (MAE), reported as the mean and standard deviation for the model achieving the highest validation accuracy. Replacing the random positional encoding in R-PEARL with our ColorUID yields a variant we call ColorPEARL, which provides further improvements on ZINC (Table 6) and indicates that the d-hop coloring scheme offers stronger performance while remaining highly compatible and plug-and-play.

Table 6: Performance of different methods on ZINC (lower is better).

	GIN	GIN+rand id	SignNet	R-PEARL	Color-PEARL
Test MAE	0.209 ± 0.018	0.279 ± 0.023	0.0853 ± 0.0026	0.0696 ± 0.004	<b>0.0679 ± 0.0024</b>

**Effect of the parameter d on performance.** We evaluate ColorGNN with different values of d on the BPP dataset(7). The results indicate that: (1) 4-hop coloring yields the nearly best performance, which is consistent with our theory that a 2-layer GNN only requires 4-hop coloring; (2) 2-hop coloring is not optimal in terms of accuracy, but it offers a pragmatic compromise between computational cost and performance; and (3) as d increases beyond 4, performance starts to deteriorate. Based on these observations, our practical guideline for choosing d is to set it such that  $d \leq 2 \cdot \text{GNN depth}$ .

Table 7: Comparative results on the BPP dataset for different d (d-hop ColorUID), evaluated across Top-m% thresholds (lower is better).

#d	Top-30%	Top-50%	Top-70%	Top-90%	Top-100%
1	6.05	10.03	14.92	18.88	28.87
2	0.00	0.01	1.02	7.71	18.27
3	0.00	0.01	0.83	7.81	18.26
4	0.00	0.00	<b>0.68</b>	5.39	<b>16.34</b>
5	0.00	0.00	0.72	<b>5.31</b>	16.67
6	0.00	0.005	0.69	5.43	16.41

## 6 CONCLUSION

This paper introduced a parsimonious feature-augmentation scheme for GNNs on ILPs based on d-hop unique coloring, along with a novel color-aware architecture ColorGNN and its lightweight variant ColorUID. Theoretically, we proved that compared with conventional Global-UID schemes, our Local-UID approach achieves the same expressive power while offering superior generalization. Extensive experiments validate its effectiveness across diverse benchmarks. Notably, our method demonstrates unexpectedly strong out-of-distribution performance, and the scheme can integrate seamlessly with other UID strategies.

486 **Reproducibility Statement.** We provide an anonymized repository with complete code and  
487 scripts to reproduce all results. One-click entry points are included for data preprocessing, train-  
488 ing, evaluation, and figure generation. The computing environment, per-table configurations, and  
489 hyperparameters are documented in Section 5.1.

490  
491 **LLM usage.** LLMs is used to aid and polish writing.

## 492 REFERENCES

- 493  
494  
495 Noga Alon, László Babai, and Alon Itai. A fast and simple randomized parallel algorithm for the  
496 maximal independent set problem. *Journal of algorithms*, 7(4):567–583, 1986.
- 497  
498 David Applegate, Mateo Díaz, Oliver Hinder, Haihao Lu, Miles Lubin, Brendan O’Donoghue, and  
499 Warren Schudy. Practical large-scale linear programming using primal-dual hybrid gradient. *Ad-  
500 vances in Neural Information Processing Systems*, 34:20243–20257, 2021.
- 501  
502 Maya Bechler-Speicher, Moshe Eliasof, Carola-Bibiane Schönlieb, Ran Gilad-Bachrach, and Amir  
503 Globerson. On the utilization of unique node identifiers in graph neural networks. *arXiv preprint  
504 arXiv:2411.02271*, 2024.
- 505  
506 Maya Bechler-Speicher, Moshe Eliasof, Carola-Bibiane Schonlieb, Ran Gilad-Bachrach, and Amir  
507 Globerson. Towards invariance to node identifiers in graph neural networks. *arXiv preprint  
508 arXiv:2502.13660*, 2025.
- 509  
510 Stephen Boyd and Jacob Mattingley. Branch and bound methods. *Notes for EE364b, Stanford  
511 University*, 2006:07, 2007.
- 512  
513 Qian Chen, Tianjian Zhang, Linxin Yang, Qingyu Han, Akang Wang, Ruoyu Sun, Xiaodong Luo,  
514 and Tsung-Hui Chang. Symilo: A symmetry-aware learning framework for integer linear opti-  
515 mization. *Advances in Neural Information Processing Systems*, 37:24411–24434, 2024.
- 516  
517 Qian Chen, Lei Li, Qian Li, Jianguhua Wu, Akang Wang, Ruoyu Sun, Xiaodong Luo, Tsung-Hui  
518 Chang, and Qingjiang Shi. When gnns meet symmetry in ilps: an orbit-based feature augmenta-  
519 tion approach. *arXiv preprint arXiv:2501.14211*, 2025.
- 520  
521 Ziang Chen, Jialin Liu, Xinshang Wang, Jianfeng Lu, and Wotao Yin. On representing linear pro-  
522 grams by graph neural networks. *arXiv preprint arXiv:2209.12288*, 2022.
- 523  
524 Ziang Chen, Jialin Liu, Xinshang Wang, and Wotao Yin. On representing mixed-integer linear  
525 programs by graph neural networks. In *The Eleventh International Conference on Learning Rep-  
526 resentations, ICLR 2023, Kigali, Rwanda, May 1-5, 2023*. OpenReview.net, 2023a.
- 527  
528 Ziang Chen, Jialin Liu, Xinshang Wang, and Wotao Yin. On representing mixed-integer linear  
529 programs by graph neural networks. In *The Eleventh International Conference on Learning Rep-  
530 resentations*, 2023b. URL <https://openreview.net/forum?id=4gc3MGZra1d>.
- 531  
532 Vijay Prakash Dwivedi and Xavier Bresson. A generalization of transformer networks to graphs.  
533 arxiv. *arXiv preprint arXiv:2012.09699*, 2020.
- 534  
535 Vijay Prakash Dwivedi, Anh Tuan Luu, Thomas Laurent, Yoshua Bengio, and Xavier Bresson.  
536 Graph neural networks with learnable structural and positional representations. *arXiv preprint  
537 arXiv:2110.07875*, 2021.
- 538  
539 Vijay Prakash Dwivedi, Chaitanya K Joshi, Anh Tuan Luu, Thomas Laurent, Yoshua Bengio, and  
540 Xavier Bresson. Benchmarking graph neural networks. *Journal of Machine Learning Research*,  
24(43):1–48, 2023.
- 541  
542 Moshe Eliasof, Fabrizio Frasca, Beatrice Bevilacqua, Eran Treister, Gal Chechik, and Haggai  
543 Maron. Graph positional encoding via random feature propagation. In *International conference  
544 on machine learning*, pp. 9202–9223. PMLR, 2023.

- 540 Yuval Emek, Christoph Pfister, Jochen Seidel, and Roger Wattenhofer. Anonymous networks: ran-  
541 domization= 2-hop coloring. In *Proceedings of the 2014 ACM symposium on Principles of dis-*  
542 *tributed computing*, pp. 96–105, 2014.
- 543  
544 Maxime Gasse, Didier Chételat, Nicola Ferroni, Laurent Charlin, and Andrea Lodi. Exact combi-  
545 natorial optimization with graph convolutional neural networks. *Advances in neural information*  
546 *processing systems*, 32, 2019.
- 547 Maxime Gasse, Simon Bowly, Quentin Cappart, Jonas Charfreitag, Laurent Charlin, Didier Chételat,  
548 Antonia Chmiela, Justin Dumouchelle, Ambros Gleixner, Aleksandr M Kazachkov, et al. The  
549 machine learning for combinatorial optimization competition (ml4co): Results and insights. In  
550 *NeurIPS 2021 competitions and demonstrations track*, pp. 220–231. PMLR, 2022.
- 551  
552 Florian Grötschla, Jiaqing Xie, and Roger Wattenhofer. Benchmarking positional encodings for  
553 gnns and graph transformers. *arXiv preprint arXiv:2411.12732*, 2024.
- 554 Qingyu Han, Linxin Yang, Qian Chen, Xiang Zhou, Dong Zhang, Akang Wang, Ruoyu Sun, and Xi-  
555 aodong Luo. A gnn-guided predict-and-search framework for mixed-integer linear programming.  
556 *arXiv preprint arXiv:2302.05636*, 2023.
- 557  
558 Kaj Holmberg, Mikael Rönnqvist, and Di Yuan. An exact algorithm for the capacitated facility  
559 location problems with single sourcing. *European Journal of Operational Research*, 113(3):544–  
560 559, 1999.
- 561 Taoan Huang, Aaron M Ferber, Arman Zharmagambetov, Yuandong Tian, and Bistra Dilkina. Con-  
562 trastive predict-and-search for mixed integer linear programs. In *Forty-first International Confer-*  
563 *ence on Machine Learning*, 2024.
- 564  
565 Yinan Huang, William Lu, Joshua Robinson, Yu Yang, Muhan Zhang, Stefanie Jegelka, and Pan Li.  
566 On the stability of expressive positional encodings for graphs. *arXiv preprint arXiv:2310.02579*,  
567 2023.
- 568 John J Irwin, Teague Sterling, Michael M Mysinger, Erin S Bolstad, and Ryan G Coleman. Zinc:  
569 a free tool to discover chemistry for biology. *Journal of chemical information and modeling*, 52  
570 (7):1757–1768, 2012.
- 571  
572 Charilaos I Kanatsoulis, Evelyn Choi, Stephanie Jegelka, Jure Leskovec, and Alejandro  
573 Ribeiro. Learning efficient positional encodings with graph neural networks. *arXiv preprint*  
574 *arXiv:2502.01122*, 2025.
- 575  
576 Nicolas Keriven, Alberto Bietti, and Samuel Vaiter. On the universality of graph neural networks on  
577 large random graphs. *Advances in Neural Information Processing Systems*, 34:6960–6971, 2021.
- 578 Elias B Khalil, Christopher Morris, and Andrea Lodi. Mip-gnn: A data-driven framework for guid-  
579 ing combinatorial solvers. In *Proceedings of the AAAI Conference on Artificial Intelligence*,  
580 volume 36, pp. 10219–10227, 2022.
- 581  
582 Markus Kruber, Marco E Lübbecke, and Axel Parmentier. Learning when to use a decomposition. In  
583 *International conference on AI and OR techniques in constraint programming for combinatorial*  
584 *optimization problems*, pp. 202–210. Springer, 2017.
- 585  
586 Wen-Yang Ku and J Christopher Beck. Mixed integer programming models for job shop scheduling:  
587 A computational analysis. *Computers & Operations Research*, 73:165–173, 2016.
- 588  
589 Fabian Kuhn and Rogert Wattenhofer. On the complexity of distributed graph coloring. In *Proceed-*  
590 *ings of the twenty-fifth annual ACM symposium on Principles of distributed computing*, pp. 7–15,  
2006.
- 591  
592 Bingheng Li, Linxin Yang, Yupeng Chen, Senmiao Wang, Haitao Mao, Qian Chen, Yao Ma, Akang  
593 Wang, Tian Ding, Jiliang Tang, and Ruoyu Sun. Pdhg-unrolled learning-to-optimize method for  
large-scale linear programming. In *Forty-first International Conference on Machine Learning, ICML 2024, Vienna, Austria, July 21-27, 2024*. OpenReview.net, 2024.

- 594 Derek Lim, Joshua Robinson, Lingxiao Zhao, Tess Smidt, Suvrit Sra, Haggai Maron, and Stefanie  
595 Jegelka. Sign and basis invariant networks for spectral graph representation learning. *arXiv*  
596 *preprint arXiv:2202.13013*, 2022.
- 597
- 598 Haoyang Liu, Jie Wang, Zijie Geng, Xijun Li, Yuxuan Zong, Fangzhou Zhu, Jianye Hao, and Feng  
599 Wu. Apollo-milp: An alternating prediction-correction neural solving framework for mixed-  
600 integer linear programming. *arXiv preprint arXiv:2503.01129*, 2025.
- 601
- 602 Andreas Loukas. What graph neural networks cannot learn: Depth vs width. In *8th International*  
603 *Conference on Learning Representations, ICLR 2020*, 2020.
- 604
- 605 Michael Luby. A simple parallel algorithm for the maximal independent set problem. In *Proceedings*  
606 *of the seventeenth annual ACM symposium on Theory of computing*, pp. 1–10, 1985.
- 607
- 608 Sanjay Melkote and Mark S Daskin. Capacitated facility location/network design problems. *Euro-*  
609 *pean journal of operational research*, 129(3):481–495, 2001.
- 610
- 611 Leilei Meng, Chaoyong Zhang, Yaping Ren, Biao Zhang, and Chang Lv. Mixed-integer linear  
612 programming and constraint programming formulations for solving distributed flexible job shop  
613 scheduling problem. *Computers & industrial engineering*, 142:106347, 2020.
- 614
- 615 Vinod Nair, Sergey Bartunov, Felix Gimeno, Ingrid Von Glehn, Pawel Lichocki, Ivan Lobov, Bren-  
616 dan O’Donoghue, Nicolas Sonnerat, Christian Tjandraatmadja, Pengming Wang, et al. Solving  
617 mixed integer programs using neural networks. *arXiv preprint arXiv:2012.13349*, 2020.
- 618
- 619 Moni Naor and Larry Stockmeyer. What can be computed locally? In *Proceedings of the twenty-fifth*  
620 *annual ACM symposium on Theory of computing*, pp. 184–193, 1993.
- 621
- 622 Max B Paulus, Giulia Zarpellon, Andreas Krause, Laurent Charlin, and Chris Maddison. Learning to  
623 cut by looking ahead: Cutting plane selection via imitation learning. In *International conference*  
624 *on machine learning*, pp. 17584–17600. PMLR, 2022.
- 625
- 626 Sujoy Ray, Andrei Soeanu, Jean Berger, and Mourad Debbabi. The multi-depot split-delivery ve-  
627 hicle routing problem: Model and solution algorithm. *Knowledge-Based Systems*, 71:238–265,  
628 2014.
- 629
- 630 Ryoma Sato, Makoto Yamada, and Hisashi Kashima. Approximation ratios of graph neural networks  
631 for combinatorial problems. *Advances in Neural Information Processing Systems*, 32, 2019.
- 632
- 633 Ryoma Sato, Makoto Yamada, and Hisashi Kashima. Random features strengthen graph neural  
634 networks. In *Proceedings of the 2021 SIAM international conference on data mining (SDM)*, pp.  
635 333–341. SIAM, 2021.
- 636
- 637 Pierre Schaus, Pascal Van Hentenryck, Jean-Noël Monette, Carleton Coffrin, Laurent Michel, and  
638 Yves Deville. Solving steel mill slab problems with constraint-based techniques: Cp, Ins, and  
639 cbls. *Constraints*, 16(2):125–147, 2011.
- 640
- 641 Anita Schöbel. A model for the delay management problem based on mixed-integer-programming.  
642 *Electronic notes in theoretical computer science*, 50(1):1–10, 2001.
- 643
- 644 Petra Schwerin and Gerhard Wäscher. The bin-packing problem: A problem generator and some  
645 numerical experiments with ffd packing and mtp. *International transactions in operational re-*  
646 *search*, 4(5-6):377–389, 1997.
- 647
- 648 Clement Vignac, Andreas Loukas, and Pascal Frossard. Building powerful and equivariant graph  
649 neural networks with structural message-passing. *Advances in neural information processing*  
650 *systems*, 33:14143–14155, 2020.
- 651
- 652 Keyulu Xu, Weihua Hu, Jure Leskovec, and Stefanie Jegelka. How powerful are graph neural  
653 networks? *arXiv preprint arXiv:1810.00826*, 2018.

## 648 A RELATED WORKS

649  
650 **Positional encodings for GNNs** The following positional-encoding (PE) methods are *complemen-*  
651 *tary* to our color-based UID scheme and are therefore best viewed as orthogonal design dimensions  
652 rather than head-to-head baselines. Graph Transformers with Laplacian-eigenvector absolute posi-  
653 tional encodings (LapPE) were introduced by Dwivedi & Bresson (2020). Lim et al. (2022) design  
654 architectures that are invariant to eigenvector sign and eigenspace basis, making spectral PEs robust  
655 to symmetry choices. Huang et al. (2023) further improve stability and expressivity by processing  
656 eigenvectors via eigenvalue-weighted, softly partitioned eigenspaces. Beyond explicit eigenvectors,  
657 Eliasof et al. (2023) obtain scalable PEs by filtering random signals with the graph Laplacian, while  
658 Kanatsoulis et al. (2025) propose permutation-invariant features with lower computational cost and  
659 improved stability. Dwivedi et al. (2021) decouple random-walk-based PEs from other structural  
660 signals to build more flexible architectures, and Bechler-Speicher et al. (2024) align latent spaces  
661 induced by different PEs via tailored losses to promote permutation equivariance.

662 Structural Message Passing (SMP) (Vignac et al., 2020) enhances expressive power by propagating  
663 a local-context matrix initialized with one-hot node identifiers. This perspective is complementary  
664 to ours: one can, for example, replace the SMP context matrix with our color features, or incorpo-  
665 rate their node identifiers into our ColorGNN architecture. Under the distributed local algorithm  
666 model, Sato et al. (2019) prove that weak 2-coloring strictly increases the capability of GNNs. In a  
667 related vein, Sato et al. (2021) use random node features that act as global UIDs to enhance GNNs.  
668 Our work generalizes this line of research by showing that a parsimonious family of *local* UID  
669 schemes (such as  $d$ -hop colorings) suffices to significantly improve GNN expressivity across a wide  
670 range of combinatorial and general-graph tasks. Complementary to these perspectives, Grötschla  
671 et al. (2024) provide a systematic benchmark of PEs for both message-passing GNNs and graph  
672 Transformers (Dwivedi & Bresson, 2020). In their framework, PEs are treated as node-feature aug-  
673 mentations that encode graph topology and can be plugged into arbitrary GNN/GT backbones. Our  
674 ColorGNN fits naturally into this view: it can be interpreted as a new family of structural positional  
675 encodings based on locally unique colorings (e.g.,  $d$ -hop unique coloring) that provide symmetry-  
676 breaking identity information while preserving permutation equivariance.

## 677 B PROOF

### 678 B.1 PROOF OF PROPOSITION 1

679  
680 *Proof.* ( $\Rightarrow$ ) Suppose the coloring is  $d$ -hop unique. Take any two vertices  $u, w$  with  $\text{dist}(u, w) \leq 2d$ .  
681 A shortest  $u$ - $w$  path has a midpoint of  $v$  with both  $\text{dist}(u, v), \text{dist}(w, v) \leq d$ , so  $u, w \in N_d(v)$  and  
682 must have distinct colors. ( $\Leftarrow$ ) Suppose the coloring is a  $(2d)$ -hop coloring. Then for any  $v$  and any  
683 distinct  $u, w \in N_d(v)$ , we have  $\text{dist}(u, w) \leq \text{dist}(u, v) + \text{dist}(v, w) \leq d + d = 2d$  by the triangle  
684 inequality. So  $u$  and  $w$  must have different colors. Thus each  $N_d(v)$  is uniquely colored.  $\square$   
685  
686

### 687 B.2 PROOF OF PROPOSITION 2

688  
689 *Proof.* Fix a coloring  $\chi$ . On the one hand, both ColorGNN and ColorUID with  $C$  can be simulated  
690 by a  $d$ -round distributed algorithm with the same coloring: one layer corresponds to one round, and  
691 the coloring  $C$  is provided as auxiliary input.

692 On the other hand, given any  $d$ -round distributed algorithm with coloring  $C$ , we can construct  $d$ -  
693 layer ColorGNN and ColorUID that faithfully simulate it. For ColorUID, by the universal ap-  
694 proximation property of MLP, the merge function  $\text{Merge}^{(k)}$  can be parameterized to implement the  
695 local computation in round  $k$ . For ColorGNN, we use the embedding mapping  $\text{emb}_c(\mathbf{v}) = (c, \mathbf{v})$   
696 to preserve the coloring, and similarly choose the merge function  $\text{Merge}^{(k)}$  to implement the local  
697 computations, with  $C$  explicitly incorporated as input features. Now, the proposition follows.  $\square$   
698

### 699 B.3 PROOF OF THEOREM 1

700  
701 *Proof.* By Proposition 2, it suffices to show that a  $d$ -round distributed algorithm with a  $d$ -hop unique  
coloring is as powerful as a  $d$ -round distributed algorithm with global unique identifiers. In a  $d$ -round

distributed algorithm with global unique identifiers, each node  $v$  bases its output solely on its  $d$ -hop neighborhood  $B_d(v)$ . A  $d$ -hop unique coloring ensures that all nodes in  $B_d(v)$  receive distinct colors; thus the induced subgraph  $B_d(v)$  can be reconstructed by the node  $v$  (see, e.g., (Naor & Stockmeyer, 1993; Kuhn & Wattenhofer, 2006)). Now, the theorem is immediate.  $\square$

#### B.4 PROOF OF LEMMA 2

We first upper bound the size of the hypothesis class  $\mathcal{H}_C$  induced by a depth- $d$  ColorGNN with a fixed  $d$ -hop unique coloring  $C$ .

**I: Bounding the size of  $\mathcal{H}_C$ .** Recall that for a fixed architecture, the network parameters consist of

- color-specific embedding parameters, one embedding function  $\text{emb}_c$  for each  $c \in C$ , contributing in total  $|C| \cdot |\boldsymbol{\theta}_{\text{emb}}|$  scalar parameters; and
- color-independent parameters in the  $d$  message-passing/merge layers  $\{\text{Merge}^{(k)}\}_{k=1}^d$ , contributing in total  $d \cdot |\boldsymbol{\theta}_{\text{Merge}}|$  scalar parameters.

Let

$$P := |C| \cdot |\boldsymbol{\theta}_{\text{emb}}| + d \cdot |\boldsymbol{\theta}_{\text{Merge}}|$$

denote the total number of scalar parameters. Under the  $p$ -bit finite-precision assumption, each scalar parameter can take at most  $2^p$  distinct values. Therefore the number of distinct parameter configurations is at most

$$(2^p)^P = 2^{pP}.$$

Each parameter configuration defines a hypothesis in  $\mathcal{H}_C$ , so we obtain the simple upper bound  $|\mathcal{H}_C| \leq 2^{pP}$

**II: Finite-class generalization inequality.** We next recall a standard finite-class generalization bound for binary classification. Let  $H$  be a finite class of hypotheses  $h : \mathcal{G} \rightarrow \{0, 1\}$ , and let  $Z = \{(G_i, y_i)\}_{i=1}^N \sim \mathcal{D}^N$  denote a set of  $N$  i.i.d. sample from an unknown distribution  $\mathcal{D}$ . For  $h \in H$ , define  $L(h)$  and  $\widehat{L}(h)$  as in Section 4.2. Then for any  $\delta \in (0, 1)$ , with probability at least  $1 - \delta$  over the draw of  $Z$ , we have

$$\sup_{h \in H} |L(h) - \widehat{L}(h)| \leq \sqrt{\frac{\log(2|H|) + \log(1/\delta)}{2N}}. \quad (1)$$

For completeness, we sketch the proof. Fix any  $h \in H$  and define the random variables  $X_i := \mathbb{1}\{h(G_i) \neq y_i\}$  so that  $L(h) = \mathbb{E}[X_i]$  and  $\widehat{L}(h) = \frac{1}{N} \sum_{i=1}^N X_i$ . By Hoeffding's inequality,

$$\Pr\left(|\widehat{L}(h) - L(h)| > \varepsilon\right) \leq 2 \exp(-2N\varepsilon^2).$$

Taking a union bound over all  $h \in H$  gives

$$\Pr(\exists h \in H : |\widehat{L}(h) - L(h)| > \varepsilon) \leq 2|H| \exp(-2N\varepsilon^2).$$

Choosing  $\varepsilon$  so that the right-hand side equals  $\delta$ , i.e.

$$2|H| \exp(-2N\varepsilon^2) = \delta \iff \varepsilon^2 = \frac{\log(2|H|) + \log(1/\delta)}{2N},$$

yields equation 1. This bound holds uniformly over all  $h \in H$ .

**III: Instantiating the bound for  $\mathcal{H}_C$ .** We now apply equation 1 with  $H = \mathcal{H}_C$ . Using the bound from Step I, we obtain

$$\log(2|\mathcal{H}_C|) \leq pP + \log 2 = pP + 1,$$

Plugging this into equation 1 gives that, with probability at least  $1 - \delta$ ,

$$\sup_{h \in \mathcal{H}_C} |L(h) - \widehat{L}(h)| \leq \sqrt{\frac{pP + \log(2/\delta)}{2N}} = \sqrt{\frac{p \cdot (|C| \cdot |\boldsymbol{\theta}_{\text{emb}}| + d \cdot |\boldsymbol{\theta}_{\text{Merge}}|) + \log(2/\delta)}{2N}}.$$

Since this bound holds uniformly over all  $h \in \mathcal{H}_C$ , it is equivalent to the claimed statement of Lemma 2 for the generalization gap  $\text{gen}(h, \mathcal{D}, N) = L(h) - \widehat{L}(h)$ .