RingFormer: A Ring-Enhanced Graph Transformer for Organic Solar Cell Property Prediction

Zhihao Ding*, Ting Zhang*, Yiran Li, Jieming Shi, Chen Jason Zhang

Department of Computing, The Hong Kong Polytechnic University, Hong Kong SAR
{tommy-zh.ding, leyla2.zhang, yi-ran.li}@connect.polyu.hk, {jieming.shi, jason-c.zhang}@polyu.edu.hk

Abstract

Organic Solar Cells (OSCs) are a promising technology for sustainable energy production. However, the identification of molecules with desired OSC properties typically involves laborious experimental research. To accelerate progress in the field, it is crucial to develop machine learning models capable of accurately predicting the properties of OSC molecules. While graph representation learning has demonstrated success in molecular property prediction, it remains under-explored for OSC-specific tasks. Existing methods fail to capture the unique structural features of OSC molecules, particularly the intricate ring systems that critically influence OSC properties, leading to suboptimal performance. To fill the gap, we present RingFormer, a novel graph transformer framework specially designed to capture both atom and ring-level structural patterns in OSC molecules. RingFormer constructs a hierarchical graph that integrates atomic and ring structures and employs a combination of local message-passing and global attention mechanisms to generate expressive graph representations for accurate OSC property prediction. We evaluate RingFormer's effectiveness on five curated OSC molecule datasets through extensive experiments. The results demonstrate that RingFormer consistently outperforms existing methods, achieving a 22.77% relative improvement over the nearest competitor on the CEPDB dataset.

Code — https://github.com/TommyDzh/RingFormer
Extended version — https://arxiv.org/abs/2412.09030

Introduction

As the demand for renewable energy sources grows, organic solar cells (OSCs) have attracted considerable interest for their ability to convert sunlight into electricity (Wang et al. 2016). Despite their potential, the development of OSCs has been hindered by the reliance on trial-and-error methods, which involve complex and time-consuming synthesis procedures (Sun et al. 2019). To accelerate progress, there is increasing interest in leveraging machine learning models to accurately predict the properties of OSC molecules, promising to expedite the development of OSCs.

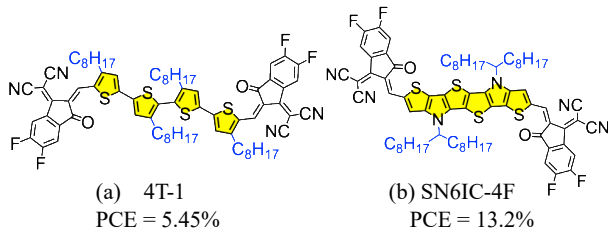In this work, we aim to predict OSC properties based on the structure of OSC molecules, i.e., organic small-molecule

---

Figure 1: Example of OSC molecules.

(a) 4T-1
PCE = 5.45%

(b) SN6IC-4F
PCE = 13.2%

semi-conducting materials for the active layer of OSCs. Such molecules function as electron acceptors or donors to create photovoltaic effects with efficacy linked to their conjugated structure, such as aromatic rings (Solak and Irmak 2023). In chemistry, rings are closed loops of atoms connected through covalent bonds (Jonathan Clayden 2012), and the design of complex ring systems has become a central focus in OSC research (Schweda et al. 2021). These systems can include various fused and non-fused ring structures (Gao et al. 2023). For example, as shown in Figure 1, the core of OSC molecule 4T-1 (highlighted in yellow) on the left comprises four non-fused thiophene rings connected by single bonds. In contrast, the core of SN6IC-4F on the right side is a fused S,N-heteroacene consisting of six rings, where adjacent rings share a pair of bonded atoms. The backbone of both molecules can be characterized by multiple inter-connected rings connecting each other in various ways, forming a complex ring system. Additionally, different alkyl functional groups (in blue) attached to the core as side chains can further influence OSC performance (Ching and Isabelle 2023). These structural differences ultimately lead to significantly different power conversion efficiency (PCE) values between 4T-1 and SN6IC-4F. Therefore, accurately predicting the properties of OSC molecules requires capturing both the high-level structure of the ring system and the local-level atomic groups within these molecules.

For predicting molecular properties, graph neural networks (GNNs) have been widely adopted, they represent chemical molecules as graphs, with nodes for atoms and edges for chemical bonds. While GNNs effectively capture local atomic structures like functional groups, they struggle to model higher-order patterns such as those found in OSC

ring systems (He et al. 2023). To address these limitations, graph pooling methods (Gao and Ji 2019; Lee et al. 2019) and motif-based approaches (Yu and Gao 2022; Zang et al. 2023) have been developed to extend graph learning beyond localized atomic features by leveraging substructures and motifs. However, they fail to emphasize the rings and their interconnections which are crucial for OSC molecules. This limitation is particularly problematic as GNNs are inherently weak in modeling ring-like structures and long-range dependencies (Loukas 2019; Chen et al. 2020; Rampášek et al. 2022), leading to suboptimal performance in OSC property prediction (Eibeck et al. 2021). Recently, Zhu *et al.* (Zhu et al. 2023) enhanced GNNs by incorporating additional ring representations, but they still neglect the crucial interconnections between rings. Thus, current graph-based models are inadequate for effectively highlighting the ring systems in OSC molecules, thereby limiting their predictive accuracy for OSC properties.

To fill the gap, we introduce RingFormer, a novel graph transformer framework designed to capture structural patterns at both the atom and ring levels within OSC molecules. As illustrated in Figure 2, RingFormer firstly constructs a hierarchical OSC graph to explicitly model atom groups and the ring system in an OSC molecule. The hierarchical OSC graph comprises three levels: the atom-level graph, the ring-level graph, and the inter-level graph. The atom-level graph describes the atomic bonding structure of the OSC molecule. Above this, the ring-level graph focuses on the rings and their interconnections, capturing the high-level ring system. Finally, the inter-level graph models the relationships between rings and atoms, representing the hierarchical structure of the molecule. By integrating these three levels, the hierarchical OSC graph provides a comprehensive depiction of OSC molecular structures, enabling more accurate prediction of OSC properties.

Subsequently, RingFormer combines the power of local message-passing and global attention to capture the distinct structural patterns in each level and learn expressive graph representations. On atom-level graphs, RingFormer layers use message-passing GNNs to encode localized structures into atom node representations. For ring-level graphs, Ring-Former layers incorporate a novel cross-attention mechanism to capture global patterns within the ring system. Specifically, the proposed mechanism effectively and efficiently captures the interconnections between rings. Ring-Former layers further facilitate interactions between ring and atom nodes through message-passing on the inter-level graph. At the end of each RingFormer layer, a hierarchical fusion strategy is implemented to allow information from different levels to complement each other. After stacking multiple layers, RingFormer aggregates both atom and ring node representations into graph representations that comprehensively encode the structure of OSC molecules. To evaluate the effectiveness of RingFormer in OSC property prediction, we compare RingFormer against 11 baselines over 5 OSC molecule datasets. Experimental results show that RingFormer outperforms baselines consistently. Notably, on the large-scale CEPDB dataset, RingFormer achieves a remarkable 22.77% relative improvement over the closest competitor. In summary, our contributions are:

- We study the important problem of OSC property prediction and present RingFormer, the first graph transformer framework capturing ring systems within OSC molecules.

- We construct a hierarchical OSC graph to depict OSC molecule at atom and ring levels. Message-passing and transformer-based attention mechanism are combined to learn expressive representations on the hierarchical graph.

- We further capture the inter-level interactions between rings and atoms in the hierarchical graph, and design a hierarchical fusion strategy to strength the quality of learned representations.

- Extensive experiments on 5 OSC property prediction datasets validate the superior performance of RingFormer.

Related Work

OSC Property Prediction. Organic solar cells (OSCs) have garnered significant research attention as one of the most promising technologies for harnessing solar energy (Eibeck et al. 2021). As conducting laboratory experiments to screen candidate OSC molecules is time and resource-intensive (Xu et al. 2022), researchers have recently turned to machine learning methods for efficient OSC property prediction. Currently, fingerprint-based approaches (Eibeck et al. 2021) are commonly employed. Typically, these methods utilize hand-crafted molecular fingerprints such as MACCS (Durant et al. 2002) and ECFP (Rogers et al. 2010) as molecular features, which are then inputted into off-the-shelf machine learning models like random forest and support vector machine. However, fingerprints represent simplified abstractions of molecular structures, which overlook crucial molecular information and interactions, particularly in OSC molecules with complex structures (Miyake and Saeki 2021). Inspired by the success of GNNs in drug discovery, Eibeck *et al.* (Eibeck et al. 2021) recently explored the application of GNNs in OSC property prediction. However, they found that conventional GNNs often perform poorly in predicting OSC properties, achieving lower accuracy compared to fingerprint methods. The development of effective models for OSC property prediction remains under-explored.

Graph Representation Learning on Molecules. As molecules can be naturally represented as graphs, graph neural networks (GNNs) (Hu et al. 2019; Zhu et al. 2023) are widely used for molecule property prediction. However, conventional GNNs struggles to capture high-order structures (He et al. 2023), including important molecular features such as rings (Chen et al. 2020; Loukas 2019). To address the limitations of GNNs, researchers have developed graph pooling methods (Gao and Ji 2019; Lee et al. 2019), motif-based methods (Yu and Gao 2022; Zang et al. 2023), and graph transformers (Rong et al. 2020; Ying et al. 2021; Kim et al. 2022). Although these models capture higher-level molecular patterns beyond localized features, they still fail to emphasize the rings and their connections, which are crucial for accurately predicting OSC molecule properties.
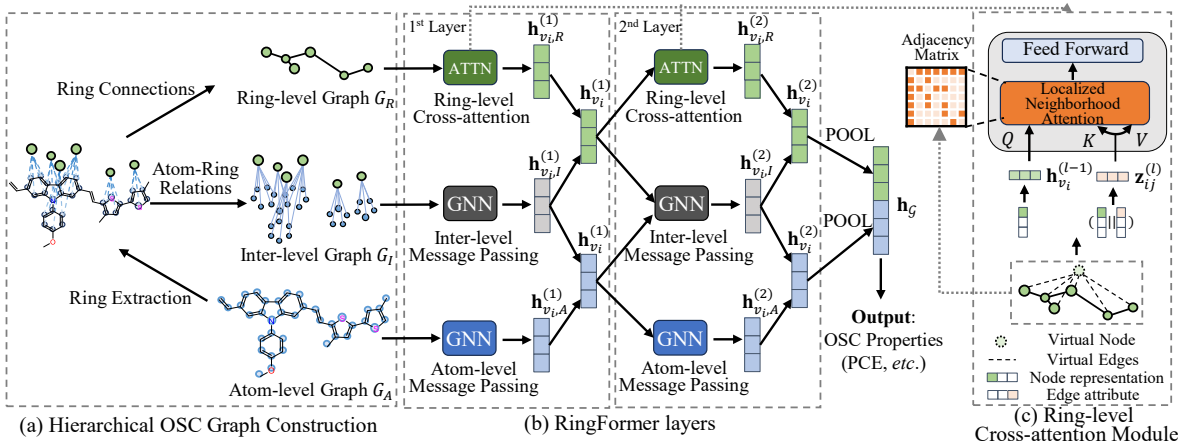
Figure 2: The RingFormer framework. For clarity, we showcase the framework with $L = 2$ RingFormer layers.

Problem Formulation

Data Model. Following previous works on OSC property prediction (Eibeck et al. 2021), an OSC molecule is represented by an atom-level molecular graph $G_A = (V_A, E_A)$, where V_A is the set of nodes representing atoms and E_A is the set of edges representing chemical bonds. Let $\mathbf{x}_{v_i} \in \mathbb{R}^{d_{V_A}}$ denote node attribute vector of node $v_i \in V_A$ and $\mathbf{e}_{ij} \in \mathbb{R}^{d_{E_A}}$ denote edge attribute vector of edge $e_{ij} \in E$, where d_{V_A} and d_{E_A} are the dimension of node and edge attributes, respectively.

Prediction Task. Given an OSC molecular graph G_A, the objective of our work is to learn a property prediction model $f : G_A \rightarrow y$ that predicts the target property value y, such as power conversion efficiency (PCE), highest occupied molecular orbital (HOMO), lowest unoccupied molecular orbital (LUMO), etc (Miyake and Saeki 2021). Notably, PCE holds particular significance for OSC molecules, as it serves as a pivotal indicator of their efficiency in converting sunlight into electrical energy (Eibeck et al. 2021). Accurate prediction of PCE is essential for assessing the performance of OSCs. In this study, we prioritize the prediction of PCE as our pivotal target property.

The RingFormer Method

Overview. Aiming to capture both atom and ring-level structural patterns in an OSC molecule, our proposed Ring-Former framework firsts constructs a hierarchical OSC graph, and then holistically encodes the hierarchical graph using RingFormer layers for property prediction. As depicted in Figure 2, the hierarchical OSC graph consists of atom, ring, and inter-level graphs that models OSC molecule structure from multiple levels. The RingFormer layers combine message-passing and global attention mechanisms to further learn node representations at each level and fuse information across hierarchies. Finally, RingFormer aggregates node representations of both atom and ring nodes to generate the graph representation which is used for predicting target OSC properties.

Hierarchical OSC Graph Construction

In this section, we introduce how to construct the hierarchical OSC graph $\mathcal{G}$ for an OSC molecule. As illustrated in Figure 1, the ring system, characterized by the types of rings and their interconnections, plays a crucial role in determining the properties of OSC molecules. To overcome the limitations of a flat molecular graph G_A that solely depicts low-level atom-based structure, our proposed hierarchical OSC graph $\mathcal{G}$ comprehensively represents an OSC molecule from three levels. As shown in Figure 2, given an input atom-level graph G_A, we construct a ring-level graph G_R above the atom-level G_A to explicitly depict the high-level ring system. After having G_A and G_R, we further incorporate a bipartite inter-level graph G_I to describe the connections between atom and ring nodes and represent graph hierarchy. Combining the three levels of graphs, we obtain the hierarchical OSC graph $\mathcal{G} = \{G_A, G_R, G_I\}$. Next, we provide more details on each level of the graph.

Ring-level Graph Construction. Given an input OSC molecule modeled by an atom-level molecular graph G_A (introduced in the problem formulaiton section), we first extract the smallest rings from G_A and represent each ring by a ring node in $V_R = \{v_1, v_2, \cdots v_{|V_R|}\}$. Here, the smallest ring is defined as a ring such that no proper subset of its nodes can form a smaller ring.

Different types of rings are characterized by their atom composition. For a ring node $v_i \in V_R$, we use the one-hot encoding of its ring type as the attribute vector $\mathbf{x}_{v_i} \in \mathbb{R}^{d_{V_R}}$, where d_{V_R} represents the total number of ring types. Between a pair of extracted ring nodes, we add an edge in the edge set E_R to suggest the connections between them when (1) the two rings share one or more atoms or (2) the two rings are connected by a single chain of one or more non-aromatic bonds. For an edge $e_{ij} \in E_R$, its edge attribute is an one-hot vector $\mathbf{e}_{ij} \in \mathbb{R}^{d_{E_R}}$ indicating the connection type, where d_{E_R} represents the total number of connection types. We differentiate the type of connection between two ring nodes based on the atoms shared by two rings in condition (1) or the atoms and bonds composition of the connecting chain in

condition (2). Having the ring nodes and their interconnections, we construct ring-level graph $G_R = (V_R, E_R)$.

Inter-level Graph Construction. We further construct an inter-level graph G_I to make G_A and G_R interact with each other and depict graph hierarchy. Specifically, for a ring node $v_i \in V_R$ and an atom node $v_j \in V_A$, we connect them using $e_{ij} \in E_I$ if the atom node is one of the constituents of the ring, represented by the ring node. In this way, we model the graph hierarchy based on the membership between atom and ring nodes in the inter-level graph and have $G_I = (V_A \cup V_R, E_I)$. Note that G_I is an undirected graph, enabling bidirectional message propagation between any pair of connected atom and ring nodes.

Finally, we combine the three levels of graphs as the hierarchical OSC graph $\mathcal{G} = \{G_A, G_R, G_I\}$ which comprehensively encapsulates the complex structure of OSC molecules, enabling accurate property prediction.

RingFormer Layer

In this section, we present RingFormer layers that learn expressive representations for both atom and ring nodes on the hierarchical OSC graph $\mathcal{G}$. Considering the hierarchical nature of $\mathcal{G}$, we combine the power of message-passing and global attention in RingFormer layers to better capture the unique patterns in each graph level. As shown in Figure 2, each RingFormer layer consists of an atom-level message passing module, a ring-level cross-attention module, and an inter-level message passing module. At the end of each RingFormer layer, we fuse the latent information learned from hierarchical levels to generate expressive node representations for atoms and rings, which are then forwarded to the next layer for further updates. Next, we introduce each module in detail.

Atom-level Message Passing Module. On the atom-level graph G_A, we aim to learn atom node representations that capture the local chemical structures in OSC molecules, such as function groups. As GNNs are good at mining patterns in local structures (He et al. 2023), we adopt GNNs to perform message passing on G_A. In the l-th RingFormer layer, GNNs are applied on the atom-level graph as follows:

$$\mathbf{h}_{v_i,A}^{(l)} = \text{GNN}_A^{(l)}(\mathbf{h}_{v_i}^{(l-1)}, \{\mathbf{h}_{v_j}^{(l-1)} | v_j \in \mathcal{N}_{G_A}(v_i)\}, \{\mathbf{e}_{ij}\}), \quad (1)$$

where $v_i, v_j \in V_A$ denote nodes in G_A, $\mathcal{N}_{G_A}(v_i)$ is the neighborhood of node v_i in G_A, $\mathbf{e}_{ij}$ is the edge attribute of edge $e_{ij} \in E_A$, and $\mathbf{h}_{v_i}^{(l-1)} \in \mathbb{R}^d$ is the node representation of node v_i from the preceding RingFormer layer. Specially, we have $\mathbf{h}_{v_i}^{(0)} = \mathbf{W}_a \mathbf{x}_{v_i}$, where $\mathbf{W}_a \in \mathbb{R}^{d \times d_{V_A}}$ are learnable parameters to transform node attribute $\mathbf{x}_{v_i}$, d is the hidden dimension. Here, $\text{GNN}_A^{(l)}$ can be arbitrary GNN architecture. Throughout the paper, we use GINE (Hu et al. 2019) as the default GNN backbone. In Appendix, we study the effect of different GNN backbones. $\mathbf{h}_{v_i,A}^{(l)} \in \mathbb{R}^d$ is the output node representation of the atom-level message passing module in the l-th RingFormer layer.

Ring-level Cross-attention Module. In the encoding of ring-level graph, we aim to learn ring node representations that capture the global structure of the ring system. To

achieve this, we introduce a novel transformer-based cross-attention module tailored for the ring-level graph. Recently, graph transformers (Ying et al. 2021; Kreuzer et al. 2021) have shown their superiority to GNNs in modeling long-range node dependencies and capturing structural patterns beyond localized graph structures (Rampášek et al. 2022).

However, existing graph transformers still exhibit limitations in capturing the chemical semantic-rich structures of ring-level graphs that represent ring systems in OSC molecules. A primary limitation of existing graph transformers is their inadequate utilization of edge attributes which contain crucial information about how two rings are connected in the ring-level graph. The graph transformers typically compute self-attention using node representations as queries, keys, and values, with edge attributes merely considered in attention biases. Consequently, valuable chemical semantics in edge attributes of the ring-level graph are not integrated into the output ring node representations.

Moreover, graph transformers further suffer from efficiency issues when integrating edges attributes. For instance, as Graphormer calculate fully-connected attention among all node pairs in a graph, it also needs to compute the shortest paths among all node pairs as edge attributes, leading to high computation complexity relative to the number of nodes (Shirzad et al. 2023).

To address the above limitations, we introduce two key designs in our proposed ring-level cross-attention module: cross-attention mechanism and localized neighborhood attention strategy. These designs optimize the utilization of edge attributes, while maintaining computational efficiency. Firstly, to better leverage the edge attributes, we propose to adopt a *cross-attention* strategy instead of the node-centric *self-attention* commonly used in graph transformers. Specifically, for a source ring node $v_j \in V_R$ connected to the target ring node $v_i \in V_R$ through edge $e_{ij} \in E_R$, we regard the node representation of v_i as the query, while the combination of the node representation of v_j and edge attribute $\mathbf{e}_{ij}$ serves as the key and value. In such a way, edge attributes are not only considered in the attention scores but also integrated into the target node representation during aggregation. Then, in calculating multi-head attention, the queries, keys, and values for the c-th attention head in the l-th layer are calculated as follows:

$$\mathbf{z}_{ij}^{(l)} = \text{MLP}_Z^{(l)}(\mathbf{h}_{v_j}^{(l-1)} || \mathbf{e}_{ij}),$$
$$\mathbf{q}_{i,c}^{(l)} = \mathbf{W}_{q,c}^{(l)} \mathbf{h}_{v_i}^{(l-1)}, \mathbf{k}_{ij,c}^{(l)} = \mathbf{W}_{k,c}^{(l)} \mathbf{z}_{ij}^{(l)}, \mathbf{v}_{ij,c}^{(l)} = \mathbf{W}_{v,c}^{(l)} \mathbf{z}_{ij}^{(l)}, \quad (2)$$

where $||$ indicates concatenation operation, $\text{MLP}_Z^{(l)}$ is a multi-layer perception network (MLP) with one hidden layer, $\mathbf{W}_{q,c}^{(l)}, \mathbf{W}_{k,c}^{(l)}, \mathbf{W}_{v,c}^{(l)} \in \mathbb{R}^{\frac{d}{C} \times d}$ are learnable parameters, and C is the number of attention heads. Specially, we have $\mathbf{h}_{v_i}^{(0)} = \mathbf{W}_r \mathbf{x}_{v_i}$, where $\mathbf{W}_r \in \mathbb{R}^{d \times d_{V_R}}$ are learnable parameters to transform node attribute $\mathbf{x}_{v_i}$.

Furthermore, to reduce the computation cost, inspired by (Shirzad et al. 2023), we replace fully-connected graph attention with localized neighborhood attention augmented by a virtual molecule node. Specifically, ring-level cross-attention module adds a virtual molecule node to G_R and connects the virtual node to all other ring nodes using virtual

edges. In this way, global information can be gathered and spread through the virtual node without the need to compute attention between all pairs of nodes. We initialize the representation of the virtual node using learnable embedding and set the edge attributes of the virtual edges as one-hot vectors different from real edges in G_R. Then, we calculate localized neighborhood attention between the target node and its immediate neighbors as follows:

$$\alpha_{ij} = \frac{\langle \mathbf{q}_{i,c}^{(l)}, \mathbf{k}_{ij,c}^{(l)} \rangle}{\sum_{v_k \in \mathcal{N}_{G_R}(v_i)} \langle \mathbf{q}_{i,c}^{(l)}, \mathbf{k}_{ik,c}^{(l)} \rangle}, \quad (3)$$

where $\mathbf{q}_{i,c}^{(l)}, \mathbf{k}_{ij,c}^{(l)}, \mathbf{v}_{ij,c}^{(l)} \in \mathbb{R}^{\frac{d}{C}}$ are the query, key, value vectors calculated in Eq.(2), $\mathcal{N}_{G_R}(v_i)$ is the neighborhood of node v_i in G_R, and $\langle \mathbf{q}, \mathbf{k} \rangle = \frac{\mathbf{q}^T \mathbf{k}}{\sqrt{d}}$. Then, we update ring node representations based on cross-attention as follows:

$$\hat{\mathbf{h}}_{v_i,R}^{(l)} = \mathbf{W}_s^{(l)} \mathbf{h}_{v_i}^{(l-1)} + \mathbf{W}_o^{(l)} \overset{C}{\underset{c=1}{\|}} (\sum_{v_j \in \mathcal{N}_{G_R}(v_i)} \alpha_{ij} \mathbf{v}_{ij,c}^{(l)}), \quad (4)$$

where $\mathbf{W}_s^{(l)}, \mathbf{W}_o^{(l)} \in \mathbb{R}^{d \times d}$ are learnable parameters, $\hat{\mathbf{h}}_{v_i,R}^{(l)} \in \mathbb{R}^d$ is the output node representation of node v_i after multi-head cross-attention. Specially, we calculate $\mathbf{h}_{v_i}^{(0)} = \mathbf{W}_r \mathbf{x}_{v_i} \| \mathbf{p}_{v_i}$, where $\mathbf{p}_{v_i} \in \mathbb{R}^{d_p}$ is a real-valued embedding vector with dimension d_p working as node v_i's degree-based position encoding (Ying et al. 2021) and $\mathbf{W}_r \in \mathbb{R}^{(d-d_p) \times d_{V_R}}$ is learnable parameter.

Following the convention of typical transformer blocks (Vaswani et al. 2017), the node representations generated by cross-attention further go through a feed-forward layer and are updated as follows:

$$\mathbf{h}_{v_i,R}^{(l)} = \text{FFN}^{(l)}(\hat{\mathbf{h}}_{v_i,R}^{(l)} + \mathbf{h}_{v_i}^{(l-1)}) \quad (5)$$

where $\text{FFN}^{(l)}$ is the feed-forward block in the l-th layer, and $\mathbf{h}_{v_i,R}^{(l)} \in \mathbb{R}^d$ is the output node representation of ring-level cross-attention module in the l-th RingFormer layer.

Inter-level Message Passing Module. In the inter-level message passing module, we further transfer knowledge between atom and ring nodes on the bipartite graph G_I, allowing atom representations to perceive the global structure of the high-level ring system, while ring representations are enriched by the local structure around their constituent atoms. Specifically, GNNs are applied on the inter-level message passing module as follows:

$$\mathbf{h}_{v_i,I}^{(l)} = \text{GNN}_I^{(l)}(\mathbf{h}_{v_i}^{(l-1)}, \{\mathbf{h}_{v_j}^{(l-1)} | v_j \in \mathcal{N}_{G_I}(v_i)\}), \quad (6)$$

where $v_i, v_j \in V_I$ denotes nodes in G_I, $\mathcal{N}_{G_I}(v_i)$ is the neighborhood of node v_i in G_I, and $\mathbf{h}_{v_i,I}^{(l)} \in \mathbb{R}^d$ is the output node representation of inter-level message passing module in the l-th RingFormer layer.

Hierarchical Messages Fusion. After learning node representations on the three levels of graphs, the l-th RingFormer layer generates two types of node representations from different graph hierarchies for every atom and ring node. Specially, an atom node $v_i \in V_A$ has $\mathbf{h}_{v_i,A}^{(l)}$ generated by atom-level message passing module on G_A and $\mathbf{h}_{v_i,I}^{(l)}$ generated

DATASET	# GRAPHS	AVG. # NODES	AVG. # EDGES	AVG. # RINGS
CEPDB	2.3M	27.6	33.3	6.7
HOPV	350	42.7	49.3	7.5
PFD	1055	77.1	84.2	8.2
NFA	654	118.2	133.0	15.8
PD	277	80.7	88.2	8.5

Table 1: Dataset statistics.

by inter-level message passing module on G_I. Similarly, an ring node $v_j \in V_R$ has $\mathbf{h}_{v_j,R}^{(l)}$ learned by ring-level cross-attention module on G_R and $\mathbf{h}_{v_j,I}^{(l)}$ learned by inter-level message passing module on G_I.

To facilitate information fusion across the hierarchies, we combine the two representations for each node:

$$\mathbf{h}_{v_i}^{(l)} = \begin{cases} \text{MLP}_A^{(l)}(\mathbf{h}_{v_i,A}^{(l)} \| \mathbf{h}_{v_i,I}^{(l)}) & \text{if } v_i \in V_A \\ \text{MLP}_R^{(l)}(\mathbf{h}_{v_i,R}^{(l)} \| \mathbf{h}_{v_i,I}^{(l)}) & \text{if } v_i \in V_R \end{cases} \quad (7)$$

where $\mathbf{h}_{v_i}^{(l)} \in \mathbb{R}^d$ is the final node representation output by the l-th RingFormer layer for node $v_i \in V_A \cup V_R$. $\mathbf{h}_{v_i}^{(l)}$ is then sent to the next RingFormer layer for further update.

Prediction Layer

After stacking L RingFormer layers, we aggregate their output node representations to generate graph representation for the hierarchical OSC graph. First, for each node $v_i \in V_A \cup V_R$, we concatenate its representations from all RingFormer layers as its final node representation to incorporate structural patterns at different scales:

$$\mathbf{h}_{v_i} = \text{CONCAT}(\mathbf{h}_{v_i}^{(0)}, \mathbf{h}_{v_i}^{(1)}, \cdots, \mathbf{h}_{v_i}^{(L)}), \quad (8)$$

where CONCAT indicates concatenation operation, and $\mathbf{h}_{v_i} \in \mathbb{R}^{d \times (L+1)}$ is the concatenated node representation. Then, we separately aggregate node representations of atoms and rings in $\mathcal{G}$ using sum pooling and concatenate the two to obtain the final graph representation:

$$\mathbf{h}_\mathcal{G} = \text{POOL}(\{\mathbf{h}_{v_i} | v_i \in V_A\}) \| \text{POOL}(\{\mathbf{h}_{v_i} | v_i \in V_R\}), \quad (9)$$

where POOL indicates sum pooling operation and $\mathbf{h}_\mathcal{G} \in \mathbb{R}^{2d \times (L+1)}$ is the final graph representation for the hierarchical OSC graph $\mathcal{G}$. As the target properties of OSC molecules considered in this paper are real values, and thus OSC property prediction can be regarded as a regression task, we project the molecular representation into the logits $\hat{y}_\mathcal{G}$ using one linear layer. Given a batch of training molecules, RingFormer is trained with mean absolute error (MAE) loss:

$$\mathcal{L} = \frac{1}{B} \sum_{b=1}^{B} |y_{\mathcal{G}_b} - \hat{y}_{\mathcal{G}_b}|, \quad (10)$$

where B is the batch size, and $\hat{y}_{\mathcal{G}_b}$ and $y_{\mathcal{G}_b}$ represent the predicted and ground-truth molecule properties, respectively.

Experiments

Experimental Settings

Datasets and Evaluation Metrics. We curate 5 OSC molecule datasets to evaluate property prediction performance, as listed in Table 1. Specially, CEPDB (Hachmann

Method	CEPDB	HOPV	PFD	NFA	PD
MACCS	0.898	1.632	**1.770**	2.614	2.594
ECFP	0.510	1.544	1.787	<u>2.377</u>	2.704
GINE	0.460	1.614	1.826	2.620	2.528
GINE-VN	0.393	1.724	1.878	3.164	2.962
AttentiveFP	0.415	2.002	1.897	2.826	2.608
O-GNN	0.267	1.727	1.868	2.587	2.866
TopKPool	0.527	1.598	1.830	2.644	2.523
SAGPool	0.536	1.607	1.841	2.648	2.557
GPS	0.247	1.942	2.395	3.233	2.690
Graphormer	-	1.609	1.799	2.689	<u>2.522</u>
GraphViT	<u>0.244</u>	<u>1.479</u>	1.887	2.467	2.856
RingFormer	**0.189**	**1.477**	<u>1.776</u>	**2.259**	**2.482**

Table 2: PCE (%) prediction performance compared between RingFormer and baselines in terms of test MAE ($\downarrow$). $\downarrow$ indicates smaller values are better. **Bold**: best. <u>Underline</u>: runner-up.

Method	PCE	HOMO	LUMO	Band Gap	Voc	Jsc
	%	eV	eV	eV	V	mA/cm^2
MACCS	0.898	0.115	0.122	0.184	0.115	35.297
ECFP	0.510	0.066	0.065	0.080	0.066	15.574
GINE	0.491	0.049	0.058	0.073	0.049	15.409
GINE-VN	0.496	0.048	0.059	0.073	0.048	15.110
AttentiveFP	0.453	0.041	0.057	0.068	0.040	14.182
O-GNN	0.259	0.026	0.030	0.036	0.026	8.039
TopKPool	0.566	0.053	0.060	0.078	0.053	16.120
SAGPool	0.581	0.056	0.050	0.074	0.056	15.510
GPS	<u>0.241</u>	<u>0.020</u>	<u>0.021</u>	<u>0.025</u>	<u>0.018</u>	<u>7.514</u>
GraphViT	0.322	0.040	0.035	0.055	0.040	12.479
RingFormer	**0.193**	**0.014**	**0.018**	**0.023**	**0.014**	**5.993**

Table 3: Multi-task learning performance on CEPDB by test MAE ($\downarrow$). The second row indicates the units of different properties. **Bold**: best. <u>Underline</u>: runner-up.

et al. 2011) dataset is generated based on density functional theory (DFT), and other four datasets, HOPV (Lopez et al. 2016), PFD (Nagasawa et al. 2018), NFA (Miyake and Saeki 2021), and PD (Miyake and Saeki 2021), are datasets consisting of different types of OSC molecules whose properties are experimentally validated. Additional dataset and target OSC properties descriptions can be found in the Appendix of the extended version of this paper. We adopted scaffold-based splitting (Wu et al. 2018), commonly used in the field, to partition molecules into training, validation, and testing sets with a ratio of 6:2:2. We use mean absolute error (MAE) as the evaluation metric following previous studies (Eibeck et al. 2021; Saleh et al. 2023).

Baselines and Implementation Details. We compare RingFormer with 11 competitors in 4 categories. (i) Fingerprint-based methods, *MACCS* (Durant et al. 2002), and *ECFP* (Rogers et al. 2010). (ii) GNN-based methods, *GINE* (Hu et al. 2019), *GINE-VN* (Gilmer et al. 2017), *AttentiveFP* (Xiong et al. 2019), and *O-GNN* (Zhu et al. 2023). (iii) Pooling-based methods, including *TopKPool* (Gao and Ji 2019) and *SAGPool* (Lee et al. 2019). (iv) Transformer-based methods, *Graphormer* (Ying et al. 2021), *GPS* (Rampášek et al. 2022), and *GraphViT* (He et al. 2023). The implementation details of RingFormer and other baselines are provided in the Appendix.

Overall Performance

PCE prediction. We report the performance of RingFormer in predicting power conversion efficiency (PCE), the primary property of interest for OSC molecules (Solak and Irmak 2023), in Table 2. Firstly, we observe that Ring-Former consistently achieves the best performance, except that RingFormer is the runner-up in PFD. For instance, on CEPDB, RingFormer achieves test MAE 0.189, which indicates 22.8% relative improvement over the best competitor with test MAE 0.244. On NFA, the dataset with the highest average number of rings (See Table 1), RingFormer outper-

forms the fingerprint-based method ECFP by 4.96%. In contrast, other deep learning models struggle to match the performance of ECFP. Furthermore, across the four experimental datasets (HOPV, PFD, NFA, PD), we observe that GNN-based methods consistently yield inferior performance compared to fingerprint-based methods, indicating a struggle in learning structural patterns from larger and more complicated OSC molecules. However, RingFormer still achieves competitive performance across these datasets, emerging as the top performer in three out of four datasets. The results demonstrate RingFormer's ability to capture structural patterns in OSC molecules for accurate property prediction.

Multi-task learning. We further evaluate the performance of RingFormer in multi-task learning using CEPDB dataset. Specifically, we aim to predict 5 target properties, resulting in 5 regression tasks. The details of the five target properties are given in Appendix. For training RingFormer and other deep neural network-based competitors, we set the output dimension to be the same as the number of target properties and train the neural network using MAE loss. As the results shown in Table 3, RingFormer consistently outperforms other competitors in all six target properties, often by a significant margin. For instance, RingFormer has test MAE 5.993 in predicting J_{sc}, achieving 20.24% relative improvement on the best competitor. The results further validate RingFormer's superiority in predicting multiple OSC properties simultaneously. Additionally, we observe GPS consistently achieve promising results across all target properties, only inferior to RingFormer. It is because both RingFormer and GPS combine the power of message passing and global attention, which validates the importance of capturing both local and global structural features in OSC molecules.

Model Analysis

Ablation on hierarchical OSC graph. We evaluate the effectiveness of hierarchical OSC graph by comparing it with all three levels $\mathcal{G} = \{G_A, G_R, G_I\}$ to a subset of $\mathcal{G}$. As shown in Table 4, we observe a significant performance drop

	CEPDB	HOPV	PFD	NFA	PD
G_A only	0.550	1.912	1.841	3.122	2.683
G_R only	0.358	1.874	2.001	2.412	2.531
$\mathcal{G}\setminus G_R$	0.606	1.526	1.831	2.304	2.540
$\mathcal{G}\setminus G_I$	0.315	1.497	1.795	2.299	2.501
$\mathcal{G}$	**0.189**	**1.477**	**1.776**	**2.259**	**2.482**

Table 4: Ablation on hierarchical OSC graph by test MAE ($\downarrow$). \ denotes exclusion.

Encoder	CEPDB	HOPV	PFD	NFA	PD
Cross-attention	0.1886	**1.4774**	**1.7757**	2.2588	**2.4819**
Cross-attention w.o. virtual	**0.1860**	1.5106	1.8082	2.2602	2.5876
GINE	0.3576	1.4796	1.7892	**2.2519**	2.6203
GINE-VN	0.3136	1.5069	1.7856	2.2665	2.5296
Vanilla Transformer	0.2319	1.5394	1.8273	2.4151	2.5805
GPS	0.2231	1.5137	1.7910	2.4069	2.5390

Table 5: PCE prediction performance of different implementations of ring-level graph encoder by test MAE ($\downarrow$).

	CEPDB	HOPV	PolymerFA	nNFA	pNFA
Motifs	0.4556	1.5203	**1.7230**	2.5013	2.5960
Rings	**0.1886**	**1.4774**	1.7757	**2.2588**	**2.4819**
Ring+Motifs	0.2440	1.4833	1.8098	2.7561	2.5908

Table 6: Comparison between using rings and BRICS-based motifs as high-level structure by test MAE ($\downarrow$).
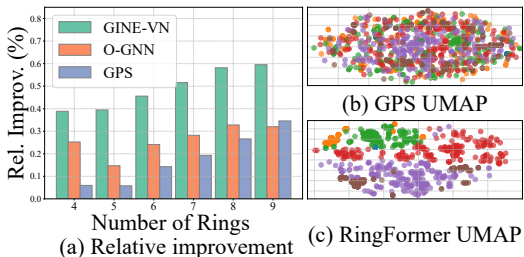

Figure 3: (a) Performance improvement on molecules with varying numbers of rings; (b) and (c) Comparison of embedding visualizations, colors representing the number of rings.

functional group nodes with ring nodes if they share atoms. As results shown in Figure 6, we observe that RingFormer generally achieves better performance by modeling rings than motifs and combination of rings and motifs. The results indicate the importance of focusing on the patterns in ring systems for OSC property prediction.

Performance *vs.* ring system complexity. To demonstrate RingFormer's advantage in leveraging ring structures within OSC molecules, we evaluate the performance of Ring-Former on OSC molecules with different numbers of rings, which reflect different levels of ring system complexity. Firstly, we examine the relative improvement in MAE achieved by RingFormer in predicting PCE for molecules with varying numbers of rings in the CEPDB test set, compared to baseline methods GINE-VN, O-GNN, and GPS. As illustrated in Figure 3 (a), the performance gain generally increases together with the number of rings in molecules, indicating a clear correlation between RingFormer's superiority and the complexity of the ring systems. Furthermore, we visualize the graph representations of OSC molecules in the CEPDB test set using UMAP (McInnes 2018). As depicted in Figure 3 (b) and (c), the embeddings generated by Ring-Former can be distinctly separated based on the number of rings in the OSC molecules, unlike those generated by GPS. These observations confirm that RingFormer excels in capturing the intricate structures of ring systems.

More experiments. In the Appendix of the extended version of this paper, we present the efficiency of RingFormer, evaluate its performance with different GNN backbones other than GINE, and report the results of varying the number of RingFormer layers.

Conclusion

This paper addresses the under-explored problem of predicting properties of organic solar cells (OSCs) and introduces RingFormer, a novel graph transformer framework designed to capture rings and their interconnections within an OSC molecule to facilitate accurate prediction. RingFormer constructs a hierarchical OSC graph that represents OSC molecular structure at both atom and ring levels, and leverages a combination of local message-passing and global attention mechanisms to learn expressive graph representations. Extensive experiments demonstrate the superiority of Ring-Former in OSC property prediction.

on the incomplete hierarchical OSC graph compared to the full one, which validates the necessity of all levels in $\mathcal{G}$. Moreover, observing that $\mathcal{G}\setminus G_I$ is better than G_A and G_R, we conclude that encoding structure of both atom and ring level is important in OSC property prediction. Comparing $\mathcal{G}$ with $\mathcal{G}\setminus G_I$, we find transferring hierarchical information between atom and ring levels can improve performance.

Effectiveness of ring-level cross-attention module. We evaluate the effectiveness of the proposed ring-level cross-attention module by replacing it with other graph representation learning layers and report the results in Table 5. Specifically, *Cross-attention w.o. virtual* is the variant of our cross-attention without virtual molecule nodes. *Vanilla Transformer* (Vaswani et al. 2017) is the transformer encoder without special designs for graphs. The results in Table 5 shows that our proposed cross-attention achieves the best performance in 3 out of 5 datasets and competitive results in the other 2 datasets, which validates its effectiveness.

Comparison to molecular motifs. As rings can be considered as a special kind of chemical motif, one question is how our method will perform when using general chemical motifs instead of focusing on rings. Hence, we explore the effect of modeling general chemical motifs. Following Zhang et al. (2021), we extract molecular motifs using BRICS (Degen et al. 2008), and replace the ring-level graph with a motif-level graph where motifs are regarded as nodes. We further combine BRICS functional group motifs with ring systems in the high-level graph (Ring+Motif), connecting

Acknowledgements

References

Chen, Z.; Chen, L.; Villar, S.; and Bruna, J. 2020. Can graph neural networks count substructures? *Advances in neural information processing systems*, 33: 10383–10395.

Ching, M.-C.; and Isabelle, K. 2023. Impact of Alkyl-Based Side Chains in Conjugated Materials for Bulk Heterojunction Organic Photovoltaic Cells—A Review. *Energies*, 16(18): 6639.

Degen, J.; Wegscheid-Gerlach, C.; Zaliani, A.; and Rarey, M. 2008. On the Art of Compiling and Using'Drug-Like'Chemical Fragment Spaces. *ChemMedChem: Chemistry Enabling Drug Discovery*, 3(10): 1503–1507.

Durant, J. L.; Leland, B. A.; Henry, D. R.; and Nourse, J. G. 2002. Reoptimization of MDL keys for use in drug discovery. *Journal of chemical information and computer sciences*, 42(6): 1273–1280.

Eibeck, A.; Nurkowski, D.; Menon, A.; Bai, J.; Wu, J.; Zhou, L.; Mosbach, S.; Akroyd, J.; and Kraft, M. 2021. Predicting power conversion efficiency of organic photovoltaics: models and data analysis. *ACS omega*, 6(37): 23764–23775.

Gao, H.; Han, C.; Wan, X.; and Chen, Y. 2023. Recent Progress in Non-Fused Ring Electron Acceptors for High Performance Organic Solar Cells. *Industrial Chemistry & Materials*, 1(1): 60–78.

Gao, H.; and Ji, S. 2019. Graph u-nets. In *international conference on machine learning*, 2083–2092. PMLR.

Gilmer, J.; Schoenholz, S. S.; Riley, P. F.; Vinyals, O.; and Dahl, G. E. 2017. Neural message passing for quantum chemistry. In *International conference on machine learning*, 1263–1272. PMLR.

Hachmann, J.; Olivares-Amaya, R.; Atahan-Evrenk, S.; Amador-Bedolla, C.; Sánchez-Carrera, R. S.; Gold-Parker, A.; Vogt, L.; Brockway, A. M.; and Aspuru-Guzik, A. 2011. The Harvard clean energy project: large-scale computational screening and design of organic photovoltaics on the world community grid. *The Journal of Physical Chemistry Letters*, 2(17): 2241–2251.

He, X.; Hooi, B.; Laurent, T.; Perold, A.; LeCun, Y.; and Bresson, X. 2023. A generalization of vit/mlp-mixer to graphs. In *International Conference on Machine Learning*, 12724–12745. PMLR.

Hu, W.; Liu, B.; Gomes, J.; Zitnik, M.; Liang, P.; Pande, V.; and Leskovec, J. 2019. Strategies for Pre-training Graph Neural Networks. In *International Conference on Learning Representations*.

Jonathan Clayden, S. W., Nick Greeves. 2012. *Organic chemistry*. Oxford University Press.

Kim, J.; Lee, S.; Kim, D.; Ahn, S.; and Park, J. 2022. Substructure-Atom Cross Attention for Molecular Representation Learning.

Kreuzer, D.; Beaini, D.; Hamilton, W.; Létourneau, V.; and Tossou, P. 2021. Rethinking graph transformers with spectral attention. *Advances in Neural Information Processing Systems*, 34: 21618–21629.

Lee et al., J. 2019. Self-attention graph pooling. In *International conference on machine learning*, 3734–3743. PMLR.

Lopez, S. A.; Pyzer-Knapp, E. O.; Simm, G. N.; Lutzow, T.; Li, K.; Seress, L. R.; Hachmann, J.; and Aspuru-Guzik, A. 2016. The Harvard organic photovoltaic dataset. *Scientific data*, 3(1): 1–7.

Loukas, A. 2019. What graph neural networks cannot learn: depth vs width. In *International Conference on Learning Representations*.

McInnes, L. 2018. Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426*.

Miyake, Y.; and Saeki, A. 2021. Machine learning-assisted development of organic solar cell materials: issues, analyses, and outlooks. *The Journal of Physical Chemistry Letters*, 12(51): 12391–12401.

Nagasawa et al., S. 2018. Computer-aided screening of conjugated polymers for organic solar cell: classification by random forest. *The Journal of Physical Chemistry Letters*, 9(10): 2639–2646.

Rampášek, L.; Galkin, M.; Dwivedi, V. P.; Luu, A. T.; Wolf, G.; and Beaini, D. 2022. Recipe for a general, powerful, scalable graph transformer. *Advances in Neural Information Processing Systems*, 35: 14501–14515.

Rogers et al., D. 2010. Extended-connectivity fingerprints. *Journal of chemical information and modeling*, 50(5): 742–754.

Rong, Y.; Bian, Y.; Xu, T.; Xie, W.; Wei, Y.; Huang, W.; and Huang, J. 2020. Self-supervised graph transformer on large-scale molecular data. *Advances in Neural Information Processing Systems*, 33: 12559–12571.

Saleh, J.; Haider, S.; Akhtar, M. S.; Saqib, M.; Javed, M.; Elshahat, S.; and Kamal, G. M. 2023. Energy Level Prediction of Organic Semiconductors for Photodetectors and Mining of a Photovoltaic Database to Search for New Building Units. *Molecules*, 28(3): 1240.

Schweda, B.; Reinfelds, M.; Hofstadler, P.; Trimmel, G.; and Rath, T. 2021. Recent Progress in the Design of Fused-Ring Non-Fullerene Acceptors Relations between Molecular Structure and Optical, Electronic, and Photovoltaic Properties. *ACS Applied Energy Materials*, 4(11): 11899–11981.

Shirzad, H.; Velingker, A.; Venkatachalam, B.; Sutherland, D. J.; and Sinop, A. K. 2023. Exphormer: Sparse transformers for graphs. In *International Conference on Machine Learning*.

Solak, E. K.; and Irmak, E. 2023. Advances in Organic Photovoltaic Cells: A Comprehensive Review of Materials, Technologies, and Performance. *RSC Advances*, 13(18): 12244–12269.

Sun, W.; Li, M.; Li, Y.; Wu, Z.; Sun, Y.; Lu, S.; Xiao, Z.; Zhao, B.; and Sun, K. 2019. The use of deep learning to fast evaluate organic photovoltaic materials. *Advanced Theory and Simulations*, 2(1): 1800116.

Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, Ł.; and Polosukhin, I. 2017. Attention is all you need. *Advances in neural information processing systems*, 30.

Wang, J.-L.; Xiao, F.; Yan, J.; Wu, Z.; Liu, K.-K.; Chang, Z.-F.; Zhang, R.-B.; Chen, H.; Wu, H.-B.; and Cao, Y. 2016. Difluorobenzothiadiazole-Based Small-Molecule Organic Solar Cells with 8.7% Efficiency by Tuning of π-Conjugated Spacers and Solvent Vapor Annealing. *Advanced Functional Materials*, 26(11): 1803–1812.

Wu, Z.; Ramsundar, B.; Feinberg, E. N.; Gomes, J.; Geniesse, C.; Pappu, A. S.; Leswing, K.; and Pande, V. 2018. MoleculeNet: a benchmark for molecular machine learning. *Chemical science*, 9(2): 513–530.

Xiong, Z.; Wang, D.; Liu, X.; Zhong, F.; Wan, X.; Li, X.; Li, Z.; Luo, X.; Chen, K.; Jiang, H.; et al. 2019. Pushing the boundaries of molecular representation for drug discovery with the graph attention mechanism. *Journal of medicinal chemistry*, 63(16): 8749–8760.

Xu, X.; Yu, L.; Meng, H.; Dai, L.; Yan, H.; Li, R.; and Peng, Q. 2022. Polymer solar cells with 18.74% efficiency: from bulk heterojunction to interdigitated bulk heterojunction. *Advanced Functional Materials*, 32(4): 2108797.

Ying, C.; Cai, T.; Luo, S.; Zheng, S.; Ke, G.; He, D.; Shen, Y.; and Liu, T.-Y. 2021. Do transformers really perform badly for graph representation? *Advances in Neural Information Processing Systems*, 34: 28877–28888.

Yu, Z.; and Gao, H. 2022. Molecular representation learning via heterogeneous motif graph neural networks. In *International Conference on Machine Learning*, 25581–25594. PMLR.

Zang et al., X. 2023. Hierarchical molecular graph self-supervised learning for property prediction. *Communications Chemistry*, 6(1): 34.

Zhang, Z.; Liu, Q.; Wang, H.; Lu, C.; and Lee, C.-K. 2021. Motif-based graph self-supervised learning for molecular property prediction. *Advances in Neural Information Processing Systems*, 34: 15870–15882.

Zhu, J.; Wu, K.; Wang, B.; Xia, Y.; Xie, S.; Meng, Q.; Wu, L.; Qin, T.; Zhou, W.; Li, H.; et al. 2023. $\mathcal{O}$-GNN: incorporating ring priors into molecular modeling. In *ICLR*.