Expert Systems With Applications 269 (2025) 126484

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Motif-driven molecular graph representation learning

Runze Wang [a], Yuting Ma [a], Xingyue Liu [b], Zijie Xing [b], Yanming Shen [a,b,*]

[a] School of Computer Science and Technology, Dalian University of Technology, 116024, Dalian, China
[b] School of Future Technology, Dalian University of Technology, 116024, Dalian, China

ARTICLE INFO

Keywords:
Graph neural network
Expressive power
Decoupled motif representation

ABSTRACT

Graph Neural Networks (GNNs) have emerged as powerful tools for molecular graph analysis. Subgraph-based GNNs focus on learning high-level local patterns beyond simple node interactions to improve GNN expressiveness. However, current subgraph-based methods lack a unified scheme for incorporating molecular motifs that can be applied consistently across various GNN frameworks. To address this, we propose Uni-Motif, a universal molecular motif integration approach that enhances GNNs' expressive power. Specifically, we decouple a motif into functional encoding and learnable structural encoding, where functional encoding serves as a unique identifier for each motif and structural encoding provides local structural context for nodes within the same motif. We further analyze the effects of seven motif extraction techniques on model performance and provide an in-depth evaluation. Experimental results demonstrate the effectiveness of Uni-Motif in improving GNN expressive power as well as its compatibility with various GNN architectures. Code is available at https://github.com/GraphMoLab/Uni-Motif.

1. Introduction

Graph machine learning has revolutionized the research paradigm in molecular science (Wieder et al., 2020). Graph Neural Networks (GNNs) empower the representation of non-structured data, profoundly revealing the elusive intricacy within molecular complex systems. By representing the molecules as graphs, where atoms serve as nodes and bonds as edges, GNNs can effectively capture the inherent structural features, such as atomic spatial arrangement, which enable high-quality molecular representations (Wu, Cui, Pei, Zhao, & Guo, 2022). High-quality molecular representations play a critical role in accurate prediction for downstream tasks (Tropsha, Isayev, Varnek, Schneider, & Cherkasov, 2024).

Although GNNs are naturally suitable for modeling molecular systems, a common issue is their lack of expressiveness (Nalwade, Marshall, Eladi, & Sharma, 2024; Zhang et al., 2023). Recent studies have shown that the expressive power of the standard GNN frameworks is bounded by the first-order Weisfeiler–Lehman (1-WL) in graph isomorphism test (Wang et al., 2022; Xu, Hu, Leskovec, & Jegelka, 2018). The fundamental limitation of expressive power lies in its reliance on local neighbor aggregation, preventing it from distinguishing between non-isomorphic graphs with identical local structures. As a result, message passing mechanism in GNNs cannot distinguish between two nodes or two graphs that have the same local structure but differ globally. Notably, accurately distinguishing non-isomorphic graphs is crucial for molecular representation learning (Atz, Grisoni, & Schneider, 2021), as it enables differentiation of molecules with similar local structures but distinct global properties, which is essential for tasks like chemical property prediction (Wieder et al., 2020). To break the expressive bottleneck of the 1-WL test, some studies have introduced higher-order Weisfeiler–Lehman (k-WL) tests, extending the 1-WL test to k-order to broaden the global receptive field of GNNs (Feng, Chen, Li, Sarkar, & Zhang, 2022; Morris et al., 2019; Xu & Zou, 2024). However, a gap persists between the theoretical advancements and practical performance (Yang, Wang, Shen, Qi, & Yin, 2022). Another line of research focuses on designing positional encodings to integrate absolute or relative positional information into GNNs (Dwivedi, Luu, Laurent, Bengio, & Bresson, 2021; Eliasof et al., 2023; Maskey et al., 2022), including Graph Transformers (Park, Chang, Lee, Kim, & Hwang, 2022; Zhao, Ma, Zhang, Deng, & Wei, 2023). By embedding the positional context, models can capture more intricate graph structural features, enhancing their capacity to differentiate between non-isomorphic graphs.

Substructures represent recurring patterns or motifs within a molecular graph, such as rings, chains, or functional groups (Zhang, Hu, Subramonian, & Sun, 2024; Zhao, Jin, Akoglu, & Shah, 2021). These patterns act as higher-level features representing more complex interactions in molecules beyond simple pairwise node connections (Zhang,

* Corresponding author.
E-mail addresses: runze_wang@mail.dlut.edu.cn (R. Wang), mayuting_0@163.com (Y. Ma), xingyueliu22@163.com (X. Liu), kokojjj@mail.dlut.edu.cn (Z. Xing), shen@dlut.edu.cn (Y. Shen).

https://doi.org/10.1016/j.eswa.2025.126484
Received 20 November 2024; Received in revised form 26 December 2024; Accepted 7 January 2025
Available online 16 January 2025

Feng, Du, He, & Wang, 2023). Recognizing the motifs as structural encodings helps GNNs overcome the inability to distinguish non-isomorphic graphs. Some studies aim to design more expressive subgraph neural networks (Wang & Zhang, 2021; Zeng et al., 2023) by identifying explicit motifs. Besides, motifs often carry diverse chemical or functional significance. For instance, an aromatic ring in a molecule implies certain stability and reactivity, serving as a functional and unique identifier. Therefore, incorporating these chemically meaningful features allows the model to better correlate structural characteristics with molecular properties (Han et al., 2023; Jin, Barzilay, & Jaakkola, 2020).

It can be observed that modeling molecular motifs into GNNs improves their expressive power to distinguish molecular structures with subtle differences. As mentioned above, motifs hold practical significance in molecular domains, acting as identifiers with specific functional roles, while in graph structures, they represent structural patterns. Therefore, building a model that treats these motifs as key indicators for predicting bioactivity, and structure–activity relationships, which are essential in fields such as drug discovery and biochemical research. However, recent subgraph-based GNNs tend to simplify motifs as either purely structural elements (Bouritsas, Frasca, Zafeiriou, & Bronstein, 2022; Zeng et al., 2023), overlooking their functional relevance in molecular contexts, or treat them as external virtual nodes (Han et al., 2023; Zang, Zhao, & Tang, 2023) that modify the original graph, thus limiting their universality. A universal GNN framework should integrate molecular motifs in a unified manner, accounting for all their roles while maintaining compatibility with any GNN architecture. Additionally, the vast and complex chemical space gives rise to a wide variety of molecule motifs. Existing works tend to fall short of a systematic analysis for the diversity of motifs (Yu & Gao, 2022; Zhang, Liu, Wang, Lu, & Lee, 2021). Given these limitations, the challenging questions arise: *(1) Can we design a universal GNN framework that flexibly integrates motif representations, adaptable to any message-passing mechanism? and (2) How to comprehensively capture and analyze the diverse functional motifs to improve the GNN performance on molecular graphs?*

To this end, this paper proposes a universal yet expressive graph representation method, named Uni-Motif, focusing on unifying molecular motifs to enhance the ability to distinguish non-isomorphism graphs. Given the unique role of motifs in molecular graphs, we propose decoupling motifs into two aspects: *structural encodings*, representing local graph structural context and *functional encodings*, reflecting unique identifiers within molecules. The decoupling enhances the distinguishing ability of node features and enables to capture the subtle differences around the nodes, so as to improve the expressive power without modifying graph convolution strategies. To fully unleash the potential of GNNs in distinguishing non-isomorphic graphs, our method includes dual workflows: one for the message-passing mechanism and the other for global attention. Theoretical derivations confirm that our method increases the expressiveness of GNNs. To give an in-depth analysis on diverse motifs, we present seven motif extraction methods under three categories: topology-driven, ring-driven, and rule-driven, with corresponding embedding schemes. Experimental results highlight the competitive performance of Uni-Motif, showcasing its expressiveness and universality. Our main contributions are as follows:

- We propose a universal and expressive graph representation method that enhances the ability to distinguish non-isomorphic graphs by decoupling molecular motifs into structural encodings and functional encodings without relying on complex graph convolution strategies.
- We devise tailored methods for integrating both functional and structural encodings into arbitrary GNNs, along with a method to enhance Graph Transformers with functional encodings. Ablation studies show these motif integration approaches significantly improve model performance.
- We introduce and evaluate seven motif extraction techniques from topology-driven, ring-driven, and rule-driven, which fill the gap in the comprehensive analysis of diverse motifs. Experiments on various motif extraction methods reveal that the ring-driven has a more pronounced effect on model performance.
- Our method achieves superior performance with extensive experiments on benchmark molecular graph datasets.

2. Related work

2.1. Message passing and 1-Weisfeiler–Lehman (1-WL) graph isomorphism test

Most traditional GNNs are based on the message-passing mechanism, namely message-passing neural networks (MPNNs), where node features are updated iteratively by aggregating information from their neighbors. This local aggregation scheme is central to popular models like Graph Convolutional Network (GCN) (Kipf & Welling, 2016), Graph Isomorphism Network (GIN) (Xu et al., 2018) and Principal Neighbourhood Aggregation (PNA) (Corso, Cavalleri, Beaini, Liò, & Veličković, 2020). These models iteratively pass information between neighboring nodes to learn node representations. The expressive power of standard GNN models is compared to the 1-WL graph isomorphism test, also known as the color refinement algorithm (Huang & Villar, 2021). The 1-WL test distinguishes non-isomorphic graphs by iteratively refining node labels based on the labels of neighboring nodes. GNNs with standard message-passing mechanisms are shown to have a similar expressive power to the 1-WL test (Morris et al., 2019), meaning they can only distinguish between graphs that the 1-WL test can differentiate. As a result, these models fail to capture more complex graph structures (Zhao et al., 2021), such as those involving higher-order dependencies, which are known as expressiveness limitations.

2.2. Graph positional encoding and graph transformer

Recent advancements in graph positional encoding (PE) have highlighted various innovative methods to represent node positions effectively. Key approaches include graph Laplacian eigenvectors, which serve as an effective analog to sinusoidal functions in traditional transformers (Kreuzer, Beaini, Hamilton, Létourneau, & Tossou, 2021; Rampášek, Galkin, Dwivedi, Luu, Wolf, & Beaini, 2022), and alternative techniques such as shortest-path distances (Ying et al., 2021) and random walk encodings (Dwivedi et al., 2021). Concurrently, Graph transformers (GT) have gained traction for addressing challenges like over-smoothing in MPNNs (Liu et al., 2024; Min et al., 2022). The notable architecture utilizes relative positional encodings based on pairwise graph distances and node degree centralities (Ying et al., 2021). And architectures like GraphGPS (Rampášek et al., 2022) and GPS++ (Masters et al., 2023) incorporate various positional encodings to improve performance on complex graph structures. Other prominent GT models consider subgraphs as structural encodings including DeepGraph (Zhao et al., 2023), SAT (Chen, O'Bray, & Borgwardt, 2022), PatchGT (Gao, Han, Huang, Wang, & Liu, 2022). These innovations illustrate the growing synergy between PE and GT, enabling more sophisticated representations and improved performance on graph-structured data.

2.3. Subgraph-based graph neural networks

An increasing number of studies focus on AI-driven advancements in biomedical research (Li et al., 2024, 2025). For graph machine learning, subgraphs play a crucial role in graph representation learning (Alsentzer, Finlayson, Li, & Zitnik, 2020). Some studies use subgraphs to capture the underlying semantic distribution in graph space (Yin, Shen, Chen, Hua, & Luo, 2024), while others incorporate them

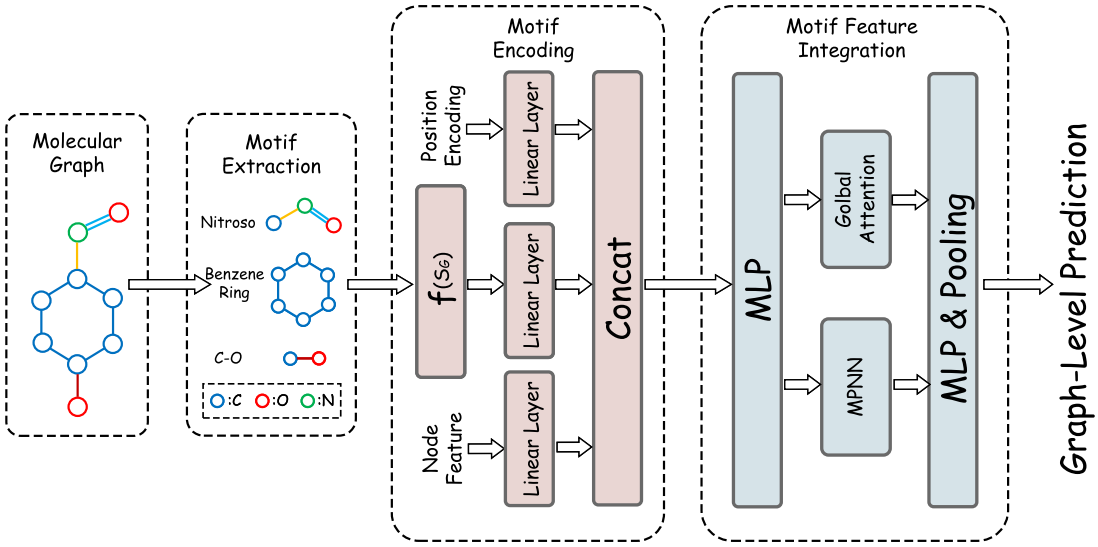


Fig. 1. Illustrations for the whole framework of Uni-Motif. Given the molecular graph, the first step is to decompose it and construct motif vocabulary. Then a corresponding motif encoder is employed to embed the diverse motifs into vector representations, which are used to enrich the node features as well as node positional information. The next step integrates the motif representations into dual workflows related to message-passing mechanisms and global graph attention. The final step readouts the molecular graph representation to perform graph-level tasks, such as property prediction.

as hierarchical structures within graphs (Gu, Luo, Chen, Deng, & Lai, 2023). Recently, subgraph-based models have emerged as a promising approach to enhance the expressiveness of GNNs by treating graphs as collections of subgraphs generated through various strategies (Bevilacqua, Eliasof, Meirom, Ribeiro, & Maron, 2023; Han et al., 2023; Zhao et al., 2021). Some studies leverage decomposed subgraph representations to enrich final graph-level features (Frasca, Bevilacqua, Bronstein, & Maron, 2022; Yu & Gao, 2022). By employing diverse graph decomposition methods, these models can effectively identify and count fundamental substructures (Bouritsas et al., 2022; Zhao et al., 2021). Recent advancements have investigated different subgraph generation policies and their implications for graph learning (Wang & Zhang, 2021; Zeng et al., 2023), demonstrating that these models often outperform traditional MPNNs by utilizing the additional structural information provided by subgraphs. Notably, HiMol (Zang et al., 2023), HimGNN (Han et al., 2023) and FragNet (Wollschläger, Kemper, Hetzel, Sommer, & Günnemann, 2024) fragment molecular graphs into subgraph sets based on chemical properties and reconstruct a hierarchical graph. FragNet (Wollschläger et al., 2024) has theoretically established that GNN models designed around constructed hierarchical graphs exhibit greater expressive power in distinguishing non-isomorphic graphs. However, unified research on integrating substructures across GNN frameworks is still lacking, along with comprehensive comparisons of various subgraph encoding techniques and their effects on overall model performance.

3. Preliminaries

Given a molecular graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, a graph decomposition function $F(\cdot)$ is employed to generate a collection of motifs, denoted as $\mathcal{S}_\mathcal{G} = F(\mathcal{G}) = \{S_1, S_2, \ldots, S_{|\mathcal{G}|}\}$, where $|\mathcal{G}|$ is the number of motifs in $\mathcal{G}$. Each motif S_i is either a node or multiple nodes with connections. Furthermore, $\mathcal{S}_\mathcal{G}$ is refined into two subsets: $\mathcal{M}_\mathcal{G}$, the set of multi-node subgraph of $\mathcal{G}$, and $\mathcal{N}_\mathcal{G}$, the set of individual node of $\mathcal{G}$. Given a molecule dataset and a motif extraction algorithm, we can then construct a motif vocabulary $\mathcal{P}$, with its size denoted by $|\mathcal{P}|$, representing the total number of motif types in this dataset, which typically varies with the datasets used and the graph decomposition function employed. Similarly, the size of $\mathcal{M}_\mathcal{G}$ and $\mathcal{N}_\mathcal{G}$ will differ across different types of graph decomposition function. For instance, $\mathcal{M}_\mathcal{G} \neq \emptyset$, $\mathcal{N}_\mathcal{G} = \emptyset$ when fragmenting the graph into motifs containing multiple nodes, and $\mathcal{M}_\mathcal{G} \neq \emptyset$, $\mathcal{N}_\mathcal{G} \neq \emptyset$ when a junction node connects two motifs.

4. The proposed method

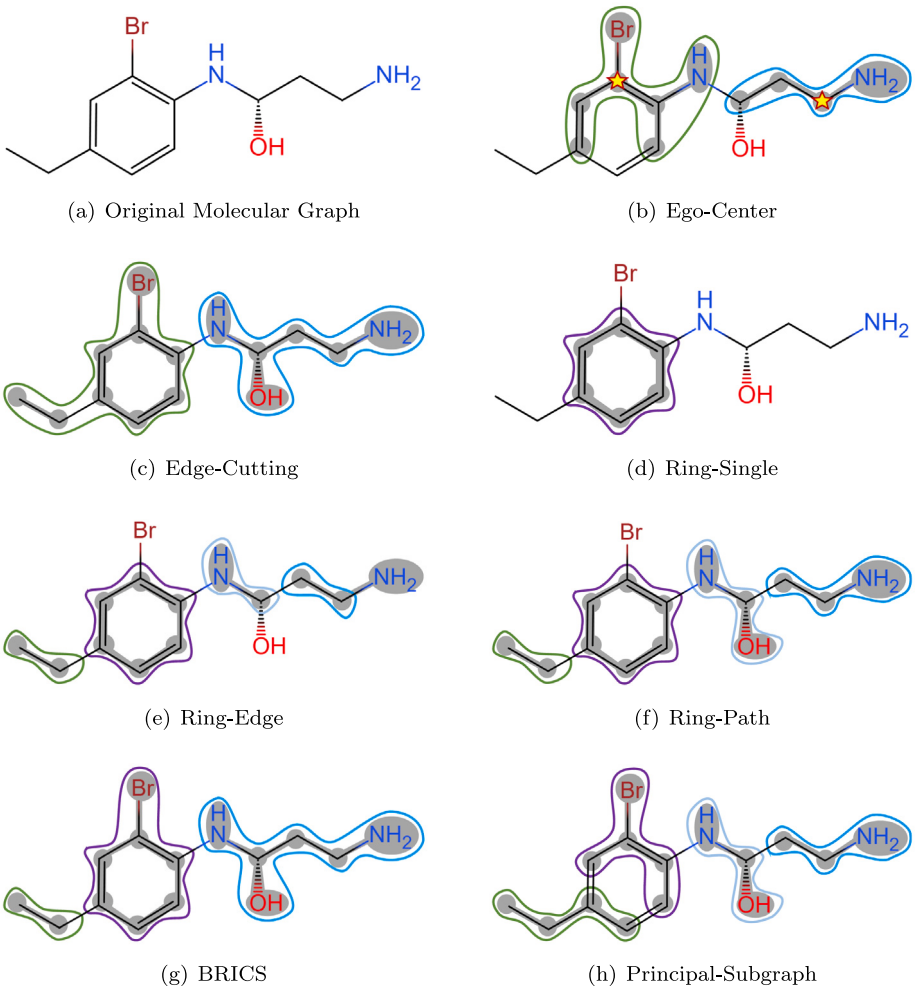
In this section, we introduce our proposed method, Uni-Motif, highlighting how it unifies molecular motif representations in a universal way, leading to a higher expressiveness. The whole framework is shown in Fig. 1.

4.1. Diverse motif extraction

Motifs are potent structural units in molecular graphs that embody predefined topological patterns and functional groups. They offer valuable insights into the graph's structural and chemical properties. Using advanced graph decomposition methods and motif-specific encoding schemes enables more precise representations, enhancing GNNs' ability to distinguish non-isomorphic molecular structures. The following sections provide a comprehensive view of diverse motif extraction techniques and detail the corresponding encoding schemes tailored to each. These schemes enable molecular motifs as unique identifiers and establish a foundation for their effective integration into GNN frameworks.

Overview. The role of a motif varies depending on the diverse motif extraction techniques. Accordingly, we classify the motif extraction techniques into three categories: topology-driven, ring-driven, and rule-driven. Each category emphasizes distinct aspects of motif structure and decomposition. The topology-driven approach aims to capture the molecular graph's topological structure through *Ego-Center*, which focuses on local neighborhoods, and *Edge-Cutting*, which targets global graph structure. The ring-driven approach identifies motifs where rings serve as key structural features, encompassing *Ring-Single*, *Ring-Edge*, and *Ring-Path* techniques. The rule-driven approach leverages predefined rules or domain knowledge about chemical structures, incorporating methods like *BRICS* and *Principal-Subgraph*. Notably, the motif vocabulary constructed in the rule-driven approach relies on both the size of the target dataset and the specific rules applied. Fig. 2 summarizes these techniques.

Ego-Center focuses on a local neighborhood of an anchor node like the yellow star in Fig. 2(b), extracting motifs based on the surrounding nodes and edges. Given the node v, its ego-center motif is defined as a multi-hop subgraph expanded with v, encompassing node v and its neighbor node set $N(v)$. To further extend this concept, the k-hop



(a) Original Molecular Graph (b) Ego-Center

(c) Edge-Cutting (d) Ring-Single

(e) Ring-Edge (f) Ring-Path

(g) BRICS (h) Principal-Subgraph

Fig. 2. Examples of motif decomposition approaches. The given molecular graph (a) is decomposed into motifs, highlighted using solid lines in different colors. Ego-Center (b) extracts k-hop subgraphs centered on yellow star nodes in a local topological view. Edge-Cutting (c) sequentially removes edges based on their computed importance from a global graph perspective. Ring-Single (d) identifies and enumerates stable cyclic structures as motifs. Ring-Edge (e) expands motifs by incorporating edges beyond pure rings. Ring-Path (f) links edges outside rings as paths, offering greater motif diversity compared to Ring-Edge. BRICS (g) segments molecular graphs into functional motifs guided by predefined chemical rules. Principal-Subgraph (h) constructs motifs by analyzing a fixed number of frequent subgraphs in the dataset.

neighbor set centered on node v is denoted as $N_k(v)$, and then the k-hop ego-center motif regarding node v can be denoted as Ego_v^k, which covers all nodes in the graph with a maximum distance of k from node v. As shown in Fig. 2(b), it presents the 2-hop ego-center motif centered on the starred node. For the entire graph $\mathcal{G}$, its motif set $\mathcal{S}_\mathcal{G}$ can be expressed as:

$$\mathcal{S}_\mathcal{G} = \text{Ego}(\mathcal{G})_k = \{\text{Ego}_v^k | v \in \mathcal{V}\}. \tag{1}$$

Previous study (Zhao et al., 2021) leverages GNNs to encode the structural features of motifs in a learnable manner. While effective, the operation inherently increases model complexity and computational costs. To alleviate this issue, we design an efficient representation approach that encodes motif structural information based on the random walk mechanism, which improves flexibility and scalability. Specifically, given a node v and its k-hop ego-center motif Ego_v^k, the corresponding encoding can be formulated as:

$$f(\text{Ego}_v^k) = [\mathcal{R}^1_{\text{Ego}_v^k}(v,v), \mathcal{R}^2_{\text{Ego}_v^k}(v,v), \ldots, \mathcal{R}^d_{\text{Ego}_v^k}(v,v)]^T, \tag{2}$$

where $\mathcal{R}^d_{\text{Ego}_v^k}(v,v)$ denotes the probability of a node v walking back to itself after d steps of traversal within the k-hop ego-center subgraph Ego_v^k.

Edge-Cutting allows us to understand global graph patterns by partitioning the graph into distinct subgraphs based on cutting edges. To enhance the stability and robustness of the model, particularly on high-density regular graphs, we adopt the Edge Betweenness Centrality (EBC) (Girvan & Newman, 2002) to measure the importance of edges. The EBC provides a way to quantify how critical an edge is to the overall connectivity of the graph by evaluating the number of shortest paths that pass through it. The higher the EBC, the more crucial the edge is in terms of maintaining the graph structure. The process of calculating EBC involves two main steps. First, the shortest paths between all pairs of nodes must be determined, which can be efficiently solved using the Floyd–Warshall algorithm. Once these shortest paths are known, they form the basis for calculating the betweenness centrality. After obtaining all the shortest paths, for each edge $e \in \mathcal{E}$, we count how many of these shortest paths pass through e. The betweenness count for an edge increases each time if it is part of the shortest path between two nodes. The raw betweenness values for each edge are typically normalized to provide a relative measure of importance. This is done by dividing each edge's betweenness count by the total number of shortest paths in the graph. The shortest paths of distance 0 or 1 are excluded, as they do not contribute to the betweenness of any edge. Then, we use the calculated EBC values to partition the graph into motifs, which is achieved by repeatedly removing the edge with the highest betweenness centrality. Since edges with higher EBC values are often critical for connecting different parts of the graph, removing them causes the graph to gradually split into more independent motifs. The

process continues until the graph is partitioned into distinct motifs with number $|\mathcal{G}|$ that are represented as:

$$S_{\mathcal{G}} = \text{Cut}(\mathcal{G}) = \{\text{Cut}_1, \text{Cut}_2, \ldots, \text{Cut}_{|\mathcal{G}|}\}. \tag{3}$$

Giving the node $v \in \text{Cut}_i$, we take an analogous encoding scheme with ego-center motifs for edge-cutting, which can be expressed as follows:

$$f(\text{Cut}_i) = [\mathcal{R}^1_{\text{Cut}_i}(v, v), \mathcal{R}^2_{\text{Cut}_i}(v, v), \ldots, \mathcal{R}^d_{\text{Cut}_i}(v, v)]^T. \tag{4}$$

Ring-Single treats individual rings as central motifs for extraction and analysis. So the constructed motif vocabulary only contains rings with various sizes. To extract ring structures from molecular graphs, we focus on identifying the minimal set of rings within the graph, which is achieved by analyzing the circuits (closed loops) within the graph to discern the underlying ring structures. We use the cycle_basis function in the networkx library (Hagberg, Swart, & Schult, 2008) for this task, as it is widely used for detecting cycles in graphs. The cycle_basis function finds a set of independent rings that form a basis for all rings within the graph, ensuring that the minimal ring structures are identified. The collection of ring motifs can be formulated as:

$$S_{\mathcal{G}} = \text{Ring}(\mathcal{G}) = \{\text{Ring}_1, \text{Ring}_2, \ldots, \text{Ring}_{|\mathcal{G}|}\}. \tag{5}$$

For Ring_i, its size is defined as the number of atom nodes it contains. For example, a ring of size 6 would consist of 6 atoms, as illustrated by the ring shown in Fig. 2(d). To encode ring motifs without introducing learnable parameters, we employ one-hot encoding to represent them as fixed vectors. The motif vocabulary, based on ring structures, has a capacity of $|\mathcal{P}|$, where $|\mathcal{P}|$ denotes the number of distinct ring sizes included in the vocabulary. The size of the largest ring in the vocabulary is $|\mathcal{P}|+2$, since the smallest ring has a size of 3. Thus, the vocabulary can encode rings starting from size 3 up to size $|\mathcal{P}|+2$. This makes the size $|\mathcal{P}|$ of the motif vocabulary variable, as it depends on the number of different ring sizes included. Consequently, the encoded feature vector for a ring motif has a dimension of $|\mathcal{P}|$, which corresponds to the number of distinct ring sizes in the vocabulary. If the size of Ring_i is l, the vector will have a value of 1 at the $(l-2)$-th position (since the smallest ring size is 3, we use $l-2$ as the index), while all other positions are set to 0.

Ring-Edge extends the motif vocabulary beyond pure ring structures by incorporating additional edges from the rest of the graph. This approach augments the motif set by integrating edge tokens along with the detected rings, creating a richer and more expressive representation of the molecular graph. Given a molecular graph $\mathcal{G}$, the ring-edge motif collection $S_{\mathcal{G}}$ can be represented as:

$$S_{\mathcal{G}} = \text{RingEdge}(\mathcal{G}) = \{\text{Ring}_1, \text{Ring}_2, \ldots, \text{Ring}_m, \text{Edge}_1, \text{Edge}_2, \ldots, \text{Edge}_{|\mathcal{G}|-m}\}, \tag{6}$$

where m is the number of ring motifs, and $|\mathcal{G}|-m$ is the number of edge motifs.

For a ring motif of size l, the one-hot encoded feature vector will have a value of 1 at the $(l-2)$-th position, with the remaining positions set to 0. For an edge motif, the $|\mathcal{P}|$-th position (the last index in the vector) is assigned a value of 1, with all other dimensions set to 0. This ring-edge encoding embeds both rings and individual edges within the same fixed-dimensional feature space, ensuring a unified representation.

Ring-Path expands upon the *Ring-Edge* by incorporating paths into the motif set. While *Ring-Edge* centers on individual edges, *Ring-Path* captures motifs representing paths, which can consist of one or multiple interconnected edges, forming more complex structures. A path motif can either represent a single edge (similar to *Ring-Edge*) or multiple edges connected through common nodes, forming linear substructures, as depicted in Fig. 2(f). The motif set $S_{\mathcal{G}}$ based on *Ring-Path* is expressed as:

$$S_{\mathcal{G}} = \text{RingPath}(\mathcal{G}) = \{\text{Ring}_1, \text{Ring}_2, \ldots, \text{Ring}_m, \text{Path}_1, \text{Path}_2, \ldots, \text{Path}_{|\mathcal{G}|-m}\}, \tag{7}$$

where $|\mathcal{G}|-m$ corresponds to the number of paths, including both single edges and longer edge chains.

The encoding of *Ring-Path* builds upon the *Ring-Edge*, with modifications to accommodate path structures. In the *Ring-Edge* approach, each edge is represented by a single token. However, *Ring-Path* requires multiple dimensions to represent paths, as they can consist of multiple interconnected edges. To capture this variety, the vocabulary size is adjusted to account for paths of different lengths. For a ring motif of size l, the one-hot encoded feature vector remains the same as in the *Ring-Edge*, with 1 at the $(l-2)$-th position. For a path motif, defined by the number of edges it contains, the encoding is based on the motif's rank k (in descending order of path lengths). The corresponding feature vector for a path motif will thus be 1 at the $(|\mathcal{P}|-k+1)$-th position.

BRICS is a rule-driven approach for fragmenting molecular graphs into functional groups. It utilizes the Breaking of Retrosynthetically Interesting Chemical Substructures (BRICS) algorithm (Degen, Wegscheid-Gerlach, Zaliani, & Rarey, 2008), which breaks chemical bonds based on a predefined set of 16 chemical rules. This approach is particularly effective for identifying key functional motifs within molecular structures, as it focuses on bonds that are most relevant to retrosynthetic analysis. Given molecular graph $\mathcal{G}$, the motif set of BRICS-derived is represented as:

$$S_{\mathcal{G}} = \text{BRICS}(\mathcal{G}) = \{\text{B}_1, \text{B}_2, \ldots, \text{B}_{|\mathcal{G}|}\}. \tag{8}$$

The *BRICS*-based motif vocabulary $\mathcal{P}$ is generated by analyzing and statistically quantifying all functional motifs across the dataset. Once the vocabulary is established, each motif is represented as a one-hot encoded vector with dimensionality equal to the vocabulary size $|\mathcal{P}|$. Each motif is assigned a unique position in the vector. If a motif is present in the molecule, its corresponding position in the feature vector is set to 1, while all other positions are 0.

Principal-Subgraph is inspired by the Byte Pair Encoding (BPE) algorithm, commonly used in word segmentation. Unlike the top-down BRICS approach, *Principal-Subgraph* builds a vocabulary of motifs using a bottom-up strategy: it begins with individual atoms and iteratively merges them to form increasingly larger motifs (Kong, Huang, Tan, & Liu, 2022). Initially, the vocabulary $\mathcal{P}$ consists of all atom types. At each step, pairs of adjacent atom nodes are merged to create larger subgraphs, with the frequency of each newly formed subgraph calculated across the dataset. The most frequent subgraph is then added to the vocabulary. With each iteration, the graph $\mathcal{G}$ is segmented into distinct motifs based on the current vocabulary, which is further refined through repeated merging of the most frequent motifs. This process continues until the vocabulary reaches the predefined size $|\mathcal{P}|$, ensuring it captures the most representative and prevalent motifs within the molecular graphs.

Once the vocabulary $\mathcal{P}$ is established, each motif is represented as a one-hot encoded vector with a dimensionality equal to the vocabulary size $|\mathcal{P}|$. Each motif is assigned a unique position within this vector.

4.2. Universal Motif integration

To ensure framework-agnostic, a method should not rely on any specific GNN architecture. To this end, we integrate motifs into the message-passing mechanism and the process for calculating global attention scores without modifying other parts.

In general GNNs, the message-passing mechanism consists of two main components: the *aggregation* function and *update* function. The aggregation function collects messages from a node's neighbors, enabling the acquisition of the local topology. Once messages from neighbors are aggregated, the update function combines the node's features with

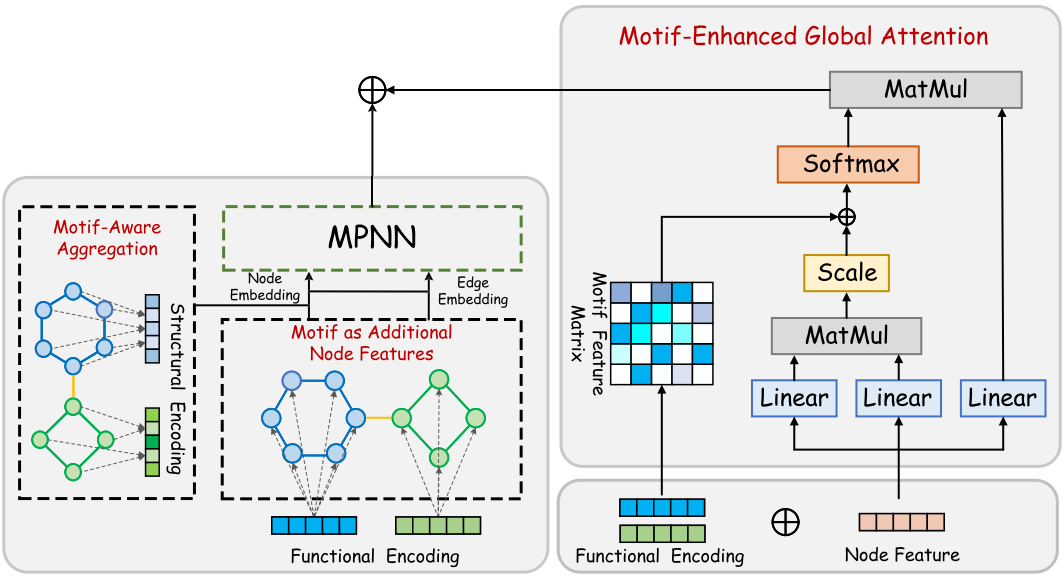


Fig. 3. The details of integrating motif representations into GNN framework. *Motif as Additional Node Features:* Functional encodings derived from motifs are incorporated as additional node features, enriching node embeddings by leveraging the unique roles of motifs. *Motif-Aware Aggregation:* Node embeddings are aggregated within the same motif at each layer, using motifs as structural identifiers to enhance the distinguishability of nodes. *Motif-Enhanced Global Attention:* Motif functional encodings are used as a bias term, integrating them into the attention matrix to improve motif-level relevance between node pairs.

the aggregated neighbor messages to generate updated features. Consequently, the message-passing mechanism refines each node's features by integrating both its state and the gathered neighbor state. In contrast, the global graph attention mechanism provides an alternative to the local message passing, allowing each node to attend to every other node in the graph rather than only their immediate neighbors. This mechanism computes weighted attention scores between each node pair, where the weights determine the extent to which a node should incorporate information from other nodes.

To devise a universal motif integration approach that is compatible with various GNN frameworks, it is essential to avoid modifying the aggregation and update functions or the calculation of global attention scores. Instead, motif features should work in parallel with the existing GNN operations. To achieve this, we propose decoupling molecular motifs into functional and structural encodings and present tailored motif integration approaches: *Motif as Additional Node Feature*, *Motif-Aware Aggregation*, *Motif-Enhanced Global Attention*. Details are illustrated in Fig. 3.

Motif as Additional Node Feature. Since the motifs within molecules carry specific chemical semantics that serve as unique identifiers, we consider enriching the original node representations via functional encodings. After extracting and embedding motifs, the resulting motif feature vectors serve as functional encodings for the nodes within those motifs. We start by mapping both node embeddings and motif representations into a shared hidden space using separate learnable linear transformations. The node embeddings and motif features are then concatenated to form the initial node representation. This combined representation enables the model to harness both low-level node features and high-level motif information during message passing. Acknowledging the significance of positional encoding in enhancing GNN expressiveness, node embeddings also consist of random walk positional encoding (RWPE). The initialization of node v_i is illustrated as follows:

$$h_{v_a}^{(0)} = W_a a_v + b_a, \tag{9}$$

$$h_{v_s}^{(0)} = \{h_{s_i}^{(0)} | v_i \in S_i\} = W_s f(S_i) + b_s, \tag{10}$$

$$h_{v_r}^{(0)} = W_r r_v + b_r, \tag{11}$$

$$h_v^{(0)} = h_{v_a}^{(0)} \oplus h_{v_r}^{(0)} \oplus h_{v_s}^{(0)}. \tag{12}$$

Here, v_a represents the node attribute, such as atomic type, charge number, etc. r_v denotes the positional encoding of v_i. The function $f(S_i)$ provides the motif encoding, and $\oplus$ denotes the concatenation of feature vectors. W_a, W_r, W_s and b_a, b_r, b_s are the trainable weights and bias terms. $h_{v_a}^{(0)}$, $h_{v_s}^{(0)}$, $h_{v_r}^{(0)}$ represent the feature vectors of node attributes, motif encodings, and position encodings in a shared hidden space, respectively. Eqs. (10) and (12) show that if node v_i is part of motif S_i, the motif representation $h_{v_s}^{(0)}$ is integrated into $h_v^{(0)}$. This approach enhances model expressiveness by capturing both local structure through node embeddings and high-level graph patterns through motifs.

Motif-Aware Aggregation. To capture the structural context between nodes in the same motif, we design a module that iteratively integrates the structural encodings of motifs:

$$h_{S_i}^{(l)} = \text{MLP}(h_{S_i}^{(l-1)} \oplus \text{AGG}(h_v^{(l-1)} | v \in S_i)). \tag{13}$$

Here, $h_{S_i}^{(l)}$ is the updated structural encoding of motif S_i in the lth graph convolutional layer. MLP$(\cdot)$ represents a multi-layer perceptron with a non-linear transformation, and AGG$(\cdot)$ is the local aggregation function used to integrate node features within the motif S_i, such as summing and averaging. The learnable structural encoding is updated based on both the structural representation from the previous layer and the updated node features in the current layer. Eq. (13) details the calculation process of updating motif structural representation, which is subsequently integrated into the MPNNs. For a node v belonging to motif S_i, we propose combining the structural encoding of motif S_i with the corresponding node embedding at each layer, which is then used to update the edge embedding. This approach allows MPNNs to maintain the use of nodes and edges as inputs without necessitating any change to the message-passing framework:

$$h_{v_M}^{(l-1)} = \text{MLP}(h_v^{(l-1)} \oplus \text{AGG}(h_{S_i}^{(l)} | v \in S_i)), \tag{14}$$

$$h_{e_M}^{(l-1)} = \text{MLP}(h_e^{(l-1)} \oplus h_v^{(l-1)} \oplus h_u^{(l-1)} | e \in (u, v)), \tag{15}$$

$$h_{v_M}^{(l)}, h_{e_M}^{(l)} = \text{MPNN}^{(l)}(h_{v_M}^{(l-1)}, h_{e_M}^{(l-1)}), \tag{16}$$

where $h_{v_M}^{(l)}$ and $h_{e_M}^{(l)}$ denote the involved node and edge embeddings from the MPNN branch at the lth layer. For any node v, AGG$(h_{S_i}^{(l)} | v \in S_i)$ indicates that the structural encodings encompassing all motifs of node

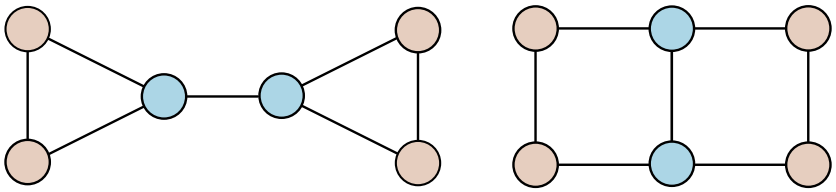


Fig. 4. An example of non-isomorphic graphs that are 1-WL indistinguishable.

v will be aggregated. If edge e connects node v and node u, Eq. (15) combines the edge embedding $h_e^{(l-1)}$ with the node embeddings $h_v^{(l-1)}$ of v and $h_u^{(l-1)}$ of u.

Motif-Enhanced Global Attention. To improve the model's ability to perceive not only node-to-node relationships but also motif-level relevance between node pairs, we devise Motif-Enhanced Global Attention, incorporating functional encodings as a bias term in the attention mechanism. For each node pair (u, v), we define a node-pair feature matrix M that encodes relative motif identifiers. Each element $M_{(u,v)}$ in the matrix represents the structural relationship between nodes u and v based on the motifs they are associated with. Let $\mathcal{S}_u$ and $\mathcal{S}_v$ denote the motifs containing nodes u and v, respectively. If an edge $e_{(u,v)} \in E$ exists between nodes u and v in the graph, the motif encodings of $\mathcal{S}_u$ and $\mathcal{S}_v$ are aggregated to capture the motif-level relationship between them. If no edge exists between nodes u and v, the feature value $M_{(u,v)}$ is set to 0, indicating an absence of direct motif relevance, which is as follows:

$$M_{(u,v)} = \begin{cases} \text{MLP}(\sum h_{\mathcal{S}_i}), & \text{if } e_{(u,v)} \in E \\ 0, & \text{otherwise,} \end{cases} \tag{17}$$

where $\mathcal{S}_i \in \{\mathcal{S}_u \cup \mathcal{S}_v | u \in \mathcal{S}_u, v \in \mathcal{S}_v\}$. The trainable linear function $\text{MLP}(\cdot)$ projects this node-pair representation to meet the required feature dimensions, serving as the bias term of attention scores matrix:

$$h_{v_T}^{(l)} = \text{GlobalAttn}^{(l)}(h_v^{(l-1)}, \text{MLP}(M^{(l-1)}). \tag{18}$$

$\text{GlobalAttn}^{(l)}$ calculates the node embedding $h_{v_T}^{(l)}$ from the branch of global attention at the lth layer. As shown in Eq. (18), the core mechanism of the global attention module remains unchanged.

4.3. Graph-level prediction and model optimization

Since Uni-Motif maximizes the topological perception of molecular graphs via dual workflows, it is necessary to fuse the outputs of MPNN and the global attention network at each layer. The fusion ensures that the local and global structural information are effectively combined at each iteration, enriching node representations for subsequent layers:

$$h_v^{(l)} = \text{MLP}(h_{v_M}^{(l)} + h_{v_T}^{(l)}). \tag{19}$$

$h_v^{(l)}$ is the fused node embedding from dual workflows. After the L layer, a pooling operation is applied to obtain the final graph representation $h_{\mathcal{G}}$:

$$h_{\mathcal{G}} = \text{POOL}(h_v^{(L)} | v \in \mathcal{G}). \tag{20}$$

The optimization of training model is tailored to the specific task, whether it involves graph regression or classification. For regression tasks, which aim to predict a continuous value (e.g., molecular properties like energy), the model minimizes the L1 loss, also known as mean absolute error (MAE). For classification tasks, where the objective is to assign graphs to discrete categories (e.g., predicting whether a molecule is active or inactive), the model minimizes cross-entropy loss.

4.4. Complexity analysis

Uni-Motif's computational complexity is determined by the applied MPNN and the attention mechanism. The identification and extraction of motifs can be pre-computed, decoupling this intensive step from the GNN's runtime. Since motif features remain fixed after pre-computation, no additional complexity is introduced during training or inference. The Motif-Aware Aggregation module aggregates node embeddings within motifs, contributing a complexity of $O(k)$, where k is the size of the largest motif. Combined with the GNN computation, the total complexity during training and inference is $O(|E| + k)$, where $O(|E|)$ arises from the GNN. Given that $k \ll |E|$, as motifs represent sparse higher-order substructures relative to the graph size, the overhead introduced by motifs is negligible as graph size grows.

5. Theoretical analysis on the expressive power

Note that the Uni-Motif framework consists of an MPNN and global attention, where the MPNN achieves the maximum expressiveness (1-WL) when its aggregation, update, and pooling functions are injective (Xu et al., 2018). The following proposition shows that Uni-Motif can achieve a higher expressive power than the applied MPNN.

Proposition 1. *Uni-Motif that integrates motif encodings is strictly more powerful than the applied MPNN.*

Proof. Given two non-isomorphic graphs $\mathcal{G}_1$ and $\mathcal{G}_2$. Assume there exists a motif difference between these two graphs. Let $\mathcal{F}(\cdot)$ denote an MPNN's output that aggregates neighbor information as follows:

$$h_v^{(l+1)} = \mathcal{F}(h_v^{(l)}, \{h_u^{(l)}\} : u \in N(v)). \tag{21}$$

With motif encodings, the node features become:

$$h_v^{(l+1)} = \mathcal{F}(h_v^{(l)} + Func(v) + Struc(v), \{h_u^{(l)} + Func(u) + Struc(u)\} : u \in N(v)), \tag{22}$$

where $Func(v)$ and $Struc(v)$ are functional encodings and structural encodings. Since decoupled motif encodings $Func(v)$ and $Struc(v)$ reflect the nodes' unique motif roles, the function $\mathcal{F}(\cdot)$ now distinguishes between nodes based on both their features and motif characteristic in the graph. Thus, the model can distinguish between previously indistinguishable nodes (and graphs), proving an increase in the expressive power of the MPNN. Furthermore, Uni-Motif is strictly more powerful than 1-WL when MPNN is 1-WL expressive. □

The two non-isomorphic graphs in Fig. 4, which differ only in the presence or absence of a triangle, are indistinguishable by 1-WL and standard GNNs. Uni-Motif can easily differentiate them by explicitly enriching the node embeddings with triangle representations.

6. Experiments

In this section, we embark on extensive experiments to assess Uni-Motif's performance in predicting molecular properties. We aim to answer the following pivotal questions: **RQ1:** Can Uni-Motif perform

Table 1
Hyperparameter settings for Uni-Motif.

Para. type	Hyperparameter	ZINC	MolHIV	MolPCBA
Model	Layers	10	10	5
	Hidden Dimensions	64	64	384
	Num of Attention Heads	4	4	4
	Readout	mean	mean	mean
	Node PE	RWPE-20	RWPE-16	RWPE-16
Motif Extraction	Hop of Ego-center	3	3	3
	Num of Edge-cut	4	4	4
	Max Size of Ring	8	8	8
	Size of Principal Subgraph Vocabulary	100	300	300
Training Process	Learning Rate	1e−3	1e−4	5e−4
	Optimization	AdamW	AdamW	AdamW
	Batch Size	32	64	256
	Epoch	2000	100	100
	Warmup Steps	50 epochs	5 epochs	5 epochs

Table 2
Results on ZINC, MolHIV, and MolPCBA datasets, where the best results are in bold, and second-best are underlined.

Method	ZINC MAE ↓	MolHIV ROC–AUC ↑	MolPCBA AP(%) ↑
MPNN Models			
GCN	0.367 ± 0.011	0.7599 ± 0.0119	24.24 ± 0.34
GIN	0.526 ± 0.051	0.7707 ± 0.0149	27.03 ± 0.23
GatedGCN	0.282 ± 0.015	–	26.20 ± 0.10
PNA	0.188 ± 0.004	0.7905 ± 0.0132	28.38 ± 0.35
PR-MPNN	0.085 ± 0.002	0.7950 ± 0.0090	–
Subgraph-Enhanced GNNs			
GSN	0.101 ± 0.010	0.8039 ± 0.0090	–
GIN-AK+	0.080 ± 0.001	0.7964 ± 0.0119	29.30 ± 0.44
SAPNA	0.073 ± 0.001	0.7944 ± 0.0144	27.84 ± 0.03
GNN-SSWL+	0.070 ± 0.005	0.7958 ± 0.0035	–
FragNet	0.077 ± 0.005	–	–
Graph Transformer			
SAN	0.139 ± 0.006	0.7785 ± 0.2470	27.65 ± 0.42
Graphormer	0.122 ± 0.006	–	–
GraphGPS	0.070 ± 0.004	0.7880 ± 0.0101	29.07 ± 0.28
Gradformer	0.069 ± 0.002	0.7915 ± 0.0089	–
Subgraph-Enhanced Graph Transformer			
k-ST SAT	0.094 ± 0.008	–	–
DeepGraph	0.072 ± 0.004	–	–
GraphTrans-MLP-Mixer	0.077 ± 0.003	0.7969 ± 0.0061	–
Uni-Motif (ours)	**0.062 ± 0.001**	**0.8072 ± 0.0053**	**29.37 ± 0.30**

well across different molecular datasets and tasks? **RQ2:** How does the decoupled motif representation improve the model's expressiveness? **RQ3:** What is the impact of using diverse motif extraction methods? **RQ4:** How does motif integration affect the model performance? **RQ5:** Can the proposed motif integration approach be universally applied across various GNN architectures?

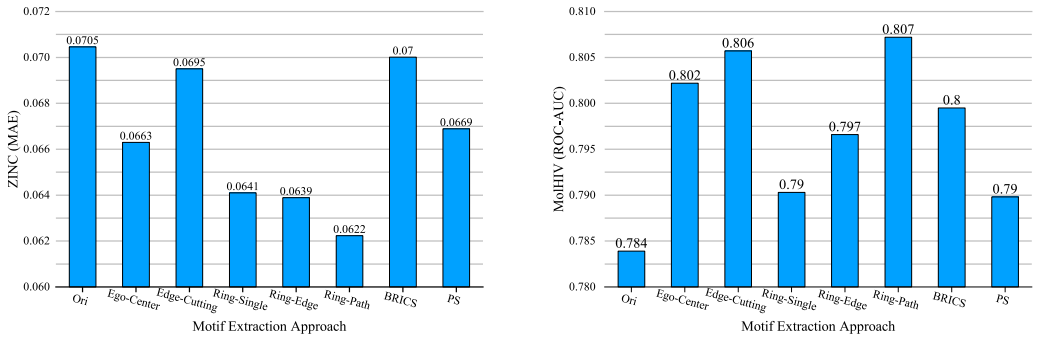
6.1. Datasets

We use three publicly available molecular datasets which are ZINC-10K (Dwivedi et al., 2023), MolPCBA and MolHIV from OGB (Hu et al., 2020) to evaluate the performance of proposed methods.

- ZINC-10K dataset, namely ZINC here, comprises 12,000 molecular instances, with solubility as the target property, measured by the logP value. We follow the 10:1:1 training, validation, and testing split defined by the previous work (Dwivedi et al., 2023).
- MolPCBA dataset contains 437,929 molecular graphs, where the objective is to predict binary active/inactive labels across 128 bioassays. We utilize the scaffold-based split with a ratio of 8:1:1.
- MolHIV dataset is used for predicting a binary label indicating whether a molecule inhibits HIV replication. It contains 41,127 molecular graphs and a scaffold split with a ratio of 8:1:1 is employed for training, validation, and testing.

6.2. Baselines

The baseline models used for comparisons include MPNN rooted methods: GCN (Kipf & Welling, 2016), GIN (Xu et al., 2018), GatedGCN (Bresson & Laurent, 2017), PNA (Corso et al., 2020), PR-MPNN (Qian et al., 2024). Subgraph-based GNN baselines include GSN (Bouritsas et al., 2022), GIN-AK+ (Zhao et al., 2021), SAPAN (Zeng et al., 2023), GNN-SSWL+ (Zhang, Feng et al., 2023), FragNet (Wollschläger et al., 2024). We also choose recent state-of-the-art Graph Transformer models: SAN (Kreuzer et al., 2021), Graphormer (Ying et al., 2021), K-ST SAT (Chen et al., 2022), GraphGPS (Rampášek et al., 2022), as well as subgraph-based Graph Transformers: DeepGraph (Zhao et al., 2023), GraphTrans-MLP-Mixer (He et al., 2023),



(a) On the ZINC dataset (b) On the MolHIV dataset

Fig. 5. Performance of different motif extraction approaches.

Gradformer (Liu et al., 2024).

6.3. Experimental setups

All experiments are conducted on NVIDIA GeForce GTX3090 GPU servers. The hyperparameter settings are summarized in Table 1. All scores are averaged over 4 runs with 4 different random seeds.

6.4. Evaluation metrics

ZINC dataset is a graph regression dataset, evaluated using Mean Absolute Error (MAE). MolPCBA is a multi-task binary graph classification dataset with a significant class imbalance, as positive samples make up only 1.4% of the dataset. To address this, Average Precision (AP) is used as the evaluation metric. MolHIV is a binary graph classification dataset, evaluated with the Receiver Operating Characteristic-Area Under the Curve (ROC–AUC) metric.

6.5. Experimental results and discussion

To answer RQ1, we conduct comprehensive experiments across three molecular graph datasets of varying scales and tasks. The detailed results are summarized in Table 2. By comparing these outcomes with baseline methods, we observe Uni-Motif's unique advantages and substantial potential. Uni-Motif achieves notable performance on all datasets, highlighting its capability as a universal framework that significantly enhances the expressiveness of existing GNNs. Specifically, Uni-Motif attains an optimal MAE of 0.062 on the ZINC dataset, reducing prediction errors by over 10% compared to the second-best baseline, which underscores its precision in molecular property prediction tasks. Additionally, Uni-Motif performs strongly on molecular classification datasets. On the MolHIV dataset, it achieves a top-ranked ROC–AUC score of 0.8072, indicating high accuracy in complex bioactivity prediction. Similarly, on the MolPCBA dataset, Uni-Motif achieves the best performance with an average precision of 0.2937, further validating its robustness in handling multi-label classification with significant class imbalance.

Notably, Uni-Motif achieves these exceptional results without altering the fundamental message-passing mechanism or graph attention architecture of the underlying GNNs. By effectively integrating functional encodings and structural motif encodings, Uni-Motif retains the model's efficiency and flexibility while significantly enhancing its predictive performance. This demonstrates that, without requiring major modifications to the core GNN components, Uni-Motif strengthens the expressive power and competitiveness of GNNs in molecular graph-level tasks.

Table 3
Ablation study of the role of motifs. The best results are in bold.

Method	ZINC MAE ↓	MolHIV ROC–AUC ↑	MolPCBA AP(%) ↑
Uni-func	0.071 ± 0.002	0.7519 ± 0.0129	28.58 ± 0.18
Uni-struc	0.064 ± 0.005	0.7860 ± 0.0138	29.35 ± 0.23
Uni-Motif (Full)	**0.062 ± 0.001**	**0.8072 ± 0.0053**	**29.37 ± 0.30**

6.6. Ablation study

In this section, we design a series of ablation experiments to thoroughly assess the effectiveness of each component of the Uni-Motif framework, compare motif extraction techniques, and examine the impact of proposed motif integration approaches.

6.6.1. Effectiveness of decoupled Motif representation

As outlined in Section 4, Uni-Motif enhances GNN expressive power by decoupling motifs into functional encodings and structural encodings. When used as functional encodings, motif features build initial node representations, while as structural encodings, they are updated in each layer to capture local and specific structures. To address RQ2, we design two Uni-Motif variants: Uni-func and Uni-struc. Uni-func uses motifs exclusively as functional encodings, impacting only initial node representations without further updates in the subsequent layers. In contrast, Uni-struc retains motifs purely as structural encodings, excluding them from initial representations but incorporating them in iterative updates. Results are summarized in Table 3.

As shown in Table 3, both Uni-func and Uni-struc underperform compared to the full Uni-Motif model, underscoring the importance of using motifs as both functional encodings and structural encodings to enhance model performance. These variants show consistent performance drops across the three molecular graph datasets, reinforcing our proposition that combining motifs as functional encodings and structural encodings increases model expressiveness. This improvement arises mainly from two aspects. First, conventional GNNs struggle to distinguish diverse structural roles of nodes, often missing structural properties like triangles or rings. By encoding motifs explicitly as part of GNNs, the model can capture these critical structural nuances. Second, iteratively integrating the structural encodings of motifs allows the GNN to better characterize the structural features of each node, encompassing not only node embeddings but also surrounding subgraph patterns and their interrelationships. The refined representation aids GNNs to capture richer information in complex graph structures, as evidenced by the significant performance declines observed across all datasets when structural encodings of motifs are removed.

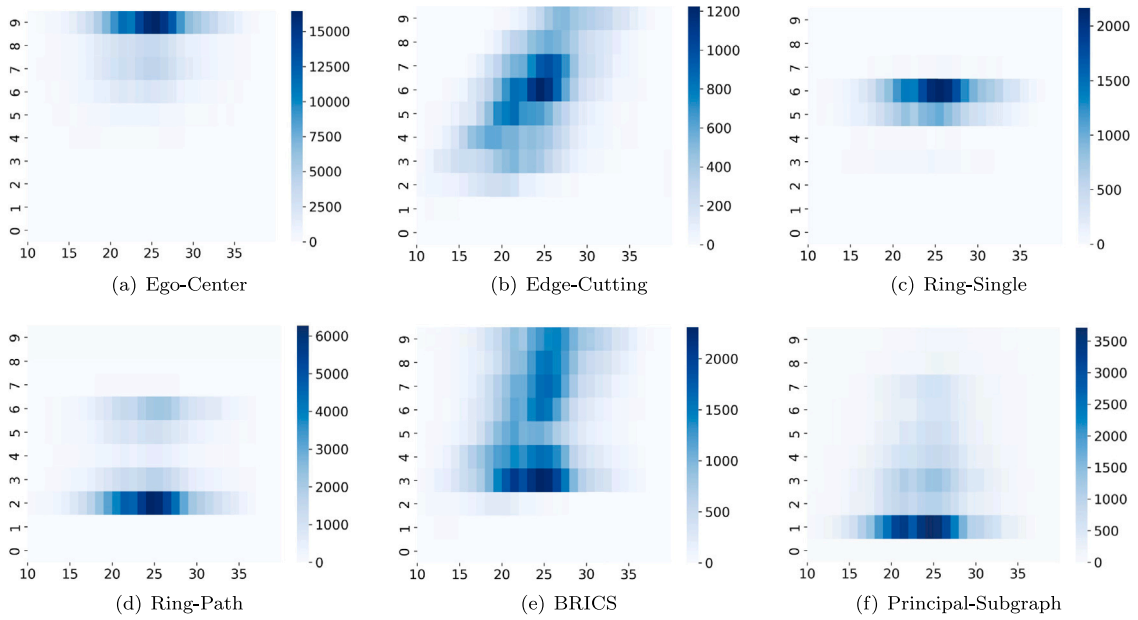


Fig. 6. The motif size distribution of different motif extraction approaches on the ZINC dataset. The x-axis represents the number of atom nodes, and the y-axis is the size of the motifs.

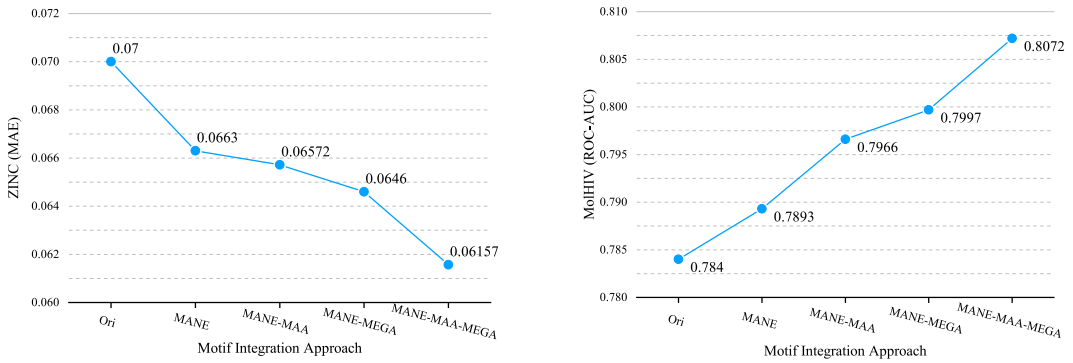
6.6.2. *Effectiveness of Motif extraction*

In response to RQ3, this section systematically investigates how different motif extraction techniques impact GNN model performance on the ZINC and MolHIV datasets. The results, illustrated in Fig. 5, highlight several key findings. First, the incorporation of motifs derived from diverse extraction rules consistently enhances model performance compared to the baseline model without additional motifs (denoted as 'Ori'). This underscores the benefit of motif-based strategies in boosting the expressive power of GNNs. Second, on the ZINC dataset, motif extraction methods focusing on ring structures, such as *Ring-Single*, *Ring-Edge* and *Ring-Path*, demonstrate notably strong performance. These methods take advantage of the ring structures in molecular graphs, which are essential for capturing local topological patterns. Similarly, *Ring-Path* performs exceptionally well on the MolHIV dataset, further affirming the effectiveness of ring-based motif extraction in graph classification tasks where biochemical relevance is high. Ring structures, as key chemical motifs in many biochemical molecules, offer substantial advantages when integrated into GNNs. Serving as distinctive structural features, they enable models to capture essential properties for each node, thereby improving the GNN's expressive capacity. Besides, the stability and functional significance of the ring profoundly influence overall molecular properties. In biochemistry, ring structures often correlate with critical molecular characteristics, such as stability and reactivity, which are crucial for tasks like predicting molecular activity, supporting a more thorough understanding of molecular behavior.

The analysis above examines motif impacts from a structural perspective. To further investigate, we visualize and discuss motif distributions across various extraction methods on the ZINC dataset, as shown in Fig. 6. This comparison offers insights into how different extraction techniques yield motifs of varying sizes and distributions, highlighting their role in capturing molecular complexity. *Ego-Center* associates motifs with multi-hop neighbors around each node, producing relatively large motifs. *Edge-Cutting* generates motif sizes that scale proportionally with molecular size. As molecular graphs grow, the corresponding motifs expand due to their global property. *Ring-Single* focuses solely on ring structures, yielding motifs within a narrow size range, regardless of molecular size. This size consistency suggests that *Ring-Single* specializes in identifying cyclic patterns but it may have limitations in capturing non-cyclic features. *Ring-Path* builds on *Ring-Single* by including both rings and paths with varying edge counts, leading to a broader motif distribution. This diversity reflects both localized (ring) and extended (path) connections. *BRICS* generates fragments by cleaving bonds, creating a wider range of substructure sizes, while *Principal-Subgraph* decomposes molecules by mining frequent subgraphs, with the distribution of motifs influenced by dataset size. *Principal-Subgraph* motif distribution is relatively concentrated, indicating both the count and scale of frequent subgraphs tend to be smaller. *Ring-Edge* builds on *Ring-Single* by only adding one type of edge, thus it is omitted in Fig. 6. Given the characteristics and average molecular graph size of the ZINC dataset, *Ring-Path* achieves a balanced and relatively uniform distribution, as depicted in Fig. 6. This balance likely explains the improved performance *Ring-Path* on the ZINC dataset, indicating that selecting a motif extraction strategy aligned with the dataset's unique properties is critical for model optimization.

6.6.3. *Effectiveness of Motif integration*

To address RQ4, this section evaluates the impact of three motif integration approaches – Motif as Additional Node Feature (MANF), Motif-Aware Aggregation (MAA), and Motif-Enhanced Global Attention (MEGA) – on GNN performance. The analysis is based on ablation experiments conducted on the ZINC and MolHIV datasets, using *Ring-Path* for motif extraction. 'Ori' denotes not using motif integration. As shown in Fig. 7, the experiments reveal several key findings. First, notable performance gains with MANF: Adding the MANF module to the 'Ori' baseline yields a substantial performance improvement, as MANF assigns unique, distinguishable motif features to nodes, enhancing their expressiveness within the GNN framework. Second, further improvements with MAA and MEGA: adding MAA and MEGA independently on top of MANF produces additional performance gains, with both contributing similarly to the enhancement. MAA improves local aggregation by incorporating motif-awareness in message passing, while MEGA strengthens the model's ability to capture global relationships by applying motif-enhanced attention, allowing the network to prioritize critical nodes and motifs across the entire graph. Third, significant gains from combining integration approaches: Applying all three motif integration approaches simultaneously provides a notable performance boost beyond that achieved by each approach alone. Uni-Motif leverages a dual workflow, combining message-passing and global attention mechanisms designed to capture both local and global node relationships within each layer. It incorporates learnable structural encodings



(a) On the ZINC dataset (b) On the MolHIV dataset

Fig. 7. Performance of different motif integration approaches.

Table 4
Ablation study of the universality of the proposed method. The best results are in bold.

Method	ZINC MAE ↓	MolHIV ROC–AUC ↑
GINE(PE)	0.085 ± 0.005	0.7802 ± 0.0155
GINE(PE)-Motif	0.076 ± 0.003	0.7860 ± 0.0110
GatedGCN(PE)	0.167 ± 0.002	0.7722 ± 0.0060
GatedGCN(PE)-Motif	0.091 ± 0.003	0.7776 ± 0.0012
Uni-Motif (GINE)	**0.062 ± 0.001**	0.8010 ± 0.0025
Uni-Motif (GatedGCN)	0.078 ± 0.002	**0.8072 ± 0.0053**

based on motifs at both local and global levels, assigning each node-specific structural attribute from surrounding nodes to enhance the expressive power of GNNs.

6.6.4. Discussion for universality

In reply to RQ5, this section evaluates the universality of the proposed motif integration approach across various GNNs within the Uni-Motif framework, with experimental results summarized in Table 4. Uni-Motif is designed to be framework-agnostic. MANF and MAA interact with different GNNs – specifically GINE and GatedGCN here – by adding functional and structural encodings to node features. This process does not depend on specific message-passing frameworks and can be integrated with both GINE and GatedGCN. MEGA introduces motif bias terms unidirectionally into the global attention matrix, making it suitable for architectures like Graph Transformers that compute pairwise node relationships. In this setup, GINE(PE) and GatedGCN(PE) denote configurations where random walk positional encoding is added to node embeddings within the GINE and GatedGCN frameworks. Extending this, GINE(PE)-Motif and GatedGCN(PE)-Motif incorporates both the MANE and MAA modules. Uni-Motif(GINE) and Uni-Motif(GatedGCN) further adapt GINE and GatedGCN as one branch within the dual-workflow setup, respectively. The findings show that motif integration substantially improves the performance of standard GNN models, even in the absence of graph attention mechanisms. The addition of the MANE and MAA modules further enhances both GINE and GatedGCN by enriching node features with decoupled motif representations, without modifying GNN structures or adding computational overhead specific to motif identification, confirming the method's universality. In the dual-workflow Uni-Motif configuration, which combines GINE and GatedGCN with global graph attention, the approach achieves state-of-the-art performance across all tested datasets. In summary, Uni-Motif not only enhances the expressive power of GNNs but also maintains universality, making it compatible with a range of GNN architectures.

7. Conclusions

In this work, we introduce Uni-Motif, a novel framework designed to enhance the expressive power of GNNs through molecular motif representations while ensuring universality across various GNN architectures. To accomplish this, we propose decoupling motif representations into functional encodings and structural encodings, creating a flexible foundation for motif integration. We present three distinct methods for integrating motifs: Motif as Additional Node Feature, Motif-Aware Aggregation, and Motif-Enhanced Global Attention, each effectively incorporating motif features into both local and global node relationships. Furthermore, we provide a synthesized analysis of seven molecular motif extraction techniques, systematically discussing their impact on model performance. Extensive experiments on diverse molecular datasets demonstrate that Uni-Motif achieves state-of-the-art results, fulfilling both universality and expressiveness for GNN models.

We believe that our work paves the way for several promising future directions. These include the development of transferable motif representations that can enhance the generalization of GNNs for unseen molecules and molecular graph tasks. Additionally, future efforts could focus on extending motif extraction and integration methods to other biological domains.

CRediT authorship contribution statement

Runze Wang: Writing – reviewing & editing, Conceptualization, Methodology. **Yuting Ma:** Data curation, Writing – original draft, Validation. **Xingyue Liu:** Validation, Testing. **Zijie Xing:** Validation, Testing. **Yanming Shen:** Writing – reviewing & editing, Conceptualization, Methodology, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the National Natural Science Foundation of China under Grant 62276044.

Data availability

I have shared the link to my code in my manuscript.

References

Alsentzer, E., Finlayson, S., Li, M., & Zitnik, M. (2020). Subgraph neural networks. *Advances in Neural Information Processing Systems*, *33*, 8017–8029.

Atz, K., Grisoni, F., & Schneider, G. (2021). Geometric deep learning on molecular representations. *Nature Machine Intelligence*, *3*(12), 1023–1032.

Bevilacqua, B., Eliasof, M., Meirom, E., Ribeiro, B., & Maron, H. (2023). Efficient subgraph gnns by learning effective selection policies. ArXiv preprint arXiv:2310.20082.

Bouritsas, G., Frasca, F., Zafeiriou, S., & Bronstein, M. M. (2022). Improving graph neural network expressivity via subgraph isomorphism counting. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *45*(1), 657–668.

Bresson, X., & Laurent, T. (2017). Residual gated graph convnets. ArXiv preprint arXiv:1711.07553.

Chen, D., O'Bray, L., & Borgwardt, K. (2022). Structure-aware transformer for graph representation learning. In *International conference on machine learning* (pp. 3469–3489). PMLR.

Corso, G., Cavalleri, L., Beaini, D., Liò, P., & Veličković, P. (2020). Principal neighbourhood aggregation for graph nets. *Advances in Neural Information Processing Systems*, *33*, 13260–13271.

Degen, J., Wegscheid-Gerlach, C., Zaliani, A., & Rarey, M. (2008). On the art of compiling and using'drug-like'chemical fragment spaces. *ChemMedChem*, *3*(10), 1503.

Dwivedi, V. P., Joshi, C. K., Luu, A. T., Laurent, T., Bengio, Y., & Bresson, X. (2023). Benchmarking graph neural networks. *Journal of Machine Learning Research*, *24*(43), 1–48.

Dwivedi, V. P., Luu, A. T., Laurent, T., Bengio, Y., & Bresson, X. (2021). Graph neural networks with learnable structural and positional representations. ArXiv preprint arXiv:2110.07875.

Eliasof, M., Frasca, F., Bevilacqua, B., Treister, E., Chechik, G., & Maron, H. (2023). Graph positional encoding via random feature propagation. In *International conference on machine learning* (pp. 9202–9223). PMLR.

Feng, J., Chen, Y., Li, F., Sarkar, A., & Zhang, M. (2022). How powerful are k-hop message passing graph neural networks. *Advances in Neural Information Processing Systems*, *35*, 4776–4790.

Frasca, F., Bevilacqua, B., Bronstein, M., & Maron, H. (2022). Understanding and extending subgraph gnns by rethinking their symmetries. *Advances in Neural Information Processing Systems*, *35*, 31376–31390.

Gao, H., Han, X., Huang, J., Wang, J.-X., & Liu, L. (2022). Patchgt: Transformer over non-trainable clusters for learning graph representations. In *Learning on graphs conference* (pp. 1–27). PMLR.

Girvan, M., & Newman, M. E. (2002). Community structure in social and biological networks. *Proceedings of the National Academy of Sciences*, *99*(12), 7821–7826.

Gu, Z., Luo, X., Chen, J., Deng, M., & Lai, L. (2023). Hierarchical graph transformer with contrastive learning for protein function prediction. *Bioinformatics*, *39*(7), btad410.

Hagberg, A., Swart, P. J., & Schult, D. A. (2008). *Exploring network structure, dynamics, and function using NetworkX*: *Technical report*, Los Alamos, NM (United States): Los Alamos National Laboratory (LANL).

Han, S., Fu, H., Wu, Y., Zhao, G., Song, Z., Huang, F., et al. (2023). Himgnn: a novel hierarchical molecular graph representation learning framework for property prediction. *Briefings in Bioinformatics*, *24*(5), bbad305.

He, X., Hooi, B., Laurent, T., Perold, A., LeCun, Y., & Bresson, X. (2023). A generalization of vit/mlp-mixer to graphs. In *International conference on machine learning* (pp. 12724–12745). PMLR.

Hu, W., Fey, M., Zitnik, M., Dong, Y., Ren, H., Liu, B., et al. (2020). Open graph benchmark: Datasets for machine learning on graphs. *Advances in Neural Information Processing Systems*, *33*, 22118–22133.

Huang, N. T., & Villar, S. (2021). A short tutorial on the Weisfeiler-Lehman test and its variants. In *ICASSP 2021-2021 IEEE international conference on acoustics, speech and signal processing* (pp. 8533–8537). IEEE.

Jin, W., Barzilay, R., & Jaakkola, T. (2020). Hierarchical generation of molecular graphs using structural motifs. In *International conference on machine learning* (pp. 4839–4848). PMLR.

Kipf, T. N., & Welling, M. (2016). Semi-supervised classification with graph convolutional networks. ArXiv preprint arXiv:1609.02907.

Kong, X., Huang, W., Tan, Z., & Liu, Y. (2022). Molecule generation by principal subgraph mining and assembling. *Advances in Neural Information Processing Systems*, *35*, 2550–2563.

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

Li, C., Liu, X., Li, W., Wang, C., Liu, H., Liu, Y., et al. (2024). U-kan makes strong backbone for medical image segmentation and generation. ArXiv preprint arXiv:2406.02918.

Li, C., Liu, X., Wang, C., Liu, Y., Yu, W., Shao, J., et al. (2025). Gtp-4o: Modality-prompted heterogeneous graph learning for omni-modal biomedical representation. In *European conference on computer vision* (pp. 168–187). Springer.

Liu, C., Yao, Z., Zhan, Y., Ma, X., Pan, S., & Hu, W. (2024). Gradformer: Graph transformer with exponential decay. ArXiv preprint arXiv:2404.15729.

Maskey, S., Parviz, A., Thiessen, M., Stärk, H., Sadikaj, Y., & Maron, H. (2022). Generalized laplacian positional encoding for graph representation learning. ArXiv preprint arXiv:2210.15956.

Masters, D., Dean, J., Klaser, K., Li, Z., Maddrell-Mander, S., Sanders, A., et al. (2023). Gps++: Reviving the art of message passing for molecular property prediction. ArXiv preprint arXiv:2302.02947.

Min, E., Chen, R., Bian, Y., Xu, T., Zhao, K., Huang, W., et al. (2022). Transformer for graphs: An overview from architecture perspective. ArXiv preprint arXiv:2202.08455.

Morris, C., Ritzert, M., Fey, M., Hamilton, W. L., Lenssen, J. E., Rattan, G., et al. (2019). Weisfeiler and leman go neural: Higher-order graph neural networks. In *Proceedings of the AAAI conference on artificial intelligence (Vol. 33)* (no. 01), (pp. 4602–4609).

Nalwade, A., Marshall, K., Eladi, A., & Sharma, U. (2024). On the expressive power of graph neural networks. ArXiv preprint arXiv:2401.01626.

Park, W., Chang, W., Lee, D., Kim, J., & Hwang, S.-w. (2022). Grpe: Relative positional encoding for graph transformer. ArXiv preprint arXiv:2201.12787.

Qian, C., Manolache, A., Ahmed, K., Zeng, Z., Van den Broeck, G., Niepert, M., et al. (2024). Probabilistically rewired message-passing neural networks. In *The twelfth international conference on learning representations*.

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

Tropsha, A., Isayev, O., Varnek, A., Schneider, G., & Cherkasov, A. (2024). Integrating QSAR modelling and deep learning in drug discovery: the emergence of deep QSAR. *Nature Reviews Drug Discovery*, *23*(2), 141–155.

Wang, Z., Cao, Q., Shen, H., Bingbing, X., Zhang, M., & Cheng, X. (2022). Towards efficient and expressive GNNs for graph classification via subgraph-aware Weisfeiler-Lehman. In *Learning on graphs conference* (pp. 1–17). PMLR.

Wang, X., & Zhang, M. (2021). GLASS: GNN with labeling tricks for subgraph representation learning. In *International conference on learning representations*.

Wieder, O., Kohlbacher, S., Kuenemann, M., Garon, A., Ducrot, P., Seidel, T., et al. (2020). A compact review of molecular property prediction with graph neural networks. *Drug Discovery Today: Technologies*, *37*, 1–12.

Wollschläger, T., Kemper, N., Hetzel, L., Sommer, J., & Günnemann, S. (2024). Expressivity and generalization: Fragment-biases for molecular GNNs. ArXiv preprint arXiv:2406.08210.

Wu, L., Cui, P., Pei, J., Zhao, L., & Guo, X. (2022). Graph neural networks: foundation, frontiers and applications. In *Proceedings of the 28th ACM SIGKDD conference on knowledge discovery and data mining* (pp. 4840–4841).

Xu, K., Hu, W., Leskovec, J., & Jegelka, S. (2018). How powerful are graph neural networks? ArXiv preprint arXiv:1810.00826.

Xu, T., & Zou, L. (2024). Rethinking higher-order representation learning with graph neural networks. In *Learning on graphs conference* (pp. 1–38). PMLR.

Yang, M., Wang, R., Shen, Y., Qi, H., & Yin, B. (2022). Breaking the expression bottleneck of graph neural networks. *IEEE Transactions on Knowledge and Data Engineering*, *35*(6), 5652–5664.

Yin, N., Shen, L., Chen, C., Hua, X.-S., & Luo, X. (2024). Sport: A subgraph perspective on graph classification with label noise. *ACM Transactions on Knowledge Discovery from Data*, *18*(9), 1–20.

Ying, C., Cai, T., Luo, S., Zheng, S., Ke, G., He, D., et al. (2021). Do transformers really perform badly for graph representation? *Advances in Neural Information Processing Systems*, *34*, 28877–28888.

Yu, Z., & Gao, H. (2022). Molecular representation learning via heterogeneous motif graph neural networks. In *International conference on machine learning* (pp. 25581–25594). PMLR.

Zang, X., Zhao, X., & Tang, B. (2023). Hierarchical molecular graph self-supervised learning for property prediction. *Communications Chemistry*, *6*(1), 34.

Zeng, D., Liu, W., Chen, W., Zhou, L., Zhang, M., & Qu, H. (2023). Substructure aware graph neural networks. In *Proceedings of the AAAI conference on artificial intelligence (Vol. 37)* (no. 9), (pp. 11129–11137).

Zhang, B., Fan, C., Liu, S., Huang, K., Zhao, X., Huang, J., et al. (2023). The expressive power of graph neural networks: A survey. ArXiv preprint arXiv:2308.08235.

Zhang, B., Feng, G., Du, Y., He, D., & Wang, L. (2023). A complete expressiveness hierarchy for subgraph gnns via subgraph Weisfeiler-Lehman tests. In *International conference on machine learning* (pp. 41019–41077). PMLR.

Zhang, S., Hu, Z., Subramonian, A., & Sun, Y. (2024). Motif-driven contrastive learning of graph representations. *IEEE Transactions on Knowledge and Data Engineering*.

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

Zhao, L., Jin, W., Akoglu, L., & Shah, N. (2021). From stars to subgraphs: Uplifting any GNN with local structure awareness. ArXiv preprint arXiv:2110.03753.

Zhao, H., Ma, S., Zhang, D., Deng, Z.-H., & Wei, F. (2023). Are more layers beneficial to graph transformers? ArXiv preprint arXiv:2303.00579.