MEGSA: Adaptive Motif-Level Structural Alignment for Graph Similarity Learning

Zhuolin Jia[1], Guangqi Wen[2,3], Lingwen Liu[4], Peng Cao[1] ✉, Jinzhu Yang[1], and Osmar R. Zaiane[5]

[1] Computer Science and Engineering, Northeastern University, Shenyang, China
[2] School of Computer Science and Artificial Intelligence, Shandong Normal University
[3] Shandong Key Laboratory of Intelligent Collaborative Perception and Computing
[4] Graduate School of Information, Production and Systems, Waseda University, Kitakyushu, Fukuoka, Japan
[5] Amii, University of Alberta, Edmonton, Alberta, Canada
caopeng@mail.neu.edu.cn

Abstract. In this study, we identify critical problems in graph similarity computation: context deficiency and spurious alignments. To solve these problems, our approach regards motifs as fundamental alignment units for capturing essential local structures while maintaining semantic context, and proposes a two-stage learning paradigm, involving Self-Supervised Motif Extraction and Motif Interaction Graph Similarity Learning. Moreover, we design multiple structural regularizations to ensure that these motifs retain their semantically meaningful properties (confidence, balance, diversity, connectivity). Theoretically, we prove that motif-level matching achieves a lower error upper bound, establishing that motif-level alignment is provably more robust than node-level alternatives. Extensive experiments on three benchmark datasets—LINUX, AIDS700nef, and IMDBMulti—demonstrate that MEGSA consistently outperforms state-of-the-art baselines across multiple evaluation metrics, achieving improvements of 9.73%, 16.62%, and 13.73% in MSE, respectively. Ablation studies and interpretability analyses further validate the effectiveness of motif-level matching, offering clear insight into the graph similarity estimate.

Keywords: Graph Similarity Computation · Self Supervised Learning · Motif Extraction · Error Bound Analysis · Interpretability

1 Introduction

Graph similarity learning (GSL) is a fundamental problem for tasks ranging from molecular analysis to social network comparison[1]. Although Graph Edit Distance (GED)[3] provides a flexible and interpretable measure of graph similarity, computing it exactly is NP-hard. To alleviate this issue, prior work has developed various GED approximation algorithms, such as A*[17], Beam Search[15], Hungarian matching[16], and VJ[22]. However, these methods remain computationally intensive, limiting their scalability in large-scale or real-time applications.

This limitation has driven increasing interest in learning-based graph similarity estimation, where graph neural networks (GNNs) can approximate graph similarity more efficiently. Existing GNN-based methods have further evolved from efficient but structurally coarse graph-level embeddings to more expressive node-level interaction paradigms.

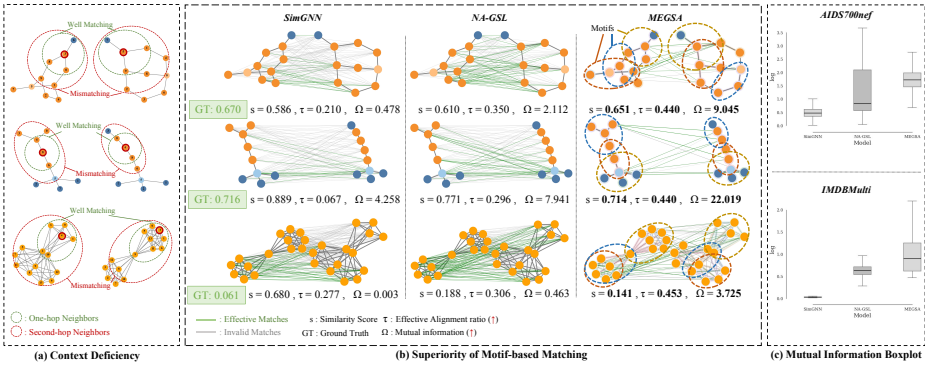


Fig. 1: In (a), when relying solely on 1-hop neighborhoods, two nodes may appear structurally similar and thus be considered a valid match. However, expanding the reception field to 2-hop neighborhoods reveals that their higher-order structures are non-isomorphic, rendering them distinguishable. In (b), we visualize the matching process for two node matching methods, *i.e.,* SimGNN (GNN-based) as well as NA-GSL(Transformer-based), and MEGSA. In (c): the mutual information is computed between the graph embedding obtained by individual methods on the two datasets. MEGSA shows larger mutual information values than the other two methods, revealing that it learns more informative graph embedding with richer semantics, which benefits graph similarity computation.

Despite their success, we argue that node-level GNN-based graph similarity learning methods remain fundamentally limited by two fundamental limitations: **context deficiency** and **spurious alignments**. **Context deficiency** arises from insufficient node representation. Individual nodes encode only localized information and lack the global structural context necessary for reliable alignment. As illustrated in Figure 1(a), standard GNNs may fail to distinguish nodes whose neighborhoods are non-isomorphic but differ only in their higher-order structural patterns. This reveals that node-level matching can be unreliable when node representations do not incorporate sufficient higher-order contextual information. On the other hand, **spurious alignments** stem from the node-level matching mechanism itself. As shown in Figure 1(b), node-level interactions often generate spurious alignments, *i.e.,* invalid matches, during cross-graph alignment. This issue is further aggravated by the quadratic growth of candidate correspondences in node-level matching, which amplifies the noise introduced by invalid matches.

As the number of candidate node pairs increases rapidly, invalid correspondences can accumulate and interfere with reliable similarity estimation.

Together, these two limitations reveal a fundamental **granularity gap** in existing graph similarity learning paradigms: although node-level interactions improve upon the coarse-grained nature of graph-level representations, they are overly fine-grained. To quantify these limitations empirically, we analyze the similarity learning process through two metrics: effective alignment ratio (τ) for measuring the proportion of valid alignments, and mutual information (Ω) between graph embeddings for quantifying the similar structural and semantic information. Preliminary results confirm that models with a higher effective alignment ratio and more similar graph structures yield better graph similarity estimates. This observation motivates the exploration of motif-based alignment that preserves richer structural context while mitigating the noise inherent in pairwise node interactions, which provides exactly an intermediate granularity.

Graph motifs introduce intermediate structural units that are well-suited for alignment, overcoming context deficiency and mitigating the spurious correspondences. To adapt motifs to task-specific structural patterns, motifs should be a principled unit rather than predefined. In the context of graph similarity learning, a motif is defined as a subgraph that is (i) *moderate in size*, (ii) *structurally informative*, (iii) *globally common*, and (iv) *connected*. However, leveraging motifs as principled alignment units introduces two challenges: extracting effective motifs and exploiting them for reliable similarity estimation. We address these with MEGSA, a two-stage framework: (i) Self-Supervised Motif Extraction for adaptive motif discovery, and (ii) Motif Interaction Graph Similarity Learning for explicit cross-graph alignment. Moreover, a serious of motif structure regulations are designed to ensure that these motifs remain the semantically meaningful properties. Experiments on LINUX, AIDS700nef, and IMDBMulti show significant performance gains. Motif analysis further demonstrates that MEGSA is interpretable, with the learned motifs capturing meaningful structural patterns that provide insights into graph similarity. In summary, our contributions are:

- We conduct a systematic analysis of existing graph similarity paradigms and identify two critical bottlenecks—context deficiency and spurious alignments—that limit the performance of node-level matching methods. We argue that fine-grained node correspondences introduce structural ambiguity and quadratic noise, necessitating a more reasonable and effective alignment mechanism via motifs: coherent substructures that capture higher-order topological semantics.
- We reformulate the concept of motifs, and propose a motif-based graph similarity learning paradigm. Our framework integrates self-supervised motif extraction with cross-graph motif interaction, enabling adaptive discovery of task-relevant motifs and robust similarity computation at an intermediate granularity that balances structural expressiveness and computational efficiency. Moreover, motifs are extracted in the pre-training task, reducing the amount of parameters in the downstream task. Furthermore, we provide

a theoretical error-bound analysis showing that motif-level matching admits a tighter upper bound on matching error.
- State-of-the-art performance and improved interpretability. Extensive experiments on three benchmark datasets demonstrate that MEGSA consistently outperforms existing methods, demonstrating that motif-level alignment is inherently more robust to structural noise and more discriminative in capturing inherent graph correspondences. Beyond quantitative gains, qualitative analyses reveal that the learned motifs capture structurally interpretable patterns, offering valuable insights into the similarity estimation process.

2 Related Work

2.1 Graph Similarity Computation

Quantifying graph similarity is a cornerstone of graph mining, traditionally dominated by Graph Edit Distance (GED) and Maximum Common Subgraph (MCS). While these combinatorial metrics offer rigorous alignment, their exponential computational complexity in the worst case renders them impractical for large-scale queries [3, 4]. Recent shifts toward Neural Graph Similarity Learning (GSL) have introduced two primary paradigms: Global Encoding and Cross-Graph Interaction. Early encoders map graphs to a joint vector space for efficient distance calculation but suffer from coarse-grained information loss. To mitigate this, interaction-based models like SimGNN [1], GMN [7], and DeepSim [9] introduce fine-grained node-level alignment. Despite their success, we argue that these methods encounter a bottleneck: node-level correspondences often produce noisy matching signals because individual nodes lack sufficient structural context to resolve topological ambiguities. Furthermore, the quadratic complexity of $O(N_1 N_2)$ node-pair interactions limits their scalability. These limitations have motivated increasing interest in developing representations that maintain strong structural expressiveness while reducing computational overhead [27, 28, 11]. This motivates our exploration of an intermediate structural scale, the motif, to balance expressiveness and efficiency.

2.2 Motif Extraction

Graph motifs, defined as recurring subgraph patterns, serve as the functional "building blocks" of complex networks [14]. Conventional discovery algorithms rely on combinatorial enumeration or subgraph census, and the number of possible patterns grows super-exponentially with motif size, making exact counting computationally prohibitive [14, 18]. The integration of motifs into GNNs has evolved from predefined templates to automated discovery. Early motif-based graph representation methods typically relied on enumerating a small set of hand-crafted structures (e.g., triangles, cliques, feed-forward loops) and then building higher-level features on top of them. More recent self-supervised frameworks such as MICRO-Graph and related MOTIF-style contrastive pretraining

automatically learn motif prototypes by clustering subgraph embeddings, but mainly target graph-level representation enhancement rather than explicit cross-graph alignment [26, 13, 10]. Our work departs from these constraints by introducing a fully adaptive, reconstruction-driven extraction mechanism, ensuring that the identified motifs are both statistically significant and task-relevant.

3 Method

3.1 Problem Statement

Given two graphs G_p and G_q, graph similarity learning aims to learn a function $f_\theta : (G_p, G_q) \to s$, that predicts a similarity score consistent with ground-truth graph edit distance (GED). We enhance this by introducing motif-level structural representations. Let $\mathcal{M}(G) = \{M_1, M_2, \ldots, M_K\}$ denote the set of motifs extracted from graph G, where each motif M_k is a connected, recurring, and structurally informative subgraph. Instead of operating directly on raw graphs, we learn a motif-enhanced similarity function $f_\theta^{\text{motif}} : (\mathcal{M}(G_p), \mathcal{M}(G_q)) \to s_{pq}$ based on alignment and interaction of motif-level representations. This formulation captures both local structural semantics and global graph context, mitigating context deficiency and spurious alignments inherent in node-level matching. Motif sets are obtained via a self-supervised extraction module that learns task-adaptive motifs without predefined templates.

3.2 Motif Definition

A motif is defined as a recurring and statistically significant pattern of interconnections within a graph [14]. In the context of graph similarity learning, we reformulate a motif as a subgraph satisfying the following properties:

(*i*) Size Constraint. Let $m = |V_M|$ denote the number of nodes in a motif. A motif should be moderate in size ($2 \leq m \ll |V|$), to represent compact higher-order interaction patterns rather than trivial nodes or the entire graph.

(*ii*) Structural Informativeness. To preserve the structural integrity of motifs, node membership is defined in a non-exclusive manner. Formally, a node may simultaneously belong to multiple motifs, i.e. $\exists\, i,\ M_a \neq M_b$ s.t. $i \in V_{M_a}$ and $i \in V_{M_b}$, This property avoids strict graph partitioning and ensures that assigning a node to one motif does not disrupt the structural completeness of another, thereby preserving informative higher-order interaction patterns.

(*iii*) Global Commonality. A motif should represent a structural pattern that is widely shared across graphs in the dataset. Specifically, a motif is considered globally common if it appears in a sufficient portion of graphs in the dataset. This property ensures that motifs capture transferable structural patterns that can be used for cross-graph comparison and alignment.

(*iv*) Connectivity. The motif M forms a connected subgraph, i.e., for any two nodes $u, v \in V_M$, there exists a path in M connecting them. This property ensures that a motif represents a coherent structural unit rather than a collection of disconnected components.

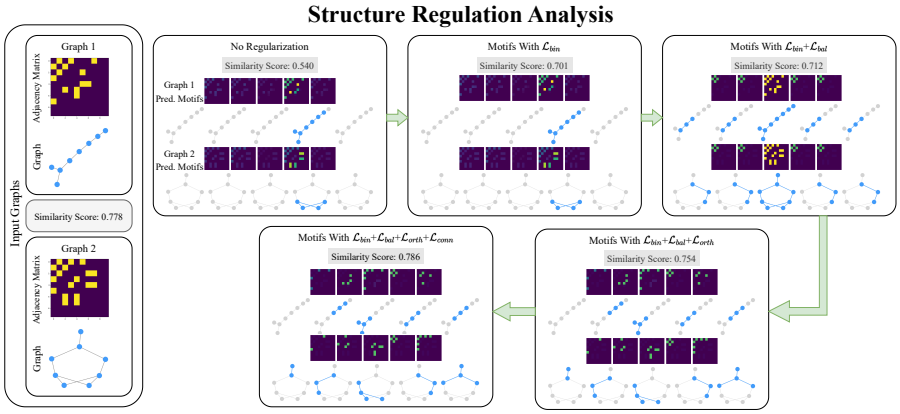


Fig. 2: Illustration of motif structure regulations in graph similarity prediction. The max number of motifs is set to 5 in these dataset.

To preserve the intrinsic properties of motifs, we impose a set of structural regularization terms that enforce motif confidence, balance, diversity, and connectivity. The effect of these constraints is visualized in Fig. 2, which validates their ability to guide motifs toward the desired structural properties.

3.3 Overview

We propose the **M**otif-**E**nhanced **G**raph **S**imilarity **A**lignment (**MEGSA**) framework (Fig. 3), which models graph similarity through motif-level alignment rather than node-level correspondence to mitigate the issues of **context deficiency** and **spurious alignments**.

3.4 Self-Supervised Motif Extraction

The objective of motif extraction is to identify structurally informative subgraph units that reliably support subsequent graph alignment. A self-supervised learning framework is designed to encourage the extracted motifs to preserve the essential topological structures of the original graph, while maintaining sufficient flexibility for downstream similarity modeling.

Motif Decomposition Rather than relying on predefined substructures, motif decomposition learns motifs adaptively from data, allowing the model to identify structurally relevant subgraphs for similarity modeling. It partitions the input graph into smaller, structurally coherent subgraphs that capture recurring connectivity patterns. By representing the graph as a composition of such higher-order units, the framework preserves essential topological characteristics while maintaining flexibility with respect to the similarity objective.

Node Encoding. To identify motifs that reflect coherent structural semantics, node representations must incorporate both local attributes and global structural

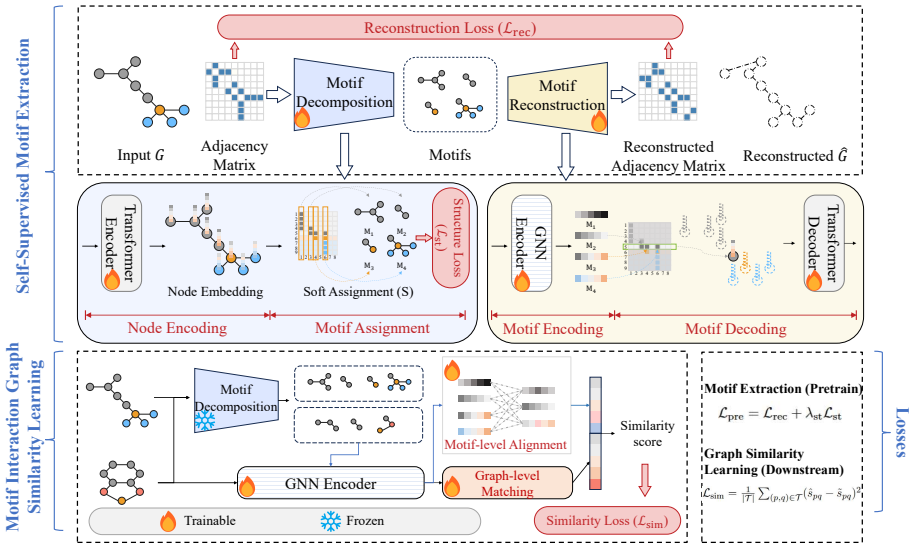


Fig. 3: MEGSA operates in two stages. First, self-supervised motif extraction identifies higher-order task-adaptive motifs via decomposition (extracting patterns) and reconstruction (preserving structure). Second, motif interaction Graph similarity learning performs supervised similarity estimation through explicit cross-graph motif alignment based on the extracted motifs.

context. We employ a Transformer-based encoder [21], denoted as $\mathcal{E}_\theta$, to compute context-aware node embeddings as $Z = \mathcal{E}_\theta(X, A)$, where $X \in \mathbb{R}^{N\times F}$ denotes the node feature matrix with N nodes and F-dimensional features, $A \in \mathbb{R}^{N\times N}$ is the adjacency matrix encoding graph connectivity, and $Z \in \mathbb{R}^{N\times d}$ represents the resulting node embeddings with dimensionality d. The self-attention mechanism enables long-range dependency modeling beyond the limited receptive fields of standard GNNs, providing a structurally informed representation basis for subsequent motif discovery.

Motif Assignment. Since motifs are learned from data rather than predefined and nodes may simultaneously belong to multiple structural patterns, motif discovery is formulated as a soft assignment process instead of hard clustering. Given node embeddings $Z \in \mathbb{R}^{N\times d}$, a soft assignment matrix is computed as $S = \sigma(ZW_s)$, where $W_s \in \mathbb{R}^{d\times K}$ is a learnable parameter matrix, K denotes the maximum number of motifs, and $S \in \mathbb{R}^{N\times K}$ is the resulting assignment matrix [25]. Each entry S_{ik} quantifies the soft membership of node i to motif k. Based on the soft assignment matrix $S \in \mathbb{R}^{N\times K}$, we construct a motif-aware adjacency matrix for each motif k by modeling pairwise node co-membership. Specifically, the k-th motif structure is obtained as $\tilde{A}_k = S_{:,k}S_{:,k}^\top$, which captures the soft co-assignment strength between nodes. To incorporate sparsification and motif weighting while preserving the original graph connectivity, a differentiable sharp-

ening operation is applied: $\mathcal{A}_k = \sigma\Big(\gamma(\tilde{A}_k - t)\Big) \odot A$, where $\mathcal{A}_k \in \mathbb{R}^{N\times N}$ denotes the motif-aware adjacency matrix for motif k, t is a sparsification threshold, γ controls the sharpness of the soft masking, and $\odot$ represents element-wise multiplication. Stacking all $\mathcal{A}_k$ yields the motif-aware adjacency tensor $\mathcal{A} \in \mathbb{R}^{K\times N\times N}$. This formulation establishes motifs as latent structural alignment units that are optimized jointly with the similarity objective within an end-to-end learning framework. To promote structurally meaningful motif identification, we impose a set of motif structure regularization terms on the assignment matrix.

1.Confidence Regularization. To encourage confident motif memberships, we penalize ambiguous soft assignments: $\mathcal{L}_{\text{bin}} = \frac{1}{NK}\sum_{i=1}^{N}\sum_{k=1}^{K} S_{ik}(1 - S_{ik})$. This term promotes near-discrete memberships, ensuring that motifs remain structurally concentrated.

2.Connectivity Regularization. To enforce structural coherence within motifs, we minimize the Laplacian quadratic form: $\mathcal{L}_{\text{conn}} = \text{Tr}(S^\top LS)$, where $L = D - A$ is the graph Laplacian. The regularization enables our model to favor connected and structurally informative subgraphs.

3.Orthogonality Regularization. To reduce redundancy across motifs, we introduce a soft orthogonality constraint: $\mathcal{L}_{\text{orth}} = \left\| \frac{S^\top S}{\|S^\top S\|_F} - \frac{I}{\|I\|_F} \right\|_F$. This regularization discourages excessive overlap between motifs and promotes structural diversity across learned substructures.

4.Balance Regularization. To avoid degenerate solutions in which motifs collapse or vanish, we impose $\mathcal{L}_{\text{bal}} = -\frac{1}{K}\sum_{k=1}^{K} \log\left(\sum_{i=1}^{N} S_{ik} + \epsilon\right)$. This constraint ensures that the motif remains non-trivial across the graph.

Overall, the motif structural regularization is defined as in Eq.(1). The proposed regularization align closely with the desired properties of motifs. Specifically, the orthogonality loss $\mathcal{L}_{\text{orth}}$ encourages structural diversity across motifs, ensuring that each motif is moderate in size and non-redundant [Properties ***(i)***]. The connectivity loss $\mathcal{L}_{\text{conn}}$ promotes internal coherence, enforcing that motifs form connected subgraphs [Properties ***(iv)***]. Finally, the soft assignment S allows nodes to participate in multiple motifs, which, combined with the confidence and balance constraints, encourages motifs to be globally common and structurally informative [Properties ***(ii)*** and ***(iii)***]. Collectively, these design choices ensure that the learned motifs satisfy the intended structural properties while being optimized end-to-end.

$$\mathcal{L}_{\text{st}} = \mathcal{L}_{\text{bin}} + \lambda_c\mathcal{L}_{\text{conn}} + \lambda_o\mathcal{L}_{\text{orth}} + \lambda_s\mathcal{L}_{\text{bal}}. \tag{1}$$

Motif Reconstruction The motif reconstruction objective enforces structural completeness by requiring that the original graph topology be recoverable from the motif-induced structures. This constraint prevents motifs from collapsing into fragmented or information-deficient subgraphs.

Motif Encoding. For each motif k, a motif-specific graph encoder is applied to propagate information over its induced adjacency matrix $\mathcal{A}_k$. The layer-wise update rule is defined as $H_k^{(l)} = \sigma\Big(\mathcal{A}_k H_k^{(l-1)} W_l + H_k^{(l-1)}\Big)$, where W_l denotes the learnable weight matrix at layer l, and $H_k^{(l-1)}$ is the input to the l-th layer. The

residual connection stabilizes optimization and enables deeper motif representations. After L layers, we obtain motif-specific embeddings $\{H_k^{(L)}\}_{k=1}^K$, which encode the intermediate structural composition of the graph.

Motif Decoding. To reconstruct node-level representations from motif embeddings, we first aggregate motif features through the assignment matrix S into the fused node representation, $H = \sum_{k=1}^K S_{:,k} \odot H_k^{(L)}$. A Transformer-based decoder, denoted by $\mathcal{D}_\phi$, is then applied to reconstruct the adjacency matrix $\hat{A} = \mathcal{D}_\phi(H)$. The reconstruction objective encourages the motif representations to preserve sufficient structural information to recover the global topology, thereby enforcing structural completeness at the motif level. To ensure reconstruction consistency, we introduce a reconstruction loss that enforces the preservation of the original graph topology within motif-aware representations: $\mathcal{L}_{\text{rec}} = \frac{1}{N^2}\|\hat{A} - A\|_F^2$. The overall pre-training objective for the motif extraction stage is defined as: $\mathcal{L}_{\text{pre}} = \mathcal{L}_{\text{rec}} + \lambda_{\text{st}}\mathcal{L}_{\text{st}}$, where λ_{st} controls the trade-off between reconstruction fidelity and structural regularization.

3.5 Motif Interaction Graph Similarity Learning

To accurately estimate graph similarity, we leverage the previously learned structural motifs as structurally stable units, providing a higher-level representation of graph topology. In this stage, similarity computation is performed by extracting motifs via a frozen motif decomposition module and capturing fine-grained correspondences through motif-level alignment to yield a global similarity score.

Motif-Level Alignment Given a pair of graphs G_p and G_q, motif embeddings $M^{(p)}, M^{(q)} \in \mathbb{R}^{K \times d_m}$ are extracted using the pretrained motif encoder. Rather than relying on node-level correspondence, cross-graph similarity is modeled at the motif level, providing richer structural context and reducing local ambiguity.

To model cross-graph interactions, the motif embeddings are first projected into Query, Key, and Value spaces: $Q_p = M^{(p)}W_Q$, $K_q = M^{(q)}W_K$, and $V_q = M^{(q)}W_V$, where W_Q, W_K, W_V are learnable projection matrices. Motif correspondence is then computed via scaled dot-product attention:

$$\alpha_{ij}^{pq} = \frac{\exp\big((Q_p)_i (K_q)_j^\top / \sqrt{d_m}\big)}{\sum_{k=1}^K \exp\big((Q_p)_i (K_q)_k^\top / \sqrt{d_m}\big)}, \tag{2}$$

where α_{ij}^{pq} measures the structural compatibility between the i-th motif in G_p and the j-th motif in G_q.

Then, the aligned motif representations are obtained as $\tilde{M}^{(p)} = \alpha^{pq} V_q$ and $\tilde{M}^{(q)} = (\alpha^{pq})^\top M^{(p)}$. This alignment emphasizes structurally consistent subgraphs while suppressing noisy or weak correspondences, enabling robust similarity estimation across graphs of varying sizes and structures.

Graph-level Matching While motif-level alignment captures fine-grained structural correspondence, certain global properties may not be fully preserved by local motifs alone. To this end, we include a lightweight graph-level matching module (Neural Tensor Network (NTN) [1]) as a complementary signal, *i.e.,*

$g^{pq} = \sigma\big(h_p^\top W^{[1:k]} h_q + V[h_p; h_q] + b\big)$, where $W^{[1:k]} \in \mathbb{R}^{D\times D\times k}$ is a third-order tensor capturing bilinear interactions between the two graph embeddings across k tensor slices, V and b denote the linear projection matrix and bias term, respectively, and $\sigma(\cdot)$ is a non-linear activation function. Finally, g^{pq} is concatenated with motif embeddings $\tilde{M}^{(p)}$ and $\tilde{M}^{(q)}$ from motif-level alignment for the final similarity prediction. Importantly, this graph-level interaction serves as an auxiliary context to stabilize similarity prediction under extreme structural variations.

Given graph pairs (G_p, G_q) from the training pair set $\mathcal{T}$, with predicted similarity $\hat{s}_{pq}$ obtained by $f_\theta^{\text{motif}}(\mathcal{M}(G_p), \mathcal{M}(G_q))$ and ground-truth similarity $\tilde{s}_{pq}$, the similarity learning loss over the training set is defined as $\mathcal{L}_{\text{sim}} = \frac{1}{|\mathcal{T}|}\sum_{(p,q)\in\mathcal{T}}(\hat{s}_{pq} - \tilde{s}_{pq})^2$, where $|\mathcal{T}|$ denotes the number of training graph pairs.

3.6 Theoretical Error Bound Analysis of Motif-Level Matching

Let $M^\star$ denote the optimal structural alignment between two graphs, and let $\hat{M}_{\text{node}}$ and $\hat{M}_{\text{motif}}$ denote the estimated matchings obtained by node-level and motif-level alignment, respectively. Based on the theoretical analysis in Supplementary A, we obtain the following proposition (the detailed proof is deferred to the supplementary material): $\Pr(\hat{M}_{\text{motif}} \neq M^\star) < \Pr(\hat{M}_{\text{node}} \neq M^\star)$. This result indicates that motif-level matching yields a tighter theoretical error upper bound than node-level matching. The reason is threefold: motif aggregation alleviates local perturbations and thus reduces error variance; motif connectivity makes incorrect candidates less likely to exhibit false similarity; and motif-level matching substantially compresses the candidate interaction space. Together, these properties make motif-level matching more robust to local noise and more discriminative in structural alignment, resulting in a lower overall upper bound on matching error.

4 Experiment

4.1 Experimental Setup

Datasets and Hyperparameter Settings. We evaluate our model on three widely used graph similarity estimation benchmarks: LINUX [1, 23], AIDS700nef [1, 23], and IMDBMulti [24]. Dataset statistics are summarized in Table 1. These datasets include diverse domains such as software engineering, molecular chemistry, and social networks, providing a comprehensive testbed for graph similarity modeling. All datasets are publicly available to ensure reproducibility. Graph Edit Distances (GED) are used as the ground-truth similarity measure. The graph similarity score $\tilde{s}_{pq}$ is computed as described in the Supplementary B.1. We tune key hyperparameters on the validation set. Dataset-specific settings are provided in the Supplementary B.2.

Competitive Methods To evaluate the proposed method, we consider representative baselines from two categories: (1) conventional Graph Edit Distance

Table 1: Statistics of Graph Similarity Datasets

Dataset	Graphs	Node range	Average nodes	Average edges	Attributes	Category
AIDS700nef	700	[2, 10]	8.9	8.4	Yes	Biochemical
LINUX	1000	[4, 10]	7.6	6.9	No	Code fragment
IMDBMulti	1500	[4, 89]	13.0	65.9	No	Social

(GED) algorithms, including A* [17], Beam [15], Hungarian [16], and VJ [22]; and (2) state-of-the-art learning-based graph similarity models, including SimGNN [1], GMN [7], GraphSim [2], GOTSim [5], MGMN [8], DeepSim [9], NA-GSL [20], CLSim [28] and AMFF [29]. Among these methods, SimGNN [1], MGMN [8], DeepSim [9], CLSim [28] and AMFF [29] adopt multi-granularity learning strategies that integrate both node-level and graph-level representations, whereas the remaining models primarily operate at the node level. To further demonstrate the effectiveness of motif-level information, we compare the full model, which integrates both motif-level and graph-level representations, with an ablation variant that uses only motif-level representations. This comparison highlights the impact of incorporating both motif-level and graph-level structural information on similarity prediction. Moreover, to quantitatively evaluate the impact of motif extraction strategies, we integrate different motif extractors into the same graph similarity framework. Clique Motif: Extracts complete subgraphs where every node is connected to each other node, and Cycle Motif: Focuses on closed-loop structures in the graph. Performance is evaluated using three complementary metrics: Mean Squared Error (MSE) to measure prediction accuracy, Spearman's rank correlation coefficient (ρ) [19] to assess ranking consistency, and precision at top 10 (p@10) to evaluate top-k retrieval quality.

4.2 Result Analysis

Table 2 summarizes the performance comparison between our proposed method and various baselines. To thoroughly evaluate the effectiveness of our approach, we conduct the analysis by addressing the following three research questions.

RQ1: Does the proposed model outperform sota methods across diverse datasets? As shown in Table 2, our method (M+G) consistently achieves superior or highly competitive results across all datasets. On the AIDS700nef dataset, our model achieves MSE of 1.600 and Spearman's correlation (ρ) of 0.898, outperforming the best learning-based methods (GraphSim and NA-GSL) by a clear margin. On the large-scale IMDBMulti dataset, our model achieves the highest MSE (0.641) and ρ (0.937). These results indicate that by integrating motif-level structures with graph-level representations, MEGSA effectively captures both local structural semantics and global topological coherence

RQ2: Does the motif-level strategy contribute to similarity computation? On the LINUX dataset, the motif-only variant (MEGSA *w/o* Graph) achieves the highest ρ (0.988) among all learning-based methods, even slightly

Table 2: Performance comparison on benchmark datasets. N, M and G indicate the node-level, motif-level and graph-level, respectively.

Methods	AIDS700nef			LINUX			IMDBMulti		
	mse	ρ	p@10	mse	ρ	p@10	mse	ρ	p@10
A*	0.000	1.000	1.000	0.000	1.000	1.000	-	-	-
Beam	9.268	0.827	0.973	12.090	0.609	0.481	2.413	0.905	0.813
Hungarian	29.805	0.638	0.913	25.296	0.510	0.360	1.845	0.932	0.825
VJ	64.863	0.581	0.287	29.157	0.517	0.310	1.831	0.934	0.815
SimGNN (N+G)	3.699	0.828	0.441	1.902	0.963	0.953	3.253	0.916	0.828
GMN (N)	4.610	0.672	0.200	2.571	0.906	0.888	4.422	0.725	0.604
GraphSim (N)	<u>1.919</u>	0.849	0.446	<u>0.471</u>	0.976	0.956	<u>0.743</u>	<u>0.926</u>	0.828
GOTSim (N)	2.360	0.860	**0.870**	4.250	0.920	0.860	5.920	0.850	0.730
MGMN (N+G)	3.162	<u>0.869</u>	0.362	4.530	0.922	0.551	13.291	0.854	0.508
DeepSim (N+G)	6.370	0.674	0.342	11.990	0.861	0.713	2.042	0.921	**0.882**
NA-GSL (N)	2.536	0.863	0.449	0.558	<u>0.984</u>	<u>0.968</u>	1.748	0.855	0.554
CLSim (N+G)	2.120	0.866	0.471	0.514	0.977	0.956	1.143	0.919	0.823
AMFF (N+G)	2.200	0.869	0.452	<u>0.452</u>	0.986	0.963	0.988	0.902	0.816
MEGSA (*w/o* Graph) (M)	4.039	0.769	0.331	0.647	**0.988**	**0.979**	0.887	0.915	0.837
MEGSA (Clique) (M+G)	2.888	0.827	0.37	0.553	0.980	0.953	2.209	0.891	0.790
MEGSA (Cycle) (M+G)	2.364	0.854	0.426	0.492	0.987	0.956	2.011	0.912	0.801
MEGSA (M+G)	**1.600**	**0.898**	<u>0.585</u>	**0.408**	**0.988**	0.963	**0.641**	**0.937**	<u>0.854</u>

surpassing the full model. This observation suggests that in software flow graphs, motif-level structures are highly discriminative for similarity estimation. Furthermore, on IMDBMulti, the motif-only model still achieves competitive performance (MSE=0.887, ρ=0.915) compared with several complex node+graph fusion models such as SimGNN and DeepSim. These results demonstrate that the motif-level strategy provides effective structural cues for similarity computation and becomes particularly beneficial for large graphs, where node-level information may be noisy.

RQ3: Does the learned motif offer a more effective and flexible way to capture graph similarity? To quantitatively evaluate the impact of motif extraction strategies,we integrate different motif extractors. The results consistently show that fixed-pattern extractors lead to degraded performance across all datasets. The gap is particularly obvious on AIDS700nef and IMDBMulti, where predefined motifs struggle to capture the task-specific structural patterns inherent in molecular chemistry or social networks. In contrast, our adaptive extractor learns motifs that are better aligned with the similarity objective, providing more discriminative representations.

4.3 Motif Analysis

The influence of Motif number To further investigate the hyperparameter sensitivity of our model, we analyze the effect of the maximum number of

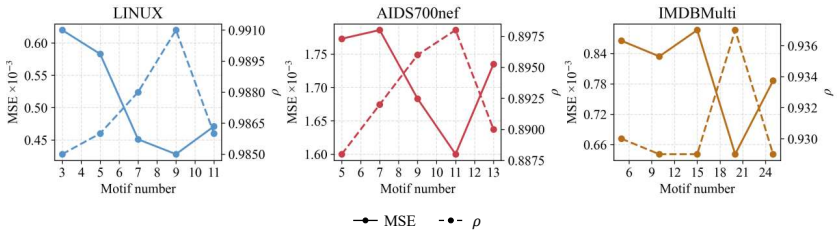


Fig. 4: The influence of maximum motif numbers.

motifs on model performance. Specifically, we vary the motif number while keeping other parameters fixed and evaluate the model using three metrics: MSE, Spearman's correlation coefficient (ρ), and precision@10 (p@10). The results are presented as line charts. As the number of motifs increases, the model performance first improves and then gradually declines. This trend indicates that a moderate number of motifs allows the model to capture richer structural patterns, whereas an excessive number introduce redundant or noisy substructures, leading to performance degradation. Notably, the optimal number of motifs differs across datasets. In our experiments, the motif number for each dataset was chosen based on performance on a validation set.

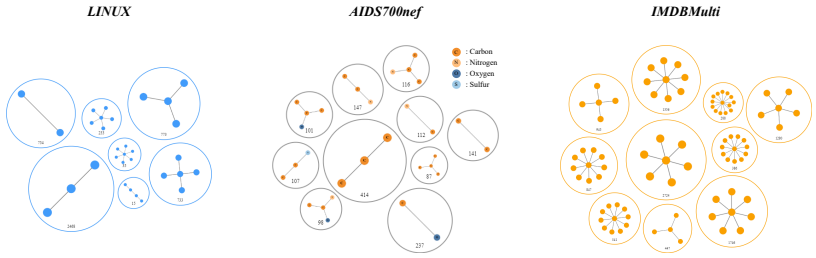


Fig. 5: High-frequency motifs extracted across three datasets. For AIDS700nef, the top frequent motifs, *i.e.,* C–C–C with branched nitrogen (N) or oxygen (O) atoms, shape the topology of molecular interactions and hydrogen-bond profiles [12]; C–C–O increases molecular polarity and reactivity, directly influencing solubility and adsorption [6]. On Linux and IMDBMulti, the learned motifs capture domain-specific structures, including recurrent code dependencies and movie–actor interaction hubs.

The identified high-frequency motifs To investigate whether the learned motifs align with domain-specific knowledge, we visualize the high-frequency motifs identified by our model. Figure 5 summarizes these patterns and their frequencies. By extracting such concise, high-order topological abstractions, MEGSA avoids black-box reasoning and provides transparent rationales for its similar-

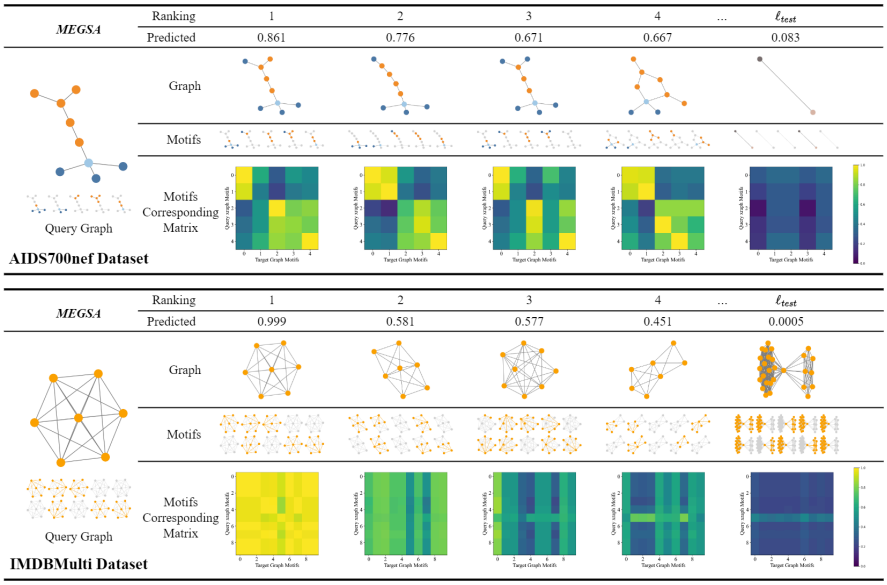


Fig. 6: Qualitative analysis of similarity retrieval across datasets. For each query graph, we present the most similar graphs retrieved by MEGSA together with their motif correspondence matrix. The graphs with higher predicted similarity scores exhibit strong structural coherence and highly consistent motif compositions with the query graph. In contrast, graphs with lower similarity scores present divergent structural layouts and distinct motif distributions. The motif correspondence matrix visualizes the similarity of motif compositions between the query graph and the retrieved graphs. Each row or column in the matrix corresponds to a specific motif learned by MEGSA, while the matrix entries represent the strength of alignment between motifs from the two graphs. High correspondence scores indicate that the two graphs share structurally consistent and semantically meaningful subgraph patterns.

ity estimates. These identified motifs enable our model to provide interpretable insights into the predicted similarity scores.

Motif Interpretability To further assess the interpretability and effectiveness of the proposed model, we conduct a qualitative case study on the two datasets in Fig. 6. These observations suggest that the proposed model captures both global structural alignment and local motif-level correspondences, while the motif correspondence matrix provides an interpretable view of how motif-level structures contribute to the similarity prediction.

5 Conclusion

In this paper, we propose a novel graph similarity learning framework that integrates both graph-level and motif-level representations to achieve coarse-to-fine structural matching. Unlike traditional methods that rely solely on global or node-level comparisons, our model introduces motif-level interaction as an intermediate granularity, effectively capturing local structural semantics while maintaining global coherence. Through comprehensive experiments on multiple benchmark datasets, our method consistently outperforms existing approaches. Ablation studies and qualitative analyses further validate the effectiveness of motif-level matching and the semantic interpretability of learned motifs. Our work establishes both theoretical and empirical foundations for motif-enhanced graph similarity learning, opening new directions for multi-granularity reasoning in graph representation learning. In future work, we plan to extend our framework to handle other graph-related regression and classification tasks. We believe that the hierarchical motif-aware design and multi-level interaction mechanism can generalize well to these problems, enabling more accurate and robust predictions across diverse graph domains.

Acknowledgments. This research was supported by the National Key Research and Development Program of China (No.2020YFC0833302), the National Natural Science Foundation of China (No.62076059), the Science and Technology Joint Project of Liaoning province (2023JH2/101700367) and the Fundamental Research Funds for the Central Universities (No. N2424010-7).

References

1. Yunsheng Bai, Hao Ding, Song Bian, Ting Chen, Yizhou Sun, and Wei Wang. Simgnn: A neural network approach to fast graph similarity computation. In *Proceedings of the twelfth ACM international conference on web search and data mining*, pages 384–392, 2019.
2. Yunsheng Bai, Hao Ding, Ken Gu, Yizhou Sun, and Wei Wang. Learning-based efficient graph similarity computation via multi-scale convolutional set matching. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, pages 3219–3226, 2020.
3. Horst Bunke. What is the distance between graphs. *Bulletin of the EATCS*, 20(1983):35–39, 1983.
4. Horst Bunke and Kim Shearer. A graph distance metric based on the maximal common subgraph. *Pattern recognition letters*, 19(3-4):255–259, 1998.
5. Khoa D Doan, Saurav Manchanda, Suchismit Mahapatra, and Chandan K Reddy. Interpretable graph similarity computation via differentiable optimal alignment of node embeddings. In *Proceedings of the 44th international ACM SIGIR conference on research and development in information retrieval*, pages 665–674, 2021.
6. Mária Jerigová, Mateusz Odziomek, and Nieves López-Salas. "we are here!" oxygen functional groups in carbons for electrochemical applications. *ACS omega*, 7(14):11544–11554, 2022.

7. Yujia Li, Chenjie Gu, Thomas Dullien, Oriol Vinyals, and Pushmeet Kohli. Graph matching networks for learning the similarity of graph structured objects. In *International conference on machine learning*, pages 3835–3845. PMLR, 2019.
8. Xiang Ling, Lingfei Wu, Saizhuo Wang, Tengfei Ma, Fangli Xu, Alex X Liu, Chunming Wu, and Shouling Ji. Multilevel graph matching networks for deep graph similarity learning. *IEEE Transactions on Neural Networks and Learning Systems*, 34(2):799–813, 2021.
9. Bo Liu, Zhihan Wang, Jidong Zhang, Jiahui Wu, and Guangzhi Qu. Deepsim: a novel deep learning method for graph similarity computation. *Soft Computing*, 28(1):61–76, 2024.
10. Youming Liu, Lin Shu, Chuan Chen, and Zibin Zheng. Fine-grained semantics enhanced contrastive learning for graphs. *IEEE Transactions on Knowledge and Data Engineering*, 36:8238–8250, 2024.
11. Zhouyang Liu, Yixin Chen, Ning Liu, Jiezhong He, and Dongsheng Li. Graph2region: Efficient graph similarity learning with structure and scale restoration. *IEEE Transactions on Knowledge and Data Engineering*, 37:7213–7225, 2025.
12. Frank Lovering, Jack Bikker, and Christine Humblet. Escape from flatland: increasing saturation as an approach to improving clinical success. *Journal of medicinal chemistry*, 52(21):6752–6756, 2009.
13. Xiao Luo, Wei Ju, Meng Qu, Yiyang Gu, C. Chen, Minghua Deng, Xiansheng Hua, and Ming Zhang. Clear: Cluster-enhanced contrast for self-supervised graph representation learning. *IEEE Transactions on Neural Networks and Learning Systems*, 35:899–912, 2022.
14. Ron Milo, Shai Shen-Orr, Shalev Itzkovitz, Nadav Kashtan, Dmitri Chklovskii, and Uri Alon. Network motifs: simple building blocks of complex networks. *Science*, 298(5594):824–827, 2002.
15. Michel Neuhaus, Kaspar Riesen, and Horst Bunke. Fast suboptimal algorithms for the computation of graph edit distance. In *Joint IAPR International Workshops on Statistical Techniques in Pattern Recognition (SPR) and Structural and Syntactic Pattern Recognition (SSPR)*, pages 163–172. Springer, 2006.
16. Kaspar Riesen and Horst Bunke. Approximate graph edit distance computation by means of bipartite graph matching. *Image and Vision computing*, 27(7):950–959, 2009.
17. Kaspar Riesen, Xiaoyi Jiang, and Horst Bunke. Exact and inexact graph matching: Methodology and applications. *Managing and mining graph data*, pages 217–247, 2010.
18. F. Ruzzenenti, F. Picciolo, Petter Holme, and R. Mastrandrea. Weighted network motifs as random walk patterns. *New Journal of Physics*, 24, 2020.
19. Charles Spearman. The proof and measurement of association between two things. 1961.
20. Wenhui Tan, Xin Gao, Yiyang Li, Guangqi Wen, Peng Cao, Jinzhu Yang, Weiping Li, and Osmar R Zaiane. Exploring attention mechanism for graph similarity learning. *Knowledge-Based Systems*, 276:110739, 2023.
21. Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.
22. Runzhong Wang, Tianqi Zhang, Tianshu Yu, Junchi Yan, and Xiaokang Yang. Combinatorial learning of graph edit distance via dynamic embedding. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5241–5250, 2021.

23. Xiaoli Wang, Xiaofeng Ding, Anthony KH Tung, Shanshan Ying, and Hai Jin. An efficient graph indexing method. In *2012 IEEE 28th International Conference on Data Engineering*, pages 210–221. IEEE, 2012.
24. Pinar Yanardag and SVN Vishwanathan. Deep graph kernels. In *Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining*, pages 1365–1374, 2015.
25. Zhitao Ying, Jiaxuan You, Christopher Morris, Xiang Ren, Will Hamilton, and Jure Leskovec. Hierarchical graph representation learning with differentiable pooling. *Advances in neural information processing systems*, 31, 2018.
26. Shichang Zhang, Ziniu Hu, Arjun Subramonian, and Yizhou Sun. Motif-driven contrastive learning of graph representations. *arXiv preprint arXiv:2012.12533*, 2020.
27. Wei Zhuo and Guang Tan. Efficient graph similarity computation with alignment regularization. *ArXiv*, abs/2406.14929, 2024.
28. Cuifang Zou, Guangquan Lu, Longqing Du, Xuxia Zeng, and Shilong Lin. Graph similarity learning for cross-level interactions. *Information Processing & Management*, 62(1):103932, 2025.
29. Cuifang Zou, Guangquan Lu, Wenzhen Zhang, Xuxia Zeng, Shilong Lin, Longqing Du, and Shichao Zhang. Enhanced graph similarity learning via adaptive multi-scale feature fusion. In *Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence*, pages 7309–7317, 2025.