SIHAN: Semantic-Instance Guided Hypergraph Attention Network With Dual-View Contrastive Learning for Heterogeneous Graph

Shujuan Wei, Huijun Tang, Pengfei Jiao, *Member, IEEE*, and Huaming Wu, *Senior Member, IEEE*

***Abstract*—Heterogeneous graphs are well-suited to modeling the diverse types of entities and their complex interactions in the real world. However, existing Heterogeneous Graph Neural Networks (HGNNs) are typically based on the binary message-passing framework, which struggles to explicitly and finely describe the higher-order semantic information encoded in meta-path instances. Recent studies have attempted to incorporate hypergraphs to capture higher-order interactions, yet most rely on co-occurrence aggregation of meta-path patterns, leading to the blending of instances with different semantics. Moreover, current hypergraph attention mechanisms fail to accurately differentiate the varying contributions of different semantic relations to node representation learning. To address these issues, we propose a novel fine-grained higher-order semantic modeling scheme based on meta-path instances, named SIHAN. First, we treat the complete sequence of intermediate nodes in each path instance as the semantic unit and construct a semantic-driven heterogeneous hypergraph, thereby structurally preserving instance-level semantics. Second, we develop a semantic-sensitive hypergraph attention mechanism that quantifies the contributions of different semantics to target node representation learning by measuring the semantic similarity between nodes and semantic hyperedges. Furthermore, to jointly model higher-order semantics and local topological structures, we construct a semantic-instance view and a structural-neighborhood view, and employ dual-view contrastive learning to encourage the model to simultaneously capture global semantic correlations and local structural regularities. Extensive experiments conducted on four real datasets demonstrate that our proposed model outperforms existing popular models across multiple downstream tasks.**

***Index Terms*—Heterogeneous graph neural network, semantic instance, heterogeneous hypergraph, hypergraph attention.**

Received 5 May 2026; revised 3 July 2026; accepted 10 September 2026. Date of publication 15 September 2026; date of current version 2 October 2026. This work was supported in part by the Emerging Frontiers Cultivation Program of Tianjin University Interdisciplinary Center, Tianjin Natural Science Foundation Project under Grant 25JCYBJC01540, in part by the Tianjin Municipal Science and Technology Major Program under Grant 25ZXZSSS00390, and in part by the Tianjin Future Industry Science and Technology Major Project under Grant 25ZXWCSY00290. Recommended for acceptance by Dr. Yang Lu. *(Corresponding author: Huaming Wu.)*

Shujuan Wei and Huaming Wu are with the Center for Applied Mathematics, KL-AAGDM, Tianjin University, Tianjin 300072, China (e-mail: weishujuan@tju.edu.cn; whming@tju.edu.cn).

Huijun Tang is with the Department of Engineering, Durham University, DH1 3LE Durham, U.K. (e-mail: huijun.tang@durham.ac.uk).

Pengfei Jiao is with the School of Cyberspace, Hangzhou Dianzi University, Hangzhou 310018, China (e-mail: pjiao@hdu.edu.cn).

Digital Object Identifier 10.1109/TNSE.2026.3733704

I. Introduction

GRAPHS, as a powerful data structure, have been widely used to model various entities and their interactions in the real world [9], [73], [97]. However, with the rapid development of information technology and the growing diversity of data types, traditional homogeneous graphs, which contain only a single type of nodes and edges, are insufficient to capture the rich and diverse semantic relationships inherent in complex real-world systems [4], [6], [26], [46], [77]. To overcome this limitation, heterogeneous graph structures that incorporate multiple types of entities and interactions have been extensively adopted, providing a more expressive framework for modeling complex networks [60]. In a heterogeneous graph, different node types are characterized by distinct features, while various edge types represent diverse interaction relationships. This enables the graph to model complex network structures more effectively, retain rich original information, and provide a more comprehensive and natural representation of real-world systems.

Nevertheless, conventional graph algorithms [22], [37], [54] are typically built upon the homogeneity assumption, which limits their generalization ability when dealing with heterogeneous relationships and multi-type features [92]. In recent years, Heterogeneous Graph Neural Networks (HGNNs) [12], [21], [41], [56], [66] have made notable progress in representation learning by propagating and aggregating features across different relation types. Despite these advances, existing methods still face two fundamental challenges in modeling complex semantics.

Challenge 1: Higher-order semantics are weakened or lost in existing heterogeneous graph modeling. Semantics in heterogeneous graphs are often manifested through meta-path instances, which not only characterize cross-type and cross-entity associations between nodes but also reveal the higher-order semantic relationships underlying these interactions [16], [45]. For example, consider the meta-path pattern Author–Paper–Topic–Paper–Author. Its instance $A_1 - P_1 - T_1 - P_2 - A_2$ indicates that authors A_1 and A_2 are associated with the same research topic T_1 via their respective papers P_1 and P_2, thereby uncovering a shared research focus between them.

However, traditional HGNNs [14], [68], [98] are typically based on the binary message-passing framework, which collapses the complex higher-order interactions carried by meta-path instances into pairwise relationships between the end nodes. As illustrated in Fig. 1(a), all meta-path instances are reduced

2327-4697 © 2026 IEEE. All rights reserved, including rights for text and data mining, and training of artificial intelligence and similar technologies. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

Authorized licensed use limited to: TIANJIN UNIVERSITY. Downloaded on October 22,2026 at 01:15:35 UTC from IEEE Xplore. Restrictions apply.

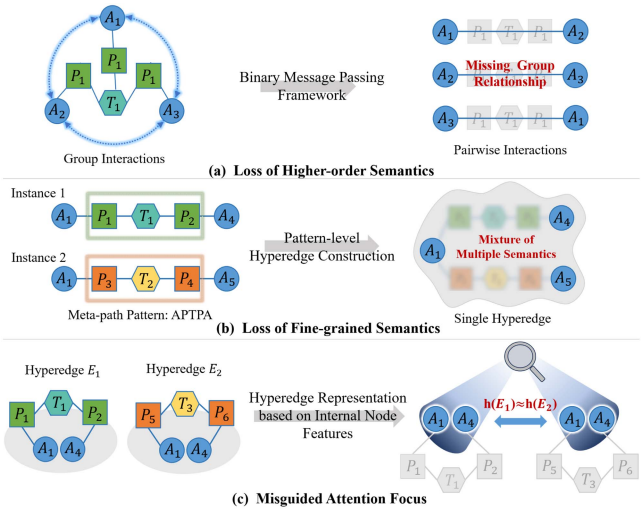


Fig. 1. Three examples of issues in existing HGNNs, where circles, squares, and hexagons correspond to author nodes, paper nodes, and topic nodes, respectively.

to direct links between author nodes. Such structural flattening forces the model to perceive only pairwise adjacency relationships, making it incapable of explicitly expressing the group interactions that pervade real-world networks, such as group collaborations formed by multiple authors around a common research topic.

To overcome the limitation of modeling based on binary relations, recent studies have begun introducing hypergraphs to explicitly characterize higher-order interactions [3], [8], [44]. These approaches connect, via hyperedges, groups of nodes that conform to a specific relational pattern, thereby capturing group-level structures. However, most existing methods model semantics only at the meta-path pattern level, grouping all nodes that satisfy the same meta-path pattern into a single hyperedge while overlooking the specific semantic differences embedded in different meta-path instances. As shown in Fig. 1(b), although the instances $A_1 - P_1 - T_1 - P_2 - A_4$ and $A_1 - P_3 - T_2 - P_4 - A_5$ share the same meta-path pattern, they correspond to entirely different papers and research topics. These semantically disparate instances are indiscriminately mapped to the same hyperedge, causing the loss of fine-grained semantic information during hypergraph construction. Some other methods [33], [38] have attempted to incorporate internal nodes of meta-path instances when constructing hyperedges. However, to reduce computational complexity, they often use only a single representative intermediate node rather than the complete sequence of intermediate nodes. Such simplified hypergraph representations still lack sufficient semantic discriminability and therefore hinder the model's ability to adequately capture the complex and diverse semantic content present in real-world heterogeneous networks.

Challenge 2: Existing hypergraph attention mechanisms struggle to accurately describe the differentiated contributions of distinct semantics. In a hypergraph, each hyperedge is commonly regarded as encoding a latent semantic, and the influence of different semantics on the representation learning of a given node can vary significantly. However, current mainstream approaches [2], [11], [81] generally follow a similar modeling paradigm: they first construct hyperedge representations by aggregating the features of nodes within each hyperedge, and then assign weights to different hyperedges based on these representations so as to regulate their overall influence during message propagation. Essentially, this paradigm assumes that the semantics of a hyperedge can be fully determined by the features of the nodes it contains, and models semantic differences merely as weight adjustments at the hyperedge level.

However, in heterogeneous hypergraphs, the semantics of a hyperedge are often determined not only by the nodes within it, but also, more fundamentally, by the contextual semantics formed by the sequence of cross-type intermediate nodes in the meta-path instance from which the hyperedge is constructed. As illustrated in Fig. 1(c), hyperedges E_1 and E_2 are derived from sequences of paper and topic nodes with entirely different semantics, and therefore encode different semantic contexts. Nevertheless, because existing methods aggregate only the features of the author nodes contained in the hyperedges, semantically distinct hyperedges may yield highly similar hyperedge representations, making it difficult to distinguish their true semantic contributions in the attention modeling process. Furthermore, such methods typically apply attention only to the hyperedge weight matrix, introducing the attention mechanism in the second stage of hypergraph information propagation, namely the "hyperedge-to-node" stage, to model the differentiated effects of different semantics on node representations. In the "node-to-hyperedge" stage, however, the model still relies on the original incidence structure for indiscriminate aggregation, failing to characterize the differentiated influence of distinct semantic information along the complete information propagation path. As a result, their ability to discriminate and represent complex heterogeneous semantics is constrained to some extent.

To address the two aforementioned challenges, this paper proposes a fine-grained higher-order semantic modeling scheme based on meta-path instances. The scheme first treats the cross-type intermediate node sequence within each meta-path instance as the fundamental semantic unit and constructs the semantic-driven heterogeneous hypergraphs accordingly. It then designs a semantic-sensitive hypergraph attention mechanism that accurately captures node-hyperedge interactions induced by different semantic contexts by measuring the semantic similarity between target node features and semantic hyperedge representations.

Different from conventional hypergraph modeling methods, the semantic hyperedges constructed in our framework explicitly preserve the complete semantic information of each meta-path instance, rather than compressing it into coarse pattern-level representations. Meanwhile, the semantic-sensitive hypergraph attention mechanism starts from the interaction strength between nodes and hyperedges, and moves the attention operation forward to the hypergraph construction stage, enabling the differentiated effects of different semantics on node representations to be modeled along the complete information propagation path.

Although the proposed method effectively models higher-order semantic associations, we argue that a comprehensive heterogeneous graph representation learning framework should not focus solely on higher-order correlations at the semantic level but must also incorporate the structural information embedded

Authorized licensed use limited to: TIANJIN UNIVERSITY. Downloaded on October 22,2026 at 01:15:35 UTC from IEEE Xplore. Restrictions apply.

in local neighborhoods. Extensive studies [47], [48], [49] have demonstrated that the direct connections and structural patterns of nodes in local neighborhoods can reflect essential properties such as homophily and interaction stability, which are crucial for enhancing the generalization capability and robustness of the model. For instance, in academic networks, higher-order semantic associations reveal shared research interests among groups of authors, whereas local topological structures describe the closeness between authors and their membership in local communities. These two perspectives provide complementary information from semantic and structural dimensions, respectively.

To achieve a collaborative modeling of these complementary factors, we construct an additional view dedicated to capturing local topological structures alongside the semantic-driven hypergraph view, thereby forming a dual-view representation. Building on this framework, we further design a dual-view contrastive learning mechanism that enforces consistency between the representations of the same node across views. This encourages the model to simultaneously learn semantic correlations and structural regularities, ultimately producing node embeddings with stronger generalization capability and greater robustness. The main contributions of this paper can be summarized as follows:

- *Semantic-Driven Hypergraph Construction:* To achieve fine-grained modeling of higher-order semantics in heterogeneous graphs, we propose a semantic-driven hypergraph construction method. This method constructs hyperedges by treating the complete intermediate node sequence within each meta-path instance as an independent semantic unit. It overcomes the limitation that existing approaches fail to fully preserve fine-grained instance-level semantic information, thereby extending structural preservation from pattern-level statistics to instance-level semantics.
- *Semantic-Sensitive Hypergraph Attention Mechanism:* To precisely characterize the differences in contributions among distinct semantics, we design a semantic-sensitive hypergraph attention mechanism. Specifically, this mechanism first generates semantic hyperedge representations based on semantic units, and then models the association strength between the target node and different semantic hyperedges by measuring the similarity between target node features and semantic hyperedge representations. Through this mechanism, semantic discriminability is introduced at the level of the hypergraph incidence structure, enabling the model to capture the differentiated effects of different semantics on node representations throughout the complete information propagation path.
- *Dual-View Contrastive Learning Framework with Semantic-Structure Synergy:* To effectively integrate higher-order semantic associations and local topological structures, we develop a dual-view contrastive learning framework that coordinates semantic instances and structural neighborhood information. Through the synergistic interaction of the two views, the framework achieves unified modeling of global semantic relations and local structural patterns, thereby learning more comprehensive and more robust node representations.
- *Effective Performance:* We conduct extensive empirical comparisons with existing related works across four real heterogeneous network datasets to validate the effectiveness of our proposed solution.

The rest of the paper is as follows. Section II introduces the related research work. In Section III, we present the key technologies and symbols used in this model. In Section IV, we put forward SIHAN and analyze the key parts. We carry out extensive experiments to verify the superiority of the proposed model in Section V, and make a summary of the paper in Section VI.

II. Related Work

This section reviews the related work from three perspectives that are closely relevant to our approach, including heterogeneous graph neural networks, hypergraph attention mechanisms, and graph contrastive learning methods.

A. Heterogeneous Graph Neural Networks

As data becomes increasingly diverse and complex, heterogeneous graph representation learning has emerged as a prominent research focus, aiming to enhance the quality of node embeddings [28], [92], [95]. Unlike homogeneous graphs, heterogeneous graphs exhibit rich semantic structures, where different types of entities interact through diverse relation types. In such networks, meta-paths, defined as sequences of node and relation types, are commonly used to characterize composite relationships and capture higher-order semantics among heterogeneous entities [39]. Consequently, extensive research [5], [24], [31], [71] has focused on meta-path-based modeling paradigms for HGNNs. Based on their modeling strategies, existing approaches can be broadly divided into random walk-based methods and hybrid relation-based methods. Recent extensions have further introduced subgraph-based and hypergraph-based modeling frameworks to capture more complex structural interactions.

Random walk-based methods [13], [27], [29], [70], [87] represent one of the earliest and most widely adopted approaches for heterogeneous graph representation learning. These methods typically design meta-path-guided random walk strategies to generate sequences of nodes that preserve heterogeneous semantics, and then employ sequence modeling techniques, such as the skip-gram model, to learn node embeddings from the generated sequences. By constraining random walks with predefined meta-paths, heterogeneous relations are implicitly transformed into homogeneous node contexts, enabling scalable embedding learning. Metapath2Vec [13] is a representative work in this category. It performs random walks constrained by predefined meta-path schemas and applies a skip-gram model to learn node embeddings, thereby effectively modeling both the heterogeneous structural and semantic contexts of the graph. Building upon this idea, HINE [29] utilizes meta-paths to guide random walks and formulates an objective function that computes the similarity between node pairs within both the meta-path context and the embedding space. The model then optimizes node representations by minimizing the KL divergence between these similarity distributions. HetGNN [87]

Authorized licensed use limited to: TIANJIN UNIVERSITY. Downloaded on October 22,2026 at 01:15:35 UTC from IEEE Xplore. Restrictions apply.

integrates a restart-based random walk mechanism into HGNNs to sample a fixed number of neighbors, followed by a neural aggregation process that employs type-specific encoders to sequentially aggregate features from same-type and cross-type neighbors. To further enhance the expressiveness of random walks in heterogeneous networks, HeteSpaceyWalk [27] proposes a heterogeneous spacey random walk that adaptively skips previous states to approximate higher-order dependencies while maintaining meta-path-guided semantics, while HHNE [70] maps meta-path-based embeddings into hyperbolic space to better capture hierarchical and scale-free properties commonly observed in heterogeneous graphs.

While effective in preserving meta-path-guided proximity, random walk-based methods typically encode heterogeneous information in an implicit manner through sampled node sequences. As a result, the semantics of different relation types and the structural roles of nodes along meta-paths are not explicitly modeled during representation learning. Moreover, the learned representations are largely independent of downstream tasks, which restricts their flexibility in modeling task-specific heterogeneous relations. To address these limitations, subsequent studies propose hybrid relation-based modeling approaches [14], [15], [16], [68], [96] that explicitly incorporate heterogeneous relations and meta-path structures into neural message-passing and attention mechanisms. Hin2Vec [15] jointly learns node embeddings and meta-path embeddings by modeling the relationships between node pairs and meta-path instances, enabling the representation to encode multi-relational semantics. HAN [68] further designs a hierarchical attention mechanism that operates at two levels: node-level attention to differentiate the importance of node neighbors under a given meta-path, and semantic-level attention to weigh the contributions of different meta-paths. MAGNN [16] extends HAN by explicitly modeling intermediate nodes within meta-path instances and performing instance-level aggregation before semantic fusion, thereby enriching the semantic representation of meta-path-based neighborhoods. More recent studies have focused on meta-path selection and learning mechanisms. PM-HGNN [96] formulates meta-path construction as a node-dependent process, dynamically generating meta-path instances conditioned on individual node contexts rather than relying on globally predefined meta-path schemas. MP-GNN [14] models relation composition as a learnable process within graph neural networks, enabling the model to identify effective multi-step relational patterns without explicitly enumerating meta-path structures.

While the aforementioned methods explicitly incorporate heterogeneous relationships and meta-path semantics into network architectures, their semantic modeling is still primarily organized around individual meta-paths or their weighted combinations. As a result, complex relational patterns that span multiple meta-paths are often fragmented into separate semantic views. To transcend the limitations of linear path-based semantics, recent studies [10], [59], [84], [89], [90] have explored subgraph-based and metagraph-based modeling paradigms, which treat richer structural patterns as fundamental modeling units. mg2vec [90] proposes metagraphs that consist of multiple interconnected meta-paths, enabling the modeling of composite semantic structures that cannot be captured by a single meta-path. HCMGNN [89] integrates causal inference principles into its modeling framework by identifying and aggregating subgraphs that are derived from causally relevant meta-paths. This approach enhances the interpretability of representations, particularly in domains such as biomedical networks. HG-SCC [84] introduces subgraph-aware convolution mechanisms tailored for few-shot learning, where subgraphs relevant to the task are utilized to enhance node classification performance under limited supervision.

Despite their enhanced expressive capability, subgraph-based and metagraph-based approaches remain grounded in ordinary graph representations, where higher-order interactions are implicitly encoded through induced substructures. Such methods focus on capturing localized structural patterns, but do not explicitly represent collective interactions among multiple entities. To overcome this limitation, recent research [3], [8], [33] has turned to hypergraph structures. By allowing hyperedges to connect arbitrary sets of nodes, hypergraphs offer a natural and explicit framework for directly modeling group-level collaborative relationships among multiple entities, thereby advancing the representation of higher-order interactions in heterogeneous graph learning. For instance, MVH^{2}GCN [3] introduces a multi-view heterogeneous hypergraph convolution framework that transforms heterogeneous information networks into multiple semantic hypergraph views and performs joint node-hyperedge message propagation to efficiently capture higher-order and complementary semantic dependencies. MGA-HHN [33] generates diverse hyperedges based on different semantic granularities to broaden the coverage of higher-order interactions. Through the joint learning of multi-granularity hyperedges, structural information from various levels of abstraction collectively optimizes node representations. HHMDA [8] proposes a hybrid framework that integrates heterogeneous hypergraph convolution for explicit higher-order interaction modeling with multi-scale graph convolution for complementary pairwise structural learning. However, these methods typically depend on pattern-level hyperedge construction strategies, where nodes participating in the same meta-path pattern or neighborhood context are grouped into hyperedges. Such constructions fail to distinguish fine-grained semantic differences among distinct meta-path instances, causing instance-level semantic information to be blurred and ultimately limiting the precision of higher-order semantic expression.

B. *Hypergraph Attention Mechanisms*

Attention mechanisms have been widely adopted in graph neural networks as a flexible and effective tool for selectively modeling relational information, enabling more expressive and interpretable message aggregation [30], [74], [76], [79], [88], [91].

Early studies [7], [35], [53], [62], [64] mainly focus on homogeneous graphs, where attention is introduced to distinguish the relative importance of node neighbors during information propagation. GAT [64] incorporates self-attention into neighborhood aggregation by computing attention coefficients based on feature interactions between a target node and its neighbors. Through multi-head attention, GAT captures diverse structural

Authorized licensed use limited to: TIANJIN UNIVERSITY. Downloaded on October 22,2026 at 01:15:35 UTC from IEEE Xplore. Restrictions apply.

patterns while alleviating the limitations of uniform aggregation. Subsequent works further extend this idea by refining neighbor selection and attention design. SuperGAT [35] proposes a self-supervised graph attention framework that predicts edges using two compatible attention forms, coupling node-level attention with edge-centric supervision to explicitly identify friendly neighborhoods and guide attention design beyond standard message-passing schemes. GAAN [7] integrates an adaptive attention mechanism to dynamically weigh neighbors and employs multi-head graph convolution to enhance feature representation for node classification. TWC-GNN [53] unifies directed graph topology with self-attention by designing direction-aware attention guided by structural roles, enabling node classification to jointly model asymmetric connectivity patterns and feature interactions within a single coherent attention framework. GATransformer [62] integrates graph structural information into the self-attention mechanism of Transformer, and utilizes graph attention to constrain token relationships, enabling structured and interpretable attention modeling for brain tumor detection.

Motivated by the success of attention mechanisms in homogeneous graphs, recent studies [16], [23], [42], [43], [52], [68], [94] have extended attention-based frameworks to heterogeneous graphs to capture complex interactions among multiple types of nodes and relations. Representative models such as HAN and MAGNN, discussed in the previous section, employ hierarchical attention mechanisms to model node-level importance and semantic-level contributions, significantly enhancing the ability to capture heterogeneous semantics. Beyond these approaches, several studies have further advanced attention modeling in heterogeneous graphs. For instance, HMAM [23] learns node representations by decomposing heterogeneous networks into subgraphs and employing hierarchical attention, including structure enhancement attention, subgraph fusion attention, and semantic attention, all of which leverage higher-order attribute and semantic neighbors for guidance. DHAN [94] models bi-typed multi-relational graphs using a dual hierarchical attention architecture that combines intra-type attention for homogeneous neighbors and inter-type attention for heterogeneous neighbors, jointly capturing complex relational dependencies. RioGNN [52] addresses noisy neighbor aggregation in multi-relational graphs by incorporating reinforced relation-aware selection, where label-guided similarity and reinforcement learning are used to identify informative relation-specific neighborhoods. From a multi-level perspective, HDGAT [43] constructs a multi-level semantic graph by introducing global nodes and a graph attention mechanism enhanced by Squeeze-and-Excitation Networks, and applies unidirectional attention to selectively propagate relevant semantic information while suppressing noisy message passing. NGGAN [42] generates multi-level neighbor graphs via adaptive sampling and applies hierarchical attention over nodes, paths, and neighbor graphs to capture higher-order structural interactions in heterogeneous graphs.

While attention mechanisms have proven effective in heterogeneous graph neural networks for differentiating relational importance, the introduction of hypergraph structures brings new challenges in modeling the varying contributions of nodes within hyperedges and the semantic diversity across different hyperedges. Consequently, recent studies [2], [33], [34], [67], [81] have extended attention mechanisms to hypergraph neural networks to enable adaptive weighting at the node-hyperedge and hyperedge-level interactions. The hypergraph attention model in [2] generalizes attention-based message passing from graphs to hypergraphs by modeling node-hyperedge interactions, where it assigns different weights to nodes within the same hyperedge, enabling weighted aggregation of node features when forming hyperedge representations. HCAN [67] models interactions between nodes and hyperedges by grouping and weighting hyperedges of different orders, and then performing order-specific aggregation. CIAH [81] constructs interaction representations using attentive hypergraph encoding, applies attribute-level attention with saliency-guided and cluster-based consistency constraints, and jointly co-clusters interactions and attention distributions for explainable modeling. In heterogeneous scenarios, attention-based heterogeneous hypergraph neural networks have been proposed to model interactions among multiple entity types. For instance, HHGSA [34] models higher-order user-item interactions via a heterogeneous hypergraph and applies attentive vertex and hyperedge aggregation to capture user and item importance for social recommendation. MGA-HHN [33] further designs a dual-layer multi-granularity attention mechanism that integrates node-level Transformer attention to capture long-range interactions and then incorporates hyperedge-level attention to adaptively fuse multi-semantic information, thereby learning node representations that encompass rich higher-order relationships. A common characteristic of these approaches is that node-hyperedge interactions are primarily characterized based on aggregations of internal node features, which may lead to semantically different higher-order structures being treated similarly.

C. Graph Contrastive Learning

Graph contrastive learning has emerged as an effective self-supervised paradigm for representation learning on graphs, aiming to enhance robustness and generalization without relying on labeled data [57], [58], [82]. Early studies [1], [25], [55], [63], [65], [78], [80], [83] primarily focus on homogeneous graphs and establish fundamental contrastive principles through different view construction strategies.

A seminal work, DGI [65], introduces a global-local contrastive objective by constructing node-level representations as local views and a graph summary as a global view, maximizing mutual information to encourage node embeddings to capture global structural characteristics. Building upon the global-local contrastive paradigm, MVGRL [25] constructs two complementary views based on graph diffusion and first-order neighbors, enabling contrastive learning across different structural perspectives to capture both local connectivity and long-range dependencies. SLGGCL [1] forms local subgraph views and global diffusion graph views through perturbation, masking and diffusion strategies, and applies multi-scale contrastive learning to jointly capture local and global structural information for node classification. Extending this idea, GALCL [80] generates a global graph representation together with K-means-based local subgraph embeddings, and applies multi-view cross-scale

Authorized licensed use limited to: TIANJIN UNIVERSITY. Downloaded on October 22,2026 at 01:15:35 UTC from IEEE Xplore. Restrictions apply.

contrastive learning to jointly encode normal and anomalous node representations for graph anomaly detection.

Another major research direction focuses on data augmentation-based contrastive learning, where different graph views are generated through stochastic perturbations. GraphCL [83] applies a set of topology-level and feature-level augmentations to construct multiple corrupted views of the same graph, enforcing invariance across augmented representations. GCC [55] constructs contrastive views through subgraph sampling and random walks, enabling pre-training of graph neural networks by pulling together representations of structurally similar subgraphs while pushing apart representations of unrelated ones. BGRL [63] learns node representations by bootstrapping graph augmentations, using an online-target network architecture to align representations across perturbed graph views without negative samples or explicit contrastive objectives. More recently, FA-GCL [78] incorporates feature augmentation strategies alongside structural perturbations, constructing contrastive views that jointly preserve attribute semantics and graph topology, thereby improving representation quality under feature noise.

Collectively, these methods establish the core principles of graph contrastive learning, including view construction, invariance modeling, and alignment across local and global representations. However, extending contrastive learning to heterogeneous graphs introduces additional challenges due to the coexistence of multiple node types, relation types, and semantic patterns. Early efforts [32], [50], [51], [69], [86] adapt contrastive objectives by explicitly constructing heterogeneous views. DMGI [51] is among the earliest works that introduce contrastive learning into heterogeneous graph representation learning. It contrasts the original node representations with those from a perturbed graph, and introduces a consistency regularization framework to integrate node representations across different relational perspectives for efficiently learning multi-attribute networks. HeCo [69] proposes a co-contrastive framework that constructs a network schema view and a meta-path view, and designs a view-masking strategy to enable mutual supervision between the views. Similarly, GCL-SS [86] explicitly separates heterogeneous information into a structural view that captures graph topology and a semantic view that encodes meta-path relationships, and employs contrastive learning to align representations learned from these complementary perspectives. M2CHGNN [50] constructs homogeneous subgraphs from meta-paths and meta-structures, hierarchically attends to neighbors and subgraphs, and uses contrastive learning to align feature-based and topology-based representations. ASHGCL [32] designs lower-order, higher-order, and attribute-similarity views for heterogeneous graphs, and applies attribute-enhanced positive sampling with local and global contrastive learning to capture multi-scale node representations.

More recently, contrastive learning has been extended to hypergraph settings to accommodate group-based and multi-entity interactions [20], [38], [75]. MMHCL [20] constructs user-user and item-item hypergraphs to capture second-order multimodal semantics, and applies synergistic contrastive learning to align first-order interactions with hypergraph representations for recommendation. OFSH [75] predicts hyperedges by propagating hyperedge orders to weight neighbors, using max-min pooling and key node guided augmentation, and aligns augmented views via triadic contrastive learning. SNHE [38] jointly models node-level and hyperedge-level representations and defines dual contrastive objectives across different structural entities, promoting consistency between nodes and hyperedges while capturing heterogeneous higher-order interactions.

These studies collectively demonstrate that multi-view contrastive learning plays a crucial role in integrating heterogeneous structural and semantic information. By constructing complementary views across different granularities, relational semantics, and structural scopes, contrastive learning provides a powerful mechanism for learning robust and discriminative representations in complex heterogeneous and higher-order graph settings.

Building on the above analysis, this paper proposes a unified heterogeneous graph representation learning framework that integrates instance-level semantic modeling, semantic contribution differentiation, and the joint learning of higher-order semantics and local structures, aiming to produce node representations that are fine-grained, interpretable, and robust. To facilitate a clear comparison, Table I summarizes representative methods and their adopted modeling components.

III. Preliminary

In this section, we introduce the fundamental concepts of heterogeneous information networks, meta-paths, and hypergraphs, with the corresponding symbols and definitions summarized in Table II.

Definition 1 (Heterogeneous Information Network): A heterogeneous information network can be defined as $\mathcal{H}_\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{A}, \mathcal{R}, \phi, \varphi)$, where $\mathcal{V}$ and $\mathcal{E}$ respectively denote the set of nodes and edges. For each node $v_i \in \mathcal{V}$ and $e_j \in \mathcal{E}$, there exists a corresponding type mapping function $\phi : \mathcal{V} \to \mathcal{A}$ and $\varphi : \mathcal{E} \to \mathcal{R}$, where $\mathcal{A}$ and $\mathcal{R}$ respectively represent the sets of node types and edge types, and $|\mathcal{A}| + |\mathcal{R}| > 2$.

Definition 2 (Meta-path Pattern): A meta-path p in the heterogeneous graph $\mathcal{H}_\mathcal{G}$ is represented as a sequence in the form of $A_1 \xrightarrow{R_1} A_2 \xrightarrow{R_2} \cdots \xrightarrow{R_l} A_{l+1}$ (abbreviated as $A_1 A_2 \cdots A_{l+1}$), where $A_i \in \mathcal{A}$ denotes the node type of the i-th position in the path, and $R_i \in \mathcal{R}$ denotes the relation type connecting nodes of type A_i and A_{i+1}. It implies a composite relation $R = R_1 \circ R_2 \circ \cdots \circ R_l$ between head node type A_1 and tail node type A_{l+1}, where $\circ$ denotes the composition operator on relations.

Definition 3 (Meta-path Instance): Given a meta-path pattern $A_1 \xrightarrow{R_1} A_2 \xrightarrow{R_2} \cdots \xrightarrow{R_l} A_{l+1}$, a meta-path instance is defined as a node sequence $(v_1, v_2, \ldots, v_{l+1})$ that satisfies the following constraints: (1) type consistency, where $\phi(v_i) = A_i$ for $\forall i = 1, \ldots, l+1$, and (2) relation consistency, where $(v_i, v_{i+1}) \in E$ and $\varphi(v_i, v_{i+1}) = R_i$ for $\forall i = 1, \ldots, l$.

Definition 4 (Hypergraph): Hypergraph, as a generalized graph, is formally defined as $\mathcal{G} = (\mathcal{V}_\mathcal{G}, \mathcal{E}_\mathcal{G})$, where $\mathcal{V}_\mathcal{G}$ denotes the set of nodes and $\mathcal{E}_\mathcal{G}$ is the set of hyperedges. The structure of a hypergraph can be equivalently encoded using a binary incidence matrix $Q \in R^{N \times S}$, where N and S denote the number of nodes and hyperedges, respectively. For $\forall v \in \mathcal{V}_\mathcal{G}$ and $e \in \mathcal{E}_\mathcal{G}$, if

Authorized licensed use limited to: TIANJIN UNIVERSITY. Downloaded on October 22,2026 at 01:15:35 UTC from IEEE Xplore. Restrictions apply.

TABLE I
COMPARISON OF REPRESENTATIVE METHODS IN TERMS OF KEY MODELING COMPONENTS

Method	Heterogeneous Graph	Meta-path Pattern	Group-Level Interaction	Local Topological Structure	Attention Mechanism	Instance-Level Semantic Modeling	Contrastive Learning
[13], [27], [29], [70], [90]	✓	✓	×	✓	×	×	×
[15]	✓	✓	×	✓	×	✓	×
[14], [23], [42], [43], [68], [84], [89], [94]	✓	✓	×	✓	✓	×	×
[16], [96]	✓	✓	×	✓	✓	✓	×
[3]	✓	✓	✓	×	✓	×	×
[33]	✓	✓	✓	×	✓	✓	×
[8]	✓	×	✓	✓	✓	×	×
[34]	✓	×	✓	×	✓	×	×
[7], [53], [62], [64]	×	×	×	✓	✓	×	×
[35]	×	×	×	✓	✓	×	✓
[52], [87]	✓	×	×	✓	✓	×	×
[2], [81]	×	×	✓	×	✓	×	×
[67]	×	×	✓	×	✓	×	✓
[1], [25], [55], [63], [65], [78], [80], [83]	×	×	×	✓	×	×	✓
[51]	✓	×	×	✓	×	×	✓
[32], [50], [69], [86]	✓	✓	×	✓	✓	×	✓
[20]	✓	×	✓	✓	×	×	✓
[38]	✓	✓	✓	×	✓	✓	✓
[75]	✓	×	✓	×	✓	×	✓
SIHAN (Ours)	✓	✓	✓	✓	✓	✓	✓

Note: ✓ indicates that the corresponding component is explicitly modeled in the method, while × denotes that the component is not involved in the original formulation. Components are defined as follows: Heterogeneous Graph: graph with multiple node and/or edge types; Meta-path Pattern: schema-level type sequence capturing semantic relations; Instance-Level Semantic Modeling: representation learning based on concrete semantic instances; Group-Level Interaction: higher-order interactions among multiple nodes explicitly defined as a group (e.g., via hyperedges); Local Topological Structure: explicit modeling of a node's local connectivity, reflecting first-order adjacency and neighborhood characteristics. (e.g., density and degree); Attention Mechanism: a mechanism that assigns different importance to elements (e.g., neighbors, hyperedges, or views) during aggregation; Contrastive Learning: representation learning by contrasting positive pairs against negative pairs.

TABLE II
SYMBOLS AND EXPLANATIONS

Symbol	Explanation
$\mathcal{H}_{\mathcal{G}}$	Heterogeneous information network
$\mathcal{G}$	Hypergraph
$\mathcal{E}$	Set of edges
$\mathcal{V} = \{v_1, v_2, \ldots, v_N\}$	Set of nodes
$\mathcal{A} = \{A_1, A_2, \ldots, A_{l+1}\}$	Set of node types
$\mathcal{R} = \{R_1, R_2, \ldots, R_l\}$	Set of edge types
$P = \{p^1, p^2, \cdots, p^M\}$	Set of meta-paths
$F = \{f_1, f_2, \ldots, f_N\}$	Initial node feature matrix
$X = \{x_1, x_2, \ldots, x_N\}$	Node feature matrix after projection
E_i	Set of hyperedges associated with v_i
N	Number of nodes
S	Number of hyperedges
$Q \in R^{N \times S}$	Incidence matrix
$A \in R^{N \times N}$	Adjacency matrix
$W \in R^{S \times S}$	Hyperedge weight matrix
H^{p^m}	Semantic-instance embedding under meta-path p^m
L^{p^m}	Structural-neighborhood embedding under meta-path p^m
β^{p^m}	Semantic-instance view weight of meta-path p^m
α^{p^m}	Structural-neighborhood view weight of meta-path p^m
$N_i^{p^m}$	Neighbor set of v_i under meta-path p^m
Z^{sem}	The final embedding of semantic-instance view
Z^{str}	The final embedding of structural-neighborhood view

node v is connected to hyperedge e, then $q(v, e) = 1$; otherwise $q(v, e) = 0$.

IV. SIHAN

The proposed SIHAN consists of three core modules: the semantic-instance view module, the structural-neighborhood view module, and the dual-view contrastive learning module. In the semantic-instance view module, we treat the complete intermediate node sequence of each meta-path instance as a semantic unit and transform the heterogeneous information network into multiple heterogeneous hypergraphs, thereby explicitly and finely modeling higher-order semantic relations. On this basis, we further design a semantic-sensitive hypergraph attention mechanism, which constructs attention associations at the node-hyperedge level by measuring the similarity between target node features and semantic hyperedge representations, so as to capture the differentiated effects of different semantic information on node representations along the complete information propagation path. In the structural-neighborhood view, we project the heterogeneous information network into multiple homogeneous graphs based on meta-paths, thereby supplementing the structural characteristics of nodes within their local neighborhoods. In the dual-view contrastive learning module, the above two views are jointly optimized through collaborative contrastive learning, enabling them to complement and supervise each other, ultimately producing higher-quality node embeddings. The framework of SIHAN is illustrated in Fig. 2.

A. Node Feature Transformation

Due to the diversity of node types in heterogeneous graphs, it is essential to map the features of all types of nodes into a shared space before learning the node embeddings. For a node v_i of type ϕ_i, the projection process can be defined as follows:

$$x_i = \sigma\left(W_{\phi_i} \cdot f_i + b_{\phi_i}\right), \tag{1}$$

Authorized licensed use limited to: TIANJIN UNIVERSITY. Downloaded on October 22,2026 at 01:15:35 UTC from IEEE Xplore. Restrictions apply.

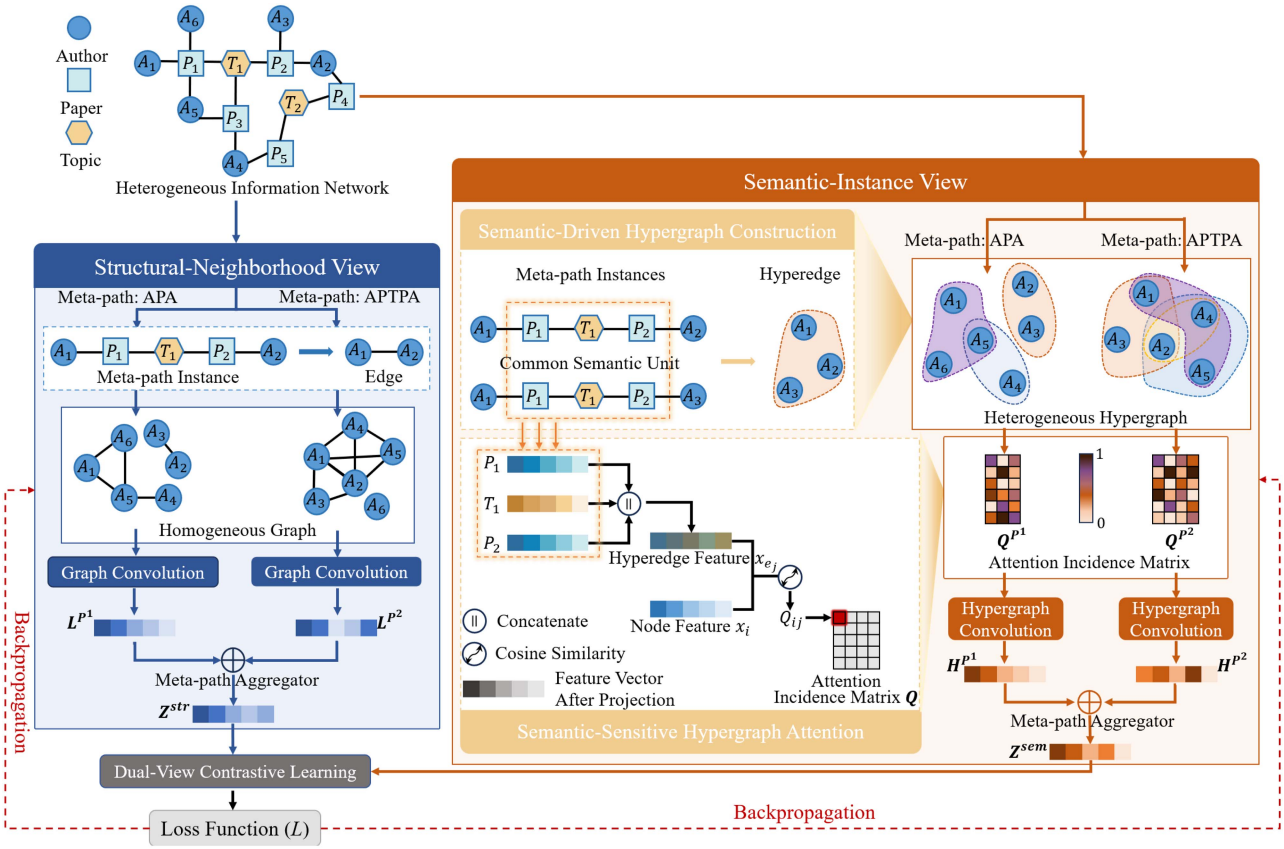


Fig. 2. The framework of SIHAN. First, two views are generated based on meta-paths. These views are then separately used to generate semantic and structural embeddings for nodes. Finally, the embeddings are projected into the same space, and contrastive optimization is performed by maximizing mutual information to enable collaborative supervision between higher-order semantics and local structures.

where W_{ϕ_i} and b_{ϕ_i} represent the corresponding mapping weight matrix and bias vector, respectively, x_i denotes the feature of node v_i after projection, f_i refers to the initial feature of node v_i, and $\sigma(\cdot)$ is the activation function applied to the result.

B. Node Embedding Encoder in Semantic-Instance View

1) Semantic-Driven Hypergraph Construction: A hypergraph characterizes the relationships among multiple objects through hyperedges, enabling it to model higher-order correlations that traditional graph structures inherently struggle to capture [18]. Meta-path instances naturally encode fine-grained semantic information, as each instance corresponds to a concrete sequence of heterogeneous nodes and relations. Therefore, in the semantic-instance view, we model the heterogeneous information network as a collection of heterogeneous hypergraphs constructed on meta-path instances, thereby capturing complex and fine-grained higher-order semantic interactions within the network.

Given a heterogeneous information network $\mathcal{H}_\mathcal{G}$ and its corresponding set of M meta-paths $\mathcal{P} = \{p^1, p^2, \ldots, p^M\}$, where p^m denotes the m-th specific meta-path pattern, we consider an arbitrary target node v_i. For this node, let $p_i^{(m)} = \{\pi_{i1}^{(m)}, \pi_{i2}^{(m)}, \ldots, \pi_{iK}^{(m)}\}$ denote the set of all meta-path instances that start from v_i and follow the meta-path p^m, where $\pi_{ik}^{(m)}$ represents the k-th specific meta-path instance. For each meta-path instance, the sequence of intermediate nodes obtained by removing the start and end nodes is defined as a semantic unit. Subsequently, the target node v_i together with all nodes that share the same semantic unit with it are jointly constructed as a hyperedge.

This process transforms the heterogeneous information network $\mathcal{H}_\mathcal{G}$ into multiple heterogeneous hypergraphs $\{\mathcal{G}^1, \mathcal{G}^2, \ldots, \mathcal{G}^M\}$, each corresponding to a different meta-path layer.

2) Semantic-Sensitive Hypergraph Attention: The contextual semantics characterized by different semantic units vary substantially, and the semantic hyperedges constructed from them contribute unequally to the representation learning of target nodes in higher-order relation modeling. To address this issue, we design a semantic-sensitive hypergraph attention mechanism. Different from existing methods that characterize semantic contributions by learning global hyperedge weights, the proposed hypergraph attention mechanism models the effects of different semantic information at the level of the hypergraph incidence structure from the perspective of node-hyperedge interaction strength. Specifically, instead of representing whether a node belongs to a hyperedge in a binary manner, the proposed mechanism first generates semantic hyperedge representations

Authorized licensed use limited to: TIANJIN UNIVERSITY. Downloaded on October 22,2026 at 01:15:35 UTC from IEEE Xplore. Restrictions apply.

based on semantic units, and then finely characterizes the associations between nodes and hyperedges by measuring the semantic similarity between target node features and semantic hyperedge representations. In this way, different semantic hyperedges can make differentiated contributions to the update of target node representations at various stages of information propagation according to their degree of semantic compatibility with the target node. As a result, the model can more accurately focus on relation patterns that are semantically relevant to the target node during information aggregation, while effectively suppressing the interference introduced by irrelevant or noisy semantics, thereby achieving fine-grained modeling of complex heterogeneous semantics.

Formally, for the target node v_i and its associated j-th hyperedge e_j, we first extract the intermediate node sequence $(n_1, n_2, \ldots, n_T)$ corresponding to the semantic unit that constructs this hyperedge. Considering that different meta-path types correspond to different semantic structures, we learn an independent set of parameters for each meta-path p^m to capture its specific semantic composition pattern. On this basis, the features of all nodes in the sequence are concatenated and then encoded into a unified hyperedge representation through a mapping function with nonlinear activation. The computation process is given as follows:

$$f_{e_j} = \sigma\left(W_c^{(p^m)}\left[x_{n_1} \| x_{n_2} \| \cdots \| x_{n_T}\right] + b_c^{(p^m)}\right), \tag{2}$$

where $W_c^{(p^m)}$ and $b_c^{(p^m)}$ denote the learnable parameters corresponding to the meta-path type p^m, and $\|$ denotes vector concatenation.

To ensure that the node and hyperedge features are aligned within a comparable semantic space, we apply separate linear projection functions to the target node features and the hyperedge features, mapping them into a shared space and yielding the transformed node representation x_i and the hyperedge representation x_{e_j}. Subsequently, we compute their semantic compatibility using the cosine similarity function, and use normalized similarity as the relative importance weight of the hyperedge to the target node. The computation is given as follows:

$$Q_{ij} = \frac{\exp\left(\sigma\left(\cos\left(x_i, x_{e_j}\right)\right)\right)}{\sum_{e_k \in E_i} \exp\left(\sigma\left(\cos\left(x_i, x_{e_k}\right)\right)\right)}, \tag{3}$$

where E_i represents the set of hyperedges associated with node v_i.

Based on the normalized semantic similarities, an attention-based incidence matrix Q can be constructed to replace the conventional binary incidence matrix in hypergraphs. Unlike the hard association scheme that merely indicates whether a node belongs to a given hyperedge, Q explicitly encodes the interaction strength between nodes and semantic hyperedges, thereby enabling the influence of different semantic information on nodes to be clearly distinguished. In addition, by moving the attention operation forward to the hypergraph construction stage, the proposed method allows the differentiated effects of distinct semantics on node representations to be modeled throughout the entire information propagation process, thus enhancing the model's ability to represent complex higher-order semantic relations.

Finally, we update the node feature representation by transferring and aggregating information from neighbor nodes using the hypergraph convolution layer [17], which is set to a single layer in this paper. The update is performed according to the following formula:

$$H^{p^m} = \sigma\left(Q^{p^m} W {Q^{p^m}}^T X \theta_1\right), \tag{4}$$

where Q^{p^m} and H^{p^m} represent the incidence matrix and output feature matrix of nodes under the meta-path p^m, respectively. X and θ_1 denote the input feature matrix and the trainable weight matrix. W indicates the hyperedge weight matrix, which is set as the identity matrix in this paper.

Through the above operations, we obtain the semantic-instance embedding $\{h_i^{p^1}, h_i^{p^2}, \ldots, h_i^{p^M}\}$ for node v_i across M meta-paths. Since each meta-path captures distinct semantic information, we introduce a meta-path aggregator to facilitate the integration of information across different meta-paths. To begin, we calculate the importance of each meta-path under the semantic-instance view:

$$W_h^{p^m} = \frac{1}{|V|} \sum_{i \in V} a_{sem}^T \cdot \tanh\left(W_{sem} \cdot h_i^{p^m} + b_{sem}\right), \tag{5}$$

$$\beta^{p^m} = \frac{\exp\left(W_h^{p^m}\right)}{\sum_{t=1}^{M} \exp\left(W_h^{p^t}\right)}, \tag{6}$$

where W_{sem} and b_{sem} are learnable parameters, a_{sem} is the attention vector at the meta-path type level, and β^{p^m} denotes the importance of meta-path p^m to the target node.

We obtain the final node embedding z_i^{sem} learned from the semantic-instance view by weighted summation of the embedding vectors obtained from all meta-paths:

$$z_i^{sem} = \sum_{m=1}^{M} \beta^{p^m} \cdot h_i^{p^m}. \tag{7}$$

C. Node Embedding Encoder in Structural-Neighborhood View

The purpose of this view is to model the local topological structures among nodes in the heterogeneous graph to learn structural regularities. Specifically, given the target node v_i and its corresponding set of M meta-paths $\mathcal{P} = \{p^1, p^2, \ldots, p^M\}$, for each meta-path p^m, we first identify all nodes that are connected to v_i via p^m in the original heterogeneous information network, thereby determining its structural-neighborhood under the constraint of this meta-path. On this basis, a corresponding homogeneous graph is constructed by taking the target node and its associated neighboring nodes as the node set. For any two nodes v_i and v_j, if there exists at least one connection path conforming to meta-path p^m in the original heterogeneous graph, an edge is established between them in the homogeneous graph, thereby transforming the original higher-order relation into a first-order adjacency relation. Through this process, multiple homogeneous graphs can be obtained under different meta-path

Authorized licensed use limited to: TIANJIN UNIVERSITY. Downloaded on October 22,2026 at 01:15:35 UTC from IEEE Xplore. Restrictions apply.

constraints, in which node types are unified, while edges characterize the structural association relationships among nodes.

Considering that GCN [37] has achieved significant success in homogeneous graph learning and possesses a powerful capability to explore key information in data, we utilize a single graph convolution layer to perform message passing and aggregation in this view, expressed by the following formula:

$$L^{p^m} = \sigma\left(\widetilde{D}^{-\frac{1}{2}}(A^{p^m}+I)\widetilde{D}^{-\frac{1}{2}}X\theta_2\right), \tag{8}$$

where X and θ_2 respectively indicate the input feature matrix and the learnable weight matrix. L^{p^m} and A^{p^m} represent the output features of nodes and adjacency matrix under the meta-path p^m. $\widetilde{D}$ denotes the degree matrix corresponding to $A^{p^m}+I$.

After obtaining the structural-neighborhood node embedding $\{l_i^{p^1}, l_i^{p^2}, \ldots, l_i^{p^M}\}$for node v_i under M meta-paths, we employ the same attention mechanism as that used in the semantic-instance view to aggregate them, thereby obtaining the final embedding representation z_i^{str} for node v_i in the structural-neighborhood view:

$$W_l^{p^m} = \frac{1}{|V|}\sum_{i\in V} a_{str}^T \cdot \tanh\left(W_{str}\cdot l_i^{p^m} + b_{str}\right), \tag{9}$$

$$\alpha^{p^m} = \frac{\exp\left(W_l^{p^m}\right)}{\sum_{t=1}^{M}\exp\left(W_l^{p^t}\right)}, \tag{10}$$

$$z_i^{str} = \sum_{m=1}^{M}\alpha^{p^m}\cdot l_i^{p^m}, \tag{11}$$

where α^{p^m} denotes the importance of meta-path p^m for the target node, W_{str} and b_{str} are learnable parameters, and a_{str} represents the attention vector at the meta-path type level.

D. Contrastive Learning

Owing to the complexity of the heterogeneous graph, some redundant information may be introduced during the process of learning node embeddings, which can degrade the quality of the learned embeddings. To address this issue, we adopt a graph contrastive learning approach to maximize the mutual information between node representations in the semantic-instance and structural-neighborhood views, so as to facilitate cooperation and supervision between the two views and enhance the accuracy of node representation in the model.

First, we project the embedding z_i^{sem} from the semantic-instance view and z_i^{str} from the structural-neighborhood view into the same contrastive loss space:

$$z_i^{sem}_p = w_2\sigma\left(w_1 z_i^{sem} + b_1\right) + b_2, \tag{12}$$

$$z_i^{str}_p = w_2\sigma\left(w_1 z_i^{str} + b_1\right) + b_2, \tag{13}$$

where $z_i^{sem}_p$ and $z_i^{str}_p$ represent the node embedding after projection, and the same learnable parameters w_1, w_2, b_1, b_2 are utilized across both perspectives.

For the selection of positive and negative samples, we consider that the greater the number of meta-paths connecting two nodes, the higher their correlation.

For nodes v_i and v_j, we use the function $N_i(\cdot)$ to compute the number of connecting meta-paths:

$$N_i(j) = \sum_{m=1}^{M} 1_{N_i^{p^m}}(j), \tag{14}$$

where $N_i^{p^m}$ denotes the neighbor set of v_i based on the meta-path p^m, $1_A(x)$ is an indicator function, and if $x \in A, 1_A(x) = 1$, otherwise 0.

Based on $N_i(j)$, we sort the neighbor nodes in descending order, set a positive sample threshold K, and select the top K nodes as positive samples P_{os} and all the remaining nodes as negative samples N_{eg}. If the number of neighbor nodes is less than K, all neighbor nodes are treated as positive samples.

After sampling the positive sample set P_{os} and negative sample set N_{eg}, we compute the loss under the semantic-instance view and the structural-neighborhood view, respectively:

$$\mathcal{L}_i^{sem} = -\log\left(\frac{\sum_{j\in P_{os}}\exp\left(\frac{\mathrm{sim}\left(z_i^{sem}_p, z_j^{str}_p\right)}{\tau}\right)}{\sum_{k\in\{P_{os}\cup N_{eg}\}}\exp\left(\frac{\mathrm{sim}\left(z_i^{sem}_p, z_k^{str}_p\right)}{\tau}\right)}\right), \tag{15}$$

$$\mathcal{L}_i^{str} = -\log\left(\frac{\sum_{j\in P_{os}}\exp\left(\frac{\mathrm{sim}\left(z_i^{str}_p, z_j^{sem}_p\right)}{\tau}\right)}{\sum_{k\in\{P_{os}\cup N_{eg}\}}\exp\left(\frac{\mathrm{sim}\left(z_i^{str}_p, z_k^{sem}_p\right)}{\tau}\right)}\right), \tag{16}$$

where $\mathrm{sim}(\cdot,\cdot)$ denotes the cosine similarity function and τ is the temperature parameter.

The total loss function is:

$$\mathcal{L} = \frac{1}{|V|}\sum_{i\in V}\left[\lambda\mathcal{L}_i^{sem} + (1-\lambda)\mathcal{L}_i^{str}\right], \tag{17}$$

where λ is used to balance the loss of the two perspectives. The algorithmic process for SIHAN is listed in Algorithm 1.

E. Complexity Analysis

The complexity of SIHAN mainly consists of three parts. First, during the feature projection stage, nodes of different types are mapped into a unified feature space, which incurs a complexity of $\mathcal{O}(|A|\cdot|V_A|\cdot F_1\cdot F_2)$. Second, in the structural-neighborhood view, SIHAN performs graph convolution on the homogeneous graphs constructed from meta-paths, leading to a complexity of $\mathcal{O}(|E_l|\cdot {F_2}^2)$, where $|E_l|$ denotes the number of pairwise edges derived from meta-paths. Third, in the semantic-instance view, the model first needs to map the heterogeneous information network into heterogeneous hypergraphs. This construction process essentially transforms each semantic unit into a hyperedge and then adds the nodes associated with that semantic unit into the hyperedge. The complexity of this step is $\mathcal{O}(\sum_{e\in E_h}|e|)$, where E_h denotes the set of hyperedges and $|e|$ represents the number of nodes contained in hyperedge e. The term $\sum_{e\in E_h}|e|$ represents the total number of nodes in all hyperedges of the hypergraph, which can be approximated

Authorized licensed use limited to: TIANJIN UNIVERSITY. Downloaded on October 22,2026 at 01:15:35 UTC from IEEE Xplore. Restrictions apply.

Algorithm 1: SIHAN: Semantic-Instance Guided Hypergraph Attention Network with Dual-View Contrastive Learning for Heterogeneous Graph.

Input: The heterogeneous information network $\mathcal{H}_\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{A}, \mathcal{R}, \phi, \varphi)$, the features of all types of nodes F, the meta-path set $P = \{p^1, p^2, \ldots, p^n\}$, the sample threshold K

Output: Semantic-instance node embeddings Z^{sem}, structural-neighborhood node embeddings Z^{str}

```
Project node features of all types into a shared space
for p^m ∈ {p^1, p^2, ..., p^n} do
    Construct heterogeneous hypergraph from H_G
    Calculate Q^{p^m} via (3)
    Construct homogeneous graph from H_G
    Calculate embeddings H^{p^m} via (4) and L^{p^m} via (8)
    Calculate meta-path weights β^{p^m} via (6) and α^{p^m} via (10)
end for
Aggregate to get Z^{sem} via (7) and Z^{str} via (11)
Project Z^{sem} and Z^{str} into common contrastive space
Sample positive set P_{os} and negative set N_{eg}
Calculate L^{sem} via (15) //Semantic-instance loss
Calculate L^{str} via (16) // Structural-neighborhood loss
Calculate L via (17) // Total loss
Backpropagate L and update parameters
```

as $|E_h||V|$, so the complexity of hypergraph construction is $\mathcal{O}(|E_h||V|)$, where $|E_h|$ is the number of hyperedges and $|V|$ is the number of nodes. The model then introduces attention computation and hypergraph convolution over the hypergraph structure, resulting in a complexity of $\mathcal{O}(|E_h|(|V| + {F_2}^2))$. Therefore, the overall complexity of SIHAN can be summarized as $\mathcal{O}(|A| \cdot |V_A| \cdot F_1 \cdot F_2 + |E_l| \cdot {F_2}^2 + |E_h|(|V| + {F_2}^2))$.

V. Experiments

This section first introduces the datasets used in the experiment and relevant experimental settings, and then compares our model with popular baselines on various downstream tasks to demonstrate its superiority. Finally, we conduct ablation experiments and analyze the parameter sensitivity.

A. Experimental Setting

1) Datasets: We carry out experiments based on the following four real datasets. Details of the datasets are shown in Table III.

- **ACM [93]**: A heterogeneous academic network dataset comprising three types of nodes: paper (P), author (A), and subject (S), connected by two types of relationships. Papers (P) serve as the target nodes and are categorized into three classes.
- **DBLP [16]**: Another heterogeneous academic network dataset that includes four types of nodes: papers (P), authors (A), conferences (C), and terms (T). These nodes are interconnected through three types of relationships. In this dataset, authors (A) are designated as the target nodes and are categorized into four distinct classes.
- **Freebase [40]**: A heterogeneous movie network dataset with four types of nodes: movies (M), actors (A), directors (D), and writers (W), connected by three types of edges. Movies (M) are the target nodes and are classified into three categories: action, comedy, and opera.
- **Aminer [69]**: A subset of the Public Aminer dataset consisting of three types of nodes: paper (P), author (A), and reference (R), connected by two types of relationships. Papers (P) are the target nodes and fall into four classes.

TABLE III
The Characteristics of the Datasets

Dataset	Classes	Nodes	Edges	Meta-paths
ACM	3	paper (P):4019 author (A):7167 subject (S):60	P-A:13407 P-S:4019	PAP PSP
DBLP	4	author (A):4057 paper (P):14328 conference (C):20 term (T):7723	P-A:19645 P-C:14328 P-T:85810	APA APCPA APTPA
Freebase	3	movie (M):3492 actor (A):33401 director (D):2502 writer (W):4459	M-A:65341 M-D:3762 M-W:6414	MAM MDM MWM
Aminer	4	paper (P):6564 author (A):13329 reference (R):35890	P-A:18007 P-R:58831	PAP PRP

To be specific, for each dataset, we only use the initial attributes of target nodes, while for nodes of other types, we employ their one-hot encodings as features.

2) Baselines: Our proposed SIHAN is compared with ten baselines, which encompass a broad range of methods, including homogeneous graph, heterogeneous graph, and hypergraph models, as well as supervised and unsupervised learning approaches.

- **GCN [37]**:[1] It is a first-order local approximation of spectral graph convolution, which realizes multi-hop information propagation by stacking several convolutional layers.
- **GAT [64]**:[2] It introduces the attention mechanism to assign different weights to neighbors of the target node, and then obtains the node representation through weighted summation of neighbor nodes.
- **DGI [65]**:[3] It maximizes mutual information between graph enhancement representation and the currently extracted graph information to attain useful node embeddings.
- **DMGI [51]**:[4] It utilizes the Deep Multiplex Graph Infomax strategy to integrate node attributes and relationship types, introduces consistency regularization to reduce divergence in the embedding of relationship types, and employs the attention mechanism to distinguish the importance of different relationship types.

[1]https://github.com/tkipf/gcn
[2]https://github.com/PetarV-/GAT
[3]https://github.com/PetarV-/DGI
[4]https://github.com/pcy1302/DMGI

Authorized licensed use limited to: TIANJIN UNIVERSITY. Downloaded on October 22,2026 at 01:15:35 UTC from IEEE Xplore. Restrictions apply.

TABLE IV
PARAMETER SETTINGS OF SIHAN.

Dataset	dropout_feat	dropout_atth	learning rate	patience	τ	λ	K
ACM	0.3	0.5	0.0008	5	0.8	0.5	5
DBLP	0.4	0.35	0.0008	30	0.9	0.5	1000
Freebase	0.1	0.3	0.001	20	0.5	0.5	80
Aminer	0.5	0.5	0.003	40	0.5	0.5	15

- **HAN [68]**:[5] It generates homogeneous subgraphs for each meta-path, performs message passing based on node-level attention in different subgraphs, and aggregates information across different meta-paths using semantic-level attention.
- **HeCo [69]**:[6] It is a self-supervised learning method for heterogeneous graphs, which utilizes the cross-view contrastive mechanism and view mask mechanism to generate node embeddings.
- **MHGCN [85]**:[7] It designs a multi-layer convolutional aggregation module to learn useful interactions from heterogeneous meta-paths of different lengths in a multi-path heterogeneous network, effectively integrating network structure and semantic properties into the learned node embeddings.
- **SNHE [38]**:[8] It captures higher-order semantic relationships in heterogeneous information networks by transforming meta-paths into hypergraphs, and employs a dual self-supervised mechanism to maximize mutual information between different views, thereby learning high-quality node and hyperedge embeddings.
- **HGCML [72]**:[9] It generates multi-views that preserve semantics through meta-paths, designs multi-level contrastive objectives for nodes and graphs across views, and combines a positive sampling strategy considering both topology and semantics.
- **GTC [61]**:[10] It combines the local information aggregation ability of GNNs and the global modeling ability of Transformer through collaborative contrastive learning, effectively capturing multi-hop neighbor information.

3) Parameter Settings: To more comprehensively assess our model, the parameters of all baselines are initialized with the same values as those reported in the original papers to reduce the impact of other factors, and we also further carefully adjust parameters to attain better results. For the proposed SIHAN, we follow the same approach as [69], initializing parameters by Glorot initialization [19] and training the model using Adam [36]. In both the semantic-instance view and the structural-neighborhood view, the hypergraph convolution and graph convolution are set to one layer for each meta-path. For all models, the embedding dimension is fixed to 64. To reduce the influence of random initialization and stochastic optimization, we independently repeat the complete model training and downstream evaluation pipeline 10 times with different random seeds and report the average performance. For homogeneous graph models like GCN, GAT and DGI, we construct the homogeneous graph based on meta-paths and report the best performance. In addition, the positive sample threshold K in SIHAN is treated as a dataset-level hyperparameter, since different datasets exhibit different meta-path-based neighborhood densities and thus a universal value of K is not appropriate. We select K through a coarse-to-fine validation search. Specifically, we first perform a coarse search over a 1-2-5 logarithmic grid, i.e., $\mathcal{K}_{\text{coarse}} = \{a \times 10^b \mid a \in \{1, 2, 5\}, b \in \mathbb{Z}\} \cap [K_{\min}, K_{\max}]$. After obtaining the best coarse value K_c, we conduct a fine-grained local search around it using $\mathcal{K}_{\text{fine}} = \{K_c + t\Delta \mid t = -10, \ldots, 10\}$, where $\Delta = \max(1, \text{round}(0.05K_c))$. The final value of K is selected according to validation performance, and the test set is not used for hyperparameter selection. The parameter details of the proposed SIHAN are provided in Table IV.

4) Evaluation Metrics: To more comprehensively measure the performance of SIHAN, we conduct experiments on two downstream tasks: node classification and node clustering.

- For node classification, we train a logistic regression classifier on the learned node embeddings. Specifically, we randomly select 20, 40, and 60 nodes from each class in the dataset to form the training set, and randomly select 1000 nodes from the entire dataset as the test set. The validation set is used only for model selection during classifier training. The logistic regression classifier is implemented as a single linear layer whose input dimension is the embedding dimension and whose output dimension is the number of classes. Its weight matrix is initialized using Xavier uniform initialization, and the bias terms are initialized to zero. The classifier is optimized using the cross-entropy loss and the Adam optimizer. The learning rate of the classifier is set to 0.05 on ACM and 0.01 on DBLP, Freebase, and Aminer, with weight decay fixed to 0 for all datasets. For each training ratio in each outer run, the logistic regression classifier is trained for 200 epochs and independently repeated 50 times on the same fixed node embeddings to reduce the variance caused by classifier initialization. The averaged result over the 50 classifier trials is used as the classification performance of the corresponding outer run, and the final results are reported over 10 outer runs. The quality of the classification results is evaluated using Macro-F1, Micro-F1, and AUC.

[5]https://github.com/Jhy1993/HAN
[6]https://github.com/liun-online/HeCo
[7]https://github.com/NSSSJSS/MHGCN
[8]https://github.com/limengran98/SNHE
[9]https://github.com/Zehong-Wang/HGCML
[10]https://github.com/PHD-lanyu/GTC

Authorized licensed use limited to: TIANJIN UNIVERSITY. Downloaded on October 22,2026 at 01:15:35 UTC from IEEE Xplore. Restrictions apply.

TABLE V
NODE CLASSIFICATION RESULTS ON DATASETS (%). (BOLD: BEST).

Dataset	Metric	Split	GCN	GAT	DGI	DMGI	HAN	HeCo	MHGCN	SNHE	HGCML	GTC	SIHAN
ACM	Macro-F1	20	80.98 ± 3.4	80.30 ± 4.0	79.21 ± 4.0	86.31 ± 0.5	84.94 ± 1.7	87.09 ± 0.9	87.39 ± 1.0	88.32 ± 0.5	88.94 ± 0.6	89.46 ± 0.5	**90.54** ± 0.5
		40	81.24 ± 3.8	81.02 ± 4.0	79.76 ± 4.7	85.24 ± 0.5	85.01 ± 0.9	85.99 ± 1.2	86.91 ± 0.7	87.13 ± 0.6	88.06 ± 0.8	88.29 ± 0.5	**89.24** ± 0.9
		60	81.67 ± 2.0	81.56 ± 3.5	80.15 ± 3.9	86.57 ± 0.6	86.33 ± 0.9	88.19 ± 0.8	88.41 ± 0.7	88.07 ± 0.6	89.26 ± 0.5	89.71 ± 0.6	**90.06** ± 0.7
	Micro-F1	20	80.69 ± 2.3	80.25 ± 3.7	79.66 ± 3.3	86.09 ± 1.0	83.89 ± 1.5	86.87 ± 1.5	87.02 ± 0.8	87.57 ± 0.9	88.15 ± 0.6	89.33 ± 0.4	**90.19** ± 0.6
		40	81.13 ± 2.9	80.98 ± 3.0	79.81 ± 3.3	85.13 ± 0.8	85.23 ± 1.0	85.80 ± 0.8	86.93 ± 1.1	87.13 ± 0.8	87.63 ± 0.4	88.40 ± 0.6	**89.25** ± 0.4
		60	81.84 ± 3.1	81.37 ± 3.4	80.32 ± 3.3	86.26 ± 0.2	86.07 ± 0.5	87.60 ± 0.4	87.27 ± 0.5	87.51 ± 0.3	88.84 ± 0.4	89.18 ± 0.7	**89.79** ± 0.3
	AUC	20	90.14 ± 3.4	90.02 ± 3.2	90.37 ± 4.1	95.32 ± 0.4	92.81 ± 0.6	95.69 ± 0.3	95.89 ± 0.3	96.02 ± 0.4	96.37 ± 0.2	96.85 ± 0.4	**97.00** ± 0.5
		40	91.33 ± 2.5	90.97 ± 2.9	90.56 ± 3.9	95.11 ± 0.5	94.32 ± 0.4	95.98 ± 0.4	95.48 ± 0.3	95.91 ± 0.3	96.51 ± 0.5	96.75 ± 0.3	**97.52** ± 0.2
		60	91.27 ± 3.9	90.68 ± 3.1	90.84 ± 3.2	95.55 ± 0.7	94.25 ± 0.8	95.84 ± 0.5	96.99 ± 0.6	96.83 ± 0.7	96.80 ± 0.3	96.98 ± 0.3	**97.27** ± 0.2
DBLP	Macro-F1	20	88.03 ± 0.9	87.78 ± 0.6	86.53 ± 0.9	89.31 ± 0.7	90.67 ± 0.5	91.10 ± 0.9	90.13 ± 0.7	90.31 ± 0.8	91.73 ± 0.8	91.86 ± 0.9	**92.35** ± 0.4
		40	88.37 ± 0.9	87.93 ± 0.7	86.95 ± 1.2	88.75 ± 1.1	89.64 ± 0.6	89.98 ± 0.6	89.75 ± 1.0	89.75 ± 0.8	90.02 ± 0.5	90.64 ± 0.6	**91.37** ± 0.5
		60	89.15 ± 0.5	88.50 ± 0.4	87.10 ± 0.8	89.01 ± 0.9	90.29 ± 0.5	90.88 ± 0.7	90.10 ± 1.1	90.99 ± 0.9	90.85 ± 1.0	91.21 ± 0.2	**91.92** ± 0.8
	Micro-F1	20	87.99 ± 1.1	87.72 ± 1.0	87.11 ± 0.6	89.95 ± 0.3	89.64 ± 0.9	91.95 ± 0.7	91.34 ± 0.6	91.25 ± 0.5	92.07 ± 0.6	92.38 ± 0.3	**92.98** ± 0.4
		40	88.27 ± 0.3	87.83 ± 0.8	87.68 ± 0.5	89.18 ± 0.6	89.26 ± 0.7	90.26 ± 0.7	90.09 ± 0.6	90.18 ± 0.7	91.26 ± 0.3	91.20 ± 0.6	**91.77** ± 0.7
		60	88.32 ± 0.4	88.14 ± 0.5	88.07 ± 0.8	89.83 ± 0.4	90.05 ± 1.1	91.74 ± 0.5	90.38 ± 0.9	90.83 ± 0.4	91.55 ± 1.0	91.67 ± 0.6	**92.76** ± 0.3
	AUC	20	96.51 ± 0.3	96.02 ± 0.5	96.09 ± 0.4	96.69 ± 0.2	98.04 ± 0.3	98.26 ± 0.1	97.69 ± 0.4	97.27 ± 0.3	98.49 ± 0.3	98.60 ± 0.2	**98.73** ± 0.2
		40	96.77 ± 0.5	96.24 ± 0.7	96.21 ± 0.4	96.27 ± 0.6	97.15 ± 0.5	98.04 ± 0.1	97.19 ± 0.5	96.87 ± 0.6	98.00 ± 0.3	98.15 ± 0.1	**98.18** ± 0.3
		60	96.83 ± 0.2	96.20 ± 0.7	96.28 ± 0.7	96.33 ± 0.5	98.46 ± 0.2	98.55 ± 0.1	98.01 ± 0.4	97.33 ± 0.6	98.51 ± 0.3	98.44 ± 0.1	**98.76** ± 0.3
Freebase	Macro-F1	20	53.64 ± 1.7	52.38 ± 1.5	52.91 ± 2.6	53.69 ± 0.9	54.80 ± 1.6	55.47 ± 1.1	56.39 ± 0.8	55.89 ± 0.9	57.67 ± 0.9	58.49 ± 0.9	**58.57** ± 1.1
		40	53.09 ± 1.3	51.99 ± 2.1	52.83 ± 2.5	50.21 ± 1.9	58.32 ± 1.5	60.29 ± 0.9	60.71 ± 0.5	58.47 ± 1.3	60.97 ± 0.9	61.08 ± 1.2	**62.35** ± 1.0
		60	53.28 ± 0.9	52.22 ± 2.0	52.14 ± 2.3	53.15 ± 0.7	57.30 ± 0.8	56.50 ± 1.3	55.49 ± 0.9	53.15 ± 0.9	55.56 ± 1.2	**57.42** ± 0.6	56.09 ± 1.9
	Micro-F1	20	52.85 ± 2.0	53.01 ± 1.7	53.71 ± 1.0	54.27 ± 1.5	58.15 ± 1.0	58.02 ± 0.7	58.86 ± 1.0	57.27 ± 1.8	60.28 ± 0.4	59.88 ± 0.9	**60.58** ± 0.9
		40	53.61 ± 2.3	53.14 ± 1.9	54.51 ± 1.4	53.98 ± 2.1	62.67 ± 0.9	62.61 ± 0.5	61.83 ± 0.8	59.98 ± 1.9	62.11 ± 0.7	62.85 ± 1.2	**65.32** ± 0.7
		60	53.57 ± 1.6	53.19 ± 1.1	54.49 ± 1.8	54.01 ± 1.7	58.01 ± 0.7	59.74 ± 0.4	58.05 ± 0.5	56.91 ± 0.7	57.12 ± 0.4	**60.03** ± 0.5	58.30 ± 1.2
	AUC	20	73.19 ± 0.9	72.96 ± 1.2	72.06 ± 1.9	72.18 ± 1.4	73.31 ± 1.0	73.04 ± 1.1	75.27 ± 0.4	74.19 ± 0.5	75.40 ± 0.6	75.27 ± 0.6	**76.28** ± 0.6
		40	73.30 ± 0.8	73.04 ± 1.2	71.18 ± 1.5	70.52 ± 1.4	76.95 ± 0.5	77.67 ± 0.4	76.53 ± 0.5	74.81 ± 0.4	75.87 ± 0.7	75.60 ± 0.8	**77.95** ± 0.7
		60	72.93 ± 0.8	73.17 ± 1.1	72.01 ± 1.7	71.97 ± 1.2	**75.00** ± 0.5	74.68 ± 0.9	74.92 ± 0.9	73.97 ± 1.2	74.45 ± 1.0	74.85 ± 1.0	74.14 ± 1.3
Aminer	Macro-F1	20	51.82 ± 4.7	53.39 ± 4.1	51.65 ± 4.1	58.42 ± 2.9	59.77 ± 3.1	67.27 ± 1.2	68.98 ± 0.7	68.52 ± 0.9	68.76 ± 0.7	68.88 ± 0.4	**70.10** ± 0.7
		40	56.08 ± 4.2	55.21 ± 3.5	53.99 ± 4.5	60.19 ± 2.0	63.92 ± 3.1	68.50 ± 2.3	70.24 ± 0.3	69.99 ± 0.7	71.91 ± 0.2	71.76 ± 0.6	**73.38** ± 0.7
		60	55.65 ± 3.5	56.70 ± 3.2	55.02 ± 3.9	60.25 ± 4.1	63.16 ± 3.4	72.34 ± 1.4	73.31 ± 0.9	72.25 ± 1.1	72.50 ± 0.8	72.24 ± 0.4	**74.56** ± 0.5
	Micro-F1	20	64.49 ± 4.0	65.24 ± 3.8	62.39 ± 4.8	63.07 ± 2.7	71.36 ± 1.6	74.28 ± 2.0	78.13 ± 0.9	76.07 ± 0.7	**78.38** ± 1.1	76.49 ± 1.0	77.53 ± 0.4
		40	67.05 ± 3.3	66.68 ± 3.8	63.85 ± 4.6	62.88 ± 2.6	76.01 ± 1.0	75.70 ± 1.3	80.27 ± 0.4	79.88 ± 0.6	80.04 ± 1.2	79.85 ± 1.2	**80.63** ± 0.9
		60	66.90 ± 3.6	66.99 ± 4.5	63.74 ± 4.8	60.65 ± 3.0	75.33 ± 1.8	79.27 ± 1.5	80.12 ± 1.1	80.51 ± 0.3	80.83 ± 0.9	80.65 ± 0.7	**82.14** ± 1.2
	AUC	20	79.54 ± 4.7	80.78 ± 2.9	78.49 ± 4.0	84.31 ± 2.3	81.96 ± 2.4	87.48 ± 0.9	88.39 ± 0.5	87.91 ± 0.3	88.57 ± 0.1	88.69 ± 1.0	**90.82** ± 0.5
		40	79.93 ± 4.3	81.42 ± 3.3	79.03 ± 4.3	86.57 ± 1.9	83.65 ± 2.3	88.75 ± 0.9	90.56 ± 0.1	89.87 ± 0.4	90.61 ± 0.6	90.44 ± 0.3	**91.02** ± 0.4
		60	80.06 ± 4.3	81.65 ± 3.0	79.25 ± 4.3	86.11 ± 2.1	84.14 ± 2.4	90.38 ± 1.2	91.52 ± 0.6	90.61 ± 0.6	91.83 ± 0.6	91.08 ± 0.4	**92.38** ± 0.4

- For node clustering, we directly feed the learned node embeddings into the k-means algorithm without using label information during clustering. The number of clusters is set to the number of ground-truth classes in each dataset. In our implementation, k-means is performed with random state fixed to 20, while the other hyperparameters, including the initialization and the number of initializations, follow the default settings of the scikit-learn implementation. For each outer run, k-means is applied to the embeddings learned in that run, and the final clustering results are averaged over the 10 outer runs. The quality of the clustering results is evaluated using Normalized Mutual Information (NMI) and Adjusted Rand Index (ARI).

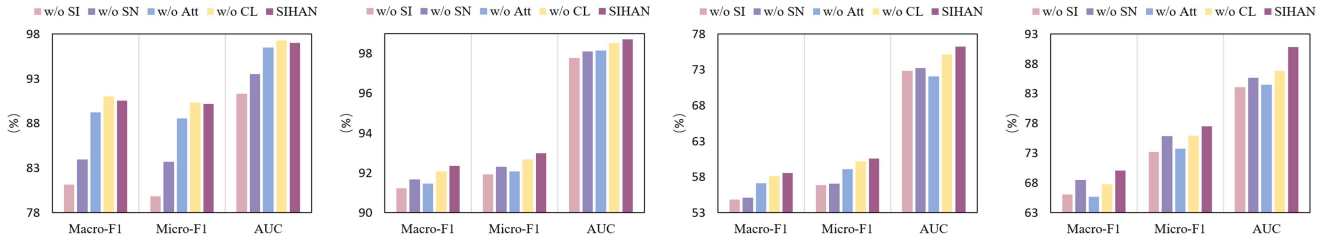
B. Node Classification

1) Result Comparison: The Macro-F1, Micro-F1, and AUC of the ten baselines and the proposed SIHAN on the four datasets are shown in Table V. From the experimental results, SIHAN outperforms all competing methods across the datasets, except for the case of the Freebase dataset with 60 training nodes per class and the case of the Aminer dataset with 20 training nodes per class in Micro-F1. Compared with homogeneous graph neural networks and conventional heterogeneous graph neural networks, SIHAN achieves substantial performance improvements. Furthermore, compared with leading self-supervised HGNN models, SIHAN also delivers the best overall performance. In addition, the standard deviations obtained from repeated experiments show that SIHAN exhibits relatively small fluctuations on the vast majority of evaluation metrics, with standard deviations consistently remaining at a low level. This further indicates that the proposed model is highly robust to random initialization and produces stable and reliable results.

2) Ablation Studies: To thoroughly validate the contribution of each key component in SIHAN, including the dual-view architecture, the semantic-sensitive attention mechanism and the contrastive learning module, we construct four ablated variants.

- **w/o SI**: A variant that removes the semantic-instance view encoder. It relies solely on the structural-neighborhood

Authorized licensed use limited to: TIANJIN UNIVERSITY. Downloaded on October 22,2026 at 01:15:35 UTC from IEEE Xplore. Restrictions apply.



(a) ACM (b) DBLP (c) Freebase (d) Aminer

Fig. 3. The comparison of SIHAN and its variants.

view to generate node embeddings, where both positive and negative samples are drawn from the structural-neighborhood view.

- **w/o SN**: A variant that removes the structural-neighborhood view encoder, encoding nodes only through the semantic-instance view. Both positive and negative samples are derived from the semantic-instance view.
- **w/o Att**: A variant that replaces the attention-based incidence matrix with a binary incidence matrix, thereby disabling the semantic-sensitive attention mechanism while retaining the dual-view structure.
- **w/o CL**: A variant that eliminates the contrastive learning module. Instead, a small amount of labeled data is used to directly train a classifier on the concatenated semantic-instance and structural-neighborhood embeddings by minimizing the cross-entropy loss.

We applied these variants to the node classification task and compared their performance with SIHAN, presenting the results of 20 labeled nodes per class. As illustrated in Fig. 3, SIHAN consistently outperforms both w/o SI and w/o SN, demonstrating the necessity of jointly modeling fine-grained higher-order semantics and local topological structures, as well as the effectiveness of the proposed dual-view design. Furthermore, w/o SN generally performs better than w/o SI, indicating that the higher-order semantic interactions captured by the semantic-instance view play a more critical role, while the local topological information provided by the structural-neighborhood view serves as an effective complement to the representation learning process. This finding highlights the particular importance of modeling rich higher-order interactions in the network.

Moreover, SIHAN significantly outperforms w/o Att, indicating that the semantic-sensitive hypergraph attention mechanism effectively characterizes the fine-grained association strength between nodes and hyperedges under different semantic contexts. The resulting soft association relationships are markedly superior to simple binary hard connections, thereby enabling the model to capture richer heterogeneous semantic patterns.

For the variant w/o CL, the incorporation of limited labeled data allows it to achieve competitive performance, even yielding a slight advantage over SIHAN on graphs with simple structures like ACM. However, on more complex graphs, SIHAN substantially outperforms w/o CL, demonstrating the superiority of the collaborative contrastive learning framework. The results highlight that contrastive alignment between the two views is essential for integrating their complementary information in realistic and challenging scenarios, ultimately enhancing both representation quality and generalization capability.

3) Sensitivity Analysis on λ and K: In this section, we analyze the impact of the balance parameter λ between the two views and different positive sample thresholds K on the model performance.

Balance Parameter λ: To analyze the impact of the balance parameter λ in the dual-view fusion mechanism, we evaluate the model on the four datasets with λ set to 0.1, 0.3, 0.5, 0.7, and 0.9, respectively, under the setting of 40 labeled nodes per class. The results are shown in Fig. 4. It can be observed that the performance in terms of Macro-F1, Micro-F1, and AUC generally follows a trend of first increasing and then decreasing with the growth of λ. The model attains the best performance in most cases when λ=0.5, indicating that the semantic-instance view and the structural-neighborhood view make similarly important contributions to the learned representations. In contrast, when the weight of either view becomes overly dominant, the model is biased toward a single information source, which weakens the complementary benefits of the other view and ultimately limits its ability to model complex interactions effectively.

Sampling Threshold K: To analyze the effect of the positive sample size in contrastive learning on model performance, we conduct a systematic sensitivity study of the positive sample threshold K on different datasets and evaluate the model under the setting of 60 labeled nodes per class. The results are shown in Fig. 5. From the overall trend, the model performance does not vary monotonically with increasing K, but instead generally follows a pattern of first increasing and then decreasing. Meanwhile, the optimal value of K differs across datasets: on the ACM, DBLP, Freebase, and Aminer datasets, the model achieves the best performance when K=5, K=1000, K=80, and K=15, respectively. This observation suggests that the appropriate choice of the positive sample size depends largely on the structural characteristics and semantic distributions of the dataset itself.

C. Node Clustering

1) Result Comparison and Ablation Studies: For the clustering task, we compare only with self-supervised baselines and variants that do not rely on manually annotated information, so as to eliminate the interference of label information on the experimental results. The clustering performance of six baseline models, three SIHAN variants, and the full SIHAN model on

Authorized licensed use limited to: TIANJIN UNIVERSITY. Downloaded on October 22,2026 at 01:15:35 UTC from IEEE Xplore. Restrictions apply.

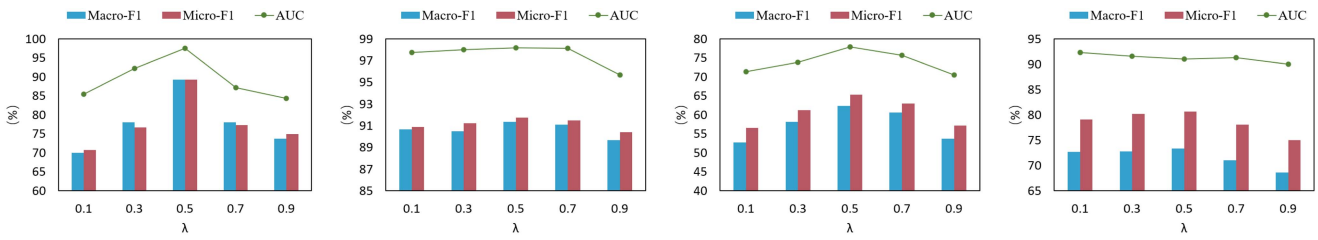


Fig. 4. The effect of balance parameter λ on the classification performance of SIHAN.

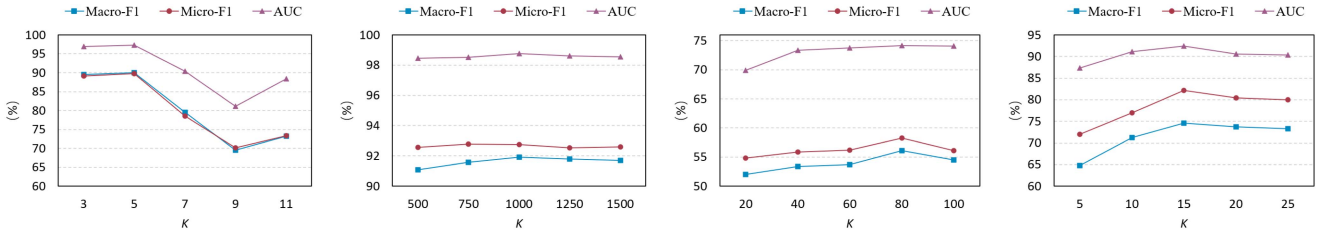


Fig. 5. The effect of positive sample threshold K on the classification performance of SIHAN.

TABLE VI
NODE CLUSTERING PERFORMANCE (%). (BOLD: BEST).

Dataset	Metric	DGI	DMGI	HeCo	SNHE	HGCML	GTC	w/o Att	w/o SI	w/o SN	SIHAN
ACM	NMI	52.94	53.27	59.20	61.07	60.91	62.15	62.43	28.43	46.26	**62.47**
	ARI	43.15	50.02	58.26	60.95	62.10	61.93	62.09	22.72	39.08	**62.98**
DBLP	NMI	59.10	70.71	69.85	69.55	71.02	71.37	70.50	71.30	70.87	**71.82**
	ARI	60.96	74.44	75.54	74.31	**75.62**	74.34	74.56	74.54	74.37	74.71
Freebase	NMI	18.31	15.81	17.67	16.84	17.95	18.09	18.24	9.57	17.51	**18.43**
	ARI	10.95	16.22	20.26	18.09	19.36	20.33	20.01	7.61	19.88	**20.59**
Aminer	NMI	24.50	21.35	32.46	30.56	27.29	30.72	27.73	31.54	30.50	**33.15**
	ARI	16.88	23.32	29.25	31.49	25.50	27.33	23.99	29.98	29.42	**34.08**

the four datasets is reported in Table VI. As shown, SIHAN achieves the best performance on the vast majority of datasets and evaluation metrics, outperforming a variety of advanced self-supervised node representation learning methods, including HeCo, SNHE, and GTC, as well as all ablation variants. Specifically, except for the ARI metric on the DBLP dataset, SIHAN obtains the best results on all remaining datasets and metrics. This demonstrates that the proposed model can not only effectively capture structural characteristics and semantic relationships in heterogeneous networks, but also learn well-separated node embeddings through the dual-view architecture and contrastive learning mechanism in a completely label-free setting, thereby exhibiting strong robustness and generalizability in clustering tasks.

In addition, the comparison with the ablation variants shows that the complete SIHAN achieves the best performance under all settings. In particular, on the ACM and Aminer datasets, removing either view leads to a significant performance degradation, which further confirms the importance of jointly modeling higher-order semantic information and local structural information for learning discriminative node representations. Meanwhile, when the semantic-sensitive attention mechanism is replaced with a binary incidence matrix, the model performance declines to varying degrees on all datasets. This result further indicates that fine-grained modeling of node-hyperedge associations is also crucial for improving clustering quality.

2) Sensitivity Test on λ and K: Furthermore, in the node clustering task, we systematically analyze the sensitivity of the model to the balance parameter λ and the positive sample threshold K. The results are shown in Figs. 6 and 7. It can be observed that the response trends of the model to these two parameters in the clustering task are generally consistent with those observed in the node classification task. As λ increases, the clustering performance also first improves and then declines, reaching the optimum when the two views are assigned equal weights. Similarly, the model performance with respect to K exhibits a trend of first increasing and then decreasing, and the optimal values on the four datasets are K=5, 1000, 80, and 15, respectively. These results further demonstrate that the proposed dual-view contrastive learning framework exhibits consistent and stable behavioral characteristics across different task settings.

Authorized licensed use limited to: TIANJIN UNIVERSITY. Downloaded on October 22,2026 at 01:15:35 UTC from IEEE Xplore. Restrictions apply.

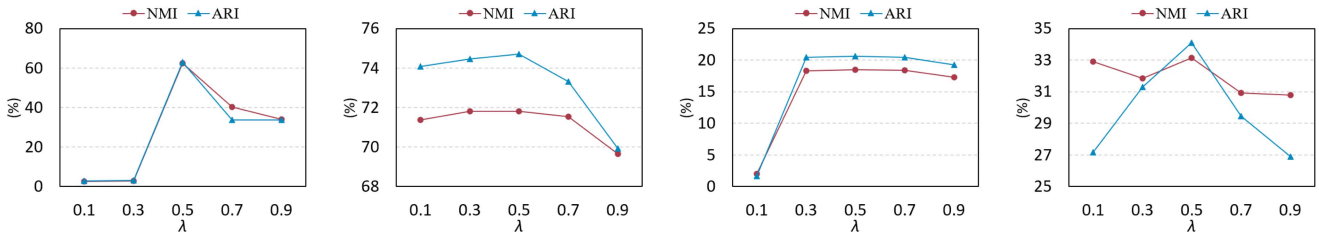


Fig. 6. The effect of balance parameter λ on the clustering performance of SIHAN.

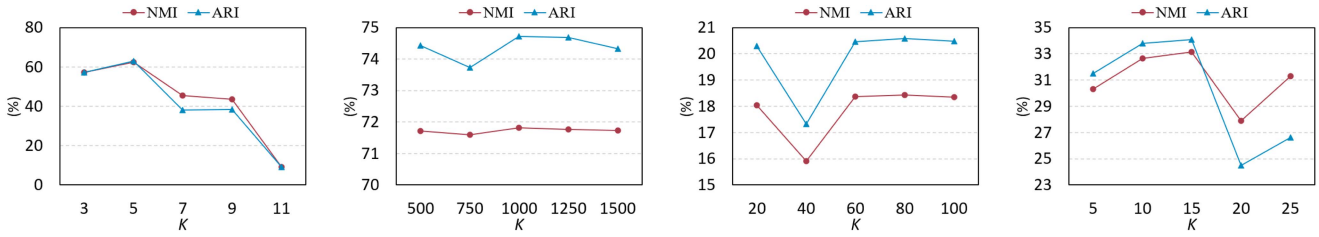


Fig. 7. The effect of positive sample threshold K on the clustering performance of SIHAN.

TABLE VII
EFFICIENCY COMPARISON OF MODELS

Model	ACM		DBLP		Freebase		Aminer	
	GPU(MB)	Training time(s)	GPU(MB)	Training time(s)	GPU(MB)	Training time(s)	GPU(MB)	Training time(s)
SNHE	1932.68	0.46	2735.81	1.30	5524.00	1.51	8406.21	1.59
MHGCN	2386.06	1.33	2916.31	3.81	5716.29	2.95	8993.33	3.23
HGCML	2860.16	0.66	4777.56	2.00	5253.98	1.96	9669.50	1.48
GTC	1742.70	0.20	2551.28	0.93	5998.82	1.27	8639.07	1.30
SIHAN	1032.98	0.35	2317.59	1.01	5223.07	1.32	8155.19	1.34

D. Computational Efficiency Analysis

To evaluate the practical computational efficiency of SIHAN, we further conduct a comparative analysis of GPU memory consumption and the average training time per epoch on the four datasets. The comparison includes several baseline models with strong overall performance, namely SNHE, MHGCN, HGCML and GTC. The results are presented in Table VII. It can be seen that SIHAN achieves the lowest memory consumption on all four datasets, demonstrating superior space efficiency. Specifically, on the ACM, DBLP, Freebase, and Aminer datasets, the GPU memory usage of SIHAN is 1032.98 MB, 2317.59 MB, 5223.07 MB, and 8155.19 MB, respectively, all of which are lower than those of the competing methods. This indicates that although SIHAN introduces a higher-order relation modeling mechanism based on semantic instances, it does not incur additional significant storage overhead; instead, it improves resource utilization efficiency through a more compact structural design. In terms of training time, the average time per epoch of SIHAN is 0.35 s, 1.01 s, 1.32 s, and 1.34 s, respectively, which also remains at a competitive level overall. Although GTC achieves slightly lower training time per epoch than SIHAN on some datasets, its memory consumption is substantially higher. In contrast, HGCML, MHGCN, and SNHE generally exhibit both longer training time and greater GPU memory usage. These results indicate that SIHAN achieves a favorable balance between training efficiency and memory consumption. Overall, SIHAN not only significantly outperforms most high-performing baselines in terms of resource consumption, but also maintains strong competitiveness in training speed, demonstrating good computational efficiency and scalability.

E. Instance-Level Semantic Discrimination Analysis

To verify whether SIHAN can distinguish the specific semantic differences among different path instances under the same meta-path, we further design an instance-level semantic discrimination experiment. Specifically, let $v_i \in \mathcal{V}$ denote a target node, and $\mathcal{E}(v_i)$ denote the set of candidate hyperedges constructed under a given meta-path p^m. For each hyperedge $e \in \mathcal{E}(v_i)$, we define the category distribution of nodes within the hyperedge as:

$$n_c(e) = \sum_{v \in e} \mathbf{1}\left(y_v = c\right), \quad c \in \mathcal{C}, \tag{18}$$

Authorized licensed use limited to: TIANJIN UNIVERSITY. Downloaded on October 22,2026 at 01:15:35 UTC from IEEE Xplore. Restrictions apply.

TABLE VIII
SEMANTIC DISCRIMINATION CAPABILITY OF SIHAN

Dataset	Meta-path	SIHAN	w/o Att
ACM	PAP	0.895	0.764
	PSP	0.883	0.699
DBLP	APA	0.897	0.685
	APCPA	0.901	0.690
	APTPA	0.924	0.658
Freebase	MAM	0.830	0.641
	MDM	0.835	0.667
	MWM	0.812	0.615
Aminer	PAP	0.890	0.792
	PRP	0.879	0.784

where y_v is the ground-truth category of node v, $\mathcal{C}$ is the set of all categories, and $\mathbf{1}(\cdot)$ is the indicator function.

The dominant semantic category of hyperedge e is defined as:

$$c^*(e) = \arg\max_{c \in \mathcal{C}} n_c(e). \tag{19}$$

Based on this, we define the semantic relevance label between node v_i and hyperedge e as:

$$r(v_i, e) = \begin{cases} 1, & \text{if } c^*(e) = y_{v_i}, \\ 0, & \text{otherwise.} \end{cases} \tag{20}$$

If $r(v_i, e) = 1$, the node-hyperedge pair is regarded as a semantically relevant pair; otherwise, it is regarded as a semantically irrelevant pair. Subsequently, all node-hyperedge pairs are ranked according to the attention-based node-hyperedge association scores Q^{p^m} produced by SIHAN, and the AUC is used to measure the ability of the model to distinguish the specific semantics of different path instances. Since the AUC measures the probability that the model ranks semantically relevant node-hyperedge pairs ahead of semantically irrelevant ones, this metric can directly reflect the model's capability to distinguish the specific semantics of different path instances.

As shown in Table VIII, SIHAN achieves high AUC values under all meta-path settings on the four datasets and significantly outperforms the variant without the semantic-sensitive hypergraph attention mechanism, namely w/o Att, indicating that the proposed method can effectively distinguish the fine-grained semantics of different path instances. Specifically, the average AUC of SIHAN reaches 0.8746, whereas that of w/o Att is only 0.6995, corresponding to an average improvement of 0.1751. Among all settings, the largest improvement is observed on APTPA in DBLP, suggesting that the advantage of SIHAN in semantic-sensitive modeling becomes more pronounced when the meta-path is longer and the internal semantic structure is more complex. Overall, this experiment quantitatively verifies that SIHAN is capable of distinguishing the specific semantics of different path instances under the same meta-path, rather than merely modeling semantics at the coarse-grained meta-path pattern level.

F. Analysis of Positive Sampling Strategies

The positive sample selection strategy plays an important role in contrastive learning, since it determines which nodes are encouraged to have similar representations. In SIHAN, we select positive samples according to meta-path connectivity, where nodes connected with the target node under more meta-paths are regarded as more semantically related. Although this strategy is simple and intuitive, it is still necessary to examine whether the performance of SIHAN is overly dependent on this specific sampling rule. Therefore, we conduct additional experiments to compare different positive sampling strategies and evaluate the robustness of the proposed contrastive learning framework. In this experiment, we compare SIHAN with three alternative positive sampling strategies, namely RandPS, FeatPS, and StruPS.

- *RandPS:* It denotes random positive sampling, where positive samples are randomly selected from the meta-path-induced neighbors of each target node.
- *FeatPS:* It denotes feature-similarity positive sampling, where positive samples are ranked according to cosine similarity in the feature representation space, and the most similar nodes are selected as positives.
- *StruPS:* It denotes structure-similarity positive sampling, where positive samples are selected according to the Jaccard similarity between meta-path-based neighborhood sets.

For all methods, we fix the positive sample size K at the dataset-level optimal value determined via validation, ensuring a fair comparison under the same neighborhood scale. Table IX reports the performance of different positive sampling strategies on node classification, measured by Macro-F1, Micro-F1, and AUC, and node clustering, measured by NMI and ARI. RandPS generally produces lower performance than the other three strategies on all four datasets. This indicates that randomly selecting positive samples from meta-path-connected neighbors may introduce noisy positive pairs into the contrastive objective, since these randomly selected nodes do not necessarily share strong semantic or structural consistency with the target node. In contrast, FeatPS, StruPS, and SIHAN achieve relatively close results across both node classification and node clustering. This

TABLE IX
PERFORMANCE COMPARISON OF DIFFERENT POSITIVE SAMPLING STRATEGIES (%)

Dataset	Metric	RandPS	FeatPS	StruPS	SIHAN
ACM	Macro-F1	87.75	90.12	90.03	90.06
	Micro-F1	87.33	89.67	89.53	89.79
	AUC	96.49	97.27	97.21	97.27
	NMI	59.98	62.51	61.99	62.47
	ARI	59.39	62.48	62.37	62.97
DBLP	Macro-F1	90.29	91.40	91.66	91.92
	Micro-F1	90.83	92.45	92.79	92.76
	AUC	97.04	98.37	98.82	98.76
	NMI	70.09	71.41	71.65	71.82
	ARI	70.88	73.96	74.24	74.71
Freebase	Macro-F1	55.57	55.95	56.13	56.09
	Micro-F1	58.04	58.18	58.26	58.30
	AUC	73.86	74.10	74.05	74.14
	NMI	18.11	18.19	18.16	18.43
	ARI	19.70	19.99	20.41	20.59
Aminer	Macro-F1	73.00	74.27	74.35	74.56
	Micro-F1	80.60	82.12	81.89	82.14
	AUC	91.63	92.30	92.07	92.38
	NMI	30.43	33.02	32.80	33.15
	ARI	30.89	34.00	33.85	34.08

Authorized licensed use limited to: TIANJIN UNIVERSITY. Downloaded on October 22,2026 at 01:15:35 UTC from IEEE Xplore. Restrictions apply.

suggests that feature similarity, structural similarity, and meta-path connectivity can all provide useful signals for identifying meaningful positive samples.

More importantly, the comparable performance among these three strategies shows that SIHAN is not overly sensitive to a specific positive sampling heuristic. Among them, the proposed meta-path connectivity-based strategy achieves the best or near-best overall performance on most datasets and metrics. This is reasonable because meta-path connectivity directly reflects heterogeneous semantic associations: if two nodes are connected under more meta-paths, they are more likely to share multiple semantic contexts. Compared with FeatPS, which depends on similarity in the feature representation space, and StruPS, which relies on neighborhood-set overlap, the proposed strategy provides a simple and direct structural prior for positive sample selection. Therefore, these results demonstrate both the robustness of SIHAN to different positive sampling strategies and the rationality of using meta-path connectivity for constructing positive pairs in heterogeneous graph contrastive learning.

VI. Conclusion

In this paper, we propose a novel semantic-instance based HGNN framework that effectively addresses the limitations of existing models in capturing fine-grained, higher-order semantic interactions. The framework adopts a dual-view representation learning architecture. In the semantic-instance view, the heterogeneous information network is transformed into semantic-driven heterogeneous hypergraphs, coupled with a semantic-sensitive hypergraph attention mechanism for accurately modeling higher-order semantic relationships. In the structural-neighborhood view, the network is transformed into multiple homogeneous graphs to complement the local topological information among nodes. By employing contrastive learning between the two views, the approach generates higher-quality embedding vectors, thereby enhancing the model's efficiency in downstream tasks. Through extensive experiments on diverse real-world datasets, it has been demonstrated that the proposed model outperforms representative HGNNs.

References

[1] S. An, B. Hong, Z. Guo, S. Zhu, K. Lin, and F. Yang, "Simultaneously local and global contrastive learning of graph representations," *Eng. Appl. Artif. Intell.*, vol. 145, 2025, Art. no. 110203.

[2] S. Bai, F. Zhang, and P. H. S. Torr, "Hypergraph convolution and hypergraph attention," *Pattern Recognit.*, vol. 110, 2021, Art. no. 107637.

[3] R. Bing, G. Yuan, Y. Zhang, S. Wang, B. Li, and Y. Zhou, "An efficient multi-view heterogeneous hypergraph convolutional network for heterogeneous information network representation learning," *IEEE Trans. Big Data*, vol. 11, no. 3, pp. 1144–1157, Jun. 2025.

[4] Y. Cao, T. Radhika, A. Chandrasekar, and A. R. Subhashri, "Nonfragile state estimation for memristor based competitive neural networks subject to sensor saturation and stochastic cyber attacks," *Sci. China Technol. Sci.*, vol. 69, 2026, Art. no. 1420401.

[5] Y. Chang, C. Chen, W. Hu, Z. Zheng, X. Zhou, and S. Chen, "Megnn: Meta-path extracted graph neural network for heterogeneous graph representation learning," *Knowl. Based Syst.*, vol. 235, 2022, Art. no. 107611.

[6] S. Chen, Y. Wan, L. Rutkowski, L. Hua, and J. Cao, "Nash equilibrium seeking for hybrid heterogeneous euler–lagrange systems in multicluster games under dos attacks: A predefined-time approach," *IEEE Trans. Comput. Soc. Syst.*, early access, Jan. 21, 2026, doi: 10.1109/TCSS.2025.3649613.

[7] C. Zhao, "Graph adaptive attention network with cross-entropy," *Entropy*, vol. 26, no. 7, 2024, Art. no. 576.

[8] W. Dai et al., "Prediction of miRNA-disease association based on heterogeneous hypergraph convolution and heterogeneous graph multi-scale convolution," *Health Inf. Sci. Syst.*, vol. 13, no. 1, 2025, Art. no. 4.

[9] L. Deng, D. Lian, Z. Huang, and E. Chen, "Graph convolutional adversarial networks for spatiotemporal anomaly detection," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 6, pp. 2416–2428, Jun. 2022.

[10] Y. Deng, W. Bai, Y. Jiang, and Y. Tang, "Subgraph-based feature fusion models for semantic similarity computation in heterogeneous knowledge graphs," *Knowl. Based Syst.*, vol. 257, 2022, Art. no. 109906.

[11] K. Ding, J. Wang, J. Li, D. Li, and H. Liu, "Be more with less: Hypergraph attention networks for inductive text classification," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2020, pp. 4927–4936.

[12] Y. Ding, Q. Yao, H. Zhao, and T. Zhang, "DiffMG: Differentiable meta graph search for heterogeneous graph neural networks," in *Proc. 27th ACM SIGKDD Conf. Knowl. Discov. Data Mining*, 2021, pp. 279–288.

[13] Y. Dong, N. V. Chawla, and A. Swami, "metapath2vec: Scalable representation learning for heterogeneous networks," in *Proc. 23rd ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2017, pp. 135–144.

[14] F. Ferrini, A. Longa, A. Passerini, and M. Jaeger, "Meta-path learning for multi-relational graph neural networks," in *Proc. 2nd Learn. Graphs Conf.*, Nov. 2024, vol. 231, pp. 2:1–2:17.

[15] T.-Y. Fu, W.-C. Lee, and Z. Lei, "HIN2Vec: Explore meta-paths in heterogeneous information networks for representation learning," in *Proc. ACM Conf. Inf. Knowl. Manage.*, 2017, pp. 1797–1806.

[16] X. Fu, J. Zhang, Z. Meng, and I. King, "MAGNN: Metapath aggregated graph neural network for heterogeneous graph embedding," in *Proc. Web Conf.*, 2020, pp. 2331–2341.

[17] Y. Gao, Y. Feng, J. Shuyi, and R. Ji, "HGNN$^+$: General hypergraph neural networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 3, pp. 3181–3199, Mar. 2023.

[18] Y. Gao, S. Ji, X. Han, and Q. Dai, "Hypergraph computation," *Engineering*, vol. 40, pp. 188–201, 2024.

[19] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proc. 13th Int. Conf. Artif. Intell. Statist.*, 2010, vol. 9, pp. 249–256.

[20] X. Guo et al., "MMHCL: Multi-modal hypergraph contrastive learning for recommendation," *ACM Trans. Multimedia Comput. Commun. Appl.*, vol. 21, no. 10, pp. 298:1–298: 23, 2025.

[21] Y. Guo, H. Zhu, J. Cao, R. Chen, and D. Zhou, "Learning heterogeneous biological interactions via meta-relation-guided dual-channel graph transformer for circular ribonucleic acid function prediction," *Knowl.-Based Syst.*, vol. 342, 2026, Art. no. 115929.

[22] W. L. Hamilton, R. Ying, and J. Leskovec, "Inductive representation learning on large graphs," in *Proc. 31st Int. Conf. Neural Inf. Process. Syst.*, 2017, pp. 1025–1035.

[23] L. Han and H. Zhao, "Heterogeneous network representation learning via multiple attention mechanisms," *Knowl. Based Syst.*, vol. 330, 2025, Art. no. 114747.

[24] J. Hang et al., "Paths2pair: Meta-path based link prediction in billion-scale commercial heterogeneous graphs," in *Proc. 30th ACM SIGKDD Conf. Knowl. Discov. Data Mining*, Barcelona, Spain, 2024, pp. 5082–5092.

[25] K. Hassani and A. H. Khasahmadi, "Contrastive multi-view representation learning on graphs," in *Proc. 37th Int. Conf. Mach. Learn.*, Jul. 2020, pp. 4116–4126.

[26] C. He, Y. Lu, R. Zhang, Y. Li, B. Ai, and D. Niyato, "GNN-enabled coordinated beamforming design for high speed railway communication systems," *IEEE Trans. Mobile Comput.*, vol. 25, no. 6, pp. 8384–8397, Jun. 2026.

[27] Y. He, Y. Song, J. Li, C. Ji, J. Peng, and H. Peng, "HeteSpaceyWalk: A heterogeneous Spacey random walk for heterogeneous information network embedding," in *Proc. 28th ACM Int. Conf. Inf. Knowl. Manage.*, 2019, pp. 639–648.

[28] B. Hu, Y. Fang, and C. Shi, "Adversarial learning on heterogeneous information networks," in *Proc. 25th ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2019, pp. 120–129.

[29] Z. Huang and N. Mamoulis, "Heterogeneous information network embedding for meta path based proximity," 2017, *arXiv:1701.05291*.

[30] Z. Huang, Y. Ren, X. Pu, S. Huang, Z. Xu, and L. He, "Self-supervised graph attention networks for deep weighted multi-view clustering," in *Proc. AAAI Conf. Artif. Intell.*, 2023, vol. 37, pp. 7936–7943.

Authorized licensed use limited to: TIANJIN UNIVERSITY. Downloaded on October 22,2026 at 01:15:35 UTC from IEEE Xplore. Restrictions apply.

[31] D. Hwang, J. Park, S. Kwon, K. M. Kim, J.-W. Ha, and H. J. Kim, "Self-supervised auxiliary learning with meta-paths for heterogeneous graphs," in *Proc. Adv. Neural Inf. Process. Syst.*, H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, and H. Lin, Eds., 2020, vol. 33, pp. 10294–10305.
[32] R. Jiang, Y. Li, H. Liu, and Y. Yu, "Incorporating attributes and multi-scale structures for heterogeneous graph contrastive learning," *Inf. Fusion*, vol. 122, 2025, Art. no. 103220.
[33] H. Jin, K. Zhou, J. Yin, L. You, and Z. Zhou, "Multi-Granular attention based heterogeneous hypergraph neural network," 2025, *arXiv:2505.04340*.
[34] B. Khan, J. Wu, J. Yang, and X. Ma, "Heterogeneous hypergraph neural network for social recommendation using attention network," *Trans. Recommender Syst.*, vol. 3, no. 3, pp. 30:1–30:22, 2025.
[35] D. Kim and A. Oh, "How to find your friendly neighborhood: Graph attention design with self-supervision," in *Proc. 9th Int. Conf. Learn. Representations*, Austria, 2021, pp. 1–25.
[36] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. 3rd Int. Conf. Learn. Representations*, Y. Bengio and Y. LeCun, Eds., San Diego, CA, USA, 2015, pp. 1–15.
[37] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," in *Proc. 5th Int. Conf. Learn. Representations*, Toulon, France, 2017, pp. 2713–2726.
[38] M. Li, Y. Zhang, W. Zhang, Y. Chu, Y. Hu, and B. Yin, "Self-supervised nodes-hyperedges embedding for heterogeneous information network learning," *IEEE Trans. Big Data*, vol. 9, no. 4, pp. 1210–1224, Aug. 2023.
[39] Q. Li et al., "MetapathVis: Inspecting the effect of metapath in heterogeneous network embedding via visual analytics," *Comput. Graph. Forum*, vol. 44, 2025, Art. no. e15285.
[40] X. Li, D. Ding, B. Kao, Y. Sun, and N. Mamoulis, "Leveraging meta-path contexts for classification in heterogeneous information networks," in *Proc. IEEE 37th Int. Conf. Data Eng.* 2021, pp. 912–923.
[41] B. Lin, L. Zhou, Z. Gao, and H. Fang, "Evolutionary dynamics of group cooperation on heterogeneous higher-order networks," *IEEE Trans. Netw. Sci. Eng.*, vol. 12, no. 6, pp. 4894–4905, 2025.
[42] K. Lin et al., "Adaptive neighbor graph aggregated graph attention network for heterogeneous graph embedding," *ACM Trans. Knowl. Discov. Data*, vol. 18, no. 1, pp. 29:1–29:21, 2024.
[43] M. Lin, T. Wang, Y. Zhu, X. Li, X. Zhou, and W. Wang, "A heterogeneous directed graph attention network for inductive text classification using multilevel semantic embeddings," *Knowl. Based Syst.*, vol. 295, 2024, Art. no. 111797.
[44] Z. Lin et al., "Unified heterogeneous hypergraph construction for incomplete multimedia recommendation," *ACM Trans. Inf. Syst.*, vol. 43, no. 5, pp. 1–31, 2025.
[45] J. Liu, L. Song, G. Wang, and X. Shang, "Meta-HGT: Metapath-aware hypergraph transformer for heterogeneous information network embedding," *Neural Netw.*, vol. 157, pp. 65–76, 2023.
[46] Q. Liu, Y. Xia, Y. Guo, and W. Li, "DiffAD: Diffusion augmentation for influenza virus antigenic distance prediction," *Expert Syst. Appl.*, vol. 298, 2026, Art. no. 129925.
[47] D. Loveland, J. Zhu, M. Heimann, B. Fish, M. T. Schaub, and D. Koutra, "On performance discrepancies across local homophily levels in graph neural networks," in *Proc. Learn. Graphs Conf.*, 2024, pp. 6:1–6:30.
[48] S. Luan et al., "When do graph neural networks help with node classification? Investigating the homophily principle on node distinguishability," in *Proc. Adv. Neural Inf. Process. Syst.*, 2023, vol. 36, pp. 28748–28760.
[49] H. Mao et al., "Demystifying structural disparity in graph neural networks: Can one size fit all?," in *Proc. Adv. Neural Inf. Process. Syst.*, 2023, vol. 36, pp. 37013–37067.
[50] M. Moradi, P. Moradi, A. Faroughi, and M. Jalili, "Two-level attention mechanism with contrastive learning for heterogeneous graph representation learning," *Expert Syst. Appl.*, vol. 273, 2025, Art. no. 126751.
[51] C. Park, D. Kim, J. Han, and H. Yu, "Unsupervised attributed multiplex network embedding," in *Proc. AAAI Conf. Artif. Intell.*, 2020, vol. 34, pp. 5371–5378.
[52] H. Peng, R. Zhang, Y. Dou, R. Yang, J. Zhang, and P. S. Yu, "Reinforced neighborhood selection guided multi-relational graph neural networks," *ACM Trans. Inf. Syst.*, vol. 40, no. 4, pp. 69:1–69: 46, 2022.
[53] Y. Peng, J. Xia, D. Liu, M. Liu, L. Xiao, and B. Shi, "Unifying topological structure and self-attention mechanism for node classification in directed networks," *Sci. Rep.*, vol. 15, no. 1, 2025, Art. no. 805.
[54] B. Perozzi, R. Al-Rfou, and S. Skiena, "Deepwalk: Online learning of social representations," in *Proc. 20th ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2014, pp. 701–710.
[55] J. Qiu et al., "GCC: Graph contrastive coding for graph neural network pre-training," in *Proc. 26th ACM SIGKDD Conf. Knowl. Discov. Data Mining*, 2020, pp. 1150–1160.
[56] M. Schlichtkrull et al., "Modeling relational data with graph convolutional networks," in *Proc. 15th Int. Conf. Semantic Web*, Crete, Greece, 2018, pp. 593–607.
[57] Q. Shao, D. Chen, and W. Yu, "A unique framework of heterogeneous augmentation graph contrastive learning for both node and graph classification," *IEEE Trans. Netw. Sci. Eng.*, vol. 11, no. 6, pp. 5818–5828, Nov./Dec. 2024.
[58] D. Sun, C. Sun, Z. Ding, C. Wang, and B. Luo, "Multi-view interactive heterogeneous graph contrastive learning," *IEEE Trans. Netw. Sci. Eng.*, vol. 13, pp. 1389–1405, 2026.
[59] W. Sun, Y. Cheng, and X. Zhao, "HGNN-NHSA: Heterogeneous graph neural network based on neighbor and high-order subgraph aggregation learning," *Neurocomputing*, vol. 626, 2025, Art. no. 129618.
[60] Y. Sun and J. Han, "Mining heterogeneous information networks: A structural analysis approach," *SIGKDD Explorations Newslett.*, vol. 14, no. 2, pp. 20–28, 2013.
[61] Y. Sun, D. Zhu, Y. Wang, Y. Fu, and Z. Tian, "GTC: GNN-transformer co-contrastive learning for self-supervised heterogeneous graph representation," *Neural Netw.*, vol. 181, 2025, Art. no. 106645.
[62] S. Tehsin, I. M. Nasir, and R. Damasevicius, "GATransformer: A graph attention network-based transformer model to generate explainable attentions for brain tumor detection," *Algorithms*, vol. 18, no. 2, 2025, Art. no. 89.
[63] S. Thakoor, C. Tallec, M. G. Azar, R. Munos, P. Veličković, and M. Valko, "Bootstrapped representation learning on graphs," in *Proc. Workshop Geometrical Topological Representation Learn.*, 2021, pp. 1–14.
[64] P. Velickovic, G. Cucurull, A. Casanova, A. Romero, P. Liò, and Y. Bengio, "Graph attention networks," in *Proc. 6th Int. Conf. Learn. Representations*, Vancouver, BC, Canada, 2018, pp. 2920–2931.
[65] P. Velickovic et al., "Deep graph infomax," in *Proc. 7th Int. Conf. Learn. Representations*, New Orleans, LA, USA, 2019 pp. 1–17.
[66] H. Wang, F. Zhang, M. Hou, X. Xie, M. Guo, and Q. Liu, "SHINE: Signed heterogeneous information network embedding for sentiment link prediction," in *Proc. 11th ACM Int. Conf. Web Search Data Mining*, 2018, pp. 592–600.
[67] J. Wang, J. Chen, Z. Wang, and M. Gong, "Hypergraph contrastive attention networks for hyperedge prediction with negative samples evaluation," *Neural Netw.*, vol. 181, 2025, Art. no. 106807.
[68] X. Wang et al., "Heterogeneous graph attention network," in *Proc. World Wide Web Conf.*, 2019, pp. 2022–2032.
[69] X. Wang, N. Liu, H. Han, and C. Shi, "Self-supervised heterogeneous graph neural network with co-contrastive learning," in *Proc. 27th ACM SIGKDD Conf. Knowl. Discov. Data Mining*, 2021, pp. 1726–1736.
[70] X. Wang, Y. Zhang, and C. Shi, "Hyperbolic heterogeneous information network embedding," in *Proc. AAAI Conf. Artif. Intell.*, 2019, vol. 33, pp. 5337–5344.
[71] Y. Wang, Z. Zhang, J. Yin, X. Li, B. Shi, and B. Dong, "HGExplainer: Toward interpretable heterogeneous graph neural networks via meta-path perturbation," *J. Signal Process. Syst.*, vol. 97, no. 5, pp. 235–245, 2025.
[72] Z. Wang, Q. Li, D. Yu, X. Han, X.-Z. Gao, and S. Shen, "Heterogeneous graph contrastive multi-view learning," in *Proc. SIAM Int. Conf. Data Mining*, 2023, pp. 136–144.
[73] Z. Wang et al., "Graph neural networks in life sciences: Opportunities and solutions," in *Proc. 28th ACM SIGKDD Conf. Knowl. Discov. Data Mining*, 2022, pp. 4834–4835.
[74] H. Wu, Z. Xu, M. Wang, and X. Lyu, "Meta-learning enhanced physics-informed graph attention convolutional network for distribution power system state estimation," *IEEE Trans. Netw. Sci. Eng.*, vol. 12, no. 2, pp. 1186–1197, Mar./Apr. 2025.
[75] H. Xie, C. Song, H. Shao, and L. Wang, "Hypergraph semi-supervised contrastive learning for hyperedge prediction based on enhanced attention aggregator," *Entropy*, vol. 27, no. 10, 2025, Art. no. 1046.
[76] C. Xu and W. Song, "Intelligent task allocation for mobile crowdsensing with graph attention network and deep reinforcement learning," *IEEE Trans. Netw. Sci. Eng.*, vol. 10, no. 2, pp. 1032–1048, Mar./Apr. 2023.
[77] J. Xu et al., "A privacy-preserving auction for task offloading and resource allocation in UAV-assisted MEC," *IEEE Trans. Mobile Comput.*, vol. 25, no. 2, pp. 2611–2626, Feb. 2026.
[78] L. Xu and H. Chen, "FA-GCL: Feature-augmented graph contrastive learning method," *Neural Netw.*, vol. 193, 2026, Art. no. 108045.

Authorized licensed use limited to: TIANJIN UNIVERSITY. Downloaded on October 22,2026 at 01:15:35 UTC from IEEE Xplore. Restrictions apply.

[79] Y. Yan, Z. Zhao, Z. Yang, Y. Yu, and C. Li, "A fast and robust attention-free heterogeneous graph convolutional network," *IEEE Trans. Big Data*, vol. 10, no. 5, pp. 669–681, Oct. 2024.
[80] Q. Yang, H. Yu, Z. Liu, P. Li, X. Chen, and X. Luo, "Synthesizing global and local perspectives in contrastive learning for graph anomaly detection," *Knowl. Based Syst.*, vol. 315, 2025, Art. no. 113289.
[81] T. Yang et al., "Co-clustering interactions via attentive hypergraph neural network," in *Proc. 45th Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval*, 2022, pp. 859–869.
[82] J. Yin, P. Wang, S. Sun, and Z. Zheng, "Incomplete multi-view clustering via multi-level contrastive learning," *IEEE Trans. Knowl. Data Eng.*, vol. 37, no. 8, pp. 4716–4727, Aug. 2025.
[83] Y. You, T. Chen, Y. Sui, T. Chen, Z. Wang, and Y. Shen, "Graph contrastive learning with augmentations," in *Proc. Adv. Neural Inf. Process. Syst.*, H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, and H. Lin, Eds, 2020, vol. 33, pp. 5812–5823.
[84] M. Yu, Y. Zhang, J. Sun, M. Huang, T. Zhang, and G. Yu, "HG-SCC: A subgraph-aware convolutional few-shot classification method on heterogeneous graphs," *IEEE Trans. Knowl. Data Eng.*, vol. 37, no. 4, pp. 1871–1884, Apr. 2025.
[85] P. Yu, C. Fu, Y. Yu, C. Huang, Z. Zhao, and J. Dong, "Multiplex heterogeneous graph convolutional network," in *Proc. 28th ACM SIGKDD Conf. Knowl. Discov. Data Mining*, 2022, pp. 2377–2387.
[86] R. Yu, Y. Xin, Y. Dong, and J. Qian, "A graph contrastive learning model based on structural and semantic view for HIN recommendation," *Neural Process. Lett.*, vol. 56, no. 1, 2024, Art. no. 16.
[87] C. Zhang, D. Song, C. Huang, A. Swami, and N. V. Chawla, "Heterogeneous graph neural network," in *Proc. 25th ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2019, pp. 793–803.
[88] H.-K. Zhang, Y.-G. Zhang, Z. Zhou, and Y.-F. Li, "HONGAT: Graph attention networks in the presence of high-order neighbors," in *Proc. AAAI Conf. Artif. Intell.*, 2024, vol. 38, pp. 16750–16758.
[89] K. Zhang et al., "Heterogeneous causal metapath graph neural network for gene-microbe-disease association prediction," in *Proc. 23rd Int. Joint Conf. Artif. Intell.*, Jeju, South Korea, 2024, pp. 6161–6169.
[90] W. Zhang, Y. Fang, Z. Liu, M. Wu, and X. Zhang, "mg2vec: Learning relationship-preserving heterogeneous graph representations via meta-graph embedding," *IEEE Trans. Knowl. Data Eng.*, vol. 34, no. 3, pp. 1317–1329, Mar. 2022.
[91] Y. Zhang, X. Wang, C. Shi, X. Jiang, and Y. Ye, "Hyperbolic graph attention network," *IEEE Trans. Big Data*, vol. 8, no. 6, pp. 1690–1701, Dec. 2022.
[92] J. Zhao, X. Wang, C. Shi, B. Hu, G. Song, and Y. Ye, "Heterogeneous graph structure learning for graph neural networks," in *Proc. AAAI Conf. Artif. Intell.*, 2021, vol. 35, pp. 4697–4705.
[93] J. Zhao, X. Wang, C. Shi, Z. Liu, and Y. Ye, "Network schema preserving heterogeneous information network embedding," in *Proc. 29th Int. Joint Conf. Artif. Intell.*, C. Bessiere, Ed., 2020, pp. 1366–1372.
[94] Y. Zhao et al., "Learning bi-typed multi-relational heterogeneous graph via dual hierarchical attention networks," *IEEE Trans. Knowl. Data Eng.*, vol. 35, no. 9, pp. 9054–9066, Sep. 2023.
[95] S. Zheng, D. Guan, and W. Yuan, "Semantic-aware heterogeneous information network embedding with incompatible meta-paths," *World Wide Web*, vol. 25, no. 1, pp. 1–21, 2022.
[96] Z. Zhong, C.-T. Li, and J. Pang, "Personalised meta-path generation for heterogeneous graph neural networks," *Data Mining Knowl. Discov.*, vol. 36, no. 6, pp. 2299–2333, 2022.
[97] F. Zhu, Y. Wang, C. Chen, G. Liu, and X. Zheng, "A graphical and attentional framework for dual-target cross-domain recommendation," in *Proc. 29th Int. Joint Conf. Artif. Intell.*, 2021, pp. 3001–3008.
[98] Y. Zhu, Y. Xu, H. Cui, C. Yang, Q. Liu, and S. Wu, "Structure-enhanced heterogeneous graph contrastive learning," in *Proc. SIAM Int. Conf. Data Mining*, 2022, pp. 82–90.

Authorized licensed use limited to: TIANJIN UNIVERSITY. Downloaded on October 22,2026 at 01:15:35 UTC from IEEE Xplore. Restrictions apply.