Bridging Feature-structural Homophily and Long-range Heterogeneity for Self-supervised Heterogeneous Graph Learning

Minda Chen[1,2], **Yujie Mo**[3,*], **Junkai Huang**[1,2], **Guoqiu Wen**[1,2,*] and **Xiaofeng Zhu**[4,5]

[1]Key Lab of Education Blockchain and Intelligent Technology, Ministry of Education, Guangxi Normal University, China

[2]Guangxi Key Lab of Multi-source Information Mining & Security, China

[3]School of Computing, National University of Singapore, Singapore

[4]Hainan University, China

[5]School of Computer Science and Engineering, University of Electronic Science and Technology of China, China

Abstract

Self-supervised heterogeneous graph learning has achieved promising results in diverse applications but still faces two issues: (i) existing methods focus on either feature similarity or meta-path to capture homophily, neglecting their inherent complementarity; (ii) existing methods rely on meta-paths to capture interactions among same-type nodes, which may introduce noise and inherently exclude long-range cross-type interactions. To address these issues, we first propose a self-expressive solver that captures the complementary homophily between meta-paths and node features to obtain homophilous representations. Meanwhile, we design separate path encoders to model diverse interactions, thus explicitly including cross-type interactions while mitigating noise via adaptive fusion. Theoretical analysis verifies that homophilous representations exhibit a high-order grouping effect to capture complementary homophily, while path encoders possess adaptive smoothness capabilities to filter noise. Extensive experiments on diverse datasets, including a large-scale dataset, demonstrate the superiority of the proposed method.

1 Introduction

Heterogeneous graph learning aims to model complex interactions among various types of nodes and edges, thereby yielding discriminative representations [Dong *et al.*, 2017; Liu *et al.*, 2022]. To avoid labor-intensive manual labeling, self-supervised heterogeneous graph learning (SHGL) has emerged as a promising paradigm with remarkable success in diverse real-world applications, including recommendation systems [Huang *et al.*, 2025b; Huang *et al.*, 2025c], and drug discovery [Yu and Gao, 2022; Ma *et al.*, 2025].

Existing SHGL methods can be broadly grouped into two categories, *i.e.,* meta-path-based methods and feature-based methods. The former typically utilizes predefined meta-paths to explicitly capture structural homophily, aggregating information from connected nodes that may share the same label [Zhu *et al.*, 2022]. For example, HeCo [Wang *et al.*, 2021] aggregates neighbor information from nodes that may belong to the same class along meta-paths and generates representations for contrastive learning. In contrast, feature-based methods focus on inferring latent homophily structures directly from node features [Mo *et al.*, 2024a]. For example, HERO [Mo *et al.*, 2024b] learns an adaptive self-expressive matrix from node features by assigning large weights to node pairs that may belong to the same class.

Although existing methods have achieved promising results, there are still some limitations to be addressed. On one hand, most methods capturing homophily rely solely on meta-paths or node features, neglecting their inherent complementarity [Lv *et al.*, 2021]. As a result, models may suffer performance degradation when encountering sparse meta-path structures or noisy node features. On the other hand, existing methods are often limited to interactions among same-type nodes [Li *et al.*, 2024], which may aggregate neighbor information from different classes when meta-paths exhibit low homophily, leading to noise accumulation. Furthermore, they fail to capture long-range heterogeneity (*i.e.,* cross-type interaction), which may carry task-related information.

Based on the above analysis, it is possible to enhance the effectiveness of SHGL by jointly exploiting homophily from both meta-paths and node features while explicitly modeling long-range heterogeneity. To achieve this, there are two key challenges to be solved: (i) How to extract the complementary homophily from meta-paths and node features, as there may be potential conflicts between them; and (ii) How to mitigate the noise introduced by meta-paths while capturing the long-range heterogeneity in the heterogeneous graph.

In this paper, to address the above challenges, we propose a new framework to bridge Feature-structural hOmophily and long-Range heteroGEneity for effective SHGL (FORGE for short), as shown in Figure 1. Specifically, we design a self-expressive solver to capture homophily from both meta-paths and node features and obtain homophilous representations, thus tackling Challenge (i). Subsequently, we capture long-

*Corresponding authors

Code is available at https://github.com/MindChen/FORGE

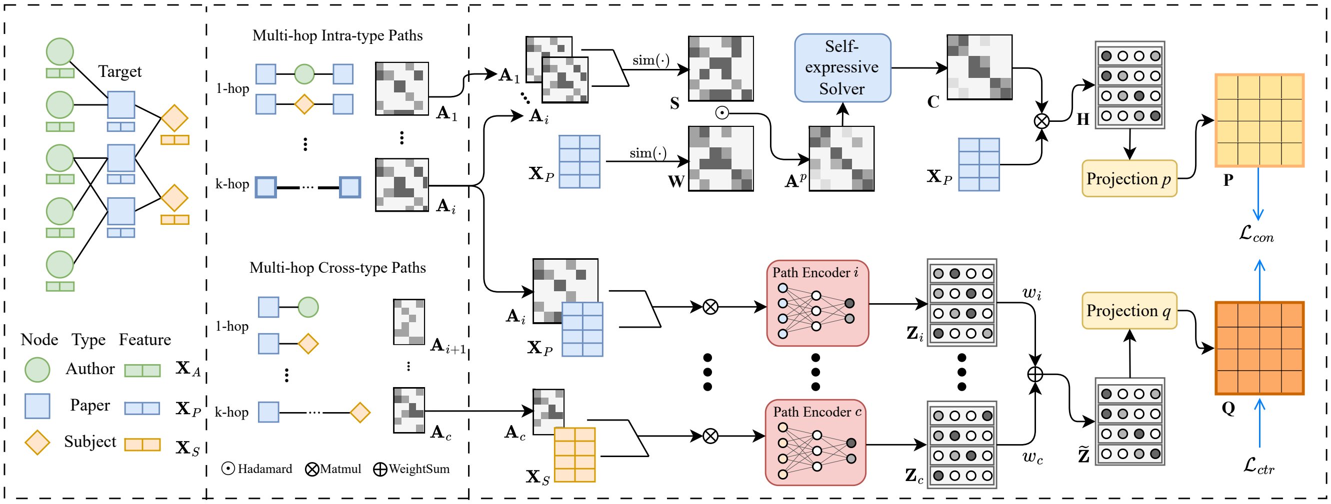


Figure 1: The flowchart of the proposed FORGE. Specifically, given a heterogeneous graph, we first generate subgraphs $\mathbf{A}_i$ from multi-hop intra-type and cross-type paths to distinguish diverse interactions. Then we design a self-expressive solver to capture the complementary homophily from the node features and intra-type paths similarity matrix (*i.e.,* $\mathbf{W}$ and $\mathbf{S}$), thus obtaining homophilous representations $\mathbf{H}$ with the self-expressive matrix $\mathbf{C}$. Meanwhile, we design separate path encoders to model both intra-type and cross-type interactions to obtain path-specific representations $\mathbf{Z}_i$ and adaptively fuse them to obtain path-fused representations $\widetilde{\mathbf{Z}}$. After that, we design a consistency loss $\mathcal{L}_{con}$ and a contrastive loss $\mathcal{L}_{ctr}$ to extract invariant information between the projected homophilous representations and path-fused representations (*i.e.,* $\mathbf{P}$ and $\mathbf{Q}$) while pushing apart nodes that may belong to different classes.

range heterogeneity by modeling multi-hop neighbors via path encoders while mitigating noise through adaptive fusion, thus tackling Challenge (ii). Finally, we theoretically verify that the learned homophilous representations have a high-order grouping effect that captures homophily from both meta-paths and node features. Moreover, we prove that path encoders have adaptive smoothness capabilities to filter noise.

Compared to previous SHGL methods, our main contributions can be summarized as follows:

- To the best of our knowledge, this is the first unified framework to bridge feature-structural homophily and long-range heterogeneity for self-supervised heterogeneous graph learning.
- We propose a self-expressive solver to capture complementary homophily between meta-paths and node features. Furthermore, we utilize path encoders and an adaptive fusion module to explicitly model long-range heterogeneity and filter noise, respectively.
- We theoretically demonstrate that homophilous representations have a high-order grouping effect that captures homophily from both meta-paths and node features. Furthermore, we validate that path encoders have adaptive capabilities for noise filtering.
- We experimentally demonstrate the superiority of the proposed method across diverse datasets, including a large-scale heterogeneous graph dataset, compared to numerous comparison methods.

2 Method

Notations. Let $\mathcal{G} = \{\mathcal{V}, \mathcal{E}, \mathbf{X}, \mathcal{T}, \mathcal{R}\}$ denote a heterogeneous graph, where $\mathcal{V}$ and $\mathcal{E}$ represent the node and edge sets, $\mathbf{X}$ denotes the node feature matrix, and $\mathcal{T}$ and $\mathcal{R}$ are the node and edge type sets, respectively.

Intra-type and cross-type paths. Given a heterogeneous graph $\mathcal{G}$, a multi-hop path $\mathcal{P}_i$ is defined as $t_1 \xrightarrow{r_1} t_2 \cdots \xrightarrow{r_k} t_{k+1}$, representing the composition relation $r_1 \circ \cdots \circ r_k$. Unlike previous methods that usually utilize meta-paths limited to intra-type connections, the multi-hop paths generalize this concept to cover cross-type connections. Then we classify multi-hop paths into two categories, *i.e.,* **intra-type paths** ($t_1 = t_{k+1}$) and **cross-type paths** ($t_1 \neq t_{k+1}$). We further compute the induced subgraph matrix for $\mathcal{P}_i$ as $\mathbf{A}_i = \hat{\mathbf{A}}_{t_1,t_2}\hat{\mathbf{A}}_{t_2,t_3}\cdots\hat{\mathbf{A}}_{t_k,t_{k+1}}$, where $\hat{\mathbf{A}}_{t_k,t_{k+1}}$ denotes the row-normalized adjacency matrix. As a result, subgraphs induced by intra-type paths can be used to capture structural homophily, while those induced by cross-type paths can be used to capture long-range heterogeneity.

2.1 Motivation

Existing methods typically capture homophily through either node features or meta-paths but often overlook their inherent complementarity, thereby degrading the model performance when facing sparse structures or noisy features. In this case, exploring the complementarity between node features and meta-paths may provide a potential solution. For instance, in Figure 2(a), low feature similarity between two nodes of the same class may be supplemented by connections from meta-paths (*e.g.,* nodes 1 and 2). Meanwhile, connections between nodes of the same class that are missing from meta-paths may be complemented by the high similarity of node features (*e.g.,* nodes 3 and 4).

In addition, existing methods generally rely on meta-paths to capture interactions among same-type nodes but may in-

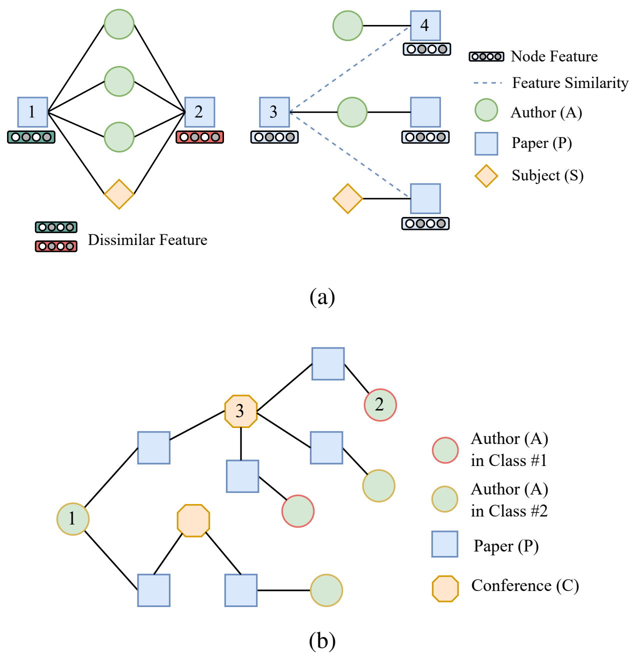


Figure 2: Example of complementary homophily and long-range heterogeneity. (a) On the left, for noisy features, meta-paths can serve as a supplement (*e.g.,* nodes 1 and 2); for sparse structures on the right, we can derive complementarity from feature similarity (*e.g.,* nodes 3 and 4). (b) Long-range heterogeneity (*e.g.,* A-P-C) can supplement discriminative information that cannot be obtained from meta-paths (*e.g.,* A-P-A, A-P-C-P-A).

troduce noise from different classes and exclude the long-range heterogeneity. For instance, in Figure 2(b), meta-paths (*i.e.,* A-P-C-P-A) may connect cross-class neighbors (*e.g.,* nodes 1 and 2) that hinder message-passing. Meanwhile, meta-paths filter out valuable heterogeneous neighbors (*e.g.,* connection between nodes 1 and 3), which may provide a rich heterogeneous context (*e.g.,* features of author-attended conferences) that helps to distinguish authors.

Therefore, it is critical to capture complementary homophily and model long-range heterogeneity in the heterogeneous graph, which are ignored in previous SHGL methods. Although some studies [Mo *et al.*, 2024a; Mo *et al.*, 2024b] attempt to capture heterogeneity through multi-layer single-hop aggregation over heterogeneous neighbors, they may face the over-smoothing issue and heavy computational costs [Yang *et al.*, 2023; Li *et al.*, 2024]. In this paper, we aim to address these issues and bridge complementary homophily and long-range heterogeneity, thereby enabling the model to generate node representations based on both of them.

2.2 Complementary Homophily Extraction

Previous SHGL methods generally capturing homophily rely on either meta-paths or node features. For instance, CPIM [Mo *et al.*, 2023b] maximizes mutual information between node representations of different meta-path-based graphs, while SCHOOL [Mo *et al.*, 2024a] incorporates rank-constrained spectral clustering on node features. However, these methods neglect the inherent complementarity between these two perspectives. To solve this issue, we propose a self-expressive solver to capture complementary homophily from intra-type paths and node features.

Traditionally, the self-expressive solver [Chapelle *et al.*, 2002; He *et al.*, 2019; Li *et al.*, 2020] can be illustrated as follows:

$$\begin{aligned} &\min_{\mathbf{C}} \|\mathbf{X}_t - \mathbf{C}\mathbf{X}_t\|_F^2 + \gamma\|\mathbf{C} - \mathbf{W}\|_F^2 + \lambda\|\mathbf{C}\|, \\ &\text{s.t. } \operatorname{diag}(\mathbf{C} = 0), \end{aligned} \tag{1}$$

where $\mathbf{X}_t$ is the feature matrix of target nodes, $\mathbf{W}$ is the similarity matrix obtained by calculating the cosine similarity of node features, $\mathbf{C}$ is the self-expressive matrix, γ and λ are non-negative parameters. In Eq. (1), the first term linearly reconstructs every node by other nodes, the second ensures that nodes with similar features contribute significantly, the third term is a regularization to promote sparsity, and the constraint $\operatorname{diag}(\mathbf{C} = 0)$ prevents nodes from reconstructing themselves by setting the diagonal elements to zero. Based on the assumption that nodes with similar features may belong to the same class, the value of the element $\mathbf{c}_{ij}$ reflects the likelihood of nodes v_i and v_j belonging to the same class, thus capturing homophily from node features.

However, the traditional self-expressive solver overlooks the structural homophily from intra-type paths. To address this issue, we constrain the self-expression matrix to leverage the structural homophily.

Specifically, we first construct the path similarity matrix $\mathbf{S}$ to capture structural homophily from subgraphs of intra-type paths, *i.e.,*

$$\begin{aligned} &\mathbf{M} = \frac{1}{p}\sum_{i=1}^{p} \mathbf{A}_i, \\ &\mathbf{S}^{(i,j)} = \operatorname{sim}(\mathbf{M}^{(i,:)}, \mathbf{M}^{(j,:)}), \end{aligned} \tag{2}$$

where p is the number of intra-type paths, $\mathbf{M}$ is the averaged subgraph over all intra-type paths, $\mathbf{M}^{(i,:)}$ denotes the i-th row of matrix $\mathbf{M}$, and $\operatorname{sim}(\cdot,\cdot)$ indicates the cosine similarity function. A larger $\mathbf{S}^{(i,j)}$ represents more shared neighbors across all intra-type paths between the nodes v_i and v_j. Since nodes sharing more neighbors often belong to the same class [Perozzi *et al.*, 2014], the path similarity matrix $\mathbf{S}$ reflects the homophily derived from intra-type paths.

After that, we incorporate $\mathbf{S}$ into the self-expression solver, enabling the self-expressive matrix $\mathbf{C}$ to leverage homophily from both node features and intra-type paths. To do this, we combine $\mathbf{S}$ and $\mathbf{W}$ to obtain a new constraint matrix $\mathbf{A}^p$, *i.e.,*

$$\mathbf{A}^p = \operatorname{TopK}(\mathbf{S} \odot \mathbf{W} - \mathbf{I}), \tag{3}$$

where $\odot$ denotes the Hadamard product and $\operatorname{TopK}(\cdot)$ means selecting the top K elements of each row, then assign the values of these selected elements to 1 and the remaining elements to 0. In Eq. (3), $\mathbf{A}^p$ weights $\mathbf{S}$ with $\mathbf{W}$, providing path similarity while considering the original constraints, thus serving as the new constraint matrix. Consequently, the self-expressive solver designed to capture complementary homophily from both node features and intra-type paths is given as follows:

$$\begin{aligned} &\min_{\mathbf{C}} \|\mathbf{X}_t - \mathbf{C}\mathbf{X}_t\|_F^2 + \gamma\|\mathbf{C} - \mathbf{A}^p\|_F^2 + \lambda\|\mathbf{C}\|, \\ &\text{s.t. } \operatorname{diag}(\mathbf{C} = 0). \end{aligned} \tag{4}$$

In Eq. (4), the first term leverages homophily from node features, while the second term leverages homophily from intra-type paths. Moreover, the second term penalizes the weights between nodes of different classes induced by the first term, while the first term captures the intra-class connections overlooked by the second term, thereby complementing each other. Consequently, the self-expressive solver captures the complementarity by balancing them.

With the self-expressive matrix $\mathbf{C}$, we then obtain homophilous representations $\mathbf{H}$, *i.e.,*

$$\mathbf{H} = \mathbf{C}\mathbf{X}. \tag{5}$$

Therefore, homophilous representations are able to capture the inherent complementary homophily between node features and intra-type paths, thus conducting the message-passing among nodes that likely belong to the same class.

In this paper, we further prove that the homophilous representations $\mathbf{H}$ exhibit a high-order grouping effect that captures homophily from both node features and intra-type paths. To do this, we first follow [Li *et al.*, 2020; Xie *et al.*, 2025] to define the grouping effect as follows.

Definition 1. *(Grouping Effect) For every node pair v_i and v_j satisfying $v_i \to v_j$ (*i.e., $\|X_i - X_j\|_2 \to 0$ *), the matrix $\mathbf{G} \in \mathbb{R}^{N\times F}$ is said to have a grouping effect if $|G_{ip} - G_{jp}| \to 0, \forall 1 \le p \le F$.*

Traditional grouping effects primarily emphasize similarities of node features, whereas real-world graphs often exhibit complex structures. To address this, we first establish an upper bound on the distance between the homophilous representations of two nodes (*i.e.,* $\|\mathbf{H}_i - \mathbf{H}_j\|$).

Proposition 1. *(Distance Bound of Homophilous Representations) For any two nodes v_i and v_j, $\|\mathbf{H}_i - \mathbf{H}_j\|$ is bounded by,*

$$\begin{aligned}
\|\mathbf{H}_i - \mathbf{H}_j\| &\le O\Big(\mathcal{M} + \mathcal{N}\Big), \\
\mathcal{M} &= \Big\|\hat{\mathbf{M}}_i - \hat{\mathbf{M}}_j\Big\|^2 + \Big\|\hat{\mathbf{X}}_i - \hat{\mathbf{X}}_j\Big\|^2, \\
\mathcal{N} &= \sum_{m \neq i,j} \Big(\|\hat{\mathbf{M}}_i^\top \hat{\mathbf{M}}_m - \hat{\mathbf{M}}_j^\top \hat{\mathbf{M}}_m\| + \\
&\qquad \|\hat{\mathbf{X}}_i^\top \hat{\mathbf{X}}_m - \hat{\mathbf{X}}_j^\top \hat{\mathbf{X}}_m|\Big),
\end{aligned} \tag{6}$$

where $\hat{\mathbf{M}} = \mathbf{M}/\|\mathbf{M}\|$ and $\hat{\mathbf{X}} = \mathbf{X}/\|\mathbf{X}\|$.

Specifically, $\mathcal{M}$ reflects the direct similarity between nodes v_i and v_j in terms of both intra-type paths (via $\hat{\mathbf{M}}$) and node features (via $\hat{\mathbf{X}}$). Meanwhile, $\mathcal{N}$ measures the consistency of their similarity to all other intermediate nodes v_m (*i.e.,* high-order similarity), thereby capturing high-order information. Based on Proposition 1, we formally derive the high-order grouping effect of the homophilous representations $\mathbf{H}$.

Theorem 1. *(High-order Grouping Effect) The homophilous representations $\mathbf{H}$ exhibit a high-order grouping effect. Specifically, for any two nodes v_i and v_j satisfying $v_i \to v_j$ (*i.e., $\mathcal{M} + \mathcal{N} \to 0$*), their representations converge,* i.e.,

$$v_i \to v_j \implies \|\mathbf{H}_i - \mathbf{H}_j\| \to 0. \tag{7}$$

Theorem 1 theoretically guarantees that if two nodes share similar features and intra-type path structures, implying they likely belong to the same class, their learned representations will be close. This indicates that the homophilous representations $\mathbf{H}$ capture homophily from both node features and intra-type paths. The proof is provided in Appendix B.1.

2.3 Multi-hop Path Encoding

Previous studies rely on meta-paths to capture interactions among same-type nodes, which may introduce noise [Yang *et al.*, 2023]. Furthermore, they neglect long-range heterogeneity, which may lose information from nodes of different types and thus affect model performance [Li *et al.*, 2024]. To solve these issues, we utilize path encoders to filter noise in intra-type paths and capture long-range heterogeneity to introduce more task-related information.

Specifically, we employ a multi-layer perceptron (MLP) as the encoder ϕ_i to obtain path-specific representations $\mathbf{Z}_i$ on each subgraph $\mathbf{A}_i$ of the multi-hop path, *i.e.,*

$$\mathbf{Z}_i = \sigma(\mathbf{A}_i \mathbf{X}_{t_{k+1}} \mathbf{W}_\phi), \tag{8}$$

where σ is the activation function, $\mathbf{W}_\phi$ denotes the trainable parameters of the encoder ϕ_i, and $\mathbf{X}_{t_{k+1}}$ is the feature matrix of node type t_{k+1} corresponding to the multi-hop paths. In this paper, the subgraph $\mathbf{A}_i$ is constructed from either intra-type or cross-type paths. When $\mathbf{A}_i$ is a subgraph of an intra-type path, $\mathbf{Z}_i$ can capture information from same-type neighbors. Unlike predefined meta-paths focusing on a few symmetric paths among same-type nodes, intra-type paths flexibly cover more paths, including asymmetric ones without predefined selection. Meanwhile, when $\mathbf{A}_i$ is a subgraph of a cross-type path, $\mathbf{Z}_i$ can capture long-range heterogeneity and aggregate information from nodes of different types.

To effectively utilize representations $\mathbf{Z}_i$ from both intra-type and cross-type paths, we then adaptively fuse them to obtain path-fused representations $\widetilde{\mathbf{Z}}$ as follows:

$$\widetilde{\mathbf{Z}} = \sum_{i=1}^{c} w_i \mathbf{Z}_i, \tag{9}$$

where w_i is a learnable weight and c is the number of all paths. By assigning different weights to the paths, the path-fused representations $\widetilde{\mathbf{Z}}$ can fuse information from both intra-type and cross-type neighbors while mitigating noise. We further demonstrate in Theorem 2 that the path encoders possess adaptive smoothness capabilities, enabling them to mitigate the noise introduced by intra-type paths [Gu *et al.*, 2025].

Theorem 2. *The design of path encoders can approximate arbitrary graph filters for adaptive smoothness capabilities extending beyond simple low- or high-pass ones, expressing the K-order polynomial graph filter ($\sum_{i=0}^{K} \theta_i \widehat{\mathbf{L}}^i$) with arbitrary coefficients θ_i.*

Theorem 2 indicates that the design of path encoders can approximate the K-order polynomial graph filters with arbitrary coefficients, which can suppress noise while preserving task-related information by assigning proper coefficients. As a result, the design of path encoders can mitigate the impact of noise. The proof can be found in Appendix B.2.

Method	ACM		Yelp		DBLP		Aminer	
	Macro-F1	Micro-F1	Macro-F1	Micro-F1	Macro-F1	Micro-F1	Macro-F1	Micro-F1
Deep Walk	73.9±0.3	74.1±0.1	68.7±1.1	73.2±0.9	88.1±0.2	89.5±0.3	54.7±0.8	59.7±0.7
GCN	86.9±0.2	87.0±0.3	85.0±0.6	87.4±0.8	90.2±0.2	90.9±0.5	64.5±0.7	71.5±0.9
GAT	85.0±0.4	84.9±0.3	86.4±0.5	88.2±0.7	91.0±0.4	92.1±0.2	63.8±0.4	70.6±0.7
Mp2vec	87.6±0.5	88.1±0.3	78.2±0.8	83.6±0.9	85.7±0.3	87.6±0.6	58.7±0.5	65.3±0.6
HAN	89.4±0.2	89.2±0.2	90.5±1.2	90.7±1.4	91.2±0.4	92.0±0.5	65.3±0.7	72.8±0.4
HGT	91.5±0.7	91.6±0.6	89.9±0.5	90.2±0.6	90.9±0.6	91.7±0.8	64.5±0.5	71.0±0.7
DMGI	89.8±0.1	89.8±0.1	82.9±0.8	85.8±0.9	92.1±0.2	92.9±0.3	63.8±0.4	67.6±0.5
DMGIattn	88.7±0.3	88.7±0.5	82.8±0.7	85.4±0.5	90.9±0.2	91.8±0.3	62.4±0.9	66.8±0.8
HDMI	90.1±0.3	90.1±0.3	80.7±0.6	84.0±0.9	91.3±0.2	92.2±0.5	65.9±0.4	71.7±0.6
HeCo	88.3±0.3	88.2±0.2	85.3±0.7	87.9±0.6	91.0±0.3	91.6±0.2	71.8±0.9	78.6±0.7
HGCML	90.6±0.7	90.7±0.5	90.7±0.8	91.0±0.7	91.9±0.8	93.2±0.7	70.5±0.4	76.3±0.6
CPIM	91.4±0.3	91.3±0.2	90.2±0.5	90.3±0.4	93.2±0.6	93.8±0.8	70.1±0.9	75.8±1.1
HGMAE	90.5±0.5	90.6±0.7	90.5±0.7	90.7±0.5	92.9±0.5	93.4±0.6	72.3±0.9	80.3±1.2
HERO	92.2±0.5	92.1±0.7	92.4±0.7	92.3±0.6	93.8±0.6	94.4±0.4	75.1±0.7	84.5±0.9
MGHC	92.3±0.3	92.4±0.4	92.5±0.4	92.3±0.3	93.0±0.3	93.6±0.5	68.8±0.6	72.0±0.7
D^{2}CMG	92.6±0.1	92.7±0.2	92.8±0.4	92.5±0.5	92.7±0.3	93.4±0.3	70.9±0.4	77.0±0.5
FORGE	**93.5±0.6**	**93.4±0.5**	**94.3±0.3**	**94.0±0.2**	**94.3±0.4**	**94.9±0.4**	**79.9±0.5**	**88.9±0.6**

Table 1: Classification performance (*i.e.,* Macro-F1 and Micro-F1) on four heterogeneous graph datasets.

2.4 Optimization Objective

Given homophilous representations $\mathbf{H}$ and path-fused representations $\widetilde{\mathbf{Z}}$, both are derived for the same node from different perspectives and share the same original node features, thus intuitively containing consistent information. Therefore, we design a consistency loss to extract the consistent information between them.

Specifically, we first propose to learn projection heads p and q to map homophilous representations $\mathbf{H}$ and path-fused representations $\widetilde{\mathbf{Z}}$ into the same latent space, *i.e.,* $\mathbf{P} = p(\mathbf{H}), \mathbf{Q} = q(\widetilde{\mathbf{Z}})$. After that, we design a consistency loss to maximize the invariance between them, *i.e.,*

$$\mathcal{L}_{con} = \frac{1}{N}\sum_{i=1}^{N}\|\mathbf{p}_i - \mathbf{q}_i\|^2. \tag{10}$$

Eq. (10) encourages the alignment between the projected homophilous and path-fused representations $\mathbf{P}$ and $\mathbf{Q}$, thereby capturing their consistent information.

However, the consistency loss lacks the constraint for nodes from different classes, which may make node representations of different classes difficult to distinguish, thereby degrading downstream task performance. To solve this issue, based on the assumption that near neighbors are likely to belong to the same class [Zhou *et al.*, 2003], we further design a contrastive loss $\mathcal{L}_{ctr}$ to push apart node representations that may belong to different classes:

$$\begin{aligned} &\mathcal{L}_{ctr} = -\frac{1}{N}\sum_{i=1}^{N}\log\frac{1}{\sum_{k\neq i} w_{ik}\cdot e^{(\mathbf{q}_i^\top \mathbf{q}_k/\tau)}}, \\ &w_{ij} = \begin{cases} a, & \mathbf{q}_j \in \mathrm{KNN}(\mathbf{q}_i), \\ b, & otherwise, \end{cases} \end{aligned} \tag{11}$$

where τ is a temperature parameter, $\mathrm{KNN}(\mathbf{q}_i)$ means the set formed by the k nearest neighbors of node v_i [Cover and Hart, 1967], a and b are non-negative parameters assigned to neighbors and non-neighbors of node v_i. Therefore, by setting $a < b$, the contrastive loss focuses more on separating node representations that may belong to different classes, thereby increasing the inter-class distance.

Integrating the consistency loss $\mathcal{L}_{con}$ in Eq. (10) and the contrastive loss $\mathcal{L}_{ctr}$ in Eq. (11), the objective function $\mathcal{J}$ is formulated as:

$$\mathcal{J} = \alpha\mathcal{L}_{con} + \beta\mathcal{L}_{ctr}, \tag{12}$$

where α and β are non-negative parameters. Finally, we concatenate homophilous representations $\mathbf{H}$ with path-fused representations $\widetilde{\mathbf{Z}}$ to obtain final representations $\hat{\mathbf{Z}}$ for downstream tasks. As a result, the final representations $\hat{\mathbf{Z}}$ contain homophily from both node features and intra-type paths, while also modeling long-range heterogeneity.

3 Experiments

In this section, we conduct experiments on five heterogeneous graph datasets to evaluate the proposed FORGE in different downstream tasks (*i.e.,* node classification and similarity search), compared to heterogeneous and homogeneous graph methods. Details of the experimental settings and additional results are shown in Appendix C.

3.1 Experimental Setup

Datasets

The used heterogeneous graph datasets include three academic datasets (*i.e.,* ACM [Wang *et al.*, 2019], DBLP [Wang *et al.*, 2019], and Aminer [Hu *et al.*, 2019]), one business dataset (*i.e.,* Yelp [Lu *et al.*, 2019]), and one large-scale academic dataset (*i.e.,* ogbn-mag [Wang *et al.*, 2020]).

Comparison Methods

The comparison methods include fourteen heterogeneous graph methods and seven homogeneous graph methods. The

Method	ACM		Yelp		DBLP		Aminer	
	Sim@5	Sim@10	Sim@5	Sim@10	Sim@5	Sim@10	Sim@5	Sim@10
Deep Walk	78.7±0.2	76.7±0.3	73.6±0.4	72.1±0.7	85.7±0.2	84.6±0.3	67.1±0.4	65.9±0.6
GCN	86.8±0.3	84.9±0.3	85.1±0.5	83.9±0.4	88.2±0.1	87.4±0.2	77.4±0.4	75.2±0.5
GAT	86.5±0.2	85.4±0.4	85.3±0.3	84.2±0.5	90.6±0.3	90.2±0.3	76.8±0.5	74.6±0.6
Mp2vec	82.3±0.4	81.2±0.5	79.1±0.6	77.9±0.7	87.1±0.5	85.8±0.6	71.8±0.3	68.6±0.4
HAN	87.2±0.2	85.6±0.3	87.5±0.7	87.4±0.4	89.3±0.3	89.3±0.5	78.2±0.4	78.0±0.5
HGT	90.1±0.5	90.0±0.6	88.6±0.7	88.0±0.5	90.3±0.4	89.9±0.7	79.2±0.7	78.3±0.6
DMGI	89.8±0.3	89.1±0.1	83.2±0.5	82.5±0.7	91.3±0.4	91.0±0.3	76.3±0.8	73.3±0.7
DMGIattn	90.1±0.2	88.9±0.4	80.7±0.4	79.6±0.2	89.0±0.2	87.9±0.4	72.8±0.6	71.5±0.5
HDMI	89.8±0.4	89.0±0.2	83.2±0.4	79.8±0.7	91.2±0.3	90.9±0.2	78.5±0.5	77.1±0.7
HeCo	89.6±0.2	88.7±0.2	86.4±0.3	85.1±0.4	90.9±0.2	90.6±0.1	82.9±0.7	81.7±0.5
HGCML	89.5±0.5	89.0±0.6	87.1±0.6	85.8±0.7	91.3±0.4	90.8±0.6	81.7±0.6	80.5±0.4
CPIM	89.6±0.6	88.8±0.4	87.9±0.7	86.8±0.5	91.1±0.6	91.0±0.3	82.1±0.6	80.9±0.8
HGMAE	89.5±0.3	88.9±0.5	88.1±0.8	86.9±0.7	90.9±0.4	90.7±0.5	80.5±0.4	79.5±0.6
HERO	89.7±0.6	89.2±0.4	89.7±0.3	88.9±0.5	92.1±0.7	91.9±0.5	86.8±0.3	85.1±0.6
MGHC	90.1±0.6	90.0±0.5	88.3±0.6	87.5±0.3	91.4±0.5	91.2±0.7	83.1±0.4	81.8±0.5
FORGE	**90.5±0.5**	**90.0±0.4**	**90.5±0.3**	**89.9±0.4**	**93.0±0.4**	**92.7±0.5**	**88.1±0.4**	**86.3±0.6**

Table 2: Similarity search performance (*i.e.,* Sim@5 and Sim@10) on heterogeneous graph datasets.

Method	Macro-F1	Micro-F1
MLP	10.7±0.2	26.9±0.3
GCN	13.6±0.2	30.1±0.3
GAT	13.9±0.2	30.5±0.3
DGI	15.0±0.2	31.4±0.3
GCA	15.1±0.2	31.4±0.3
GIC	15.3±0.1	31.7±0.2
SUGRL	16.0±0.1	32.4±0.1
LatGRL-S	19.4±0.2	36.8±0.5
FORGE*	**21.4±0.2**	**40.6±0.5**

Table 3: Classification performance (*i.e.,* Macro-F1 and Micro-F1) on ogbn-mag datasets. * indicates the mini-batch version.

former includes two semi-supervised methods (*i.e.,* HAN [Wang *et al.*, 2019] and HGT [Hu *et al.*, 2020]), one traditional unsupervised method (*i.e.,* Mp2vec [Dong *et al.*, 2017]), and eleven self-supervised methods (*i.e.,* DMGI [Park *et al.*, 2020], DMGIattn [Park *et al.*, 2020], HeCo [Wang *et al.*, 2021], HDMI [Jing *et al.*, 2021], CPIM [Mo *et al.*, 2023b], HGCML [Wang *et al.*, 2023], HGMAE [Tian *et al.*, 2023], HERO [Mo *et al.*, 2024b], MGHC [Huang *et al.*, 2025a], LatGRL-S [Shen and Kang, 2025] and $\mathrm{D^2CMG}$ [Wang *et al.*, 2025]). The latter includes two semi-supervised methods (*i.e.,* GCN [Kipf and Welling, 2017] and GAT [Veličković *et al.*, 2018]), one traditional unsupervised method (*i.e.,* DeepWalk [Perozzi *et al.*, 2014]), and four self-supervised methods (*i.e.,* DGI [Veličković *et al.*, 2019], GIC [Mavromatis and Karypis, 2021], GCA [Zhu *et al.*, 2021], SUGRL [Mo *et al.*, 2022]). For a fair comparison, we select meta-paths for previous meta-path-based SHGL methods following [Dong *et al.*, 2017; Wang *et al.*, 2019; Lu *et al.*, 2019; Lv *et al.*, 2021]. Moreover, for homogeneous graph methods on heterogeneous graph datasets, we follow the setting of [Mo *et al.*, 2023a] by learning meta-path-based representations separately and concatenating them for downstream tasks.

3.2 Results Analysis

Effectiveness on Heterogeneous Graph

We first evaluate the effectiveness of the proposed method on four heterogeneous graph datasets by reporting the results of node classification (*i.e.,* Macro-F1 and Micro-F1) in Table 1 and similarity search (*i.e.,* Sim@5 and Sim@10) in Table 2. Obviously, our method achieves the best performance on both node classification and similarity search tasks.

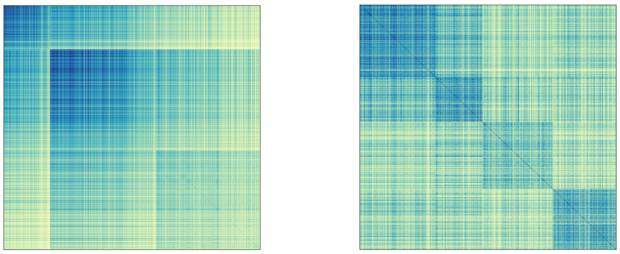
Firstly, for the node classification task, the proposed method achieves superior performance over comparison methods. For example, on four heterogeneous graph datasets, the proposed method improves by 2.1% on average compared to the best comparison method (*i.e.,* HERO). This can be attributed to the fact that the proposed method captures homophily from both node features and intra-type paths as well as long-range heterogeneity, thus producing discriminative representations to enhance classification performance. Secondly, for the similarity search task, the proposed method also achieves promising improvements. For example, the proposed method achieves an average improvement of 1.0% compared to the best comparison method (*i.e.,* HERO) on four heterogeneous graph datasets. This is due to the homophilous representations having a high-order grouping effect to capture homophily, which enhances intra-class node similarity. As a result, the effectiveness of the proposed method is verified on the heterogeneous graph datasets in terms of different downstream tasks.

Effectiveness on Large-scale Heterogeneous Graph

We further evaluate the proposed method on the large-scale heterogeneous graph dataset (*i.e.,* ogbn-mag) and report the results of node classification (*i.e.,* Macro-F1 and Micro-F1) in Table 3. We can observe that the proposed method also achieves the best results on the large-scale heterogeneous graph dataset, compared to the comparison methods. For ex-

Method	**ACM**		**Yelp**		**DBLP**		**Aminer**	
	Macro-F1	Micro-F1	Macro-F1	Micro-F1	Macro-F1	Micro-F1	Macro-F1	Micro-F1
w/o $\widetilde{\mathbf{Z}}$	90.3±0.2	90.2±0.1	92.4±0.2	91.7±0.2	89.6±0.1	90.8±0.1	64.4±0.1	71.1±0.2
w/o $\mathbf{H}$	93.1±0.3	93.1±0.4	89.6±0.5	89.1±0.4	94.0±0.3	94.6±0.4	77.5±0.6	86.8±0.6
FORGE	**93.5±0.6**	**93.4±0.5**	**94.3±0.3**	**94.0±0.2**	**94.3±0.2**	**94.9±0.3**	**79.9±0.5**	**88.9±0.6**

Table 4: Classification performance (*i.e.,* Macro-F1 and Micro-F1) of final representations from different variants.



(a) Yelp (b) DBLP

Figure 3: Node correlation maps of Yelp and DBLP datasets.

ample, the proposed method improves by 2.9% on average compared to the best comparison method (*i.e.,* LatGRL-S). This indicates that the proposed method can also be extended to the large-scale heterogeneous graph while capturing homophily and long-range heterogeneity. As a result, the effectiveness of the proposed method is further verified on the large-scale heterogeneous graph dataset.

Visualization

To verify the high-order grouping effect of homophilous representations $\mathbf{H}$, we visualize the node correlation maps on the Yelp and DBLP datasets in Figure 3, with both rows and columns permuted according to node labels. In these maps, pixel intensity reflects the correlation strength between two nodes. As shown in Figure 3, the correlation maps display a block-diagonal structure, where nodes inside the same block belong to the same class. This indicates that nodes from the same class are well aligned in the representation space. Overall, these observations confirm the high-order grouping effect of the homophilous representations that effectively capture homophily in heterogeneous graphs.

3.3 Ablation Study

In this section, we conduct ablation experiments on several designs of the proposed FORGE.

To evaluate the benefits of the concatenated representations $\hat{\mathbf{Z}}$, we conduct experiments using only homophilous or path-fused representations, with results summarized in Table 4. We observe that these variants achieve inferior performance compared to the proposed method. This indicates that the concatenated representations integrating homophily and long-range heterogeneity may capture more task-related information, thus enhancing performance.

To evaluate the sensitivity of the top-K value in the self-expressive solver, we conduct experiments on different K values, with results shown in Figure 4. We observe that the proposed method achieves optimal performance with a relatively

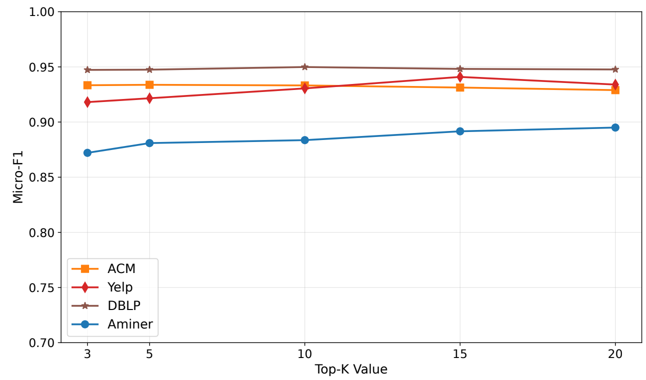


Figure 4: Performance of different top-K values.

small K value, indicating that a large number of neighbors is not necessary. Furthermore, the results show only minor performance variation across different K values. This indicates that the proposed method is robust to the choice of K, which enhances its practical applicability.

4 Conclusion

In this paper, we proposed a self-supervised heterogeneous graph learning framework to capture both homophily and long-range heterogeneity. To do this, we first designed a self-expressive solver to capture the complementary homophily between node features and intra-type paths, then utilized path encoders to capture long-range intra-type and cross-type interactions while filtering noise via adaptive fusion. Theoretical analysis indicates that homophilous representations have a high-order grouping effect that captures homophily from both node features and intra-type paths, and path encoders have adaptive smoothness capability to mitigate noise, while bridging homophily and heterogeneity provides more task-related information. Extensive experiments demonstrate the effectiveness of the proposed method on five heterogeneous graph datasets, including one large-scale graph.

Acknowledgements

This work was supported in part by the Guangxi Natural Science Foundation (2026GXNSFAA00640998, 2023GXNSFBA026010), the Joint Cultivation Projects of the National Natural Science Foundation of China for Guangxi Normal University (2025PY039), and the Education and Teaching Reform Project of Guangxi Normal University (Grant No. 2025JGZ14).

References

[Chapelle *et al.*, 2002] Olivier Chapelle, Jason Weston, and Bernhard Schölkopf. Cluster kernels for semi-supervised learning. In *NeurIPS*, volume 15, 2002.

[Cover and Hart, 1967] T. Cover and P. Hart. Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, 13(1):21–27, 1967.

[Dong *et al.*, 2017] Yuxiao Dong, Nitesh V. Chawla, and Ananthram Swami. metapath2vec: Scalable representation learning for heterogeneous networks. In *SIGKDD*, page 135–144, 2017.

[Gu *et al.*, 2025] Ming Gu, Zhuonan Zheng, Sheng Zhou, Meihan Liu, Jiawei Chen, Qiaoyu Tan, Liangcheng Li, and Jiajun Bu. Making classic gnns strong baselines across varying homophily: A smoothness–generalization perspective. In *NeurIPS*, 2025.

[He *et al.*, 2019] Li He, Nilanjan Ray, Yisheng Guan, and Hong Zhang. Fast large-scale spectral clustering via explicit feature mapping. *IEEE Transactions on Cybernetics*, 49(3):1058–1071, 2019.

[Hu *et al.*, 2019] Binbin Hu, Yuan Fang, and Chuan Shi. Adversarial learning on heterogeneous information networks. In *SIGKDD*, pages 120–129, 2019.

[Hu *et al.*, 2020] Ziniu Hu, Yuxiao Dong, Kuansan Wang, and Yizhou Sun. Heterogeneous graph transformer. In *WWW*, pages 2704–2710, 2020.

[Huang *et al.*, 2025a] Yudi Huang, Ci Nie, Hongqing He, Yujie Mo, Yonghua Zhu, Guoqiu Wen, and Xiaofeng Zhu. Multiplex graph representation learning with homophily and consistency. In *AAAI*, volume 39, pages 11835–11842, 2025.

[Huang *et al.*, 2025b] Zhirong Huang, Debo Cheng, Lin Liu, Jiuyong Li, Guangquan Lu, and Shichao Zhang. Interaction-data-guided conditional instrumental variables for debiasing recommender systems. In *IJCAI*, pages 2955–2963, 2025.

[Huang *et al.*, 2025c] Zhirong Huang, Shichao Zhang, Debo Cheng, Jiuyong Li, Lin Liu, Guangquan Lu, and Guixian Zhang. Learning instrumental variable representation for debiasing in recommender systems. *Neural Networks*, page 107977, 2025.

[Jing *et al.*, 2021] Baoyu Jing, Chanyoung Park, and Hanghang Tong. Hdmi: High-order deep multiplex infomax. In *WWW*, pages 2414–2424, 2021.

[Kipf and Welling, 2017] Thomas N. Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. In *ICLR*, 2017.

[Li *et al.*, 2020] Xiang Li, Ben Kao, Caihua Shan, Dawei Yin, and Martin Ester. Cast: a correlation-based adaptive spectral clustering algorithm on multi-scale data. In *SIGKDD*, pages 439–449, 2020.

[Li *et al.*, 2024] Chao Li, Zijie Guo, Kun He, et al. Long-range meta-path search on large-scale heterogeneous graphs. In *NeurIPS*, volume 37, pages 44240–44268, 2024.

[Liu *et al.*, 2022] Nian Liu, Xiao Wang, Deyu Bo, Chuan Shi, and Jian Pei. Revisiting graph contrastive learning from the perspective of graph spectrum. In *NeurIPS*, volume 35, pages 2972–2983, 2022.

[Lu *et al.*, 2019] Yuanfu Lu, Chuan Shi, Linmei Hu, and Zhiyuan Liu. Relation structure-aware heterogeneous information network embedding. In *AAAI*, volume 33, pages 4456–4463, 2019.

[Lv *et al.*, 2021] Qingsong Lv, Ming Ding, Qiang Liu, Yuxiang Chen, Wenzheng Feng, Siming He, Chang Zhou, Jianguo Jiang, Yuxiao Dong, and Jie Tang. Are we really making much progress? revisiting, benchmarking and refining heterogeneous graph neural networks. In *SIGKDD*, pages 1150–1160, 2021.

[Ma *et al.*, 2025] Yunjie Ma, Fei Wang, Qiyu Feng, Zuocheng Wang, and Luyu Xie. HeMDAP: Heterogeneous Graph Self-Supervised Learning for MiRNA-Disease Association Prediction . *IEEE Transactions on Computational Biology and Bioinformatics*, 22(02):769–780, 2025.

[Mavromatis and Karypis, 2021] Costas Mavromatis and George Karypis. Graph infoclust: Maximizing coarse-grain mutual information in graphs. In *PAKDD*, pages 541–553, 2021.

[Mo *et al.*, 2022] Yujie Mo, Liang Peng, Jie Xu, Xiaoshuang Shi, and Xiaofeng Zhu. Simple unsupervised graph representation learning. In *AAAI*, volume 36, pages 7797–7805, 2022.

[Mo *et al.*, 2023a] Yujie Mo, Yajie Lei, Jialie Shen, Xiaoshuang Shi, Heng Tao Shen, and Xiaofeng Zhu. Disentangled multiplex graph representation learning. In *ICML*, pages 24983–25005, 2023.

[Mo *et al.*, 2023b] Yujie Mo, Zongqian Wu, Yuhuan Chen, Xiaoshuang Shi, Heng Tao Shen, and Xiaofeng Zhu. Multiplex graph representation learning via common and private information mining. In *AAAI*, volume 37, pages 9217–9225, 2023.

[Mo *et al.*, 2024a] Yujie Mo, Zhihe Lu, Runpeng Yu, Xiaofeng Zhu, and Xinchao Wang. Revisiting self-supervised heterogeneous graph learning from spectral clustering perspective. In *NeurIPS*, volume 37, pages 43133–43163, 2024.

[Mo *et al.*, 2024b] Yujie Mo, Feiping Nie, Ping Hu, Heng Tao Shen, Zheng Zhang, Xinchao Wang, and Xiaofeng Zhu. Self-supervised heterogeneous graph learning: a homophily and heterogeneity view. In *ICLR*, 2024.

[Park *et al.*, 2020] Chanyoung Park, Donghyun Kim, Jiawei Han, and Hwanjo Yu. Unsupervised attributed multiplex network embedding. In *AAAI*, volume 34, pages 5371–5378, 2020.

[Perozzi *et al.*, 2014] Bryan Perozzi, Rami Al-Rfou, and Steven Skiena. Deepwalk: Online learning of social representations. In *SIGKDD*, pages 701–710, 2014.

[Shen and Kang, 2025] Zhixiang Shen and Zhao Kang. When heterophily meets heterogeneous graphs: Latent graphs guided unsupervised representation learning. *IEEE Transactions on Neural Networks and Learning Systems*, 2025.

[Tian *et al.*, 2023] Yijun Tian, Kaiwen Dong, Chunhui Zhang, Chuxu Zhang, and Nitesh V Chawla. Heterogeneous graph masked autoencoders. In *AAAI*, volume 37, pages 9997–10005, 2023.

[Veličković *et al.*, 2019] Petar Veličković, William Fedus, William L. Hamilton, Pietro Liò, Yoshua Bengio, and R Devon Hjelm. Deep Graph Infomax. In *ICLR*, 2019.

[Veličković *et al.*, 2018] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. Graph attention networks. In *ICLR*, 2018.

[Wang *et al.*, 2019] Xiao Wang, Houye Ji, Chuan Shi, Bai Wang, Yanfang Ye, Peng Cui, and Philip S Yu. Heterogeneous graph attention network. In *WWW*, pages 2022–2032, 2019.

[Wang *et al.*, 2020] Kuansan Wang, Zhihong Shen, Chiyuan Huang, Chieh-Han Wu, Yuxiao Dong, and Anshul Kanakia. Microsoft academic graph: When experts are not enough. *Quantitative Science Studies*, 1(1):396–413, 2020.

[Wang *et al.*, 2021] Xiao Wang, Nian Liu, Hui Han, and Chuan Shi. Self-supervised heterogeneous graph neural network with co-contrastive learning. In *SIGKDD*, pages 1726–1736, 2021.

[Wang *et al.*, 2023] Zehong Wang, Qi Li, Donghua Yu, Xiaolong Han, Xiao-Zhi Gao, and Shigen Shen. Heterogeneous graph contrastive multi-view learning. In *SDM*, pages 136–144, 2023.

[Wang *et al.*, 2025] Xin Wang, Liang Peng, Rongyao Hu, Ping Hu, and Xiaofeng Zhu. Unsupervised multiplex graph representation learning via maximizing coding rate reduction. *Pattern Recognition*, 165:111557, 2025.

[Xie *et al.*, 2025] Xuanting Xie, Bingheng Li, Erlin Pan, Zhao Kang, and Wenyu Chen. Aggregation-aware mlp: An unsupervised approach for graph message-passing. *ArXiv*, abs/2507.20127, 2025.

[Yang *et al.*, 2023] Xiaocheng Yang, Mingyu Yan, Shirui Pan, Xiaochun Ye, and Dongrui Fan. Simple and efficient heterogeneous graph neural network. In *AAAI*, volume 37, pages 10816–10824, 2023.

[Yu and Gao, 2022] Zhaoning Yu and Hongyang Gao. Molecular representation learning via heterogeneous motif graph neural networks. In *ICML*, pages 25581–25594, 2022.

[Zhou *et al.*, 2003] Dengyong Zhou, Olivier Bousquet, Thomas Lal, Jason Weston, and Bernhard Schölkopf. Learning with local and global consistency. In *NeurIPS*, volume 16, 2003.

[Zhu *et al.*, 2021] Yanqiao Zhu, Yichen Xu, Feng Yu, Qiang Liu, Shu Wu, and Liang Wang. Graph contrastive learning with adaptive augmentation. In *WWW*, pages 2069–2080, 2021.

[Zhu *et al.*, 2022] Yanqiao Zhu, Yichen Xu, Hejie Cui, Carl Yang, Qiang Liu, and Shu Wu. Structure-enhanced heterogeneous graph contrastive learning. In *SDM*, pages 82–90. SIAM, 2022.