

DiffGraph: Heterogeneous Graph Diffusion Model

Zongwei Li
University of Hong Kong
Hong Kong, China
conwayli@tencent.com

Lianghao Xia
University of Hong Kong
Hong Kong, China
aka_xia@foxmail.com

Hua Hua
Tencent
Shenzhen, Guangdong, China
mikehua@tencent.com

Shijie Zhang
Guangdong Engineering Center for
Social Computing and Mental Health
Shenzhen, Guangdong, China
zhang.shijie1101@gmail.com

Shuangyang Wang
Tencent
Shenzhen, Guangdong, China
feymanwang@tencent.com

Chao Huang*
University of Hong Kong
Hong Kong, China
chaohuang75@gmail.com

ABSTRACT

Recent advances in Graph Neural Networks (GNNs) have revolutionized graph-structured data modeling, yet traditional GNNs struggle with complex heterogeneous structures prevalent in real-world scenarios. Despite progress in handling heterogeneous interactions, two fundamental challenges persist: noisy data significantly compromising embedding quality and learning performance, and existing methods' inability to capture intricate semantic transitions among heterogeneous relations, which impacts downstream predictions. To address these fundamental issues, we present the Heterogeneous Graph Diffusion Model (DiffGraph), a pioneering framework that introduces an innovative cross-view denoising strategy. This advanced approach transforms auxiliary heterogeneous data into target semantic spaces, enabling precise distillation of task-relevant information. At its core, DiffGraph features a sophisticated latent heterogeneous graph diffusion mechanism, implementing a novel forward and backward diffusion process for superior noise management. This methodology achieves simultaneous heterogeneous graph denoising and cross-type transition, while significantly simplifying graph generation through its latent-space diffusion capabilities. Through rigorous experimental validation on both public and industrial datasets, we demonstrate that DiffGraph consistently surpasses existing methods in link prediction and node classification tasks, establishing new benchmarks for robustness and efficiency in heterogeneous graph processing. The model implementation is publicly available at: <https://github.com/HKUDS/DiffGraph>.

CCS CONCEPTS

• Information systems → Data mining; • Mathematics of computing → Graph algorithms.

*Chao Huang is the Corresponding Author.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

WSDM '25, March 10–14, 2025, Hannover, Germany

© 2025 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 979-8-4007-1329-3/25/03...\$15.00
<https://doi.org/10.1145/3701551.3703590>

KEYWORDS

Graph Learning, Diffusion Model, Heterogeneous Graph

ACM Reference Format:

Zongwei Li, Lianghao Xia, Hua Hua, Shijie Zhang, Shuangyang Wang, and Chao Huang. 2025. DiffGraph: Heterogeneous Graph Diffusion Model. In *Proceedings of the Eighteenth ACM International Conference on Web Search and Data Mining (WSDM '25)*, March 10–14, 2025, Hannover, Germany. ACM, New York, NY, USA, 10 pages. <https://doi.org/10.1145/3701551.3703590>

1 INTRODUCTION

Learning with heterogeneous graphs has emerged as a pivotal paradigm in modern data science, reflecting the complexity of real-world systems. Unlike homogeneous graphs with uniform node and edge types, heterogeneous graphs encapsulate rich, multi-faceted interactions among diverse entities, enabling more expressive data representations. These advanced graph structures have demonstrated remarkable effectiveness across numerous domains, from bibliographic academic data [16, 39], medical data [5], to recommender systems [44]. The fundamental objective lies in leveraging this inherent diversity to enhance various analytical tasks. By harnessing these heterogeneous relationships, researchers have achieved significant breakthroughs in node classification [33], link prediction [8], and graph classification [31], consistently demonstrating superior performance compared to traditional homogeneous approaches.

Recent years have witnessed remarkable advancements in heterogeneous graph neural networks (HGNNs). Through relation-aware message passing frameworks, earlier study [49] has improved graph learning tasks by effectively capturing both complex structural patterns and diverse semantic information within heterogeneous graphs. Notable research works include HGT [33], which leverages graph attention mechanisms to dynamically assess the importance of various heterogeneous paths, thereby enhancing both semantic-level and node-level representation learning. Similarly, HeteGNN [49] introduces a framework that simultaneously models heterogeneous structures and their associated content features. Building upon the success of self-supervised learning (SSL) [26, 46] in addressing data scarcity and noise challenges, recent research efforts [23, 33, 40] have effectively incorporated SSL techniques to further advance heterogeneous graph learning capabilities.

Despite significant advances in heterogeneous graph learning, two critical challenges remain insufficiently addressed. **First**, current HGNN approaches demonstrate limited capability in handling

heterogeneous noise. While self-supervised learning techniques, particularly contrastive learning [40], attempt to mitigate data noise through self-supervisory signals, they lack sophisticated mechanisms for processing noisy data with complex heterogeneous semantics. These methods primarily rely on simple random data augmentations (e.g., feature masking, random walk), overlooking the crucial interactions between different relation types that could significantly impact noise distribution. **Second**, existing approaches struggle to effectively model complex transition patterns among heterogeneous relations. Although some methods employ attention mechanisms [43] or meta paths [10] to capture dependencies, these approaches are insufficient for encoding sophisticated semantic transition processes between heterogeneous relation types, potentially compromising the extraction of task-relevant information and degrading downstream prediction performance. These observations lead us to two fundamental research questions:

- How can we develop a robust learning paradigm that effectively addresses and mitigates data noise in heterogeneous graphs?
- How can we accurately capture the complex semantic transition processes across diverse relation types in heterogeneous graphs?

To address these challenges, we propose a dual-component solution. **First**, we introduce an adaptive parametric function that systematically filters noisy structures while preserving task-critical information for downstream predictions. This function is specifically designed to encode heterogeneous semantics and structural patterns, enabling effective characterization of complex noise distributions. **Second**, we develop an advanced semantic transition model to capture intricate relationships across different relation types, employing a fine-grained approach to accommodate diverse heterogeneous semantic patterns and their evolutionary processes.

Inspired by the remarkable capabilities of diffusion models in capturing complex data generation processes, we propose a novel generative diffusion framework for heterogeneous graphs. Our approach implements a bi-directional process: a forward phase for systematic noise addition and a backward phase for noise removal, jointly enhancing model robustness against heterogeneous data noise. Extending beyond conventional diffusion models, our method uniquely facilitates the transfer between auxiliary and target graph views through multiple fine-grained diffusion steps, enabling precise modeling of semantic transitions at a granular level.

Following these ideas, we present a Heterogeneous Graph Diffusion Model (DiffGraph). Our approach begins by identifying the target subgraph containing node and edge types most crucial to the prediction task, while designating the remaining structure as the auxiliary graph for feature enhancement. A specialized heterogeneous graph encoder, built on graph convolutions, projects both target and auxiliary graphs into a latent representation space. We then introduce a novel latent heterogeneous graph diffusion module that orchestrates forward and backward diffusion processes on these encoded representations, with auxiliary graph data serving as the diffusion source and target graph data providing denoising training signals. Our proposed diffusion framework accomplishes two key objectives: capturing complex noise distributions in heterogeneous graph data and modeling semantic transitions between diverse relation types. By conducting noise addition and removal

operations in latent space rather than the original graph space, DiffGraph effectively addresses the challenges of generating sparse and discrete heterogeneous graph data. This design choice significantly enhances the model’s capability for unbiased heterogeneous graph modeling, overcoming traditional limitations in graph generation. The main contributions of this paper can be summarized as follows:

- We propose the DiffGraph, a novel approach that enhances model performance by systematically filtering non-essential semantic information from heterogeneous graph structures.
- We develop an innovative hidden-space diffusion mechanism that effectively removes noisy information through a sophisticated multi-step process of controlled noise addition and denoising within heterogeneous embeddings.
- We present extensive experimental evaluations of DiffGraph across diverse datasets, demonstrating its superior effectiveness in both link prediction and node classification tasks.

2 METHODOLOGY

This section outlines the details of the proposed DiffGraph framework. The architecture of DiffGraph is illustrated in Figure 1.

2.1 Heterogeneous Graph Learning

Heterogeneous Graphs. In a heterogeneous graph, the node set is denoted as $\mathcal{V} = \{v_i\}$, with the set of node types denoted as \mathbb{V} . There exists a node type mapping function $\phi : \mathcal{V} \rightarrow \mathbb{V}$, where $\phi(v_i)$ represents the type of node v_i . Similarly, the set of edges is denoted as $\mathcal{E} = \{(v_i, v_j)\}$, with an edge type mapping function $\psi : \mathcal{E} \rightarrow \mathbb{E}$. Here, \mathbb{E} denotes the set of edge types, and $\psi(v_i, v_j)$ represents the type of edge (v_i, v_j) . With these notations, we formally define the heterogeneous graph as $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \phi, \psi)$. For convenience, a heterogeneous graph is recorded by the binary adjacency matrix \mathbf{A} of size $|\mathcal{V}| \times |\mathcal{V}| \times |\mathbb{E}|$. An entry $a_{i,j}^r \in \mathbf{A}$ equals 1 if there is a link of type $r \in \mathbb{E}$ between nodes v_i and v_j , otherwise $a_{i,j}^r = 0$.

Heterogeneous Graph Prediction. A predictive model f for heterogeneous graphs can be divided into an encoding phase and a prediction phase, formally as: $f(\mathcal{G}) = \text{Pred} \circ \text{Enc}(\mathcal{G})$. The encoding phase $\text{Enc}(\cdot)$ learns d -dimensional hidden representations $\mathbf{E} \in \mathbb{R}^{|\mathcal{V}| \times d}$ for nodes $v_i \in \mathcal{V}$, while the prediction phase $\text{Pred}(\cdot)$ generates task-specific predictions based on the learned embeddings. Typical graph forecasting tasks include node classification and link prediction. For node classification, the prediction network $\text{Pred}(\cdot)$ of DiffGraph employs a multilayer perceptron (MLP) to infer nodes’ classes from their learned embeddings. For link prediction, the dot-product operator is utilized to infer the existence of edges based on the embeddings of the connected nodes in $\text{Pred}(\cdot)$.

Motivated by the simplicity and effectiveness of Graph Convolutional Networks (GCNs) [12], our DiffGraph adopts a GCN-based encoding process for $\text{Enc}(\cdot)$, which is formally defined as follows:

$$\begin{aligned} \mathbf{E} &= \text{Pooling}(\{\mathbf{E}_r | r \in \mathbb{E}\}), \quad \mathbf{E}_r = \sum_{l=0}^L \mathbf{E}_{r,l} \\ \mathbf{E}_{r,l} &= \text{norm}(\delta(\mathbf{D}^{-1/2} \mathbf{A}^r \mathbf{D}^{-1/2} \mathbf{E}_{r,l-1})) \end{aligned} \quad (1)$$

where $\mathbf{A}^r \in \mathbb{R}^{|\mathcal{V}| \times |\mathcal{V}|}$ denotes the specific adjacency matrix for each heterogeneous relation $r \in \mathbb{E}$, and \mathbf{D} denotes its corresponding diagonal degree matrix. $\mathbf{E}_{r,l} \in \mathbb{R}^{|\mathcal{V}| \times d}$ represents the embedding

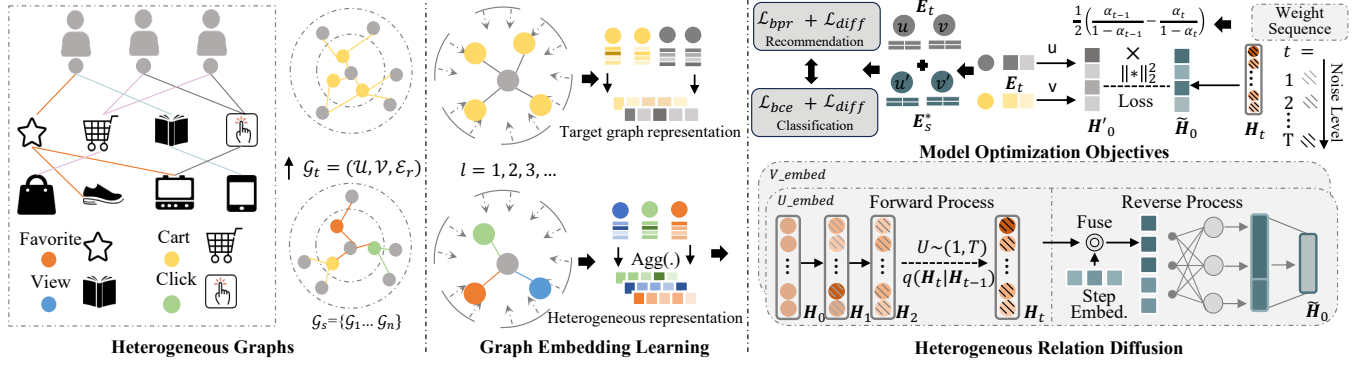


Figure 1: Overall architecture of the proposed DiffGraph framework.

matrix at the l -th GCN iteration. For datasets without node features, the initial embeddings $E_{r,0}$ are randomly initialized learnable parameters. For datasets with node features, $E_{r,0}$ utilizes the initial node features. During each iteration, the existing embeddings are disseminated to adjacent nodes through the Laplacian-normalized adjacency matrix, utilizing the activation function $\delta(\cdot)$ and the l_2 embedding normalization function $\text{norm}(\cdot)$. Following L iterations, the final node embeddings E_r are generated from an element-wise aggregation of multi-order node embeddings. And the final node embeddings E is obtained by aggregate node embeddings E_r of different relation types with a pooling function.

2.2 Cross-view Heterogeneous Graph Denoising

2.2.1 Denoising Auxiliary Graphs using Target Graph. To enhance the performance of heterogeneous graph prediction, we propose a cross-view denoising approach to maximally extract task-relevant information from heterogeneous graphs. Initially, we identify the task-relevant subgraph within the compound heterogeneous graph as the target graph \mathcal{G}_t . An example of such a target graph is the purchase behavior graph between users and items in an e-commerce recommender system. Once this target graph \mathcal{G}_t is removed from the entire graph \mathcal{G} , the remaining graph is referred to as the auxiliary graph \mathcal{G}_s (also known as the source graph).

Our cross-view denoising framework for heterogeneous graphs aims to maximize the mutual information between the auxiliary view and the target view by learning a parametric function g that maps the auxiliary graph space to the task-relevant data space. This learnable denoising function g takes the source graph data \mathcal{G}_s as the input and utilizes the target graph data \mathcal{G}_t as labels. Specifically, this denoising framework can be formulated as follows:

$$\arg \min_{\theta_f, \theta_g} \mathcal{L}_{main}(\mathcal{Y}, f(\mathcal{G})) + \lambda \cdot \mathcal{L}_{deno}(\mathcal{G}_t, g(\mathcal{G}_s)) \quad (2)$$

where \mathcal{L}_{main} represents the loss for the main prediction task, such as BPR loss [29] for link prediction and cross-entropy loss for node classification. Here, \mathcal{Y} denotes the task-specific labels. Meanwhile, \mathcal{L}_{deno} denotes the learning objective for our denoising function g , which measures the differences between the task-relevant target graph \mathcal{G}_t and the denoising output of function g . The learning process tunes the parameter sets θ_f and θ_g , which are parameters for the forecasting model f and the denoising network g , respectively. A weighting parameter λ is used to balance the denoising task.

2.2.2 Latent Heterogeneous Graph Diffusion. Inspired by the impressive performance of diffusion models in capturing complex data generation processes across diverse data types [4], we propose to instantiate our denoising function as a heterogeneous graph diffusion model. This model aims to distill task-relevant information from the auxiliary graph by iteratively refining the graph data. Given the complexity of graph-structured data, we propose conducting this diffusion process in the latent representation space of the heterogeneous graph data. Specifically, our diffusion model aims to achieve the following step-by-step transformation:

$$\mathcal{G}_s^* \stackrel{\pi}{\rightleftharpoons} \mathbf{E}_s^* \xrightarrow{\varphi} \tilde{\mathbf{E}}_s^* \xrightarrow{\varphi'} \hat{\mathbf{E}}_s^* \stackrel{\pi'}{\rightleftharpoons} \hat{\mathcal{G}}_s^*, \quad \mathcal{G}_s^* = \{\mathcal{G}_s^1, \dots, \mathcal{G}_s^n\} \quad (3)$$

where π and π' denote the bidirectional mapping between the graph-structured data and the latent representations. By utilizing the heterogeneous GCN-based encoder in the prediction function f , \mathbf{E}_s^* captures heterogeneous semantic information from the source view. Function φ represents the forward diffusion process, which incrementally adds noise to \mathbf{E}_s^* . Correspondingly, φ' refers to the reverse diffusion process, which removes the noise step by step.

Rationale for Latent Diffusion over Graph Diffusion. In the context of graph-based task learning, the role of the graph encoder is to aggregate the embeddings of neighboring nodes, transforming the structural semantics of the graph into low-dimensional information within the latent space. This characteristic suggests that by learning the denoising process φ' within the latent space, our hidden-space diffusion model can effectively filter out the imprecise semantic differences within heterogeneous graphs, thereby better supporting the learning of the target graph.

Moreover, graph data exhibits several challenging characteristics that hinder effective diffusion, including its sparse and discrete nature, as well as its large data space of exponential complexity. In contrast, the latent space is dense and continuous, and as a compressed representation of the graph, it usually has a smaller size. These key advantages motivate us to conduct graph diffusion in the latent representation space, instead of in the graph space.

2.2.3 Forward and Reverse Diffusion. Forward Process. The forward process of our diffusion model consists of T steps of gradual noise addition. This process begins with the encoded hidden representations, denoted as $\mathbf{H}_0 = \mathbf{E}_s^*$. It then iteratively increases

the noise ratio following the recursive formula below:

$$q(\mathcal{H}_t|\mathcal{H}_{t-1}) = \mathcal{N}(\mathcal{H}_t; \sqrt{1 - \beta_t}\mathcal{H}_{t-1}, \beta_t \mathbf{I}) \quad (4)$$

where β_t controls the noise scale. \mathcal{N} denotes the Gaussian distribution used to generate the noise, and the noise data \mathbf{H}_t will progressively increase as β_t increases, ultimately converging to pure Gaussian noise [15]. This property allows our noise resolution methodology to cover a wide range of noise intensities within the training data. The additive nature of the Gaussian distribution enables us to directly compute the data at step t using \mathbf{H}_0 and pre-computed values related to the β_t sequence. Specifically, we can pre-calculate the following values to expedite the diffusion process:

$$\alpha_t = 1 - \beta_t, \quad \bar{\alpha}_t = \prod_{t'=1}^t \alpha_{t'}, \quad \beta_t = 1 - b_t/b_{t-1} \quad (5)$$

To determine β_t at each step, we introduce two hyperparameters, \bar{b}_{max} and \bar{b}_{min} . These parameters generate a linear interpolation sequence $b = (1, \bar{b}_{max}, \dots, \bar{b}_{min})$. Using the calculations from Eq 5, we derive the following results:

$$\begin{aligned} \mathcal{H}_t &= \sqrt{\bar{\alpha}_t}\mathcal{H}_{t-1} + \sqrt{1 - \bar{\alpha}_t}\xi_1 \\ &\Rightarrow \sqrt{\bar{\alpha}_t}(\sqrt{\bar{\alpha}_{t-1}}\mathcal{H}_{t-2} + \sqrt{1 - \bar{\alpha}_{t-1}}\xi_2) + \sqrt{1 - \bar{\alpha}_t}\xi_1 \\ &\Rightarrow \sqrt{\bar{\alpha}_t}\mathcal{H}_0 + \sqrt{1 - \bar{\alpha}_t}\xi'_t, \quad \xi \mapsto \mathcal{N}(0, \mathbf{I}) \end{aligned} \quad (6)$$

where ξ_* represents independent Gaussian distributions. By pre-calculating α_t , we can leverage the additive properties to infer \mathcal{H}_t efficiently, bypassing the need for recursive computations.

Reverse Process. The purpose of the reverse process is to enable DiffGraph to learn the denoising ability of noisy heterogeneous semantic data. Specifically, we aim to reconstruct the heterogeneous relations in the hidden space from the noisy data \mathcal{H}_t . A learnable neural network is employed to estimate the generation of conditional probabilities for the following:

$$p(\mathcal{H}_{t-1}|\mathcal{H}_t) = \mathcal{N}(\mathcal{H}_{t-1}; \mu_\theta(\mathcal{H}_t, t), \Sigma_\theta(\mathcal{H}_t, t)) \quad (7)$$

where $\mu_\theta(\cdot)$ and $\Sigma_\theta(\cdot)$ are obtained from a two-layer neural network with learning parameter θ , which is intended to be used for estimating Gaussian distributions. The framework is defined:

$$\mu_\theta(\mathbf{h}_t, t) = \text{MLP}^2(\mathbf{h}_t || \mathbf{s}_t), \quad \text{MLP}(\mathbf{x}) = \sigma(\mathbf{W}\mathbf{x} + \mathbf{b}) \quad (8)$$

where $\text{MLP}^2(\cdot)$ represents two continuous MLP layers. We concatenate the t -step hidden vector \mathbf{h}_t with a time step-specific embedding \mathbf{s}_t input data x . σ , \mathbf{W} and \mathbf{b} denotes the activation function, linear transformation and bias for the neural network respectively.

2.3 Model Training for DiffGraph

2.3.1 Diffusion Loss Function. To optimize the hidden-space diffusion module, we adopt the embedding \mathbf{E}_t of the target graph as the evidence lower bound (ELBO). Based on the data specific to each side, $\mathbf{h}'_0 = \mathbf{E}_u, \mathbf{E}_v \in \mathbf{E}_t$, the paradigm for maximizing the

ELBO can be summarized as follows:

$$\begin{aligned} \log p(\mathbf{h}'_0) &\geq \underbrace{\mathbb{E}_{q(\mathbf{h}_1|\mathbf{h}_0)} \left[\log p_\theta(\mathbf{h}'_0|\mathbf{h}_1) \right]}_{\text{(reconstruction term)}} \\ &\quad - \underbrace{\sum_{t=2}^T \mathbb{E}_{q(\mathbf{h}_t|\mathbf{h}_0)} [D_{\text{KL}}(q(\mathbf{h}_{t-1}|\mathbf{h}_t, \mathbf{h}_0) \| p_\theta(\mathbf{h}_{t-1}|\mathbf{h}_t))]}_{\text{(denoising comparing term)}} \end{aligned} \quad (9)$$

The model incorporates two optimization terms. The objective of the denoising comparing term is to align the true distribution $q(\mathbf{h}_{t-1}|\mathbf{h}_t, \mathbf{h}_0)$ more approximate our heterogeneous semantics denoiser $p_\theta(\mathbf{h}_{t-1}|\mathbf{h}_t)$, thereby minimizing the KL divergence. Drawing on the conclusions of previous works [15, 37], we can simplify the study of the standard deviation as $\Sigma_\theta(\mathbf{h}_t, t) = \sigma^2(t)\mathbf{I}$. The \mathcal{L}_t for denoising comparing term can be transformed as follow:

$$\mathcal{L}_t = \mathbb{E}_{q(\mathbf{h}_t|\mathbf{h}_0)} \left[\frac{1}{2} \left(\frac{\bar{\alpha}_{t-1}}{1 - \bar{\alpha}_{t-1}} - \frac{\bar{\alpha}_t}{1 - \bar{\alpha}_t} \right) \|\hat{\mathbf{h}}_\theta(\mathbf{h}_t, t) - \mathbf{h}'_0\|_2^2 \right] \quad (10)$$

where $\hat{\mathbf{h}}_\theta(\mathbf{h}_t, t)$ denotes the data \mathbf{h}_0 predicted by \mathbf{h}_t and t via a Multi-Layer Perceptron (MLP), which indicated in Eq 8. From the inference of Eq8, the reconstruction term can also be approximated as the squared loss between the optimization $\hat{\mathbf{h}}_\theta(\mathbf{h}_t, t)$ and the target graph vector \mathbf{h}'_0 . We define it as:

$$\mathcal{L}_1 = \mathbb{E}_{q(\mathbf{h}_1|\mathbf{h}_0)} \left[\|\hat{\mathbf{e}}_\theta(\mathbf{h}_1, 1) - \mathbf{h}'_0\|_2^2 \right] \quad (11)$$

It is noteworthy that the semantic information in the latent space of the source heterogeneous graphs serves as the input for the diffusion module, and the denoised embedding is optimized by forming a control group with the semantic vector \mathbf{h}'_0 of the target graph. In determining the number of steps t , we employ a uniform sampling strategy. Formally, the diffusion loss \mathcal{L}_{deno} is as follows:

$$\mathcal{L}_{deno} = \mathbb{E}_{t \sim \mathcal{N}(1, t)} \mathcal{L}_t \quad (12)$$

2.3.2 Prediction and Optimization. For the link prediction task, we combine the denoised heterogeneous semantic relations of the source graphs with the semantic relations of the target graph to get the final embedding to make predictions. And we employ the BPR loss function to optimize predictions \hat{y} :

$$\begin{aligned} \hat{y}_{i,j} &= \tilde{\mathbf{e}}_i^\top \tilde{\mathbf{e}}_j, \quad \tilde{\mathbf{E}} = \mathbf{E}^t + \tilde{\mathbf{E}}^*, \\ \mathcal{L}_{main} &= \sum_{u, v^+, v^-} -\log \text{sigm}(\hat{y}_{u, v^+} - \hat{y}_{u, v^-}) \end{aligned} \quad (13)$$

where u, v^+, v^- stands for the positive and negative sample sampling strategy [29]. For the node classification task, we utilize the following cross entropy loss function:

$$p_{i,c} = \text{softmax}(\text{MLP}(\tilde{\mathbf{e}}_i))_c, \quad \mathcal{L}_{main} = - \sum_{v_i \in \mathcal{V}} \log p_{i, y_i} \quad (14)$$

2.3.3 Model Complexity Analysis. This section gives a thorough analysis on the time and space complexity of our DiffGraph. **Time Complexity.** Initially, DiffGraph performs graph-level information propagation on both the target collaborative graph \mathcal{G}_t and the auxiliary heterogeneous graphs $\mathcal{G}_* = \{\mathcal{G}_1, \dots, \mathcal{G}_n, \dots, \mathcal{G}_N\}$. This standard graph convolutional process requires $\mathcal{O}((|\mathcal{E}_r| + \sum_{l=1}^N |\mathcal{E}_n|) \times d)$ calculations for message passing and $\mathcal{O}((|\mathcal{U}| + |\mathcal{V}|) \times d^2)$ for embedding transformation. For recommendation task, our diffusion

Table 1: Statistics of experimental datasets.

Dataset	User #	Item #	Link #	Interaction Types
Tmall	31882	31232	1,451,29	View, Favorite, Cart, Purchase
Retail Rocket	2174	30113	97,381	View, Cart, Transaction
IJCAI	17435	35920	799,368	View, Favorite, Cart, Purchase
Industry	1M	361	23,890,445	Purchase, Friend, Complete Task

	Node	Metapath		Node	Metapath
DBLP	Author:4057	APA	AMiner	Paper:6564	PAP
	Paper:14328	APCPA		Author:13329	PRP
	Conference:20	APTPA		Reference:35890	POS
	Term:7723				

process operates on the user side and item side at each training epoch respectively. The forward diffusion process costs $O(|\mathcal{U}| \times d)$ and $O(|\mathcal{V}| \times d)$ computations, while the reverse process costs $O(|\mathcal{U}| \times (d^2 + dd'))$ and $O(|\mathcal{V}| \times (d^2 + dd'))$. For node classification task, the diffusion process only performs on user-tie nodes, which costs $O(|\mathcal{U}| \times d)$ for forward diffusion process and $O(|\mathcal{U}| \times (d^2 + dd'))$ for reverse process. In conclusion, DiffGraph achieves comparable time costs to common heterogeneous recommenders. **Memory Complexity.** The graph encoding process of our DiffGraph model requires a similar number of parameters as conventional graph-based heterogeneous models. The hidden-space diffusion network employs $O(d^2 + dd')$ parameters for the denoiser.

3 EVALUATION

We evaluate the performance of our DiffGraph framework by studying the following Research Questions (RQs):

- **RQ1:** How does the proposed DiffGraph framework perform on link prediction, node classification, and industry datasets?
- **RQ2:** How effective are the designed modules in DiffGraph?
- **RQ3:** How do different settings of key hyperparameters impact the graph prediction accuracy of our DiffGraph method?
- **RQ4:** How is the efficiency of DiffGraph compared to baselines?
- **RQ5:** How effectively can our DiffGraph approach perform to alleviate the issue of data sparsity for graph data?
- **RQ6:** How well can DiffGraph handle noisy heterogeneous graphs?
- **RQ7:** Can DiffGraph provide explanations in specific cases?

3.1 Experimental Settings

3.1.1 Datasets. We evaluate DiffGraph on link prediction and node classification tasks. For link prediction, we utilize **Tmall**, **Retailrocket**, and **IJCAI** datasets. For node classification, we employ **DBLP** and **AMiner** (focusing on academic networks), along with an **Industry** dataset from a gaming platform. Detailed statistics are presented in Table 1, with dataset descriptions below.

Link Prediction Dataset: • **Tmall:** An E-commerce dataset with user views, favorites, cart additions, and purchases, filtered to users with ≥ 3 purchases [43]. • **Retailrocket:** A dataset with user page views, cart additions, and transactions, where purchases are primary and other interactions auxiliary [44]. • **IJCAI:** A dataset from IJCAI15 with heterogeneous user behaviors similar to Tmall.

Node Classification Dataset: • **Industry:** A game platform dataset with retention as target behavior and three auxiliary behaviors (purchases, friendships, tasks), forming user-item, user-user, and user-task graphs. • **DBLP:** A dataset subset of authors categorized into

four research areas (Database, Data Mining, AI, and Information Retrieval) for node classification. • **AMiner:** A dataset containing papers in four categories, utilizing meta-path relationships (author, paper, conference, reference) for heterogeneous node classification.

3.1.2 Baseline Methods. For a thorough evaluation of DiffGraph, we include the following 21 baselines from different research lines.

Recommendation Models with Heterogeneous Relations:

- **NMTR** [11]: It uses a multi-task learning framework with redefined cascading relationships to explore diverse user behaviors.
- **MATN** [43]: This method uses a memory-augmented attention mechanism for the modeling of multiple user behaviors.
- **MBGCN** [18]: It utilizes a GCN-based model to analyze multiple patterns of behavior within the user-item interaction graph.

Heterogeneous Graph Neural Networks:

- **Mp2vec** [6]: This approach learns embedding vectors for meta path in heterogeneous networks.
- **HERec** [30]: This method models recommendation and its auxiliary information as a heterogeneous information network.
- **HetGNN** [49]: It combines a two-module architecture and random walk sampling process for heterogeneous graph learning.
- **HGT** [17]: This is a graph transformer architecture for encoding diverse relationships in heterogeneous graphs.
- **HAN** [39]: This method uses a hierarchical attention mechanism to aggregate information from nodes of multiple meta-paths.
- **DMGI** [27]: This heterogeneous graph learning method maximizes the mutual information between local and global views.
- **HeCo** [40]: It proposes a cross-view contrastive learning approach between two heterogeneous information networks.

Homogeneous Graph Convolutional Networks:

- **GraphSage** [13]: This model accomplishes graph node aggregation using a node-centric neighbor sampling approach.
- **GCN** [20]: It aggregates the neighboring nodes embedded after linear projection and finally performs the average computation.
- **GAE** [19]: This is a variational auto-encoding approach for unsupervised graph representation learning.
- **GAT** [34]: This is a graph attention network that differentiates weights of different neighbors for the central nodes.
- **LightGCN** [14]: It employs a lightweight graph convolutional network architecture for the effective aggregation.
- **DGI** [35]: This method learns node embeddings by maximizing the mutual information between local and global graph views.
- **GraphMAE** [35]: It applies the masked auto-encoding training objective to learn node embeddings for downstream tasks.
- **PinSAGE** [45]: This approach facilitate efficient graph neural networks for web-scale recommendataion.
- **NGCF** [38]: This is one of the early work that applies graph neural networks to modeling recommendation graphs.

Non-Graph Neural Networks:

- **BPR** [29]: This baseline enhances matrix factorization by utilizing the Bayesian personalized ranking loss.
- **DNN:** This is a deep neural network for node classification.

3.1.3 Evaluation Protocols. The evaluation protocols for our DiffGraph and the baselines in our experiments are as follows:

- For link prediction, we used leave-one-out strategy by generating the test set from users' last interacted items under the target behavior. The evaluation metrics used were Recall@20 and NDCG@20, widely adopted in ranking-based evaluations.
- For node classification using public datasets DBLP and Aminer, we use 20, 40 and 60 labeled nodes for each class as the training set and 1000 nodes as the validation set and test set, respectively. In our experiments, we used the widely used metrics MicroF1, Macro-F1 and AUC for evaluation.
- For the node classification task on the Industry dataset, it adopts a more realistic data split scheme, in which the training set and the test set are made of users and interactions of different time periods. Both the training set and the test set contain 1 million user nodes, respectively, and the evaluation metric is AUC with application to the prediction of user retention.

3.1.4 Hyperparameter Settings. The experiments were conducted on a device with an NVIDIA TITAN RTX GPU and an Intel Xeon W-2133 CPU. The hyperparameter settings are as follows:

- **Link Prediction Task:** We adjusted the learning rate among $\{1e-2, 1e-3, 1e-4\}$. For graph models, the depth of graph propagation layers was tuned within $\{1, 2, 3, 4\}$. We use l_2 regularization when training, with the strength coefficient being tuned across $\{1e-1, 3e-2, 1e-2, 1e-3\}$. For diffusion-related parameters, we varied the number of noise steps from 0 to 250, and the noise scale was adjusted among $\{1e-3, 1e-4, 1e-5\}$. The dimensionality of embeddings is tuned from the range $\{8, 16, 32, 64\}$, and the batch size is selected between 512 and 4096.
- **Node Classification Task:** For the public datasets DBLP and Aminer, we follow the settings of HeCo. We set the hidden embedding size to 64 and searched for the learning rate from $1e-4$ to $5e-3$. We varied the number of noise steps from 0 to 250, and the noise scale was tuned among $\{1e-4, 5e-5, 1e-5, 5e-6, 1e-6\}$.

3.2 Overall Performance Comparison (RQ1)

This section examines whether DiffGraph outperforms existing baselines in different graph tasks. The results are presented in Table 2. From the outcomes, we draw the following conclusions:

- **Superior performance of DiffGraph.** Across both link prediction and node classification, DiffGraph consistently outperforms all baselines, demonstrating superior performance. These results underscore the advanced graph forecasting capabilities of the DiffGraph framework. Additionally, the performance advantages of DiffGraph on the industry dataset further validate its effectiveness in real-world applications. We attribute this success to the model's ability to eliminate inaccurate and spurious semantics from heterogeneous graph structures through the latent heterogeneous graph diffusion paradigm. This semantic refactoring process enables the model to learn more precise and meaningful heterogeneous relations, allowing DiffGraph to effectively mine valuable heterogeneous information and enhance specific tasks.
- **Effectiveness of heterogeneous relationships.** The results indicate that methods based on heterogeneous graph learning [17,

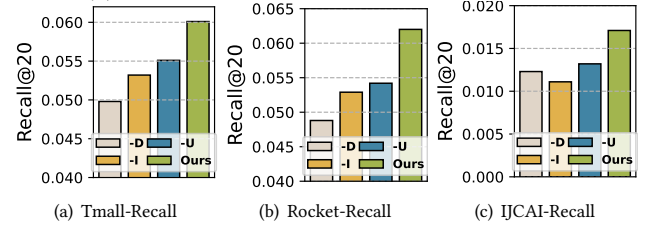


Figure 2: Ablation study for modules in DiffGraph.

18] generally outperform those using traditional homogeneous graph neural networks [20, 34, 38]. This advantage highlights the positive role of auxiliary heterogeneous information in graph forecasting tasks. By referring to more information with heterogeneous semantics, graph models can capture richer structural information to improve the prediction accuracy. This advantage is evident in both link prediction and node classification.

- **Effectiveness of hidden-diffusion-based denoising.** The performance of DiffGraph surpasses that of state-of-the-art heterogeneous graph learning networks, a superiority attributed to the efficacy of our hidden-space diffusion paradigm in denoising and accurately reconstructing complex heterogeneous graph relationships. The denoised source graph heterogeneous semantics serve as a valuable aid in modeling the target graph relations, enabling stronger optimization for the final prediction task.

3.3 Ablation Study (RQ2)

To study the impact of DiffGraph's sub-modules, we remove or replace essential modules and evaluate the resulting performance. These ablated models include: **i) -D:** Omits the holistic diffusion module. **ii) -U:** Disregards user-side information of recommendation data. **iii) -I:** Similar to -U, eliminates item-side information. **iv) -H:** Excludes auxiliary heterogeneous graphs when training. **v) DAE:** Replace the diffusion module with a denoising autoencoder.

The evaluation results for link prediction is listed in Figure 2. We also conduct ablation study for node classification on the Industry data. The results in terms of AUC, is as follows: **-H**, 0.7840, **-D**, 0.7901, **DAE**, 0.7911, **DiffGraph**, 0.8025. Through meticulous examination, we make the following noteworthy observations:

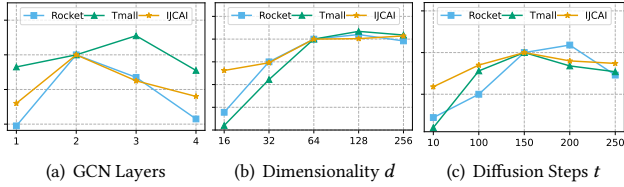
- Removing the diffusion module "-D" leads to performance degradation, highlighting both the adverse effects of noise in auxiliary heterogeneous data and demonstrating the effectiveness of our latent feature-level diffusion model in its denoising function.
- Comparing "-U" to DiffGraph demonstrates the value of heterogeneous user aggregation. Similarly, "-I" and "-H" variants confirm the benefits of heterogeneous information. Notably, "-D" outperforming "-I" suggests that removing item-side diffusion both loses valuable signals and introduces noise.
- The "DAE" variant shows only limited performance improvement over completely removing the denoising module (*i.e.*, "-D"). This demonstrates the difficulty of denoising heterogeneous graph data using existing DAE approaches. Consequently, it validates the superiority of the stepwise noise addition and removal process in our latent heterogeneous graph diffusion module.

Table 2: Overall performance comparison on the link prediction and node classification tasks.

Link Prediction Performance on Public Datasets																				
Data	BPR		Pinsage		NGCF		NMTR		MBGCN		HGT		MATN		DiffGraph		Improv		p-val	
	R@20	N@20	R@20	N@20	R@20	N@20	R@20	N@20	R@20	N@20	R@20	N@20	R@20	N@20	R@20	N@20	R@20	N@20	R@20	N@20
Tmall	0.0248	0.0131	0.0368	0.0156	0.0399	0.0169	0.0441	0.0192	0.0419	0.0179	0.0431	0.0192	0.0463	0.0197	0.0589	0.0274	27.21%	39.09%	9.8-10	1.8-11
Rocket	0.0308	0.0237	0.0423	0.0248	0.0405	0.0257	0.0460	0.0265	0.0492	0.0258	0.0413	0.0250	0.0524	0.0302	0.0620	0.0367	18.32%	21.52%	1.65-9	3.5-9
IJCAI	0.0051	0.0037	0.0101	0.0041	0.0091	0.0035	0.0108	0.0048	0.0112	0.0045	0.0126	0.0051	0.0136	0.0054	0.0171	0.0063	25.74%	16.67%	3.1-11	2.0-4

Node Classification Performance on Industrial Dataset, in terms of AUC										
Data	DNN	GraphSage	GCN	GAT	LightGCN	HAN	HGT	HeCo	HGMAE	DiffGraph
Ind.	0.7778	0.7836	0.7912	0.7871	0.7904	0.7909	0.7982	0.7915	0.7831	0.8025

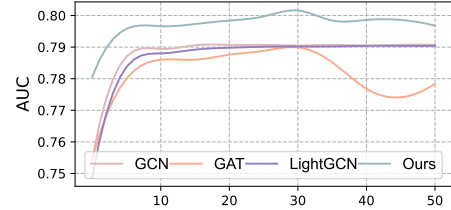
Node Classification performance on public datasets													
Dataset	Metric	GraphSage	GAE	Mp2vec	HERec	HetGNN	HAN	DGI	DMGI	HeCo	GraphMAE	DiffGraph	
DBLP	Micro-F1	20	71.44±8.7	91.55±0.1	89.67±0.1	90.24±0.4	90.11±1.0	90.16±0.9	88.72±2.6	90.78±0.3	91.97±0.2	93.30±0.4	
		40	73.61±8.6	90.00±0.3	89.14±0.2	90.15±0.4	89.03±0.7	89.47±0.9	89.22±0.5	89.92±0.4	90.76±0.3	87.80±0.5	93.05±0.3
		60	74.05±8.3	90.95±0.2	89.14±0.2	91.01±0.3	90.43±0.6	90.34±0.8	90.35±0.8	90.66±0.5	91.59±0.2	89.82±0.4	93.81±0.3
	Macro-F1	20	71.97±8.4	90.90±0.1	88.98±0.2	89.57±0.4	89.51±1.1	89.31±0.9	87.93±2.4	89.94±0.4	91.28±0.2	87.94±0.7	93.03±0.4
		40	73.69±8.4	89.60±0.3	88.68±0.2	89.73±0.4	88.61±0.8	88.87±1.0	88.62±0.6	89.25±0.4	90.34±0.3	86.85±0.7	92.81±0.3
		60	73.86±8.1	90.08±0.2	90.25±0.1	90.18±0.3	89.56±0.5	89.20±0.8	89.19±0.9	89.46±0.6	90.64±0.3	88.07±0.6	93.16±0.3
	Auc	20	90.59±4.3	98.15±0.1	97.69±0.0	98.21±0.2	97.96±0.4	98.07±0.6	96.99±1.4	97.75±0.3	98.32±0.1	92.23±3.0	99.20±0.1
		40	91.42±4.0	97.85±0.1	97.08±0.0	97.93±0.1	97.70±0.3	97.48±0.6	97.12±0.4	97.23±0.2	98.06±0.1	91.76±2.5	98.84±0.1
		60	91.73±3.8	98.37±0.1	98.00±0.0	98.49±0.1	97.97±0.2	97.96±0.5	97.76±0.5	97.72±0.4	98.59±0.1	91.63±2.5	99.21±0.1
AMiner	Micro-F1	20	49.68±3.1	65.78±2.9	60.82±0.4	63.64±1.1	61.49±2.5	68.86±4.6	62.39±3.9	63.93±3.3	78.81±1.3	68.21±0.3	80.55±0.6
		40	52.10±2.2	71.34±1.8	69.66±0.6	71.57±0.7	68.47±2.2	76.89±1.6	63.87±2.9	63.60±2.5	80.53±0.7	74.23±0.2	83.29±1.3
		60	51.36±2.2	67.70±1.9	63.92±0.5	69.76±0.8	65.61±2.2	74.73±1.4	63.10±3.0	62.51±2.6	82.46±1.4	72.28±0.2	82.10±1.0
	Macro-F1	20	42.46±2.5	60.22±2.0	54.78±0.5	58.32±1.1	50.06±0.9	56.07±3.2	51.61±3.2	59.50±2.1	71.38±1.1	62.64±0.2	71.98±1.0
		40	45.77±1.5	65.66±1.5	64.77±0.5	64.50±0.7	58.97±0.9	63.85±1.5	54.72±2.6	61.92±2.1	73.75±0.5	68.17±0.2	75.57±1.2
		60	44.91±2.0	63.74±1.6	60.65±0.3	65.53±0.7	57.34±1.4	62.02±1.2	55.45±2.4	61.15±2.5	75.80±1.8	68.21±0.2	74.57±0.7
	Auc	20	70.86±2.5	85.39±1.0	81.22±0.3	83.35±0.5	77.96±1.4	78.92±2.3	75.89±2.2	85.34±0.9	90.82±0.6	86.29±4.1	90.19±0.7
		40	74.44±1.3	88.29±1.0	88.82±0.2	88.70±0.4	83.14±1.6	80.72±2.1	77.86±2.1	88.02±1.3	92.11±0.6	89.98±0.0	94.41±0.8
		60	74.16±1.3	86.92±0.8	85.57±0.2	87.74±0.5	84.77±0.9	80.39±1.5	77.21±1.4	86.20±1.7	92.40±0.7	88.32±0.0	94.25±0.9

**Figure 3: Hyperparameter study in terms of Recall@20.**

3.4 Impact of Hyperparameters (RQ3)

This section examines the impact of hyperparameter settings. The results are presented in Figure 3. We make the following analysis.

- **Graph propagation iterations L .** Tested within 1, 2, 3, 4. Performance improves up to $L = 3$; beyond this, additional layers may introduce noise and result in over-smoothing issues [3, 43].
- **Embedding dimensionality d .** Angles tested were 16, 32, 64, 128, 256. As d increases, there's a general improvement in performance, peaking before $d = 256$, where performance slightly declines due to potential overfitting.
- **Maximum diffusion steps T .** Varied from 10 to 250. Increasing the diffusion steps T typically enhances the performance for DiffGraph, but very high values degrade it, likely due to the excessive noise disrupting social information integrity.
- **Noise scale S .** Evaluated at scales $1e-3$, $1e-4$, $1e-5$, $1e-6$ with respective AUC scores of 0.7966, 0.8025, 0.7988, 0.7974 on the Industry dataset. Optimal noise scales improve model performance by enhancing denoising effectiveness; however, excessively high scales diminish performance by obscuring key graph semantics.

**Figure 4: Test performance v.s. training epochs.**

3.5 Efficiency Study (RQ4)

This section assesses the computational efficiency of DiffGraph. On industry data, we conducted a convergence analysis shown in Figure 4. We also compared the per-epoch running time of DiffGraph and baselines on two link datasets with the following results: **Tmall:** HGT, 16.824s, MBGCN, 12.644s, **DiffGraph**, 6.558s. **Retail_rocket:** HGT, 2.451s, MBGCN, 2.312s, **DiffGraph**, 1.811s. The results show DiffGraph consistently outperforms baselines in training efficiency, benefiting from its effective hidden diffusion.

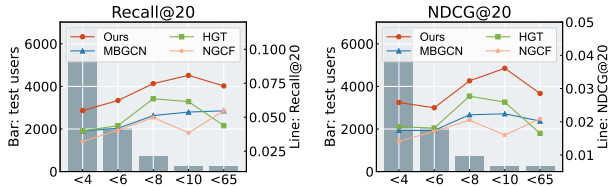
3.6 Performance w.r.t. Data Sparsity (RQ5)

We further assess how the diffusion model in DiffGraph address the data sparsity issue in heterogeneous graph learning. For the Tmall dataset, we split nodes into five groups based on interaction counts (e.g., " <8 ", " <65 "), and evaluate the performance on each group. The results are shown in Figure 5. Our observations include:

- **Trend in Performance.** As edges per node increase, so does performance across all methods, indicating better embeddings

Table 3: Performance decay caused by different noise ratios.

Noise Ratios		10%		30%		50%	
		Recall	NDCG	Recall	NDCG	Recall	NDCG
pv	DiffGraph	98.28%	96.31%	98.80%	97.05%	97.42%	96.68%
	HGT	96.98%	94.79%	97.45%	96.35%	95.59%	90.10%
fav	DiffGraph	97.93%	95.57%	99.14%	95.57%	98.62%	96.31%
	HGT	98.14%	93.75%	97.68%	91.67%	97.22%	89.06%
cart	DiffGraph	98.97%	98.15%	98.28%	97.42%	96.73%	92.62%
	HGT	98.61%	96.88%	96.29%	95.31%	95.82%	92.19%

**Figure 5: Performance w.r.t. different data sparsity**

from richer edge data. However, performance dips in groups with "<10" and "<65" links, likely due to fewer test samples.

- **Effect of Heterogeneity Learning.** Heterogeneous models (MBGCN, HGT) outperform the homogeneous NGCF, showing that multiple context fusion effectively combats data sparsity.
- **Superiority of DiffGraph.** DiffGraph consistently surpasses all baselines, suggesting its diffusion-based approach effectively preserves and utilizes heterogeneous behavioral data, thus offering a robust solution to data sparsity challenges.

3.7 Exploration of Anti-Noise Capacity (RQ6)

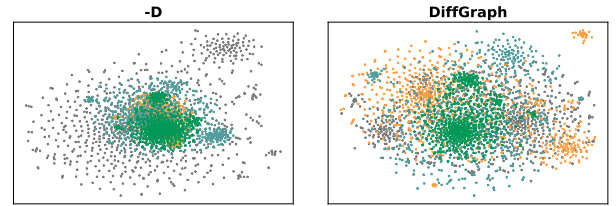
To study the model robustness against data noise, we evaluate the percentage of performance degradation for each heterogeneous relation under different noise ratios (Table 3). In our experiment, 0, 10%, 30% and 50% of the heterogeneous relations are randomly replaced by noise signals. We tested the performance of DiffGraph and HGT on the Tmall dataset. The results demonstrate that our model experiences a markedly less pronounced decline in performance in comparison to the baseline model. Such findings substantiate the enhanced denoising capability of our DiffGraph approach.

3.8 Case Study (RQ7)

We utilized the t-SNE technique to visualize the embeddings of different relation types from the Tmall dataset, color-coding each type for distinction. By comparing the behavior embedding of our DiffGraph to its variant "-D" (which lacks the denoising module), it's evident that our model achieves clearer separation, drawing embeddings with similar semantics closer together. This demonstrates our model's proficiency in capturing and refining heterogeneous data through hidden-space diffusion and denoising, effectively reducing noise and enhancing semantic accuracy. By optimizing the semantic diffusion from source to target, our method efficiently processes and highlights diverse information across graph structures, significantly improving learning outcomes.

4 RELATED WORK

Graph Neural Networks. Graph learning has evolved substantially, with GNNs demonstrating significant impact across domains including spammer detection [22, 42, 50], recommender systems [9,

**Figure 6: Embedding visualization on the Tmall dataset.**

47], and machine translation [25, 41]. While traditional GNNs like Graph Convolutional Networks (GCNs) [12, 52] and Graph Attention Networks (GATs) [24, 51] have advanced representation learning on graphs, they primarily focus on homogeneous or simple bipartite structures. To address the limitations in modeling complex heterogeneous graph relations, we propose a denoised heterogeneous graph neural network based on the diffusion model.

Heterogeneous Graph Learning. Heterogeneous graphs, featuring diverse node and link types, are crucial in various real-world applications. Heterogeneous graph neural networks (HGNNs) develop node embeddings that capture complex semantics for tasks like node classification, edge classification, and link prediction [32, 49]. Methods like HetGNN [49] and MAGNN [10] incorporate attention mechanisms and meta-path aggregations to refine these embeddings. HGT [17] leverages a transformer-based architecture for modeling diverse relationships. While these methods enhance the capture of heterogeneous semantics, they often ignore noise and irrelevant information. Our model introduces a diffusion-based denoising mechanism to extract genuine, task-relevant information.

Generative Models for Graph Learning. Graph generation aims to uncover structural patterns, mitigate anomalies, and simulate new graphs, garnering significant interest. Traditional models [1, 7, 21] were limited to generating graphs with specific statistical traits. However, advancements in deep learning have expanded possibilities. Techniques like Generative Adversarial Networks (GANs) [36] and Variational Autoencoders (VAEs) [48] play key roles in data generation and graph task learning. Recent efforts, including diffusion models [2, 4, 28], focus on generative recommendation with enhanced capabilities. Despite these advancements, the use of these models in graph denoising is limited. This work pioneers the application of generative models to transition from auxiliary heterogeneous graphs to targeted semantic spaces.

5 CONCLUSION

This paper presents DiffGraph, a heterogeneous graph diffusion model that advances graph learning through two key innovations: 1) A latent diffusion mechanism that progressively filters noise while preserving task-relevant signals from source data, and 2) An enhanced semantic transition framework that better captures relationships across different interaction types, enabling more nuanced heterogeneous graph modeling. Extensive experiments conducted on both public benchmarks and industrial datasets demonstrate our approach's superiority in terms of prediction accuracy and model robustness. The results consistently validate the effectiveness of our proposed techniques across various scenarios. Future work could explore extending our model to dynamic heterogeneous graphs, where both node attributes and graph structure evolve over time.

REFERENCES

- [1] R. Albert and A.-L. Barabási. Statistical mechanics of complex networks. *Reviews of modern physics*, 74(1):47, 2002.
- [2] J. Austin, D. D. Johnson, J. Ho, D. Tarlow, and R. Van Den Berg. Structured denoising diffusion models in discrete state-spaces. *Advances in Neural Information Processing Systems (NeurIPS)*, 34:17981–17993, 2021.
- [3] D. Chen, Y. Lin, W. Li, P. Li, J. Zhou, and X. Sun. Measuring and relieving the over-smoothing problem for graph neural networks from the topological view. In *AAAI Conference on Artificial Intelligence (AAAI)*, volume 34, pages 3438–3445, 2020.
- [4] F.-A. Croitoru, V. Hondru, R. T. Ionescu, and M. Shah. Diffusion models in vision: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 2023.
- [5] A. P. Davis, C. J. Grondin, R. J. Johnson, D. Sciaky, B. L. King, R. McMorran, J. Wiegiers, T. C. Wiegiers, and C. J. Mattingly. The comparative toxicogenomics database: update 2017. *Nucleic acids research*, 45(D1):D972–D978, 2017.
- [6] Y. Dong, N. V. Chawla, and A. Swami. metapath2vec: Scalable representation learning for heterogeneous networks. In *Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining*, pages 135–144, 2017.
- [7] P. ERDdS and A. R&w. On random graphs i. *Publ. math. debrecen*, 6(290-297):18, 1959.
- [8] S. Fan, J. Zhu, X. Han, C. Shi, L. Hu, B. Ma, and Y. Li. Metapath-guided heterogeneous graph neural network for intent recommendation. In *International Conference on Knowledge Discovery & Data Mining (KDD)*, pages 2478–2486, 2019.
- [9] W. Fan, Y. Ma, Q. Li, Y. He, E. Zhao, J. Tang, and D. Yin. Graph neural networks for social recommendation. In *The world wide web conference*, pages 417–426, 2019.
- [10] X. Fu, J. Zhang, Z. Meng, and I. King. Magnn: Metapath aggregated graph neural network for heterogeneous graph embedding. In *Proceedings of The Web Conference 2020*, pages 2331–2341, 2020.
- [11] C. Gao, X. He, D. Gan, X. Chen, F. Feng, Y. Li, T.-S. Chua, and D. Jin. Neural multi-task recommendation from multi-behavior data. In *International conference on data engineering (ICDE)*, pages 1554–1557. IEEE, 2019.
- [12] H. Gao, Z. Wang, and S. Ji. Large-scale learnable graph convolutional networks. In *KDD*, pages 1416–1424, 2018.
- [13] W. Hamilton, Z. Ying, and J. Leskovec. Inductive representation learning on large graphs. *Advances in neural information processing systems*, 30, 2017.
- [14] X. He, K. Deng, X. Wang, Y. Li, Y. Zhang, and M. Wang. Lightgcn: Simplifying and powering graph convolution network for recommendation. In *International Conference on Research and Development in Information Retrieval (SIGIR)*, pages 639–648, 2020.
- [15] J. Ho, A. Jain, and P. Abbeel. Denoising diffusion probabilistic models. *Advances in Neural Information Processing Systems (NeurIPS)*, 33:6840–6851, 2020.
- [16] W. Hu, B. Liu, J. Gomes, M. Zitnik, P. Liang, V. Pande, and J. Leskovec. Strategies for pre-training graph neural networks. In *International Conference on Learning Representations (ICLR)*, 2020.
- [17] Z. Hu, Y. Dong, K. Wang, and Y. Sun. Heterogeneous graph transformer. In *ACM Web Conference (WWW)*, pages 2704–2710, 2020.
- [18] B. Jin, C. Gao, X. He, D. Jin, and Y. Li. Multi-behavior recommendation with graph convolutional networks. In *International Conference on Research and Development in Information Retrieval (SIGIR)*, pages 659–668, 2020.
- [19] T. N. Kipf and M. Welling. Variational graph auto-encoders. 2016.
- [20] T. N. Kipf and M. Welling. Semi-supervised classification with graph convolutional networks. In *International Conference on Learning Representations (ICLR)*, 2017.
- [21] J. Kleinberg. The small-world phenomenon: An algorithmic perspective. In *ACM Symposium on Theory of Computing (STOC)*, pages 163–170, 2000.
- [22] A. Li, Z. Qin, R. Liu, Y. Yang, and D. Li. Spam review detection with graph convolutional networks. In *International Conference on Information and Knowledge Management (CIKM)*, pages 2703–2711, 2019.
- [23] Z. Li, L. Xia, and C. Huang. Reddiff: Diffusion model for social recommendation. In *International Conference on Information and Knowledge Management (CIKM)*, pages 1346–1355, 2024.
- [24] R. Liao, Y. Li, Y. Song, et al. Efficient graph generation with graph recurrent attention networks. *Advances in Neural Information Processing Systems (NeurIPS)*, 32, 2019.
- [25] M.-T. Luong, H. Pham, and C. D. Manning. Effective approaches to attention-based neural machine translation. In *Empirical Methods in Natural Language Processing (EMNLP)*, 2015.
- [26] I. Misra and L. v. d. Maaten. Self-supervised learning of pretext-invariant representations. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 6707–6717, 2020.
- [27] C. Park, D. Kim, J. Han, and H. Yu. Unsupervised attributed multiplex network embedding. In *AAAI Conference on Artificial Intelligence (AAAI)*, volume 34, pages 5371–5378, 2020.
- [28] V. Popov, I. Vovk, V. Gogoryan, T. Sadekova, and M. Kudinov. Grad-tts: A diffusion probabilistic model for text-to-speech. In *International Conference on Machine Learning (ICML)*, pages 8599–8608. PMLR, 2021.
- [29] S. Rendle, C. Freudenthaler, Z. Gantner, and L. Schmidt-Thieme. Bpr: Bayesian personalized ranking from implicit feedback. In *Conference on Uncertainty in Artificial Intelligence (UAI)*, pages 452–461, 2009.
- [30] C. Shi, B. Hu, W. X. Zhao, and S. Y. Philip. Heterogeneous information network embedding for recommendation. *Transactions on Knowledge and Data Engineering (TKDE)*, 31(2):357–370, 2018.
- [31] X. Sun, H. Yin, B. Liu, H. Chen, J. Cao, Y. Shao, and N. Q. Viet Hung. Heterogeneous hypergraph embedding for graph classification. In *International Conference on Web Search and Data Mining (WSDM)*, pages 725–733, 2021.
- [32] J. Tang, Y. Yang, W. Wei, L. Shi, L. Xia, D. Yin, and C. Huang. Higt: Heterogeneous graph language model. In *International Conference on Knowledge Discovery and Data Mining (KDD)*, pages 2842–2853, 2024.
- [33] Y. Tian, K. Dong, C. Zhang, C. Zhang, and N. V. Chawla. Heterogeneous graph masked autoencoders. In *AAAI Conference on Artificial Intelligence (AAAI)*, volume 37, pages 9997–10005, 2023.
- [34] P. Velicković, G. Cucurull, A. Casanova, A. Romero, P. Lio, and Y. Bengio. Graph attention networks. In *International Conference on Learning Representations (ICLR)*, 2018.
- [35] P. Velickovic, W. Fedus, W. L. Hamilton, P. Lio, Y. Bengio, and R. D. Hjelm. Deep graph infomax. In *International Conference on Learning Representations (ICLR)*, 2019.
- [36] J. Wang, L. Yu, W. Zhang, Y. Gong, Y. Xu, B. Wang, P. Zhang, and D. Zhang. Irgan: A minimax game for unifying generative and discriminative information retrieval models. In *International Conference on Research and Development in Information Retrieval (SIGIR)*, pages 515–524, 2017.
- [37] W. Wang, Y. Xu, F. Feng, X. Lin, X. He, and T.-S. Chua. Diffusion recommender model. In *International Conference on Research and Development in Information Retrieval (SIGIR)*, 2023.
- [38] X. Wang, X. He, M. Wang, F. Feng, and T.-S. Chua. Neural graph collaborative filtering. In *international conference on Research and development in Information Retrieval (SIGIR)*, pages 165–174, 2019.
- [39] X. Wang, H. Ji, C. Shi, B. Wang, Y. Ye, P. Cui, and P. S. Yu. Heterogeneous graph attention network. In *ACM Web Conference (WWW)*, pages 2022–2032, 2019.
- [40] X. Wang, N. Liu, H. Han, and C. Shi. Self-supervised heterogeneous graph neural network with co-contrastive learning. In *International Conference on Knowledge Discovery & Data Mining (KDD)*, pages 1726–1736, 2021.
- [41] Y. Wu. Google’s neural machine translation system: Bridging the gap between human and machine translation. *arXiv preprint arXiv:1609.08144*, 2016.
- [42] Y. Wu, D. Lian, Y. Xu, L. Wu, and E. Chen. Graph convolutional networks with markov random field reasoning for social spammer detection. In *AAAI Conference on Artificial Intelligence (AAAI)*, volume 34, pages 1054–1061, 2020.
- [43] L. Xia, C. Huang, Y. Xu, P. Dai, B. Zhang, and L. Bo. Multiplex behavioral relation learning for recommendation via memory augmented transformer network. In *International Conference on Research and Development in Information Retrieval (SIGIR)*, pages 2397–2406, 2020.
- [44] Y. Yang, C. Huang, L. Xia, Y. Liang, Y. Yu, and C. Li. Multi-behavior hypergraph-enhanced transformer for sequential recommendation. In *International Conference on Knowledge Discovery and Data Mining (KDD)*, pages 2263–2274, 2022.
- [45] R. Ying, R. He, K. Chen, P. Eksombatchai, W. L. Hamilton, and J. Leskovec. Graph convolutional neural networks for web-scale recommender systems. In *International Conference on Knowledge Discovery & Data Mining (KDD)*, pages 974–983, 2018.
- [46] Y. You, T. Chen, Y. Sui, et al. Graph contrastive learning with augmentations. *Advances in Neural Information Processing Systems (NeurIPS)*, 33:5812–5823, 2020.
- [47] J. Yu, H. Yin, J. Li, Q. Wang, N. Q. V. Hung, and X. Zhang. Self-supervised multi-channel hypergraph convolutional network for social recommendation. In *ACM Web Conference (WWW)*, pages 413–424, 2021.
- [48] X. Yu, X. Zhang, Y. Cao, and M. Xia. Vaegan: A collaborative filtering framework based on adversarial variational autoencoders. In *International Joint Conference on Artificial Intelligence (IJCAI)*, pages 4206–4212, 2019.
- [49] C. Zhang, D. Song, C. Huang, A. Swami, and N. V. Chawla. Heterogeneous graph neural network. In *International Conference on Knowledge Discovery & Data Mining (KDD)*, pages 793–803, 2019.
- [50] S. Zhang, H. Yin, T. Chen, Q. V. N. Hung, Z. Huang, and L. Cui. Gcn-based user representation learning for unifying robust recommendation and fraudster detection. In *International Conference on Research and Development in Information Retrieval (SIGIR)*, pages 689–698, 2020.
- [51] W. Zhang, Z. Yin, Z. Sheng, Y. Li, O. Wen, X. Li, Y. Tao, Z. Yang, and B. Cui. Graph attention multi-layer perceptron. In *International Conference on Knowledge Discovery and Data Mining (KDD)*, pages 4560–4570, 2022.
- [52] Y. Zhang, X. Wang, C. Shi, N. Liu, and G. Song. Lorentzian graph convolutional networks. In *ACM Web Conference (WWW)*, pages 1249–1261, 2021.

ETHICAL CONSIDERATIONS

In this section, we examine the ethical implications of the DiffGraph framework, which is designed to distill essential information from auxiliary relational data for enhancing prediction tasks in heterogeneous graphs. Our discussion addresses potential ethical challenges and considerations that may arise during both the implementation and deployment phases of the DiffGraph framework.

Data Privacy and Consent. The DiffGraph framework processes auxiliary relational data to enhance predictions in heterogeneous graphs, which inherently raises important privacy and consent considerations. When handling relational data, particularly those containing personal information, strict compliance with data protection regulations (e.g., GDPR, CCPA) is paramount. The potential for unintended inference or exposure of sensitive information through relational analysis presents a significant privacy concern, especially if explicit user consent is not properly obtained. To address these challenges, implementations must incorporate robust safeguards, including strict access control mechanisms, advanced

data anonymization techniques, and transparent data usage policies. These measures are essential for protecting user privacy while maintaining the framework's analytical capabilities.

Transparency and Explainability. The sophisticated nature of the DiffGraph framework, particularly its mechanisms for integrating and interpreting auxiliary relational data, presents significant challenges in terms of transparency and explainability. In high-stakes environments, stakeholders need clear understanding of how the model arrives at its predictions and subsequent recommendations. The inherent complexity of the model can potentially obscure decision-making processes, making it difficult to identify and rectify logical errors while potentially eroding stakeholder trust. To address these challenges, it is crucial to develop robust explainability methods, such as interpretable feature importance metrics and comprehensive decision rationale documentation. These transparency measures not only foster trust among stakeholders but also ensure compliance with regulatory requirements and maintain accountability throughout the model's deployment lifecycle.