MIGDiff: Multi-attributes Imputations for Attribute-missing Graphs via Graph Denoising Diffusion Model

Ye Liu, Yang Chen, Hongmin Cai*

South China University of Technology, Guangzhou, China
yliu03@scut.edu.cn, cyang810@163.com, hmcai@scut.edu.cn

Abstract

The missing of graph attributes poses a significant challenge in graph representation learning. Some existing graph attribute completion methods adopt the shared-space hypothesis or employ end-to-end frameworks to perform single-attribute imputation. However, these models can only generate one single attribute with a few specific patterns that either adhere to prior knowledge or are optimal for downstream tasks, making it difficult to capture the full range of variations in the target attribute distribution. This limitation negatively impacts the model's generalizability and efficiency. Therefore, to address this issue, we proposed a new method based on graph denoising diffusion model, called **M**ulti-attribute **I**mputation **G**raph Denoising **Diff**usion Model (**MIGDiff**), which can generate multiple high-quality attributes. Specifically, it employs a **Dual-source Auto-encoder** on existing attributes and graph topology to extract reliable knowledge, which serves as a condition for training the diffusion module. Within diffusion, noise is added to the structural embeddings of nodes without attributes in the forward process. In the reverse process, a **Structure-aware Denoising Network** is devised to integrate feature and structural information via attention mechanism and to perform neighbor-guided refinement based on graph connectivity, thereby enhancing denoising and accurately recovering missing attributes while effectively maintaining structural consistency and distributional fidelity. During generation, multiple initial values are sampled to produce diverse attribute imputations, avoiding focusing on a few easy-to-learn patterns. Extensive experiments conducted on four public datasets highlight the state-of-the-art performance of MIGDiff in both attribute imputation and node classification tasks.

Code — https://github.com/YeLiu-Lab/MIGDiff.git

Introduction

Graph neural networks (GNNs) have emerged as a powerful tool for graph representation learning, utilizing a message-passing mechanism to aggregate information from neighboring nodes and edges. The performance of GNNs heavily relies on the quality of the initial node and edge attributes. However, in real-world graph data, missing attributes are common due to privacy concerns or the high demands of data collection (Moepya et al. 2016; Krasnova et al. 2010), as illustrated in Fig. 1 (a). Attribute-missing issue introduces significant uncertainty to the graph representation learning, which degrades the performance of GNNs.

Traditional imputation methods such as statistical approaches (Xia et al. 2017), and matrix completion (Zhang and Chen 2020) rely only on observed data statistics and ignore graph structure. However, GNNs (Kipf and Welling 2016; Veličković et al. 2017) exploit structure but are not explicitly designed for missing-attribute imputation. Recently, several advanced methods have been proposed. For example, SAT (Chen et al. 2022) employs an autoencoder to reconstruct structure and attributes separately, leveraging a shared-space assumption to enforce their consistency. ITR (Tu et al. 2022) initializes embeddings using structural information and refines them by adaptively aggregating observed information. AMGCL (Zhang et al. 2023) leverages Dirichlet energy minimization for feature pre-coding and employs a graph augmentation based contrastive learning method to learn latent variables. AIAE (Xia et al. 2024) employs a knowledge distillation-based encoder and a multi-scale decoder with masking mechanisms to integrate attribute and structural information, further generating graph representations. However, these approaches typically focus on the imputation of single attribute, as depicted in Fig. 1 (b). This narrow focus hinders their ability to capture the inherent uncertainty of missing information, thus limiting their capacity to model the full spectrum of data variations. Therefore, a critical question emerges: ***Is single-attribute imputation sufficient to support effective graph representation learning?***

To explore this question, we conduct experiments on two single-attribute imputation methods, AmGCL (Zhang et al. 2023) and AIAE (Xia et al. 2024), and analyze their disparities by visualizing the distribution of real missing attributes, on the Cora (McCallum et al. 2000) and Citeseer (Sen et al. 2008) datasets. Notably, compared to the ground-truth distribution in Figure 2(4), AmGCL exhibits overlapping regions between certain classes while forms indistinct clusters. Meanwhile, AIAE results in heavily intermixed node distributions, making class separation difficult. Consequently, these single-attribute imputation methods fail to effectively recover the original attribute distribution, leading to de-

*The corresponding author.

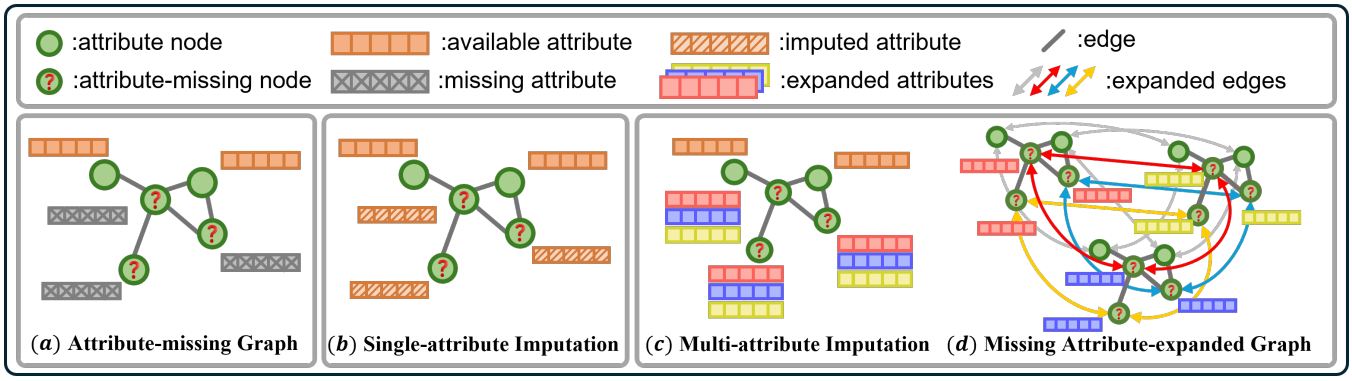

Figure 1: Attribute-missing graph and the proposed imputation strategies.

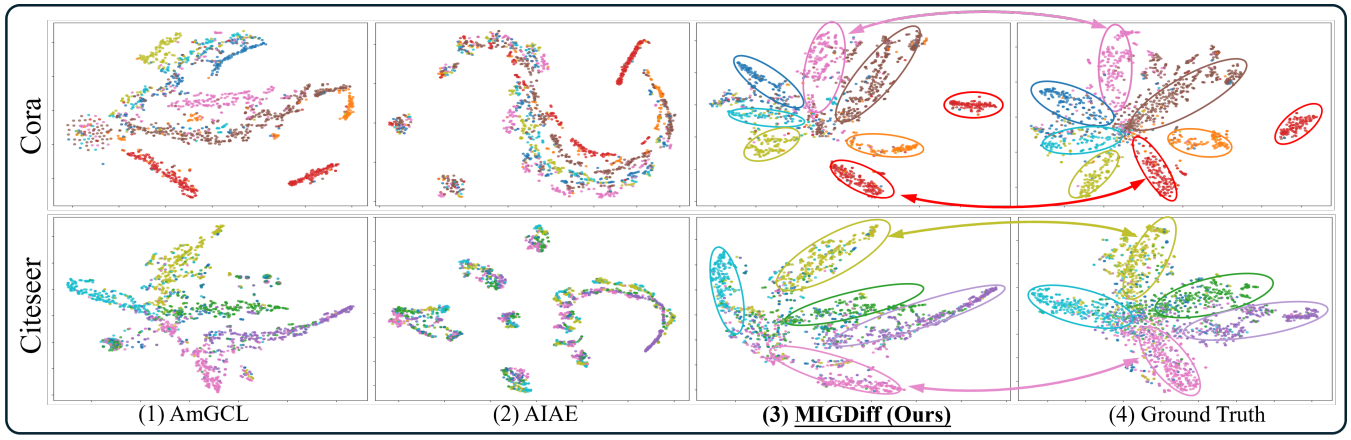

Figure 2: Visualization of representations for nodes with imputed attributes generated by a supervised, pretrained GCN–MLP framework: a two-layer GCN produced embeddings and a one-layer MLP yielded classification probabilities. 60% of nodes were used for training and 40% for testing. Test-node attributes were masked to simulate missing data and subsequently imputed by different methods. Embeddings of the imputed nodes were generated by the pretrained GCN, and visualized via t-SNE.

graded graph representations.

To address the above issue, we introduce an innovative framework called **M**ulti-attribute **I**mputation **G**raph Denoising **Diff**usion Model (**MIGDiff**), which supports multi-attribute graph imputation, as shown in Fig. 1(c). Specifically, MIGDiff employs a **Dual-source Auto-encoder** to extract complementary information from both the graph structure and observed node attributes, thereby obtaining topological embeddings and attribute embeddings. These embeddings are used as condition in the diffusion model. During the diffusion process, we introduce a **Structure-aware Denoising Network**, a novel denoising model designed to improve the generation of missing node attributes. It first effectively integrates topological and attribute embeddings through an attention mechanism, and then leverages the aggregated information to guide the denoising of noisy nodes based on graph connectivity. During training, the topological embeddings of nodes without attributes serve as the starting point of the diffusion process, which is recovered by the denoising model. During inference, MIGDiff samples from a Gaussian distribution multiple times to generate diverse imputations for the missing attributes. These imputations are decoded to recover the missing node attributes, enabling MIGDiff to effectively handle the uncertainty of missing data and produce robust, diverse, and accurate imputations. Analogous to the experiments in Fig. 2(1) and (2), the proposed MIGDiff framework reconstructs missing attributes by averaging attributes inferred from three independent sampling iterations. The reconstructed attributes were then processed by the pretrained GCN to generate nodes embeddings, which are visualized in Fig. 2(3). Obviously, MIGDiff achieves superior inter-class separation and generates embeddings highly consistent with the ground truth, demonstrating its effectiveness in recovering meaningful node features and improving graph representation quality.

In conclusion, the main contributions of this paper can be summarized as follows:

- Inspired by diffusion models, we propose MIGDiff, a novel method that performs multi-attribute imputation for graphs with missing attributes. By imputing multiple attributes, MIGDiff provides more robust and accurate imputations that better captures data distribution,

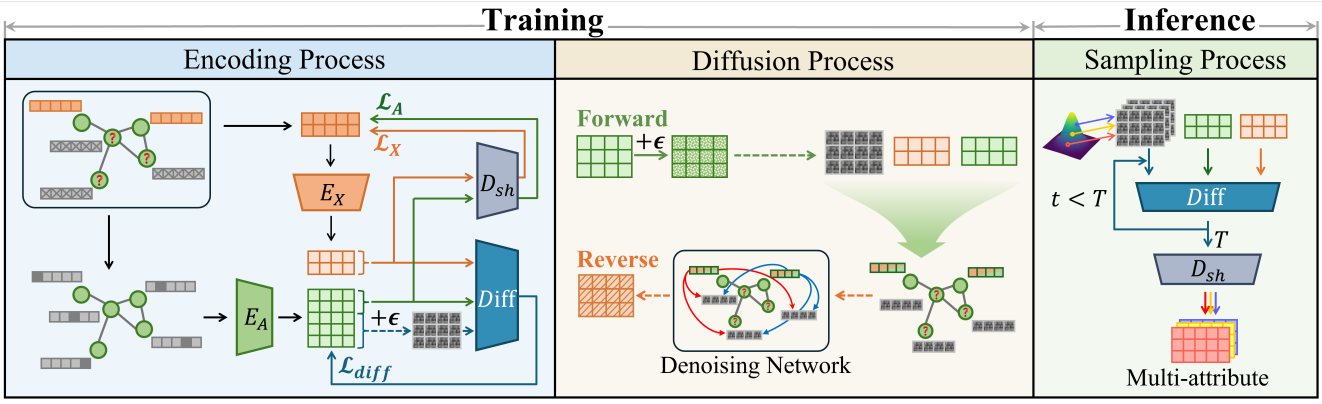

Figure 3: An illustration of MIGDiff architecture.

hence improving the performances of graph representation learning and downstream tasks.

- A novel structure-aware denoising network is developed to enhance missing node attributes generation. It integrates attribute and structural information via attention and exploits graph-structure guidance to improve denoising performance.
- Extensive experiments on four publicly available datasets demonstrate that MIGDiff not only performs well in single-attribute imputation but also produces multiple high-quality imputations, which enhances attribute completion and node classification performances.

Methodology

Our proposed method MIGDiff for missing attribute graph imputation consists of two parts: training and inference, as illustrated in Fig. 3, where the training phase includes both the encoding and diffusion processes, and inference corresponds to the sampling process. Assume an attribute-missing graph is defined as $\mathcal{G}_m(\mathcal{V}, \mathbf{A}, \mathbf{X}^{(\text{obs})})$, where $\mathcal{V} = \{v_1, v_2, \ldots, v_N\}$ is the set of N nodes, and the adjacency matrix $\mathbf{A} \in \mathbb{R}^{N\times N}$ encodes the graph's connectivity. The observed attributes corresponding to the nodes $\mathcal{V}^{(\text{obs})}$ are denoted as $\mathbf{X}^{(\text{obs})} \in \mathbb{R}^{N_o\times D}$, where N_o is the number of nodes with attributes and D is the dimension of the original features. The target of attribute-missing graph imputation is to recover the missing node attributes $\mathbf{X}^{(\text{miss})} \in \mathbb{R}^{N_m\times D}$ associated with $\mathcal{V}^{(\text{miss})}$, where N_m is the number of nodes with missing attributes, and $N = N_o + N_m$, $\mathcal{V} = \mathcal{V}^{(\text{obs})} \cup \mathcal{V}^{(\text{miss})}$.

MIGDiff Training

Dual-source Auto-encoder. To efficiently map the original high-dimensional data to a compact low-dimensional space, a dual-source autoencoder is employed. Considering the challenge of missing attributes in incomplete graphs, this framework utilizes two separate encoders to extract complementary information from the graph topology and the existing attributes respectively. To capture the graph topology, an encoder E_A, consisting of two graph attention network (GAT) (Veličković et al. 2017) layers, is applied on the graph $\mathcal{G}_a = (\mathcal{V}, \mathbf{A}, \mathbf{I})$ to generate the topological embeddings $\mathbf{Z}_a \in \mathbb{R}^{N\times D_{\text{emb}}}$, where the initial attribute is set as the identity matrix $\mathbf{I}$. The resulting embeddings can be divided as $\mathbf{Z}_a = [\mathbf{Z}_a^{(\text{obs})}, \mathbf{Z}_a^{(\text{miss})}]$, where $\mathbf{Z}_a^{(\text{obs})}$ corresponds to the structural embeddings of nodes with observed attributes. Simultaneously, an attribute encoder E_X, composed of two MLP layers, maps the observed attributes $\mathbf{X}^{(\text{obs})}$ to the corresponding embeddings $\mathbf{Z}_x^{(\text{obs})} \in \mathbb{R}^{N_o\times D_{\text{emb}}}$. Since the information of nodes with available attributes are reliable, both $\mathbf{Z}_a^{(\text{obs})}$ and $\mathbf{Z}_x^{(\text{obs})}$ are used for reconstruction. Besides, the graph structure and the attributes share similar semantics, so a shared two-layer MLP decoder D_{sh} is employed to reconstruct the existing attributes from these latent embeddings. Mathematically, the reconstruction process is expressed as follows:

$$\begin{cases} \mathbf{Z}_x^{(\text{obs})} = E_X\left(\mathbf{X}^{(\text{obs})}\right) \\ \widehat{\mathbf{X}}_X^{(\text{obs})} = D_{\text{sh}}\left(\mathbf{Z}_x^{(\text{obs})}\right), \end{cases} \quad \begin{cases} \mathbf{Z}_a = E_A\left(\mathbf{I}, \mathbf{A}\right) \\ \widehat{\mathbf{X}}_A^{(\text{obs})} = D_{\text{sh}}\left(\mathbf{Z}_a^{(\text{obs})}\right) \end{cases} \tag{1}$$

where $\widehat{\mathbf{X}}_A^{(\text{obs})}$ and $\widehat{\mathbf{X}}_X^{(\text{obs})}$ denote the reconstructed existing attributes from topology and existing attributes respectively. Two loss function are derived to minimize the difference between observed attributes and reconstructed attributes:

$$\mathcal{L}_X = \left\|\mathbf{X}^{(\text{obs})} - \widehat{\mathbf{X}}_X^{(\text{obs})}\right\|_2^2, \mathcal{L}_A = \left\|\mathbf{X}^{(\text{obs})} - \widehat{\mathbf{X}}_A^{(\text{obs})}\right\|_2^2. \tag{2}$$

Forward and Reverse Processes. After capturing useful information in the incomplete graph by the dual-source auto-encoder, we construct a conditional diffusion model to fully exploit this information for modeling the true distribution of the missing attribute embeddings.

During the forward process, $\mathbf{Z}_a^{(\text{miss})}$ serves as the initial state $\mathbf{Z}_0$ for the conditional diffusion. The embeddings then undergo a series of parameterized transformations, defined as follows:

$$q\left(\mathbf{Z}_t \mid \mathbf{Z}_{t-1}\right) = \mathcal{N}\left(\mathbf{Z}_t; \sqrt{1-\beta_t}\mathbf{Z}_{t-1}, \beta_t\mathbf{I}\right) \tag{3}$$

$$q\left(\mathbf{Z}_t \mid \mathbf{Z}_0\right) = \mathcal{N}\left(\mathbf{Z}_t; \sqrt{\bar{\alpha}_t}\mathbf{Z}_0, \left(1-\bar{\alpha}_t\right)\mathbf{I}\right) \tag{4}$$

where $\beta_t \in (0,1)$ represents the noise level at the t-th time step with $t \in (0,T)$, controlling the degree of data degradation during the diffusion process. Additionally, $\alpha_t = 1 - \beta_t$ and $\bar{\alpha}_t = \prod_{i=1}^{t} \alpha_i$. The reverse process starts from $\mathbf{Z}_T$, as shown in the middle of Fig. 3, where a learnable structure-aware denoising network leverages observed node information and graph structure to iteratively denoise the embeddings of nodes with missing attributes. Specifically, the noisy features are progressively restored via a conditional probability distribution, which is conditioned on the observed attribute embeddings $\mathbf{Z}_x^{(\text{obs})}$, the topological embeddings $\mathbf{Z}_a^{(\text{obs})}$, and the adjacency matrix $\mathbf{A}$:

$$p_\theta\left(\mathbf{Z}_{t-1} \mid \mathbf{Z}_t, \mathbf{Z}_x^{(\text{obs})}, \mathbf{Z}_a^{(\text{obs})}, \mathbf{A}\right) = \mathcal{N}\left(\mathbf{Z}_{t-1}; \boldsymbol{\mu}_\theta, \Sigma_\theta\right) \tag{5}$$

where $\mathcal{N}$ represents a Gaussian distribution parameterized by the mean $\boldsymbol{\mu}_\theta$ and variance Σ_θ, θ represents the learnable parameters of the denoising model.

Structure-aware Denoising Network. During the denoising process, a new graph-based denoising model shown in Fig. 4 is designed to achieve accurate missing attribute imputations. To integrate topology and attribute information, the observed latent attribute embeddings $\mathbf{Z}_x^{(\text{obs})}$ and the topological features $\mathbf{Z}_a^{(\text{obs})}$ are first concatenated and then processed through a single-layer MLP. Then, a self-attention mechanism is introduced to refine the aggregated features:

$$\mathbf{Z}_{\text{agg}}^{(\text{obs})} = \text{Self-Attn}\left(\text{MLP}\left(\mathbf{Z}_x^{(\text{obs})} \| \mathbf{Z}_a^{(\text{obs})}\right)\right) \tag{6}$$

where Self-Attn$(\cdot)$ denotes the self-attention operations. Eq. (6) enables the model to focus on critical relationships and dependencies, thereby improving the accuracy of the denoising process and ensuring better reconstruction of the missing attributes. After obtaining the aggregated embeddings $\mathbf{Z}_{\text{agg}}^{(\text{obs})} \in \mathbb{R}^{N_o \times D_{\text{emb}}}$, the noisy graph $\widetilde{\mathcal{G}}_m = (\mathcal{V}, \mathbf{A}, \mathbf{Z}_{\text{m}})$ is constructed by combining these embeddings with the noisy embeddings as the completed attributes, i.e. $\mathbf{Z}_{\text{m}} = [\mathbf{Z}_{\text{agg}}^{(\text{obs})}, \mathbf{Z}_t]$. To denoise the noisy graph, $\widetilde{\mathcal{G}}_m$ is processed through a residual graph autoencoder, where the reliable information in $\mathbf{Z}_{\text{agg}}^{(\text{obs})}$ is leveraged to guide the denoising of noisy nodes, ensuring that the embeddings of nodes with observed attributes effectively refine the denoising process. The residual connections in residual graph autoencoder further preserve the structural integrity of the graph, facilitating robust denoising and improving the quality of the reconstructed attributes.

Model Optimization Strategy. To optimize θ in Eq. (5), the denoising model minimizes the variational upper bound (Liu et al. 2024). By applying Bayes' rule and taking our context into account, $q(\mathbf{Z}_{t-1}|\mathbf{Z}_t, \mathbf{Z}_0)$ can be expressed in the following form:

$$\begin{cases} \boldsymbol{\mu}_\theta = \dfrac{\sqrt{\alpha_t}\left(1-\bar{\alpha}_{t-1}\right)}{1-\bar{\alpha}_t}\mathbf{Z}_t + \dfrac{\sqrt{\bar{\alpha}_{t-1}}\left(1-\alpha_t\right)}{1-\bar{\alpha}_t}\hat{\mathbf{Z}}_\theta \\ \sigma^2(t) = \dfrac{\left(1-\alpha_t\right)\left(1-\bar{\alpha}_{t-1}\right)}{1-\bar{\alpha}_t} \end{cases} \tag{7}$$

where $\boldsymbol{\mu}_\theta(t)$ and $\sigma^2(t)$ represent the mean and variance of the Gaussian distribution at step t, and $\hat{\mathbf{Z}}_\theta \in \mathbb{R}^{N_m \times D_{\text{emb}}}$ is the output of the graph denoising model corresponding to the nodes with missing attribute. In addition, as noted in (Liu et al. 2024), predicting initialized data rather than noise improves data imputation performance. Consequently, we propose to reconstruct the initial data $\mathbf{Z}_0$, the loss function for the diffusion part of the MIGDiff Model can be rewritten as:

$$\mathcal{L} = \mathbb{E}_{q(\mathbf{Z}_t|\mathbf{Z}_0)}\left[\frac{1}{2}\left(\frac{\bar{\alpha}_{t-1}}{1-\bar{\alpha}_{t-1}} - \frac{\bar{\alpha}_t}{1-\bar{\alpha}_t}\right)\left\|\hat{\mathbf{Z}}_\theta - \mathbf{Z}_0\right\|_2^2\right] \tag{8}$$

Following (Ho, Jain, and Abbeel 2020), the coefficient is omitted to enhance computational efficiency, allowing Eq. 8 to be simplified as follows:

$$\mathcal{L}_{\text{diff}} = \mathbb{E}\left[\left\|\mathbf{Z}_0 - \hat{\mathbf{Z}}_\theta\left(\mathbf{Z}_t, \mathbf{Z}_x^{(\text{obs})}, \mathbf{Z}_a^{(\text{obs})}, \mathbf{A}\right)\right\|^2\right] \tag{9}$$

where $\mathbf{Z}_0$ is $\mathbf{Z}_a^{(\text{miss})}$, and $\mathbf{Z}_x^{(\text{obs})}, \mathbf{Z}_a^{(\text{obs})}, \mathbf{A}$ can be regarded as conditions for the diffusion model. Combining the reconstruction loss $\mathcal{L}_{\text{X}}$ and $\mathcal{L}_{\text{A}}$ in Eq. (2), the overall loss of MIGDiff is as follows:

$$\mathcal{L} = \mathcal{L}_{\text{diff}} + \alpha\mathcal{L}_{\text{X}} + \beta\mathcal{L}_{\text{A}} \tag{10}$$

where α and β are the hyperparameters.

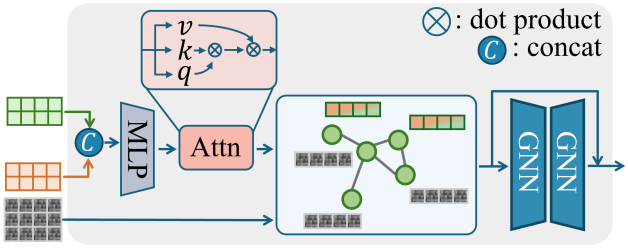

Figure 4: Structure-aware Denoising Network.

MIGDiff Inference

During inference, MIGDiff initiates the reverse diffusion process by sampling noise from a standard Gaussian distribution for multiple times. This noise is progressively transformed through the reverse diffusion steps to recover embeddings of missing nodes, which are then decoded to generate multiple imputed attributes. Concretely, for the i-th imputation, a random initial value $\mathbf{Z}_T^{(i)} \sim \mathcal{N}(0, \mathbf{I})$ is sampled, and inference at each time step t is performed by using trained denoising model, i.e., $\hat{\mathbf{Z}}_{t-1}^{(i)} = \mu_\theta\left(\mathbf{Z}_t^{(i)}, \mathbf{Z}_x^{(\text{obs})}, \mathbf{Z}_a^{(\text{obs})}, \mathbf{A}\right)$. Finally, the embedding $\mathbf{Z}_0^{(i)}$ of all nodes with missing attributes can be iteratively recovered, and its corresponding attributes can be reconstructed through the trained shared decoder, as formulated by $\widehat{\mathbf{X}}^{(i)} = D_{sh}\left(\hat{\mathbf{Z}}_0^{(i)}\right)$. This process is repeated k times to obtain a set of k plausible missing attribute matrices $\widehat{\mathcal{X}}_k = \left\{\widehat{\mathbf{X}}^{(1)}, \widehat{\mathbf{X}}^{(2)}, \ldots, \widehat{\mathbf{X}}^{(k)}\right\}$, resulting in multi-attribute imputations for all nodes with missing attributes, as shown in Fig. 1(c).

Dataset	Metric (%)	NA	GNN*	Hers	GRNA	ARWMF	VAE/VAE-MI	SAT/SAT-MI	ITR/ITR-MI	AmGCL	AIAE	**MIGDiff**
Cora	R@10	9.19	12.78	12.37	12.43	8.62	8.84/12.44	16.29/12.32	16.40/15.46	17.20	16.87	**17.65**
	R@20	14.28	18.09	17.12	17.29	12.73	12.42/17.18	23.46/17.04	23.49/22.50	24.60	23.69	**25.02**
	R@50	20.08	29.78	28.26	28.22	20.38	21.44/28.31	36.33/28.18	36.06/35.89	37.37	34.44	**37.71**
	N@10	12.26	17.39	17.00	17.04	10.66	12.15/17.03	22.57/16.89	23.01/21.47	23.82	23.85	**24.49**
	N@20	15.60	20.97	20.26	20.29	13.35	14.59/20.26	27.33/20.11	27.75/26.15	28.73	28.34	**29.45**
	N@50	18.75	27.13	26.07	26.04	17.37	19.30/26.09	34.12/25.95	34.42/33.22	35.50	34.12	**36.16**
Citeseer	R@10	5.19	6.24	5.63	5.52	2.42	3.71/5.60	7.88/5.66	9.72/8.45	9.76	9.67	**10.07**
	R@20	9.16	10.93	10.14	10.01	4.55	6.62/10.12	13.31/10.12	15.51/14.08	15.66	15.10	**15.99**
	R@50	15.09	20.47	19.28	19.44	9.72	12.99/19.56	24.21/19.55	26.79/25.38	27.27	25.50	**27.77**
	N@10	8.34	10.21	8.80	8.77	3.84	5.86/8.75	13.60/8.78	16.45/14.48	16.61	16.40	**17.18**
	N@20	11.66	14.11	12.54	12.51	5.61	8.28/12.50	18.15/12.49	21.30/19.18	21.54	20.95	**22.14**
	N@50	15.70	20.36	18.52	18.68	8.98	12.46/18.69	25.29/18.67	28.71/26.61	29.18	27.79	**29.89**
Amac	R@10	3.23	2.73	2.59	2.73	1.77	2.57/2.70	4.23/2.63	4.44/4.06	4.24	4.54	**4.58**
	R@20	5.94	5.31	5.11	5.30	3.51	5.05/5.29	7.49/5.21	7.81/7.15	7.46	7.90	**7.96**
	R@50	13.08	12.79	12.32	12.72	8.33	12.12/12.77	15.78/12.66	16.12/15.28	15.69	16.34	**16.40**
	N@10	7.93	6.75	6.45	6.74	4.53	6.38/6.71	10.33/6.55	10.83/9.97	10.35	11.08	**11.15**
	N@20	11.60	10.29	9.90	10.28	6.94	9.76/10.27	14.66/10.08	15.29/14.12	14.64	15.48	**15.63**
	N@50	19.27	18.29	17.64	18.25	12.25	17.41/18.28	23.48/18.08	24.12/22.78	23.41	24.32	**24.59**
Amap	R@10	3.28	2.95	2.77	2.95	1.86	2.76/2.94	4.26/2.78	4.34/4.09	4.23	4.41	**4.42**
	R@20	6.16	5.74	5.38	5.74	3.66	5.27/5.74	7.68/5.53	7.76/7.40	7.56	7.82	**7.89**
	R@50	13.66	13.26	12.84	13.26	8.82	12.70/13.25	16.32/13.15	16.40/15.81	16.10	16.37	**16.53**
	N@10	8.07	7.16	6.85	7.16	4.72	6.76/7.18	10.46/6.89	10.72/10.06	10.36	10.81	**10.86**
	N@20	11.91	10.96	10.42	10.93	7.21	10.19/10.96	15.00/10.61	15.28/14.47	14.81	15.36	**15.47**
	N@50	19.98	19.06	18.43	19.04	12.84	18.19/19.06	24.19/18.78	24.47/23.45	23.90	24.47	**24.66**

Table 1: Attribute reconstruction results with 0.4 observed rate. Best and sub-optimal results are highlighted in **bold** and underlined, respectively. (R denotes Recall, and N denotes NDCG.)

Experiments

Datasets and Baselines

We conduct experiments on four publicly available real-world datasets, including **Cora** (McCallum et al. 2000), **Citeseer**(Sen et al. 2008), **Amac** and **Amap** (Shchur et al. 2018). To evaluate the performance of MIGDiff, we compare it with 12 state-of-the-art methods, which are capable of generating imputations for nodes with missing attributes. These baseline methods can be categorized into 4 groups: classical methods and recent methods designed for attribute-missing graphs. The classical methods include **NeighAggre (NA)** (Şimşek and Jensen 2008), **GCN** (Kipf and Welling 2016), **GAT** (Veličković et al. 2017), and **GraphSage** (Hamilton, Ying, and Leskovec 2017). "GNN*" denotes the best result among GCN, GraphSage, and GAT. A method designed to address the cold-start problem in recommender systems, which can also be applied to attribute imputation, is **Hers**(Hu et al. 2019). Methods based on attributed random walking for attribute imputation include **GraphRNA (GRNA)**(Huang et al. 2019) and **ARWMF** (Chen et al. 2019). The imputation methods tailored for attribute-missing graphs include **VAE** (Chen et al. 2022), **SAT** (Chen et al. 2022), **ITR** (Tu et al. 2022), **AmGCL** (Zhang et al. 2023) and **AIAE** (Xia et al. 2024). Since existing methods only support single-attribute imputation, we enhance them with VAE-style latent-probabilistic modeling and reparameterized sampling, producing VAE-MI, SAT-MI, and ITR-MI for fair comparison with MIGDiff. Unlike the prior VAE-based method in (Chen et al. 2022), which encodes observed node attributes into latent codes, aggregates neighbor information in the latent space, and reconstructs attributes via a decoder without leveraging generative modeling, while VAE-MI can generate multiple attributes.

Experimental Settings

For the imputation of missing attributes in MIGDiff, three independent inferences were performed, the resulting attributes were averaged and used for downstream tasks. To ensure a fair comparison, VAE-MI, SAT-MI, and ITR-MI similarly performed three imputations and averaged their attributes before evaluation. Baselines designed for single-attribute imputation were directly evaluated on the attribute reconstruction and node classification tasks.

To efficiently assess the effectiveness of these imputations on the node classification task. Specifically, two complementary experimental settings are adopted. The first setting evaluates the discriminative power of the reconstructed attributes by training a two-layer MLP using only these attributes. The second setting further incorporates structural information by applying a two-layer GCN to the subgraph composed of attribute-missing nodes and their corresponding recovered attributes. Moreover, to further explore the potential of MIGDiff, we construct a missing

Experiments	NA	GNN*	Hers	GRNA	ARWMF	VAE/VAE-MI	SAT/SAT-MI	ITR/ITR-MI	AmGCL	AIAE	**MIGDiff**
MLP-Cora	62.45 ±0.46	41.25 ±0.14	29.39 ±0.00	29.53 ±0.06	29.39 ±0.00	29.39/29.39 ±0.00/±0.00	80.49/29.39 ±0.41/±0.00	80.98/78.29 ±0.32/±0.23	<u>82.78</u> ±0.37	80.10 ±0.39	**84.97** ±0.64
MLP-Citeseer	55.34 ±0.32	35.97 ±0.30	21.21 ±0.30	23.83 ±0.66	24.33 ±0.15	23.41/21.76 ±0.13/±0.42	64.78/21.95 ±0.36/±0.41	65.85/64.03 ±0.32/±0.43	**68.91** ±0.29	64.83 ±0.44	<u>68.73</u> ±0.35
MLP-Amac	83.73 ±0.11	37.48 ±0.00	37.48 ±0.00	37.06 ±0.02	37.48 ±0.00	37.49/37.48 ±0.03/±0.00	83.25/37.48 ±0.09/±0.00	<u>84.84</u>/78.04 ±0.10/±0.15	81.31 ±0.11	71.62 ±0.13	**88.18** ±0.11
MLP-Amap	88.50 ±0.12	26.01 ±0.04	26.50 ±0.03	26.02 ±0.05	26.02 ±0.05	26.26/25.99 ±0.23/±0.01	89.66/26.05 ±0.11/±0.07	<u>90.65</u>/88.36 ±0.08/±0.05	89.99 ±0.06	79.09 ±0.17	**90.99** ±0.12
GCN-Cora	65.10 ±0.35	45.33 ±0.30	30.38 ±0.17	81.73 ±0.20	33.23 ±0.45	32.06/31.94 ±0.22/±0.22	85.09/35.65 ±0.20/±0.31	84.30/83.37 ±0.13/±0.13	<u>85.29</u> ±0.17	83.30 ±0.20	**85.81** ±0.21
GCN-Citeseer	54.72 ±0.28	38.89 ±0.23	26.50 ±0.24	67.54 ±0.25	28.28 ±0.32	29.48/26.21 ±0.33/±0.26	68.75/26.54 ±0.25/±0.31	68.57/67.32 ±0.22/±0.30	<u>69.75</u> ±0.19	66.73 ±0.32	**69.90** ±0.20
GCN-Amac	87.10 ±0.10	40.50 ±0.51	40.40 ±0.31	84.44 ±0.14	40.50 ±0.51	40.42/40.21 ±0.37/±0.36	87.06/39.96 ±0.13/±0.35	<u>87.95</u>/84.10 ±0.08/±0.14	86.69 ±0.07	82.12 ±0.29	**89.80** ±0.07
GCN-Amap	90.06 ±0.08	36.43 ±0.20	36.38 ±0.17	89.55 ±0.08	36.43 ±0.20	36.30/36.19 ±0.16/±0.21	90.38/36.08 ±0.08/±0.19	<u>91.97</u>/90.83 ±0.05/±0.08	91.59 ±0.06	88.62 ±0.14	**92.28** ±0.05

Table 2: Node classification results (mean accuracy±std, %) with 0.4 observed rate. Best and sub-optimal results are highlighted in **bold** and <u>underlined</u>, respectively.

attribute-expanded graph by integrating three imputed attributed graphs with additional edges connecting identical nodes, as demonstrated in Fig. 1 (d). Additionally, the missing attribute-expanded graph is evaluated by applying the same two-layer GCN to its subgraph of attribute-missing nodes and associated recovered attributes, this setting is referred to as MIGDiff-expanded in the experiments. Similarly, for VAE-MI, SAT-MI, and ITR-MI, their evaluation experiments on the attribute-expanded graph are referred to as VAE-expanded, SAT-expanded, and ITR-expanded respectively in the experiments. The training and test sets were then divided following (Chen et al. 2022).

Performance Comparison

The performances of attribute reconstruction are shown in Table 1. Firstly, compared with classical methods, methods designed for attribute imputation show better performances, it demonstrates that missing attributes do degrades the performance of graph representation learning. Secondly, compared to advanced methods designed for single-attribute imputation, several multi-attribute imputation approaches yield higher Recall and NDCG scores. For instance, the VAE-MI extension improves Recall@10 on the Cora dataset by 40.7% over the baseline VAE from (Chen et al. 2022), which relies solely on neighbor aggregation and does not fully exploit the generative capacity of VAE. However, SAT-MI and ITR-MI perform worse than their single-attribute counterparts (SAT and ITR), primarily due to the limitations introduced by the VAE component. Specifically, VAE imposes a prior distribution constraint on the latent space, which forces the learned embedding distribution to remain close to this prior. This restriction reduces the model's flexibility in capturing the true underlying distribution of the data. In contrast, our proposed method, MIGDiff, leverages a diffusion-based generative framework that denoises missing attributes through graph-aware generation without requiring any prior assumptions. This enables it to more accurately recover the target attribute distribution and leads to superior performance across all evaluation metrics.

The node classification performance are shown in Table 2, MIGDiff demonstrates remarkable superiority than compared methods across most datasets with both MLP and GCN classifiers. These results confirm that high-quality imputed attributes produced by MIGDiff effectively support downstream classification tasks. Moreover, the results of MIGDiff-expanded and three multi-imputation methods on the attribute-expanded graph with three-attributes imputations , are illustrated in Fig. 5. MIGDiff-expanded achieves the highest accuracy across all scenarios, with substantial improvements over the baseline methods. Compared with results in Table. 2, MIGDiff-expanded achieves higher accuracy than the original MIGDiff. These results indicate that augmenting the original graph with enriched attribute information from multiple imputations can lead to improved performance on downstream tasks.

Figure 5: Comparison results by attributed graph expansion with three-attributes imputations

Ablation Study

The effectiveness of MIGDiff in recovering missing node attributes lies in its ability to fully leverage both structural and attribute information of observed nodes, guided by the underlying graph topology. To assess the contribution of each component, we design three ablation experiments: removing

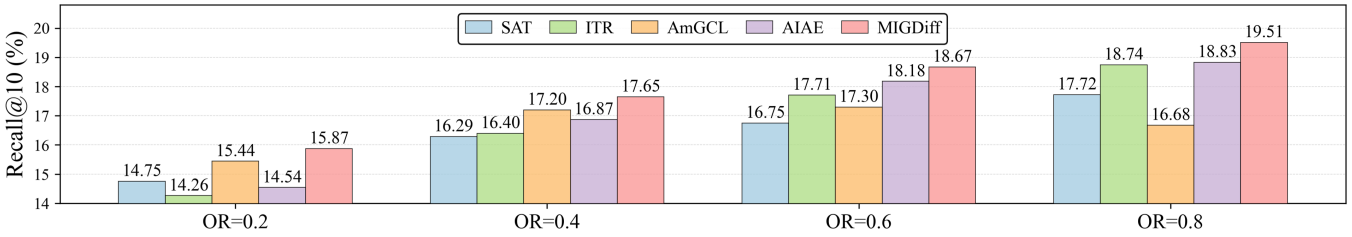

Figure 6: Performance of attribute reconstruction with different observed rate (OR).

Experiments	Recall@10/20/50	NDCG@10/20/50
MIGDiff-af	10.78/15.63/25.76	15.09/18.34/23.66
MIGDiff-a	13.82/20.24/32.25	19.66/23.90/30.30
MIGDiff-f	17.37/24.50/36.79	24.23/28.99/35.03
MIGDiff	**17.65/25.02/37.71**	**24.49/29.45/36.15**

Table 3: The ablation study of MIGDiff on Cora.

Dataset	Method	2 times	3 times	4 times	5 times
Cora	VAE-MI	0.0038	0.0048	0.0054	0.0057
	SAT-MI	0.0043	0.0056	0.0064	0.0068
	ITR-MI	0.0243	0.0315	0.0354	0.0378
	MIGDiff	**1.8388**	**2.7019**	**2.8678**	**2.9455**
Citeseer	VAE-MI	0.0027	0.0035	0.0040	0.0043
	SAT-MI	0.0063	0.0084	0.0095	0.0101
	ITR-MI	0.0354	0.0479	0.0542	0.0581
	MIGDiff	**0.0562**	**0.0864**	**0.0919**	**0.0918**

Table 4: Diversity analysis results with different number of attribute imputations (k)

attribute embeddings (MIGDiff-a), removing structural embeddings (MIGDiff-f), and removing both while replacing the graph auto-encoder with a simple MLP (MIGDiff-af).

Table 3 reveals the ablation study results of MIGDiff. Excluding structural embeddings (MIGDiff-a) causes a sharp 21.7% drop in Recall, highlighting the critical role of graph topology in guiding attribute imputation. Structural embeddings capture node relationships, enabling effective information propagation, which is essential for accurate imputation. In contrast, removing only attribute embeddings (MIGDiff-f) leads to a smaller decrease in Recall (1.6%), indicating that while attribute information enriches the imputation, its impact is less significant than structural information. The most severe performance drop occurs in MIGDiff-af, where both embeddings are removed and the graph auto-encoder is replaced by an MLP, resulting in a 38.9% decrease in Recall. This underscores the necessity of both structural and attribute information, and the vital role of the graph-specific encoder in achieving optimal performance.

Diversity Analysis

To evaluate the diversity of multiple attribute imputations generated by the comparison methods and MIGDiff, we conduct the following experiment on the Cora and Citeseer datasets. For each model, we perform k ($k \in \{2, 3, 4, 5\}$) independent samplings of the imputation process, thereby obtaining k distinct imputed attributes for each node. We then compute the variance of these k vectors for each feature dimension and average over all dimensions to obtain a single diversity score.

The results are summarized in Table 4, it is evident that MIGDiff produces substantially higher average variance across multiple attribute imputations compared with VAE-MI, SAT-MI and ITR-MI. For example, on Cora with five samplings, MIGDiff achieves an average variance of 2.9455, which is two orders of magnitude greater than the second best method (ITR-MI at 0.0378). On Citeseer, MIGDiff similarly outperforms all baselines, further validating its superior capacity to generate diverse imputations. Consequently, the attributes imputed by MIGDiff exhibit substantial diversity, better capturing richer variations within data distribution. Moreover, the diversity score increases steadily with the number of sampled imputations ($k = \{1, \ldots, 5\}$).

Sensitivity Analysis of the Observed Ratio

MIGDiff was evaluated on Cora with observed rates of 0.2, 0.4, 0.6 and 0.8 to assess its robustness and effectiveness across varying degrees of attribute missingness. As Figure 6 shows, MIGDiff consistently demonstrates competitive performance across all observation rates. Notably, even when only 20% of attributes are observed, it slightly outperforms the second-best baseline. As the observation rate increases, the performance gain becomes more pronounced, indicating that MIGDiff is highly effective at recovering missing attributes, even under severely limited observed information.

Conclusion

In this paper, we addressed the critical challenge of multi-attributes imputations in graphs by proposing **MIGDiff**, a novel denoising diffusion imputation model tailored for attribute-missing graph. Based on the diffusion model, MIGDiff generates diverse and high-quality attribute imputations by leveraging the integrated topological and attribute information to improve the denoising process of nodes without attributes according to graph connectivity. Through extensive experiments on four public datasets, MIGDiff demonstrated superior performances in both node classification and attribute reconstruction task.

Acknowledgments

This work was supported in part by the National Science and Technology Major Project under Grant 2022ZD0117700, the National Natural Science Foundation of China (62306118, 62325204), the Fundamental Research Funds for the Central Universities (2025ZYGXZR054).

References

Chen, L.; Gong, S.; Bruna, J.; and Bronstein, M. 2019. Attributed random walk as matrix factorization. In *neural information processing systems, graph representation learning workshop*.

Chen, X.; Chen, S.; Yao, J.; Zheng, H.; Zhang, Y.; and Tsang, I. W. 2022. Learning on Attribute-Missing Graphs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(2): 740–757.

Hamilton, W.; Ying, Z.; and Leskovec, J. 2017. Inductive representation learning on large graphs. *Advances in neural information processing systems*, 30.

Ho, J.; Jain, A.; and Abbeel, P. 2020. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33: 6840–6851.

Hu, L.; Jian, S.; Cao, L.; Gu, Z.; and Amirbekyan, A. 2019. HERS: Modeling Influential Contexts with Heterogeneous Relations for Sparse and Cold-Start Recommendation. 3830–3837.

Huang, X.; Song, Q.; Li, Y.; and Hu, X. 2019. Graph recurrent networks with attributed random walks. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 732–740.

Kipf, T. N.; and Welling, M. 2016. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*.

Krasnova, H.; Spiekermann, S.; Koroleva, K.; and Hildebrand, T. 2010. Online social networks: why we disclose. *Journal of Information Technology*, 25(2): 109–125.

Liu, Y.; Ajanthan, T.; Husain, H.; and Nguyen, V. 2024. Self-supervision improves diffusion models for tabular data imputation. In *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management*, 1513–1522.

McCallum, A. K.; Nigam, K.; Rennie, J.; and Seymore, K. 2000. Automating the construction of internet portals with machine learning. *Information Retrieval*, 3: 127–163.

Moepya, S. O.; Akhoury, S. S.; Nelwamondo, F. V.; and Twala, B. 2016. THE ROLE OF IMPUTATION IN DETECTING FRAUDULENT FINANCIAL REPORTING. *International Journal of Innovative Computing Information and Control*, 12(1): 333–356.

Sen, P.; Namata, G.; Bilgic, M.; Getoor, L.; Galligher, B.; and Eliassi-Rad, T. 2008. Collective classification in network data. *AI magazine*, 29(3): 93–93.

Shchur, O.; Mumme, M.; Bojchevski, A.; and Günnemann, S. 2018. Pitfalls of graph neural network evaluation. *arXiv preprint arXiv:1811.05868*.

Şimşek, Ö.; and Jensen, D. 2008. Navigating networks by using homophily and degree. *Proceedings of the National Academy of Sciences*, 105(35): 12758–12762.

Tu, W.; Zhou, S.; Liu, X.; Liu, Y.; Cai, Z.; Zhu, E.; Zhang, C.; and Cheng, J. 2022. Initializing Then Refining: A Simple Graph Attribute Imputation Network. In *IJCAI*, 3494–3500.

Veličković, P.; Cucurull, G.; Casanova, A.; Romero, A.; Lio, P.; and Bengio, Y. 2017. Graph attention networks. *arXiv preprint arXiv:1710.10903*.

Xia, J.; Zhang, S.; Cai, G.; Li, L.; Pan, Q.; Yan, J.; and Ning, G. 2017. Adjusted weight voting algorithm for random forests in handling missing values. *Pattern Recognition*, 69: 52–60.

Xia, R.; Zhang, C.; Li, A.; Liu, X.; and Yang, B. 2024. Attribute imputation autoencoders for attribute-missing graphs. *Knowledge-Based Systems*, 291: 111583.

Zhang, M.; and Chen, Y. 2020. Inductive Matrix Completion Based on Graph Neural Networks. In *International Conference on Learning Representations*.

Zhang, X.; Li, M.; Wang, Y.; and Fei, H. 2023. Amgcl: Feature imputation of attribute missing graph via self-supervised contrastive learning. *arXiv preprint arXiv:2305.03741*.