Adversarial Masked Graph Modeling for Robust Graph Autoencoders

Qiqi Zhang[1], Chuanjin Liu[1], Gen Liu[1], Chao Li[2], Zhongying Zhao[1*]
[1]College of Computer Science and Engineering, Shandong University of Science and Technology, Qingdao 266590, China
[2]College of Electronic and Information Engineering, Shandong University of Science and Technology, Qingdao 266590, China
zqqstu0818@163.com, liuchuanjin2026@163.com, sdust_lg@163.com, lichao@sdust.edu.cn, zzysuin@163.com

Abstract

Graph Masked AutoEncoders (GMAEs) have achieved notable success in diverse downstream tasks. However, their superior performance usually relies on reliable and unperturbed input graphs. In real-world scenarios, graphs are vulnerable to adversarial attacks (e.g., perturbations on node features or topology). As a result, existing GMAEs tend to overfit these perturbed inputs, leading to poor robustness. Motivated by this, we propose an **A**dversarial masked g**r**aph **mo**deling for **r**obust **G**raph **A**uto**E**ncoders, named **ArmorGAE**. Specifically, we first design a *dual-consistency graph augmentation* method that enriches the graph topology to mitigate structural vulnerabilities. Theoretically, we establish a *connection between ArmorGAE and contrastive learning*. It demonstrates that our augmentation strategy in GMAEs is equivalent to effectively expanding the set of high-quality positive samples in contrastive learning, providing richer and more reliable supervisory signals. We then present an *adversarial masked autoencoder* to mitigate model overfitting to the perturbed graph. The generator and discriminator are jointly trained under an adversarial paradigm, where the former aims to recover the underlying true data distribution, while the latter distinguishes reconstructed and real edges. Experimental results on five datasets demonstrate that ArmorGAE outperforms state-of-the-art methods in terms of classification accuracy on clean graphs and adversarial robustness under 16 distinct adversarial scenarios. The codes are publicly available at https://github.com/ZZY-GraphMiningLab/ArmorGAE.

1 Introduction

Graph Masked AutoEncoders (GMAEs) have emerged as a prominent self-supervised method for graph representation learning [Hou *et al.*, 2022; Shen *et al.*, 2025], achieving notable success in diverse scenarios such as traffic prediction and anomaly detection [Ma *et al.*, 2024; Liu *et al.*, 2024].

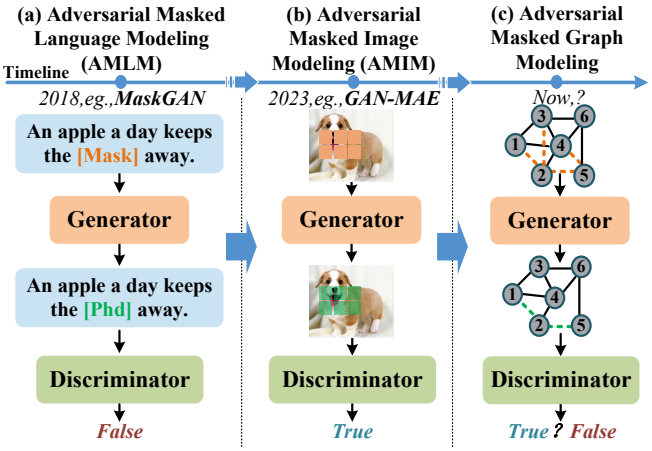

Figure 1: From left to right: illustrative examples of (a) adversarial masked language modeling (AMLM), (b) adversarial masked image modeling (AMIM), and (c) adversarial masked graph modeling.

Their goal is to reconstruct the masked portions of node features or edges from partial observations [Hou *et al.*, 2023; Li *et al.*, 2023].

Despite their satisfactory performance, most existing GMAEs rely on the assumption that input graphs are reliable and unperturbed. In practice, this assumption does not always hold, especially in adversarial attack scenarios [Dai *et al.*, 2018; Sun *et al.*, 2022]. As a result, GMAEs are not robust enough and are still facing the following two challenges:

(1) Overfitting to Adversarial Noises. Existing GMAEs typically rely on simple reconstruction objectives (e.g., Mean Squared Error (MSE) [Hou *et al.*, 2022] or Binary Cross Entropy (BCE) [Jung and Park, 2025]). These methods indiscriminately treat all observed edges as ground truth [Hu *et al.*, 2025; Hu *et al.*, 2024]. It causes the model to fit spurious statistical correlations rather than distinguishing inherent structural signals from adversarial attacks [Zügner *et al.*, 2018; Ganin *et al.*, 2016].

(2) Neglect of Implicit Structural Guidance. Current GMAEs are often limited to recovering masked components based solely on visible graph [Huo *et al.*, 2023; Zheng *et al.*, 2018]. They fail to capture latent correlations between nodes and their unconnected yet semantically similar counterparts. As a result, these models lack sufficient contextual guidance

*Corresponding Author.

to yield robust embeddings, especially when the graph structure is subjected to adversarial attacks [Xia *et al.*, 2025].

To address these challenges, we draw inspiration from the adversarial masking paradigm, which has proven effective in natural language processing [Fedus *et al.*, 2018] and computer vision [Fei *et al.*, 2023], as illustrated in Figure 1. These methods enhance model performance by reconstructing masked parts from inputs that are both partially masked and adversarially perturbed. It motivates a compelling yet unexplored question: **Can the adversarial masking modeling paradigm effectively enhance the robustness of graph masked autoencoders?**

To this end, we propose **ArmorGAE**, an **A**dversarial masked g**r**aph **mo**deling method for **r**obust **G**raph **A**uto**E**ncoders. We establish the connection between masked autoencoders and contrastive learning, demonstrating that our strategy is theoretically equivalent to expanding the positive samples in implicit contrastive learning. Specifically, we first enrich the graph topology by adding semantically relevant edges based on feature similarity and degree constraint. We then design an adversarial masked autoencoder to alleviate the risk of overfitting to perturbed graphs. Within this adversarial framework, the generator and discriminator are jointly optimized. The former aims to recover the underlying true data distribution, while the latter distinguishes reconstructed and real edges. The contributions of this paper are as follows.

- **New Method:** We propose ArmorGAE, the first adversarial masked graph modeling method that synergizes adversarial training and graph masked autoencoders to enhance the robustness of learned representations.
- **Theoretical Analysis:** We present a dual-consistency augmentation method and prove its equivalence to expanding positive samples in implicit contrastive learning, which provides enriched supervisory signals.
- **Superior Performance:** Extensive experiments demonstrate that ArmorGAE significantly outperforms state-of-the-art baselines under various adversarial attacks, confirming its superior performance and robustness.

2 Related Work

Graph Masked Autoencoders. Recent advancements in GMAEs have established them as a powerful generative self-supervised paradigm. These methods are typically categorized into feature reconstruction (*FeatRec*) and topology reconstruction (*TopoRec*). In the realm of *FeatRec*, GraphMAE [Hou *et al.*, 2022] introduced scaled cosine error and remask decoding to stabilize generative training, while GraphMAE2 [Hou *et al.*, 2023] further refined this via multi-view latent prediction. AUGMAE [Wang *et al.*, 2024] attempted to increase task difficulty using an adversarial masking strategy. Simultaneously, *TopoRec* methods such as MaskGAE [Li *et al.*, 2023] and S2GAE [Tan *et al.*, 2023] focused on recovering missing edges to capture structural dependencies. Bandana [Zhao *et al.*, 2024] further improved structural awareness via layer-wise bandwidth prediction. Despite their notable success, these methods assume the availability of clean graphs, leaving them vulnerable to structural noise and incomplete connectivity.

Adversarial Training. Adversarial training on graphs has traditionally focused on the interplay between attacks and defenses [Sun *et al.*, 2022; Chang *et al.*, 2020]. Attack methods like Nettack [Zügner *et al.*, 2018] and Metattack [Sun *et al.*, 2020] have exposed the fragility of GNNs by perturbing graph spectra or poisoning training structures. To counter these, adversarial training has emerged as a powerful paradigm for learning robust representations. In computer vision, recent studies have demonstrated that integrating adversarial training into the MAE framework effectively enhances adversarial robustness [Fei *et al.*, 2023]. In the graph domain, ARGA [Pan *et al.*, 2018] regularized the latent space against a prior distribution to prevent representation collapse. However, the intersection of generative self-supervised learning and adversarial training on graphs remains underexplored.

3 Proposed Method

3.1 Preliminaries

Definition 1. Attributed Graph [Zhao *et al.*, 2022]. Let $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathbf{X})$ be an attributed graph, where $\mathcal{V} = \{v_1, v_2, \ldots, v_N\}$ is the set of nodes, and $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ is the set of edges. The node feature matrix is denoted as $\mathbf{X} \in \mathbb{R}^{N \times D}$, where $\mathbf{x}_v \in \mathbb{R}^D$ represents the attribute vector of node v. The structure is represented by an adjacency matrix $\mathbf{A} \in \{0,1\}^{N \times N}$, where $\mathbf{A}_{vu} = 1$ if $e_{vu} \in \mathcal{E}$, and $\mathbf{A}_{vu} = 0$ otherwise.

Definition 2. Adversarial Training [Ganin *et al.*, 2016]. Adversarial training aims to enhance model robustness by formulating the learning process as a min-max optimization problem. For a model f_θ, the objective is defined as:

$$\min_{\theta} \mathbb{E}_{\mathcal{G} \sim \mathcal{D}} \left[\max_{\mathcal{G}' \in \Phi(\mathcal{G})} \mathcal{L}(f_\theta(\mathcal{G}'), \mathcal{G}) \right], \tag{1}$$

where $\mathcal{G}$ represents the input graph and $\mathcal{G}' \in \Phi(\mathcal{G})$ denotes a perturbed graph within a feasible constraint set Φ (e.g., structural perturbations or feature noise). The inner maximization identifies the worst-case perturbation example, while the outer minimization optimizes θ to maintain performance.

Problem. Robust Representation Learning. Given an attributed graph $\mathcal{G}$, the goal is to learn a graph encoder $f_\theta(\mathbf{X}, \mathbf{A})$ that maps nodes into the embedding space. The learned representations $H = f_\theta(X, A)$ should remain effective for downstream tasks, even when the graph is subjected to perturbations. In particular, we define this robustness as the ability to resist interference from noise and missing data, which is evaluated against structure, feature, and dual attacks under both targeted and non-targeted scenarios.

3.2 Theoretical Analysis

Recent studies suggest that generative methods, such as GraphMAE, can be interpreted as performing implicit context-level graph contrastive learning [Wang *et al.*, 2024]. Building upon this perspective, we theoretically demonstrate that our augmentation strategy is equivalent to strengthening the implicit GCL by expanding the set of positive samples.

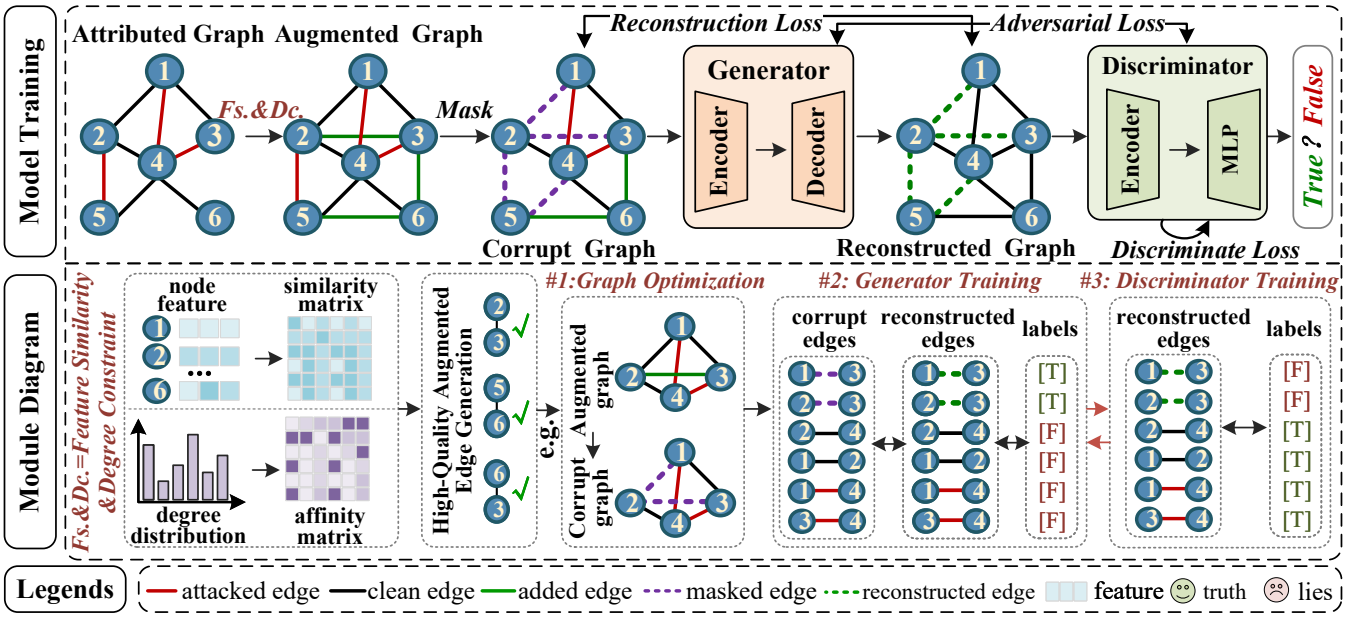

Figure 2: The framework of the proposed ArmorGAE.

Let the input graph be defined as $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathbf{X})$. Its augmented counterpart, $\mathcal{G}' = (\mathcal{V}, \widetilde{\mathcal{E}}, \mathbf{X})$, is generated via dual-consistency graph augmentation, where $\widetilde{\mathcal{E}} = \mathcal{E} \cup \mathcal{E}^+$.

Assumption 1. Edge reconstruction capability. For the augmented graph $\mathcal{G}'$ and any edge decoder g, we assume there exists a pseudo-inverse encoder f_g such that the resulting pseudo-autoencoder $h_g = g \circ f_g$ satisfies Eqn. (2).

$$\mathbb{E}_{(v_i,v_j)\in\widetilde{\mathcal{E}}}\left[\mathrm{Dist}\left(h_g(\mathbf{x}_i,\mathbf{x}_j), \widetilde{A}_{ij}\right)\right] \le \varepsilon^+, \tag{2}$$

where $h_g(\mathbf{x}_i, \mathbf{x}_j)$ denotes the predicted edge probability, $\widetilde{A}_{ij} \in \{0, 1\}$ is the ground-truth entry in the augmented graph, and $\mathrm{Dist}(\cdot)$ is a distance metric (e.g., mean squared error or cosine similarity). The parameter ε^+ serves as the upper bound of the expected reconstruction error.

The edge reconstruction loss on the augmented graph is formulated as shown in Eqn. (3).

$$\begin{aligned}\mathcal{L}^{\mathrm{Aug}}_{\mathrm{Struct}}(h) = -\mathbb{E}_{(v_i,v_j)\in\widetilde{\mathcal{E}}}\big[&\widetilde{A}_{ij}\log\sigma(s_{ij}) \\ &+(1-\widetilde{A}_{ij})\log(1-\sigma(s_{ij}))\big],\end{aligned} \tag{3}$$

where $s_{ij} = h(\boldsymbol{x}_i)^\top h(\boldsymbol{x}_j)$ represents the similarity score for a given node pair, and $\sigma(\cdot)$ denotes the sigmoid function.

To establish a connection with contrastive learning, we define a pretext alignment loss based on the theoretical framework in AUGMAE [Wang *et al.*, 2024]. The pretext loss is related to the context-level alignment loss in GCL as follows:

$$\mathcal{L}^{\mathrm{Aug}}_{\mathrm{Pretext}}(h) = -\mathbb{E}_{(v_i,v_j)\in\widetilde{\mathcal{E}}}\left[h_g(\boldsymbol{x}_i)^\top h(\boldsymbol{x}_j)\right], \tag{4}$$

$$\mathcal{L}^{\mathrm{Aug}}_{\mathrm{Pretext}}(h) \ge \frac{1}{2}\mathcal{L}^{\mathrm{GCL}}_{\mathrm{Align}}(h) + \mathrm{const}, \tag{5}$$

where $\mathcal{L}^{\mathrm{GCL}}_{\mathrm{Align}}(h)$ is the alignment objective enforcing representation consistency between positive nodes.

Lemma 1. Properties of the augmented graph. To ensure the quality of the augmented graph, we posit that the enriched edge set based on feature similarity and degree constraints must satisfy the following properties: (i) the number of edges increases, i.e., $|\widetilde{\mathcal{E}}| > |\mathcal{E}|$; (ii) each augmented edge $(v_i, v_j) \in \mathcal{E}^+$ exhibits high similarity, i.e., $Sim(\mathbf{x}_i, \mathbf{x}_j) \ge \tau$ for a threshold τ, where a relatively high τ is adopted to ensure the selection of high-confidence semantic neighbors; and (iii) the expected similarity score between connected representations satisfies Eqn. (6):

$$\begin{aligned}&\mathbb{E}_{(v_i,v_j)\in\mathcal{E}^+}\left[h_g(\boldsymbol{x}_i)^\top h(\boldsymbol{x}_j)\right] - \\ &\mathbb{E}_{(v_i,v_j)\in\mathcal{E}}\left[h_g(\boldsymbol{x}_i)^\top h(\boldsymbol{x}_j)\right] \ge \delta,\end{aligned} \tag{6}$$

where $\delta \ge 0$ is a small constant indicating that the reconstruction expectations of newly augmented edges are consistent with those of the original structural edges.

Theorem 1. Incorporating graph augmentation module in MAE is equivalent to introducing the high-quality positive samples in GCL. Under Assumption 1, minimizing the BCE loss for edge reconstruction on the augmented graph $\mathcal{G}'$ provides a lower bound for the pretext alignment loss:

$$\mathcal{L}^{\mathrm{Aug}}_{\mathrm{Struct}}(h) \ge \frac{\gamma}{2}\mathcal{L}^{\mathrm{Aug}}_{\mathrm{Pretext}}(h) - \frac{\gamma}{2}\varepsilon' + \mathrm{const}, \tag{7}$$

where $\gamma = 1$ and ε' is a function of bound ε^+. Please refer to Appendix A.2 for the detailed proof of Theorem 1.

3.3 Framework

In this section, we present an adversarial masked graph modeling method for robust graph autoencoders, named ArmorGAE. As illustrated in Figure 2, ArmorGAE comprises two key components: (i) Dual-Consistency Graph Augmentation and (ii) Adversarial Masked Graph Autoencoder.

3.4 Dual-Consistency Graph Augmentation

To enhance the input graph, we propose a dual-consistency graph augmentation method that supplements $\mathcal{G}$ with an auxiliary edge set $\mathcal{E}^+$, comprising node pairs with high feature and structure similarities.

Feature Similarity Score. To quantify the attitude-level proximity between node pairs that are not directly connected, we compute the cosine similarity using $s_f(v_i, v_j) = \frac{\mathbf{x}_i^\top \mathbf{x}_j}{\|\mathbf{x}_i\|\|\mathbf{x}_j\|}$, where $\mathbf{x}_i$ denotes the feature vector of node v_i.

We further apply min-max normalization to project these scores into the range [0, 1] according to Eqn. (8).

$$\tilde{s}_f(v_i, v_j) = \frac{s_f(v_i, v_j) - \min(\mathbf{S}_f)}{\max(\mathbf{S}_f) - \min(\mathbf{S}_f) + \epsilon}, \tag{8}$$

where $\mathbf{S}_f = \{s_f(v_i, v_j) \mid (v_i, v_j) \notin \mathcal{E}\}$ represents the ensemble of raw similarity scores for all candidate pairs, and ϵ is a small constant for numerical stability.

Degree Constraint Score. We define a structural affinity score to capture local connectivity patterns. For each node v_i, we compute its normalized degree to represent its relative importance within the topology $\tilde{d}_i = \frac{\deg(v_i)}{\max(\mathbf{d})+\epsilon}$, where $\mathbf{d}$ is the vector containing degrees of all nodes in $\mathcal{V}$.

To quantify the structural proximity between two non-adjacent nodes, we calculate the product of their normalized degrees, formulated as Eqn. (9).

$$s_d(v_i, v_j) = \begin{cases} \tilde{d}_i \cdot \tilde{d}_j & \text{if } i \neq j \text{ and } (v_i, v_j) \notin \mathcal{E} \\ 0 & \text{otherwise} \end{cases}. \tag{9}$$

Subsequently, we apply min-max normalization to obtain the final structural score $\tilde{s}_d(v_i, v_j)$.

Finally, we integrate the normalized feature and structural consistency metrics into a unified augmentation score:

$$s_{\text{aug}}(v_i, v_j) = \tilde{s}_f(v_i, v_j) + \tilde{s}_d(v_i, v_j). \tag{10}$$

We then rank all non-adjacent node pairs in descending order of s_{aug}. The number of added edges is determined by an augmentation ratio p, such that $|\mathcal{E}^+| = \lfloor p \cdot |\mathcal{E}| \rfloor$. Finally, the augmented graph is constructed as $\mathcal{G}_{\text{aug}} = (\mathcal{V}, \mathcal{E} \cup \mathcal{E}^+, \mathbf{X})$.

3.5 Adversarial Masked Graph Autoencoder

To enhance the robustness of node representations, we design an adversarial masked graph autoencoder that engages a generator and a discriminator in a min-max game.

Generator: Masked Graph Autoencoder. The generator Gen is designed as a masked graph autoencoder aimed at reconstructing the structure of $\mathcal{G}_{\text{aug}}$. Specifically, given $\mathcal{G}_{\text{aug}} = (\mathcal{V}, \widetilde{\mathcal{E}}, \mathbf{X})$, we randomly mask a part of edges from $\widetilde{\mathcal{E}}$ with a masking ratio m_e. This process partitions the edge set into a remaining set $\widetilde{\mathcal{E}}_{\text{remain}}$ and a masked set $\widetilde{\mathcal{E}}_{\text{mask}}$.

Subsequently, the encoder f_{enc} maps the node features and the observed structure into latent representations.

$$\mathbf{H}^t = f_{\text{enc}}(\mathbf{X}, \widetilde{\mathcal{E}}_{\text{remain}}). \tag{11}$$

To recover the masked structure, the edge decoder calculates the existence probability $\hat{p}_{uv}$ for any node pair as shown in Eqn. (12).

$$\hat{p}_{uv} = \sigma\left(\text{MLP}(\mathbf{h}_u^t \odot \mathbf{h}_v^t)\right), \tag{12}$$

where $\odot$ denotes the Hadamard product.

The generator is optimized by minimizing the loss $\mathcal{L}_{\text{recon}}^e$.

$$\begin{aligned} \mathcal{L}_{\text{recon}}^e = &- \frac{1}{|\widetilde{\mathcal{E}}_{\text{mask}}|} \sum_{(v_i,v_j)\in\widetilde{\mathcal{E}}_{\text{mask}}} \log \hat{p}_{uv} \\ &- \frac{1}{|\widetilde{\mathcal{E}}_{\text{neg}}|} \sum_{(v_i,v_j)\in\widetilde{\mathcal{E}}_{\text{neg}}} \log(1-\hat{p}_{uv}), \end{aligned} \tag{13}$$

where $\widetilde{\mathcal{E}}_{\text{neg}}$ is a set of randomly sampled negative edges.

Discriminator: Edge-Level Discriminator. We introduce an edge-level discriminator $Disc$ to distinguish between real edges and those reconstructed by the generator. $Disc$ is a weight-aware GCN that operates on a synthetic graph structure $\widetilde{\mathcal{E}} = \widetilde{\mathcal{E}}_{\text{remain}} \cup \widetilde{\mathcal{E}}_{\text{mask}}$.

The discriminator learns node embeddings via $\mathbf{H} = \text{GCN}(\mathbf{X}, \widetilde{\mathcal{E}}, \mathbf{w}_{\text{disc}})$, where $\mathbf{w}_{\text{disc}}$ is a composite confidence vector. For an observed edge $e \in \widetilde{\mathcal{E}}_{\text{remain}}$, its weight is fixed to 1; for a reconstructed edge $e \in \widetilde{\mathcal{E}}_{\text{mask}}$, its weight is assigned as its predicted probability $\hat{p}_{uv}$. It enables the discriminator to perceive generative quality with a fine-grained sensitivity. Finally, $Disc$ computes the realism logit s_{uv} for each edge.

$$s_{uv} = \text{MLP}([\mathbf{h}_u \| \mathbf{h}_v]) + b, \tag{14}$$

where $\|$ is the concatenation operation, b is a learnable bias.

Generator Training. The generator Gen is optimized to produce high-fidelity reconstructions that 'fool' the discriminator while maintaining structural integrity.

$$\mathcal{L}_G = \mathcal{L}_{\text{recon}}^e + \lambda_{adv}\mathcal{L}_G^{\text{adv}}, \tag{15}$$

where $\mathcal{L}_G^{\text{adv}} = -\mathbb{E}_{e\in\widetilde{\mathcal{E}}_{\text{mask}}}[\log(\sigma(s_e))]$ represents the adversarial loss that compels the reconstructed edges to be indistinguishable from the original ones in $Disc$.

Discriminator Training. We assign binary labels $y_e = 1$ for observed edges $e \in \widetilde{\mathcal{E}}_{\text{remain}}$ and $y_e = 0$ for reconstructed edges $e \in \widetilde{\mathcal{E}}_{\text{mask}}$. The discriminator $Disc$ is optimized to minimize a confidence-weighted binary cross-entropy loss:

$$\begin{aligned} \mathcal{L}_D = -\frac{1}{|\widetilde{\mathcal{E}}|}\sum_{e\in\widetilde{\mathcal{E}}} \alpha_e \Big[& y_e \log(\sigma(s_e)) \\ & + (1-y_e)\log(1-\sigma(s_e)) \Big], \end{aligned} \tag{16}$$

where α_e is the confidence weight associated with edge e.

4 Experiments

In this section, we conduct extensive experiments to evaluate the robustness of ArmorGAE. We aim to answer the following **R**esearch **Q**uestion**s** (RQs):

Datasets	Nodes	Edges	Features	Classes
Cora	2,485	5,069	1,433	7
Citeseer	2,110	3,668	3,703	6
Pubmed	19,717	44,338	500	3
Amazon-Photo	7,650	119,081	745	8
Amazon-Computers	13,752	245,861	767	10

Table 1: The statistics of datasets used in this paper.

Datasets	Methods	Clean	Poisoning (Acc.)					Evasive (Acc.)				
			5%	10%	15%	20%	25%	5%	10%	15%	20%	25%
Cora	GraphMAE'22	81.21±0.11	79.07±0.39	71.66±0.52	66.88±0.21	53.38±0.58	49.44±0.75	80.12±0.09	72.94±0.13	71.68±0.09	63.23±0.18	53.92±0.13
	MaskGAE'23	85.05±0.13	80.65±0.64	75.67±0.33	67.87±0.32	54.26±1.17	52.05±0.55	65.36±0.84	60.90±0.83	51.54±1.23	45.81±1.05	41.71±1.35
	Bandana'24	84.95±0.61	79.74±1.15	75.45±1.08	68.29±0.85	55.57±0.92	51.63±0.88	79.22±0.98	74.01±0.75	67.45±1.12	53.67±1.04	49.45±1.31
	AUGMAE'24	83.25±0.00	77.73±0.35	72.30±0.33	65.10±0.31	54.72±0.45	48.80±0.38	80.35±1.00	**78.03±0.29**	73.27±0.28	**66.08±0.35**	**60.76±0.57**
	DGMAE'25	83.45±0.08	78.48±0.16	74.08±0.11	68.31±0.15	57.85±0.10	50.68±0.19	81.05±0.11	77.45±0.13	74.12±0.17	64.70±0.14	58.47±0.12
	ArmorGAE	**85.53±0.32**	**81.25±0.33**	**77.16±0.51**	**71.37±0.45**	**59.62±1.54**	**52.28±1.37**	**81.87±0.23**	77.16±0.49	**75.16±1.35**	60.81±1.47	57.79±0.44
Citeseer	GraphMAE'22	70.80±0.21	68.85±0.32	67.90±0.36	62.59±0.38	54.44±0.42	53.95±0.82	69.75±0.09	68.84±0.09	65.52±0.29	63.39±0.22	61.61±0.09
	MaskGAE'23	75.29±0.62	74.38±0.43	72.35±0.51	71.46±0.77	63.63±1.24	62.69±0.58	73.31±0.73	70.11±1.01	70.28±0.60	55.91±1.78	64.77±0.81
	Bandana'24	72.96±0.28	73.13±1.31	71.90±0.50	69.60±0.81	63.93±0.57	**64.15±1.61**	72.69±0.67	71.30±0.98	69.44±0.90	60.02±0.96	62.07±2.51
	AUGMAE'24	73.25±0.22	71.47±0.20	66.02±0.41	66.18±0.36	57.61±0.23	56.62±0.33	72.59±0.16	71.74±0.17	71.02±0.29	**66.85±0.32**	**66.72±0.27**
	DGMAE'25	74.82±0.00	72.93±0.00	71.27±0.00	69.37±0.00	57.70±0.00	59.60±0.00	73.82±0.00	**73.10±0.00**	70.26±0.00	64.16±0.00	65.88±0.00
	ArmorGAE	**76.05±0.15**	**75.59±0.14**	**73.50±0.12**	**71.94±0.32**	**66.04±0.43**	64.14±1.65	**74.97±0.10**	73.02±0.70	**72.32±0.36**	64.78±0.63	66.18±0.58
Pubmed	GraphMAE'22	79.10±0.01	70.27±0.08	73.80±0.47	69.24±0.38	65.76±0.26	59.61±0.60	80.86±0.16	78.94±0.12	76.44±0.19	74.04±0.32	71.05±0.36
	MaskGAE'23	84.50±0.02	80.55±0.03	77.32±0.02	75.64±0.09	75.37±0.08	72.89±0.08	81.60±0.08	78.48±0.04	76.10±0.06	**74.23±0.07**	**72.28±0.04**
	Bandana'24	84.93±0.11	82.24±0.08	79.63±0.08	77.51±0.06	74.83±0.11	72.63±0.07	80.90±0.09	78.25±0.18	75.90±0.20	73.46±0.18	71.60±0.07
	AUGMAE'24	82.47±0.02	80.59±0.27	79.60±0.01	77.81±0.35	76.39±0.69	74.07±0.62	81.00±0.02	78.93±0.02	76.30±0.03	73.60±0.10	71.13±0.04
	DGMAE'25	83.51±0.01	**82.68±0.03**	**81.73±0.05**	**81.55±0.05**	**79.71±0.06**	**77.85±0.04**	79.83±0.12	**79.83±0.12**	**77.70±0.20**	73.90±0.03	71.30±0.03
	ArmorGAE	**85.19±0.06**	82.14±0.06	80.93±0.09	78.73±0.21	77.56±0.11	76.08±0.15	**81.73±0.09**	79.38±0.13	76.93±0.14	73.91±0.11	72.27±0.14
Photo	GraphMAE'22	85.06±3.56	69.39±17.32	77.93±8.17	80.98±1.72	80.52±1.96	79.77±2.58	79.28±12.83	77.99±13.28	76.78±9.89	77.77±10.35	77.09±11.55
	MaskGAE'23	93.10±0.23	90.10±0.15	89.76±0.14	89.00±0.22	88.52±0.27	88.00±0.25	90.25±0.18	89.75±0.12	89.47±0.21	89.20±0.13	**89.08±0.15**
	Bandana'24	93.23±0.07	90.23±0.25	88.69±0.09	88.58±0.14	88.63±0.12	88.42±0.19	90.25±0.16	89.46±0.27	88.41±0.14	88.89±0.21	89.01±0.13
	AUGMAE'24	93.10±0.36	**90.49±0.28**	89.63±0.39	89.02±0.24	88.23±0.82	88.15±0.51	**90.63±0.35**	89.79±0.47	**89.72±0.40**	89.02±0.52	88.85±0.49
	DGMAE'25	92.58±0.66	90.47±0.60	89.81±0.79	**89.35±0.88**	88.76±1.27	88.49±1.38	90.23±0.50	89.39±0.40	89.23±0.56	88.83±0.46	88.61±0.51
	ArmorGAE	**93.46±0.26**	89.77±0.60	**90.07±0.23**	89.33±0.32	**89.17±0.38**	**89.11±0.26**	90.34±0.31	**89.87±0.26**	89.69±0.24	**89.47±0.25**	88.93±0.27
Computers	GraphMAE'22	84.06±3.56	64.72±6.33	65.94±4.02	64.78±3.74	62.04±3.51	61.60±3.24	70.10±1.95	68.63±1.78	67.87±1.74	67.21±1.83	66.02±1.23
	MaskGAE'23	89.41±0.23	86.48±0.15	86.40±0.14	86.00±0.09	85.56±0.10	84.87±0.09	86.32±0.12	85.77±0.18	85.51±0.19	85.00±0.26	84.83±0.22
	Bandana'24	**89.57±0.13**	86.04±0.09	85.69±0.19	85.48±0.21	84.60±0.25	85.80±0.17	86.23±0.25	85.69±0.09	**85.58±0.14**	84.63±0.12	84.42±0.19
	AUGMAE'24	87.70±0.59	85.37±0.26	83.28±0.42	77.93±0.47	75.97±0.36	72.44±0.92	84.95±0.37	83.98±0.18	83.62±0.28	83.06±0.89	82.54±0.41
	DGMAE'25	87.50±0.41	84.62±1.72	81.84±3.62	82.98±1.56	82.78±0.58	80.79±2.75	85.08±0.55	84.01±0.45	83.37±0.42	82.77±0.39	82.23±0.55
	ArmorGAE	89.53±0.26	**89.35±0.13**	**88.17±0.04**	**88.38±0.27**	**88.63±0.14**	**88.25±0.16**	**86.66±0.31**	**86.08±0.33**	85.49±0.28	**85.34±0.21**	**85.09±0.30**

Table 2: Node classification accuracy (Acc.) of ArmorGAE and five competitive methods under non-targeted attack (Metattack) in poisoning and evasive scenarios. The best result is highlighted in **bold**, and the runner-up is underlined.

- **RQ1**: How robust is ArmorGAE against structural attacks in poisoning and evasive scenarios?
- **RQ2**: Can ArmorGAE maintain stability under feature and dual attacks?
- **RQ3**: What are the contributions of different components to the performance of ArmorGAE?
- **RQ4**: What is the computational efficiency of ArmorGAE in terms of training time?
- **RQ5**: How do key hyperparameters influence the performance of ArmorGAE?

4.1 Datasets and Baselines

We evaluate the robustness of ArmorGAE on five real-world datasets, including citation networks (Cora, Citeseer, and Pubmed) [Sen *et al.*, 2008], and Amazon co-purchase networks (Photo and Computers) [Shchur *et al.*, 2018]. The statistical results of the datasets are shown in Table 1.

To demonstrate the superiority of ArmorGAE, we compare it with five competitive baselines: GraphMAE [Hou *et al.*, 2022] focuses on node feature masking and reconstruction to capture node representations, while MaskGAE [Li *et al.*, 2023] centers on edge masking and reconstruction to learn topological patterns. Bandana [Zhao *et al.*, 2024] advances these approaches by considering the continuity and context of masked information to enhance representation coherence. For robustness, AUGMAE [Wang *et al.*, 2024] introduces an adversarial masking strategy to generate hard-to-align pairs with a uniformity regularizer, and DGMAE [Zheng *et al.*, 2025] employs an adaptive discrepancy selection module to preserve discriminative node information.

4.2 Implementation Details

We implement ArmorGAE using PyTorch, with training accelerated on an NVIDIA RTX 4060 GPU (12GB). All models are evaluated using classification accuracy (Acc.) under both clean and adversarial scenarios, including structural, feature, and dual attacks. For structural robustness, we employ Metattack with perturbation rates ranging from 5% to 25% and Nettack with perturbations per node varying in $\{1, 2, 3\}$. Regarding feature robustness, we introduce Gaussian noise $\boldsymbol{\delta} \sim \mathcal{N}(0, \mathbf{I})$ to node features [Shen *et al.*, 2025]. To ensure statistical significance, we report the average performance over ten independent trials with different random seeds.

4.3 Experimental Results and Analysis under Structural Attacks (RQ1)

Robustness against Metattack. Table 2 presents the node classification performance under Metattack in poisoning and evasive settings. We obtain the following findings: (i) ArmorGAE achieves the highest performance on all datasets in both clean and attacked scenarios. This superiority demonstrates that the dual-consistency graph augmentation module and adversarial learning mechanism effectively defend against structural noise and learn more robust latent representations. (ii) ArmorGAE exhibits superior robustness against increas-

Datasets	Methods	Clean	Poisoning (Acc.)			Evasive (Acc.)		
			1	2	3	1	2	3
Cora	GraphMAE	82.17	76.02	70.26	67.10	77.11	74.70	74.85
	MaskGAE	83.86	78.52	70.12	68.19	68.39	63.28	59.88
	Bandana	82.05	77.83	75.06	72.29	80.12	77.35	75.06
	AUGMAE	82.75	79.36	74.36	71.69	81.45	**78.41**	**76.11**
	DGMAE	83.78	76.89	**75.78**	72.26	80.12	76.29	73.25
	ArmorGAE	**85.30**	**80.12**	75.06	**73.37**	**83.98**	77.87	75.18
Citeseer	GraphMAE	80.63	69.84	63.65	49.21	69.36	55.70	57.91
	MaskGAE	82.38	81.43	73.65	61.90	80.95	79.37	77.14
	Bandana	81.75	80.63	68.73	61.59	81.82	80.05	79.32
	AUGMAE	80.95	80.95	77.46	**65.34**	79.37	77.41	65.24
	DGMAE	**82.54**	80.95	74.60	61.90	80.95	79.37	66.67
	ArmorGAE	**82.54**	**81.79**	**80.63**	64.92	**82.86**	**80.32**	**79.37**
Pubmed	GraphMAE	86.96	83.87	82.80	79.03	85.99	85.59	82.73
	MaskGAE	86.77	82.80	77.96	70.91	82.96	78.55	73.98
	Bandana	87.42	84.78	80.59	75.75	84.68	78.28	74.41
	AUGMAE	87.79	85.22	84.81	**83.33**	87.10	85.48	**84.41**
	DGMAE	87.25	86.48	84.78	81.72	87.17	**86.83**	83.87
	ArmorGAE	**88.15**	**87.01**	**85.17**	80.32	**87.31**	86.39	83.35

Table 3: Node classification accuracy of ArmorGAE and five competitive methods under targeted attack (Nettack) in poisoning and evasive scenarios.

ing perturbation rates, incurring the minimal performance drop among all evaluated models. ArmorGAE achieves an accuracy of 52.28% on Cora dataset under the poisoning rate of 25%, outperforming state-of-the-art baselines such as GraphMAE and AUGMAE. Moreover, ArmorGAE demonstrates near-immunity to structural perturbations on the Computers dataset, with a marginal 1.1% performance decline as the attack rate escalates from 5% to 25%. (iii) In contrast to representative generative baselines like GraphMAE and MaskGAE, which suffer from performance degradation under high-intensity attacks (20%–25% perturbation), ArmorGAE demonstrates superior robustness. Even when compared to competitive baselines like AUGMAE and DGMAE, ArmorGAE delivers superior consistency across graphs with varying homophily and sparsity. This robustness is attributed to ArmorGAE's capability to effectively mitigate the distributional shift caused by perturbations, aligning the learned representations with the underlying clean data distribution.

Robustness against Nettack. We further evaluate the robustness of ArmorGAE against targeted attacks (Nettack) in Table 3, from which we draw the following observations: (i) ArmorGAE achieves the best average ranking on Cora and Citeseer datasets, significantly surpassing traditional generative self-supervised methods. It validates the efficacy of our adversarial masking mechanism in neutralizing malicious edge perturbations. (ii) In poisoning attacks, most baselines suffer from sharp performance drops as intensity increases, as the corrupted training topology misguides the representation learning process. In evasive attacks, ArmorGAE achieves 83.98% on Cora dataset at intensity 1, outperforming AUGMAE by approximately 2.5%. (iii) On the Pubmed dataset, ArmorGAE achieves superior performance in both clean settings and low-to-medium intensity attacks (ratios 1 and 2). While AUGMAE and DGMAE exhibit slightly better resilience at the highest perturbation ratio, ArmorGAE remains highly competitive, maintaining a comparable performance with only a marginal gap to the optimal results.

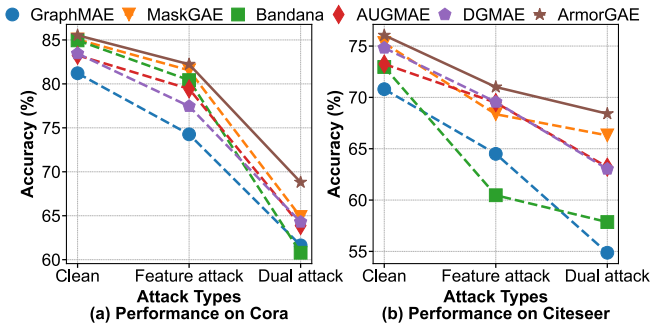

Figure 3: Accuracy comparison under adversarial attacks.

4.4 Experimental Results and Analysis under Feature and Dual Attacks (RQ2)

Robustness against Feature Attacks. As illustrated in Figure 3, all models suffer performance degradation under feature attack, yet ArmorGAE incurs the minimal performance degradation. Taking the Cora dataset as an example, while the accuracy of GraphMAE plummets from 81.21% to 74.26%, ArmorGAE maintains a superior accuracy of 82.23%, markedly outperforming the runner-up (MaskGAE at 81.55%). This robustness is attributed to the adversarial game, which recovers the underlying true data distribution and distinguishes between reconstructed and real edges.

Robustness against Dual Attacks. In most dual-attack scenarios, the defensive advantages of ArmorGAE become even more pronounced. On Citeseer dataset, while the accuracies of representative baselines such as Bandana and GraphMAE drop below the 60% threshold, ArmorGAE sustains a robust performance of 68.41%, maintaining a significant margin over MaskGAE (66.30%). As depicted in the performance trends, the degradation curve of ArmorGAE remains significantly flatter compared to the competing baselines as attack intensity escalates. By formulating an adversarial game to establish a defensive barrier within the latent feature space, ArmorGAE effectively mitigates the impact of highly uncertain joint perturbations on feature and structure.

4.5 Ablation Studies (RQ3)

To investigate the contribution of each component in ArmorGAE, we conduct ablation studies on four datasets in both clean and attacked scenarios (ptb=0.15), as illustrated in Figure 4. We compare ArmorGAE with four variants: (1) '-r/p-rand': replaces our augmentation methods with random edge selection; (2) '-w/o-aug': removes the dual-consistency augmentation; (3) '-w/o-adv': removes the adversarial learning mechanism; (4) '-w/o-er': eliminates the edge masking-reconstruction module. We obtain the following insights: (i) '-w/o-er' suffers the most performance degradation in all scenarios. It confirms that structural reconstruction is the cornerstone of our model to recover the intrinsic graph topology. (ii) ArmorGAE outperforms the variant '-w/o-adv'. This gap is particularly evident in attack scenarios (e.g., a drop of nearly

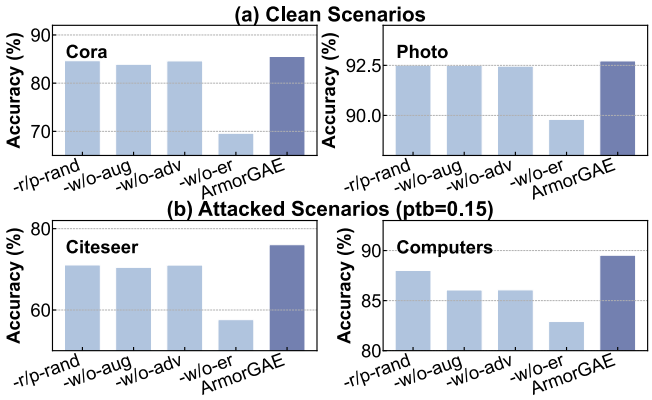

Figure 4: Ablation studies of model components.

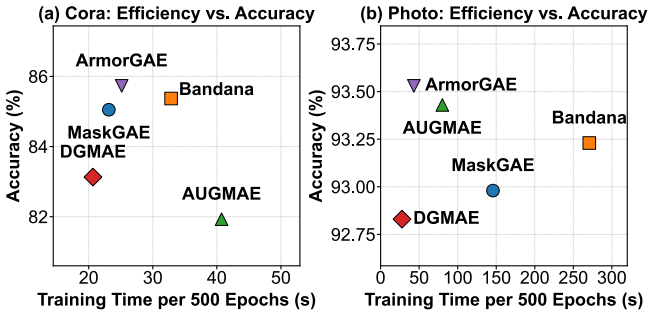

Figure 5: Performance-efficiency comparison of different methods.

3.5% on Computers dataset). It validates that the adversarial mechanism effectively strengthens the model's defensive barrier. (iii) The performance decline observed in the variant '-w/o-aug' underscores the importance of our designed augmentation strategy. (iv) The variant '-r/p-rand' consistently underperforms ArmorGAE, especially in perturbed environments. It demonstrates that our augmentation strategy enriches the graph with semantically relevant edges, whereas random perturbations tend to introduce noise.

4.6 Training Efficiency (RQ4)

Figure 5 presents the performance-efficiency trade-off of five methods on Cora and Photo datasets. To ensure a fair comparison, we report the total training time under a fixed budget of 500 epochs for all models. ArmorGAE achieves state-of-the-art accuracy of 85.53% with a training time of 25.15s on Cora dataset. It outperforms the best baseline (Bandana, 85.37%) in terms of both accuracy and efficiency. Similarly, ArmorGAE reaches the highest accuracy of 93.53% with a training duration of 43.14s on Photo dataset. It is significantly more efficient than the competitive baseline Bandana. These results validate that the integration of the adversarial masked autoencoder and augmentation modules does not introduce excessive computational overhead.

4.7 Parameter Sensitivity Analysis (RQ5)

We conduct a sensitivity analysis to evaluate the stability of ArmorGAE. The results are summarized in Figure 6.

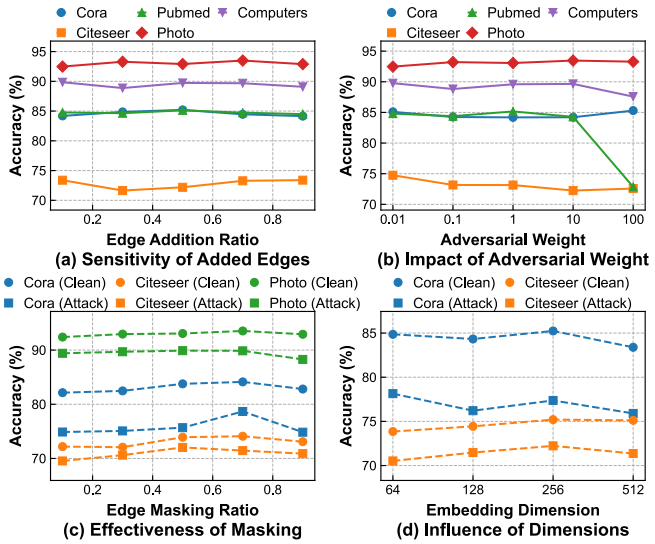

Figure 6: The experimental of hyper-parameter analysis.

Edge Addition Ratio (p)**:** As shown in Figure 6(a), the performance remains relatively stable on different edge addition ratios. The optimal accuracy is achieved when the ratio is in the range of 0.3-0.5. It indicates that moderate structural augmentation provides sufficient signals, while excessively high ratios may introduce detrimental structural noise.

Adversarial Weight (λ_{adv})**:** As illustrated in Figure 6(b), ArmorGAE exhibits consistently robust performance on a range of adversarial weights ($\lambda_{adv} \in [0.01, 10]$). However, a significant performance drop occurs on the Pubmed dataset when λ_{adv} increases to 100. It suggests that an excessive adversarial penalty may dominate the total loss, which hinders the model from learning effective node features.

Edge Masking Ratio (m_e)**:** In Figure 6(c), the accuracy peaks at a masking ratio of 0.7 on most datasets. Higher masking rates compel the model to recover more complex structural information. It enhances the robustness of the learned representations in both clean and attacked scenarios.

Embedding Dimension (h_d)**:** Figure 6(d) shows that ArmorGAE is relatively insensitive to the hidden dimension size. While a dimension of 256 offers the best balance, further increasing it to 512 leads to marginal performance drops, likely due to slight overfitting on smaller datasets like Cora.

5 Conclusion

In this paper, we propose ArmorGAE, a robust graph masked autoencoder tailored to mitigate the vulnerability of existing GMAEs against adversarial attacks. ArmorGAE synergizes a dual-consistency graph augmentation with an adversarial masked modeling paradigm to counteract structural perturbations while enriching topological semantics. Theoretically, we provide an analysis revealing that our augmentation strategy is equivalent to expanding high-quality positive samples in contrastive learning. Extensive experiments demonstrate that ArmorGAE achieves state-of-the-art performance in clean scenarios and exhibits superior robustness against diverse adversarial attacks.

Acknowledgments

This research is supported by the National Natural Science Foundation of China (Grant No. 62472263, 62502288, 62506216), the Natural Science Foundation of Shandong Province (Grant No. ZR2024MF034), the Ministry of Education in China Foundation for Humanities and Social Sciences (Grant No. 24YJAZH058, 24YJCZH461, 24YJCZH149).

References

[Chang *et al.*, 2020] Heng Chang, Yu Rong, Tingyang Xu, Wenbing Huang, Honglei Zhang, Peng Cui, Wenwu Zhu, and Junzhou Huang. A restricted black-box adversarial framework towards attacking graph embedding models. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, pages 3389–3396, 2020.

[Dai *et al.*, 2018] Hanjun Dai, Hui Li, Tian Tian, Xin Huang, Lin Wang, Jun Zhu, and Le Song. Adversarial attack on graph structured data. In *International conference on machine learning*, pages 1115–1124, 2018.

[Fedus *et al.*, 2018] William Fedus, Ian Goodfellow, and Andrew M Dai. Maskgan: Better text generation via filling in the -. In *Proceedings of the International Conference on Learning Representations*, 2018.

[Fei *et al.*, 2023] Zhengcong Fei, Mingyuan Fan, Li Zhu, Junshi Huang, Xiaoming Wei, and Xiaolin Wei. Masked auto-encoders meet generative adversarial networks and beyond. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 24449–24459, 2023.

[Ganin *et al.*, 2016] Yaroslav Ganin, Evgeniya Ustinova, Hana Ajakan, Pascal Germain, Hugo Larochelle, François Laviolette, Mario March, and Victor Lempitsky. Domain-adversarial training of neural networks. *Journal of machine learning research*, 17(59):1–35, 2016.

[Hou *et al.*, 2022] Zhenyu Hou, Xiao Liu, Yukuo Cen, Yuxiao Dong, Hongxia Yang, Chunjie Wang, and Jie Tang. Graphmae: Self-supervised masked graph autoencoders. In *Proceedings of the 28th ACM SIGKDD conference on knowledge discovery and data mining*, pages 594–604, 2022.

[Hou *et al.*, 2023] Zhenyu Hou, Yufei He, Yukuo Cen, Xiao Liu, Yuxiao Dong, Evgeny Kharlamov, and Jie Tang. Graphmae2: A decoding-enhanced masked self-supervised graph learner. In *Proceedings of the ACM web conference 2023*, pages 737–746, 2023.

[Hu *et al.*, 2024] Yulan Hu, Sheng Ouyang, Zhirui Yang, Ge Chen, Junchen Wan, Xiao Wang, and Yong Liu. Exploring task unification in graph representation learning via generative approach. *arXiv preprint arXiv:2403.14340*, 2024.

[Hu *et al.*, 2025] Yulan Hu, Zhirui Yang, Sheng Ouyang, and Yong Liu. Adversarial masked graph autoencoders for improved graph representation learning. In *Proceedings of the 2025 International Conference on Multimedia Retrieval*, pages 478–486, 2025.

[Huo *et al.*, 2023] Cuiying Huo, Di Jin, Yawen Li, Dongxiao He, Yu-Bin Yang, and Lingfei Wu. T2-GNN: Graph neural networks for graphs with incomplete features and structure via teacher-student distillation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pages 4339–4346, 2023.

[Jung and Park, 2025] Heesoo Jung and Hogun Park. Balancing graph embedding smoothness in self-supervised learning via information-theoretic decomposition. In *Proceedings of the ACM on Web Conference 2025*, pages 2621–2632, 2025.

[Li *et al.*, 2023] Jintang Li, Ruofan Wu, Wangbin Sun, Liang Chen, Sheng Tian, Liang Zhu, Changhua Meng, Zibin Zheng, and Weiqiang Wang. What's behind the mask: Understanding masked graph modeling for graph autoencoders. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 1268–1279, 2023.

[Liu *et al.*, 2024] Chuang Liu, Yuyao Wang, Yibing Zhan, Xueqi Ma, Dapeng Tao, Jia Wu, and Wenbin Hu. Where to mask: Structure-guided masking for graph masked autoencoders. In *Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence*, pages 2180–2188, 2024.

[Ma *et al.*, 2024] Xiaoxiao Ma, Fanzhen Liu, Jia Wu, Jian Yang, Shan Xue, and Quan Z Sheng. Rethinking unsupervised graph anomaly detection with deep learning: Residuals and objectives. *IEEE Transactions on Knowledge and Data Engineering*, 2024.

[Pan *et al.*, 2018] Shirui Pan, Ruiqi Hu, Guodong Long, Jing Jiang, Lina Yao, and Chengqi Zhang. Adversarially regularized graph autoencoder for graph embedding. *arXiv preprint arXiv:1802.04407*, 2018.

[Sen *et al.*, 2008] Prithviraj Sen, Galileo Namata, Mustafa Bilgic, Lise Getoor, Brian Galligher, and Tina Eliassi-Rad. Collective classification in network data. *AI magazine*, 29(3):93–93, 2008.

[Shchur *et al.*, 2018] Oleksandr Shchur, Maximilian Mumme, Aleksandar Bojchevski, and Stephan Günnemann. Pitfalls of graph neural network evaluation. *arXiv preprint arXiv:1811.05868*, 2018.

[Shen *et al.*, 2025] Weixuan Shen, Xiaobo Shen, and Shirui Pan. Adversarial contrastive graph masked autoencoder against graph structure and feature dual attacks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pages 12515–12523, 2025.

[Sun *et al.*, 2020] Yiwei Sun, Suhang Wang, Xianfeng Tang, Tsung-Yu Hsieh, and Vasant Honavar. Adversarial attacks on graph neural networks via node injections: A hierarchical reinforcement learning approach. In *Proceedings of the Web Conference 2020*, pages 673–683, 2020.

[Sun *et al.*, 2022] Lichao Sun, Yingtong Dou, Carl Yang, Kai Zhang, Ji Wang, Philip S Yu, Lifang He, and Bo Li. Adversarial attack and defense on graph data: A survey. *IEEE Transactions on Knowledge and Data Engineering*, 35(8):7693–7711, 2022.

[Tan *et al.*, 2023] Qiaoyu Tan, Ninghao Liu, Xiao Huang, Soo-Hyun Choi, Li Li, Rui Chen, and Xia Hu. S2gae: Self-supervised graph autoencoders are generalizable learners with graph masking. In *Proceedings of the sixteenth ACM international conference on web search and data mining*, pages 787–795, 2023.

[Wang *et al.*, 2024] Liang Wang, Xiang Tao, Qiang Liu, and Shu Wu. Rethinking graph masked autoencoders through alignment and uniformity. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 15528–15536, 2024.

[Xia *et al.*, 2025] Riting Xia, Huibo Liu, Anchen Li, Xueyan Liu, Yan Zhang, Chunxu Zhang, and Bo Yang. Incomplete graph learning: A comprehensive survey. *Neural Networks*, page 107682, 2025.

[Zhao *et al.*, 2022] Zhongying Zhao, Zhan Yang, Chao Li, Qingtian Zeng, Weili Guan, and MengChu Zhou. Dual feature interaction-based graph convolutional network. *IEEE Transactions on Knowledge and Data Engineering*, 35(9):9019–9030, 2022.

[Zhao *et al.*, 2024] Ziwen Zhao, Yuhua Li, Yixiong Zou, Jiliang Tang, and Ruixuan Li. Masked graph autoencoder with non-discrete bandwidths. In *Proceedings of the ACM Web Conference 2024*, pages 377–388, 2024.

[Zheng *et al.*, 2018] Wei Zheng, Xiaofeng Zhu, Yonghua Zhu, and Shichao Zhang. Robust feature selection on incomplete data. In *Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence*, pages 3191–3197, 2018.

[Zheng *et al.*, 2025] Ziyu Zheng, Yaming Yang, Ziyu Guan, Wei Zhao, and Weigang Lu. Discrepancy-aware graph mask auto-encoder. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V. 2*, pages 4038–4049, 2025.

[Zügner *et al.*, 2018] Daniel Zügner, Amir Akbarnejad, and Stephan Günnemann. Adversarial attacks on neural networks for graph data. In *Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining*, pages 2847–2856, 2018.