Few-Shot Fine-Tuning of Pretrained Masked Graph Autoencoders for Graph Classification

1st Kopimenan Loganathan
department of Computer Science and Engineering
University of Moratuwa
Colombo, Sri Lanka
loganathan.21@cse.mrt.ac.lk

2nd Uthayasanker Thayasivam*
department of Computer Science and Engineering
University of Moratuwa
Colombo, Sri Lanka
rtuthaya@cse.mrt.ac.lk
*Co-author

***Abstract*—Few-shot graph classification is a fundamental task in graph-centric domains, including bioinformatics, social networks, and drug discovery. Graph neural networks (GNNs) have achieved success in many graph mining tasks, but typically require large labeled datasets to avoid overfitting. In this paper, we explore a two-stage learning paradigm: self-supervised pretraining using Graph Masked Autoencoders (GraphMAE) and meta-level fine-tuning under few-shot settings, analyzing representational transferability, the impact of episodic sampling and masking strategies, and overall adaptation performance under limited supervision. Our results show that combining masked-node reconstruction and episodic meta-learning objectives provides robust graph representations across diverse tasks. Moreover, our findings highlight the importance of considering label-space complexity in graph learning, highlighting models that perform well on low-granularity tasks may still require careful adaptation for richer, multi-way classification.**

***Index Terms*—Graph Classification, Few-Shot learning, Self-Supervised Learning, Graph Masked Autoencoders**

I. Introduction

Learning with limited labeled graphs is essential in domains where label acquisition is expensive or infeasible, such as drug and enzyme discovery. Traditional graph classification methods rely on abundant labeled data and task-specific training, which fail to generalize under label-scarcity. Few-shot graph learning aims to overcome this limitation by transferring knowledge from base classes to new ones throughoughough meta-learning or transfer learning [1]–[5].

Graph-structured data are ubiareitous across diverse domains, including bioinformatics, social network analysis, and chemistry. Graph classification—the task of assigning labels to entire graphs—plays a central role in predicting protein functions, identifying molecular properties, and recognizing structural motifs in complex networks. However, despite the success of Graph Neural Networks (GNNs), their performance strongly depends on large-scale labeled datasets. In many scientific domains, such annotations are costly, time-consuming, or impractical to obtain, posing a fundamental challenge for supervised GNN training.

Few-shot learning offers a promising pathway toward alleviating the challenges by enabling models to generalize from only a handful of labeled examples. Recent advances integrate self-supervised pretraining with few-shot fine-tuning, where masked graph autoencoders (GraphMAE) learn transferable representations without explicit supervision.Usingg the expressive Graph Isomorphism Network (GIN) architecture, such pretrained encoders can be effectively adapted to new tasks through episodic fine-tuning.

This work explores a two-stage learning paradigm—self-supervised pretraining followed by few-shot adaptation—on the ENZYMES dataset. We investigate how pretrained GIN-based GraphMAE models can be fine-tuned for few-shot graph classification, analyzing representational transferability, the impact of episodic sampling and masking strategies, and overall adaptation performance under limited supervision. By bridging pretraining and few-shot transfer, this study contributes to advancing sample-efficient graph representation learning for practical, label-scarce domains.

Masking-based pretraining (e.g., GraphMAE) and GNN architectures such as GIN have proven effective for representation learning on graphs. However, extending them for rapid adaptation in few-shot regimes remains an open challenge. This work positions itself at the intersection of self-supervised graph pretraining and episodic

few-shot fine-tuning, targeting robust graph classification with minimal labeled examples per class [3], [5], [6].

The source code for the implementation and experiments is publicly available at: **https://github.com/Kopimenan/Enhancing-GraphMAE**.

II. Literature Review

Graph neural networks (GNNs) have achieved success in many graph mining tasks, typicallyally require large labeled datasets to avoid overfitting. Recent work on self-supervised graph representation learning has shown that generative pretraining (autoencoder-style) can yield powerful features. In particular, the GraphMAE approach uses a masked graph autoencoder: it randomly corrupts (masks) a portion of node features and trains a GNN encoder/decoder to reconstruct the original features. Unlike contrastive methods, this generative SSL is trained with a scaled cosine error (SCE) loss that down-weights easy examples. GraphMAE's authors demonstrated strong transfer learning results in graph classification tasks: for example, with GIN backbones on datasets such as MUTAG, PROTEINS, and COLLAB, GraphMAE outperformed previous self-supervised baselines in most cases.

Few-shot learning on graphs has garnered significant attention due to the widespread presence of data scarcity in graph-centric domains, including bioinformatics, social networks, and drug discovery. Unlike traditional supervised learning requiring large labeled datasets, few-shot learning seeks to generalize from limited labeled examples across novel classes or tasks.

A. Few-shot Learning on Graphs

Survey Perspective: Recent reviews categorize few-shot graph learning based on problem granularity (node-, edge-, and graph-level tasks) as well as methodological axes—metric-based, optimization-based, and hybrid meta-learning. Yu et al. provide an IEEE survey highlighting meta-learning (MAML, Prototypical Networks), pre-training (autoencoding, contrastive learning), and prompt-based techniques as orthogonal strategies for label and structure scarcity. The rise of hybrid methods, which combine self-supervised graph pre-training (e.g., GraphMAE, contrastive coding) with meta-finetuning for fast adaptation, reflects the state-of-the-art trajectory [3], [6], [8].

Metric-based approaches in FSLG, notably Prototypical Networks, compare query embeddings to class prototypes averaged from support embeddings. Optimization-based models (e.g., MAML, Meta-GNN) learn initialization parameters for rapid gradient-based adaptation to new tasks. More recent augmentations leverage graph-structured subgraph mining or adaptive prompt-tuning for improved transferability [2], [3], [5], [6].

B. Graph Masked Autoencoders and Pre-training

GraphMAE represents a family of masked autoencoders for graphs, pre-training GNNs by reconstructing masked node features. The pre-trained representations can be adapted to downstream tasks, such as graph classification, using further fine-tuning under supervised or episodic setups. Hybridization of masked reconstruction loss with discriminative or contrastive objectives is increasingly adopted to improve few-shot generalization [5], [6].

C. Meta-learning and Fine-tuning for Few-shot Graph Classification

Transferring meta-learning paradigms (episodic, N-way K-shot) from images to graphs has proven beneficial but is challenged by graph domain-specific factors—irregular topologies, heterogeneity, and permutation invariance. Works such as Meta-MGNN, SMF-GIN, and the adaptive methods reviewed by Crisostomi et al. demonstrate that episodic fine-tuning with GIN-based graph encoders outperforms vanilla transfer learning or kernel methods for few-shot graph classification [4], [5].

Task-adaptive modules (e.g., conditional embeddings, MixUp-based augmentations) and reconstruction losses can further enrich representations for few-shot transfer. Recent findings affirm that hybrid methods leveraging both unsupervised pre-training and episodic meta-finetuning yield superior few-shot accuracy, especially on biological datasets like ENZYMES [3], [5].

D. Domain-Specific Applications and Transfer Learning

Few-shot graph learning has found particular success in specialized domains where labeled data is inherently scarce or expensive to obtain. In molecular property prediction, methods like Meta-MGNN demonstrate that few-shot approaches can effectively predict molecular properties with minimal experimental validation. These methods leverage molecular graph structure to encode chemical knowledge and enable rapid adaptation to new molecular properties [4].

Knowledge graph completion represents another critical application domain, where few-shot methods address the challenge of predicting relations for entities with limited connectivity information. Recent approaches combine graph neural networks with meta-learning to enable rapid adaptation to new entity types or relation patterns in evolving knowledge bases [1].

Social network analysis benefits from few-shot approaches when analyzing emerging communities or predicting behaviors in newly formed networks. The heterogeneous nature of social graphs makes few-shot learning particularly valuable for handling diverse node types and dynamic network structures [2].

E. Challenges

Despite these advances, challenges remain in developing parameter-efficient adaptation, scalable episode generation, and interpretable graph prompts, as highlighted in recent IEEE surveys. The literature identifies the need for benchmarks with unified splits and reporting, further motivating experiments like those in this paper [6].

Contemporary challenges include handling graph heterogeneity in few-shot scenarios, developing robust evaluation protocols that account for graph structural diversity, and creating interpretable models that can explain their few-shot predictions. The integration of large language models with graph neural networks presents both opportunities and challenges for few-shot learning on text-rich graphs.

The development of theoretical foundations for few-shot graph learning remains an active area of research, with recent work beginning to establish generalization bounds and sample complexity analysis for graph-based few-shot learning scenarios. Understanding the fundamental limits and capabilities of few-shot graph learning will be crucial for future methodological advances.

III. Methodology

A. Model Framework

The framework follows a two-stage process—self-supervised pre-training using Graph Masked Autoencoders (GraphMAE) and meta-level fine-tuning under few-shot settings.During pre-training, the encoder-decoder model is optimized to reconstruct masked node attributes from partial graph observations, thereby learning structure-aware latent representations. The encoder employs a Graph Isomorphism Network (GIN), known for its injective aggregation that distinguishes graph structures more effectively than standard GCN or GAT models.

- Encoder: Utilizes a pre-trained GraphMAE backbone with a GIN encoder. The GIN architecture is chosen for its strong injective aggregation properties, facilitating expressive learning over small graphs common in biological domains [5], [7].
- Decoder: A GIN-based decoder reconstructs masked node features from pooled graph-level embeddings, supporting masked autoencoding pre-training.

Pre-training: The model employs a scaled cosine error loss (1), consistent with GraphMAE, and when labels are unavailable, node degrees or other standard attributes are used; it is first trained on TUDataset using a masking strategy to optimize node feature reconstruction, thereby priming the GIN encoder to capture robust node-, edge-, and graph-level structural information, with all pre-training conducted under fixed hyperparameters such as learning rate and masking rate.

$$\mathcal{L}_{\text{SCE}} = \frac{1}{|\tilde{\mathcal{V}}|} \sum_{v_i \in \tilde{\mathcal{V}}} \left(1 - \frac{\mathbf{x}_i^\top \mathbf{z}_i}{\|\mathbf{x}_i\| \cdot \|\mathbf{z}_i\|}\right)^\gamma, \quad \gamma \geq 1 \tag{1}$$

B. Few-shot Fine-tuning

For each few-shot learning task, a copy of the pre-trained GraphMAE–GIN model is initialized, and a new linear classification layer is added to adapt to the current N-way problem. From the dataset, N classes are randomly chosen, with K labeled graphs per class forming the support set and Q additional graphs per class forming the query set. During fine-tuning, the model is trained on the support set for a fixed number of epochs using a standard optimization procedure with a cross-entropy classification objective. Graph representations are obtained by passing node features through the encoder, aggregating them into graph-level embeddings, and mapping them to class probabilities via the classifier. Once the support training phase is complete, the fine-tuned model is evaluated on the query set to assess its generalization to unseen examples. The overall performance is reported as the average micro-F1 score across multiple randomly sampled tasks. This episodic fine-tuning process enables efficient adaptation of the pre-trained model to new classes with limited labeled data, aligning with the principles of few-shot graph learning.

- Episodic Setup: Adopts the N-way K-shot Q-query protocol (e.g., 5-way, 5-shot, 15-query) where each episode samples N classes, with K labeled support graphs per class and Q query graphs for evaluation.
- Objective: During meta-finetuning, the model optimizes episodic classification cross-entropy loss. Optionally, an auxiliary reconstruction loss is added to retain pre-trained knowledge.
- Hyperparameters: Episodes are constructed for [N=5, K=5, Q=15]. Optimizer: Adam, learning rate 1e-4. Batch size: 16 episodes per batch. The encoder is optionally either partially or fully fine-

tuned depending on adaptation constraints. Dropout and/or MixUp-style augmentation is optionally applied during fine-tuning.

The base model follows GraphMAE, employing its original two-layer GIN architecture, node masking strategy, and cosine reconstruction loss. For downstream adaptation, a trainable linear head maps mean-pooled graph embeddings to K classes. Performance is reported as mean±standard deviation of accuracy and micro-F over 100 randomly sampled few-shot tasks per scenario [7].

The methodology involves adapting a pretrained masked graph autoencoder model to few-shot graph classification tasks through fine-tuning with a task-specific classification head. Initially, the pretrained encoder weights are retained while appending a new linear classification layer corresponding to the number of classes in the few-shot task, enabling dynamic adjustment to the task's label space. During fine-tuning, the entire model, including the encoder and new classification head, is optimized on a small support set of labeled graphs to learn class-discriminative graph embeddings. Graph-level embeddings are obtained by applying a pooling operation (mean, max, or sum) over node embeddings output by the encoder. After task-adaptive fine-tuning for a fixed number of epochs, the updated model is evaluated on a disjoint query set to measure performance using metrics like micro-averaged F1 score. This pipeline allows the pretrained model's general graph representation capabilities to be specialized efficiently to new classification tasks with limited labeled data, improving adaptation and accuracy while leveraging the learned structural knowledge encoded in the masked graph autoencoder.

IV. Experiments

A. Dataset

ENZYMES: A standard TUDataset benchmark for biochemical graph classification. ENZYMES contains 600 protein graphs, each labeled into one of 6 classes, representing structural enzyme families. Each graph has node labels, continuous attributes, and a varying number of nodes [5].

SYNTHETIC, DD, PROTEINS, OHSU: all sourced from the TUDataset collection, each contain two graph classes.

MSRC_9 (8 classes) and MSRC_21 (20 classes) datasets, both from the computer vision domain and available in TUDataset

B. Results

The experimental results present accuracy values for various N-way K-shot training configurations conducted on different TUDatasets. The results are compared against the baseline GraphMAE model, which was trained using its original, complex setup with default, fixed hyperparameters (e.g., learning rate, optimizer, masking rate)

TABLE I
RESULTS ON MSRC_9

Method	Accuracy
GraphMAE	0.91
3-way 5-shot	0.97
5-way 5-shot	0.94
3-way 10-shot	0.98
5-way 10-shot	0.96

TABLE II
RESULTS ON MSRC_21

Method	Accuracy
GraphMAE	0.91
3-way 5-shot	0.93
4-way 5-shot	0.92
3-way 10-shot	0.97
3-way 10-shot	0.96

TABLE III
RESULTS ON ENZYMES

Method	Accuracy
GraphMAE	0.22
4-way 5-shot	0.31
5-way 10-shot	0.28

TABLE IV
ACCURACIES ON DATASETS WITH 2 CLASSES

Method	SYNTHETIC	DD	PROTEINS	OHSU
GraphMAE	0.54	0.68	0.59	0.58
2-way 5-shot	0.50	0.56	0.56	0.50
2-way 10-shot	0.50	0.59	0.58	0.51

V. Discussion

The experiments indicate that GraphMAE pretraining followed by few-shot fine-tuning yields robust graph representations across diverse tasks. In particular, combining

the masked-node reconstruction objective of GraphMAE with episodic meta-learning produces a competitive few-shot learning framework. However, the benefits of this two-stage strategy are not uniform: they depend critically on the number of classes (class granularity) in the target task. We observe a clear pattern in which multi-class tasks gain substantially more from fine-tuning than binary tasks do, reflecting how class diversity interacts with the pretrained features.

On datasets with many classes, GraphMAE's unsupervised backbone alone underperforms, indicating an underfitting to the complex label space. For example, on the MSRC-9 (8-way) and MSRC-21 (20-way) image-graph datasets, the GraphMAE encoder achieved only about 0.91 accuracy without fine-tuning. (Similarly low baseline was seen on ENZYMES, a 6-way dataset, where GraphMAE yielded 0.22 accuracy.) These results suggest that the masked reconstruction loss, being label-agnostic, does not fully capture inter-class distinctions when many classes are present. The pretraining objective thus learns general structural patterns but leaves some class-specific variation unexplained, effectively underfitting the multi-class problem.

Few-shot fine-tuning remedies this gap dramatically. When even a handful of labeled examples per class are provided, the model can specialize its features to the target classes. In MSRC-9, for instance, few-shot adaptation raised accuracy from 0.91 to as high as 0.98 in a 10-shot setting; on MSRC-21 it rose from 0.91 to 0.97 under 10-shot training. These large gains (often 5–7 percentage points) demonstrate that scarce supervision is sufficient to calibrate GraphMAE's representations for fine-grained classification. In short, task-specific tuning unlocks significant performance on multi-class tasks, even though the label data is extremely limited.

In contrast, on simple binary-class benchmarks, GraphMAE's pretrained features were already quite discriminative, and fine-tuning had little effect. For SYNTHETIC, DD, PROTEINS, and OHSU (all two-class problems), the GraphMAE baseline achieved moderate-to-high accuracy (for example, 0.68 forrr DD and 0.5fororor PROTEINS). Few-shot training did not improve these results appreciably – in fact, it often yielded a slight drop. For example, DD accuracy fell from 0.68 to 0.59 after 10-shot fine-tuning, and OHSU fell from 0.58 to 0.51. These marginal or negative gains indicate that the pretrained GraphMAE embeddings already align well with the binary decision boundary. In essence, when the task has only two classes, the general graph features learned by masking are sufficient to separate them, and additional labeled data does not change the decision surface substantially.

Together, these patterns imply that class granularity fundamentally mediates fine-tuning effectiveness. When class diversity is low, general-purpose GraphMAE representations are enough for classification, so adaptation offers little improvement. But when the number of classes grows, the unlabeled pretraining fails to partition the embedding space finely, and few-shot supervision becomes crucial to refine the representations for each class. This insight has practical implications: it suggests that masked graph autoencoders may need complementary objectives or strategies when targeting highly multi-class problems. For instance, future work could integrate class-contrastive or pseudo-label losses during pretraining to encourage inter-class separation. More broadly, our findings highlight the importance of considering label-space complexity in few-shot graph learning – models that perform well on low-granularity tasks may still require careful adaptation for richer, multi-way classification.

REFERENCES

[1] J. Chen et al., "Zero-Shot and Few-Shot Learning With Knowledge Graphs: A Comprehensive Survey," Proceedings of the IEEE, vol. 111, no. 6, pp. 653–685, Jun. 2023, doi: https://doi.org/10.1109/jproc.2023.3279374.

[2] S. Wang, C. Chen, and J. Li, "Graph Few-shot Learning with Task-specific Structures," arXiv.org, 2022. https://arxiv.org/abs/2210.12130 (accessed Aug. 17, 2025).

[3] C. Zhang et al., "Few-Shot Learning on Graphs," arXiv.org, 2022. https://arxiv.org/abs/2203.09308 (accessed Aug. 20, 2025).

[4] Z. Guo et al., "Few-Shot Graph Learning for Molecular Property Prediction," arXiv (Cornell University), Apr. 2021, doi: https://doi.org/10.1145/3442381.3450112.

[5] D. Crisostomi, S. Antonelli, V. Maiorca, L. Moschella, R. Marin, and E. Rodolà, "Metric Based Few-Shot Graph Classification," arXiv.org, 2022. https://arxiv.org/abs/2206.03695 (accessed Aug. 18, 2025).

[6] X. Yu et al., "A Survey of Few-Shot Learning on Graphs: from Meta-Learning to Pre-Training and Prompt Learning," arXiv.org, 2024. https://arxiv.org/abs/2402.01440 (accessed Aug. 16, 2025).

[7] Z. Hou et al., "GraphMAE: Self-Supervised Masked Graph Autoencoders," arXiv.org, 2022. https://arxiv.org/abs/2205.10803

[8] H. Yao et al., "Graph Few-shot Learning via Knowledge Transfer," arXiv.org, 2019. https://arxiv.org/abs/1910.03053 (accessed Aug. 24, 2025).

[9] "Datasets," TUDataset, May 20, 2023. https://chrsmrrs.github.io/datasets/docs/datasets/ (accessed Sep. 5, 2025).