CS4681 - Advanced Machine Learning

Progress Report
Enhancing GraphMAE with Few-Shot Graph Learning

Kopimenan L.

210296X

1 Introduction

Graphs are a powerful data structure used to represent complex relationships in real-world systems such as social networks, molecular structures, knowledge graphs, recommendation platforms, and transportation systems. Traditional machine learning methods often assume independent and identically distributed data, which fails to capture the relational dependencies that exist in such domains. Graph Learning, and particularly Graph Neural Networks (GNNs), address this limitation by leveraging both node features and structural information to learn expressive representations, enabling improved performance in downstream tasks such as classification, link prediction, and clustering.

Recent advances in self-supervised graph learning have significantly reduced the reliance on large labeled datasets. One notable contribution is GraphMAE, a masked graph autoencoder framework that learns node and graph-level representations by reconstructing masked node features. This approach has proven highly effective in extracting generalizable features and has established state-of-the-art performance on several graph learning benchmarks. However, GraphMAE and similar self-supervised frameworks still rely on large amounts of unlabelled data for effective pretraining and do not directly address scenarios where only a small number of labelled examples are available for fine-tuning. In many real-world applications, labelled data is expensive or time-consuming to acquire, such as in drug discovery, fraud detection, or specialized knowledge graphs. This makes few-shot graph learning — the ability to generalize to new tasks or classes with only a few labelled examples—an increasingly important research direction. Integrating few-shot learning mechanisms into self-supervised frameworks like GraphMAE has the potential to improve adaptability in low-data regimes while preserving the advantages of pretraining.

This research aims to investigate how GraphMAE can be enhanced to support few-shot graph learning. Specifically, the study will explore combining self-supervised masked feature reconstruction with few-shot adaptation strategies, such as prototypical networks or meta-learning approaches, to enable more effective generalization from limited labelled samples. The expected outcome is a framework that maintains the strong representation learning capabilities of GraphMAE while improving its performance in few-shot settings.

2 Literature Review

2.1 GraphMAE

Self-supervised learning (SSL) has become an essential paradigm across domains such as computer vision (CV) and natural language processing (NLP), with generative methods like BERT, GPT, and Masked Autoencoders (MAE) demonstrating remarkable success. However, in the graph learning domain, progress has been dominated by contrastive SSL approaches, which rely heavily on negative sampling, complex training strategies, and heuristic graph augmentations. While effective for node and graph classification, these methods often suffer from instability and limited scalability. In contrast, generative self-supervised graph autoencoders (GAEs) offer a more direct approach by reconstructing graph inputs but have historically lagged behind contrastive methods due to several challenges, including overemphasis on structure reconstruction, weak robustness in feature recovery, reliance on mean square error loss, and limited decoder expressiveness.

GraphMAE was introduced to address these limitations by rethinking the design of GAEs. Instead of focusing on reconstructing graph structures, it employs masked feature reconstruction, where node attributes are partially masked and then predicted by the model. This is combined with a scaled cosine error to stabilize training and improve robustness, along with a more expressive decoder architecture. These innovations allow GraphMAE to overcome the shortcomings of prior GAEs while avoiding the complexity of contrastive frameworks. Extensive experiments across 21 datasets demonstrated that GraphMAE consistently outperforms existing generative and contrastive SSL methods on node classification, graph classification, and transfer learning benchmarks, and in many cases even matches supervised baselines [1].

2.2 Few-Shot Graph Learning

Few-shot learning on graphs addresses the critical challenge of data scarcity, which manifests in two primary forms: label scarcity, where limited annotated data exists for node, edge, or graph-level tasks, and structure scarcity, which arises from sparse topological connectivity like in long-tailed distributions or cold-start scenarios. To tackle these issues, the field has developed three major technical paradigms. Meta-learning approaches learn transferable prior knowledge from a set of base tasks with ample labels but are constrained by the need for extensive annotations and an i.i.d. task assumption. Pre-training paradigms leverage self-supervised learning on unlabelled

graphs through contrastive or generative pretext tasks to learn general-purpose representations, which are then adapted to downstream tasks via fine-tuning or more parameter-efficient strategies like prompt tuning to avoid overfitting. Finally, hybrid methods synergistically combine the strengths of both meta-learning and pre-training. A key recent advancement is the shift towards parameter-efficient adaptation techniques, such as learnable prompts or adapter modules, which fine-tune only a small subset of parameters, making them particularly effective for few-shot settings. The survey concludes by outlining future directions, including scaling these methods to large-scale and complex graphs, improving interpretability, and developing foundational models for graphs [2].

To address the challenge of molecular property prediction with limited data, Guo et al. propose Meta-MGNN, a few-shot graph learning model that integrates meta-learning with a pre-trained molecular graph neural network. The model employs a MAML-based framework to learn transferable knowledge from various property prediction tasks, enabling rapid adaptation to novel properties with only a few examples. To further enhance learning, Meta-MGNN incorporates a self-supervised module—comprising bond reconstruction and atom-type prediction tasks—to exploit unlabelled structural information, and a self-attentive task-weighting mechanism to prioritize more informative tasks during meta-training. Evaluated on the Tox21 and Sider datasets, Meta-MGNN outperforms state-of-the-art baselines, demonstrating the effectiveness of combining pre-training, meta-learning, and self-supervision for few-shot molecular property prediction [3].

V. Gracia *et al.* introduced a graph neural network (GNN) framework for few-shot learning, which formulates the problem as supervised message passing on a fully-connected graph where nodes represent images and edges are learned similarity measures. The model generalizes several existing few-shot learning approaches—such as Siamese Networks, Matching Networks, and Prototypical Networks — by unifying them within a flexible GNN architecture that supports end-to-end training and naturally handles permutation invariance. The method achieves competitive performance on Omniglot and Mini-ImageNet with significantly fewer parameters than contemporary models, and is seamlessly extended to semi-supervised and active learning settings, demonstrating the utility of graph-based models in relational and low-label regimes [4].

3 Methodology Outline

Framework Design

- Extend GraphMAE to support few-shot graph learning.
- Core idea: pretrain self-supervised GraphMAE embeddings, then perform few-shot adaptation on limited labelled graphs.
- Two complementary strategies:

1. Prototype-based Few-Shot Learning
    - Compute class prototypes from embeddings of few labeled graphs (support set).
    - Classify query graphs based on distance to prototypes in embedding space.
2. Meta-Learning with MAML
    - Learn encoder initialization that can adapt rapidly to new few-shot graph tasks
    - Use task-specific gradients to fine-tune the encoder for each episode.

Model Architecture

- GraphMAE Encoder-Decoder: masks partial node features, reconstructs them, produces rich node/graph embeddings.
- Few-Shot Adaptation Module:
    - Prototype Computation Layer: aggregates support set embeddings to form class prototypes.
    - Meta-Learning Adaptation Layer: updates encoder parameters per few-shot task using gradients from support set loss.
- Integration: embeddings from GraphMAE serve as input to few-shot modules, enabling rapid generalization to unseen classes.

Training Procedure

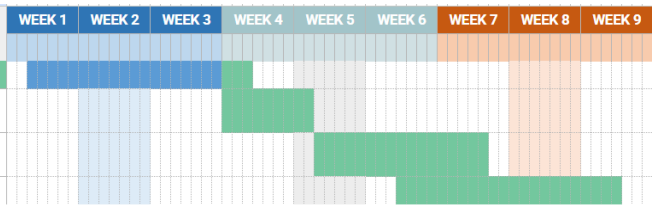
- Phase 1: Self-Supervised Pretraining
    - Train GraphMAE on large unlabeled graph dataset.
    - Loss: masked node feature reconstruction using scaled cosine error.
- Phase 2: Few-Shot Fine-Tuning
    - Use episodic training: sample N-way K-shot tasks.
    - Prototype-based approach: compute prototypes, classify query graphs, minimize cross-entropy loss.
    - Meta-learning approach: perform inner-loop updates on support set, outer-loop update on query set.

Evaluation Metrics

- Few-shot classification accuracy (N-way K-shot tasks)
- F1-score for imbalanced classes
- Generalization across tasks: measure performance on unseen classes

4 Project Timeline

TASK TITLE	START DATE	DUE DATE	DURATION DAYS	PCT OF TASK COMPLETE
Literature Review	06/08/25	27/08/25	22	75%
Setup basline GraphMAE and validation	25/08/25	03/09/25	9	0%
Few-shot adoption design and implementation	04/09/25	20/09/25	17	0%
Benchmarking & Evaluation	12/09/25	03/10/25	22	0%



- Literature review on GNN, GAE, GMAE, Few-Shot Graph Learning, Graph Classification
- End of the evaluation, final paper drafting and submission to conference work.

5 References

[1] Z. Hou *et al.*, "Graphmae: Self-supervised masked graph autoencoders," presented at the Proceedings of the 28th ACM SIGKDD conference on knowledge discovery and data mining, 2022, pp. 594–604.

[2] X. Yu *et al.*, "A Survey of Few-Shot Learning on Graphs: from Meta-Learning to Pre-Training and Prompt Learning," *arXiv preprint arXiv:2402.01440*, 2024.

[3] Z. Guo *et al.*, "Few-shot graph learning for molecular property prediction," presented at the Proceedings of the web conference 2021, 2021, pp. 2559–2567.

[4] V. Garcia and J. Bruna, "Few-shot learning with graph neural networks," *arXiv preprint arXiv:1711.04043*, 2017.