Dynamic GNNs vs. Static GNNs Accuracy Trade-offs on T-GNN Node Classification

Assignee Research

June 2, 2026

Abstract

This report synthesises findings from 10 peer-reviewed papers addressing the following research question: What is the accuracy trade-off between dynamic GNNs and static GNNs when evaluated on node classification tasks using the T-GNN benchmark with varying graph sizes (10K–100K nodes), measured by. Graph neural networks (GNNs) have emerged as powerful models for learning representations of graph data showing state of the art results in various tasks. Nevertheless, the superiority of these methods is usually supported by either evaluating their performance on small subset. 13 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 2.2/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: GNN-MultiFix: Addressing the pitfalls for GNNs for multi-label node classification. Research question: What is the accuracy trade-off between dynamic GNNs and static GNNs when evaluated on node classification tasks using the T-GNN benchmark with varying graph sizes (10K–100K nodes), measured by F1-score under limited labeled data conditions?.

2 Methodology

Systematic literature search across multiple databases yielded 10 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 2.2/10.

3 Results

10 papers retrieved. 13 claims extracted; 0 independently verified. Quality review score: 2.2/10.

4 Limitations

This report is a machine-generated literature synthesis and does not constitute original research. Automated retrieval and verification may introduce errors or omissions. Review scores reflect automated assessment, not human peer review. Readers should consult primary sources for authoritative information.

5 Extracted Claims

Claim	Verified	Confidence
On the DBLP dataset, GNN-MultiFix-MLP3 achieved an Average Precision of 0.942.	×	0.05
On the Yelp dataset, the LANC method resulted in an Out Of Memory (OOM) error.	×	0.02
On the Yelp dataset, the GNN-LSPE method resulted in an Out Of Memory (OOM) error.	×	0.02
On the BlogCat dataset, GNN-MultiFix-Linear achieved an Average Precision of 0.225.	×	0.05
On the PCG dataset, GNN-MultiFix-MLP1 achieved an Average Precision of 0.254.	×	0.04
Removing the Feature Refinement (FR) component from GNN-MultiFix-Linear reduced the Average Precision on the DBLP datase	×	0.04
Removing the Label Refinement (LR) component from GNN-MultiFix-Linear reduced the Average Precision on the DBLP dataset	×	0.05
Removing the Positional Encoding (PE) component from GNN-MultiFix-Linear reduced the Average Precision on the BlogCat da	×	0.04
On synthetic datasets with original feature fraction (rori_feat) of 0.0, the GCN method achieved an Average Precision of	×	0.04
On synthetic datasets with label homophily (rhomo) of 1.0, the DeepWalk method achieved an Average Precision of 0.552.	×	0.06
GNN-LSPE was not run on the synthetic datasets due to its large runtime.	×	0.05
On the DBLP dataset, the MajorityVote baseline achieved an Average Precision of 0.869.	×	0.02
On the BlogCat dataset, the FSGNN method achieved an Average Precision of 0.075.	×	0.02

References

- http://arxiv.org/abs/2411.14094v1
- http://arxiv.org/abs/2504.20421v1
- http://arxiv.org/abs/2102.11485v3