

System-Level Optimizations for Mini-Batch GNN Training in Multimodal Vulnerability Detection

Assignee Research

June 1, 2026

Abstract

This report synthesises findings from 12 peer-reviewed papers addressing the following research question: To what extent do system-level optimizations for mini-batch GNN training improve inference throughput and memory efficiency when deploying multimodal vulnerability detectors on resource-constrained. Graph Neural Networks (GNNs) enable learning on realworld graphs and mini-batch training has emerged as the de facto standard for training GNNs because it can scale to very large graphs and improve convergence. Current mini-batch construction policies largely ignore efficiency. 13 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 2.5/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: Efficient GNN Training Through Structure-Aware Randomized Mini-Batching. Research question: To what extent do system-level optimizations for mini-batch GNN training improve inference throughput and memory efficiency when deploying multimodal vulnerability detectors on resource-constrained devices?.

2 Methodology

Systematic literature search across multiple databases yielded 12 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 2.5/10.

3 Results

12 papers retrieved. 13 claims extracted; 0 independently verified. Quality review score: 2.5/10.

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
The study evaluates four training metrics: final validation accuracy, per-epoch training time, number of epochs until convergence.	×	0.08
The setting (RAND-ROOTS & p=0.5) represents the uniform random mini-batching baseline.	×	0.08
The setting (NORAND-ROOTS & p=1.0) represents the entirely community-based mini-batching scheme.	×	0.07
The reddit dataset contains 232,965 nodes and 114,615,892 edges.	×	0.01
The igb-small dataset contains 1,000,000 nodes and 13,070,502 edges.	×	0.01
The ogbn-products dataset contains 2,449,029 nodes and 123,718,280 edges.	×	0.03
The ogbn-papers100M dataset contains 111,059,956 nodes and 3,228,124,712 edges.	×	0.01
Across the evaluated graphs, validation accuracy with varying graph structure bias is always within 1.79 percentage point.	×	0.10
The entirely community-based mini-batching scheme for ogbn-papers100M incurs a 4% point accuracy loss compared to the baseline.	×	0.08
The accuracy loss for entirely community-based mini-batching on ogbn-papers100M is attributed to its shorter maximum epoch.	×	0.07
There is often a trade-off between per-epoch training speedup and the number of epochs until convergence.	×	0.05
The COMM-RAND-MIX-k% scheme shuffles communities as blocks, combines a fixed number of communities (k% of total) into subgraphs.	×	0.05
The NORAND-ROOTS scheme involves no shuffling of the training set before dividing nodes across batches, resulting in stable training.	×	0.03

References

- <http://arxiv.org/abs/2504.18082v1>

- <http://arxiv.org/abs/2406.00552v4>
- <http://arxiv.org/abs/2605.09402v1>