

Mini-Batch vs Full-Graph Training in GNNs for Code Vulnerability Detection

Assignee Research

June 1, 2026

Abstract

This report synthesises findings from 16 peer-reviewed papers addressing the following research question: How does mini-batch training impact the convergence speed and final accuracy of Graph Neural Networks for code vulnerability detection compared to full-graph training on large-scale software. Full-graph and mini-batch Graph Neural Network (GNN) training approaches have distinct system design demands, making it crucial to choose the appropriate approach to develop. A core challenge in comparing these two GNN training approaches lies in characterizing their model. 14 claims were extracted from source literature; 1 was independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 4.2/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: Full-Graph vs. Mini-Batch Training: Comprehensive Analysis from a Batch Size and Fan-Out Size Perspective. Research question: How does mini-batch training impact the convergence speed and final accuracy of Graph Neural Networks for code vulnerability detection compared to full-graph training on large-scale software repositories?.

2 Methodology

Systematic literature search across multiple databases yielded 16 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 4.2/10.

3 Results

16 papers retrieved. 14 claims extracted; 1 independently verified. Quality review score: 4.2/10.

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
The accuracy of the Reddit dataset is 96.13% for full-graph training and 96.32% for mini-batch training.	×	0.11
The accuracy of the Ogbn-arxiv dataset is 70.96% for full-graph training and 71.16% for mini-batch training.	×	0.11
The accuracy of the Ogbn-products dataset is 77.92% for full-graph training and 78.80% for mini-batch training.	×	0.11
The accuracy of the Ogbn-papers100M dataset is 59.54% for full-graph training and 58.52% for mini-batch training.	×	0.11
n represents the number of nodes of the entire graph.	×	0.03
n_train and n_test represent the number of nodes in the training set and the testing set, respectively.	×	0.02
n_min represents the minimal value between training and testing sets.	×	0.01
X represents the node feature matrix and x_i represents the i-th row of the node feature matrix.	×	0.03
y_i represents the ground truth label of node i and \hat{y}_i represents the estimated outcomes of node i.	×	0.03
r represents the feature size.	×	0.05
b represents the batch size.	×	0.13
β represents the fan-out size.	✓	0.21
A_train and A_full represent the adjacency matrix in each mini-batch and full-graph training iteration, respectively.	×	0.08
D_train and D_full represent the diagonal in-degree matrices in each mini-batch and full-graph training iteration, respe	×	0.08

References

- <http://arxiv.org/abs/2605.22480v1>
- <http://arxiv.org/abs/2601.22678v3>

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