

# Graph Contrastive and Supervised Anomaly Detection Scaling on Heterophilic Graphs

Assignee Research

June 1, 2026

## Abstract

This report synthesises findings from 13 peer-reviewed papers addressing the following research question: How do graph contrastive anomaly detection models scale with graph size compared to supervised methods when evaluated on heterophilic graphs using the F1 score as the primary metric. With a long history of traditional Graph Anomaly Detection (GAD) algorithms and recently popular Graph Neural Networks (GNNs), it is still not clear (1) how they perform under a standard comprehensive setting, (2) whether GNNs can outperform traditional algorithms such as tree. 6 claims were extracted from source literature; 5 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 7.9/10. This report is a machine-generated literature synthesis and does not constitute original research.

## 1 Introduction

This paper examines: GADBench: Revisiting and Benchmarking Supervised Graph Anomaly Detection. Research question: How do graph contrastive anomaly detection models scale with graph size compared to supervised methods when evaluated on heterophilic graphs using the F1 score as the primary metric?.

## 2 Methodology

Systematic literature search across multiple databases yielded 13 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 7.9/10.

### 3 Results

13 papers retrieved. 6 claims extracted; 5 independently verified. Quality review score: 7.9/10.

### 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
GADBench is a benchmark tool dedicated to supervised anomalous node detection in static graphs.	✓	0.31
GADBench facilitates a detailed comparison across 29 distinct models.	✓	0.26
GADBench evaluates models on ten real-world GAD datasets.	✓	0.16
The datasets used in GADBench encompass from thousands to approximately 6 million nodes.	×	0.05
Tree ensembles with simple neighborhood aggregation can outperform the latest GNNs tailored for the GAD task.	✓	0.36
GADBench is open-sourced at <a href="https://github.com/squareRoot3/GADBench">https://github.com/squareRoot3/GADBench</a> .	✓	0.27

### References

- <https://doi.org/10.1145/3679018>
- <https://doi.org/10.48550/arxiv.2306.12251>
- <https://doi.org/10.1109/tkde.2026.3680353>