Adversarial Robustness of XSimGCL and Graph Contrastive Learning Methods on Amazon Data

Assignee Research

June 1, 2026

Abstract

This report synthesises findings from 11 peer-reviewed papers addressing the following research question: How does the adversarial robustness of XSimGCL compare to other GCL methods when evaluated on the Amazon dataset using NDCG@10 and NDCG@20 under similar perturbation levels. Contrastive learning (CL) has emerged as a powerful framework for learning representations of images and text in a self-supervised manner while enhancing model robustness against adversarial attacks. More recently, researchers have extended the principles of contrastive learning. 13 claims were extracted from source literature; 1 was independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 4.7/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: On the Adversarial Robustness of Graph Contrastive Learning Methods. Research question: How does the adversarial robustness of XSimGCL compare to other GCL methods when evaluated on the Amazon dataset using NDCG@10 and NDCG@20 under similar perturbation levels?.

2 Methodology

Systematic literature search across multiple databases yielded 11 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 4.7/10.

3 Results

11 papers retrieved. 13 claims extracted; 1 independently verified. Quality review score: 4.7/10.

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 empirical evaluation includes multiple node and graph classification datasets, as well as a variety of graph contras	✓	0.22
The evaluation measures both clean accuracy and perturbed accuracy under various static and adaptive attack scenarios.	×	0.05
The relative drop in accuracy compared to the clean accuracy is reported.	×	0.01
The model architectures used in the evaluation are as reported in the reference implementations, including the type and	×	0.03
The primary objective is to assess the extent to which GCL methods are affected by adversarial attacks.	×	0.13
Hyperparameter tuning is not performed; instead, the best hyperparameters for each model as documented in their original	×	0.02
The Adam optimizer is utilized across all models, and the learning rate is set as defined in the reference papers.	×	0.02
Results are averaged over 15 different initializations of the GCL encoder and linear classifier.	×	0.02
For non-contrastive models such as GCN and GIN, the network is trained in a supervised manner, and then Step 3 of the ro	×	0.12
The Relative Adversarial Accuracy Drop (Rclean_adv) is computed for all models.	×	0.03
The accuracy for different contrastive and non-contrastive models, and different datasets under the attack scheme that m	×	0.06
The attacks are allowed to change 5% of edges.	×	0.02
Models achieving the three highest relative drops are highlighted as first, second, and third in Table 1.	×	0.02

References

- http://arxiv.org/abs/2206.07869v1
- http://arxiv.org/abs/2104.09369v1

- http://arxiv.org/abs/2311.17853v2