

Homophily-Guided Self-Supervision and Reasoning Accuracy in Large Language Models on Perturbed Social Graphs

Assignee Research

May 31, 2026

Abstract

This report synthesises findings from 15 peer-reviewed papers addressing the following research question: How does homophily-guided self-supervision impact the reasoning accuracy of billion-parameter LLMs on perturbed social graph datasets compared to standard pre-training. High-level automation is increasingly critical in AI, driven by rapid advances in large language models (LLMs) and AI agents. However, LLMs, despite their general reasoning power, struggle significantly in specialized, data-sensitive tasks such as designing Graph Neural Networks. 13 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 3.9/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: Proficient Graph Neural Network Design by Accumulating Knowledge on Large Language Models. Research question: How does homophily-guided self-supervision impact the reasoning accuracy of billion-parameter LLMs on perturbed social graph datasets compared to standard pretraining?.

2 Methodology

Systematic literature search across multiple databases yielded 15 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 3.9/10.

3 Results

15 papers retrieved. 13 claims extracted; 0 independently verified. Quality review score: 3.9/10.

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
Existing benchmarks for model-performance relationships lack a data dimension, which limits their use for knowledge-driven	×	0.08
Only a few methods, specifically cited as [5, 31, 60], have attempted to design feasible models for unseen tasks using m	×	0.06
Existing methods for designing models for unseen tasks rely primarily on semi-supervised or supervised techniques that r	×	0.06
Auto2Graph translates descriptive user inputs, such as 'Cora is a citation network', into configuration specifications.	×	0.02
GPT4GNAS and GHGNAS iteratively refine model designs through crafted prompts focusing on different graph heterogeneity.	×	0.02
Current LLM-based methods for GNN design suffer from over-reliance on external descriptive user inputs, resulting in poo	×	0.08
Raw LLMs inherently lack deep expertise in fine-grained graph-GNN-performance relationships.	×	0.07
Existing methods refine model designs solely based on feedback from testing models, lacking actionable, knowledge-driven	×	0.04
The Cora dataset is a citation network among machine learning papers.	×	0.05
The Cora dataset contains 2,708 nodes and 5,278 edges (inferred from fragmented text '#G Nra op dh e sP : r 1o 7p ,e 2r	×	0.02
LLMs propose overly generic GNNs when provided with purely semantic and property descriptions, failing to induce data-sp	×	0.07
The benchmark datasets listed include Cora, PubMed, Physics, and Computers.	×	0.03
The proposed method ('Ours') is compared against GPT4GNAS, GHGNAS, GNAS, and Auto-GNN.	×	0.02

References

- <http://arxiv.org/abs/2408.06717v3>

- <http://arxiv.org/abs/2502.14293v2>
- <http://arxiv.org/abs/2206.07869v1>