SOVEREIGN: Generalizing GNNs with Tokenized Mixture of Experts

SOVEREIGN Research Kernel
Autonomous draft — Owner review required before publication

May 27, 2026

Abstract

Deployed graph neural networks (GNNs) are frozen at deployment yet must fit clean data, generalize under distribution shifts, and remain stable to perturbations. We show that static inference induces a fundamental tradeoff: improving stability requires reducing reliance on shift-sensitive features, leaving an irreducible worst-case generalization floor. Instance-conditional routing can break this ceiling, but is fragile because shifts can mislead routing and perturbations can make routing fluctuate. We capture these effects via two decompositions separating coverage vs selection, and base sens

1 Introduction

Analysis of: Generalizing GNNs with Tokenized Mixture of Experts. Research goal: What is the impact of SMoES routing on cross-dataset generalization robustness measured by ANLS accuracy on unseen document understanding benchmarks (e.g., InfographicsVQA, ChartQA) compared to hard-routed MoE-VLMs?.

2 Methodology

Multi-query arXiv search (4 parallel queries, Relevance-sorted). TF-IDF cosine semantic verification (bigrams, threshold=0.15). NIM nv-embedqa-e5-v5 (dim=1024) for semantic indexing. Tribunal v2: 3-role parallel review (SKEPTIC/VALIDATOR/SYNTHESIZER) with revision round if score $<$ 6.5.

3 Results

9 papers retrieved. 5 claims extracted, 5 verified. Tribunal: 7.5/10 $\rightarrow$ APPROVE (revision_round=0). Policy: AUTO_APPROVE.

4 Uncertainties

NIM free tier latency varies. TF-IDF verification is a weak signal. arXiv Relevance ranking is query-dependent. Tribunal consensus is LLM-based and prompt-sensitive.

5 Extracted Claims

Claim	Verified	Confidence
Deployed graph neural networks (GNNs) are frozen at deployment yet must fit clean data, generalize under distribution sh	✓	0.38
Static inference induces a fundamental tradeoff: improving stability requires reducing reliance on shift-sensitive featu	✓	0.39
Instance-conditional routing can break this ceiling, but is fragile because shifts can mislead routing and perturbations	✓	0.33
STEM-GNN is a pretrain-then-finetune framework with a mixture-of-experts encoder for diverse computation paths, a vector	✓	0.48
Across nine node, link, and graph benchmarks, STEM-GNN achieves a stronger three-way balance, improving robustness to de	✓	0.45

References

- http://arxiv.org/abs/2603.11114v1
- http://arxiv.org/abs/2410.17954v2
- http://arxiv.org/abs/2602.09258v1