

What is the reduction in LLM API call overhead achieved by GNN-RAG compared to relation-path generation method

Assignee Research

June 10, 2026

Abstract

Knowledge Graphs (KGs) represent human-crafted factual knowledge in the form of triplets (head, relation, tail), which collectively form a graph. Question Answering over KGs (KGQA) is the task of answering natural questions grounding the reasoning to the information provided by the KG. Large Language Models (LLMs) are the state-of-the-art models for QA tasks due to their remarkable ability to understand natural language. On the other hand, Graph Neural Networks (GNNs) have been widely used for KGQA as they can handle the complex graph information stored in the KG. In this work, we introduce GN

1 Introduction

This paper examines: GNN-RAG: Graph Neural Retrieval for Large Language Model Reasoning. Research question: What is the reduction in LLM API call overhead achieved by GNN-RAG compared to relation-path generation methods on the WebQSP dataset?.

2 Methodology

Systematic literature search across multiple databases yielded 12 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 3.5/10.

3 Results

12 papers retrieved. 0 claims extracted; 0 independently verified. Quality review score: 3.5/10.

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.

References

- <http://arxiv.org/abs/1609.04846v1>
- <http://arxiv.org/abs/2606.05868v1>
- <http://arxiv.org/abs/2405.20139v1>