

Advanced Graph Data Augmentation Impacts on Multimodal Model Robustness and Efficiency

Assignee Research

June 2, 2026

Abstract

This report synthesises findings from 11 peer-reviewed papers addressing the following research question: In what ways do advanced graph data augmentation techniques influence the inference efficiency and robustness of multimodal models trained on sparse, large-scale knowledge graphs compared to. The rise of graph-structured data has driven major advances in Graph Machine Learning (GML), where graph embeddings (GEs) map features from Knowledge Graphs (KGs) into vector spaces, enabling tasks like node classification and link prediction. However, since GEs are derived from. 12 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 2.8/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: Improving Graph Embeddings in Machine Learning Using Knowledge Completion with Validation in a Case Study on COVID-19 Spread. Research question: In what ways do advanced graph data augmentation techniques influence the inference efficiency and robustness of multimodal models trained on sparse, large-scale knowledge graphs compared to traditional baseline augmentations?.

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 2.8/10.

3 Results

11 papers retrieved. 12 claims extracted; 0 independently verified. Quality review score: 2.8/10.

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 study used Neo4j 5.24.0 GDB1 and the Graph Data Science (GDS) library 2.12.02 to compute measures.	×	0.03
The dataset describes face-to-face interactions between individuals within an office building in France.	×	0.01
The dataset is structured in two files: one containing temporal edges of the form $t i j$, and another mapping each indivi	×	0.04
The dataset was converted into a Knowledge Graph (KG) with nodes representing users and departments, and edges represent	×	0.06
Each HAS_CONTACT_WITH edge has parameters for total contact time and timestamp.	×	0.02
Each node in the KG is associated with contagion probability, degree centrality, total contact time, and average contact	×	0.02
The transitive closure algorithm identified connected components of individuals at each 20-second interval and generated	×	0.02
The number of contacts between users increased from 1694 to 1882, an increase of 11.8%.	×	0.00
The PageRank values for individuals changed when comparing the results without KC and with KC.	×	0.02
Recent models like BiGI and SeHGNN emphasize the importance of global information or preaggregated multi-hop neighborhood	×	0.03
There is a lack of methods that are agnostic to both the embedding phase and the hyperparameter tuning process.	×	0.04
The proposed pre-processing strategy of Knowledge Completion (KC) has not been explored in prior literature.	×	0.08

References

- <http://arxiv.org/abs/2201.03812v3>
- <http://arxiv.org/abs/2511.12071v1>

- <http://arxiv.org/abs/2203.07969v1>