Metapath Sampling in MECCH vs Random Walks for Heterogeneous Graph Classification

Assignee Research

May 31, 2026

Abstract

This report synthesises findings from 8 peer-reviewed papers addressing the following research question: How do different metapath sampling strategies in MECCH compare to traditional random walk sampling in heterogeneous graph node classification benchmarks like ACM or DBLP in terms of classification. Heterogeneous graph neural networks (HGNNs) were proposed for representation learning on structural data with multiple types of nodes and edges. To deal with the performance degradation issue when HGNNs become deep, researchers combine metapaths into HGNNs to associate nodes. 11 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 3.5/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: MECCH: Metapath Context Convolution-based Heterogeneous Graph Neural Networks. Research question: How do different metapath sampling strategies in MECCH compare to traditional random walk sampling in heterogeneous graph node classification benchmarks like ACM or DBLP in terms of classification accuracy and training efficiency?.

2 Methodology

Systematic literature search across multiple databases yielded 8 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

8 papers retrieved. 11 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.

5 Extracted Claims

Claim	Verified	Confidence
MECCH was evaluated on five heterogeneous graph datasets: IMDB, ACM, DBLP, LastFM, and PubMed.	×	0.13
For node classification, MECCH was tested on IMDB, ACM, and DBLP datasets.	×	0.07
For link prediction, MECCH was tested on LastFM and PubMed datasets.	×	0.10
The IMDB dataset contains 12,722 nodes of 3 types and 37,288 edges of 4 types.	×	0.02
The ACM dataset contains 8,994 nodes of 3 types and 25,922 edges of 4 types.	×	0.03
The DBLP dataset contains 18,405 nodes of 3 types and 67,946 edges of 4 types.	×	0.02
The LastFM dataset contains 20,612 nodes of 3 types and 201,908 edges of 5 types.	×	0.03
The PubMed dataset contains 63,109 nodes of 4 types and 368,245 edges of 16 types.	×	0.02
On the ACM dataset, MECCH achieved a training time of 0.049 seconds, validation time of 4.0 seconds, and testing times o	×	0.04
On the DBLP dataset, MECCH achieved a training time of 0.181 seconds, validation time of 4.2 seconds, and testing times	×	0.04
MECCH's performance in terms of Macro-F1 and Micro-F1 scores varies with metapath length (K) and number of layers (L).	×	0.05

References

- http://arxiv.org/abs/2110.13324v1
- http://arxiv.org/abs/1102.4599v1
- http://arxiv.org/abs/2211.12792v2