Multi-Hop Feature Quality Estimation for Unsupervised Graph Representation Learning

25-2-R-10

Supervisors:
Dr. Renata Avros
Prof. Zeev Volkovich

Submitters:
Gad Azriel
Lidor Kupershmid

Introduction

Problem:

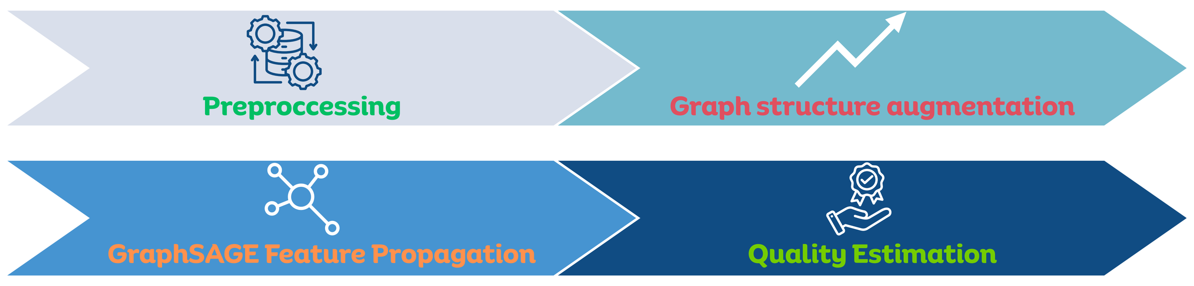
Graph Neural Networks (**GNNs**) achieve strong performance but are highly sensitive to noisy node features.
Noise propagates across hops, amplifying errors and degrading representations.
Existing methods cannot reliably estimate feature quality in an unsupervised and scalable way.

Project Goal:

To build a scalable, noise-aware graph learning system that identifies unreliable node features and produces robust embeddings by combining Multi-Hop Quality Estimation with GraphSAGE.

Outcome

MQE + GraphSAGE achieves an average classification accuracy of 87.5%, improving standalone MQE performance by 7.5%.



Proposed Model

- Learn node embeddings using GraphSAGE-based inductive multi-hop propagation with neighbor sampling.
- Estimate feature reliability at each hop using Multi-Hop Quality Estimation (MQE) via Gaussian modeling (mean μ, variance σ).
- Suppress unreliable or noisy features by down-weighting high-uncertainty (σ) representations during propagation.
- Produce noise-aware node embeddings that improve robustness under feature corruption.

Why GraphSAGE?

GraphSAGE enables scalable, inductive learning on large graphs, while MQE explicitly models feature uncertainty, preventing noise amplification across hops.

Model Validation

- Injected synthetic feature noise into benchmark citation graphs to simulate real-world corruption.
- Evaluated node classification accuracy under increasing noise levels.
- MQE-GraphSAGE maintained stable performance while standard baselines degraded.
- Demonstrated robustness across datasets and multiple propagation depths.

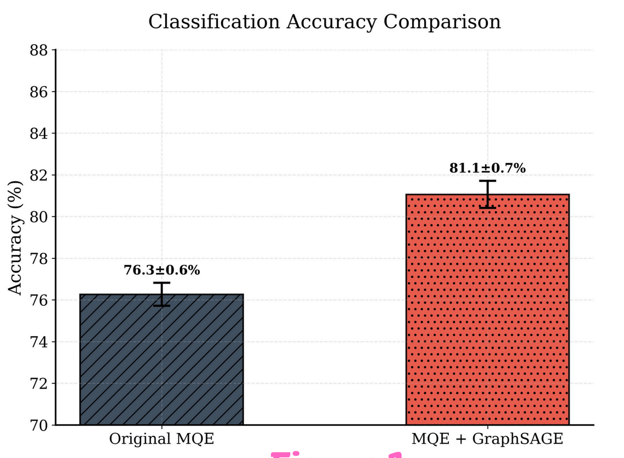


Figure 1
MQE-GraphSAGE achieves higher classification accuracy than the original MQE, showing improved feature propagation and noise suppression.

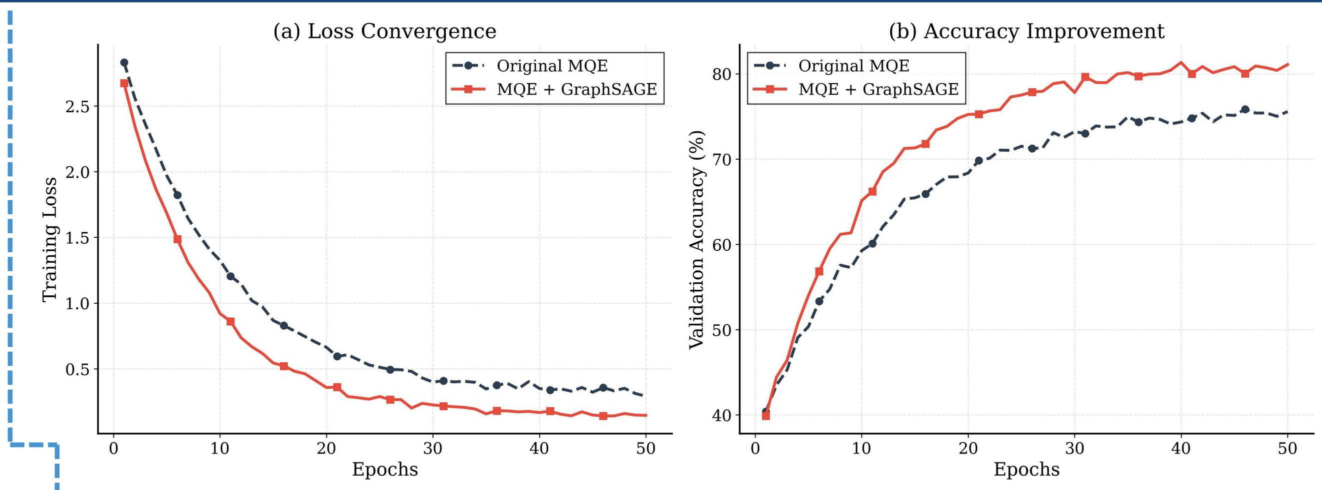


Figure 2a (Loss Convergence)
MQE-GraphSAGE converges faster and reaches lower training loss, indicating more stable and efficient optimization.

Figure 2b (Accuracy Improvement)
Validation accuracy of MQE-GraphSAGE consistently exceeds the original MQE, demonstrating stronger robustness to noisy features.

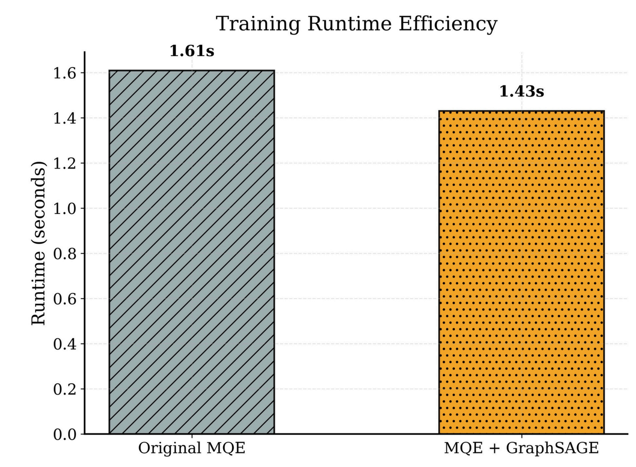
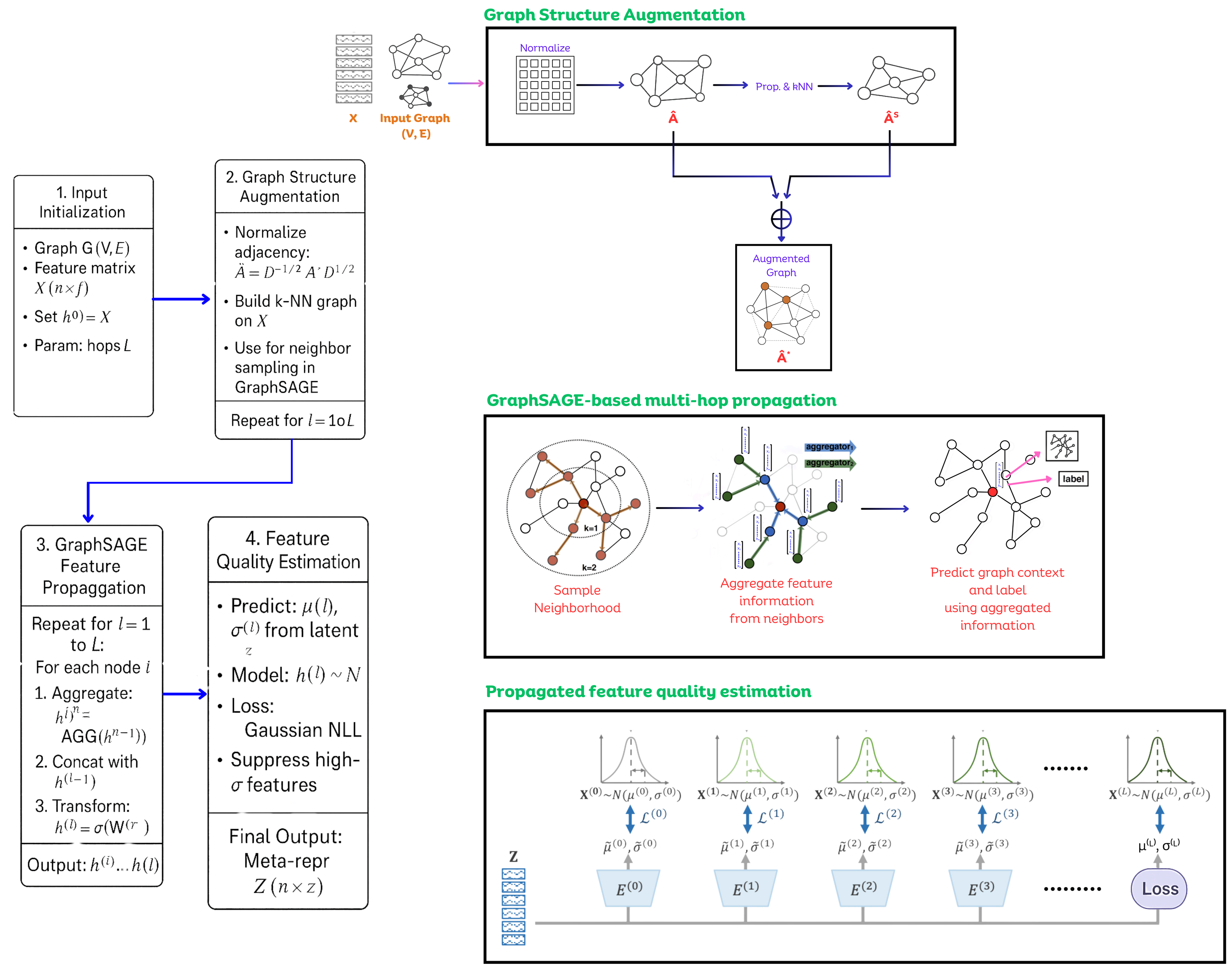


Figure 3
MQE-GraphSAGE reduces training runtime through mini-batch learning and neighbor sampling, enabling scalable graph training.



Impact Factor Analysis with Noise-Aware Graph Learning

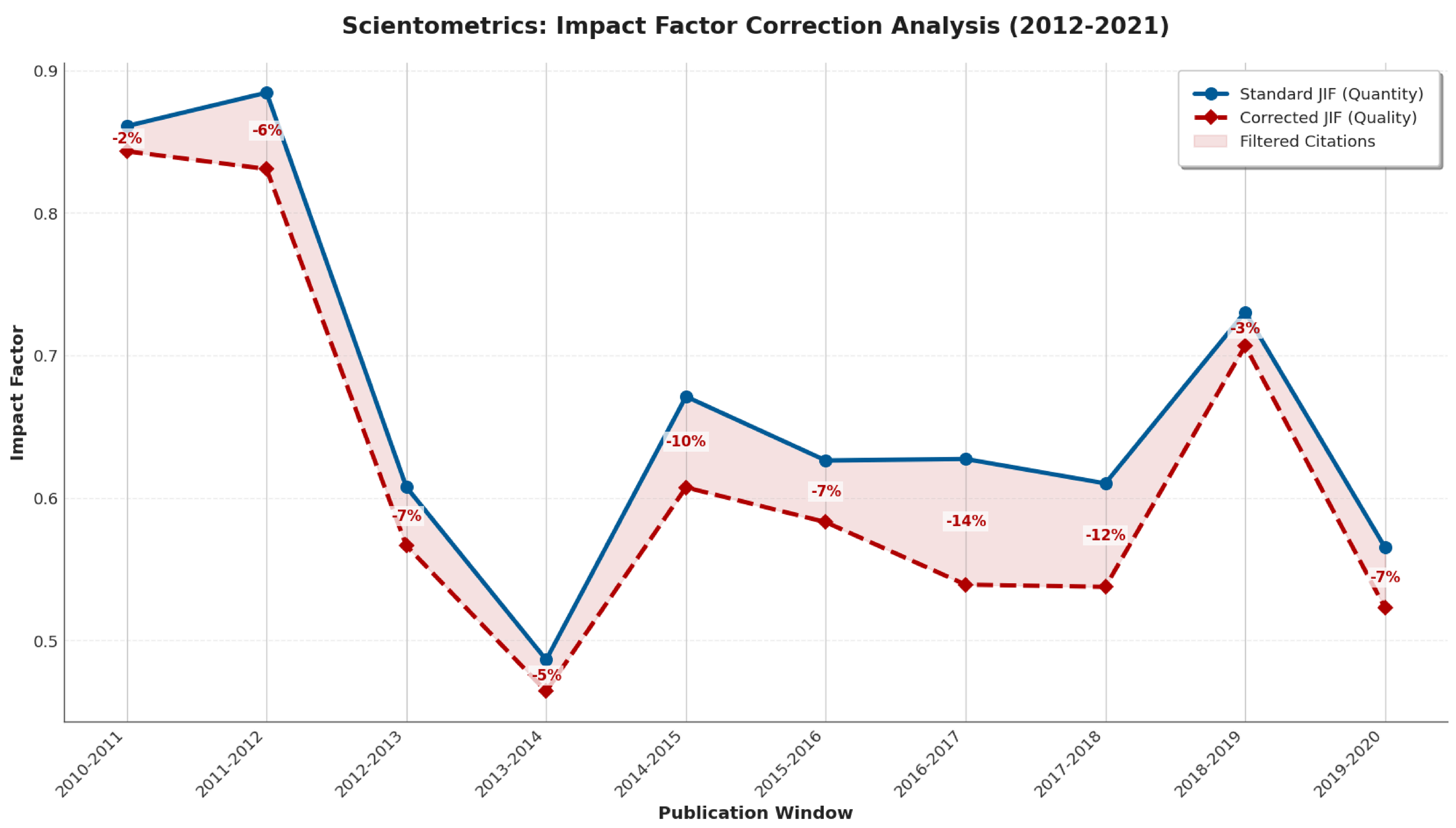


Figure 4 - Impact Factor Correction Using MQE-GraphSAGE

We analyze the evolution of impact factor scores over time using sliding publication windows, comparing standard citation-based estimates with noise-aware corrections produced by MQE-GraphSAGE.
By estimating feature uncertainty and filtering unreliable citation signals, the corrected impact factor exhibits smoother trends and reduced sensitivity to noisy or short-lived citation spikes.

Key observations:

- MQE-GraphSAGE consistently reduces inflated impact estimates caused by noisy citation patterns.
- Temporal fluctuations are moderated, yielding more stable and interpretable impact trajectories.
- The correction effect varies across time windows, highlighting periods where citation noise has the strongest influence.

This analysis demonstrates that **noise-aware graph modeling can improve the reliability of bibliometric indicators**, offering a principled alternative to purely quantity-driven impact metrics.

Conclusions

The project successfully achieved its goal of improving robustness in graph representation learning under noisy node features by integrating Multi-Hop Quality Estimation with GraphSAGE.

Key Results:

- Classification Performance: MQE-GraphSAGE consistently outperformed the original MQE model in node classification accuracy under feature noise.
- Robustness to Noise: The model effectively suppressed unreliable features, maintaining stable performance as noise intensity increased.
- Scalability & Efficiency: Inductive learning with neighbor sampling enabled faster convergence and reduced training runtime.

Minor challenges, such as hyperparameter tuning and noise calibration, were addressed through controlled experimentation.
Overall, this work demonstrates that variance-aware, inductive modeling provides a practical and effective solution for robust learning on real-world graph data.