Think Global And Local:
Graph Learning with Full Fusion / Interleaved Architectures on Tabular Data

Author: Andrei Stefan (andreistefan@tudelft.nl)
Supervisors: Dr. Kubilay Atasu, Çağrı Bilgi

TU Delft

01 Introduction

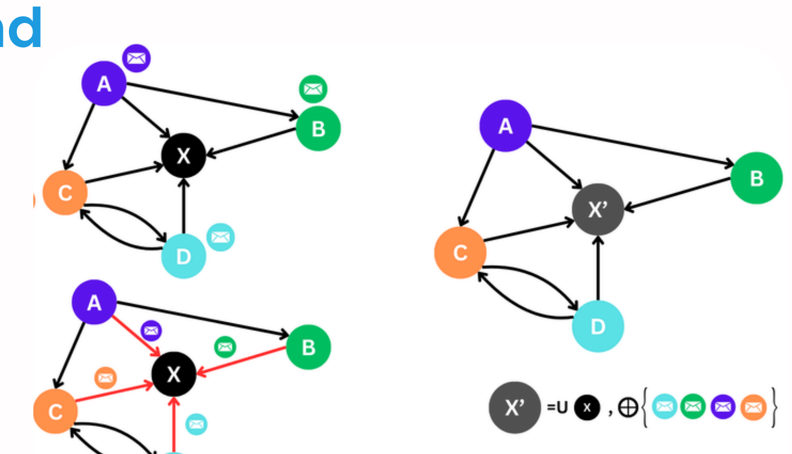
This project explores how machine learning models, specifically Graph Neural Networks (GNNs) and Transformers, can be applied to relational data that has been transformed into graph structures. The focus is on the financial sector, where fraud detection is becoming more critical than ever due to the increasing sophistication of fraudulent schemes.

02 Research Question

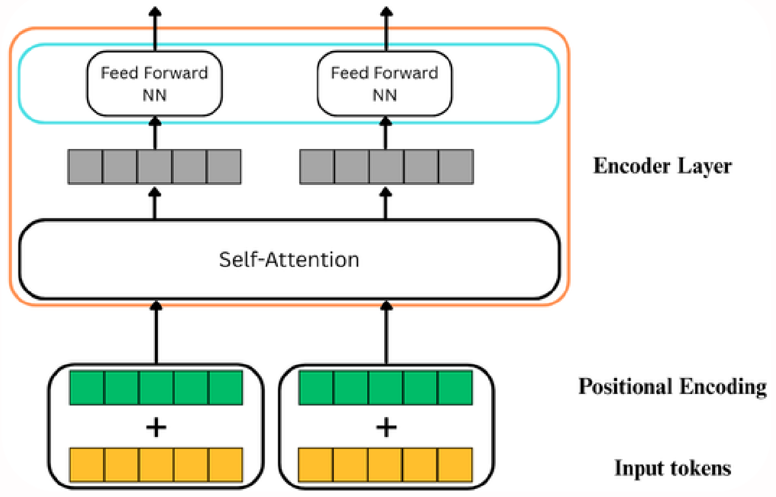
How well does a Full-Fusion or Interleaved architecture using Transformers on the edges of sampled subgraph and GNNs on nodes perform on the IBM Transactions for Anti Money Laundering (AML) [1] dataset, compared to existing literature?

03 Background

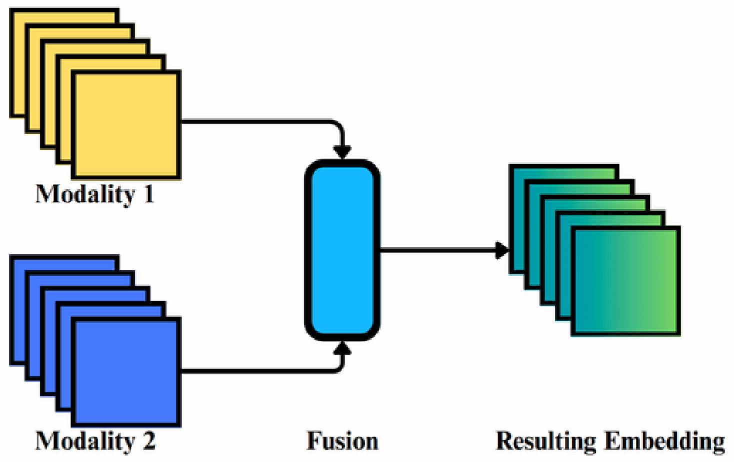
Message Passing Neural Network (MPNN)
Generates embeddings capturing local structure by aggregating neighbouring data.



Transformer Encoder
Uses multi-head Attention to bypass edge constraints and gathers a global view of the input. Is heavily helped by positional encoding schemes, which give a greater perspective of the data.



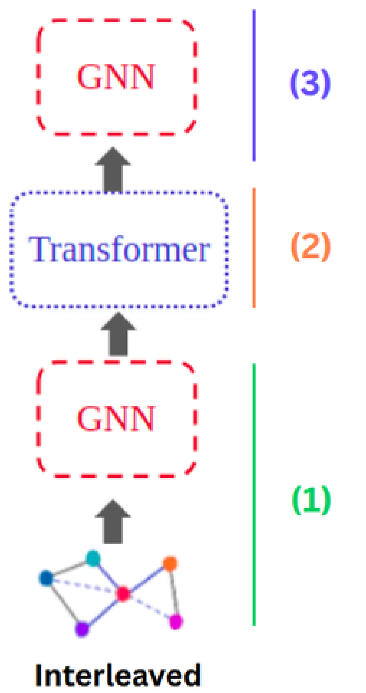
Fusion Layers
Merge different modalities using aggregation methods, from as simple as concatenating embeddings, to learnable methods using memory gates.



04 Full Fusion and Interleaved Architecture on edges

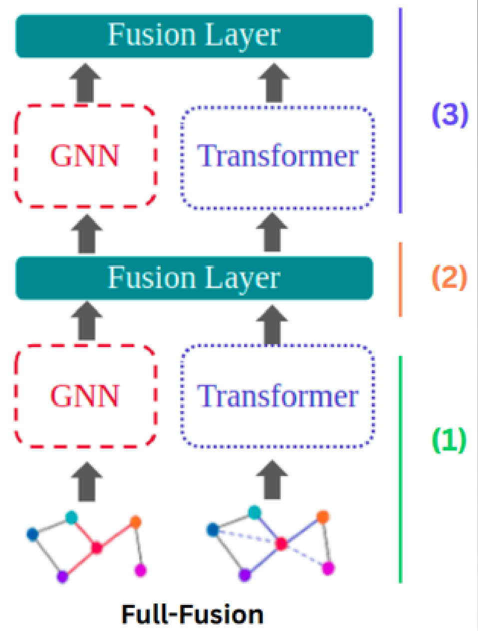
Interleaved:

- The input first passes through a GNN to compute local node and edge embeddings.
- These embeddings are then refined by a Transformer that introduces global edge-aware attention.
- The result is passed through another GNN to further integrate the global context with local structure.



Full Fusion:

- Inputs are processed simultaneously by a GNN (capturing local node structure) and a Transformer (capturing global structure via edge-focused attention).
- Their outputs are merged using fusion techniques (we experiment with Concatenate+MLP and Gated Multimodal Units).
- The fused representation is then passed through the same dual-path process to uncover deeper patterns



05 Methodology

Dataset & Task
Experiments were conducted on the IBM AML-Small datasets, targeting **binary edge classification**: predicting whether a financial transaction is illicit. Data was reformatted into tabular structure and then converted to graphs. A 60/20/20 temporal split ensured balanced daily coverage across training, validation, and test sets.

Training Setup
Models were implemented in PyTorch/PyG and trained for up to 60 epochs using the AdamW optimizer with cosine warmup scheduling and gradient clipping (max norm 1.0). Dropout, activation functions, and learning rates were tuned empirically. Due to memory constraints, effective batch size was capped at 4096 using accumulation when needed. All runs were seeded.

Evaluation
Performance was measured using F1-score due to class imbalance.

Baselines

- PNA GNN (Corso et al.)[2]: widely used AML benchmark
- MEGA-PNA (Bilgi et al.)[3]: current SOTA for edge classification on AML

Hardware
Experiments ran on DAIC HPC with NVIDIA A40 and L40 GPUs.

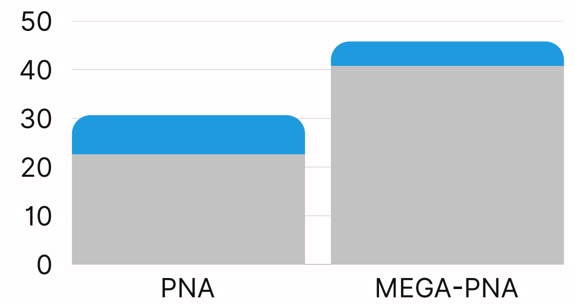
06 Results & Discussion

The Interleaved and Full Fusion architectures demonstrate consistent improvements over both the PNA baseline and MEGA-PNA with F1 score increases reaching up to 10% on datasets characterized by lower levels of illicit activity.

- Incorporating transformer-based edge embeddings leads to **reliable performance gains**, particularly in low-illicit datasets, even when parameter counts remain comparable.
- The inclusion of the **GMU Fusion** layer improves embedding prioritization across fusion stages, resulting in enhanced robustness, lower variance across runs, and reduced model complexity.
- The use of **R-PEARL** positional encodings yields mixed outcomes: performance gains of up to 5% F1 are observed in low-illicit datasets, while results in high-illicit settings show reduced consistency, indicating sensitivity to initialization and hyperparameters.

Overall, the proposed architectures **enhance** both **performance** and result **stability**, particularly in datasets with a low amount of illicit transactions.

Scan me for the paper and full results!



Improvements on the LI dataset visualized (Baseline VS Best model)

Model\Dataset	Small_LI	Small_HI	#Params
PNA + edge updates	22.66 ± 2.97	65.29 ± 2.33	32,197
MEGA-PNA	40.80 ± 2.95	73.20 ± 0.45	41,837
Interleaved	30.87 ± 3.21	68.27 ± 3.27	63,297
Interleaved w/ R-PEARL	30.14±2.87	65.21 ± 4.03	64,463
Interleaved w/ MEGA	**46.05 ± 0.55**	**74.36±0.67**	82,577
Full-Fusion w/ GMU	29.34 ± 1.77	69.71 ± 1.75	71,137
Full-Fusion w/ GMU, MEGA	**45.85 ± 1.59**	**74.97 ± 0.78**	90,417

Minority class F1-score on the AML datasets (%). Best results are highlighted.

07 Future Work

Future work:

- Test the Full-Fusion architectures on the small datasets and report results;
- Improve the models with more sophisticated approaches;
- Test the final models on all datasets considered, report data and compare with PNA and other baselines

References

[1] IBM Research. *IBM Transactions for Anti Money Laundering (AML)*. https://www.kaggle.com/datasets/ealtman2019/ibm-transactions-for-anti-money-laundering-aml. 2023.(Visited on 04/22/2025).
[2] Gabriele Corso et al. *Principal Neighbourhood Aggregation for Graph Nets*. arXiv:2004.05718 [cs]. Dec. 2020. doi: 10.48550/arXiv.2004.05718. url: http://arxiv.org/abs/2004. 05718 (visited on 04/24/2025).
[3] H. Cagrı Bilgi, Lydia Y. Chen, and Kubilay Atasu. Multigraph message passing with bi-directional multi-edge aggregations, 2024.