AUTHOR: Sorin-Catalin Enachioiu
(senachioiu@tudelft.nl)

SUPERVISORS: Kubilay Atasu, Cagri Bilgi

GRAPH LEARNING ON TABULAR DATA
CASCADE AND INTERLEAVED MODELS

TU Delft

01. Introduction

This project explores how **Graph Neural Networks** (GNNs) that capture **local** information within a graph can be combined with **Transformers**, that capture **global** information within a graph in order to **perform fraud detection** on a series of **synthetically generated transactions** [1] that are stored in a tabular format. This is possible because tabular data has an implicit graph structure.

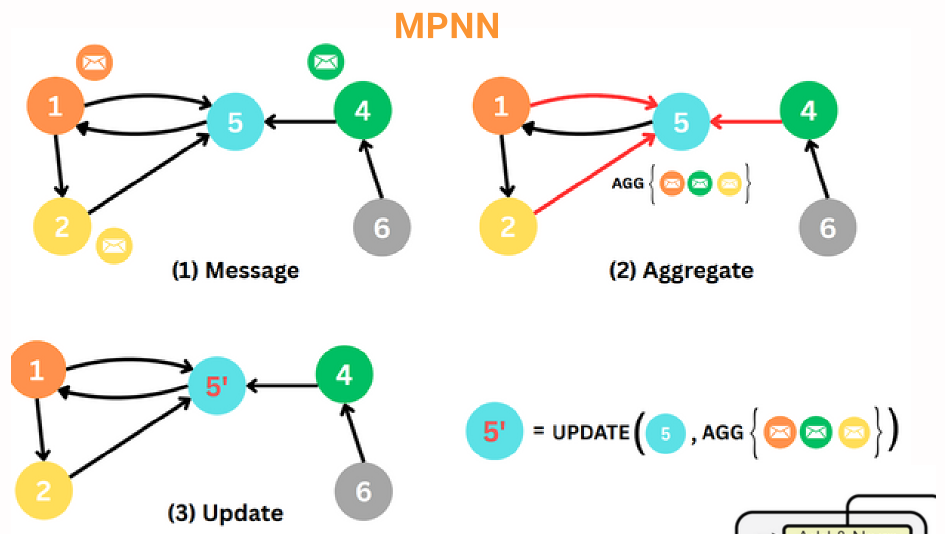
RESEARCH QUESTION

*How well do **Cascade** and **Interleaved** model architectures perform on the IBM **Anti Money Laundering** (AML) **datasets** compared to a **PNA baseline** ?*

02. Background

Message Passing Neural Networks (MPNN)

Framework for neural networks on graph data, in which **node embeddings** are updated in three steps.



Transformer Encoder

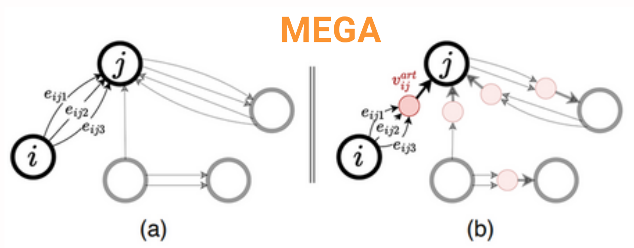
A model that outputs new embeddings that have values based on a **weighted combination** of the **initial embeddings with other embeddings that are similar in the context**.

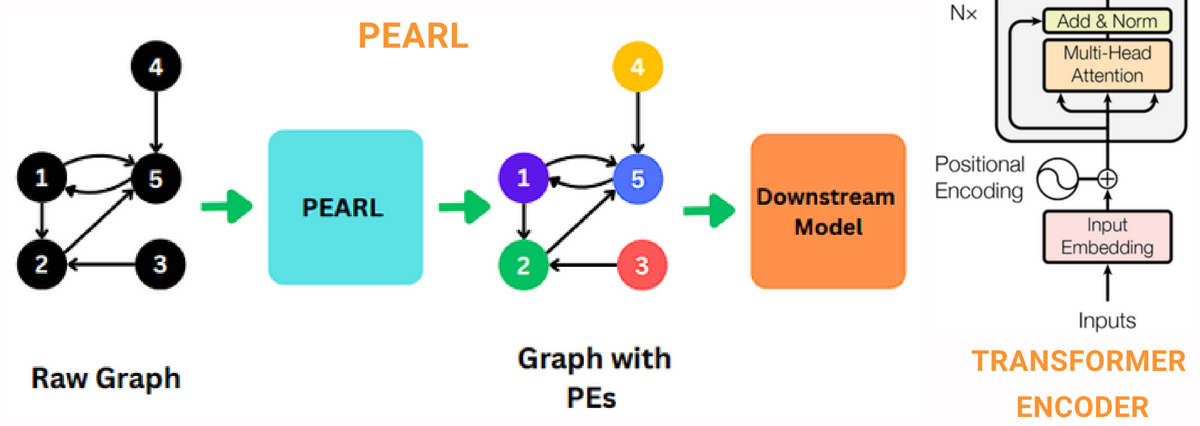
PEARL

Learnable **Positional Encodings** model.

MEGA

Multigraph adaptation for MP with two-stage aggregation.

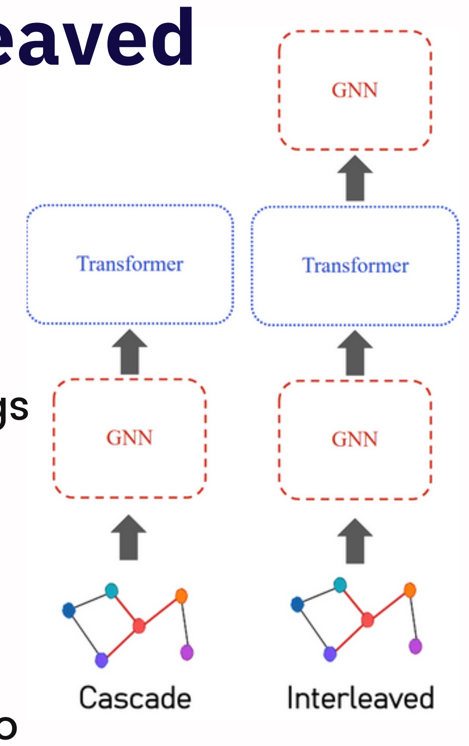




03. Cascade & Interleaved

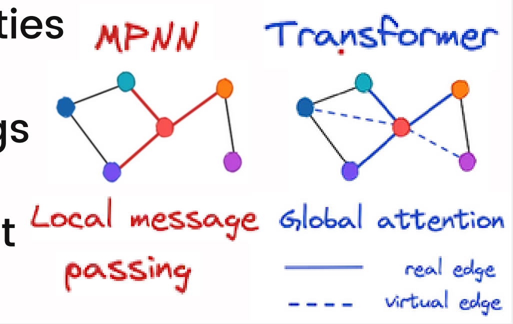
Cascade:

- **GNN:** Takes a graph as input and makes node embeddings that encode **local** neighborhood information
- **Transformer:** Uses these embeddings to look for **global** similarities.
- These similarities are then used to predict which transactions are illicit



Interleaved:

- **Cascade** : Use the cascade model to get embeddings based on similarities of local neighborhoods
- **GNN**: Tries to refine the embeddings with local information
- Predict which transactions are illicit



04. Experimental Setup and Results

Datasets

- **Small_HI –** high illicit transactions count
- **Small_LI –** low illicit transactions count
- 60-20-20 temporal split train-test-validation

Implementation

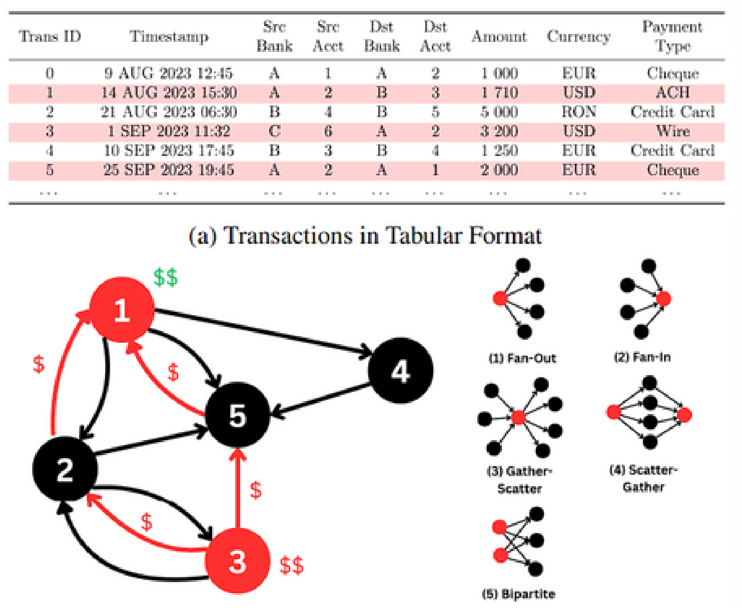
- Used **Pytorch & Pytorch Geometric** for models
- **mean ± standard** deviation for each experiment, computed over five runs, and initialized with random seeds, run for 80 epochs.

Baselines

- PNA, PNA + PEARL, PNA + MEGA, PNA + PEARL + MEGA.

Evaluation

- **F1 score to assess performance –** good for **highly imbalanced datasets**
- We **report** the **F1 test score corresponding to the highest F1 score in validation**



(a) Transactions in Tabular Format

(b) Transactions in Graph Format

(c) Fraud Patterns

PNA Ablation study

- The addition of **PEARL** to PNA improves mostly the LI results with almost 4%.
- The addition of **MEGA** brings massive improvements on both HI and LI.
- The combination of **MEGA** and **PEARL** sees the biggest improvements, showing they bring complementary benefits.

Cascade & Interleaved vs PNA

- **Cascade** and **Interleaved** improve the performance of PNA on most of the configurations.
- The addition of **PEARL** mostly improves the performance of all the models. Yet, it leads to a decreases in results for Cascade + PEARL on HI.
- The addition of **MEGA** has given significant improvements for all the models.
- The combination of **MEGA** and **PEARL** has made PNA and Cascade achieve the best results. Yet, Interleaved achieved worse results than with just **MEGA**.

Models	Small-HI	Small-LI	#params
PNA	63.48 ± 3.56	21.67 ± 1.96	32,197
+ PEARL	65.57 ± 3.97	25.35 ± 2.88	33,547
+ MEGA	72.58 ± 1.35	43.71 ± 1.14	41,837
+ MEGA + PEARL	$\underline{74.45 \pm 0.89}$	45.00 ± 1.02	43,187
Cascade	64.99 ± 2.31	25.42 ± 0.37	37,257
+ PEARL	61.84 ± 5.55	27.81 ± 3.42	38,607
+ MEGA	73.25 ± 0.66	45.50 ± 0.98	46,897
+ MEGA + PEARL	73.64 ± 1.82	46.07 ± 0.93	48,247
Interleaved	68.40 ± 2.86	31.30 ± 3.17	64,937
+ PEARL	67.63 ± 1.15	34.86 ± 2.17	66,287
+ MEGA	75.31 ± 0.45	$\underline{46.05 \pm 0.88}$	84,217
+ MEGA + PEARL	73.28 ± 2.18	44.15 ± 0.89	85,567

05. Discussion

- **PEARL** is able to improve model performance at low cost, both memory and space.
- **PNA** improves with both the addition of **PEARL** and **MEGA** and their combination as well.
- **Cascade** and **Interleaved** improve consistently the performance of PNA.

06. Conclusion and Future Work

- This research shows that local MP performance can be enhanced by combining it with global MP.
- Learnable positional encodings, PEARL, show promise in enhancing the results of the models on the AML datasets.

Future work can focus on:

- Use FraudGT instead of just a Transformer Encoder.
- integrate PEARL into other models
- integrate FraudGT model into Cascade and Interleaved