FinTech Industrial Project Planning Report

Student Name & ID: DU Jiaxing, 1155204280

Project title: Graph Neural Network Implementation

Sponsoring Organisation: CEFAR

Industrial Supervisor: Cobi XU, A.I. Phoenix Technology Co

# Project Goal

Employ Graph Neural Network (GNN) in the field of Fintech to enhance the modeling of complex relationships and interactions within financial systems. It involves learning from non-structural data via message passing between the nodes and graphs, as well as training a model collaboratively while keeping data locally store and secure. The project aims to develop an improved algorithm with usage guidance, which can effectively address various challenges, like data representation and computational complexity in Fintech applications.

# Significance

Graph neural networks (GNN) have gained attention as a powerful tool for modeling complex relationships and interactions. First, relationships among financial entities can be muti-modal and heterogeneous data representation using flexible hyperedges. To handle graphs with a huge number of modes and multiple types of relations, it is necessary to develop an efficient and scalable implementation of GNN. Hyper Graph Neural Network (HGNN), a variant of GNN, can embed high-order correlation beyond pairwise connections for data modeling. Second, rich data such as branches of financial institutions is personal or proprietary, and not meant to be shared, making privacy a critical concern and barrier to centralized data collection and model training. Federate learning networks allow multiple financial entities to train a model collaboratively by communicating gradients without sharing source data. This approach is robust to unbalanced and non-IID data distributions, and the communication costs are reduced significantly.

# Problem Statement

1. Quality and consistency of raw data: financial datasets in different scenarios often suffer from inconsistencies, missing values, noisy data, etc.
2. Training efficiency of large-scale datasets: efficient training and inference on large-scale graphs with vast nodes and edges require novel techniques to handle computational complexity, memory requirements, and optimization challenges.
3. Model complexity and overfitting: The number of layers and number of hidden units of GNNs play a crucial role in their performance. It's essential to find a balance between better representation capacity and overfitting.
4. Robustness and adversarial attacks: GNNs can be easily fooled by small perturbations on the input, called adversarial attacks.

# Proposed solutions

1. Preprocess and clean the raw data, including normalization, extreme value handling, etc. It helps to improve stability and convergence of GNN models.
2. Perform experimental analysis using existing HGNN frameworks, with the help of federated learning to train on decentralized data collaboratively.
3. Adopt regularization techniques, careful hyperparameter tuning to valid models, in order to mitigate overfitting and ensure generalization performance.
4. Improve current algorithms based on gap between the models can achieve and the needs of actual scenarios. Repeat this step for several times, record and analyze the validation process.
5. Develop a prototype for FinTech application with basic API tools and complete usage and development guidelines.

# Proposed timelines

*Sep 23 – Oct 23*

Related knowledge preparation, experiment on classic datasets

*Nov 23 – Dec 23*

Related knowledge preparation, formulate potential problems in a Fintech scenario (e.g., commercial bank and fund company)

*Jan 24*

Design the plan for 2nd semester, datasets preparation, and model selection

*Feb 24 – Apr 24*

GNN model construction, data feeding, model training and validation, parameters tuning, summarize the solution and complete the final report

*Mid Apr 24*

Wrap up, present results and findings