Project Proposal: Classifying Illicit Bitcoin Transactions using Graph Neural Networks

Team Members: Xuhui Zhan, Siyu Yang, Tianhao Qu

Github Repo: https://github.com/xuhuizhan5/BTCGraphGuard

Overview:
Our project focuses on classifying Bitcoin transactions as "licit" or "illicit" using Graph Neural Networks (GNNs) with the Elliptic Dataset. This dataset contains transaction graphs from the Bitcoin blockchain with nodes labeled as licit, illicit, or unknown. Since approximately 70% of nodes are labeled unknown, we will train and validate GNN models on known labels and predict unknown labels. We plan to compare various GNN architectures (Graph Convolutional Networks, Graph Attention Networks, and Temporal GNNs) and evaluate their performance.

Dataset:
The Elliptic Data Set (https://www.kaggle.com/datasets/ellipticco/elliptic-data-set/data) includes:

- 203,769 nodes and 234,355 edges

- 166 features per node (transactional and aggregated features)

- Labels: Illicit (2%), Licit (21%), Unknown (77%)

Methods & Innovations:
We will:

- Train different GNN models (GCN, GAT, Temporal GNN) on labeled data.

- Predict labels for unknown transactions.

- Analyze graph-level properties (degree distributions, clusters, centrality).

- Assess similarity and differences among subgraphs predicted by each GNN model.

Weekly Timeline:

Date	Task
Mar 25 (Proposal)	Submit proposal; Data exploration
Mar 25 - Apr 1	Data preprocessing; implement baseline GCN
Apr 1 (Project Update)	Preliminary results with GCN; basic dataset analysis
Apr 2 - Apr 8	Implement Graph Attention Networks (GAT)
Apr 9 - Apr 14	Implement Temporal GNN; Initial comparisons
Apr 15 & 17 (Presentation)	Present methods and preliminary findings
Apr 16 - Apr 22	Predict unknown nodes; compare predicted subgraphs
Apr 23 - Apr 25	Analyze graph similarity across methods; finalize results
Apr 25 (Final Submission)	Submit code, documentation, and final paper

This analysis will enhance understanding of GNN performance in financial anomaly detection and provide insights into structural similarities between predicted illicit transaction networks.