

Project Proposal

Inwon Kang

April 2022

For this project, I would like to use this dataset from kaggle about fraudulent transactions happening on the Bitcoin Network ¹. This dataset, original presented by [2], contains information about 203,769 transactions that took place on the Bitcoin network as nodes, and the asset flow between these transactions as the edges, of which there are 234,355.

Each node has 166 features including the timestamp, rest of which are not specifically labeled due to intellectual property issues. There are 49 evenly spaced timestamps in the dataset, and each timestamp is associated with a single connected component of the whole graph.

As for the labels, there are 4,545 fraudulent transactions, which take up around 2% of the nodes. There are 42,019 (27%) verified non-fraudulent transactions, and the rest are labelled as unknown. In the original paper that this dataset was introduced in, the authors made use of GCN models to classify nodes, in combination with RNNs to account for the temporal nature of the transaction.

Using this dataset, I would like to:

1. Apply variations of GAT models to classify the nodes (3-class problem)
2. If time allows, attempt to combine the GAT model with a temporal-aware model (LSTM or transformers models) to complete the final model.

GAT [1] are a new class of graph neural networks, and incorporate an attention mechanism to allow weights of neighboring nodes to influence the weight of each node. Basing from the assumption that some fraudsters may be working with each other and could be sending money between them, I believe this model choice can work well with this particular dataset, which is why I would like to use this dataset with GATs to try and solve the problem.

References

- [1] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua Bengio. Graph attention networks. *arXiv preprint arXiv:1710.10903*, 2017.
- [2] Mark Weber, Giacomo Domeniconi, Jie Chen, Daniel Karl I Weidele, Claudio Bellei, Tom Robinson, and Charles E Leiserson. Anti-money laundering in bitcoin: Experimenting with graph convolutional networks for financial forensics. *arXiv preprint arXiv:1908.02591*, 2019.

¹<https://www.kaggle.com/datasets/ellipticco/elliptic-data-set>