Graph-Based Deep Learning for Fraud Detection in ETH Transaction Networks

Stephen Gelinas · sgelinas@ucsd.edu · Kazuma Yamamoto · kayamamo@ucsd.edu · Ethan Zhou · ezhou@ucsd.edu

Background & Research Question

- According to the FTC, cryptocurrency scams have cost online users over $1 Billion since 2021.
- With access to Ethereum transaction networks, we can model and train phishing detection as a node classification problem.
- How do non-graph supervised learning algorithms compare to graph-based deep learning approaches for fraud detection?

Why Graph?

- Graph algorithms can numerically represent information that is inherent to a network.
- Graph neural networks take advantage of learning the structural information within a graph and embedding information about neighboring nodes in the network.
- This allows for graph models to heavily outperform traditional learning algorithms.

Data Source

The XBlock dataset containing transactions of 890 Ethereum accounts

1. Collect subgraphs by K-order sampling with K-in = 1, K-out = 3 for each of the 890 objective nodes
2. Splice into a large-scale network with 86,623 nodes and 106,083 edges

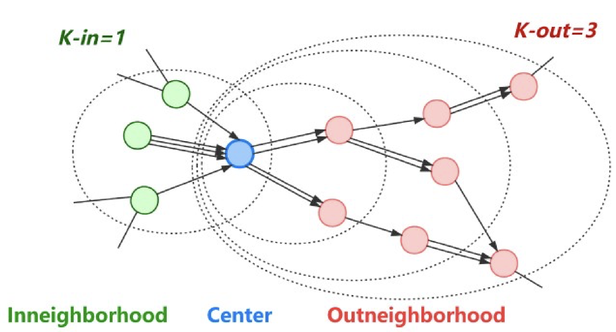

Figure 1: Schematic Illustration of a Directed K-Order Subgraph for Node Classification

On the figure above, based on the assumption that a typical money transfer flow is centered on a phishing node, the previous node of the phishing node may be a victim, and the next one to three nodes may be the bridge nodes with money laundering behaviors.

Recent Advancements in the Field

- The field of graph data science is relatively new and very active, and key advancements happen very frequently
- For example, all of the models we used were developed in the last several years. GCN and N2V are from 2016, GAT, GraphSage, and TA-GCN are from 2017.

Figure 2: An example of how our graph schema looks, at the basic level

Model	Avg Testing Accuracy	Type
TAGCN	82.2	GNN
SAGE	81.9	GNN
XGB	81.6	Non-Graph
GCN	79.6	GNN
GAT	78.5	GNN
N2V	76.6	GNN
kNN	74.6	Non-Graph
SVM	60.5	Non-Graph

Figure 3: Performance of each model

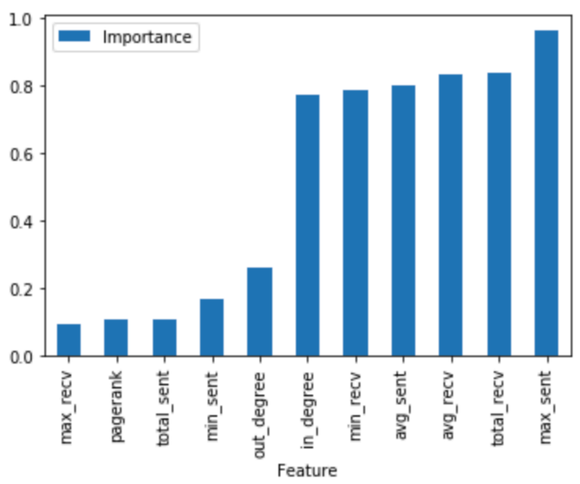

Figure 4: The relative importance of each feature to our models

Methods Overview

- The transaction network is represented as a directed graph, where each node represents a wallet in the network, and each directed edge between wallets represents the transfer of currency.
- Features are assigned to each node, including in-degree, out-degree, total transaction value, and more.
- The fraud detection task is ran as a node classification problem, and the performance of the models will be evaluated.

Conclusion and Summary of Findings

- Graph-based features improves overall model performance for both graph-based and non-graph-based models.
- Graph neural networks, specifically TA-GCN, performed best in the fraud detection task, as GNNs are able to learn the networks' structural information.
- The most important features for predicting fraudulent wallets are pagerank and the maximum amount of ETH sent between wallets.

References

1. Alin Deutsch, Yu Xu, Mingxi Wu, Victor Lee: "TigerGraph: A Native MPP Graph Database" , 2019; arXiv:1901.08248
2. Jiajing Wu, Qi Yuan, Dan Lin, Wei You, Weili Chen, Chuan Chen, Zibin Zheng: "Who Are the Phishers? Phishing Scam Detection on Ethereum via Network Embedding" , 2019, TSMC.2020.3016821; arXiv:1911.09259
3. Jiajing Wu, Dan Lin, Zibin Zheng, Qi Yuan: "T-EDGE: Temporal Weighted MultiDiGraph Embedding for Ethereum Transaction Network Analysis" , 2019, Front. Phys. 8:204 (2020); arXiv:1905.08038
4. Panpan Li, Yunyi Xie, Xinyao Xu, Jiajun Zhou, Qi Xuan: "Phishing Fraud Detection on Ethereum using Graph Neural Network" , 2022; arXiv:2204.08194
5. Jian Du, Shanghang Zhang, Guanhang Wu, Jose M. F. Moura, Soummya Kar: "Topology Adaptive Graph Convolutional Networks" , 2017; arXiv:1710.10370
6. Mark Cheung, John Shi, Lavender Yao Jiang, Oren Wright, José M. F. Moura: "Pooling in Graph Convolutional Neural Networks" , 2020; arXiv:2004.03519