Abstract

Effective illicit behaviour detection on Bitcoin transation networks

Crypto transaction networks are decentralized systems that allow the transfer of cryptocurrencies between users. These networks operate on a peer-to-peer (P2P) basis, where transactions are validated by a network of computers known as nodes. Each node maintains a copy of the blockchain, which is a public ledger that records all transactions in a secure and tamper-proof manner.

Fraud detection on crypto transaction networks is essential to prevent illicit activities such as money laundering, terrorist financing, and drug trafficking. The anonymity and decentralization of the cryptocurrency system provide an opportunity for fraudsters to exploit the system. Therefore, it is crucial to have fraud detection mechanisms in place to prevent these illegal activities.

The main goal of our thesis is to give an effective and applicable way to detect transactions with money laundering intent since it is a topic related to several malicious acts affecting human lives and causing financial damages to a high extent. For this, we use a famous and often analyzed dataset provided by Elliptic which contains information about Bitcoin transactions in graph structure on which we try to detect illicit nodes.

For this node classification problem we apply 2 different types of machine learning algorithms. One of the branches consists of classical methods such as gradient boosting trees and their numerous variants. The other methodology is neural network based which is more suitable for capturing the potential of the graph structure.

Our results are in line with preceding studies in the sense that booster algorithms still proved to be more effective for this kind of exercise than neural networks but we manage to improve performance on both fronts, due to:

- For XGBoost, CatBoost and LightGBM the better indicators can be explained by the difference in train-test splitting. Since the data contains independent networks corresponding to different timestamps, in our case, we break the habit of using time-based splitting and do randomized instead.
- For the neural network side we apply a model dedicated to discovering camouflaged fraudsters, it is called CARE-GNN. For this purpose, we have to analyze different clustering methods and come up with an edge coloring scheme. It results in a neural network workflow that outperforms previous (neural network based) experiments.