Graph embeddings for blockchain-induced networks

Máté Szőke
Budapest, Hungary 2023.

Abstract

The emergence of blockchain technology in recent years has generated massive interest and transformed various industries. Ethereum, one of the numerous blockchain platforms has garnered significant attention for its decentralized applications and smart contracts. However, comprehending Ethereum and its network interactions is a challenging task. Fortunately, with the utilization of various methods such as graph embeddings, valuable insights into its workings can be gained. Graph embeddings are powerful techniques to capture complex structures of graphs.

This Bachelor's thesis explores the intersection between graph embeddings and blockchain-induced Ethereum networks, with a specific focus on addressing a binary classification problem on Ethereum account transactions represented as graphs. Furthermore, the thesis investigates graph-level properties and the employment of dimensionality reduction methods to visualize these graphs within an embedding space.

A notable aspect of this thesis is the independent data collection process from the Ethereum blockchain by using Etherscan to retrieve the two-transaction radius around accounts banned by the USDC or USDT smart contracts. The primary objective of the binary graph classification task is to distinguish the different types of banned accounts with the help of the *Graph2Vec* graph embedding algorithm, followed by the application of various downstream classifiers.

Through a series of experiments, the performance of five distinct classification models is evaluated with diverse metrics. The effectiveness of *Graph2Vec*'s data-driven representation learning method for entire graphs is showcased, alongside the separability of the two graph classes (USDC or USDT banned) within the embedding space. All classifiers achieved good results in general, while a fine-tuned *Support Vector Classifier (SVC)* performed exceptionally well in my measurements.

Furthermore, the separability of the two graph classes within the *Graph2Vec* embedding space using Principal Component Analysis (PCA) and the global graph properties captured by *Graph2Vec* algorithm along the principal components are visually demonstrated.

In conclusion, this thesis demonstrates the potential of graph embeddings and PCA for analyzing and visualizing blockchain-induced networks, performed on my own data from the Ethereum blockchain.