traditional network tools. Recent works have analyzed the Ethereum network [3] and ERC20 token networks [19, 20] in terms of degree distribution, power laws and clustering. However, beyond these straightforward applications of traditional network analysis, there is not yet a research work on Ethereum that combines graph analysis with price prediction or anomaly detection.

Price prediction. Analyzing transactions and addresses to track the Bitcoin economy has become an important research direction [22]. An early time series prediction attempt by McNally uses a Bayesian optimised recurrent neural network (RNN) and Long Short Term Memory (LSTM) network with varying degrees of success [13]. Research works have also used blockchain features, such as average transaction amount and number of transactions to predict crypto-currency price with mixed results [8]. Blockchain graphs [2] can also provide network information that can be used as learned features. In [1], Akcora et al. use counts of blockchain subgraphs, termed chainlets, as features to predict Bitcoin price and compare their predictions to a baseline model that uses past price history only. All of the mentioned prediction works are carried out for crypto-currencies, and they track a single coin. In contrast, we track multiple cryptoassets at the same time. Furthermore, to the best of our knowledge, this is the first work that predicts the price of a token traded on a Blockchain platform.

Anomaly detection. Blockchain addresses can be linked and clustered to identify people behind suspicious transaction patterns in crypto-currencies [21]. The pattern is usually defined as a repeating shape that involves moving coins from a (black) address to an online exchange address, where the coins can be cashed out without being confiscated by authorities. The black address that starts the transaction chain may be related to money laundering [15] and ransomware payment [10] efforts. Researchers have found ample evidence in the transactions network for these anomalies [4, 12]. Differing from these work, we do not assume prior knowledge about pattern shapes. Our unsupervised depth approach detects anomalies by tracking individual token behavior. In these aspects, this work is the first attempt at detecting anomalies in a blockchain platform.

3 Background on Ethereum and CryptoTokens

The Ethereum project [23] was created in July 2015 to provide Smart Contract functionality on a blockchain. Smart Contracts are self-executing Turing complete contracts written in software code which is replicated across a distributed, decentralized blockchain network. Smart contracts are created and put to a blockchain address by its developers. Smart Contracts ensure unstoppable, deterministic code execution that can be verified publicly. Some smart contracts implement mechanisms that allow trading or sharing digital assets, known as crypto-tokens, on the blockchain. We will refer to such a smart contract as a **crypto-token** contract, and use the term token interchangeably. A token is traded among blockchain nodes, and can have an associated dollar value per each token unit. This dollar value is listed in online exchange websites such as CoinMarketCap.com. The value is arbitrated by token demand and supply. Most tokens have a fixed supply that is set at the time when a token contract is created. As the supply is fixed, the value of a token is mostly determined by its demand.

The utility of a token is a hotly debated topic; although some tokens are used to buy services in real life, most tokens exist as online entities that have no intrinsic value. For example, the Cryptokitties token is used to purchase, collect, breed and sell various types of virtual cats.

As most blockchains, the Ethereum blockchain attaches an *input data* field to each transaction to store log data. Smart contract transactions use this input data field to transmit messages; creating a transaction with input data to a Smart Contract is analogous to passing variables to a function. The earliest versions of Smart Contracts were developed without a common standard for transactions, input messages and functions contained in the contract. As such, each time a user wanted to transmit a message to the smart contract, she needed to know which message structure to use. Recently, the community has proposed standards, such as ERC20[1] (Ethereum Request for Comments 20) and ERC223, which define a common list of rules for tokens to follow within the larger Ethereum ecosystem.

An example in these standards is the name of commonly used functions. By the ERC20 standard, each token must implement a number of standard functions, such as *totalSupply()*. This naming standard allows exchanges, users and developers to create transactions for tokens in an automated manner. For example, when address A_1 wants to transfer its tokens (e.g., 20 OmiseGo

[1] `https://theethereum.wiki/w/index.php/ERC20_Token_Standard`

Table 1: Symbols table.

Symbol	Explanation
e	edge
v	vertex
t	token
h	horizon
$\Delta_{\mathcal{G}}$	total token supply
$\mathcal{T}(e)$	transferred token value
δ	min price change for anomaly
ρ_t	prediction accuracy for token t

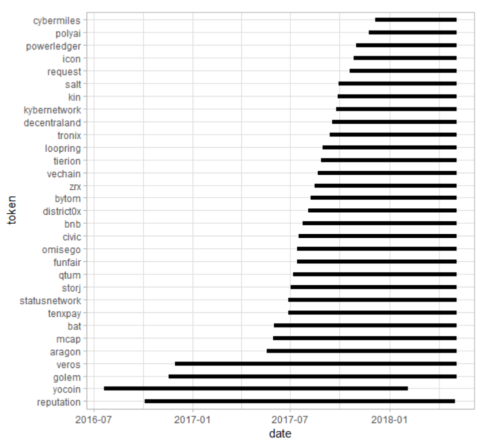

Figure 2: Token start dates on the Ethereum blockchain.

tokens) to address A_2, it creates a transaction where the "to" address is the OmiseGo address, and the input data contains the message "transfer(A_2, x=20)". Once the token discovers that a mined block contains this transaction, it updates the balance of both addresses.

Pre-processing. We created our Ethereum dataset by installing the official Ethereum Wallet [2] and downloading all blocks. Afterwards, we used the EthR [9] library to query Ethereum blocks through the Go Ethereum Client (i.e., Geth). Our dataset contains all the Ethereum data from 2015 July to 2018 May 6th, with a total of 5.5 million blocks.

By parsing the data, we discovered 1.7K ERC20 tokens who had more than 10K transactions. Furthermore, we selected 31 of the most valuable ERC20 tokens by market value, as reported by the EtherScan [3] online explorer.

In average, each token has a history of 297 days, with minimum and maximum of 151 and 576 days, respectively. The first dates of tokens on the Ethereum blockchain is given in Figure 2. Among the ERC20 tokens, we focus our attention on transfer and approve+transferFrom functions which transfer tokens between two addresses. Each time these functions are employed in a transaction, we create an edge between the two involved addresses. Formally, we define a token network as a directed multigraph $\mathcal{G} = (V, E)$ of order V and size E with a function f such that

$$f : e \rightarrow \{\{u, v\} : u, v \in V, u \neq v \text{ and } 0 < \mathcal{T}(e) \leq \Delta_{\mathcal{G}}\},$$

where $\mathcal{T}(e)$ denotes the transferred token value and $\Delta_{\mathcal{G}}$ is the total supply of the token that is set at the token contract. Typically, total supply values of tokens range between 10^6 to 10^{13}.

[2] https://www.ethereum.org/
[3] EtherScan.io