ADS Fall 2018 – Project Proposal

# Team Info

**Team Name:** Cryptonians

**Team Members:**

* David Robinson
* Martin Hu
* Scott Nicholson
* W. Mike Herrington

# Need for the Project

The cryptocurrency market is growing in market capitalization and becoming more ingrained in investment strategies supported by even traditional investment institutions, but as seen in the incredible growth and subsequent tumble of valuations from 2017 into early 2018, it is still an immature and volatile market. [A University of Texas researcher](https://www.cnbc.com/2018/06/13/much-of-bitcoins-2017-boom-was-market-manipulation-researcher-says.html) has suggested that 50% of the gain in Bitcoin value during 2017 is accounted for during 1% of all trading hours, during which transactions were timed to inflate Bitcoin’s value, which market manipulators capitalized on by transferring the Bitcoins for another cryptocurrency to lock in their gains. This manipulation poses financial risk to other investors by distorting the true value of their assets.

# Hypotheses

**Identifying Manipulation Events**

By focusing on periods of elevated trading activity, particularly those that are followed by unusually large changes in cryptocurrency valuations (as measured by their price in USD or other traditional currencies), we can identify instances of market manipulation. We also suspect that attempts at layering that are carried out by placing large volume bid orders at prices below the market exchange rate (designed to entice buyers without actually closing the transaction) may be identified by observing aggregate order data relative to trade volumes and prices in the preceding trading days. We are open to comparing current trading activity to prior, confirmed cryptocurrency manipulation events that were registered by investigators.

**Clustering Bitcoin Address Nodes to Identify Manipulators**

By investigating the Bitcoin addresses active during the elevated levels of trading that are identified in part one, above, we intend to find clusters of Bitcoin addresses engaged in coordinated action that represent participants in the manipulation.

# Technical Approaches to be Taken

Unsupervised or semi-supervised learning algorithms such as hierarchical clustering and unsupervised decision trees will be used to identify market manipulation events because our assumption is that most market manipulation incidents are currently unidentified. To the extent that we can label some market manipulation events, synthetic minority oversampling may also be valuable. This may also involve a time-series anomaly detection component: first attempting to predict future market behavior, after accounting for trends and seasonality in market fluctuations, then using methods such as Statistical Process Control (SPC) to identify statistically significant diversions from our predicted behavior.

Clustering algorithms such as KNN will be used to relate Bitcoin addresses that are active during market manipulation periods, as identified through the efforts described above.

# Data

The CryptoCompare API will be used to obtain time-series cryptocurrency exchange rate and transaction volume data. Individual transaction details will be obtained using the BlockCypher API. Both of these Python APIs pull data in JSON format.

The primary focus of this project will be Bitcoin, though we intend to attempt to apply our methodology to other cryptocurrencies such as Ethereum, Ripple, Litecoin, Dogecoin and DASH after demonstrating some level of success with Bitcoin analysis.

# Project Execution Risks and Mitigation Strategies

Overfitting the unsupervised learning algorithms could lead to mis-identifying market manipulation events or Bitcoin address clusters based on irrelevant criteria. Establishing a solid theoretical framework that can explain how factors and patterns in our data could relate to market manipulation can guard against the acceptance of a high-performing but misguided algorithm. Performing cross-validation of results will also be valuable.

# Goals

We believe that by identifying the manipulation events and participating Bitcoin address clusters we can learn more about the trading patterns that are implemented to manipulate cryptocurrency markets, identify participants in this fraudulent behavior and ultimately counter manipulation efforts. Identification of participants suspected of engaging in market manipulation could allow for further scrutiny by regulators and law enforcement. This scrutiny could possibly result in penalties that would have a deterrent effect on manipulative activity, thereby calming market volatility.

Given that research has attributed as much as $40 billion of Bitcoin’s 2017 market capitalization growth to market manipulation, even a modest real-time detection capability of 5% of events and/or participating Bitcoin address clusters could have tremendous value. This detection capability could be used to inform trading strategies, allowing customers to identify when cryptocurrency values are artificially inflated or deflated and avoid or participate in the market appropriately.

# Deliverables

We will deliver a summary of findings in a markdown file that includes visualizations that represent important trends in the underlying data. We will deliver D3.js visualizations mapping an active Bitcoin address cluster during a manipulation event.