# Title Page

The paper that I will be presenting today is “Inferring the Interplay Between Network Structure and Market Effects in Bitcoin”. This paper was published in the New Journal of Physics in 2014 and it was wrote by Daniel Kondor, Istvan Csabai, Janos Szule, Marton Posfai, and Gabor Vattay.

# Presentation Overview

# 1.0 Paper Overview

The main research question this paper attempts to solve is “aim to identify relevant changes in network structure over time and to uncover the relation of network structure and macroeconomic indicators of the system”. Their main data source is Bitcoin’s public transaction network. They used principal component analysis to identify important features in the variation of the network over time.

# 2.0 Bitcoins

Bitcoin is a form of cryptocurrency that is created and held electronically. It is decentralized as no one controls it. Every machine that mines Bitcoin make up the network so no one person has power over the network. This currency isn’t printed and is produced by people and computers around the world solving mathematical problems. Bitcoin is a finite currency. There are only a finite number of Bitcoins available, 21 million to be exact, so people can’t produce more Bitcoins to de-value the currency. Bitcoins are produced via “mining” using computational power in a distributed network. “Miners” are rewarded for the “mining” blocks. The reward however halves every 210,000 blocks. About 77% of the total blocks have been mined and the current reward is 12 Bitcoins per block. This network of “miners” also process transactions made with Bitcoin, making it a standalone payment network. Bitcoin stores details of every single transaction that ever happened on the network in a general ledger in the form of a blockchain. Everyone knows how much bitcoins are stored at a given address. Users can use multiple addresses to protect their identity.

# 3.0 Methodology

This is a general overview of the main methodology that was employed in this paper. To do their analysis, they first extracted the core from the Bitcoin network over two years. Then they constructed daily snapshots of transactions for the core network. Next, they concatenated all the snapshots into a matrix where each row corresponded to a day’s transactions and each column corresponding to a transaction. Finally, for analysis, they used Principal Component Analysis to identify the key features in the evolution of the network over time. Finally, they analyzed the base-network decay, the time-varying contribution of each base network, and the correlation between the network and the exchange rate.

# 4.0 Data Mining

The data for this paper was acquired from a slightly modified version of the open-source bitcoind client. They collected the entire list of transactions for 2012 and 2013 on March 3, 2014. The schema of their data was sending address, the receiving address, the value sent, and the time of transaction.

# 5.0 Extracting the Core Network

Their goal here was the extract the subgraph of the most active users. They did this by first contracting the graph. They did this by first identifying all the transactions with multiple inputs and then they assumed that if multiple addresses were involved in the same transaction, they belonged to the same user. They justified this assumption because users needed an account’s private key to use an account in a transaction. They, then identified 2 active cores within this subgraph. The first active core was defined as the long-term core which consisted of users active in greater than individual transactions, and active greater that 60 consecutive days. The second active core was defined as the all-users core which consisted of the 2000 most active users. The users that were associated with the Satoshi gambling site were excluded because they were considered statistical outliers and not related to the normal operation of the Bitcoin network.

# 6.0 Detecting Structural Changes

The main goal here was to extract important structural changes from the graph by comparing successive daily snapshots using Principal Component Analysis. Here, the daily snapshot is defined a weighted network where the weight of each link is equal to the number of transactions that occurred between node u and v. Finally, the used Principal Component Analysis to determine the significant basis vectors on a matrix of concatenated daily snapshots.

## 6.1 Daily Snapshots: Weighted Adjacency Matrix

For each day, a weighted adjacency matrix was constructed with all the nodes as columns and rows. The weight was equivalent to the number of transactions between the nodes. For example, user 1 had 100 transactions with user 2, 200 transactions with user 3, and 50 transactions with user 4. This weighted adjacency matrix was then rearranged into a long vector and then all the snapshot vectors were concatenated into a matrix.

## 6.2 Principal Component Analysis

They then analyze the constructed matrix using Principal Component Analysis. Principal Component Analysis is a dimensionality reduction technique that is currently implemented in Spark MLlib. The purpose of this technique is to extract the most important information from the given data set. This is done by finding new features that can accurately explain the variance in the data set. By doing so, we can compress the size of the data set by keeping only the important features. The first basis vector will explain as much variance as possible and each subsequent basis vector will explain as much of the variance left as possible. It explains the variance by creating new features that are linear combinations of the existing features. For example, in this data set, the basis vector will be along the center of these data points since variance is maximized in this direction. Here the new basis vector is linear combination of both x and y.

#The direction of the first principal component is given by the first eigenvector of the covariance matrix.

# 7.0 Key Results

The key result from their paper was they found a correspondence between the exchange rate time series and the LT core. To calculate the price, they first subtracted the average value of the price time series, and estimated this as a linear combination of singular vectors. They could approximate the exchange rate using both the first four base vectors ranked by variance explained, and the first four base vectors ranked by correlation to the price.

# 8.0 Relation to MIE1512H Research Project

The Ethereum network is like the Bitcoin network and has the same information in terms of transactions. The same research question can be answered but in terms of the Ethereum network. Daily snapshot vectors can be constructed and concatenated into a matrix. Then PCA can be performed to identify the key features of the network. Then the same method could be used to see if the network structure can be related the Ethereum exchange rate.

# 9.0 References

[1] Kondor, Daniel, et al. “Inferring the Interplay between Network Structure and Market Effects in Bitcoin.” New Journal of Physics, vol. 16, no. 12, Feb. 2014, p. 125003., doi:10.1088/1367-2630/16/12/125003

[2] <http://www.bitcoinblockhalf.com/>

[3] <https://bitcoin.org/img/icons/opengraph.png>

[4] <http://www.coindesk.com/information/what-is-bitcoin/>

[5] http://bitcoin.stackexchange.com/questions/tagged/bitcoind

[6] http://cpsc.ualr.edu/srini/DM/chapters/review5.3.html

[7] <https://pdfs.semanticscholar.org/53b9/966a0333c9c9198cdf03efc073e991647c12.pdf>

[8] http://stats.stackexchange.com/questions/2691/making-sense-of-principal-component-analysis-eigenvectors-eigenvalues

[9] https://www.ethereum.org/images/logos/ETHEREUM-LOGO\_LANDSCAPE\_Black.png