## Presentation Overview

* Now here’s an overview for my presentation
* First I will give you an overview of the paper and some background on Bitcoin in general
* Then I will provide a quick overview of the methodology employed in this paper
* Then I will explain where they got their data and how they preprocessed and analyzed their data
* Finally, I will show their key results and link this back to my project for this course

## Paper Overview

* This paper aims to identify relevant changes in the network structure over time and to uncover the relationship between the network structure and macroeconomic indicators of the system
* Their data source is Bitcoin’s transaction network and..
* Use Principal Component Analysis to identify important features in the variation of the network over time

## Bitcoin

* Now let me provide you with some background on Bitcoins
* Bitcoin is a digital currency that is created and held electronically
* It is decentralized as no one entity controls it.
  + Instead every machine that mines Bitcoin make up its network
* It is a finite currency so…
  + People can’t produce more Bitcoin to de-value the currency
  + Miners get rewarded Bitcoins for mining blocks by solving computational mathematical problems. But…
  + This reward halves over time
* One of the main characteristics of this network is that its transparent
  + All transactions that have ever happened are stored in public ledger in the form of a blockchain
  + Users are anonymized on this ledger via addresses
  + Users can have multiple addresses to hide their identity

## Methodology

* Here is an overview of the methodology employed in this paper
* To do their analysis, they first extracted the core network from the bitcoin transaction network
* They then constructed daily snapshots of transactions for the core network
* Next they concatenated all the snapshots into a matrix where each row corresponded to a daily snapshot and each column corresponded to a transaction
* Then, for analysis, they used Principal Component Analysis to identify 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 base network and the exchange rate
* I will now provide more details about each of these steps

## Data Mining

* The first step was data mining
* They got their data from an open source client known as Bitcoind or Bitcoin daemon which is a command line interface tool for Bitcion
* They collected all the transactions for the years 2012 and 2013 in 2014
* The schema of the data they collected include
  + The sending address
  + The receiving address
  + The value sent
  + The time of the transaction

## Extract the Core Network

* The second step was extracting the core network from the graph
* They first contracted the graph
  + They used a simple heuristic to identify address that belong to the same user
  + They did this by first identifying all the transactions with multiple inputs and assuming if multiple addresses were involved in the same transaction, they belong to the same user
  + They justified this assumption because user’s need a private key associated with an address in order to use it in a transaction
* Secondly, the identified 2 active cores within their subgraph
  + The first core was called the long term core
    - It consisted of users that were active for greater than 60 consecutive days and involved in greater than 100 individual transactions
  + The second core was called the all users core
    - It consisted of the 2000 most active users in the network
* They excluded any user associated with the Satoshi gambling site because..
  + They were considered statistical outliers, and not related to the normal operation of the Bitcoin network

## Detecting Structural Changes

* The third step was detecting structural changes
* The main goal here was to compare successive daily snapshots of the active core with PCA to extract important changes in the graph structure
* Here, the daily snapshots were weighted adjacency networks where the weight of each link was equal to the number of transactions between users u and v
* PCA was used to determine the significant basis vectors of the network

## Daily Snapshot

* Now I will further elaborate on the daily snapshots
* Each day will have a weighted adjacency graph constructed for it where the nodes will be users and the links will be transactions
* The users or nodes will be listed along the columns and rows
* The weight is equivalent to the number of transactions between the nodes
* For example, user 1 had 100 transactions with 2, 200 w/ 3, and 50 w/ 4.
* The constructed matrix is then rearranged into a long vector and then all the snapshot vectors were concatenated into a matrix with days as rows and transactions as columns

## PCA

* Then they analyzed the constructed matrix using Principal Component Analysis
* PCA 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 are a linear combination of existing features that can accurately explain the variance in the dataset
* The first basis vector will explain as much variance as possible, and each subsequent matrix will explain as much variance left as possible
* For example, in this dataset, the first basis vector will be along the center of this data pints since variance is maximized in this direction
* Here the new basis vector is a linear combination of both x and y

## Key Results

* Here are the key results from this paper
* From PCA, they found the time varying contribution of each base network at time t
* They plotted the first 6 networks and saw a high resemblance between the first base network of the LT core and the Bitcoin exchange rate from Mt.Gov
* The then normalized the Bitcoin exchange rate by removing the mean and estimated it as a linear combination of singular vectors
* They then improved the accuracy by using the top four base networks ranked by covariance and correlation to the price as seen here

## Relation to MIE1512H Research Project

* Ethereum network is very much like the Bitcoin network
* It is also based on a blockchain with public 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
* The same method can be used to see if the network can be related to the Ethereum exchange rate