# Understanding Latent Group Structure of Cryptocurrencies: Covariate-assisted Spectral Clustering for Dynamic Network

Li GUO & Yubo TAO

Singapore Management Univeristy

Crypto-Currencies in a Digital Economy

November, 2017

# Outline

# Agenda

# Motivation

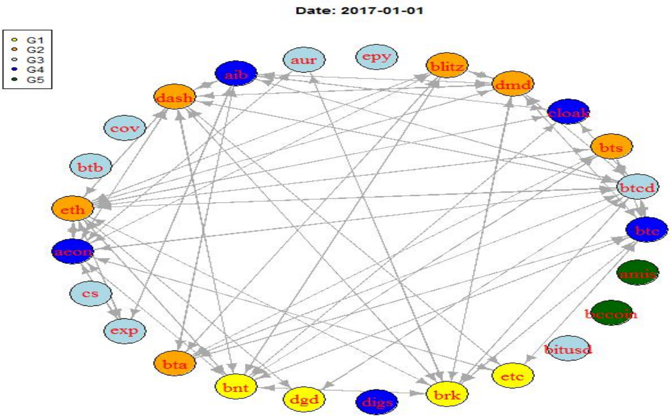
- **Background:**
  - Emerging of Altcoin Market:
    - 787 cryptocurrencies; US$102.6 Billion circulating supply and US$1.9 Trillian at the end of June 2017;
    - Network Effect (attention from central banks and institutional investors)
  - Investors Behavior Bias (return reversal)
  - Innovation of Blockchain Technology (new algorithm and proof types)
- **RQ: How fundamental information and return structure jointly determine a market segmentation?**

# Agenda

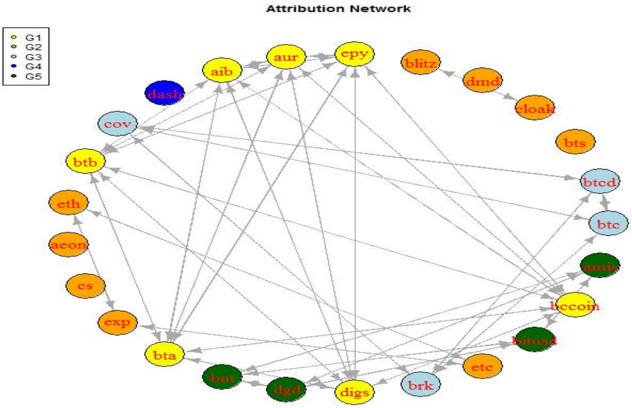
# Data

- Data Source: Cryptocompare
- Sample Period:
  - In-sample Estimation: from 2015-Jun-01 to 2016-Dec-31.
  - Out-of-Sample Tests: from 2017-01-01 to 2017-06-30.
- Cryptocurrency Daily Return:
  - 191 Cryptos Sorted on Market Cap, Age, Maximum Price and Doller Volume;
- Contract Information:
  - Algorithm
  - Proof Types
  - Premined Indicator
  - Total Coin Supply

# Return Network Structure from Adaptive LASSO



# Visualization: Node Features (Attribution Network Structure)




Attribution

# Agenda

# Modelling Undirected Graphs

- **Dynamic Stochastic Blockmodel**:

$$A_t(i,j) = \begin{cases} \text{Bernoulli}(P_t(i,j)), & \text{if } i < j \\ 0, & \text{if } i = j \\ A_t(j,i), & \text{if } i > j \end{cases} \quad (1)$$

$$\mathcal{A}_t := \mathbb{E}(A_t | Z_t) = Z_t B_t Z_t^\top, \quad (2)$$

where $Z_t \in \{0,1\}^{N \times K}$ is the *clustering matrix* such that there is only one 1 in each row and at least one 1 in each column.

- **Node Covariates**:
  - **Regularized Graph Laplacian**:

$$L_{\tau,t} = D_{\tau,t}^{-1/2} A_t D_{\tau,t}^{-1/2}, \quad (3)$$

  where $D_{\tau,t} = D_t + \tau_t I$ and $D$ is a diagonal matrix with $D_t(i,i) = \sum_{j=1}^{N} A_t(i,j)$, and $\tau_t = N^{-1} \sum_{i=1}^{N} D_t(i,i)$.

  - **Similarity Matrices** (Covariate-assisted Graph Laplacian):

$$S_t = L_{\tau,t} + \alpha_t X X^\top. \quad (4)$$

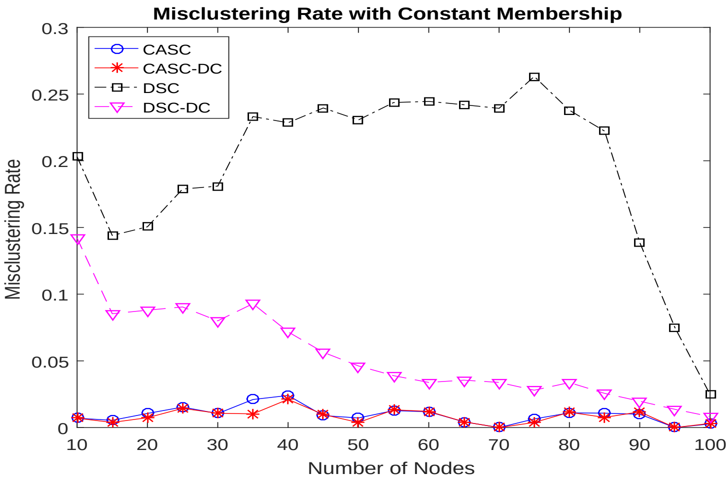
  where $\alpha_t \in [0, \infty)$ is a tuning parameter

# Agenda

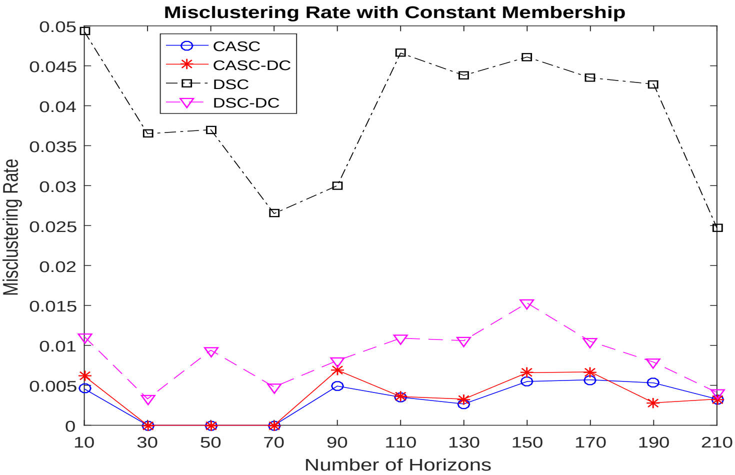
# Simulation Settings

- Misclustering Rate with Number of Nodes:
  - Block Probability:
    $$B_t = \frac{3}{N}\begin{bmatrix} \sqrt{N} & 0.1\log(N)^{t/T} & 0.1\log(N)^{t/T} \\ 0.1\log(N)^{t/T} & \log(N)^{t/T+1} & 0.1\log(N)^{t/T} \\ 0.1\log(N)^{t/T} & 0.1\log(N)^{t/T} & 0.80.1\log(N)^{t/T} \end{bmatrix};$$
  - $N = \{10, 15, \cdots, 100\}$;
  - $T = 10$, $s = N^{1/2}$, # of Replication: 100;
- Misclustering Rate with Number of Membership Changes:
  - Block Probability:
    $$B_t = \frac{3}{N}\begin{bmatrix} \sqrt{N} & 0.1\log(N)^{t/T} & 0.1\log(N)^{t/T} \\ 0.1\log(N)^{t/T} & \log(N)^{t/T+1} & 0.1\log(N)^{t/T} \\ 0.1\log(N)^{t/T} & 0.1\log(N)^{t/T} & 0.80.1\log(N)^{t/T} \end{bmatrix};$$
  - Maximum number of membership changes: $s = [0, 2, 4, 5, 10, 20, 25, 50, 100]$
  - $N = 100$, $T = 10$, # of Replication: 100;
- Misclustering Rate with Number of Horizons:
  - Block Probability:
    $$B_t = \frac{3}{N}\begin{bmatrix} \sqrt{N} & 0.1\log(N)^{t/T} & 0.1\log(N)^{t/T} \\ 0.1\log(N)^{t/T} & \log(N)^{t/T+1} & 0.1\log(N)^{t/T} \\ 0.1\log(N)^{t/T} & 0.1\log(N)^{t/T} & 0.80.1\log(N)^{t/T} \end{bmatrix};$$
  - $T = \{10, 30, 50, \cdots, 210\}$;
  - $N = 100$, $s = N^{1/2}$, # of Replication: 100;

# Performance with Growing Number of Vertices

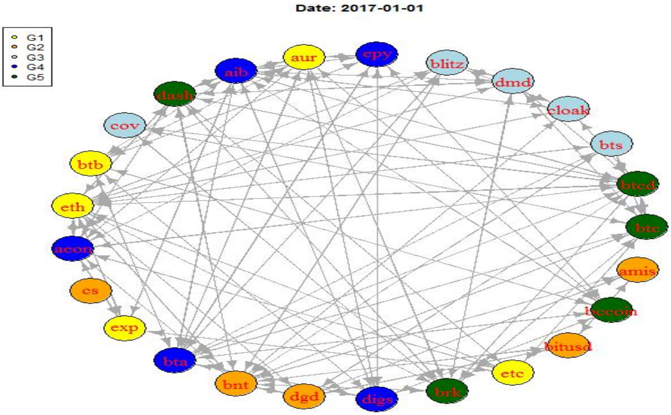


# Performance with Growing Sample Periods



# Agenda

# Visualization: Combined Network Structure



# Grouping Results

Table: Group Members

| | **Cryptocurrency** |
|---|---|
| Group 1 | **Ethereum**, Bitbar, Ethereum Classic, Expanse, Aurora coin... |
| Group 2 | Cryptospots, Amis, Bitusd, Digix Dao, Bancor Network Token... |
| Group 3 | Bitshares, Cloakcoin, Covencoin, Diamond, Blitzcoin... |
| Group 4 | Empyrean, Diggits, Advanced Internet Block, Aeoncoin, Bata... |
| Group 5 | **Bitcoin**, Breakout coin, Digital Cash, Bitcoin Dark, Bitconnect Coin... |

# DSBM Evaluation I

- *Within Group Connection*$_g$ $= \frac{\text{\# of Degrees within Group g}}{N_g}$
- *Cross Group Connection*$_g$ $= \frac{\text{\# of Degrees between Group g and other Groups}}{\bar{N}_g}$

Table: Evaluation Criteria: **Return Inferred Adjacency Matrix**

| Group ID | Within Group Connection | Cross Group Connection | Diff (W - C) |
|---|---|---|---|
| G1 | 0.073 | 0.066 | 0.007*** |
| G2 | 0.234 | 0.125 | 0.111*** |
| G3 | 0.041 | 0.064 | -0.02*** |
| G4 | 0.149 | 0.097 | 0.052*** |
| G5 | 0.015 | 0.015 | 0.000 |
| All | 0.103 | 0.073 | 0.030*** |

# DSBM Evaluation II

- *Within Group Connection*$_g$ $= \frac{\text{\# of Degrees within Group g}}{N_g}$
- *Cross Group Connection*$_g$ $= \frac{\text{\# of Degrees between Group g and other Groups}}{\bar{N}_g}$

Table: Evaluation Criteria: **Algorithm**

| Group ID | Within Group Connection | Cross Group Connection | Diff (W - C) |
|---|---|---|---|
| G1 | 0.131 | 0.155 | **-0.024** |
| G2 | 0.163 | 0.170 | **-0.006** |
| G3 | 0.179 | 0.175 | 0.004 |
| G4 | 0.161 | 0.170 | **-0.009** |
| G5 | 0.142 | 0.153 | **-0.011** |
| All | 0.155 | 0.165 | **-0.009** |

# DSBM Evaluation III

- *Within Group Connection*$_g$ $= \frac{\text{\# of Degrees within Group g}}{N_g}$
- *Cross Group Connection*$_g$ $= \frac{\text{\# of Degrees between Group g and other Groups}}{\bar{N}_g}$

Table: Evaluation Criteria: **Proof Types**

| Group ID | Within Group Connection | Cross Group Connection | Diff (W - C) |
|---|---|---|---|
| G1 | 0.273 | 0.300 | **-0.027** |
| G2 | 0.314 | 0.322 | **-0.008** |
| G3 | 0.303 | 0.310 | **-0.007** |
| G4 | 0.311 | 0.310 | 0.001 |
| G5 | 0.222 | 0.273 | **-0.050** |
| All | 0.284 | 0.303 | **-0.018** |

# Covariate-assisted Spectral Clustering Evaluation I

- *Within Group Connection*$_g$ $= \frac{\text{\# of Degrees within Group g}}{N_g}$
- *Cross Group Connection*$_g$ $= \frac{\text{\# of Degrees between Group g and other Groups}}{\bar{N}_g}$

Table: Evaluation Criteria: **Return Inferred Adjacency Matrix**

| Group ID | Within Group Connection | Cross Group Connection | Diff (W - C) |
|---|---|---|---|
| G1 | 0.064 | 0.074 | -0.010*** |
| G2 | 0.078 | 0.078 | 0.001 |
| G3 | 0.066 | 0.076 | -0.010*** |
| G4 | 0.111 | 0.091 | 0.020*** |
| G5 | 0.098 | 0.087 | 0.012*** |
| **All** | 0.083 | 0.081 | **0.002*** ** |

# Covariate-assisted Spectral Clustering Evaluation II

- *Within Group Connection*$_g$ $= \frac{\text{\# of Degrees within Group g}}{N_g}$
- *Cross Group Connection*$_g$ $= \frac{\text{\# of Degrees between Group g and other Groups}}{\bar{N}_g}$

Table: Evaluation Criteria: **Algorithm**

| Group ID | Within Group Connection | Cross Group Connection | Diff (W - C) |
|---|---|---|---|
| G1 | 0.227 | 0.164 | 0.062 |
| G2 | 0.622 | 0.039 | 0.583 |
| G3 | 0.162 | 0.122 | 0.040 |
| G4 | 0.522 | 0.176 | 0.347 |
| G5 | 0.183 | 0.140 | 0.043 |
| All | 0.343 | 0.128 | 0.215 |

# Covariate-assisted Spectral Clustering Evaluation III

- *Within Group Connection*$_g$ $= \frac{\text{\# of Degrees within Group g}}{N_g}$
- *Cross Group Connection*$_g$ $= \frac{\text{\# of Degrees between Group g and other Groups}}{\bar{N}_g}$

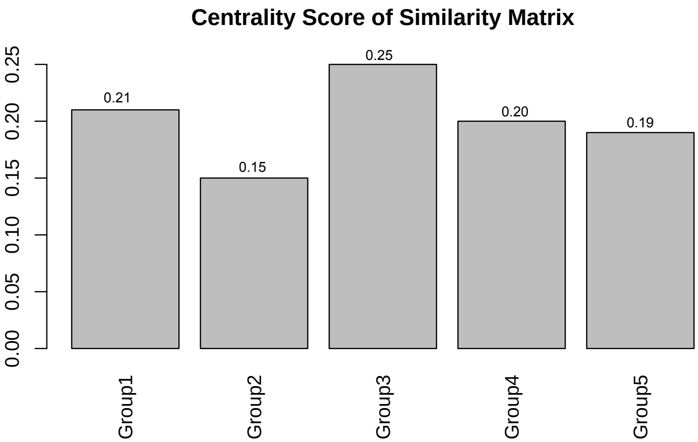
Table: Evaluation Criteria: **Proof Types**

| Group ID | Within Group Connection | Cross Group Connection | Diff (W - C) |
|---|---|---|---|
| G1 | 0.514 | 0.312 | 0.202 |
| G2 | 0.302 | 0.116 | 0.186 |
| G3 | 0.579 | 0.213 | 0.366 |
| G4 | 0.810 | 0.242 | 0.568 |
| G5 | 0.514 | 0.323 | 0.191 |
| All | 0.544 | 0.241 | 0.302 |

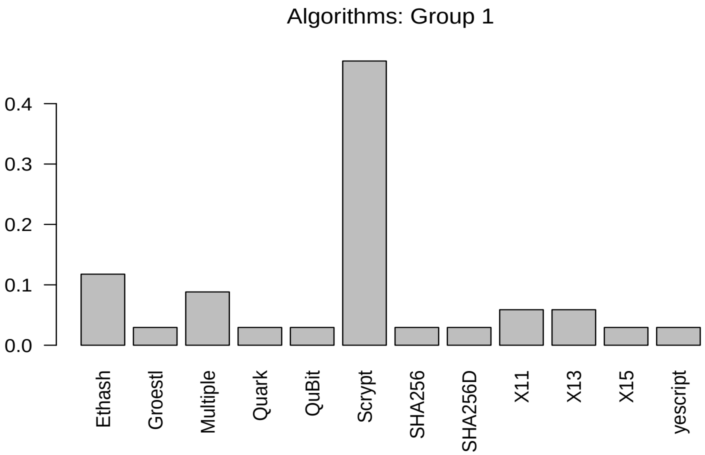
# Asset Pricing Inference: Hypothesis Development

- Rational Price: Information Diffusion
  - Within Group Return Predictability
  - Lead-Lag Effects across Groups
- Behavior Bias: Overreaction.
  - **Time Series Return Predictability**: *Stronger return predictability of lagged return for groups with low centrality than high centrality groups.*
  - **Cross Sectional Return Predictability**: *Contrarian strategy should be more profitable for those groups with lower centrality score.*

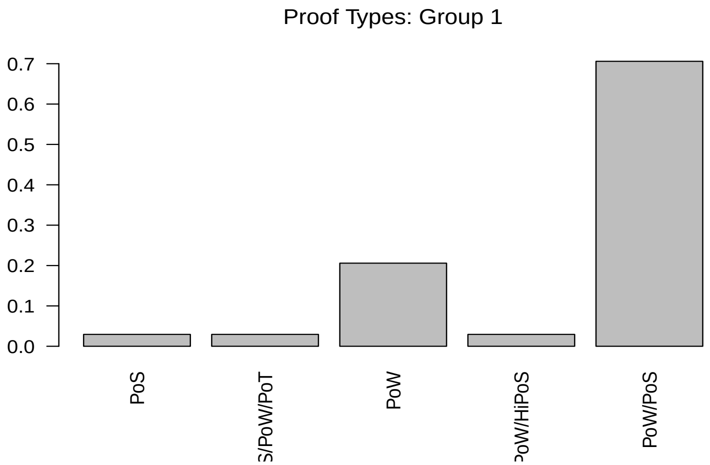
# Asset Pricing Inference: Group Centrality



# Fundamental Comparison under Different Centrality Score: Algorithm



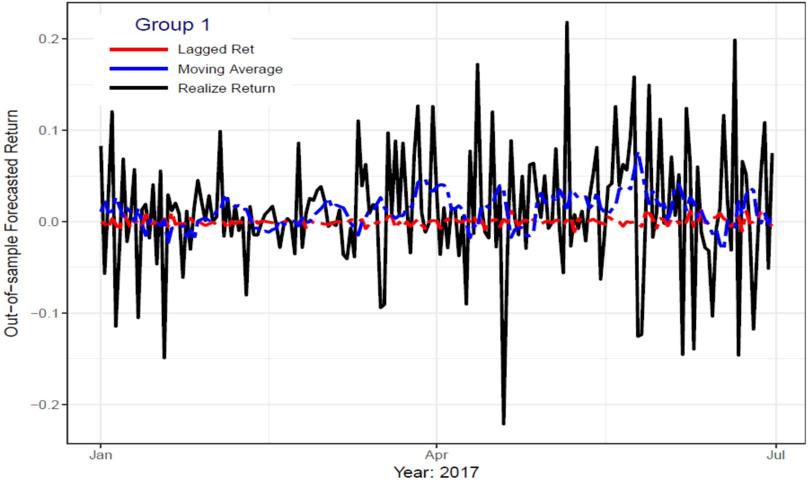
# Fundamental Comparison under Different Centrality Score: Proof Types



# Time Series Return Predictability: Daily

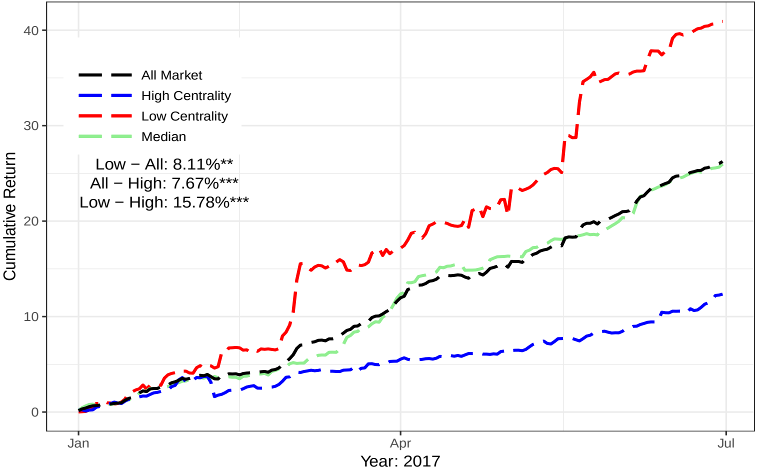
| | G1 | G2 | G3 | G4 | G5 | G2 - G3 |
|---|---|---|---|---|---|---|
| $Ret_{t-1}$ | -0.11*** | **-0.25*** ** | **-0.06** | -0.24*** | -0.20*** | **-0.19*** ** |
| | (-2.73) | (-6.65) | (-1.46) | (-6.32) | (-5.00) | |
| $Ret^2_{t-1}$ | -0.21 | 0.63*** | 0.44 | -0.10 | 0.17 | |
| | (-0.72) | (5.67) | (1.11) | (-0.49) | (0.88) | |
| Monday | -0.01 | 0.00 | 0.00 | 0.01 | -0.00 | |
| | (-0.62) | (0.32) | (0.21) | (1.12) | (-0.14) | |
| Tuesday | -0.01 | 0.00 | 0.01 | 0.01 | 0.01 | |
| | (-0.74) | (0.10) | (0.71) | (0.99) | (1.49) | |
| Wednesday | -0.00 | 0.01 | -0.00 | 0.01 | 0.01 | |
| | (-0.52) | (0.87) | (-0.53) | (0.50) | (0.80) | |
| Thursday | 0.01 | 0.00 | 0.00 | 0.02* | 0.01 | |
| | (0.99) | (0.17) | (0.47) | (1.90) | (1.54) | |
| Friday | -0.00 | -0.00 | 0.01 | 0.02* | -0.00 | |
| | (-0.48) | (-0.03) | (0.65) | (1.82) | (-0.03) | |
| Saturday | 0.01 | 0.01 | -0.00 | 0.02 | 0.00 | |
| | (0.66) | (0.67) | (-0.59) | (1.60) | (0.19) | |
| $R^2$ | 0.020 | 0.099 | 0.011 | 0.065 | 0.046 | |

# Time Series Return predictability Comparison I



# Time Series Return predictability Comparison II

# Cross Sectional Return predictability Comparison



# Agenda

# Conclusions

- **Covariate-assisted Spectral Clustering is useful to Understand Cryptocurrency Market Structure.**
  1. Attribution Matrix provides most valuable information to connect within group members.
  2. Return based adjacency matrix reveal connections across different groups.
  3. Within group connection is stronger than cross group connection.
  4. Behavior bias is stronger for those groups with low market centrality.
     a. Time Series Return Predictability based on lagged return shows high $\Delta CSFE$ for groups with low centrality.
     b. Contrarian strategy reveal higher investment value for the low centrality groups than the high centrality groups.