Covariate-assisted Spectral Clustering in Dynamic Networks: An Application to Cryptocurrencies Market

Li Guo[1] Yubo Tao[1] Wolfgang Härdle[1,2,3]

[1]Singapore Management University
[2]Humboldt-Universität zu Berlin
[3]Xiamen University

2018 China Meeting of the Econometric Society

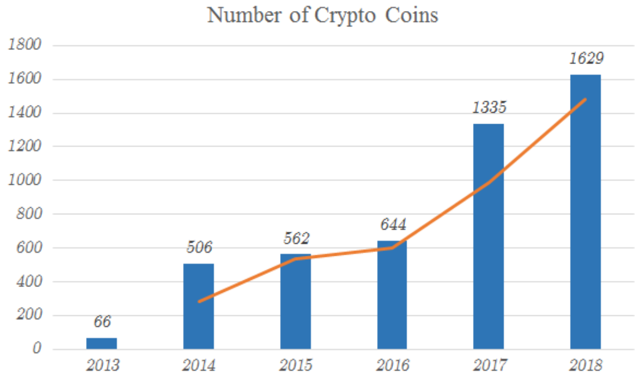
June, 2018

Outline

Emerging of Cryptocurrencies

As of June 15, 2018, by CoinMarketCap.com

- Actively Trading: 997 Coins
- Total Market Cap: $284,515,878,686



Why Cryptos Network?

- **Peer Effect**
  - Open Source of Blockchain — *Clonecoins*
  - Lack of Fundamental Valuation
- **Value of Technology**
  - *Cryptography* determines the security of the coin transactions;
  - *Proof Types* determines the mining activity of the coin developers.
  - Comovement or not?
- **RQ: How fundamental information and return structure jointly determine a market segmentation?**

Methods in Equity Market

Projection of Similarity

- SIC system (Fama and French (1997), Clarke (1989)).
- GICS system: firm's operational characteristics + investors' perceptions (Bhojraj et al. (2003)).
- Investment Style: Farrell (1974), Elton and Gruber (1970) and Brown and Goetzmann (1998).
- Return Comovement: King (1966), Lessard (1974), Grinold, Rudd and Stefek (1989), Roll (1992)), Connor (1997).
- **Product Similarity**: Hoberg and Phillips (2016)

Community Detection

- **Modelling**: Stochastic Blockmodel (Undirected), Stochastic Co-blockmodel (Directed).
- **Existing Methods**: Spectral Clustering, Maximum Likelihood, Bayesian, Modularity Maximization, etc.
- **Difficulties**:
  - **Dynamic Structure**: Bhattacharyya&Chatterjee (2017), Matias&Miele (JRSSB, 2017), Pensky&Zhang (2017), Wilson et al. (2016), etc.
  - **Node features**: Binkiewicz et al. (Biometrika, 2017), Weng&Feng (2017), Yan&Sarkar (2016), Zhang et al. (2017), etc.
  - **Sparsity**: Amini et al. (AoS, 2013), Qin&Rohe (2013), etc.
  - **Degree heterogeneity**: Zhao et al. (AoS, 2012), Qin&Rohe (2013), etc.
  - **Directionality**: Rohe&Yu (2012), Rohe et al. (PNAS, 2016), etc.

Dynamic Stochastic Blockmodel

- **Dynamic Stochastic Blockmodel**:

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

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

  - Probability of a Connection between i and j: $P_t(i,j)$.
  - Clustering Matrix: $Z_t \in \{0,1\}^{N \times K}$.
  - Block Probability Matrix: $B_t \in \mathcal{M}^{K \times K}$ and $P_t(i,j) = B_t(k,k')$..

Dealing with Degree Heterogeneity

- **Dynamic Degree Corrected Stochastic Blockmodel**:

$$A_t(i,j) = \begin{cases} \text{Bernoulli}(P_t(i,j)), & \text{if } i < j \\ 0, & \text{if } i = j \\ A_t(j,i), & \text{if } i > j \end{cases} \quad (3)$$

$$\mathcal{A}_t := \mathbb{E}(A_t | Z_t) = \Psi Z_t B_t Z_t^\top \Psi, \quad (4)$$

  - *Degree Parameter:* $\psi = (\psi_1, \cdots, \psi_N)$.

$$P_t(i,j) = \psi_i \psi_j B_t(k, k').$$

  - Identifiability Restriction:

$$\sum_{i \in \mathcal{G}_k} \psi_i = 1, \quad \forall k \in \{1, 2, \cdots, K\}. \quad (5)$$

Dealing with Sparcity

- **Regularized Graph Laplacian**:

$$L_{\tau,t} = D_{\tau,t}^{-1/2} A_t D_{\tau,t}^{-1/2}, \tag{6}$$

  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 $\tau_t = N^{-1} \sum_{i=1}^{N} D_t(i,i)$.

- **Intuition of Regularization:**
  - Adds a weak edge on every pair of nodes with edge weight τ_t / N.
  - Spectral Clustering: Sparse and stochastic graphs create a lot of small trees that are connected to the core of the graph by only one edge.
  - Regularized Spectral Clustering: leads to a "deeper cut" into the core of the graph thanks to these weak edges.

Incorporating Covariates

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

$$S_t = L_{\tau,t} + \alpha_t C_t^w. \tag{7}$$

where $C_t^w = XW_tX^\top$ and $\alpha_t \in [0, \infty)$ is a tuning parameter

- Example:

$$X = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \end{bmatrix} \longrightarrow XX^\top = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}.$$

- Interpretation of W_t
  - Introduce time-varying interaction between different covariates.
  - Select covariates by setting certain elements of W_t to zero.
  - Relax assumption that similarity in covariates leads to high probability of node connection.
- Choice of W_t: $W_t = X^\top L_{\tau,t} X$.
  - No linkage between i and j: $\mathbb{E}(x^\top L_{\tau,t} x) = 0$;
  - Linkage between i and j: $\mathbb{E}(x^\top L_{\tau,t} x) = \sum_{i,j:A_t(i,j)=1} \frac{x_i x_j}{\sqrt{D_{\tau,t}(i,i) D_{\tau,t}(j,j)}}$.

Dealing with Dynamics

- Discrete Kernel Function

$$\begin{aligned} \mathcal{F}_{r,1} &= \{0, \cdots, r\}, & \mathcal{D}_{r,1} &= \{1, \cdots, r\}; \\ \mathcal{F}_{r,2} &= \{-r, \cdots, r\}, & \mathcal{D}_{r,2} &= \{r+1, \cdots, T-r\}; \\ \mathcal{F}_{r,3} &= \{-r, \cdots, 0\}, & \mathcal{D}_{r,3} &= \{T-r+1, \cdots, T\}. \end{aligned}$$

$$\frac{1}{|\mathcal{F}_{r,j}|} \sum_{i \in \mathcal{F}_{r,j}} i^k W^j_{r,l}(i) = \begin{cases} 1, & \text{if } k = 0, \\ 0, & \text{if } k = 1, 2, \cdots, l. \end{cases} \tag{8}$$

- Discrete Kernel Estimator

$$\widehat{\mathcal{S}}_{t,r} = \sum_{j=1}^{3} \mathbb{1}_{\{t \in \mathcal{D}_{r,j}\}} \left\{ \frac{1}{|\mathcal{F}_{r,j}|} \sum_{i \in \mathcal{F}_{r,j}} W^j_{r,l}(i) S_{t+i} \right\}. \tag{9}$$

Shape of Kernel Function

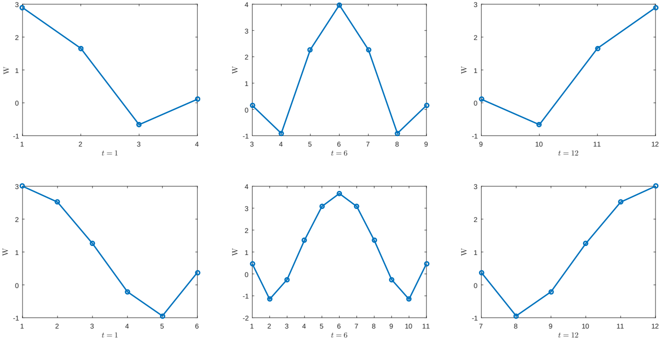


Figure: Discrete kernel functions under bandwidth $r = 3$ and $r = 5$. The horizon is $T = 12$, and the smoothing parameter is $L = 4$.

Choice of Tuning Parameters

- Choice of α_t:

$$\alpha_{\min} = \frac{\lambda_K(L_{\tau,t}) - \lambda_{K+1}(L_{\tau,t})}{\lambda_1(C_t^w)}.$$
$$\alpha_{\max} = \frac{\lambda_1(L_{\tau,t})}{\lambda_K(C_t^w) - \lambda_{K+1}(C_t^w)}.$$
$$\alpha_t = (\alpha_{\min} + \alpha_{\max})/2.$$

- Choice of r:

$$r^* = \text{argmin}_{0 \le r \le T/2} \left(\|\widehat{\mathcal{S}}_{t,r} - \mathcal{S}_{t,r}\| + \|\mathcal{S}_{t,r} - \mathcal{S}_t\| \right).$$

$$\widehat{r} = \max\left\{ 0 \le r \le T/2 : \left\|\widehat{\mathcal{S}}_{t,r} - \widehat{\mathcal{S}}_{t,\rho}\right\| \le 4W_{\max}\sqrt{\frac{N\|\mathcal{S}_t\|_\infty}{\rho \vee 1}}, \text{ for any } \rho < r \right\}.$$

- Determination of K.

Algorithm for Undirected Graphs

Algorithm 1: Covariate-Assisted Spectral Clustering in the Dynamic DCBM

Input : Adjacency matrices A_t for $t = 1, \cdots, T$; Covariates matrix X; Number of communities K; Approximation parameter ε.

Output: Membership matrices Z_t for any $t = 1, \cdots, T$.

1 Calculate regularized graph Laplacian $L_{\tau,t}$ and estimate $\mathcal{S}_t$ by $\widehat{\mathcal{S}}_{t,r}$ defined in (9).

2 Let $\widehat{U}_t \in \mathbb{R}^{N \times K}$ be a matrix representing the first K eigenvectors of $\widehat{\mathcal{S}}_{t,r}$.

3 Let N_+ be the number of nonzero rows of $\widehat{U}_t$, then obtain $\widehat{U}^+ \in \mathbb{R}^{N_+ \times K}$ consisting of normalized nonzero rows of $\widehat{U}_t$, i.e. $\widehat{U}_t^+(i,*) = \widehat{U}_t(i,*) / \left\| \widehat{U}_t(i,*) \right\|$ for i such that $\left\| \widehat{U}_t(i,*) \right\| > 0$.

4 Apply the $(1+\varepsilon)$-approximate k-medians algorithm to the row vectors of $\widehat{U}_t^+$ to obtain $\widehat{Z}_t^+ \in \mathcal{M}_{N_+,K}$.

5 Extend $\widehat{Z}_t^+$ to obtain $\widehat{Z}_t$ by arbitrarily adding $N - N_+$ many canonical unit row vectors at the end, such as, $\widehat{Z}_t(i) = (1, 0, \cdots, 0)$ for i such that $\left\| \widehat{U}_t(i,*) \right\| = 0$.

6 Output $\widehat{Z}_t$.

Assumptions

Assumption (1)

The dynamic network is composed of a series of assortative graphs that are generated under the stochastic block model with covariates whose block probability matrix B_t is positive definite for all $t = 1, \cdots, T$.

Assumption (2)

There are at most $s < \infty$ number of nodes can switch their memberships between any consecutive time instances.

Assumption (3)

For $1 \leq k \leq k' \leq K$, there exists a function $f(\cdot; k, k')$ such that $B_t(k, k') = f(\varsigma_t; k, k')$ and $f(\cdot; k, k') \in \Sigma(\beta, L)$, where $\Sigma(\beta, L)$ is a Hölder class of functions $f(\cdot)$ on $[0, 1]$ such that $f(\cdot)$ are ℓ times differentiable and

$$|f^{(\ell)}(x) - f^{(\ell)}(x')| \leq L|x - x'|^{\beta-\ell}, \text{ for any } x, x' \in [0, 1], \tag{10}$$

with ℓ being the largest integer smaller than β.

Assumptions

Assumption (4)

Let $\lambda_{1,t} \geq \lambda_{2,t} \geq \cdots \geq \lambda_{K,t} > 0$ *be the* K *largest eigenvalues of* $\mathcal{S}_t$ *for each* $t = 1, \cdots, T$*. In addition, assume that*

$$\underline{\delta} = \inf_t \{\min_i \mathcal{D}_{\tau,t}(i,i)\} > 3\log(8NT/\epsilon) \quad \text{and} \quad \alpha_{\max} = \sup_t \alpha_t \leq \frac{a}{NRJ^2\xi},$$

with

$$a = \frac{3\log(8NT/\epsilon)}{\underline{\delta}} \quad \text{and} \quad \xi = \max(\sigma^2 \|L_\tau\|_F \sqrt{\log(TR)}, \sigma^2 \|L_\tau\| \log(TR), NRJ^2/\underline{\delta}),$$

where $\sigma = \max_{i,j} \|X_{ij} - \mathcal{X}_{ij}\|_{\phi_2}$, $L_\tau = \sup_t L_{\tau,t}$.

Consistency for CASC in Dynamic SC-DCBM

Definition of Misclustering:

$$\mathbb{M}_t = \left\{ i \colon \left\| C_{i,t} \mathcal{O}_t^\top - \mathcal{C}_{i,t} \right\| > \left\| C_{i,t} \mathcal{O}_t^\top - \mathcal{C}_{j,t} \right\|, \text{ for any } j \neq i \right\},$$

Theorem

Let clustering be carried out according to Algorithm 1 on the basis of an estimator $\widehat{\mathcal{S}}_{t,r}$ of $\mathcal{S}_t$. Let $Z_t \in \mathcal{M}_{N,K}$ and $P_{\max} = \max_{i,t}(Z_t^\top Z_t)_{ii}$ denote the size of the largest block over the horizons. Then, under Assumption 1-4, as $N, T, R \to \infty$ with $R = o(N)$, the misclustering rate satisfies

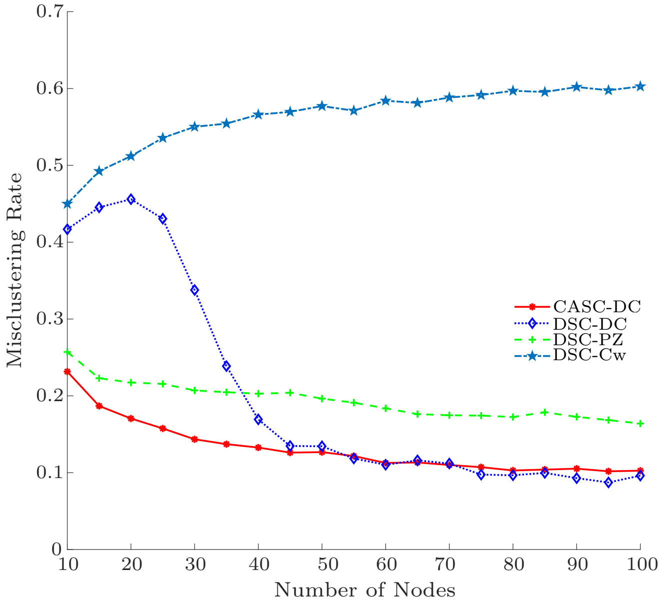
$$\sup_t \frac{|\mathbb{M}_t|}{N} \leq \frac{c(\varepsilon) K W_{\max}^2}{m_z^2 N \lambda_{K,\max}^2} \left\{ (4 + 2c_w) \frac{b}{\underline{\delta}^{1/2}} + \frac{2K}{b} (\sqrt{2P_{\max} rs} + 2P_{\max}) + \frac{NL}{b^2 \cdot l!} \left(\frac{r}{T} \right)^\beta \right\}^2.$$

with probability at least $1 - \epsilon$, where $\lambda_{K,\max} = \max_t\{\lambda_{K,t}\}$ with $\lambda_{K,t}$ being the Kth largest absolute eigenvalue of $\mathcal{S}_t$, where $b = \sqrt{3\log(8NT/\epsilon)}$, $\lambda_{K,\max} = \max_t\{\lambda_{K,t}\}$ and $c(\varepsilon) = 2^9(2+\varepsilon)^2$.

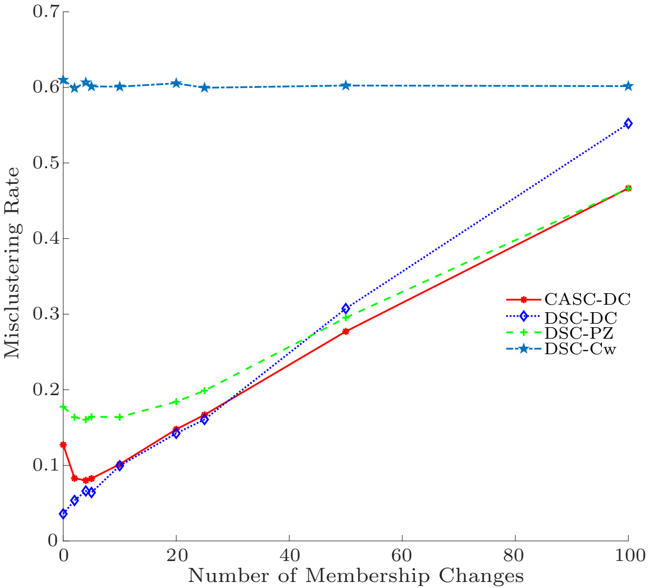
Simulation Settings

- Misclustering Rate with Number of Nodes:
  - Block Probability: $B_t = \frac{t}{T}\begin{bmatrix} 0.9 & 0.6 & 0.3 \\ 0.6 & 0.3 & 0.4 \\ 0.3 & 0.4 & 0.8 \end{bmatrix}$;
  - $R = \lfloor \log(N) \rfloor$, $X(i,j) \overset{i.i.d}{\sim}$ U(0,10);
  - $N = \{10, 15, \cdots, 100\}$;
  - $T = 10$, $s = N^{1/2}$, # of Replication: 100;
- Misclustering Rate with Number of Membership Changes:
  - Block Probability: $B_t = \frac{t}{T}\begin{bmatrix} 0.9 & 0.6 & 0.3 \\ 0.6 & 0.3 & 0.4 \\ 0.3 & 0.4 & 0.8 \end{bmatrix}$;
  - Maximum number of membership changes: $s = [0, 2, 4, 5, 10, 20, 25, 50, 100]$
  - $R = \lfloor \log(N) \rfloor$, $X(i,j) \overset{i.i.d}{\sim}$ U(0,10);
  - $N = 100$, $T = 10$, # of Replication: 100;

Performance with Growing Number of Vertices



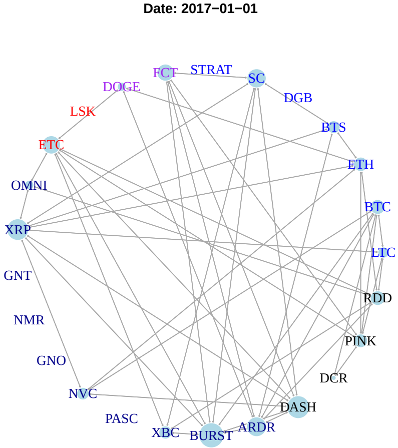
Performance with Growing Number of Membership Changes



Data

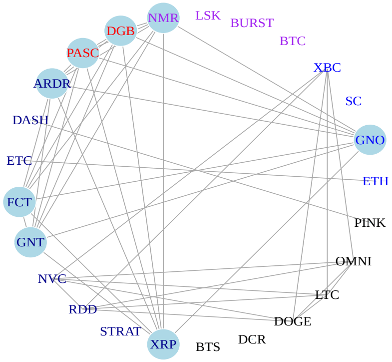
- Data Source: Cryptocompare
- Sample Period:
  - In-sample Estimation: from 2015-08-31 to 2017-12-31.
  - Out-of-Sample Tests: from 2018-01-01 to 2018-03-30.
- Cryptocurrency Daily Return:
  - Top 200 Cryptos Sorted on Market Cap, Age, Maximum Price and Dollar Volume;
- Contract Information:
  - Algorithm
  - Proof Types

Return Network Structure from Adaptive LASSO

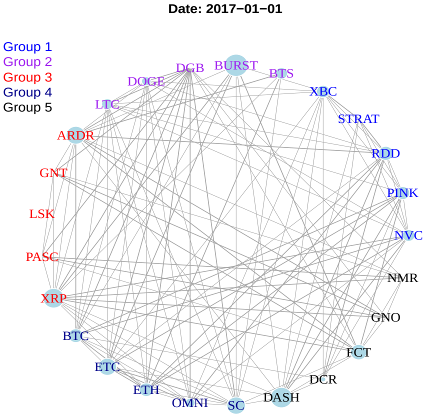


Visualization: Node Features (Attribution Network Structure)

Attribution: Algorithm



Visualization: Combined Network Structure



Grouping Results

Table: Top 5 Group Member

	Cryptocurrency
Group 1	Novacoin, Pinkcoin, Reddcoin, Stratis, Bitcoinplus
Group 2	Litecoin, Dogecoin, Bitshares, Burstcoin, Digibyte
Group 3	Ripple, Ardor, Golem Network Token, Lisk, Pascal Coin
Group 4	Bitcoin, Ethereum, Ethereum Classic, Omni, Siacoin
Group 5	Digital Cash, Decred, Factoids, Gnosis, Numerai

DSBM (Bhattacharyya&Chatterjee, 2017) Evaluation I

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

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

Group ID	Within-Group	Cross-Group	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 (Bhattacharyya&Chatterjee, 2017) Evaluation II

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

Table: Evaluation Criteria: **Algorithm**

Group ID	Within-Group	Cross-Group	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 (Bhattacharyya&Chatterjee, 2017) Evaluation III

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

Table: Evaluation Criteria: **Proof Types**

Group ID	Within-Group	Cross-Group	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

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

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

Group ID	Within-Group	Cross-Group	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

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

Table: Evaluation Criteria: **Algorithm**

Group ID	Within-Group	Cross-Group	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

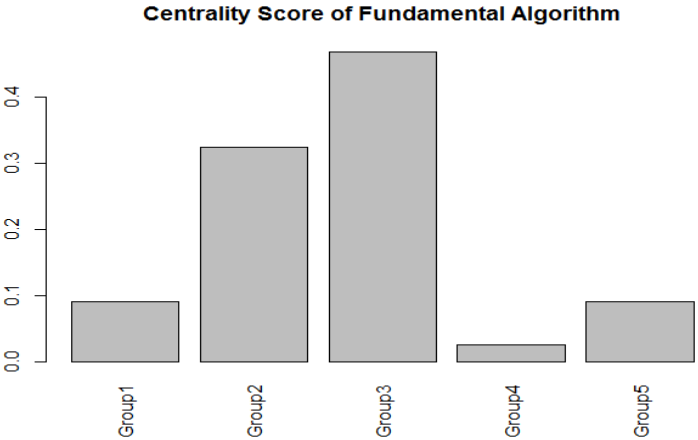
Covariate-assisted Spectral Clustering Evaluation III

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

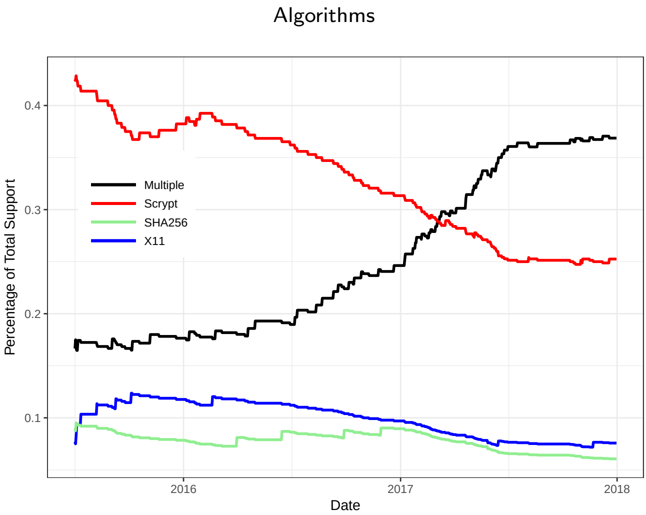
Table: Evaluation Criteria: **Proof Types**

Group ID	Within-Group	Cross-Group	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

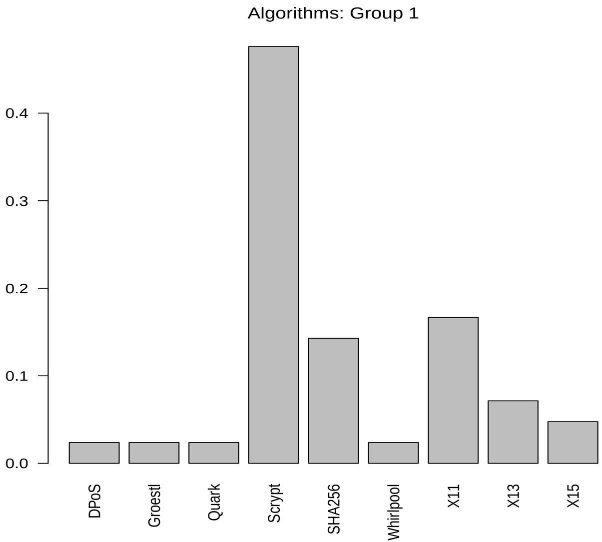
Asset Pricing Inference: Group Centrality



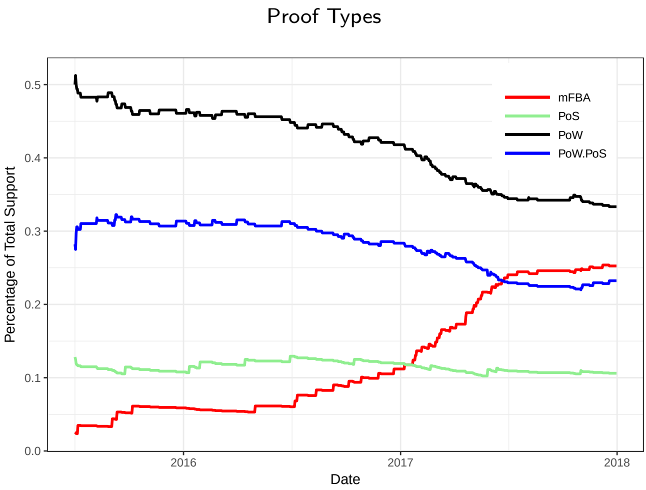
Algorithms Evolution Over Time



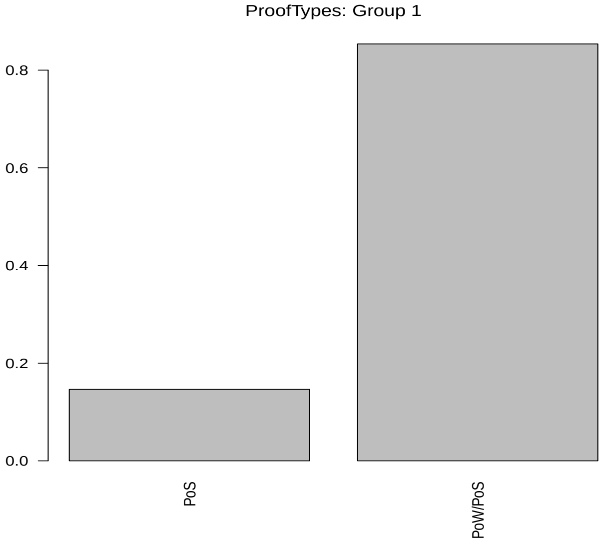
Fundamental Comparison under Different Centrality Score: Algorithm



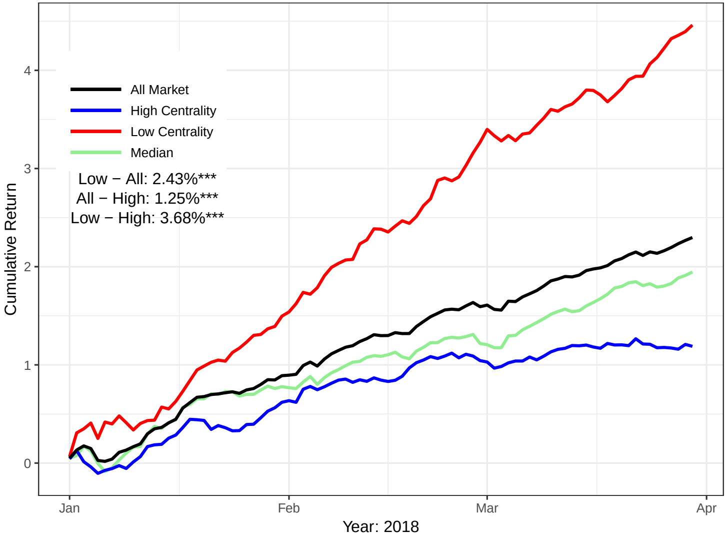
Proof Types Evolution Over Time



Fundamental Comparison under Different Centrality Score: Proof Types



Cross Sectional Return predictability Comparison



Conclusions

What we do:

- Extends regularized spectral clustering methods to analysing dynamic networks (both directed and undirected), especially when there are membership changes.
- Incorporate node covariates into the network to assist community detection in dynamic networks.

Takeaways:

1. Attribution Matrix provides valuable information to connect within group members.
2. Return-based Adjacency Matrix reveal connections across different groups.
3. Behavioral bias is stronger for those groups with low fundamental centrality.