SONIC: Social Networks with Influencers and Communities

Cathy Y.H. Chen
Wolfgang Karl Härdle
Yegor Klochkov

Ladislaus von Bortkiewicz Chair of Statistics
Humboldt–Universität zu Berlin
http://lvb.wiwi.hu-berlin.de

Analysis of networks

- Social network produces high-dimensional time series
  - Daily sentiment as quantification of one's opinion
  - Missing observations
- Adjacency matrix must be estimated
- Problem: network size is immense
- Smart data analytics based on StockTwits

StockTwits sentiment

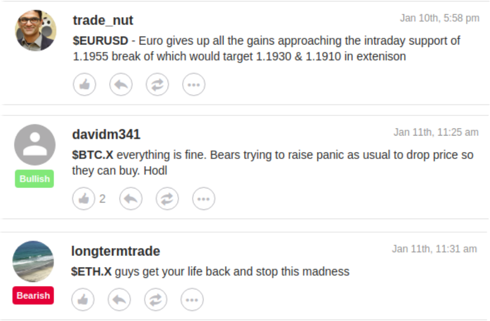


Figure 1: https://www.stocktwits.com message examples

Sentiment weight: $tf \cdot idf$ scheme

the brown cow

For each term t,

$$SW(t) = \frac{tf \cdot idf_{pos}(t) - tf \cdot idf_{neg}(t)}{tf \cdot idf_{pos}(t) + tf \cdot idf_{neg}(t)}$$

where

$$tf \cdot idf_{pos}(t) = freq_{pos}(t) \cdot \log \frac{\text{positive messages}}{\text{positive occurences of } t}$$

$$tf \cdot idf_{neg}(t) = freq_{neg}(t) \cdot \log \frac{\text{negative messages}}{\text{negative occurences of } t}$$

Crypto-specific terms

Term	Sentiment weight
🚀	0.90
💩	−0.91
💩💩	−0.98
hodl	0.32
hodl !	0.64
hackers	−0.83
miner	0.62
tulip mania	−0.94
bitcoin 😂	−0.73
scam	−0.77
f***ing scam	−0.86

CYH Chen Dictionary

@AAPL

Numbers

Figure 2: SWs constructed from @AAPL messages

@BTC

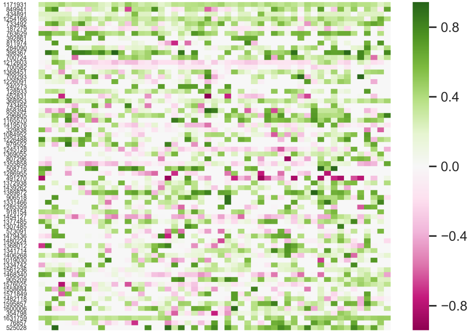


Figure 3: SWs constructed from @BTC messages

Modeling opinion networks

- Sentiment weights (SW) for N users during T days

$$Z_{it} = \text{average of } SW\text{s for user } i \text{ during day } t$$
$$Z_t \in \mathbb{R}^N$$

- Missing observations

$$Z_{it} = \delta_{it} Y_{it}, \qquad \text{i.i.d. } \delta_{it} \sim \text{Bernoulli}(p_i)$$

where Y_{it} is the *opinion*, Z_{it} — *expressed opinion*

Modeling opinion networks

- Network interactions through VAR

$$Y_t = \Theta Y_{t-1} + W_t, \qquad \mathsf{E}[W_t | \mathcal{F}_{t-1}] = 0,$$
$$\Theta \in \mathbb{R}^{N \times N},$$

  - Unknown adjacency matrix
  - Curse of dimensionality $T \lesssim N$

Zhu, X., Pan, R., Li, G., Liu, Y. and Wang, H.
Network vector autoregression
Annals of Statistics, 2017

more Literature

\Theta_ij = \beta*\A_ij/sum(A_ik, k=1..N), known A!

Influencer

- Relationships expressed by VAR parameters

$$\Theta_{ij} \neq 0 \Rightarrow \; i \;\text{ follows }\; j$$

- Influencer — followed by a significant part of network
- The amount of influencers is much smaller than N
  - motivated by real life social networks
  - sparsity constraints reduce sample complexity

Research question

- Each user is affected at most by s others

$$\max_i \sum_j \mathbf{1}(\Theta_{ij} \neq 0) \leq s;$$

- Sparsity grows up to $\|\Theta\|_0 \leq Ns$, so lasso requires

$$\frac{(sN)\log N}{T} \ll 1$$

($\|\Theta\|_0 = \sum_{ij} \mathbf{1}(\Theta_{ij} \neq 0)$)

- *Structural assumptions appropriate for social networks?*

Outline

Stochastic Block Model

- Partition of nodes into K disjoint *communities*

$$C_1 \cup \cdots \cup C_K = \{1, \ldots, N\}, \qquad C_i \cap C_j = \emptyset$$

- Independent edges $\mathrm{P}(a_{ij} = 1) = \Omega_{ij}$ with

$$\Omega_{ij} = B_{l_i l_j}, \qquad \text{for } i \in C_{l_i},\ j \in C_{l_j}$$

  (usually arbitrary diagonal elements Ω_{ii} allowed)

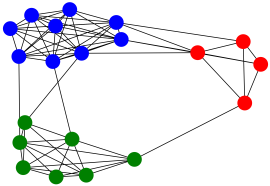
- Low rank assumption: $\mathrm{Rank}(\Omega) \leq K$

- Example for $N = 5,\ K = 2$

$$\Omega = \begin{pmatrix} a & a & b & b & b \\ a & a & b & b & b \\ b & b & c & c & c \\ b & b & c & c & c \\ b & b & c & c & c \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{pmatrix} \overbrace{\begin{pmatrix} a & b \\ b & c \end{pmatrix}}^{B} \underbrace{\begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 \end{pmatrix}}_{\text{index matrix}}$$

- Realization for $N = 20,\ K = 3$

$$\Omega = \begin{pmatrix} 1.0 & 0.1 & 0.05 \\ 0.1 & 0.9 & 0.05 \\ 0.05 & 0.05 & 0.95 \end{pmatrix}$$



New structural approach

- Few influencers: row-wise sparsity

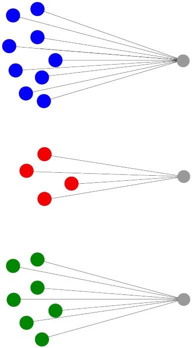
$$\max_i \sum_j \mathbf{1}(\Theta_{ij} \neq 0) \leq s$$

- Communities $C_1, \ldots, C_K$ with shared dependencies

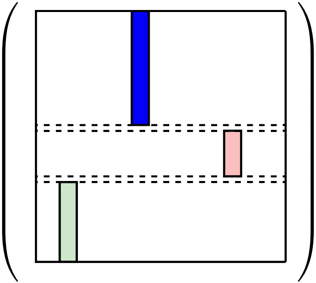
$$\Theta_{i\cdot} = \Theta_{i'\cdot}, \qquad i, i' \in C_l$$

Chen, Y., Trimborn, S., Zhang, J.
Discover Regional and Size Effects in Global Bitcoin Blockchain via Sparse-Group Network AutoRegressive Modeling
preprint, 2018

Influencers and communities



$\Theta =$



Clustering

- Via user *labels*: $\mathcal{C} = (l_1, \ldots, l_N)$, where $l_i \in [K]$

$$C_l = \{i : \; l_i = l\}$$

- Relabeling $\mathcal{C} \sim \mathcal{C}'$ iff there is π

$$l_i = \pi(l_i'), \qquad i = 1, \ldots, N$$

- *Equivalent* distance,

$$\begin{aligned} d(\mathcal{C}, \mathcal{C}') &= \min_{\pi} \sum_{i=1}^{N} \mathbf{1}(l_i \neq \pi(l_i')) \\ &= \min_{\pi} \sum_{j=1}^{K} |C_j \setminus C'_{\pi(j)}| \end{aligned}$$

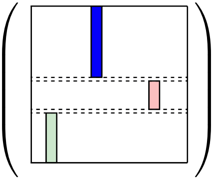
Block structure

- Shared dependencies in each community

$$l_i = l_{i'} \Rightarrow \Theta_{ij} = \Theta_{i'j}, \qquad j = 1, \dots, N$$

- Example $K = 3$

$\Theta =$



(up to a permutation)

Block structure 2

- Each column of Θ is a span of

$$z_{C_j} = \frac{1}{\sqrt{|C_j|}} \begin{pmatrix} 0 \\ 0 \\ \boxed{\begin{matrix} 1 \\ \vdots \\ 1 \end{matrix}} \\ 0 \end{pmatrix} \dashrightarrow C_j \qquad j = 1, \dots, K$$

- Factor representation

$$\Theta = Z_{\mathcal{C}} V^\top, \qquad V \in \mathbb{R}^{N \times K}$$

where $Z_{\mathcal{C}} = [z_{C_1}, z_{C_2}, \dots, z_{C_K}]$

Influencers and sparsity

- In social media users are influenced by a small group of people (e.g. celebrities)

$$\text{user } j \text{ is } \text{influencer} \text{ iff } \Theta_{ij} \neq 0 \text{ for some i}$$

- Induce row-wise sparsity on $\Theta = Z_{\mathcal{C}} V^\top$, i.e.

$$\max_j \sum_{i=1}^{N} \mathbf{1}(V_{ij} \neq 0) \leq s$$

- Sparsity + clusterisation = dimensionality reduction

Penalized loss function

- Define

$$R_\lambda(V;\mathcal{C}) = \frac{1}{2}\sum_{t=2}^{T} \| Y_{t+1} - Z_{\mathcal{C}} V^\top Y_t \|^2 + \lambda \| V \|_{1,1}$$

- ℓ_1 penalty $\| V \|_{1,1} = \sum_{ij} |V_{ij}|$ with a tuning parameter λ
- Minimum contrast estimator

$$(\hat{V}_\lambda, \hat{\mathcal{C}}_\lambda) = \arg\min R_\lambda(V;\mathcal{C}), \qquad \hat{\Theta}_\lambda = Z_{\hat{\mathcal{C}}_\lambda} \hat{V}_\lambda^\top$$

LASSO estimator for V

- Penalized risk minimization with a given clustering $\mathcal{C}$

$$\hat{V}_{\mathcal{C},\lambda} = \arg\min_{V} R_\lambda(V;\mathcal{C})$$

- Convex problem for V
- Parallelization is possible: K independent subproblems due to $Z_{\mathcal{C}}^\top Z_{\mathcal{C}} = I$

$$\hat{v}_j = \arg\min_{v\in\mathbb{R}^\top} \frac{1}{2}\sum_{t=1}^{T-1}\left\{(z_{C_j}^\top Y_{t+1}) - v^\top Y_t\right\}^2 + \lambda\|v\|_1$$

Greedy procedure

Minimize the risk for clustering

$$F_\lambda(\mathcal{C}) = \min_V R_\lambda(V; \mathcal{C}) \to \min_{\mathcal{C}}$$

1. randomly initialize $C_1, \ldots, C_K$;
2. for each $i = 1, \ldots, N$ change the label of the ith user

$$F_\lambda(\mathcal{C}) \to \min_{l_i}$$

   (i.e. $d(\mathcal{C}^{old}, \mathcal{C}^{new}) \le 1$)
3. repeat (2) until clustering does not change;

Alternating procedure

see p4.3

Joint risk

$$R_\lambda(V;\mathcal{C}) = \frac{1}{2}\mathrm{Tr}(V^\top \hat{\Sigma} V) - \mathrm{Tr}(V^\top \hat{A} Z_{\mathcal{C}}) + \lambda \|V\|_{1,1}$$

1. randomly initialize $\mathcal{C} = (C_1, \ldots, C_K)$;
2. estimate $\hat{V}_{\mathcal{C},\lambda}$ using LASSO;
3. repeat:
   3.1 perform greedy procedure for

   $$-\mathrm{Tr}(\hat{V}^\top A Z_{\mathcal{C}}) \to \min_{\mathcal{C}}$$

   3.2 update $\hat{V}_{\mathcal{C},\lambda}$ using the new clustering;
   3.3 repeat until does not change

Missing observations

Unobserved "opinion" process

$$Y_t = \Theta^* Y_{t-1} + W_t$$

- true parameter Θ^*
- innovations W_t with $\mathrm{E}(W_t | \mathcal{F}_{t-1}) = 0$

Observed variables

$$Z_{it} = \delta_{it} Y_{it}, \qquad \delta_{it} \sim \mathrm{Bernoulli}(p_i)$$

- user i makes a post with probability p_i every day
- still allows estimation of the covariance of Y

Loss decomposition

$$
\begin{aligned}
L(\Theta) =& \frac{1}{2T} \sum_{t>1} \| Y_t - \Theta Y_{t-1} \|_2^2 \\
=& \frac{1}{2} \mathrm{Tr}(\Theta \widetilde{\Sigma} \Theta^\top) - \mathrm{Tr}(\Theta \widetilde{A}) + \frac{1}{2T} \sum_{t>1} \| Y_t \|^2
\end{aligned}
$$

where

$$
\widetilde{\Sigma} = T^{-1} \sum_{t>1} Y_{t-1} Y_{t-1}^\top, \qquad \widetilde{A} = T^{-1} \sum_{t>1} Y_{t-1} Y_t^\top
$$

Probabilities of non-zero observation

$$\hat{p}_i = T^{-1} \sum_t \mathbf{1}(Z_{it} \neq 0)$$

Observed sample covariance

$$\Sigma^* = T^{-1} \sum_t Z_t Z_t^\top, \qquad A^* = T^{-1} \sum_{t>1} Z_{t-1} Z_t^\top$$

Covariance estimation

$$\begin{aligned} \hat{\Sigma} =& \operatorname{diag}(\hat{p})^{-1} \operatorname{Diag}(\Sigma^*) + \operatorname{diag}(\hat{p})^{-1} \operatorname{Off}(\Sigma^*) \operatorname{diag}(\hat{p})^{-1} \\ \hat{A} =& \operatorname{diag}(\hat{p})^{-1} A^* \operatorname{diag}(\hat{p})^{-1} \end{aligned}$$

▸ Upper bound

Lounici, K.
High-dimensional covariance matrix estimation with missing observations
Bernoulli, 2014

Local result

- Recall the definition

$$d(\mathcal{C}, \mathcal{C}') = \sum_{j=1}^{K} |C_j \setminus C_j'|$$

(1 if only one label differs)

- Greedy algorithm changes one label at each step
- If $\mathcal{C}$ is such that

$$\min_{d(\mathcal{C}, \mathcal{C}')=1} F_\lambda(\mathcal{C}') \geq F_\lambda(\mathcal{C}),$$

the algorithm stops at $\mathcal{C}$ — "locally optimal";

Conditions

- $\Theta^* = Z^*[V^*]^\top$ with $Z^* = Z_{\mathcal{C}^*}$ and
$$V = [v_1^*, \ldots, v_K^*], \qquad \|v_j^*\|_0 \le s,$$
where $\|x\|_0 = \sum \mathbf{1}(x_i \ne 0)$;
- $\|\Theta^*\|_\infty = \|V^*\|_\infty \le \gamma < 1$;
- condition number of $[V^*]^\top \Sigma V^*$ bounded by κ_0;
- significant size of clusters
$$\min_j |C_j^*| / \max_j |C_j^*| \ge \alpha \in (0, 1]$$

ERC condition

Denote *exact recovery coefficient* (ERC)

$$\mathrm{ERC}(\Lambda) = 1 - \|\Sigma_{\Lambda^c \Lambda} \Sigma_{\Lambda,\Lambda}^{-1}\|_{1,\infty},$$

where $\|A\|_{1,\infty} = \max_i \sum_j |A_{ij}|$

- Suppose,

$$\mathrm{ERC}(\Lambda_j) \geq 3/4$$

  for each $\Lambda_j = \mathrm{supp}(v_j^*)$

Tropp, J.
Just relax: Convex programming methods for identifying sparse signals in noise
IEEE Transactions on Information Theory, 2006

Network size limits

We work in the regime

$$\frac{s n^* \log N}{T p_{\min}^2} \le c$$

with $c > 0$ not depending on N, s, K, T, δ_i;

- *largest cluster size* n^* within the range

$$\frac{N}{K} \le n^* \le \frac{\alpha^{-1} N}{K}$$

- allows $N > T$ for sufficiently large K

Local result

Theorem

There are constants c, C such that if

- *the tuning parameter satisfies*

$$C\sqrt{\frac{\log N}{Tp_{\mathrm{min}}^2}} \le \lambda \le c\left\{s^{-1} \vee (\sqrt{s}K)^{-1}\right\},$$

- $N \ge c'\lambda^2 sK,$

then with probability at least $1 - 1/N$ there is a locally optimal $\hat{\mathcal{C}}$ such that $\hat{\Theta}_\lambda = Z_{\hat{\mathcal{C}}}\hat{V}_{\hat{\mathcal{C}},\lambda}$ satisfies

$$\|\hat{\Theta}_\lambda - \Theta^*\|_F \lesssim \lambda K\sqrt{s}$$

Proof sketch

Ideally we choose

$$\lambda^* \sim \sqrt{\frac{\log N}{T p_{\min}^2}}$$

In this case the bound is

$$\|\hat{\Theta}_{\lambda^*} - \Theta^*\|_F \lesssim \sqrt{\frac{sK^2 \log N}{T p_{\min}^2}}$$

Simulations

- $N = 100$, $T = 100$
- Construct Θ^* such that
  - $K = 2..30$ with C_j having equal (± 1) sizes;
  - for each $j = 1, \ldots, K$

$$\text{supp}(v_j^*) = 1;$$

  - $\|\Theta^*\|_{op} = 0.5$
- Simulate

$$Y_t = \sum_{k \geq 0} [\Theta^*]^k W_{t-k}, \qquad W_t \sim \mathrm{N}(0, I_N)$$

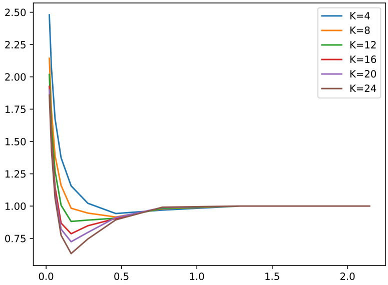


Figure 4: Normalized error $E\|\hat{\Theta}_\lambda - \Theta^*\|_F / \|\Theta^*\|_F$ against λ

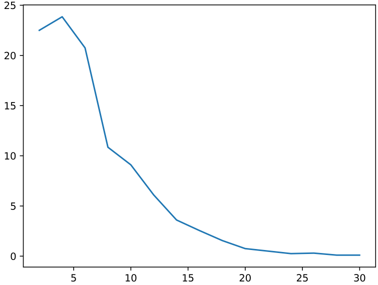


Figure 5: Cluster difference for optimal λ against $K = 2, \ldots, 30$

Choice of λ

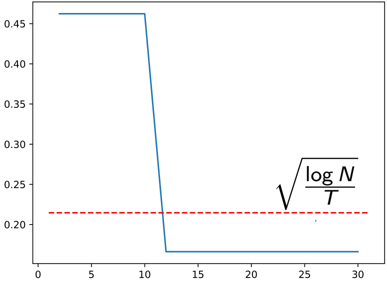


Figure 6: Optimal λ for $K = 2, \ldots, 30$

- Best choice appears to be

$$\lambda^* \approx \sigma \sqrt{\frac{\log N}{T p_{\min}^2}};$$

- In case of unknown σ take

$$\hat{\sigma} = \lambda_{K+1}(\hat{\Sigma});$$

Experiment with StockTwits

- Preprocessing
  - pick users with $\hat{p}_i \geq 0.5$ (small p_i produce too much error)
  - persistence: covariance estimator requires stationarity of (δ_{it})
  - result: 46 users & 72 days
- Estimation
  - 100 iterations with 100 initializations

@AAPL

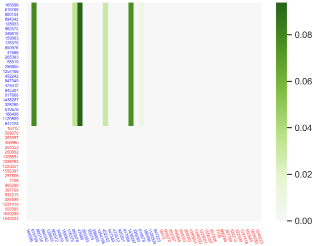


Figure 7: Estimated Θ for AAPL daily sentiment

Opinion_Networks_in_Social_Media

How to choose K?

@AAPL

ChartMill

ChartMill

Follow Message

About

Chartmill.com, the number 1 technical stock screener incl. VWAP, Relative strength and a lot more! Additional tools are added on a regular basis. Made by real traders, used by real traders!

Belgium

Technical, Swing Trader, Professional

http://www.chartmill.com

Member since January 26th 2011

536.9k	905	10.7k	32	20
Ideas	Following	Followers	Liked	Watchlist

OptionsPlayers

Steve /OptionsDayTrader77

Follow Message

About

Founder/CEO of OptionsPlayers.com. The #1 Options Trading Community with live chat, training, forums, and trading systems. I alert my plays (open/close) live! Check us out! All my posts are MY opinion. Do your own DD when trading.

Worldwide

Options, Momentum, Day Trader, Professional

https://www.optionsplayers.com

Member since October 22nd 2015

21.5k	6.0k	5.8k	4.9k	30
Ideas	Following	Followers	Liked	Watchlist

MariaC82

Maria Chaudhry

Follow Message

About

Former Broker (Schwab, Scottrade, TD Ameritrade Institutional)🔷 Options Scalper🔷 Swing Trader🔷 Pro Sharpie Thrower🔷 14+ Yrs Experience🔷 Posts Are NOT Trade Recommendations🔷 #BlackBoxStocks🔷 #StayGreen⬛ ✅ 💯 💖

Dallas, Texas

Equities, Options, Technical, Day Trader, Professional

http://bit.ly/BlackBoxStocks2

Member since September 30th 2016

30.3k	137	30.9k	41.2k	297
Ideas	Following	Followers	Liked	Watchlist

EquityTom

EquityTom

Follow Message

About

Trading stocks that setup at / near important levels. Provide Price Action Analysis on chart review.

USA

Equities, Options, Technical, Day Trader, Professional

http://bit.ly/ProfitActionRoom

Member since March 9th 2018

1.1k	3	1.4k	0	22
Ideas	Following	Followers	Liked	Watchlist

@BTC

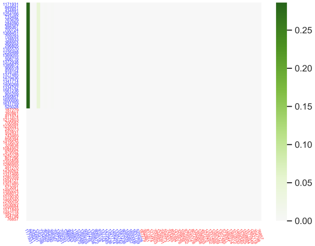
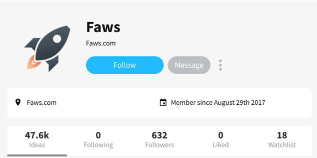


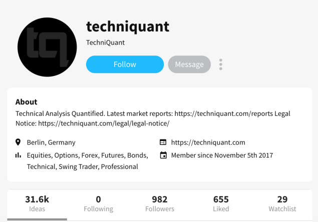
Figure 8: Estimated Θ for BTC daily sentiment

Opinion_Networks_in_Social_Media

How to choose K?

@BTC





Cha, M., Haddadi, H., Benevenuto, F., Gummadi, K.P.
Measuring User Influence in Twitter: The Million Follower Fallacy
4th AAAI conference on weblogs and social media, 2010

Outlook

- Network autoregression for social media
- Application to StockTwits sentiment
  - identify clusters and influencers
- How to verify our model?
  - follower/followee relationship unavailable in StockTwits
  - analysis of cluster stability

Literature

Zhu, X., Pan, R., Li, G., Liu, Y. and Wang, H.
Network vector autoregression
Annals of Statistics, 2017

Chernozhukov, V., Härdle, W.K., Huang, C., Wang, W.
LASSO-driven Inference in Time and Space
preprint, 2018

Chen, C.Y.H, Härdle, W, Okhrin, Y.
Tail event driven networks of SIFIs
Journal of Econometrics, 2019
DOI: 10.1016/j.jeconom.2018.09.016

@AAPL

- Sample period: 2017/05/22 to 2019/01/27 ($\sim$ 600 days)
- 449,761 messages from 26,521 users
  - 29.6% bullish / 10.7% bearish / 59.7% unlabelled
  - training dataset 99,985 positive / 36,100 negative
- Lexicon from @AAPL messages
  - 543 positive terms
  - 786 negative terms

Back

Zhu, X., Wang, W., Wang, H. and Härdle, W.K.
Network quantile autoregression
Journal of Econometrics, 2019

Chernozhukov, V., Härdle, W.K., Huang, C., Wang, W.
LASSO-driven Inference in Time and Space
Ann. Stat., to appear

Chen, C. H.-Y., Härdle, W.K., Liu, K.
Financial Risk Meter
Empirical Economics, to appear

Chen, Y., Trimborn, S., Zhang, J.
Discover Regional and Size Effects in Global Bitcoin Blockchain via Sparse-Group Network AutoRegressive Modeling
preprint, 2018

subgaussian innovations

$$\|\langle u, W_t\rangle\|_{\psi_2} \lesssim \|\langle u, W_t\rangle\|_{L_2}$$

where

$$\|X\|_{\psi_2} = \inf\{C > 0 : \ \mathsf{E}\exp(|X|^2/C) \leq 2\}$$
$$\|X\|_{L_2} = \mathsf{E}^{1/2}|X|^2$$

Lemma

Suppose,

- W_t *are subgaussian;*
- $\|\Theta^*\|_{op} \leq \gamma < 1$*;*
- $P, Q \in \mathbb{R}^{N\times N}$ *are projectors of ranks* $\leq M$*.*

It holds with probability at least $1 - e^{-u}$ *for* $u \geq 1$

$$\begin{aligned}&\|P(\hat{\Sigma}-\Sigma)Q\|_{op}\\&\qquad\leq C\|\Sigma\|_{op}\left(\sqrt{\frac{M(\log N + u)}{Tp_{\min}^2}} \bigvee \frac{M(\log N + u)\log T}{Tp_{\min}^2}\right),\end{aligned}$$

where $C = C(\gamma)$

Lemma

Suppose,

- $Y_1, \dots, Y_T$ *are subgaussian;*
- $\|\Theta^*\|_{op} \leq \gamma < 1$*;*
- $P, Q \in \mathbb{R}^{N \times N}$ *are projectors of ranks* $\leq M$*.*

It holds with probability at least $1 - e^{-u}$ *for* $u \geq 1$

$$\begin{aligned}&\|P(\hat{A} - A)Q\|_{op} \\ &\quad \leq C\|\Sigma\|_{op}\left(\sqrt{\frac{M(\log N + u)}{Tp_{\min}^2}} \bigvee \frac{M(\log N + u)\log T}{Tp_{\min}^2}\right),\end{aligned}$$

where $C = C(\gamma)$

Back

Theorem (Chapter 4)

Let $X_1, \ldots, X_T \in \mathbb{R}^{d\times d}$ *are independent with* $\left\|\|X_i\|\right\|_{\psi_1} < \infty$. *Set*

- $\sigma^2 = \left\| \mathrm{E} \sum_{i=1}^T X_i^2 \right\|$
- $U = \left\| max_{i\le T} \|X_i\| \right\|_{\psi_1}$

Then for each $t \geq 1$

$$\mathrm{P}\left(\left\| \sum_{i=1}^N X_i - \mathrm{E}\, X_i \right\| \lesssim \sigma\sqrt{t} + Ut \right) \leq de^{-t}$$

Here $\|Y\|_{\psi_1} = \inf\{C > 0 : \ \mathrm{E}\exp(|Y|/C) \leq 2\}$

Back

How to choose number K?

- Analyze stability of cluster estimation
- Consider few shorter windows (say, of length $\frac{3T}{4}$)

$$\mathcal{I}_1 = \left[0, \frac{3}{4}T\right], \mathcal{I}_2 = \left[\frac{1}{20}T, \left(\frac{3}{4} + \frac{1}{20}\right)T\right], \ldots, \mathcal{I}_6 = \left[\frac{1}{4}T, T\right]$$

- Compare resulting clusterings

$$d(\hat{\mathcal{C}}(\mathcal{I}_1), \hat{\mathcal{C}}(\mathcal{I}_j)), \qquad j = 2, \ldots, 6,$$

where $\hat{\mathcal{C}}(\mathcal{I})$ is estimated using data from time interval $\mathcal{I}$

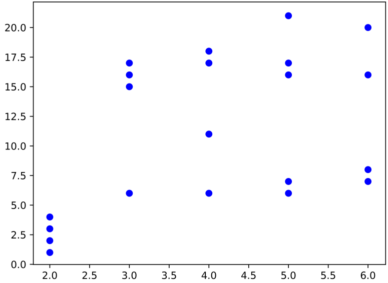


Figure 9: Cluster differences for $K = 2, 3, 4, 5, 6$ for the BTC dataset

Back

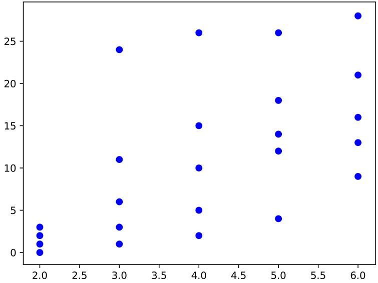


Figure 10: Cluster differences for $K = 2, 3, 4, 5, 6$ for the BTC dataset

Back

Proof sketch

- Exact recovery $\mathrm{supp}(v_j) = \Lambda_j$ in the neighbourhood of $\mathcal{C}^*$ (w.h.p.);
- Explicit expression for $F_\lambda(\mathcal{C})$;
- Quadratic deviation of deterministic part v.s. linear growth of stochastic part

Gribonval, R., Jenatton, R., Bach, F.
Sparse and spurious: dictionary learning with noise and outliers
IEEE Transactions on Information Theory, 2015

Back

V-step

For arbitrary $\mathcal{C} = (C_1, \dots, C_K)$ we solve for each $j = 1, \dots, K$

$$\hat{v}_j = \arg\min \frac{1}{2} v^\top \hat{\Sigma} v - v^\top \hat{A} z_j + \lambda \|v\|_1,$$

where $Z_{\mathcal{C}} = [z_1, \dots, z_K]$

Lemma

Denote, $\hat{c} = \hat{A} z_j$. *Suppose,*

$$\|\hat{\Sigma}_{\Lambda_j^c,\Lambda_j} \hat{\Sigma}^{-1}_{\Lambda_j,\Lambda_j} \hat{c}_{\Lambda_j} - \hat{c}_{\Lambda_j^c}\|_\infty \leq \lambda(1 - \|\hat{\Sigma}_{\Lambda_j^c,\Lambda_j} \hat{\Sigma}^{-1}_{\Lambda_j,\Lambda_j}\|_{1,\infty})$$

where $\|A\|_{1,\infty} = \max_i \sum_j |A_{ij}|$. *Then, supp*$(\hat{v}_j) \subset \Lambda_j$.

Solution with $\mathrm{supp}(\hat{v}_j) \subset \Lambda_j$

$$\hat{v}_j = \Sigma^{-1}_{\Lambda_j,\Lambda_j}(\hat{A}_{\Lambda_j,\cdot} z_j - \lambda g)$$

with some $g \in \mathbb{R}^{|\Lambda_j|}$, $\|g\|_\infty \le 1$

If $\|\hat{v}_j - v_j^*\|_\infty < \min_{i \in \Lambda_j} |V_{ij}^*|$ then follows *explicit form*

$$\hat{v}_j = \hat{\Sigma}^{-1}_{\Lambda_j,\Lambda_j}(\hat{A}_{\Lambda_j,\cdot} z_j - \lambda (s_j^*)_{\Lambda_j})$$

where $s_j^* = \mathrm{sign}(v_j^*)$

Exact recovery yields

$$
\begin{aligned}
F_\lambda(\mathcal{C}) &= -\frac{1}{2}\sum_{j=1}^{K} \hat{v}_j^\top \hat{\Sigma} \hat{v}_j \\
&= \underbrace{-\frac{1}{2}\sum_{j=1}^{K} \left(\hat{A}_{\Lambda_j,\cdot} z_j - \lambda (s_j^*)_{\Lambda_j}\right)^\top \hat{\Sigma}^{-1}_{\Lambda_j,\Lambda_j} (\hat{A}_{\Lambda_j,\cdot} z_j - \lambda (s_j^*)_{\Lambda_j})}_{=\Phi_\lambda(\mathcal{C})}
\end{aligned}
$$

Lemma

Let $\lambda, \bar{r} > 0$ *be such that*

$$C\sqrt{\frac{\log N}{Tp_{\min}^2}} \le \lambda \le c\, s^{-1}, \qquad \sqrt{\frac{n^* \log N}{Tp_{\min}^2}}\,\bar{r}^2 \le c\lambda.$$

Then, with probability $\ge 1 - N^{-\beta}$

$$F_\lambda(\mathcal{C}) = \Phi_\lambda(\mathcal{C}), \qquad \forall \mathcal{C} : \ \|Z_{\mathcal{C}} - Z_{\mathcal{C}^*}\|_{\mathsf{F}} \le \bar{r}$$

Moreover,

$$\|\hat{V}_{\mathcal{C},\lambda} - V^*\|_{\mathsf{F}} \lesssim \lambda\sqrt{Ks}$$

Estimation of clusters

Define,

$$\hat{\mathcal{C}} = \arg \min_{\mathcal{C}:\, \|Z_{\mathcal{C}} - Z_{\mathcal{C}*}\|_F \le \bar{r}} \Phi_\lambda(\mathcal{C})$$

quadratic vs. (at most) linear for $r = \|Z_{\mathcal{C}} - Z_{\mathcal{C}*}\|_F \le \bar{r}$

$$\Phi_\lambda(\mathcal{C}) - \Phi_\lambda(\mathcal{C}^*) \ge \left(c - C\sqrt{\frac{sn^* \log N}{Tp_{\min}^2}} \right) r^2 - C\lambda\sqrt{s}Kr$$

Define,

$$\bar{\Phi}_\lambda(\mathcal{C}) = -\frac{1}{2}\sum_{j=1}^{K}\left(A_{\Lambda_j,\cdot} z_j - \lambda(s_j^*)_{\Lambda_j}\right)^\top \Sigma_{\Lambda_j,\Lambda_j}^{-1}(A_{\Lambda_j,\cdot} z_j - \lambda(s_j^*)_{\Lambda_j})$$

when $r = \|Z_{\mathcal{C}} - Z_{\mathcal{C}^*}\|_F \leq 0.3$

$$\bar{\Phi}_\lambda(\mathcal{C}) - \bar{\Phi}_\lambda(\mathcal{C}^*) \geq \frac{a_0 r^2}{4}(1 - 10\alpha^{-1} r^2) - \lambda\sqrt{Ks}\|V^*\|_F r$$

With probability at least $1 - N^{-\beta}$

$$
\begin{aligned}
|\Phi_\lambda(\mathcal{C}) - & \bar{\Phi}_\lambda(\mathcal{C}) - \Phi_\lambda(\mathcal{C}^*) + \bar{\Phi}_\lambda(\mathcal{C}^*)| \\
& \lesssim \sqrt{\frac{sK \log N}{Tp_{\min}^2}} r + \sqrt{\frac{sn^* \log N}{Tp_{\min}^2}} r^2
\end{aligned}
$$

$$\begin{aligned}\Phi_\lambda(\mathcal{C}) - \Phi_\lambda(\mathcal{C}^*) \geq & \bar{\Phi}_\lambda(\mathcal{C}) - \bar{\Phi}_\lambda(\mathcal{C}^*) \\ & - |\Phi_\lambda(\mathcal{C}) - \bar{\Phi}_\lambda(\mathcal{C}) - \Phi_\lambda(\mathcal{C}^*) + \bar{\Phi}_\lambda(\mathcal{C}^*)| \\ \geq & \left(c - C\sqrt{\frac{sn^* \log N}{Tp_{\min}^2}} \right) r^2 - C\lambda\sqrt{s}Kr\end{aligned}$$

Hence $\Phi_\lambda(\hat{\mathcal{C}}) \leq \Phi_\lambda(\mathcal{C}^*)$ yields

$$r \leq \lambda\sqrt{s}K$$

Back