Exploring Dynamic Complex Systems Using Time-Varying Networks

Mladen Kolar

The University of Chicago
Booth School of Business

January 29, 2014
Networks are mathematical abstractions of complex systems

Networks are useful for

- visualization
- discovery of regularity patterns
- exploratory analysis
- ...

of complex systems.
Interactions between variables are not always observable
Interactions between variables are not always observable.

Data collected over a period of time is easily accessible.

Drosophila Life Cycle

<table>
<thead>
<tr>
<th>t=1</th>
<th>2</th>
<th>3</th>
<th>...</th>
<th>T</th>
</tr>
</thead>
</table>

M. Kolar (Chicago Booth)
Estimating time-varying networks

M. Kolar (Chicago Booth)
Drosophila Life Cycle

Data from Arbeitman et al. (2002)

66 microarray measurements across full life cycle

Four stages in the life cycle
- embryo
- larva
- pupal
- adult

Analyze subset of 588 genes related to development
Estimated Dynamic Network

biological process

molecular function

cellular component
Transient Group Interactions

(b) $t = 1$  (c) $t = 4$  (d) $t = 8$  (e) $t = 12$  (f) $t = 16$

(g) $t = 20$  (h) $t = 24$  (i) $t = 28$  (j) $t = 32$  (k) $t = 35$

(l) $t = 38$  (m) $t = 41$  (n) $t = 44$  (o) $t = 47$  (p) $t = 50$

(q) $t = 53$  (r) $t = 56$  (s) $t = 59$  (t) $t = 62$  (u) $t = 65$
Known Gene Interactions

M. Kolar (Chicago Booth)

Time-Varying Networks

January 29, 2014 12
Voting records from 109th congress (2005 - 2006)
(Banerjee, El Ghaoui, and d’Aspremont (2008))

There are 100 senators whose votes were recorded on the 542 bills, each vote is a binary outcome

Kolar, Song, Ahmed, and Xing (2010)
Senate Network  109th Congress

March 2005  January 2006  August 2006
How to recover *changing* interactions between objects from data collected over time?
Talk Objective

How to recover changing interactions between objects from data collected over time?

Challenges:
- Number of samples small
- Large number of objects
- Noisy data
- Data may contain missing values
- ...
Outline

1. Estimating Conditional Independence Relationships
   - Representation – Markov Networks
   - Estimating Graph Structure

2. Time-Varying Networks
   - Smoothly Varying Networks
   - Networks With Jumps

3. Other Challenges In Network Estimation
   - Dynamic Directed Networks
   - Missing Data
   - Multi-attribute Data
   - Conditional Networks

4. Future Directions
Outline

1 Estimating Conditional Independence Relationships
   - Representation – Markov Networks
   - Estimating Graph Structure

2 Time-Varying Networks
   - Smoothly Varying Networks
   - Networks With Jumps
Outline

1. Estimating Conditional Independence Relationships
   - Representation – Markov Networks
   - Estimating Graph Structure

2. Time-Varying Networks
   - Smoothly Varying Networks
   - Networks With Jumps

3. Other Challenges In Network Estimation
   - Dynamic Directed Networks
   - Missing Data
   - Multi-attribute Data
   - Conditional Networks
Outline

1. Estimating Conditional Independence Relationships
   - Representation – Markov Networks
   - Estimating Graph Structure

2. Time-Varying Networks
   - Smoothly Varying Networks
   - Networks With Jumps

3. Other Challenges In Network Estimation
   - Dynamic Directed Networks
   - Missing Data
   - Multi-attribute Data
   - Conditional Networks

4. Future Directions
Outline

1. Estimating Conditional Independence Relationships
   - Representation – Markov Networks
   - Estimating Graph Structure

2. Time-Varying Networks
   - Smoothly Varying Networks
   - Networks With Jumps

3. Other Challenges In Network Estimation
   - Dynamic Directed Networks
   - Missing Data
   - Multi-attribute Data
   - Conditional Networks

4. Future Directions
Markov Networks

Random vector $\mathbf{X} = (X_1, \ldots, X_p)'$

Graph $G = (V, E)$ with $p$ nodes
- represents conditional independence relationships between nodes

Useful for exploring associations between measured variables

$$(a, b) \not\in E \iff X_a \perp X_b \mid X_{\overline{ab}} \quad (\overline{ab} := V \setminus \{a, b\})$$

$$\mathbb{P}[X_a \mid X_b, X_{\overline{ab}}] = \mathbb{P}[X_a \mid X_{\overline{ab}}]$$

(Koller and Friedman, 2009)
Two Common Markov Networks

Gaussian Markov Network: \( \mathbf{X} \sim \mathcal{N}(\mu, \Sigma) \)

\[
p(\mathbf{x}) \propto \exp \left( -\frac{1}{2} (\mathbf{x} - \mu)^T \Sigma^{-1} (\mathbf{x} - \mu) \right)
\]

The precision matrix \( \Omega = \Sigma^{-1} \) encodes both parameters and the graph structure

\[
\begin{pmatrix}
* & * & * & * & * & 0 \\
* & * & * & * & * & 0 \\
* & * & * & 0 & 0 & 0 \\
* & * & 0 & * & 0 & 0 \\
* & * & 0 & 0 & * & 0 \\
0 & 0 & 0 & 0 & 0 & 0 \\
\end{pmatrix}
\]

(Koller and Friedman, 2009; Lauritzen, 1996)
Two Common Markov Networks

Gaussian Markov Network: \( \mathbf{X} \sim \mathcal{N}(\mu, \Sigma) \)

\[
p(\mathbf{x}) \propto \exp \left( -\frac{1}{2} (\mathbf{x} - \mu)^T \Sigma^{-1} (\mathbf{x} - \mu) \right)
\]

The precision matrix \( \Omega = \Sigma^{-1} \) encodes both parameters and the graph structure.

Discrete Markov network: \( \mathbf{X} \in \{-1, 1\}^p \) (Ising model)

\[
p(\mathbf{x}; \Theta) \propto \exp \left( \sum_{a \in V} x_a \theta_{aa} + \sum_{a,b \in V \times V} x_a x_b \theta_{ab} \right)
\]

\( \Theta = (\theta_{ab})_{ab} \) encodes the conditional independence relationships.

(Koller and Friedman, 2009; Lauritzen, 1996)
Structure Learning Problem

Given an i.i.d. sample $\mathcal{D}_n = \{x_i\}_{i=1}^n$ from a distribution $\mathbb{P} \in \mathcal{P}$

Learn the set of conditional independence relationships

$$\hat{G} = \hat{G}(\mathcal{D}_n)$$
Structure Learning Problem

Given an \( i.i.d. \) sample \( \mathcal{D}_n = \{x_i\}_{i=1}^n \) from a distribution \( \mathbb{P} \in \mathcal{P} \),

Learn the set of conditional independence relationships

\[
\hat{G} = \hat{G}(\mathcal{D}_n)
\]

Gaussian Markov Networks (Drton and Perlman, 2007)

- Form the maximum likelihood estimator for the covariance matrix
- Test for zeros in the precision matrix
Structure Learning Problem

Given an i.i.d. sample $\mathcal{D}_n = \{x_i\}_{i=1}^n$ from a distribution $\mathbb{P} \in \mathcal{P}$

Learn the set of conditional independence relationships

$$\hat{G} = \hat{G}(\mathcal{D}_n)$$

Gaussian Markov Networks (Drton and Perlman, 2007)
- Form the maximum likelihood estimator for the covariance matrix
- Test for zeros in the precision matrix

Discrete Markov Networks (Chickering, 1996)
- Hard to learn structure, since the log partition function cannot be evaluated efficiently
Penalized Pseudo-Likelihood Estimation

- Neighborhood Selection
- Useful for learning the structure of Gaussian and discrete Markov Networks

\[ \hat{\theta}_a = \arg \max_{\theta_a \in \mathbb{R}^p} \sum_{i \in [n]} \gamma(\theta_a; x_i) - \lambda \| \theta_a \|_1 \]

Conditional likelihood: \( \gamma(\theta_a; x_i) = \log \mathbb{P}[x_{i,a} \mid x_{i,a}; \theta_a] \)

(Meinshausen and Bühlmann, 2006)
(Ravikumar, Wainwright, and Lafferty, 2009)
Neighborhood Selection

Local structure estimation

\[ \hat{\theta}_a = \arg \max_{\theta_a \in \mathbb{R}^p} \ell(\theta_a; D_n) - \lambda \|\theta_a\|_1 \]

Estimated neighborhood

\[ \hat{N}_a = \{ b \in V \mid \hat{\theta}_{ab} \neq 0 \} \]
Neighborhood Selection

Local structure estimation

\[ \hat{\theta}_a = \arg \max_{\theta_a \in \mathbb{R}^p} \ell(\theta_a; \mathcal{D}_n) - \lambda \| \theta_a \|_1 \]

Estimated neighborhood

\[ \hat{\mathcal{N}}_a = \{ b \in V \mid \hat{\theta}_{ab} \neq 0 \} \]
Neighborhood Selection

Local structure estimation

\[ \hat{\theta}_a = \arg \max_{\theta_a \in \mathbb{R}^p} \ell(\theta_a; D_n) - \lambda \| \theta_a \|_1 \]

Estimated neighborhood

\[ \hat{N}_a = \{ b \in V \mid \hat{\theta}_{ab} \neq 0 \} \]
Neighborhood Selection

Local structure estimation

\[ \hat{\theta}_a = \arg \max_{\theta_a \in \mathbb{R}^p} \ell(\theta_a; D_n) - \lambda \| \theta_a \|_1 \]

\[ \hat{\theta}_1 = \big( \ast \ast 0 \ast 0 000 \big) \]

Estimated neighborhood

\[ \hat{N}_a = \{ b \in V \mid \hat{\theta}_{ab} \neq 0 \} \]
Neighborhood Selection

Local structure estimation

\[ \hat{\theta}_a = \arg \max_{\theta_a \in \mathbb{R}^p} \ell(\theta_a; D_n) - \lambda \| \theta_a \|_1 \]

\[ \hat{\theta}_1 = \begin{pmatrix} * & * & 0 & * & 0 & 0 & 0 \end{pmatrix} \]

Estimated neighborhood

\[ \hat{N}_a = \{ b \in V \mid \hat{\theta}_{ab} \neq 0 \} \]

\[ \hat{N}_a = \{2, 3, 5\} \]
Neighborhood Selection

Local structure estimation

$$\hat{\theta}_a = \arg \max_{\theta_a \in \mathbb{R}^p} \ell(\theta_a; D_n) - \lambda \|\theta_a\|_1$$

$$\hat{\theta}_1 = ( \ast \ast 0 \ast 0 0 0 0 )$$

Estimated neighborhood

$$\hat{N}_a = \{ b \in V \mid \hat{\theta}_{ab} \neq 0 \}$$

$$\hat{N}_a = \{ 2, 3, 5 \}$$
Neighborhood Selection

Local structure estimation

$$\hat{\theta}_a = \arg \max_{\theta_a \in \mathbb{R}^p} \ell(\theta_a; D_n) - \lambda ||\theta_a||_1$$

Estimated neighborhood

$$\hat{N}_a = \{b \in V \mid \hat{\theta}_{ab} \neq 0\}$$
Neighborhood Selection

Local structure estimation

\[ \hat{\theta}_a = \arg \max_{\theta_a \in \mathbb{R}^p} \ell(\theta_a; D_n) - \lambda \|\theta_a\|_1 \]

Estimated neighborhood

\[ \hat{N}_a = \{ b \in V \mid \hat{\theta}_{ab} \neq 0 \} \]
Neighborhood Selection

Local structure estimation

\[ \hat{\theta}_a = \arg \max_{\theta_a \in \mathbb{R}^p} \ell(\theta_a; D_n) - \lambda \| \theta_a \|_1 \]

Estimated neighborhood

\[ \hat{N}_a = \{ b \in V \mid \hat{\theta}_{ab} \neq 0 \} \]
Properties of Neighborhood Selection

Graph structure can be recovered consistently
- provable guarantees in a high-dimensional setting
  - Meinshausen and Bühlmann (2006); Ravikumar, Wainwright, and Lafferty (2009)
  - Peng, Wang, Zhou, and Zhu (2009)

Fast estimation procedures
- efficient solvers for $\ell_1$ penalized problems
  - Beck and Teboulle (2009); Friedman, Hastie, and Tibshirani (2008)
1. Estimating Conditional Independence Relationships
   - Representation – Markov Networks
   - Estimating Graph Structure

2. Time-Varying Networks
   - Smoothly Varying Networks
   - Networks With Jumps

3. Other Challenges In Network Estimation
   - Dynamic Directed Networks
   - Missing Data
   - Multi-attribute Data
   - Conditional Networks

4. Future Directions
Outline

1. Estimating Conditional Independence Relationships
   - Representation – Markov Networks
   - Estimating Graph Structure

2. Time-Varying Networks
   - Smoothly Varying Networks
   - Networks With Jumps

3. Other Challenges In Network Estimation
   - Dynamic Directed Networks
   - Missing Data
   - Multi-attribute Data
   - Conditional Networks

4. Future Directions
Estimating Time-Varying Networks

[Images of network diagrams and biological illustrations]
\[ x^t \sim P(\theta^t; G^t) \]
Estimating Time-Varying Networks

\[ \mathbf{x}^t \sim \mathbb{P}(\mathbf{\theta}^t; G^t) \]

Nodal observations
Estimating Time-Varying Networks

\[ \mathbf{x}^t \sim \mathbb{P}(\theta^t; G^t) \]

Nodal observations

Varying Coefficients
Estimating Time-Varying Networks

\[ \mathbf{x}_t \sim \mathbb{P}(\boldsymbol{\theta}_t; G_t) \]

Nodal observations

Varying Coefficients

Varying Structure
Estimating Time-Varying Networks

\[ \mathbf{x}^t \sim P(\boldsymbol{\theta}^t; G^t) \]

Nodal observations
Varying Structure
Varying Coefficients

\[ E^t = \{(a, b) \in V \times V \mid \theta_{ab}^t \neq 0\} \]
General Estimation Framework

Data: \( D_n = \{ x_t \mid x_t \sim \mathbb{P}(\theta^t; G^t) \}_{t \in T_n}, \quad T_n = \{1/n, 2/n, \ldots, 1\} \)

\[
\arg \max \ell(D_n, \{\theta^t\}) - \text{pen}(\{\theta^t\})
\]

Loss: \( \ell(D_n, \{\theta^t\}) \)
- measures the fit of model to data

Penalty: \( \text{pen}(\{\theta^t\}) \)
- balances the complexity of model and the fit to data
- encodes structural assumptions about model class
Two scenarios
Two scenarios

1. Smooth Networks

![Image of smooth networks over time]

- Song et al., 2009
- Kolar et al., 2010
- Kolar and Xing, 2011
- Kolar and Xing, 2012c

2. Networks With Jumps

- Time-Varying Networks

![Image of networks with jumps over time]

- Smooth Change
- Kernel Reweighting

Time
Two scenarios

1. Smooth Networks

2. Networks With Jumps
Two scenarios

1. **Smooth Networks**
   - (Song et al., 2009)
   - (Kolar et al., 2010)
   - (Kolar and Xing, 2011)
   - (Kolar and Xing, 2012c)

2. **Networks With Jumps**
Smoothly Evolving Networks

\[
\gamma(\theta; x_t) = \log \mathbb{P}[x_{t,a} \mid x_{t,a}; \theta]
\]

\[
w_\tau(t) = \frac{K_h(t - \tau)}{\sum_{t \in T_n} K_h(t - \tau)}
\]

\[
\hat{\theta}_a(\tau) = \arg \max_\theta \sum_{t \in T} w_\tau(t) \gamma(\theta; x_t) - \lambda \|\theta\|_1
\]

Kolar, Song, Ahmed, and Xing (2010)
\[ \gamma(\theta; x_t) = \log P[ x_{t,a} | x_{t,\bar{a}}; \theta] \]

\[ w_\tau(t) = \frac{K_h(t - \tau)}{\sum_{t \in \mathcal{T}_n} K_h(t - \tau)} \]

\[ \hat{\theta}_a(\tau) = \arg \max_\theta \sum_{t \in \mathcal{T}} w_\tau(t) \gamma(\theta; x_t) - \lambda \| \theta \|_1 \]

Kolar, Song, Ahmed, and Xing (2010)
\[ \gamma(\theta; x_t) = \log \mathbb{P}[x_{t,a} | x_{t,a}; \theta] \]

\[ w_\tau(t) = \frac{K_h(t - \tau)}{\sum_{t \in \mathcal{T}_n} K_h(t - \tau)} \]

\[ \hat{\theta}_a(\tau) = \arg \max_\theta \sum_{t \in \mathcal{T}} w_\tau(t) \gamma(\theta; x_t) - \lambda \|\theta\|_1 \]

Kolar, Song, Ahmed, and Xing (2010)
\[
\gamma(\theta; x_t) = \log P[x_{t,a} | x_{t,a}; \theta]
\]

\[
w_\tau(t) = \frac{K_h(t - \tau)}{\sum_{t \in \mathcal{T}_n} K_h(t - \tau)}
\]

\[
\hat{\theta}_a(\tau) = \arg \max_\theta \sum_{t \in \mathcal{T}} w_\tau(t) \gamma(\theta; x_t) - \lambda \|\theta\|_1
\]
Theoretical Properties

**Theorem (Kolar and Xing (2012c))**

Under suitable technical assumptions the graph $G^\tau$ is recovered with exponentially high probability for any fixed point $\tau \in [0, 1]$.

Fisher information matrix:

$$Q^\tau_a := \mathbb{E}[\nabla^2 \log P_{\theta^\tau_a} [X_a | \mathbf{X}_{\bar{a}}]], a \in V, \tau \in [0, 1]$$

- bounded eigenvalues
- incoherence condition
Theoretical Properties

Theorem (Kolar and Xing (2012c))

*Under suitable technical assumptions the graph $G^\tau$ is recovered with exponentially high probability for any fixed point $\tau \in [0, 1]$.*

Fisher information matrix:

$$Q^\tau_a := \mathbb{E}[(\nabla^2 \log \mathbb{P}_{\theta_a^\tau}[X_a | X_{\bar{a}}]), a \in V, \tau \in [0, 1]$$

- bounded eigenvalues
- incoherence condition

Smoothness: $\Sigma^t = (\sigma^t_{ab})$ are smooth functions of time
Theoretical Properties

Theorem (Kolar and Xing (2012c))

Under suitable technical assumptions the graph $G^\tau$ is recovered with exponentially high probability for any fixed point $\tau \in [0, 1]$.

Fisher information matrix:

$$Q^\tau_a := \mathbb{E}[\nabla^2 \log P_{\theta^\tau_a}[X_a|X_{\bar{a}}]], a \in V, \tau \in [0, 1]$$

- bounded eigenvalues
- incoherence condition

Smoothness: $\Sigma^t = (\sigma^t_{ab})$ are smooth functions of time

Kernel satisfies regularity conditions
Theoretical Properties

Theorem (Kolar and Xing (2012c))

Under suitable technical assumptions the graph $G^\tau$ is recovered with exponentially high probability for any fixed point $\tau \in [0, 1]$. 

Parameters:

- $\lambda \asymp \sqrt{\log p n} / 3$
- $h \asymp n - 1 / 3$

Sparsity:

- $s \asymp \log p n^2 / 3 = o(1)$ (maximal node degree)

Signal strength:

- $\theta_{\min} = \min_{e \in E} |\theta_{\tau e}| = \Omega(\sqrt{s \log p n^{1/3}})$

$P[\text{graph not recovered}] = O(\exp(-Cs^{-3/4}n^{1/3} + C') \log p n)$

$n, p \to \infty \implies P[\text{graph not recovered}] \to 0$
Theoretical Properties

Theorem (Kolar and Xing (2012c))

Under suitable technical assumptions the graph $G^\tau$ is recovered with exponentially high probability for any fixed point $\tau \in [0, 1]$.

Parameters: $\lambda \asymp \frac{\sqrt{\log p}}{n^{1/3}}$, $h \asymp n^{-\frac{1}{3}}$
Theoretical Properties

Theorem (Kolar and Xing (2012c))

Under suitable technical assumptions the graph $G^\tau$ is recovered with exponentially high probability for any fixed point $\tau \in [0, 1]$.

Parameters: $\lambda \asymp \frac{\sqrt{\log p}}{n^{1/3}}$, $h \asymp n^{-\frac{1}{3}}$

Sparsity: $\frac{s^3 \log p}{n^{2/3}} = o(1)$ \hspace{5mm} ($s$ – maximal node degree)
Theoretical Properties

Theorem (Kolar and Xing (2012c))

*Under suitable technical assumptions the graph $G^\tau$ is recovered with exponentially high probability for any fixed point $\tau \in [0, 1]$.\*

Parameters: $\lambda \asymp \frac{\sqrt{\log p}}{n^{1/3}}$, $h \asymp n^{-\frac{1}{3}}$

Sparsity: $\frac{s^3 \log p}{n^{2/3}} = o(1)$ \hspace{1cm} ($s$ – maximal node degree)

Signal strength: $\theta_{\text{min}} = \min_{e \in E^\tau} |\theta_e^\tau| = \Omega \left( \frac{\sqrt{s \log p}}{n^{1/3}} \right)$
Theoretical Properties

Theorem (Kolar and Xing (2012c))

Under suitable technical assumptions the graph $G^\tau$ is recovered with exponentially high probability for any fixed point $\tau \in [0, 1]$.

Parameters: $\lambda \approx \frac{\sqrt{\log p}}{n^{1/3}}$, $h \approx n^{-\frac{1}{3}}$

Sparsity: $\frac{s^3 \log p}{n^{2/3}} = o(1)$ (s – maximal node degree)

Signal strength: $\theta_{\text{min}} = \min_{e \in E^\tau} |\theta^\tau_e| = \Omega \left( \frac{\sqrt{s \log p}}{n^{1/3}} \right)$

$\mathbb{P} \left[ \text{graph not recovered} \right] = \mathcal{O} \left( \exp \left( -Cs^{-3}nh + C' \log p \right) \right) \xrightarrow{n,p \to \infty} 0$
Simulation Results

Chain Graph

Scaled sample size $n/(s^{4.5} \log^{1.5}(p))$

Hamming distance

$p = 60$

$p = 100$

$p = 140$
Two scenarios

1. Smooth Networks

2. Networks With Jumps

Structure Variation: $\Delta_t = |\beta_{t+1} - \beta_t|$
Two scenarios

1. Smooth Networks

2. Networks With Jumps
   (Kolar et al., 2010)
   (Kolar, Song, and Xing, 2009)
   (Kolar and Xing, 2012a)
Networks With Jumps

\[
\max_{\{\theta^t\}_{t \in T_n}} \sum_t \gamma(\theta^t; x^t) - \lambda_1 \sum_t ||\theta^t||_1 - \lambda_2 \sum_t ||\theta^t - \theta^{t-1}||_2
\]

Kolar, Song, Ahmed, and Xing (2010)
Networks With Jumps

\[
\max_{\{\theta^t\}_{t \in T_n}} \sum_t \gamma(\theta^t; x^t) - \lambda_1 \sum_t \|\theta^t\|_1 - \lambda_2 \sum_t \|\theta^t - \theta^{t-1}\|_2
\]

Kolar, Song, Ahmed, and Xing (2010)
Networks With Jumps

\[
\max_{\{\theta^t\}_{t \in \mathcal{T}_n}} \sum_t \gamma(\theta^t; x^t) - \lambda_1 \sum_t ||\theta^t||_1 - \lambda_2 \sum_t ||\theta^t - \theta^{t-1}||_2
\]

Fused Penalty (Tibshirani et al., 2005)

Kolar, Song, Ahmed, and Xing (2010)
Networks With Jumps

\[
\max_{\{\theta^t\}_{t \in T_n}} \sum_t \gamma(\theta^t; x^t) - \lambda_1 \sum_t ||\theta^t||_1 - \lambda_2 \sum_t ||\theta^t - \theta^{t-1}||_2
\]

Fused Penalty (Tibshirani et al., 2005)

Kolar, Song, Ahmed, and Xing (2010)
Model

Partition of $1, \ldots, n$: $\{B^j\}, \ j = 1, \ldots, B$ \hspace{1cm} (B fixed but unknown)

Partition boundaries: $\mathcal{T} := \{T_0 = 1 < T_1 < \ldots < T_B = n + 1\}$

Data: $x_i \sim \mathcal{N}_p(0, \Sigma^j)$, $i \in B^j$
Theoretical Properties

Theorem (Kolar and Xing (2012a))

Under suitable technical assumptions the partition boundaries are consistently estimated,

\[ \mathbb{P} \left[ \max_{j \in [B]} |T_j - \hat{T}_j| \leq n\delta_n \right] \xrightarrow{n \to \infty} 1, \]

where \( \delta_n = \mathcal{O}\left((\log n)^{\gamma}/n\right) \) and \( \gamma = 1 + \epsilon \).
Theoretical Properties

Theorem (Kolar and Xing (2012a))

Under suitable technical assumptions the partition boundaries are consistently estimated,

\[ \mathbb{P} \left[ \max_{j \in [B]} |T_j - \hat{T}_j| \leq n\delta_n \right] \xrightarrow{n \to \infty} 1, \]

where \( \delta_n = O((\log n)^\gamma / n) \) and \( \gamma = 1 + \epsilon \).

Parameters: \( \lambda_1 \asymp \lambda_2 \asymp \sqrt{\log(n)/n} \)
Theoretical Properties

Theorem (Kolar and Xing (2012a))

Under suitable technical assumptions the partition boundaries are consistently estimated,

\[ \mathbb{P} \left[ \max_{j \in [B]} |T_j - \hat{T}_j| \leq n\delta_n \right] \xrightarrow{n \to \infty} 1, \]

where \( \delta_n = \mathcal{O}((\log n)^\gamma/n) \) and \( \gamma = 1 + \epsilon \).

Parameters: \( \lambda_1 \asymp \lambda_2 \asymp \sqrt{\log(n)/n} \)

Size of change: \( \xi_{\text{min}} := \min_{a \in V} \min_{j \in [B-1]} \| \theta^{a,j+1} - \theta^{a,j} \|_2 \)

\[ \xi_{\text{min}} = \Omega(\sqrt{\log n/(\log n)^\gamma}) \]
Theoretical Properties

Theorem (Kolar and Xing (2012a))

Under suitable technical assumptions the partition boundaries are consistently estimated,

\[ \mathbb{P} \left[ \max_{j \in [B]} |T_j - \hat{T}_j| \leq n\delta_n \right] \xrightarrow{n \to \infty} 1, \]

where \( \delta_n = \mathcal{O}((\log n)^{\gamma}/n) \) and \( \gamma > 1 \).
Theoretical Properties

**Theorem (Kolar and Xing (2012a))**

*Under suitable technical assumptions the partition boundaries are consistently estimated,*

\[
P \left[ \max_{j \in [B]} |T_j - \hat{T}_j| \leq n\delta_n \right] \xrightarrow{n \to \infty} 1,
\]

where \(\delta_n = O((\log n)^{\gamma}/n)\) and \(\gamma > 1\).

---

The graph structure is consistently estimated at \(\Theta(n)\) time points.
Summary

Introduced models for time-varying networks
  - semi-parametric extension to Gaussian and Ising models
  - varying-coefficient models

Estimation procedures based on penalized pseudo-likelihood maximization
  - ideas borrowed from non-parametric regression
  - fast numerical procedures

Network structure consistently recovered
Outline

1. Estimating Conditional Independence Relationships
   - Representation – Markov Networks
   - Estimating Graph Structure

2. Time-Varying Networks
   - Smoothly Varying Networks
   - Networks With Jumps

3. Other Challenges In Network Estimation
   - Dynamic Directed Networks
   - Missing Data
   - Multi-attribute Data
   - Conditional Networks

4. Future Directions
Outline

1. Estimating Conditional Independence Relationships
   - Representation – Markov Networks
   - Estimating Graph Structure

2. Time-Varying Networks
   - Smoothly Varying Networks
   - Networks With Jumps

3. Other Challenges In Network Estimation
   - Dynamic Directed Networks
   - Missing Data
   - Multi-attribute Data
   - Conditional Networks

4. Future Directions
Dynamic Directed Networks naturally represent temporal processes

\[ X_t = A \cdot X_{t-1} + \epsilon, \quad \epsilon \sim N(0, \sigma^2 I) \]

Song, Kolar, and Xing (2009)
Dynamic Directed Networks naturally represent temporal processes

Vector Autoregressive Model

\[ X^t = A \cdot X^{t-1} + \epsilon, \quad \epsilon \sim \mathcal{N}(0, \sigma^2 I) \]

Song, Kolar, and Xing (2009)
Time Varying Dynamic Directed Networks

Dynamic Directed Networks naturally represent temporal processes

\[ X_t = A_t \cdot X_{t-1} + \epsilon, \quad \epsilon \sim \mathcal{N}(0, \sigma^2 I) \]

Song, Kolar, and Xing (2009)
Handling Data With Missing Values

Missing completely at random:

\[ P[R | X, \phi] = P[R | \phi] \text{ for all } X \text{ and } \phi \]

- Consistently recovers the graph structure
- Faster than the expectation-maximization procedure

Kolar and Xing (2012b)
Handling Data With Missing Values

Missing completely at random: $P[R \mid X, \varphi] = P[R \mid \varphi]$ for all $X$ and $\varphi$

Kolar and Xing (2012b)
Handling Data With Missing Values

Missing completely at random: \( P[R \mid X, \varphi] = P[R \mid \varphi] \) for all \( X \) and \( \varphi \)

- Consistently recovers the graph structure
- Faster than the expectation-maximization procedure

Kolar and Xing (2012b)
Handling Data With Multiple Attributes

[Diagram with nodes and connections]

M. Kolar (Chicago Booth)  Time-Varying Networks  January 29, 2014
Handling Data With Multiple Attributes

Attribute 1

Attribute 2
Handling Data With Multiple Attributes

Attribute 1

Attribute 2
Handling Data With Multiple Attributes

Attribute 1

Attribute 2
Handling Data With Multiple Attributes
Handling Data With Multiple Attributes

\[
\min_{\Omega > 0} \left\{ \text{tr } \mathbf{S}\Omega - \log |\Omega| + \lambda \sum_{a,b} ||\Omega_{ab}||_F \right\}
\]
Handling Data With Multiple Attributes

\[
\min_{\Omega > 0} \left\{ \text{tr} \mathbf{S} \Omega - \log |\Omega| + \lambda \sum_{a,b} ||\Omega_{ab}||_F \right\}
\]

Current iterate \( \Omega = \left( \begin{array}{cc} \Omega_{aa} & \Omega_{a,\bar{a}} \\ \Omega_{\bar{a},a} & \Omega_{\bar{a},\bar{a}} \end{array} \right) \)

Next iterate \( \hat{\Omega} = \tilde{\Omega} + \left( \begin{array}{cc} \Delta_{aa} & \Delta_{a,\bar{a}} \\ \Delta_{\bar{a},a} & 0 \end{array} \right) = \left( \begin{array}{cc} \hat{\Omega}_{aa} & \hat{\Omega}_{a,\bar{a}} \\ \hat{\Omega}_{\bar{a},a} & \hat{\Omega}_{\bar{a},\bar{a}} \end{array} \right) \)

\[
\hat{\Omega}_{aa} = \argmin_{\Omega_{aa}} \left\{ \text{tr}(\mathbf{S}_{aa} - \tilde{\Sigma}_{aa}) \Omega_{aa} + \frac{1}{2t} ||\Omega_{aa} - \tilde{\Omega}_{aa}||_F^2 + \lambda ||\Omega_{aa}||_F \right\}
\]

\[
\hat{\Omega}_{ab} = \argmin_{\Omega_{ab}} \left\{ \text{tr}(\mathbf{S}_{ab} - \tilde{\Sigma}_{ab}) \Omega_{ba} + \frac{1}{2t} ||\Omega_{ab} - \tilde{\Omega}_{ab}||_F^2 + \lambda ||\Omega_{ab}||_F \right\}
\]
Handling Data With Multiple Attributes

Data: \( \{X_i\}_{i \in [n]} \)
- \( X_i \) independent copy of a random vector \( X = (X_1^T, \ldots, X_p^T)^T \)
- \( X_a \in \mathbb{R}^{k_a} \) for each \( a \in \{1, \ldots, p\} \)

Goal: Learn a network \( G = (V, E) \) that represents conditional independence assumptions between nodes

\[
\rho_c(X_a, X_b; X_{ab}) = \max_{u \in \mathbb{R}^{k_a}, v \in \mathbb{R}^{k_b}} \text{Corr}(u'(X_a - \hat{A}X_{ab}), v'(X_b - \hat{B}X_{ab}))
\]

\[
\rho_c(X_a, X_b; X_{ab}) \neq 0 \iff \max_{u \in \mathbb{R}^{k_a}, v \in \mathbb{R}^{k_b}} u'\Omega_{ab}^* v \neq 0
\]
Estimating Conditional Networks

Kolar, Parikh, and Xing (2010)

low oil price

high oil price

M. Kolar (Chicago Booth)
Outline

1. Estimating Conditional Independence Relationships
   - Representation – Markov Networks
   - Estimating Graph Structure

2. Time-Varying Networks
   - Smoothly Varying Networks
   - Networks With Jumps

3. Other Challenges In Network Estimation
   - Dynamic Directed Networks
   - Missing Data
   - Multi-attribute Data
   - Conditional Networks

4. Future Directions
Outline

1. Estimating Conditional Independence Relationships
   - Representation – Markov Networks
   - Estimating Graph Structure

2. Time-Varying Networks
   - Smoothly Varying Networks
   - Networks With Jumps

3. Other Challenges In Network Estimation
   - Dynamic Directed Networks
   - Missing Data
   - Multi-attribute Data
   - Conditional Networks

4. Future Directions
Future Directions

- Network estimation under assumptions more relevant to real world problems
  - Non-parametric models
  - Confidence intervals for estimated networks
    - Needed for many high-dimensional estimation procedures
- Automatic discovery of interesting patterns
- Networks for Internet scale data
  - Efficient estimation procedures
  - Mixed type of data
Future Directions

Network estimation under assumptions more relevant to real world problems
- Non-parametric models
Future Directions

Network estimation under assumptions more relevant to real world problems
- Non-parametric models

Confidence intervals for estimated networks
- Needed for many high-dimensional estimation procedures
Future Directions

Network estimation under assumptions more relevant to real world problems
- Non-parametric models

Confidence intervals for estimated networks
- Needed for many high-dimensional estimation procedures

Automatic discovery of interesting patterns
Future Directions

Network estimation under assumptions more relevant to real world problems
  - Non-parametric models

Confidence intervals for estimated networks
  - Needed for many high-dimensional estimation procedures

Automatic discovery of interesting patterns

Networks for Internet scale data
  - Efficient estimation procedures
  - Mixed type of data
Thank you!


Kolar, Mladen and Hai Liu (2012a). “Optimal ROAD For Feature Selection in High-Dimensional Classification”. In: *Submitted*.


References IV


Simulation Results Smooth

Chain Graph

Nearest Neighbor Graph

Scaled sample size $n/(s^{4.5} \log^{1.5}(p))$

Hamming distance

$p = 60$

$p = 100$

$p = 140$
Penalized Maximum Likelihood Estimation

An i.i.d. sample $D_n = \{x_i\}_{i \in [n]}$ from the multivariate normal distribution $\mathcal{N}(0, \Sigma)$

Sample covariance matrix $S = n^{-1} \sum_{i \in [n]} x_i^T x_i$

$$\hat{\Omega} = \arg \max_{\Omega \succ 0} \log |\Omega| - \text{tr} S \Omega - \lambda ||\Omega||_1$$
Penalized Pseudo-Likelihood Estimation

Conditional likelihood:

\[ \gamma(\theta_a; x_i) = \log \mathbb{P}[x_{i,a} \mid x_{i,a}; \theta_a] \]

Loss:

\[ \ell(\theta_a; D_n) = \sum_{i \in [n]} \gamma(\theta_a; x_i) \]

Local structure estimation (neighborhood selection)

\[ \hat{\theta}_a = \arg \max_{\theta_a \in \mathbb{R}^p} \ell(\theta_a; D_n) - \lambda \| \theta_a \|_1 \]

Estimated neighborhood

\[ \hat{N}_a = \{ b \in V \mid \hat{\theta}_{ab} \neq 0 \} \]
Suppose $x_i \overset{iid}{\sim} \mathcal{N}(0, \Sigma)$ for all $i \in [n]$. 
Let $\Omega = \Sigma^{-1} = (\omega_{ab})_{ab}$.

**Lemma**

The following representation $x_{i,a} = \sum_{b \neq a} \theta_{ab} x_{i,b} + \epsilon_i$ holds with $\epsilon_i \perp x_{i,a}$ if and only if $\theta_{ab} = -\frac{\omega_{ab}}{\omega_{aa}}$.

**Local structure estimation**

$$\hat{\theta}_a = \arg \min_{\theta_a \in \mathbb{R}^p} \sum_{i \in [n]} (x_{i,a} - \theta_a^T x_{i,a})^2 + \lambda ||\theta_a||_1$$

Estimated neighborhood: $\hat{N}_a = \{b \in V \mid \hat{\theta}_{ab} \neq 0\}$
Neighborhood Selection: Discrete MRF

Suppose $x_i \sim \mathbb{P}_\Theta$ with density

$$p(x; \Theta) \propto \exp\left(\sum_{a \in V} x_a \theta_{aa} + \sum_{a, b \in V \times V} x_a x_b \theta_{ab}\right)$$

$$\gamma(\theta_a; x_i) = \log \mathbb{P}(x_{i,a} \mid x_{i,a}; \theta_a)$$

$$= x_{i,a} \langle \theta_a, x_{i,a} \rangle - \log (\exp(\langle \theta_a, x_{i,a} \rangle) + \exp(-\langle \theta_a, x_{i,a} \rangle))$$

Local structure estimation

$$\hat{\theta}_a = \arg \max_{\theta_a \in \mathbb{R}^p} \sum_{i \in [n]} \gamma(\theta_a; x_i) - \lambda \|\theta_a\|_1$$

- penalized sparse logistic regression
Dynamic Bayesian Networks naturally represent temporal processes:

\[
p(X^1, \ldots, X^T) = p(X^1) \prod_{t=2}^{T} p(X^t | X^{t-1})
\]

\[
= p(X^1) \prod_{t=2}^{T} \prod_{i=1}^{p} p(X^t_i | X^{t-1}_{\pi_i})
\]

Enough to specify the transition model \( p(X^t | X^{t-1}) \)

\[
X^t = A \cdot X^{t-1} + \epsilon, \quad \epsilon \sim \mathcal{N}(0, \sigma^2 I)
\]
Time Varying Dynamic Bayesian Networks

Time varying extension to AR model

\[
X^t = A^t \cdot X^{t-1} + \epsilon, \quad \epsilon \sim \mathcal{N}(0, \sigma^2 I)
\]
Time Varying Dynamic Bayesian Networks

Time varying extension to AR model

\[ X^t = A^t \cdot X^{t-1} + \epsilon, \quad \epsilon \sim \mathcal{N}(0, \sigma^2 I) \]

Associated network

\[ \mathcal{E}^t = \{(i, j) \in \mathcal{V} \times \mathcal{V} \mid i \neq j, A_{ij}^t \neq 0\} \]
Time Varying Dynamic Bayesian Networks

Time varying extension to AR model

\[ X^t = A^t \cdot X^{t-1} + \epsilon, \quad \epsilon \sim \mathcal{N}(0, \sigma^2 I) \]

Associated network

\[ \mathcal{E}^t = \{(i, j) \in \mathcal{V} \times \mathcal{V} \mid i \neq j, A^t_{ij} \neq 0\} \]

Estimation procedure

\[ \hat{A}^T_{i.} = \arg\min_{A^T_{i.} \in \mathbb{R}^{1 \times p}} \frac{1}{T} \sum_{t=1}^{T} w^T(t)(x^t_i - A^T_{i.} \cdot x^{t-1})^2 + \lambda \| A^T_{i.} \|_1 \]
Handling Data With Missing Values

\[ P[R|X, \phi] = P[R|\phi] \text{ for all } X \text{ and } \phi \]

\[ \hat{\sigma}_{ab} = \sum_{i \in [n]} r_{ia} r_{ib} (x_{ia} - \hat{\mu}_a)(x_{ib} - \hat{\mu}_b) \]

\[ \hat{\Omega} = \text{arg max} \Omega \succ 0 \log |\Omega| - tr S\Omega - \lambda ||\Omega||_1 \]

Properties of the procedure
- Consistently recovers the graph structure
- Faster than the expectation-maximization procedure
Handling Data With Missing Values

Missing completely at random

\[ \mathbb{P}[\mathbf{R} \mid \mathbf{X}, \varphi] = \mathbb{P}[\mathbf{R} \mid \varphi] \quad \text{for all } \mathbf{X} \text{ and } \varphi \]
Handling Data With Missing Values

Missing completely at random

\[ \mathbb{P}[\mathbf{R} \mid \mathbf{X}, \varphi] = \mathbb{P}[\mathbf{R} \mid \varphi] \quad \text{for all } \mathbf{X} \text{ and } \varphi \]

\[ \hat{\sigma}_{ab} = \frac{\sum_{i \in [n]} r_{ia} r_{ib} (x_{ia} - \hat{\mu}_a) (x_{ib} - \hat{\mu}_b)}{\sum_{i \in [n]} r_{ia} r_{ib}} \]
Handling Data With Missing Values

Missing completely at random

\[ P[R \mid X, \varphi] = P[R \mid \varphi] \quad \text{for all } X \text{ and } \varphi \]

\[ \hat{\sigma}_{ab} = \frac{\sum_{i\in[n]} r_{ia} r_{ib} (x_{ia} - \hat{\mu}_a) (x_{ib} - \hat{\mu}_b)}{\sum_{i\in[n]} r_{ia} r_{ib}} \]

\[ \hat{\Omega} = \arg \max_{\Omega \succ 0} \log |\Omega| - \text{tr } S\Omega - \lambda \|\Omega\|_1 \]
Handling Data With Missing Values

Missing completely at random

\[ P[R \mid X, \varphi] = P[R \mid \varphi] \quad \text{for all } X \text{ and } \varphi \]

\[ \hat{\sigma}_{ab} = \frac{\sum_{i \in [n]} r_{ia} r_{ib} (x_{ia} - \hat{\mu}_a)(x_{ib} - \hat{\mu}_b)}{\sum_{i \in [n]} r_{ia} r_{ib}} \]

\[ \hat{\Omega} = \arg \max_{\Omega \succ 0} \log |\Omega| - \text{tr } S\Omega - \lambda \|\Omega\|_1 \]

Properties of the procedure

- Consistently recovers the graph structure
- Faster than the expectation-maximization procedure
Handling Data With Multiple Attributes

Data:

\[ X_i \in [n] \]

- \( X_i \) independent copy of a random vector

\[ X = (X_1^T, \ldots, X_p^T)^T \]

- \( X_a \in R^k_a \) for each \( a \in \{1, \ldots, p\} \)

Goal: Learn a network \( G = (V, E) \) that represents conditional independence assumptions between nodes

\[ \rho_c(X_a, X_b; X_{ab}) = \max_{u \in R^{ka}}, v \in R^{kb}} \text{Corr}(u'(X_a - \hat{A}X_{ab}), v'(X_b - \hat{B}X_{ab})) \]

\[ \hat{A} = \arg\min \mathbb{E}[||X_a - AX_{ab}||_2^2] \]

\[ \hat{B} = \arg\min \mathbb{E}[||X_b - BX_{ab}||_2^2] \]

Kolar, Liu, and Xing (2012)
Handling Data With Multiple Attributes

Data: \( \{X_i\}_{i \in [n]} \)
- \( X_i \) independent copy of a random vector \( X = (X_1^T, \ldots, X_p^T)^T \)
- \( X_a \in \mathbb{R}^{k_a} \) for each \( a \in \{1, \ldots, p\} \)
Handling Data With Multiple Attributes

Data: \( \{X_i\}_{i \in [n]} \)

- \( X_i \) independent copy of a random vector \( X = (X_1^T, \ldots, X_p^T)^T \)
- \( X_a \in \mathbb{R}^{k_a} \) for each \( a \in \{1, \ldots, p\} \)

Goal: Learn a network \( G = (V, E) \) that represents conditional independence assumptions between nodes
Handling Data With Multiple Attributes

Data: \( \{X_i\}_{i \in [n]} \)
- \( X_i \) independent copy of a random vector \( X = (X_1^T, \ldots, X_p^T)^T \)
- \( X_a \in \mathbb{R}^{k_a} \) for each \( a \in \{1, \ldots, p\} \)

Goal: Learn a network \( G = (V, E) \) that represents conditional independence assumptions between nodes

\[
\rho_c(X_a, X_b; X_{ab}) = \max_{u \in \mathbb{R}^{k_a}, v \in \mathbb{R}^{k_b}} \text{Corr}(u'(X_a - \hat{A}X_{ab}), v'(X_b - \hat{B}X_{ab}))
\]

\[
\hat{A} = \text{argmin} \ \mathbb{E}[||X_a - AX_{ab}||_2^2]
\]
\[
\hat{B} = \text{argmin} \ \mathbb{E}[||X_b - BX_{ab}||_2^2]
\]

Kolar, Liu, and Xing (2012)
Theoretical Properties

Theorem (Kolar and Xing (2012c))

Under suitable technical assumptions the graph $G^\tau$ is recovered with exponentially high probability for any fixed point $\tau \in [0, 1]$.

Fisher information matrix:

$$Q^\tau_a := \mathbb{E}[\nabla^2 \log \mathbb{P}_{\theta^\tau_a}[X_a|X_{\bar{a}}]], a \in V, \tau \in [0, 1]$$

- bounded eigenvalues
- incoherence condition: $\|Q^\tau_{NT}(Q^\tau_{TT})^{-1}\|_{\infty} \leq 1 - \alpha$
Theoretical Properties

Theorem (Kolar and Xing (2012c))

Under suitable technical assumptions the graph $G^\tau$ is recovered with exponentially high probability for any fixed point $\tau \in [0, 1]$.

Fisher information matrix:

$$Q^\tau_a := \mathbb{E}[\nabla^2 \log \mathbb{P}_{\theta^\tau_{a}} [X_a | X_{\bar{a}}]], \, a \in V, \, \tau \in [0, 1]$$

- bounded eigenvalues
- incoherence condition: $\|Q^\tau_{NT}(Q^\tau_{TT})^{-1}\|_{\infty} \leq 1 - \alpha$

Smoothness: $\Sigma^t = (\sigma^t_{ab})$ are smooth functions of time
Theoretical Properties

Theorem (Kolar and Xing (2012c))

Under suitable technical assumptions the graph $G^\tau$ is recovered with exponentially high probability for any fixed point $\tau \in [0, 1]$.

Fisher information matrix:

$$Q^\tau_a := \mathbb{E}[\nabla^2 \log \mathbb{P}_{\theta^\tau_a}[X_a|X_{\overline{a}}]], a \in V, \tau \in [0, 1]$$

- bounded eigenvalues
- incoherence condition: $\|Q_{NT}^{-1}(Q_{TT})^{-1}\|_\infty \leq 1 - \alpha$

Smoothness: $\Sigma^t = (\sigma^t_{ab})$ are smooth functions of time

Kernel satisfies regularity conditions
Simulation Results – Nearest Neighborhood

Precision

Recall

$F_1$
\[
\max_{\{\theta^t\}_{t \in T}} \ell(D_n; \{\theta^t\}_{t \in T}) - \lambda_1 \sum_{t} ||\theta^t||_1 - \lambda_2 \sum_{t} ||\theta^t - \theta^{t-1}||_2
\]
Optimization

\[
\max_{\{\theta^t\}_{t \in T}} \ell(D_n; \{\theta^t\}_{t \in T}) - \lambda_1 \sum_t ||\theta^t||_1 - \lambda_2 \sum_t ||\theta^t - \theta^{t-1}||_2
\]

Non-smooth penalty term
Optimization

\[
\max_{\{\theta^t\}_{t \in \mathcal{T}_n}} \ell(D_n; \{\theta^t\}_{t \in \mathcal{T}_n}) - \lambda_1 \sum_t ||\theta^t||_1 - \lambda_2 \sum_t ||\theta^t - \theta^{t-1}||_2
\]

Nesterov Smoothing Technique \text{(Nesterov, 2005)}

\[
\max_{\{\theta^t\}_{t \in \mathcal{T}_n}} \ell(D_n; \{\theta^t\}_{t \in \mathcal{T}_n}) - \lambda_1 \sum_t ||\theta^t||_1 - \lambda_2 \sum_t \Psi_\mu (\theta^t - \theta^{t-1})
\]

Non-smooth penalty term
Optimization

\[
\max_{\{\theta^t\}_{t \in T_n}} \ell(D_n; \{\theta^t\}_{t \in T_n}) - \lambda_1 \sum_t ||\theta^t||_1 - \lambda_2 \sum_t ||\theta^t - \theta^{t-1}||_2
\]

Nesterov Smoothing Technique \cite{Nesterov05}

\[
\max_{\{\theta^t\}_{t \in T_n}} \ell(D_n; \{\theta^t\}_{t \in T_n}) - \lambda_1 \sum_t ||\theta^t||_1 - \lambda_2 \sum_t \Psi_\mu (\theta^t - \theta^{t-1})
\]
Optimization

\[
\max_{\{\theta^t\}_{t \in T_n}} \ell(D_n; \{\theta^t\}_{t \in T_n}) - \lambda_1 \sum_t ||\theta^t||_1 - \lambda_2 \sum_t ||\theta^t - \theta^{t-1}||_2
\]

Nesterov Smoothing Technique (Nesterov, 2005)

\[
\max_{\{\theta^t\}_{t \in T_n}} \ell(D_n; \{\theta^t\}_{t \in T_n}) - \lambda_1 \sum_t ||\theta^t||_1 - \lambda_2 \sum_t \Psi_\mu (\theta^t - \theta^{t-1})
\]

Solved using accelerated gradient descent (Beck and Teboulle, 2009)

\[
\min_{\theta \in \mathbb{R}^{(p-1) \times n}} \frac{1}{2} \left\| \theta - \left( \theta_0 - \frac{1}{L} \left( \nabla \ell + \nabla \Psi_\mu \right)(\theta_0) \right) \right\|_F^2 + \frac{2\lambda_1}{L} \left\| \theta \right\|_1
\]
Distance between two sets: \( h(A, B) := \sup_{b \in B} \inf_{a \in A} |a - b| \)

Suppose we have an upper bound on the number of partitions
\( B_{\text{max}} > B \)

If \( B < \hat{B} < B_{\text{max}} \), then
\[
\mathbb{P}[h(\hat{T}, T) \leq n\delta_n] \xrightarrow{n \to \infty} 1.
\]
Changing Associations Between Stock Prices

low oil price \rightarrow high oil price
Analyzing the S&P 500

Examine associations among stocks

- Help an economist studying the market
- Assist an investor building a diverse portfolio

Stock prices from Jan 1, 2003 to Dec 31, 2005.

Condition on oil price, an economic indicator

Kolar, Parikh, and Xing (2010)
Edge weights (proportional to partial correlations) reflect changes in associations.