Estimating Time-Varying Networks

Mladen Kolar

The University of Chicago
Booth School of Business

November 20, 2013

Acknowledgments


E. Xing


L. Song


A. Ahmed

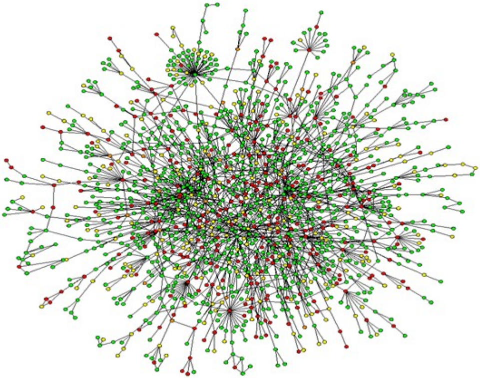

A. Parikh

Networks are mathematical abstractions of complex systems

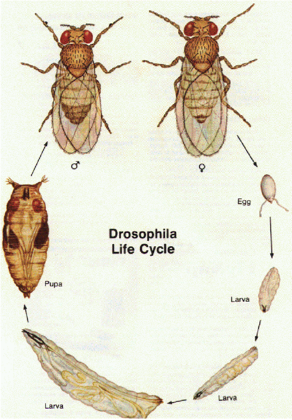
Networks are useful for

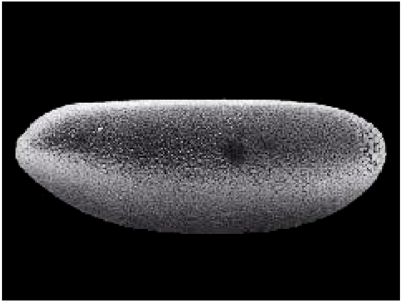
- visualization
- discovery of regularity patterns
- exploratory analysis
- ...

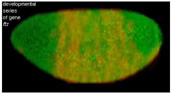
of complex systems.

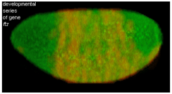


Interactions between variables are not always observable

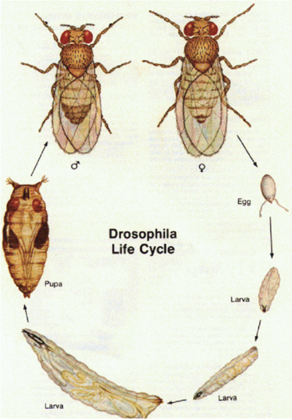




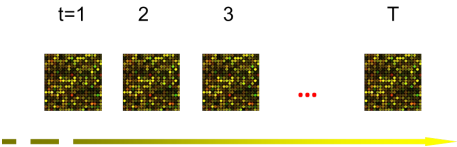




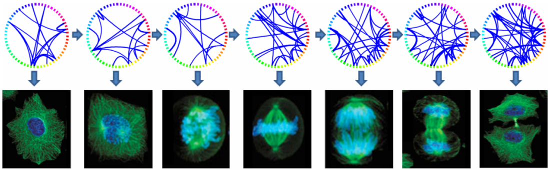
Interactions between variables are not always observable

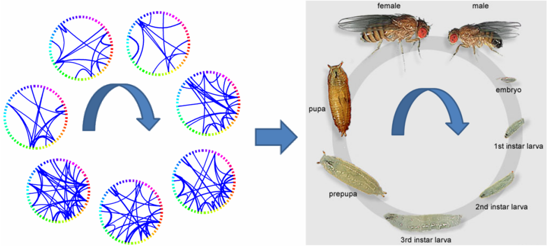


Data collected over a period of time is easily accessible



Estimating time-varying networks





Talk Objective

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

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. An Application

Outline

1. Estimating Conditional Independence Relationships
   - Representation – Markov Networks
   - Estimating Graph Structure
2. Time-Varying Networks
   - Smoothly Varying Networks
   - Networks With Jumps
3. An Application
4. Some theoretical results

Outline

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

$$(a,b) \notin E \Longleftrightarrow X_a \perp X_b \mid X_{\overline{ab}} \qquad \left(\overline{ab} := V \backslash \{a,b\} \right)$$

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

(Koller and Friedman, 2009)

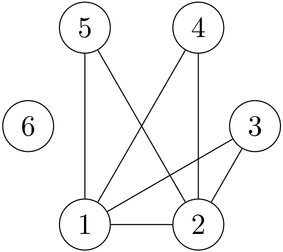
Two Common Markov Networks

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

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

The precision matrix $\boldsymbol{\Omega} = \boldsymbol{\Sigma}^{-1}$ 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: $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

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

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

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

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

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

(Koller and Friedman, 2009; Lauritzen, 1996)

Structure Learning Problem

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

Learn the set of conditional independence relationships

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

Structure Learning Problem

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

Learn the set of conditional independence relationships

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

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 $\mathcal{D}_n = \{\mathbf{x}_i\}_{i=1}^n$ from a distribution $\mathbb{P} \in \mathcal{P}$

Learn the set of conditional independence relationships

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

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

Structure Learning in High-Dimensions

Penalized Pseudo-Likelihood Estimation

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

$$\widehat{\boldsymbol{\theta}}_a = \arg\max_{\boldsymbol{\theta}_a \in \mathbb{R}^p} \sum_{i \in [n]} \gamma(\boldsymbol{\theta}_a; \mathbf{x}_i) - \lambda ||\boldsymbol{\theta}_a||_1$$

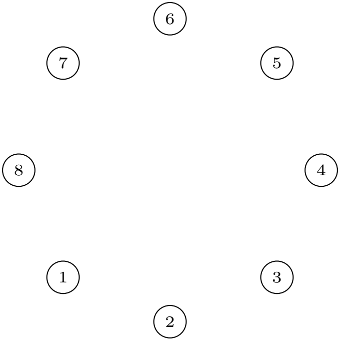
Conditional likelihood: $\gamma(\boldsymbol{\theta}_a; \mathbf{x}_i) = \log \mathbb{P}[x_{i,a} \mid \mathbf{x}_{i,\overline{a}}; \boldsymbol{\theta}_a]$

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

Neighborhood Selection

Local structure estimation

$$\widehat{\boldsymbol{\theta}}_a = \arg\max_{\boldsymbol{\theta}_a \in \mathbb{R}^p} \ell(\boldsymbol{\theta}_a; \mathcal{D}_n) - \lambda ||\boldsymbol{\theta}_a||_1$$



Estimated neighborhood

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

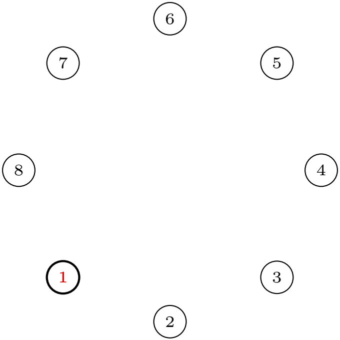
Neighborhood Selection

Local structure estimation

$$\widehat{\boldsymbol{\theta}}_a = \arg\max_{\boldsymbol{\theta}_a \in \mathbb{R}^p} \ell(\boldsymbol{\theta}_a; \mathcal{D}_n) - \lambda ||\boldsymbol{\theta}_a||_1$$

Estimated neighborhood

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



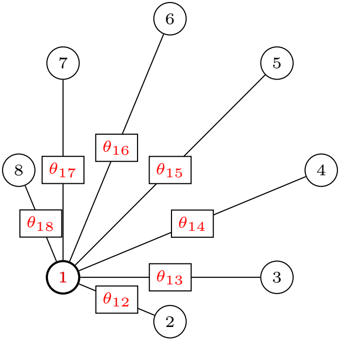
Neighborhood Selection

Local structure estimation

$$\widehat{\boldsymbol{\theta}}_a = \arg\max_{\boldsymbol{\theta}_a \in \mathbb{R}^p} \ell(\boldsymbol{\theta}_a; \mathcal{D}_n) - \lambda ||\boldsymbol{\theta}_a||_1$$

Estimated neighborhood

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



Neighborhood Selection

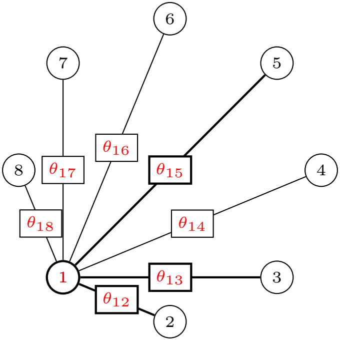
Local structure estimation

$$\widehat{\boldsymbol{\theta}}_a = \arg \max_{\boldsymbol{\theta}_a \in \mathbb{R}^p} \ell(\boldsymbol{\theta}_a; \mathcal{D}_n) - \lambda ||\boldsymbol{\theta}_a||_1$$

$$\widehat{\boldsymbol{\theta}}_1 = \left(\begin{array}{ccccccc} * & * & 0 & * & 0 & 0 & 0 \end{array} \right)$$

Estimated neighborhood

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



Neighborhood Selection

Local structure estimation

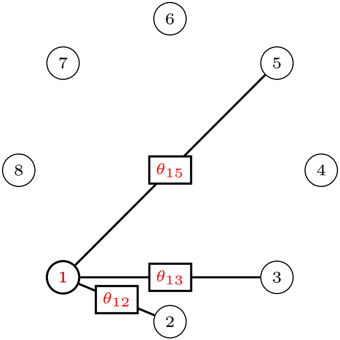
$$\widehat{\boldsymbol{\theta}}_a = \arg \max_{\boldsymbol{\theta}_a \in \mathbb{R}^p} \ell(\boldsymbol{\theta}_a; \mathcal{D}_n) - \lambda ||\boldsymbol{\theta}_a||_1$$

$$\widehat{\boldsymbol{\theta}}_1 = \begin{pmatrix} * & * & 0 & * & 0 & 0 & 0 \end{pmatrix}$$

Estimated neighborhood

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

$$\widehat{\mathbf{N}}_a = \{2, 3, 5\}$$



Neighborhood Selection

Local structure estimation

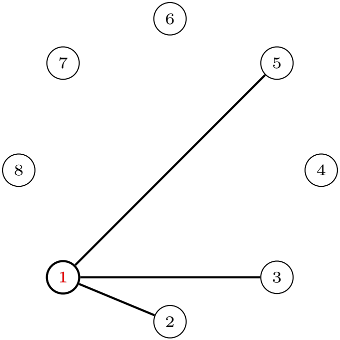
$$\widehat{\boldsymbol{\theta}}_a = \arg\max_{\boldsymbol{\theta}_a \in \mathbb{R}^p} \ell(\boldsymbol{\theta}_a; \mathcal{D}_n) - \lambda ||\boldsymbol{\theta}_a||_1$$

$$\widehat{\boldsymbol{\theta}}_1 = \begin{pmatrix} * & * & 0 & * & 0 & 0 & 0 \end{pmatrix}$$

Estimated neighborhood

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

$$\widehat{\mathbf{N}}_a = \{2, 3, 5\}$$



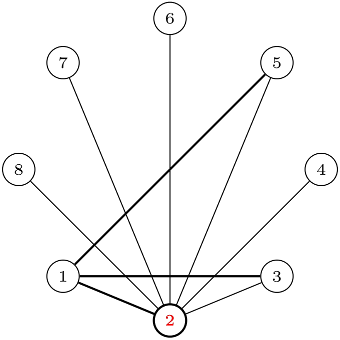
Neighborhood Selection

Local structure estimation

$$\widehat{\boldsymbol{\theta}}_a = \arg\max_{\boldsymbol{\theta}_a \in \mathbb{R}^p} \ell(\boldsymbol{\theta}_a; \mathcal{D}_n) - \lambda ||\boldsymbol{\theta}_a||_1$$

Estimated neighborhood

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



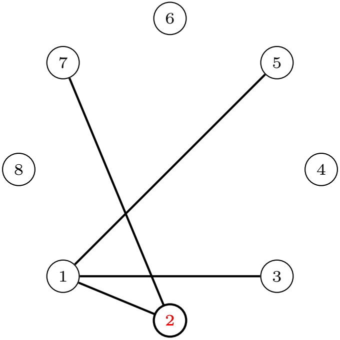
Neighborhood Selection

Local structure estimation

$$\widehat{\boldsymbol{\theta}}_a = \arg\max_{\boldsymbol{\theta}_a \in \mathbb{R}^p} \ell(\boldsymbol{\theta}_a; \mathcal{D}_n) - \lambda ||\boldsymbol{\theta}_a||_1$$

Estimated neighborhood

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



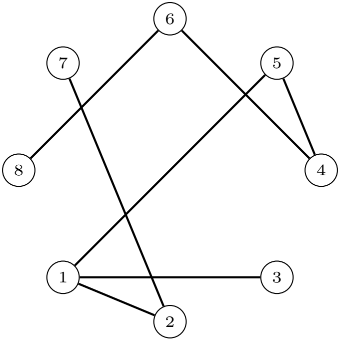
Neighborhood Selection

Local structure estimation

$$\widehat{\boldsymbol{\theta}}_a = \arg\max_{\boldsymbol{\theta}_a \in \mathbb{R}^p} \ell(\boldsymbol{\theta}_a; \mathcal{D}_n) - \lambda ||\boldsymbol{\theta}_a||_1$$

Estimated neighborhood

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



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 ℓ_1 penalized problems
- Beck and Teboulle (2009); Friedman, Hastie, and Tibshirani (2008)

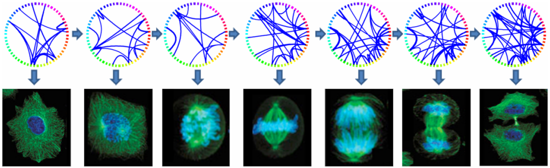
Outline

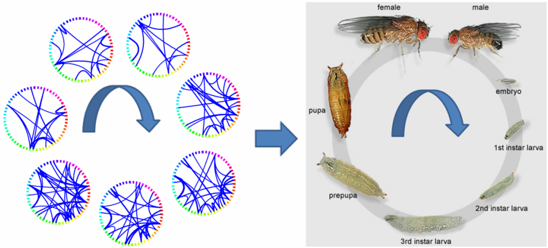
1. Estimating Conditional Independence Relationships
   - Representation – Markov Networks
   - Estimating Graph Structure
2. Time-Varying Networks
   - Smoothly Varying Networks
   - Networks With Jumps
3. An Application
4. Some theoretical results

Outline

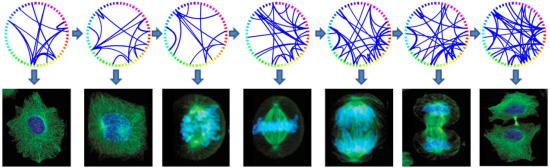
1. Estimating Conditional Independence Relationships
   - Representation – Markov Networks
   - Estimating Graph Structure
2. **Time-Varying Networks**
   - Smoothly Varying Networks
   - Networks With Jumps
3. An Application
4. Some theoretical results

Estimating Time-Varying Networks



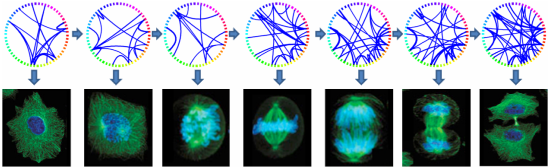


Estimating Time-Varying Networks



$$\mathbf{x}^t \sim \mathbb{P}(\boldsymbol{\theta}^t; G^t)$$

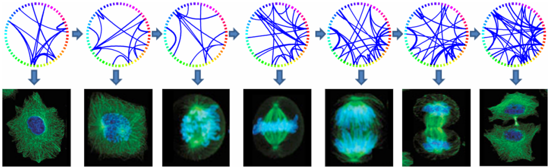
Estimating Time-Varying Networks



$$\underline{\mathbf{x}}^t \sim \mathbb{P}(\boldsymbol{\theta}^t; G^t)$$

Nodal observations

Estimating Time-Varying Networks

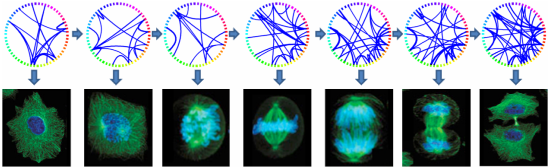


$$\underline{\mathbf{x}}^t \sim \mathbb{P}(\underline{\boldsymbol{\theta}}^t; G^t)$$

Nodal observations

Varying Coefficients

Estimating Time-Varying Networks

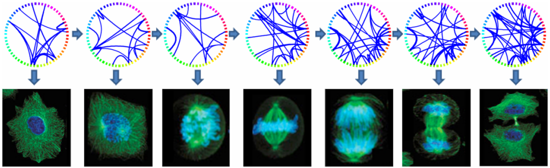


$$\underline{\mathbf{x}}^t \sim \mathbb{P}(\underline{\boldsymbol{\theta}}^t; \underline{G}^t)$$

Nodal observations

Varying Coefficients

Varying Structure



$$\underline{\mathbf{x}}^t \sim \mathbb{P}(\underline{\boldsymbol{\theta}}^t; \underline{G}^t)$$

Nodal observations

Varying Coefficients

Varying Structure

$$E^t = \{(a,b) \in V \times V \mid \theta^t_{ab} \neq 0\}$$

General Estimation Framework

Data: $\mathcal{D}_n = \{\mathbf{x}^t \mid \mathbf{x}^t \sim \mathbb{P}(\boldsymbol{\theta}^t; G^t)\}_{t \in \mathcal{T}_n}, \quad \mathcal{T}_n = \{1/n, 2/n, \ldots, 1\}$

$$\arg\max \; \ell(\mathcal{D}_n, \{\boldsymbol{\theta}^t\}) - \text{pen}\left(\{\boldsymbol{\theta}^t\}\right)$$

Loss: $\ell(\mathcal{D}_n, \{\boldsymbol{\theta}^t\})$

- measures the fit of model to data

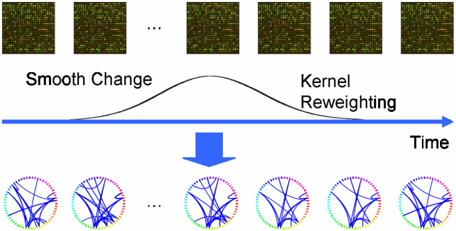
Penalty: $\text{pen}\left(\{\boldsymbol{\theta}^t\}\right)$

- balances the complexity of model and the fit to data
- encodes structural assumptions about model class

Two scenarios

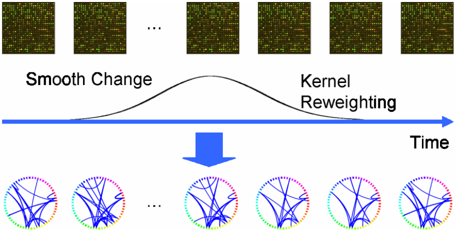
Two scenarios

1. Smooth Networks

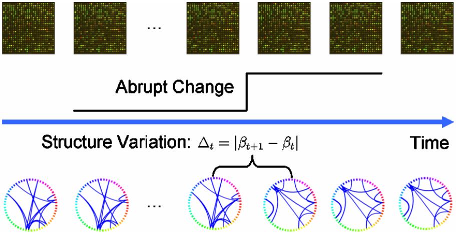


Two scenarios

1. Smooth Networks

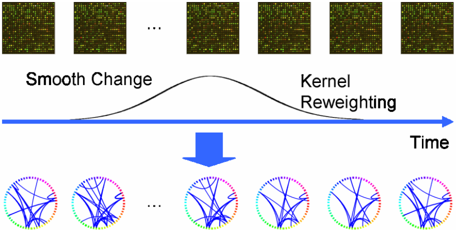


2. Networks With Jumps

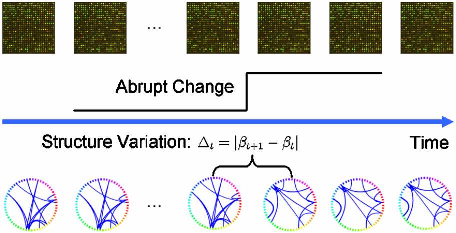


Two scenarios

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



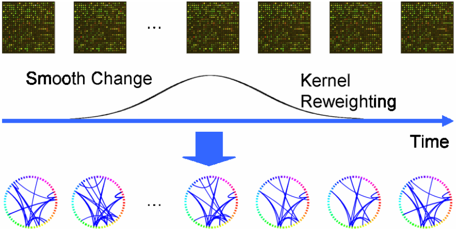
2. Networks With Jumps



Smoothly Evolving Networks

$$\gamma(\boldsymbol{\theta}; \mathbf{x}_t) = \log \mathbb{P}[x_{t,a} \mid \mathbf{x}_{t,\bar{a}}; \boldsymbol{\theta}]$$

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



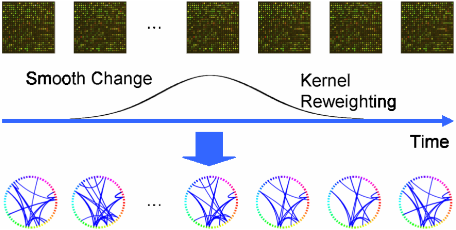
$$\widehat{\boldsymbol{\theta}}_a(\tau) = \arg\max_{\theta} \sum_{t \in \mathcal{T}} w_\tau(t)\gamma(\boldsymbol{\theta}; \mathbf{x}_t) - \lambda ||\boldsymbol{\theta}||_1$$

Kolar, Song, Ahmed, and Xing (2010)

Smoothly Evolving Networks

$$\gamma(\boldsymbol{\theta}; \mathbf{x}_t) = \log \mathbb{P}[x_{t,a} \mid \mathbf{x}_{t,\bar{a}}; \boldsymbol{\theta}]$$

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



$$\widehat{\boldsymbol{\theta}}_a(\tau) = \arg\max_{\theta} \sum_{t \in \mathcal{T}} w_\tau(t) \gamma(\boldsymbol{\theta}; \mathbf{x}_t) - \lambda ||\boldsymbol{\theta}||_1$$

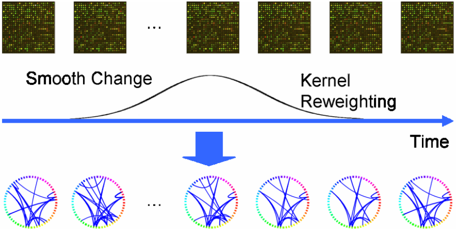
Smoothness

Kolar, Song, Ahmed, and Xing (2010)

Smoothly Evolving Networks

$$\gamma(\boldsymbol{\theta}; \mathbf{x}_t) = \log \mathbb{P}[x_{t,a} \mid \mathbf{x}_{t,\bar{a}}; \boldsymbol{\theta}]$$

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



$$\widehat{\boldsymbol{\theta}}_a(\tau) = \arg\max_{\theta} \sum_{t \in \mathcal{T}} w_\tau(t)\gamma(\boldsymbol{\theta}; \mathbf{x}_t) - \lambda ||\boldsymbol{\theta}||_1$$

Smoothness

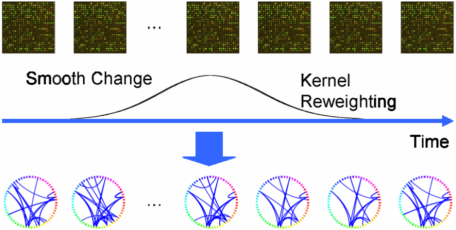
Sparsity

Kolar, Song, Ahmed, and Xing (2010)

Smoothly Evolving Networks

$$\gamma(\boldsymbol{\theta}; \mathbf{x}_t) = \log \mathbb{P}[x_{t,a} \mid \mathbf{x}_{t,\bar{a}}; \boldsymbol{\theta}]$$

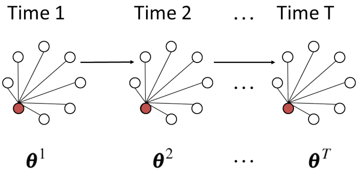
$$w_\tau(t) = \frac{K_h(t-\tau)}{\sum_{t \in \mathcal{T}_n} K_h(t-\tau)}$$



$$\widehat{\boldsymbol{\theta}}_a(\tau) = \arg\max_{\theta} \sum_{t \in \mathcal{T}} w_\tau(t)\gamma(\boldsymbol{\theta}; \mathbf{x}_t) - \lambda ||\boldsymbol{\theta}||_1$$

Smoothness

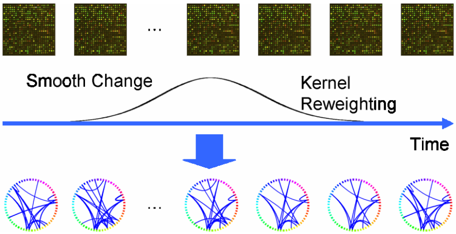
Sparsity



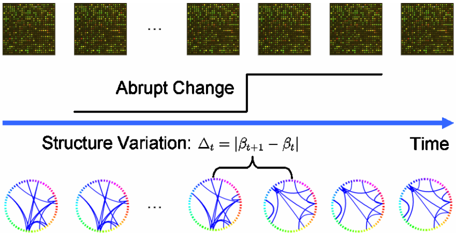
Kolar, Song, Ahmed, and Xing (2010)

Two scenarios

1. Smooth Networks

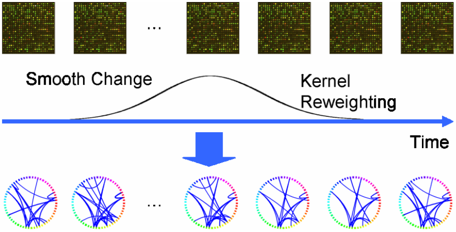


2. Networks With Jumps



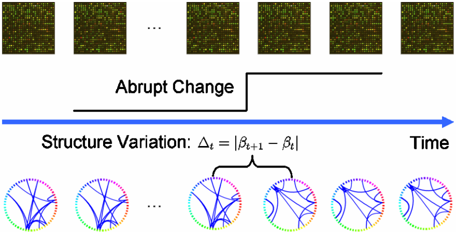
Two scenarios

1. Smooth Networks

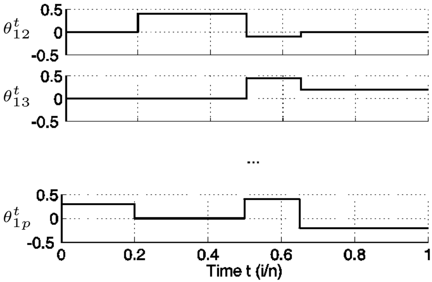


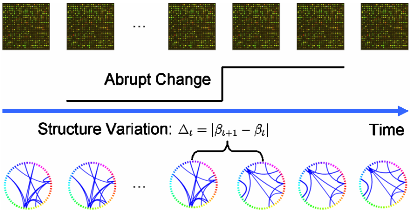
2. **Networks With Jumps**

   (Kolar et al., 2010)
   (Kolar, Song, and Xing, 2009)
   (Kolar and Xing, 2012a)



Networks With Jumps

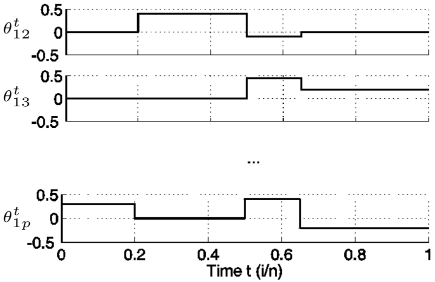


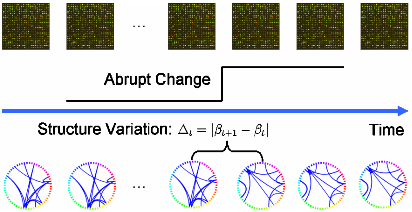


$$\max_{\{\boldsymbol{\theta}^t\}_{t \in \mathcal{T}_n}} \sum_t \gamma(\boldsymbol{\theta}^t; \mathbf{x}^t) - \lambda_1 \sum_t ||\boldsymbol{\theta}^t||_1$$
$$- \lambda_2 \sum_t ||\boldsymbol{\theta}^t - \boldsymbol{\theta}^{t-1}||_2$$

Kolar, Song, Ahmed, and Xing (2010)

Networks With Jumps



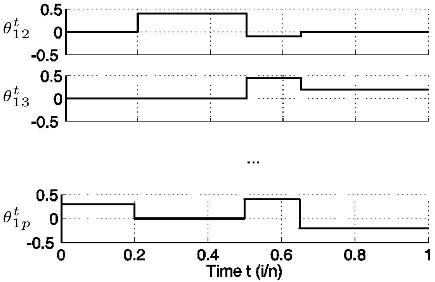


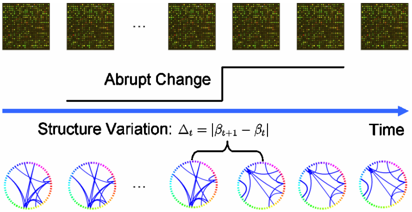
$$\max_{\{\boldsymbol{\theta}^t\}_{t \in \mathcal{T}_n}} \sum_t \gamma(\boldsymbol{\theta}^t; \mathbf{x}^t) - \lambda_1 \sum_t ||\boldsymbol{\theta}^t||_1 - \lambda_2 \sum_t ||\boldsymbol{\theta}^t - \boldsymbol{\theta}^{t-1}||_2$$

Sparsity

Kolar, Song, Ahmed, and Xing (2010)

Networks With Jumps





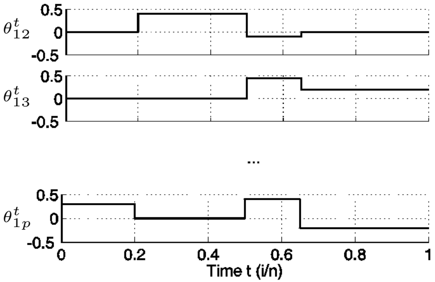
$$\max_{\{\boldsymbol{\theta}^t\}_{t \in \mathcal{T}_n}} \sum_t \gamma(\boldsymbol{\theta}^t; \mathbf{x}^t) - \lambda_1 \sum_t ||\boldsymbol{\theta}^t||_1 - \lambda_2 \sum_t ||\boldsymbol{\theta}^t - \boldsymbol{\theta}^{t-1}||_2$$

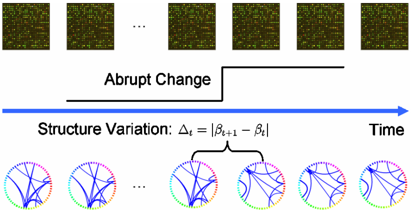
Sparsity

Structural Changes

Kolar, Song, Ahmed, and Xing (2010)

Networks With Jumps





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

Sparsity

Structural Changes

Fused Penalty (Tibshirani et al., 2005)

Kolar, Song, Ahmed, and Xing (2010)

Outline

1. Estimating Conditional Independence Relationships
   - Representation – Markov Networks
   - Estimating Graph Structure
2. Time-Varying Networks
   - Smoothly Varying Networks
   - Networks With Jumps
3. An Application
4. Some theoretical results

Outline

Drosophila Life Cycle

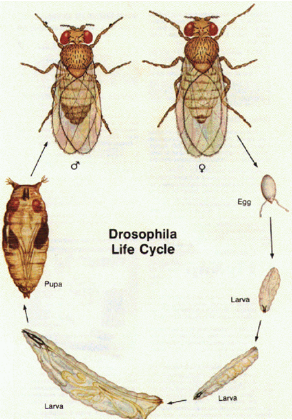
Data from Arbeitman et al. (2002)

66 microarray measurements across full life cycle

Four stages in the life cycle

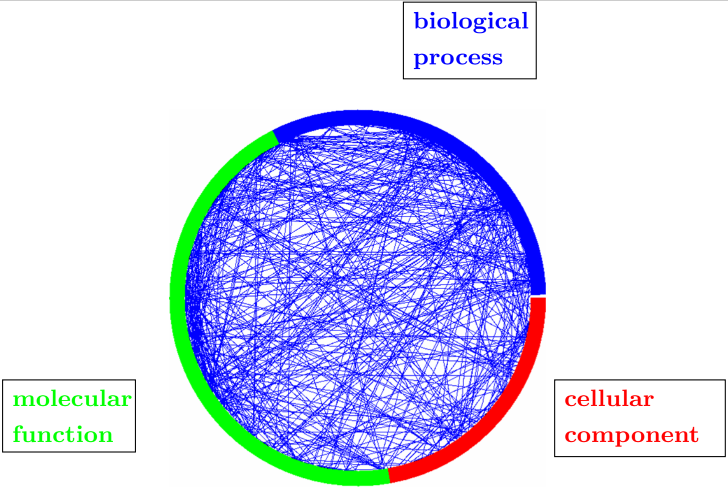
- embryo
- larva
- pupal
- adult

Analyze subset of 588 genes related to development

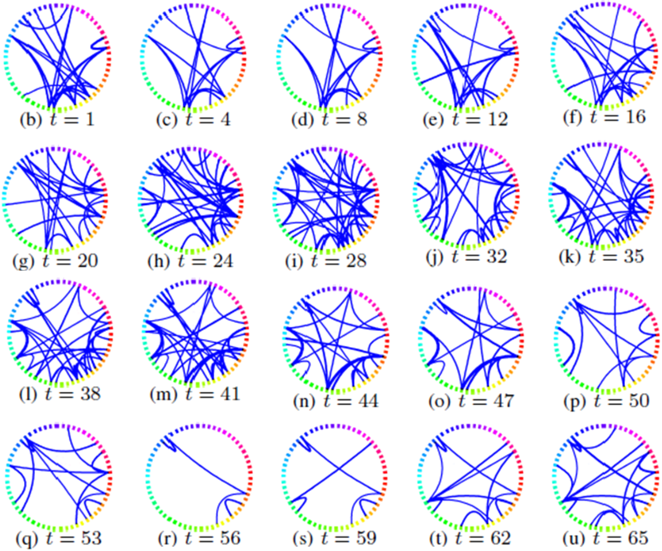


Kolar, Song, Ahmed, and Xing (2010)

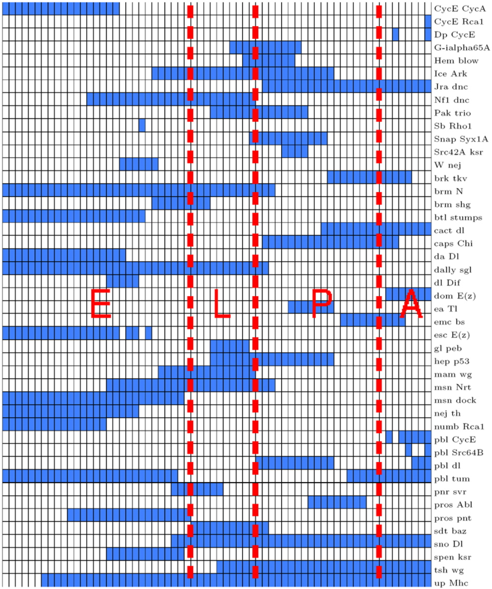
Estimated Dynamic Network



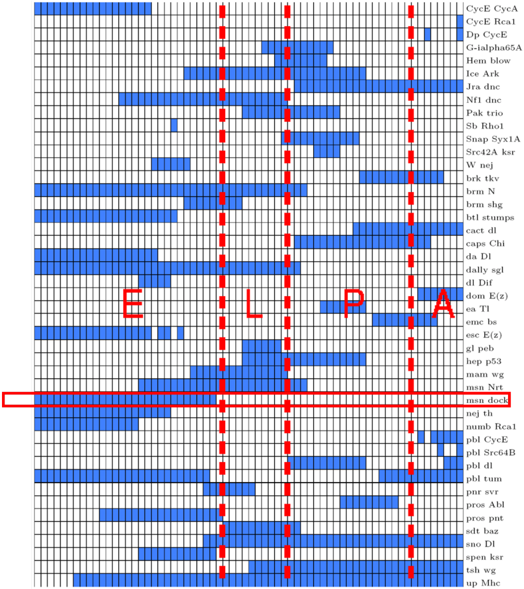
Transient Group Interactions



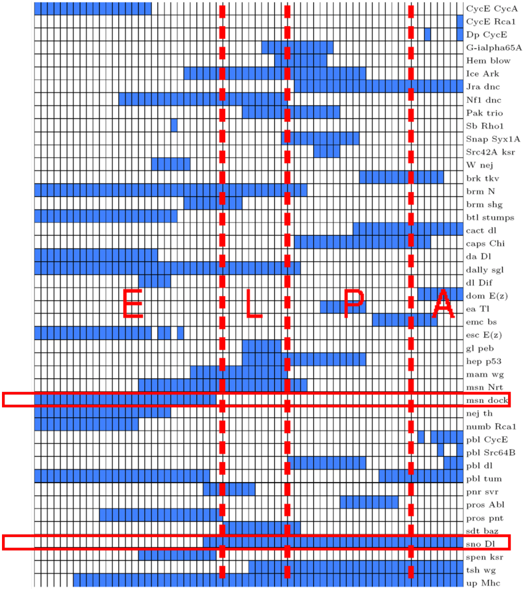
Known Gene Interactions



Known Gene Interactions



Known Gene Interactions



Outline

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

Outline

1. Estimating Conditional Independence Relationships
   - Representation – Markov Networks
   - Estimating Graph Structure
2. Time-Varying Networks
   - Smoothly Varying Networks
   - Networks With Jumps
3. An Application
4. **S**ome theoretical results

Theoretical Properties

Theorem (Kolar and Xing (2012c))

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

Fisher information matrix:

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

- bounded eigenvalues
- incoherence condition

Theoretical Properties

Theorem (Kolar and Xing (2012c))

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

Fisher information matrix:

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

- bounded eigenvalues
- incoherence condition

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

Theoretical Properties

Theorem (Kolar and Xing (2012c))

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

Fisher information matrix:

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

- bounded eigenvalues
- incoherence condition

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

Kernel satisfies regularity conditions

Theoretical Properties

Theorem (Kolar and Xing (2012c))

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

Theoretical Properties

Theorem (Kolar and Xing (2012c))

Under suitable technical assumptions the graph G^{τ} 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^{τ} 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: $\frac{s^3 \log p}{n^{2/3}} = o(1)$ $\quad$ (s – maximal node degree)

Theoretical Properties

Theorem (Kolar and Xing (2012c))

Under suitable technical assumptions the graph G^τ 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: $\frac{s^3 \log p}{n^{2/3}} = o(1)$ $\qquad$ (s – maximal node degree)

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

Theoretical Properties

Theorem (Kolar and Xing (2012c))

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

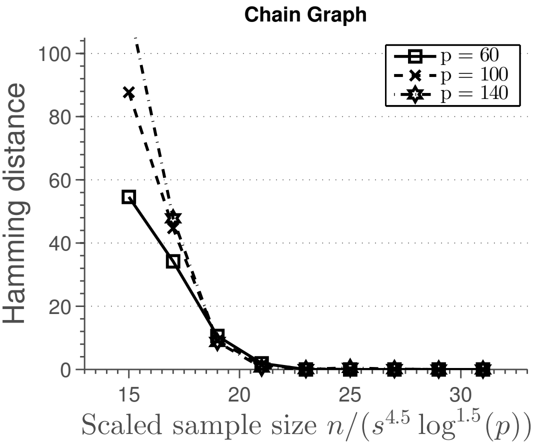
Parameters: $\lambda \asymp \frac{\sqrt{\log p}}{n^{1/3}}$, $h \asymp n^{-\frac{1}{3}}$

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

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

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

Simulation Results



Thank you!

References I

Arbeitman, M., E. Furlong, F. Imam, E. Johnson, B. Null, B. Baker, M. Krasnow, M. Scott, R. Davis, and K. White (2002). "Gene Expression During the Life Cycle of *Drosophila melanogaster*". In: *Science* 297, pp. 2270–2275.

Balakrishnan, S., M. Kolar, A. Rinaldo, and A. Singh (2012). "Recovering Block-structured Activations Using Compressive Measurements". In: *arXiv preprint arXiv:1209.3431*.

Banerjee, Onureena, Laurent El Ghaoui, and Alexandre d'Aspremont (2008). "Model Selection Through Sparse Maximum Likelihood Estimation for Multivariate Gaussian or Binary Data". In: *J. Mach. Learn. Res.* 9, pp. 485–516. ISSN: 1533-7928.

Beck, A. and M. Teboulle (2009). "A fast iterative shrinkage-thresholding algorithm for linear inverse problems". In: *SIAM Journal on Imaging Sciences* 2.1, 183202.

Chickering, David Maxwell (1996). "Learning Bayesian Networks is NP-Complete". In: *Learning from Data: Artificial Intelligence and Statistics V*. Springer-Verlag, pp. 121–130.

Drton, Mathias and Michael D. Perlman (2007). "Multiple Testing and Error Control in Gaussian Graphical Model Selection". In: *Statistical Science* 22.3, pp. 430–449. URL: `doi:10.1214/088342307000000113`.

Friedman, J., T. Hastie, and R. Tibshirani (2008). "Regularization paths for generalized linear models via coordinate descent". In: *Department of Statistics, Stanford University, Tech. Rep.*

References II

Kolar, M., H. Liu, and E. P. Xing (Oct. 2012). "Graph Estimation From Multi-attribute Data". In: *ArXiv e-prints*. arXiv:1210.7665 [stat.ML].

Kolar, M. and J. Sharpnack (May 2012). "Variance function estimation in high-dimensions". In: *ArXiv e-prints*. arXiv:1205.4770 [stat.ML].

Kolar, M. and E.P. Xing (2011). "On Time Varying Undirected Graphs". In: *Proceedings of the 14th International Conference on Artifical Intelligence and Statistics AISTATS*.

— (2012a). "Estimating Networks With Jumps". In: *Electronic Journal of Statistics*.

— (2012b). "Estimating Sparse Precision Matrices from Data with Missing Values". In:

Kolar, M., S. Balakrishnan, A. Rinaldo, and A. Singh (2011). "Minimax localization of structural information in large noisy matrices". In: *Advances in Neural Information Processing Systems*.

Kolar, Mladen, John Lafferty, and Larry Wasserman (July 2011). "Union Support Recovery in Multi-task Learning". In: *J. Mach. Learn. Res.* 12, pp. 2415–2435. ISSN: 1532-4435.

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

— (2012b). "Union Support Recovery in Multi-task Learning". In: *J. Mach. Learn. Res. (AISTATS track) WCP Vol 22*, pp. 647–655.

References III

Kolar, Mladen, Ankur P. Parikh, and Eric P. Xing (2010). "On Sparse Nonparametric Conditional Covariance Selection". In: *ICML '10: Proceedings of the 27th Annual International Conference on Machine Learning*. Hiafa, Israel.

Kolar, Mladen, Le Song, and Eric Xing (2009). "Sparsistent Learning of Varying-coefficient Models with Structural Changes". In: *Advances in Neural Information Processing Systems 22*. Ed. by Y. Bengio, D. Schuurmans, J. Lafferty, C. K. I. Williams, and A. Culotta, pp. 1006–1014.

Kolar, Mladen and Eric P. Xing (2010). "Ultra-high Dimensional Multiple Output Learning With Simultaneous Orthogonal Matching Pursuit: Screening Approach". In: *AISTATS 2010: Proceedings of the 13th International Conference on Artifical Intelligence and Statistics*, pp. 413–420.

Kolar, Mladen and Eric P Xing (July 2012c). "Sparsistent Estimation of Time-Varying Discrete Markov Random Fields". In: *0907.2337*.

Kolar, Mladen, Le Song, Amr Ahmed, and Eric P. Xing (2010). "Estimating Time-Varying Networks". In: *Annals of Applied Statistics* 4.1, pp. 94–123.

Koller, D. and N. Friedman (2009). *Probabilistic graphical models: principles and techniques*. MIT press.

Lauritzen, S. L. (1996). *Graphical Models (Oxford Statistical Science Series)*. Oxford University Press, USA.

Meinshausen, Nicolai and Peter Bühlmann (2006). "High-dimensional graphs and variable selection with the Lasso". In: *Annals of Statistics* 34.3, pp. 1436–1462.

Nesterov, Yu. (May 2005). "Smooth minimization of non-smooth functions". In: *Mathematical Programming* 103.1, pp. 127–152. DOI: 10.1007/s10107-004-0552-5. URL: http://dx.doi.org/10.1007/s10107-004-0552-5.

Peng, Jie, Pei Wang, Nengfeng Zhou, and Ji Zhu (2009). "Partial Correlation Estimation by Joint Sparse Regression Models". In: *Journal of the American Statistical Association* 104.486, pp. 735–746. DOI: 10.1198/jasa.2009.0126. eprint: http://pubs.amstat.org/doi/pdf/10.1198/jasa.2009.0126. URL: http://pubs.amstat.org/doi/abs/10.1198/jasa.2009.0126.

Ravikumar, P., M. J. Wainwright, and J. D. Lafferty (2009). "High-Dimensional Ising Model Selection Using ℓ_1 Regularized Logistic Regression". In: *Annals of Statistics* to appear.

Song, L., M. Kolar, and E.P. Xing (2009). "Time-varying dynamic bayesian networks". In: *Advances in Neural Information Processing Systems* 22, pp. 1732–1740.

References V

Song, Le, Mladen Kolar, and Eric P. Xing (2009). "KELLER: Estimating Time-Evolving Interactions Between Genes". In: *Proceedings of the 16th International Conference on Intelligent Systems for Molecular Biology*.

Tibshirani, Robert, Michael Saunders, Saharon Rosset, Ji Zhu, and Keith Knight (2005). "Sparsity and smoothness via the fused lasso". In: *Journal Of The Royal Statistical Society Series B* 67.1, pp. 91–108. URL: http://ideas.repec.org/a/bla/jorssb/v67y2005i1p91-108.html.