Laboratory for Statistical Artificial InteLligence & INtegrative Genomics

On Sparse Nonparametric Conditional Covariance Selection

Mladen Kolar, Ankur Parikh, Eric Xing

{mladenk,apparikh,epxing}@cs.cmu.edu

Machine Learning Department

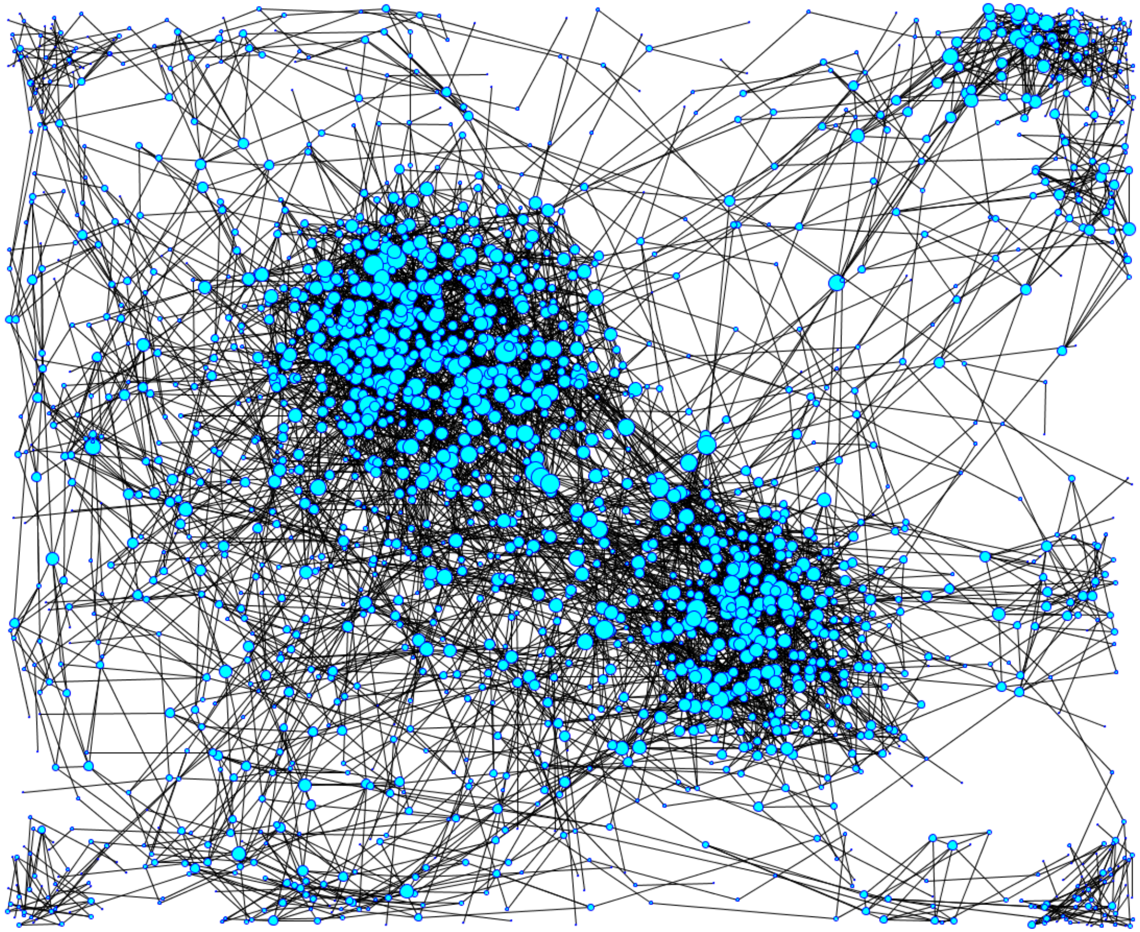
Carnegie Mellon University

Estimating Associations among Variables

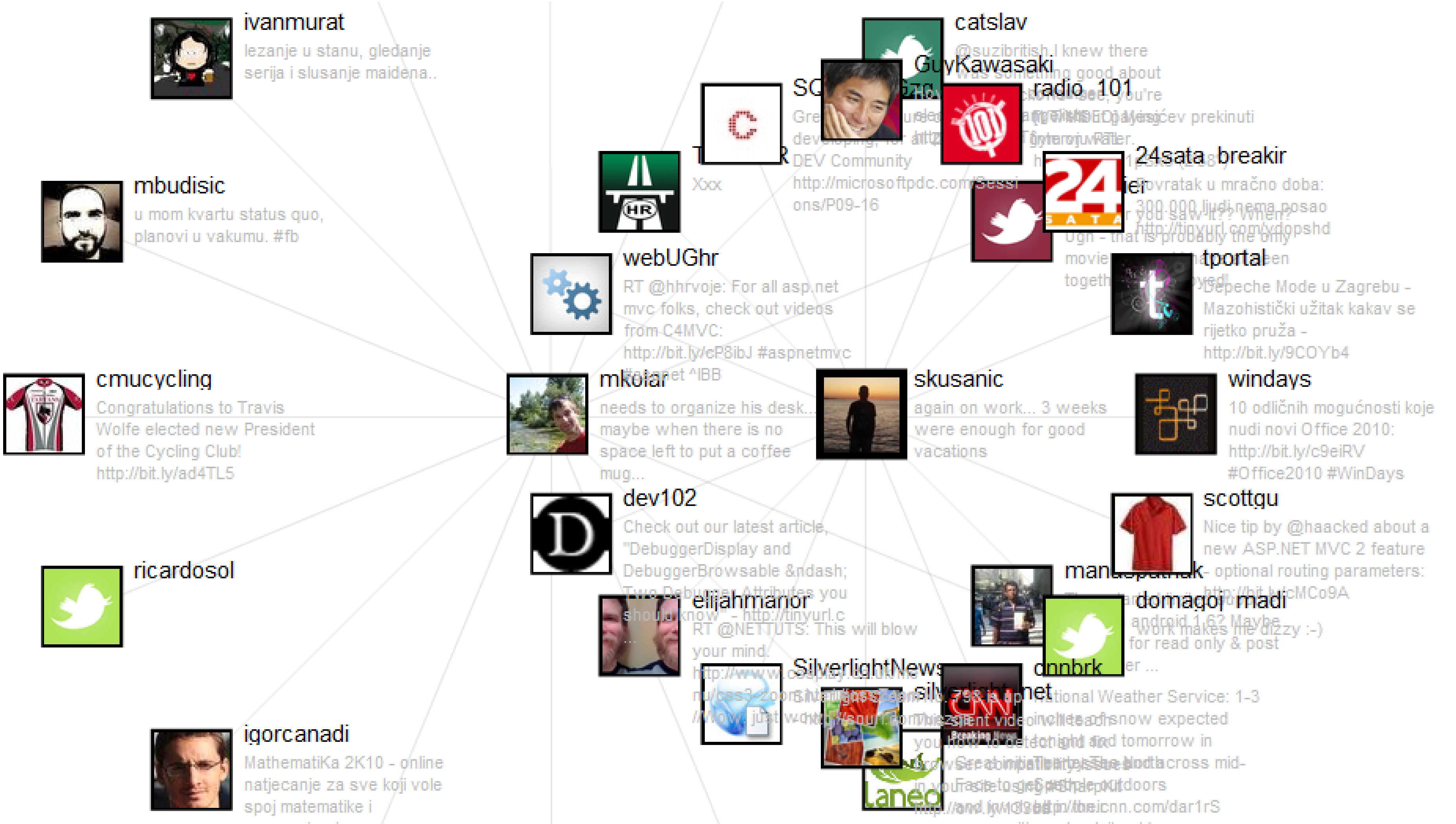
Important to many researchers in……

SAILING LAB
Laboratory for Statistical Artificial InteLligence & INtegrative Genomics

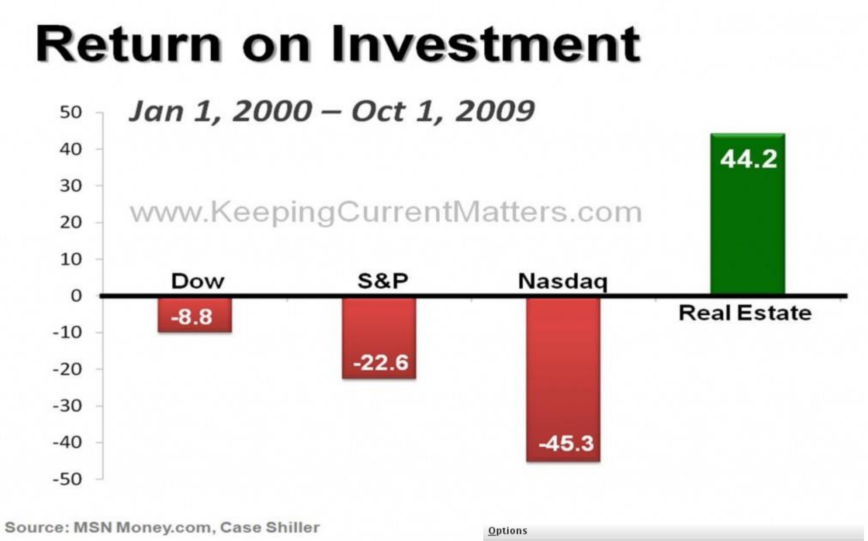
Biology: Gene Regulatory Networks

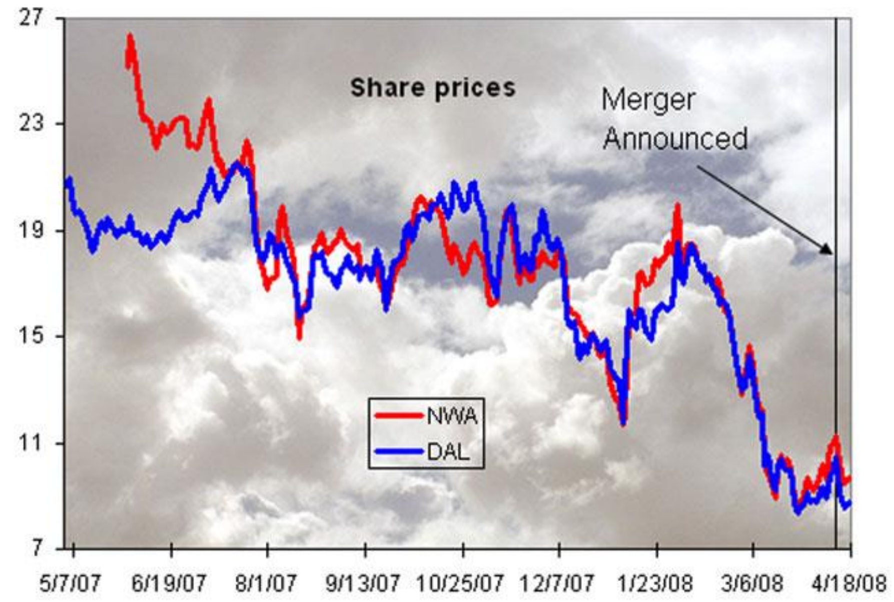


Social Networks: Twitter

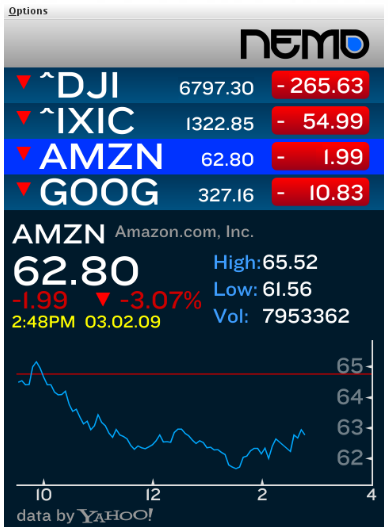


Finance: Stock Associations





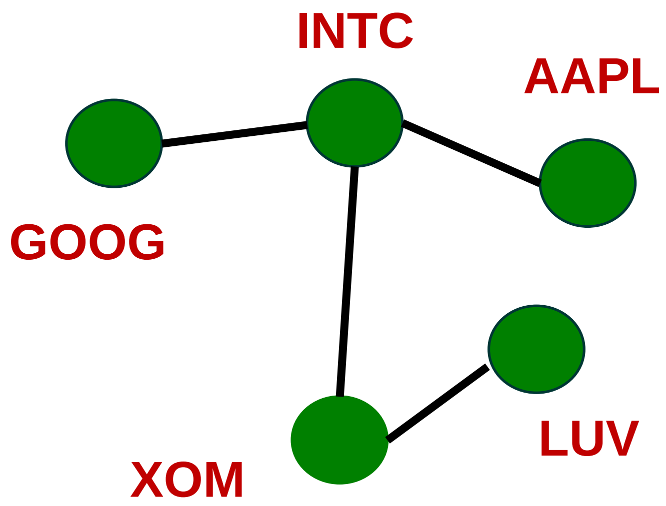
http://minnesota.publicradio.org/display/web/2008/04/15/nwa2/



http://www.xconomy.com/boston/2009/03/31/the-xconomy-mobile-innovation-showcase/attachment/stock-ticker/

Covariance Selection

- Assume data is iid and system is isolated
- Estimate non-zero elements of inverse covariance matrix
  [Meinshausen, Buhlmann 06, Ravikumar et al. 08]
- Can be visualized as network where edges are non-zero precision matrix elements

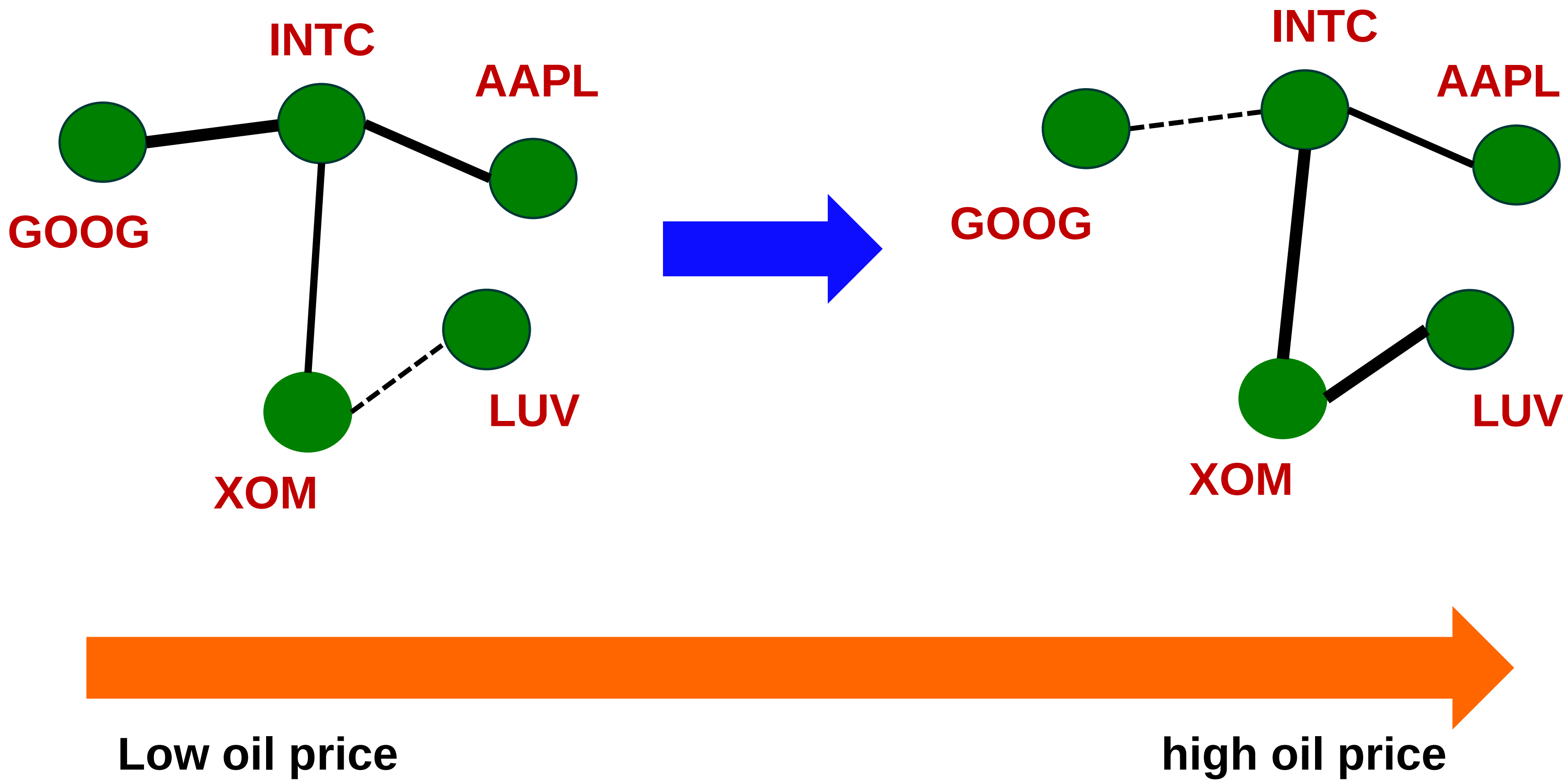


Environmental Variables

- More realistic to model the associations as **functions of the environment**
  - **Genes** in a regulatory network **conditioned on blood pressure**
  - **Stocks** in the NYSE **conditioned on economic indicator**
  - **People** in a social network **conditioned on time**
- Environmental variable can be **continuous** and **random**

Conditional Covariance Estimation

- Assume that set of edges remains fixed, but the values can change.



Conditional Covariance Estimation

$\mathbf{X} \in \mathbb{R}^p$ Set of variables (stocks)

$Z \in \mathbb{R}$ Environmental variable (oil price)

$\mathbf{\Sigma}(z) := \mathrm{Cov}(\mathbf{X}|Z = z)$ Conditional covariance matrix

$\mathbf{\Omega}(z) := \mathbf{\Sigma}(z)^{-1}$ Conditional precision matrix

Goal is to select non-zero elements of conditional precision matrix (network edges)

Partial Correlation

- Partial Correlation - correlation between two variables conditioned on the rest

- Related to precision matrix elements

Precision matrix element

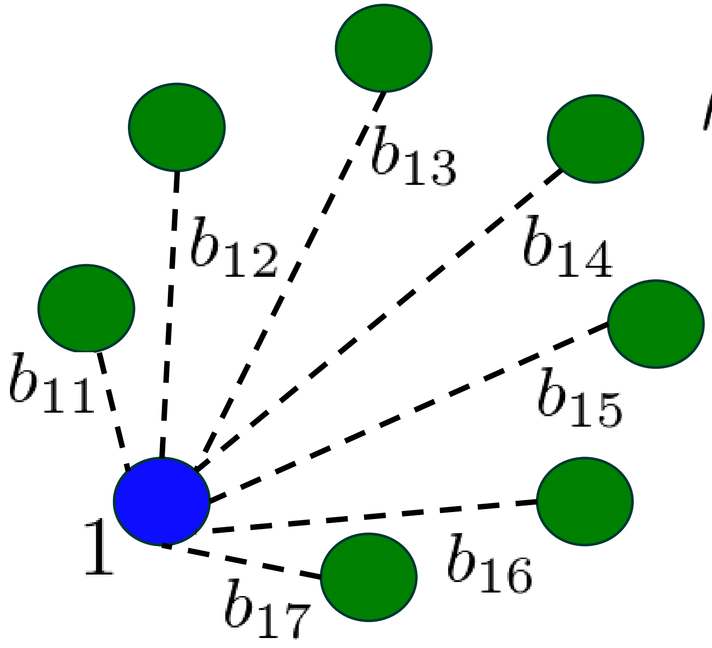
$$\rho_{uv}(z) = -\frac{\omega_{uv}(z)}{\sqrt{\omega_{uu}(z)\omega_{vv}(z)}}$$

Partial correlation coefficient

Neighborhood selection

$$X_u = \sum_{v \neq u} X_v b_{uv}(z) + \epsilon_u(z), \quad u \in [p]$$

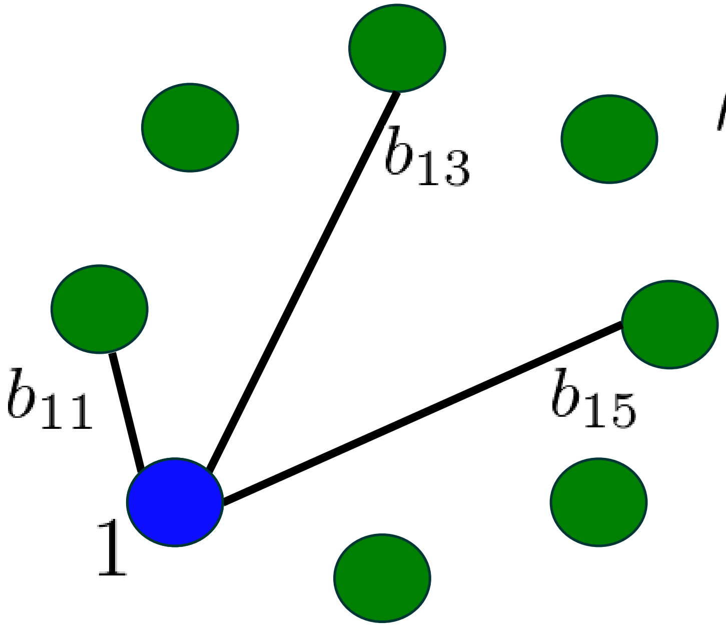
$$\rho_{uv}(z) = \text{sign}(b_{uv}(z)) \sqrt{b_{uv}(z) b_{vu}(z)}$$



Laboratory for Statistical Artificial InteLligence & INtegrative Genomics

Neighborhood selection

$$X_u = \sum_{v \neq u} X_v b_{uv}(z) + \epsilon_u(z), \quad u \in [p]$$

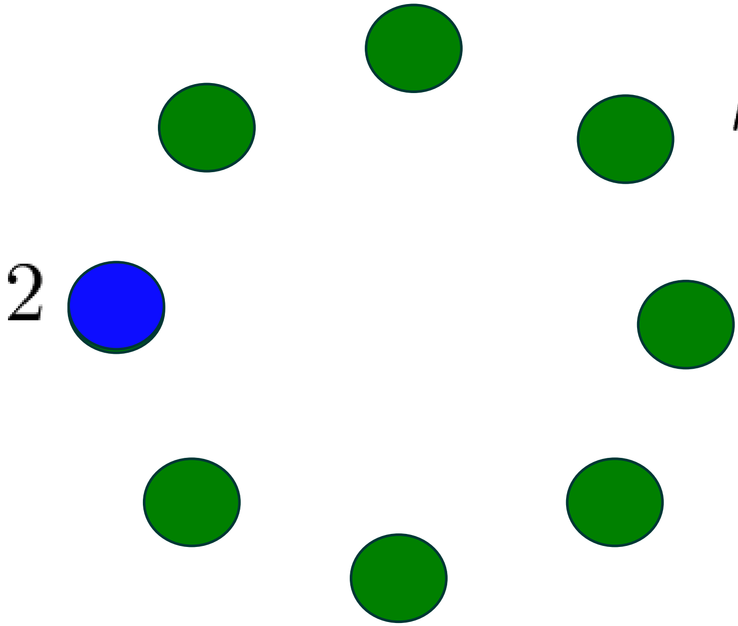


$$\rho_{uv}(z) = \text{sign}(b_{uv}(z))\sqrt{b_{uv}(z) b_{vu}(z)}$$

Neighborhood selection

$$X_u = \sum_{v \neq u} X_v b_{uv}(z) + \epsilon_u(z), \quad u \in [p]$$

$$\rho_{uv}(z) = \mathrm{sign}(b_{uv}(z))\sqrt{b_{uv}(z) b_{vu}(z)}$$



Loss Function

RSS: residual sum of squares

$$\mathcal{L}_u(\mathbf{B}_u; \mathcal{D}^n) := \sum_{z \in \{z^j\}_{j \in [n]}} \sum_{i \in [n]} \left(x_u^i - \sum_{v \neq u} x_v^i b_{uv}(z)\right)^2 K_h(z - z^i) + 2\lambda \sum_{v \neq u} \sqrt{\sum_{z \in \{z^j\}_{j \in [n]}} b_{uv}(z)^2}$$

Kernel: smooth across z

Group penalty: fixed set of non-zero elements

Determining Structure

Structure is fixed for all **z**, so

$$\hat{S} := \{(u, v) : \max\{\|\hat{b}_{uv}(\cdot)\|_2, \|\hat{b}_{vu}(\cdot)\|_2\} > 0\}$$

Optimization

- Modified active shooting algorithm /coordinate descent [Friedman et al. 2010]

- Optimize one group while holding others fixed.

$$\mathcal{L}_u^v(\{b_{uv}(z^j)\}_{j\in[n]};\mathcal{D}^n) := \sum_{z\in\{z^j\}_{j\in[n]}} \sum_{i\in[n]} \left(r_{uv}^i(z) - x_v^i b_{uv}(z)\right)^2 K_h(z-z^i) + 2\lambda \, \|b_{uv}(\cdot)\|_2,$$

$$r_{uv}^i(z) = x_u^i - \sum_{v'\neq u,v} x_{v'}^i \tilde{b}_{uv'}(z)$$

Optimization

- Fast way to check if group is zero

$$\frac{1}{\lambda^2} \sum_{z \in \{z^j\}_{j \in [n]}} \left(\sum_{i \in [n]} x_v^i r_{uv}^i(z) K_h(z - z^i) \right)^2 \leq 1.$$

- Otherwise optimize block (i.e. using gradient descent)

Consistent Structure Estimation

Let $h = \mathcal{O}(n^{-1/5})$, $\lambda = \mathcal{O}(n^{7/10}\sqrt{\log p})$ and $n^{-9/5}\lambda \to 0$.

If $\frac{n^{11/10}}{\sqrt{\log p}} \min_{u,v \in S} \|b_{uv}(\cdot)\|_2 \to \infty$, then $\mathbb{P}[\hat{S} = S] \to 1$.

Selecting Regularization

Extra penalty (Chen & Chen 08)

$$\text{BIC}_u(\lambda) = \log(\text{RSS}_u(\lambda)) + \frac{\widehat{df}_{u,\lambda}(\log(nh) + 2\log p)}{nh}$$

log residual sum of squares

Number of non-zero elements in row

Selecting Regularization

$$\hat{\lambda} = \operatorname{argmin}_{\lambda} \sum_{u \in [p]} \mathrm{BIC}_u(\lambda)$$

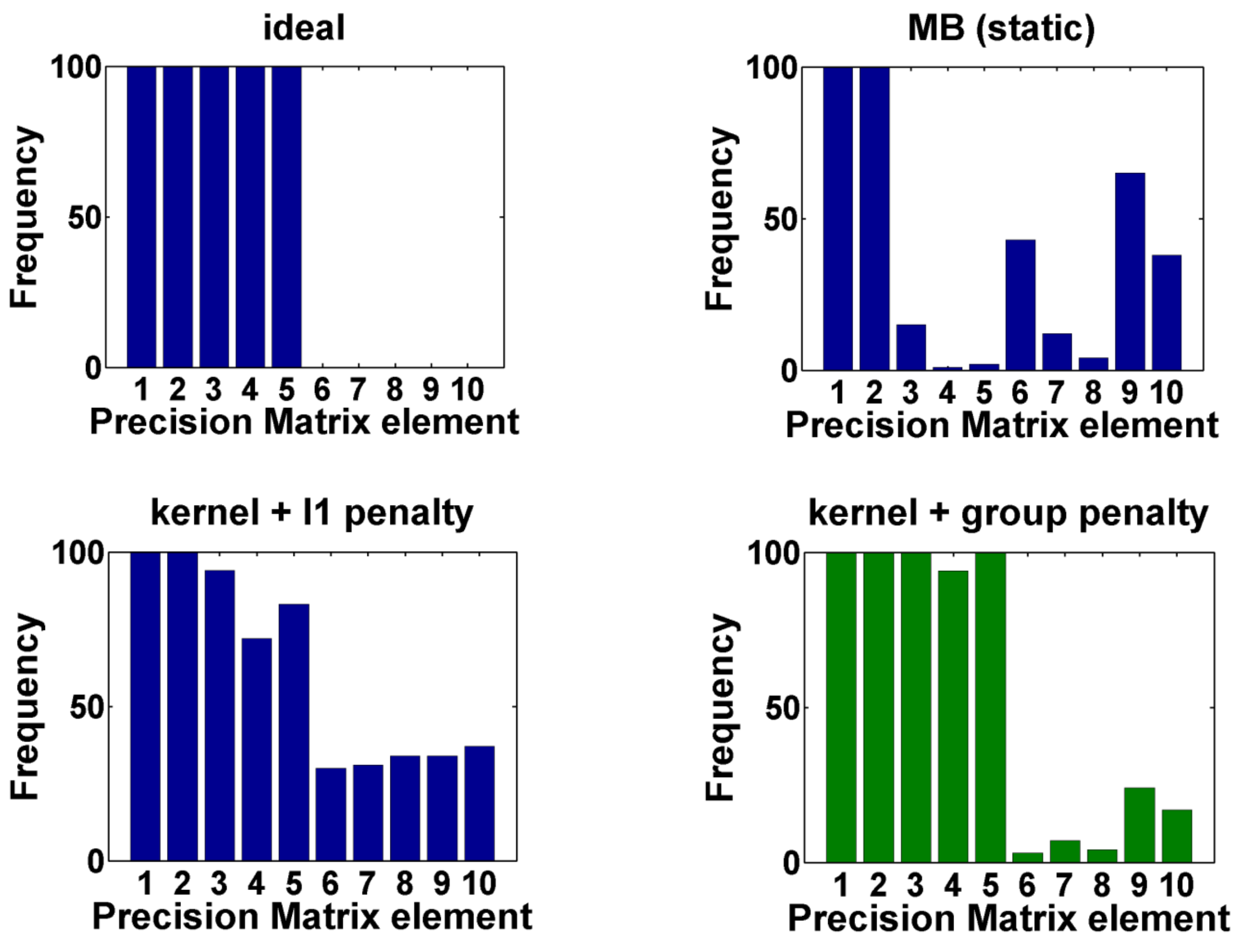
$$\mathbb{P}[\hat{S}(\hat{\lambda}) = S] \to 1.$$

Simulations: A Toy Example

1) Constant
2) Constant
3) Piecewise Constant
4) Linear
5) Sinusoid

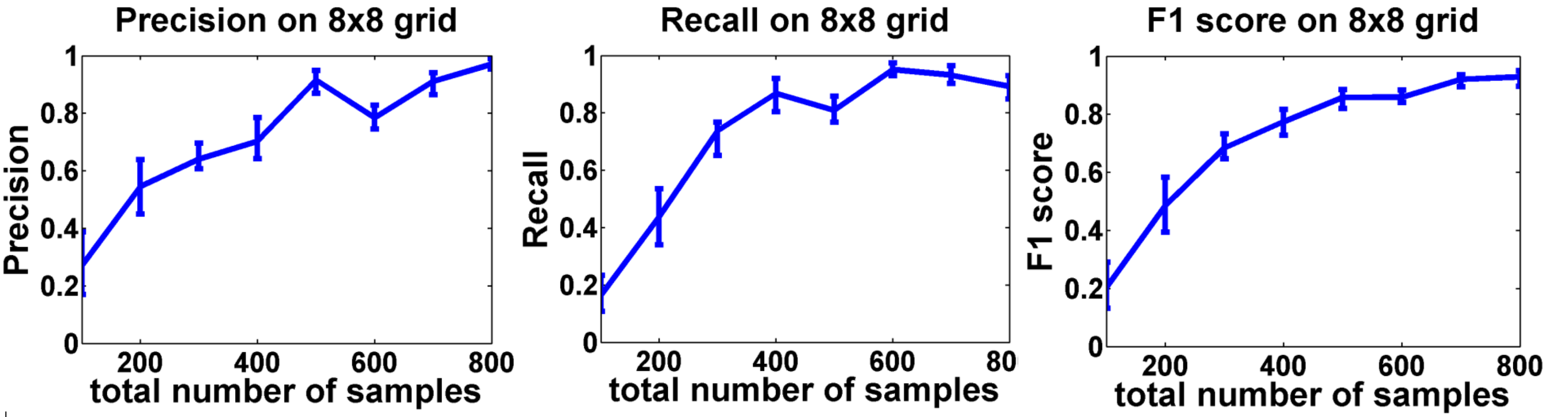
$$\Omega(z) = \begin{bmatrix} & 1 & & & 4 \\ 1 & & & 3 & \\ & & & 2 & 5 \\ & 3 & 2 & & \\ 4 & & 5 & & \end{bmatrix}$$

Toy Example Results



Larger Simulations

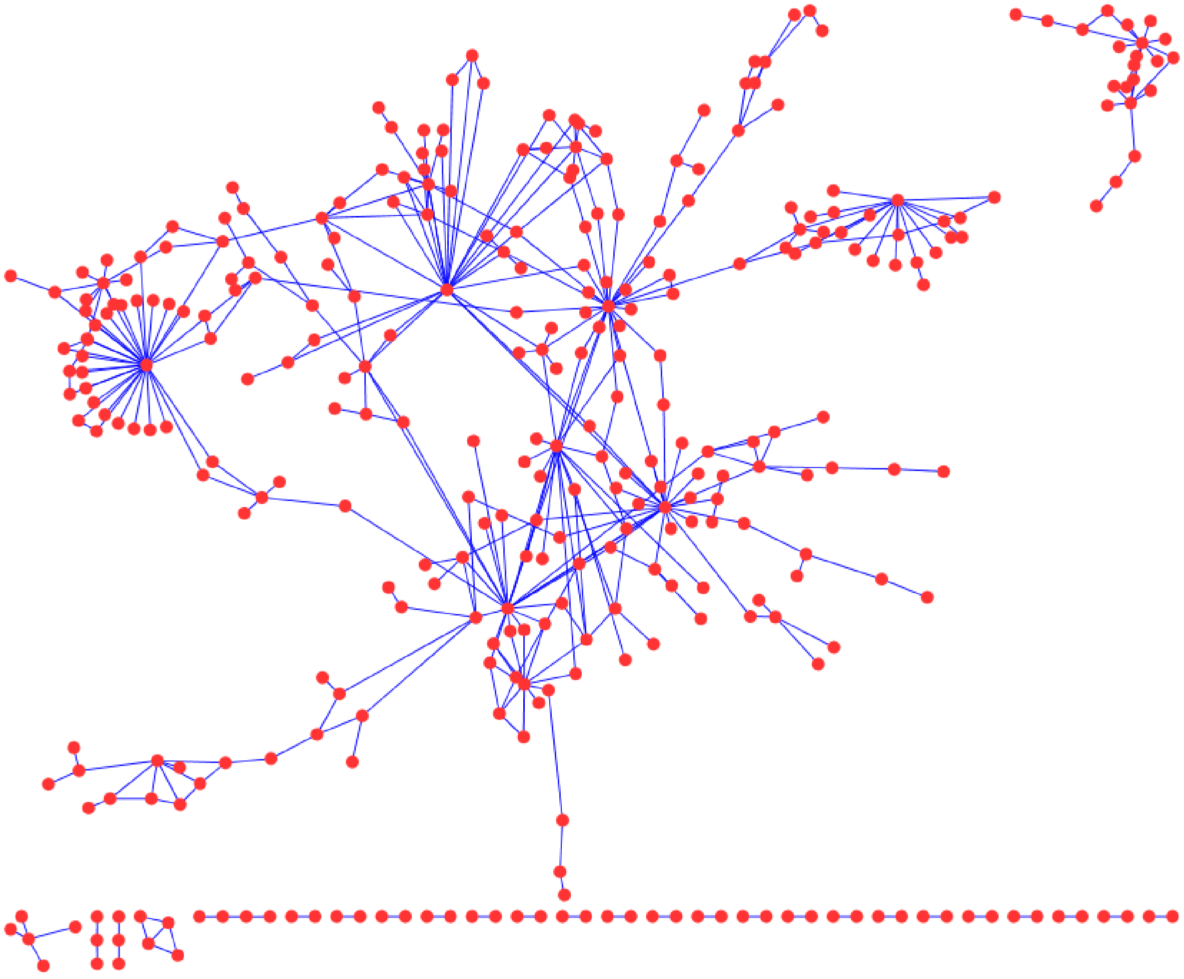
- 8x8 grid
- All non-zero precision matrix elements are sinusoids.



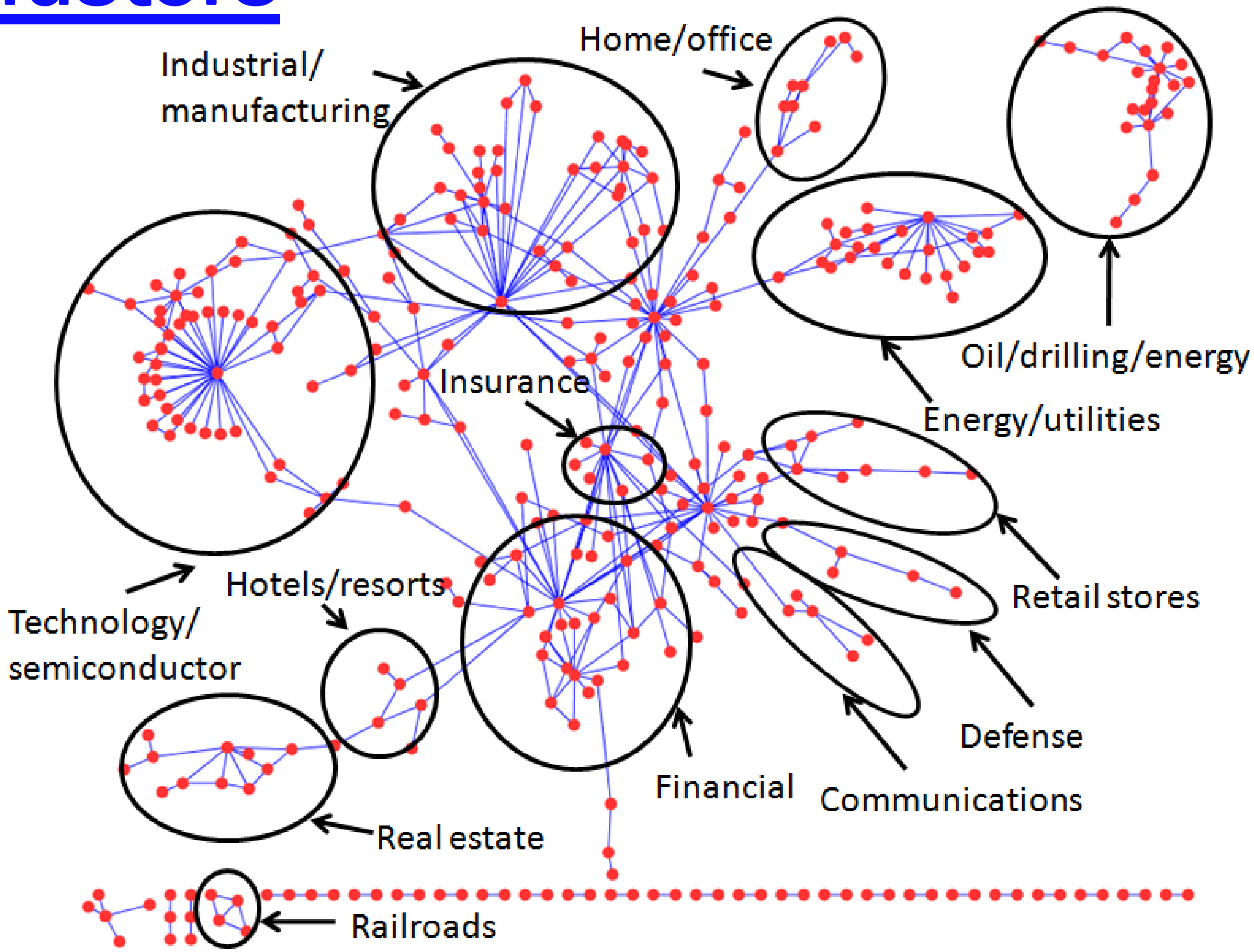
Analyzing the S&P 500

- Examine associations among stocks
  - Help an economist studying the market
  - Assist an investor building a diverse portfolio
- Stock prices from Jan 1, 2003 – Dec 31, 2005.
- Condition on oil price, an economic indicator

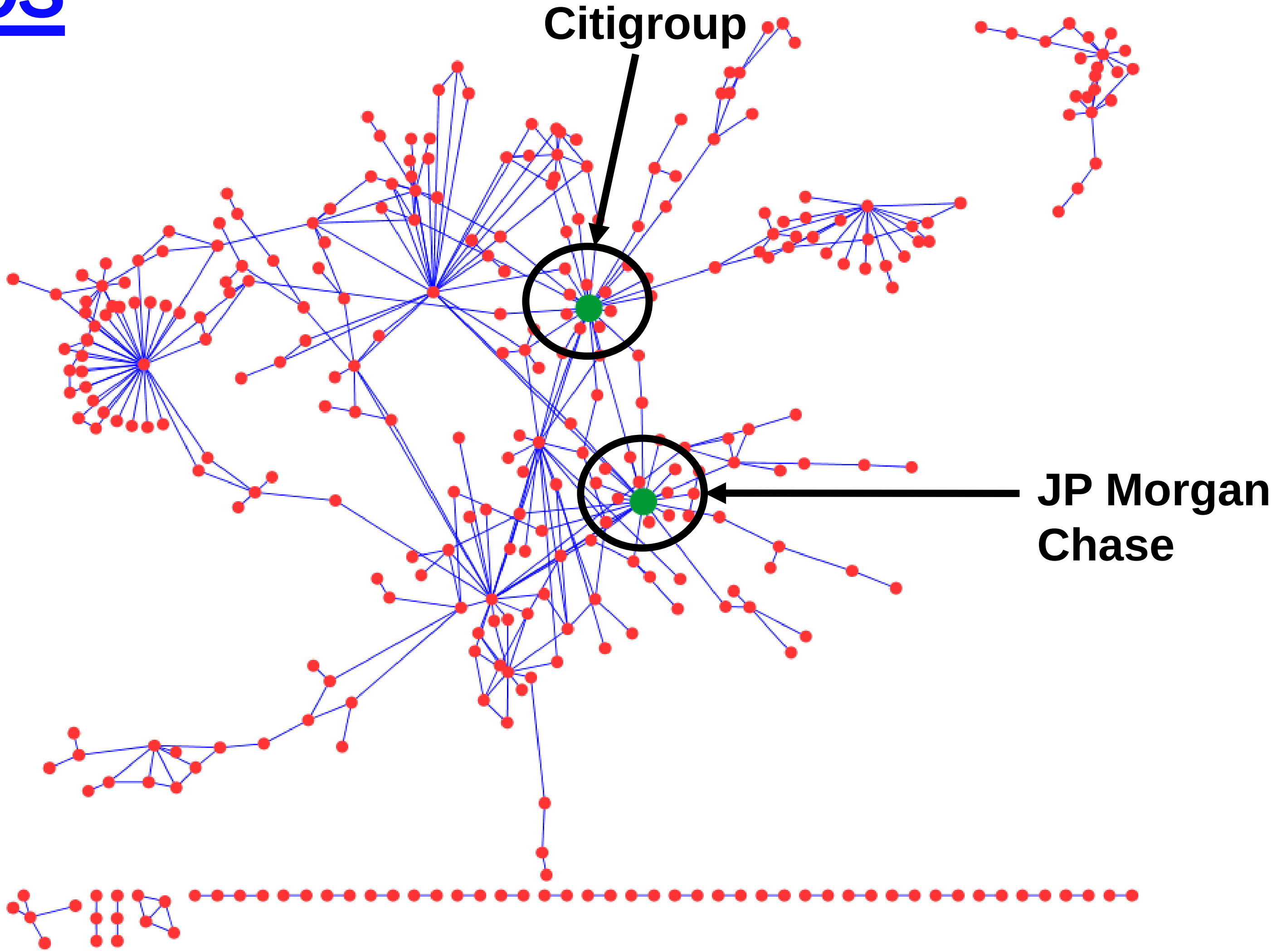
Network



Clusters

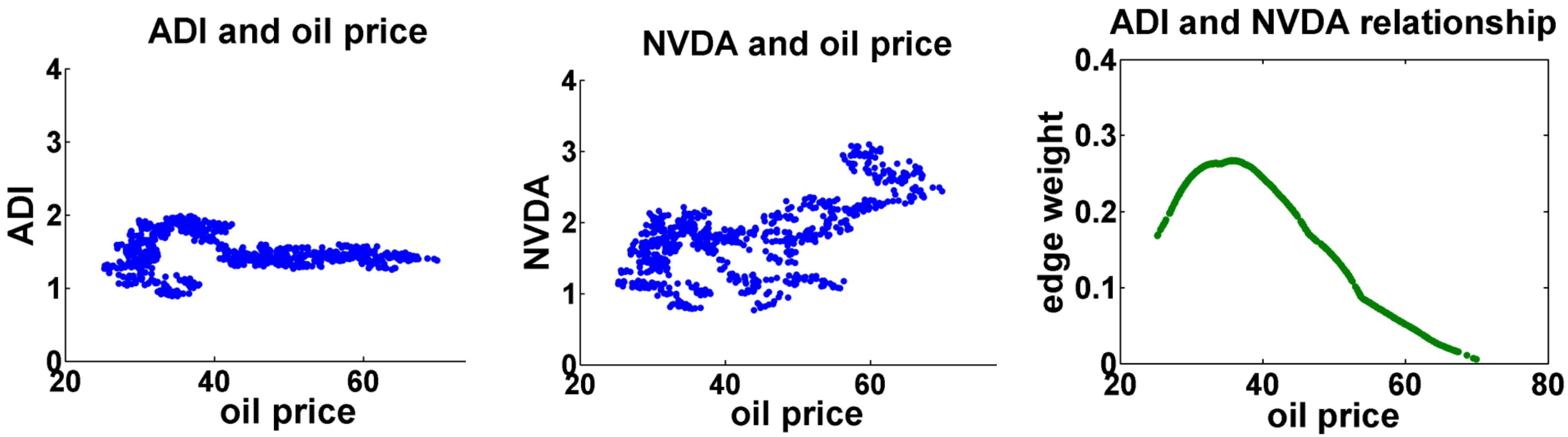

Industrial/
manufacturing
Home/office
Oil/drilling/energy
Energy/utilities
Insurance
Technology/
semiconductor
Hotels/resorts
Retail stores
Defense
Financial
Communications
Real estate
Railroads

Hubs



Edge Weights

- Edge weights (proportional to partial correlations) reflect changes in associations.



Discussion

- Estimate associations among variables conditioned on the environment
- Applicable to biology, finance, social networks etc.
- Our method is simple, nonparametric, and works well in high dimensions

Acknowledgements

ONR

NSF

NIH

Alfred P. Sloan Fellowship to EPX

Thank you!