Identifying graph-structured activation patterns in networks

James Sharpnack[1,2] Aarti Singh[1]

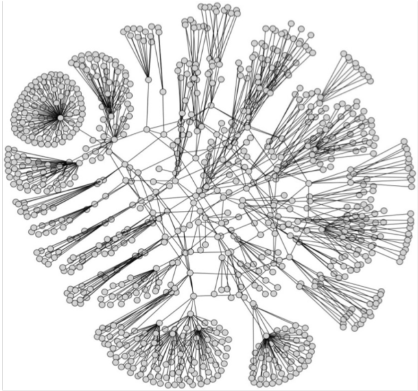
[1]Machine Learning Dept.
CMU, Pittsburgh, PA

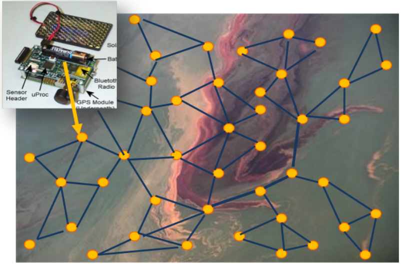
[2]Statistics Dept.
CMU, Pittsburgh, PA

8 Dec. 2010

This work was supported in part by AFOSR grant FA9550-10-1-0382.

Activation Patterns in Networks

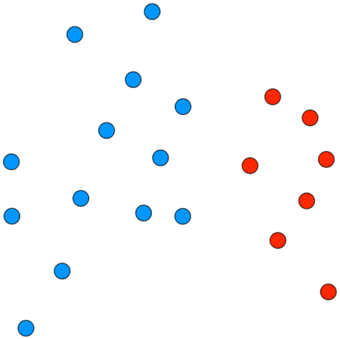




1. Localizing router congestion
2. Detecting water contamination

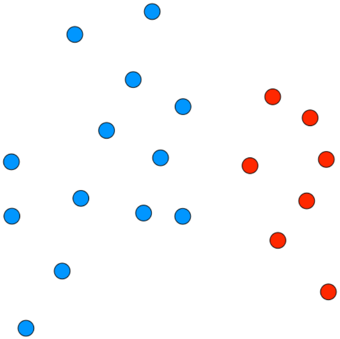
Normal Means Estimation

$\mathbf{x} \in \mathbb{R}^p$ (or $\mathbf{x} \in \{0, 1\}^p$)

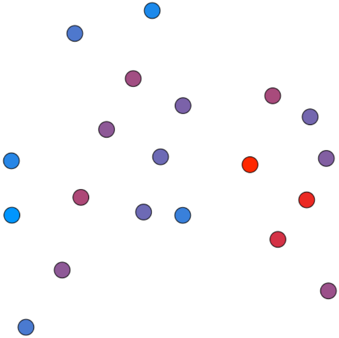


Normal Means Estimation

$\mathbf{x} \in \mathbb{R}^p$ (or $\mathbf{x} \in \{0,1\}^p$)



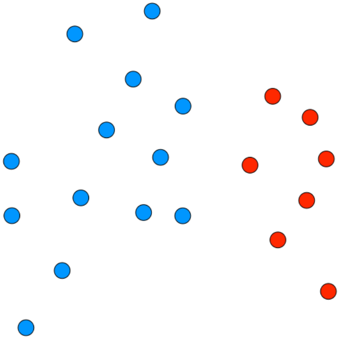
$\mathbf{y} = \mathbf{x} + \zeta$, $\zeta \sim N(0, \sigma^2 I_p)$



Task: reconstruct $\mathbf{x}$ from $\mathbf{y}$

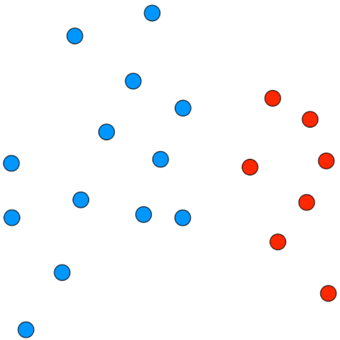
Structured Normal Means Estimation

$\mathbf{x} \in \mathbb{R}^p$ (or $\mathbf{x} \in \{0,1\}^p$)

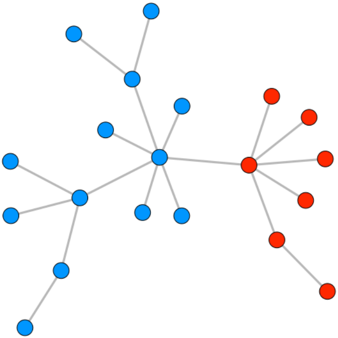


Structured Normal Means Estimation

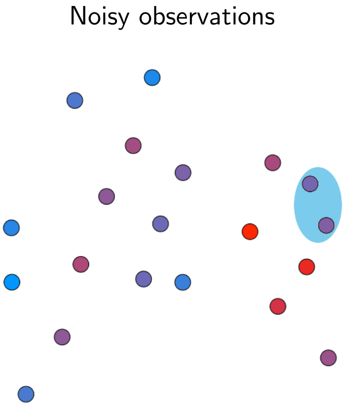
$\mathbf{x} \in \mathbb{R}^p$ (or $\mathbf{x} \in \{0,1\}^p$)

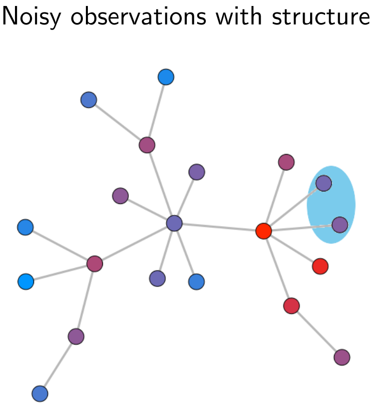


Graph: $G = (V, E, W)$



Structured Normal Means Estimation

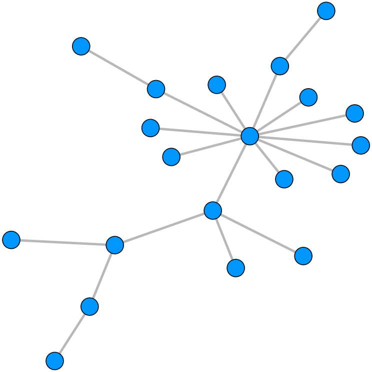




Task: reconstruct $\mathbf{x}$ from $\mathbf{y}$ exploiting dependencies (given by G)

The Model

1 Graph: $G \sim \mathcal{G}_p$ with p nodes.



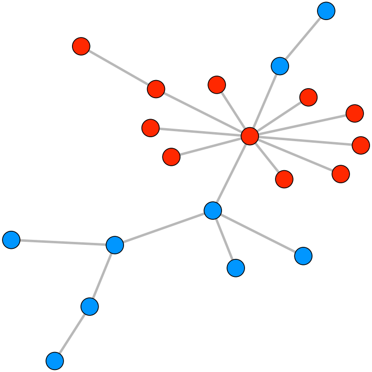
The Model

1 Graph: $G \sim \mathcal{G}_p$ with p nodes.

2 Signal: $\mathbf{x} \sim f_{\mathbf{L}} d\nu$ with

$$f_{\mathbf{L}}(\mathbf{x}) \propto e^{-\mathbf{x}^T \mathbf{L} \mathbf{x}}$$

GGM: ν is Lebesgue ($\Sigma^{-1} = L$)
Ising: ν is Counting
$L = D - W$
$\mathbf{x}^T L \mathbf{x} = \sum_{i \sim j} W_{i,j}(\mathbf{x}_i - \mathbf{x}_j)^2$



The Model

1 Graph: $G \sim \mathcal{G}_p$ with p nodes.

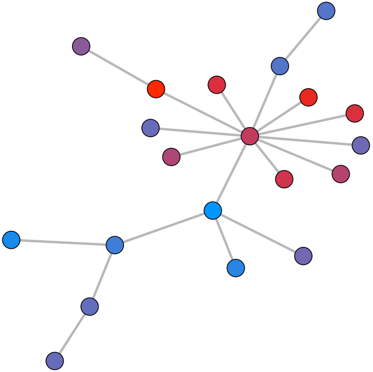
2 Signal: $\mathbf{x} \sim f_{\mathbf{L}} d\nu$ with

$$f_{\mathbf{L}}(\mathbf{x}) \propto e^{-\mathbf{x}^T \mathbf{L} \mathbf{x}}$$

GGM: ν is Lebesgue ($\Sigma^{-1} = L$)
Ising: ν is Counting
$L = D - W$
$\mathbf{x}^T L \mathbf{x} = \sum_{i \sim j} W_{i,j} (\mathbf{x}_i - \mathbf{x}_j)^2$

3 Observations: draw iid. noise $\zeta \sim N(0, \sigma^2 I_p)$

$$\mathbf{y} = \mathbf{x} + \zeta$$



Estimation of Graph-structured patterns

Bayes Optimal Rules:

Mean square error: posterior mean
Hamming distance: posterior centroid
} hard to implement

0/1-loss: posterior max (MAP) implementation via min-cut

Optimal estimator and risk have no closed form - analysis intractable
computing posterior requires knowledge of signal parameters

Estimation of Graph-structured patterns

Bayes Optimal Rules:

Mean square error: posterior mean } hard to implement
Hamming distance: posterior centroid } hard to implement
0/1-loss: posterior max (MAP) — implementation via min-cut

Optimal estimator and risk have no closed form - analysis intractable
computing posterior requires knowledge of signal parameters

Graph-based Regularization:

[Smola-Kondor '03, Belkin-Niyogi '04, Ando-Zhang '06]

Mainly justified in the embedded (manifold) setting
results focus on importance of second eigenvalue of Laplacian

Laplacian Eigenmaps Estimator

Define eigenvalue, eigenvector pairs $\{\lambda_i, u_i\}$ of Laplacian, $\mathbf{L}$, with $\lambda_i \leq \lambda_{i+1}$

Estimator of $\mathbf{x}$ given $k \in \{1, ..., p\}$:

$$\hat{\mathbf{x}} = U_{[k]} U_{[k]}^T \mathbf{y} = \sum_{i=1}^{k} (u_i^T \mathbf{y}) u_i$$

1. Easy to analyze asymptotic risk
2. Easy to implement

Laplacian Eigenmaps Estimator

Define eigenvalue, eigenvector pairs $\{\lambda_i, u_i\}$ of Laplacian, $\mathbf{L}$, with $\lambda_i \leq \lambda_{i+1}$

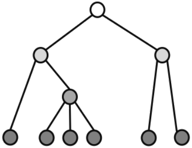
Estimator of $\mathbf{x}$ given $k \in \{1, ..., p\}$:

$$\hat{\mathbf{x}} = U_{[k]} U_{[k]}^T \mathbf{y} = \sum_{i=1}^{k} (u_i^T \mathbf{y}) u_i$$

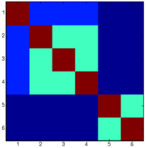
1. Easy to analyze asymptotic risk
2. Easy to implement

```
# 3 lines in R
L = graph.laplacian(g) # igraph package
U = eigen(L)$vectors
Xhat = U[,(p-k):p] %*% t(U[,(p-k):p]) %*% Y
```

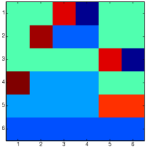
Hierarchical Graph



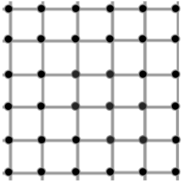
Hierarchical **L**



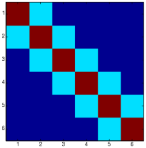
Haar Wavelet



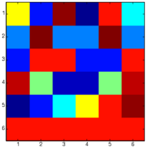
Lattice Graph



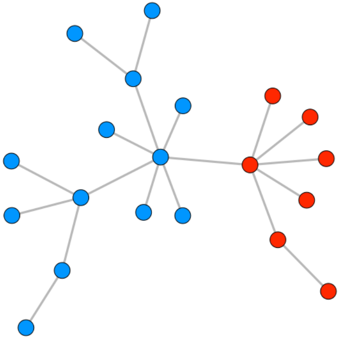
Lattice **L**



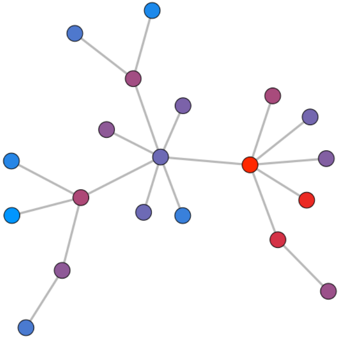
Fourier Basis



Network activation pattern: **x**

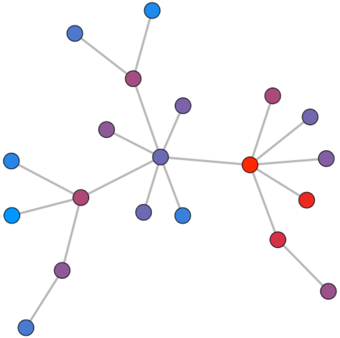


Noisy observations: $\mathbf{y}$ $(\sigma^2 = \frac{1}{2})$

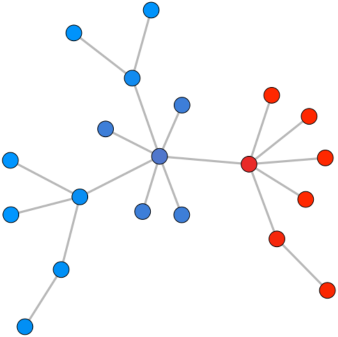


Laplacian Eigenmaps Estimator

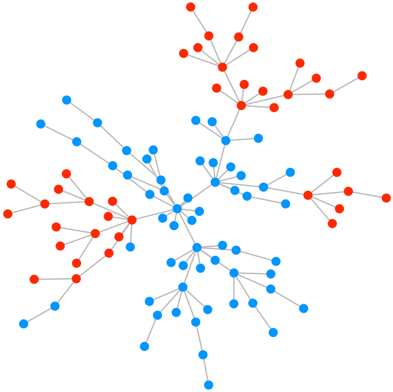
Noisy observations: $\mathbf{y}$ ($\sigma^2 = \frac{1}{2}$)



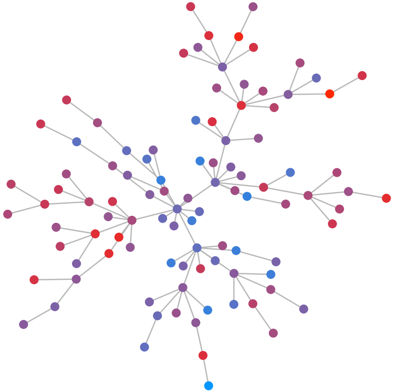
Eigenmaps estimator: $\hat{\mathbf{x}}$ ($k = 3$)



Large real-world graph ($p = 100$)

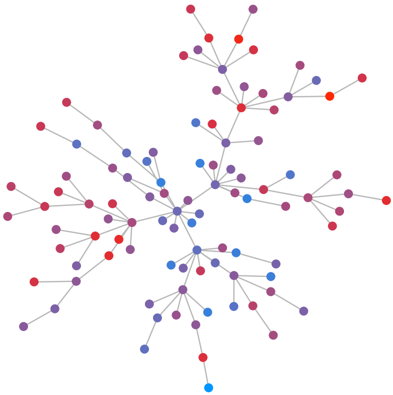


Noisy observations: $\mathbf{y}$ $(\sigma^2 = \frac{4}{5})$

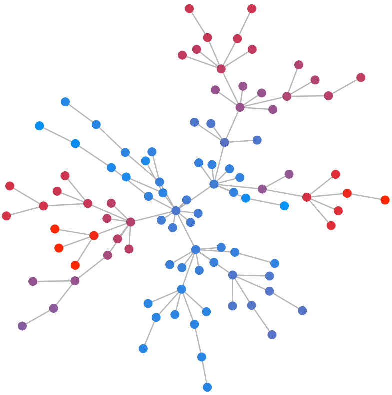


Laplacian Eigenmaps Estimator

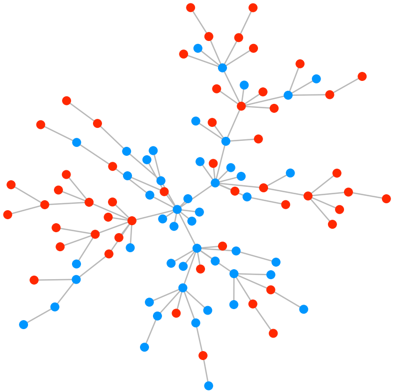
Noisy observations: $\mathbf{y}$ ($\sigma^2 = \frac{4}{5}$)



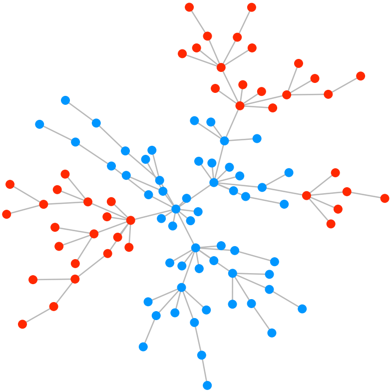
Eigenmaps estimator: $\hat{\mathbf{x}}$ ($k = 10$)



Thresholded observations: $\mathbf{y} > \tau$

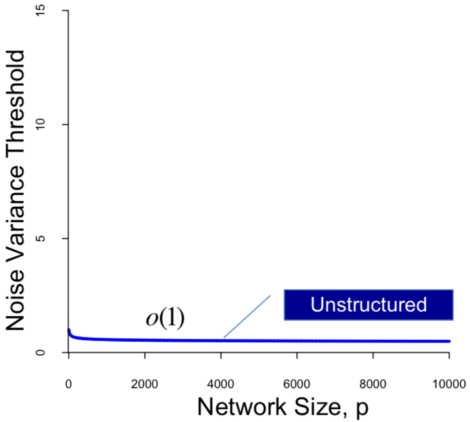


Thresholded eigenmaps estimator: $\hat{\mathbf{x}} > \tau$



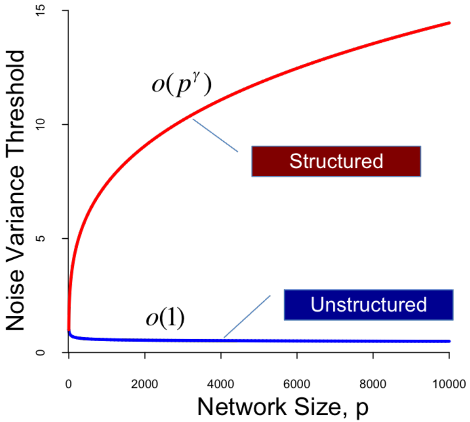
Big Picture

Consistent estimation: $R_B = \mathbb{E}_{\mathbf{x}} \frac{1}{p} ||\widehat{x}_k - \mathbf{x}||^2 \underset{p\to\infty}{\longrightarrow} 0$



Big Picture

Consistent estimation: $R_B = \mathbb{E}_{\mathbf{x}} \frac{1}{p} ||\widehat{x}_k - \mathbf{x}||^2 \underset{p\to\infty}{\longrightarrow} 0$



Tolerable noise: $\sigma^2 = o(p^\gamma) \Rightarrow$ consistent estimation

γ depends on the network evolution model

Main Result

Theorem

Let $\mathbf{x}$ *be drawn from the Ising with graph Laplacian* $\mathbf{L}$*.*

$$R_B := \frac{1}{p}\ \mathbb{E}[\|\widehat{\mathbf{x}}_k - \mathbf{x}\|^2] \leq e^{-p} + \min\left(1, \frac{\delta}{\lambda_{k+1}}\right) + \frac{k\sigma^2}{p}$$

where $0 < \delta < 2$ *is a constant and* λ_{k+1} *is the* $(k+1)^{th}$ *smallest eigenvalue of* $\mathbf{L}$*.*

Main Result

Theorem

Let $\mathbf{x}$ *be drawn from the Ising with graph Laplacian* $\mathbf{L}$*.*

$$R_B := \frac{1}{p}\ \mathbb{E}[\|\widehat{\mathbf{x}}_k - \mathbf{x}\|^2] \leq e^{-p} + \min\left(1, \frac{\delta}{\lambda_{k+1}}\right) + \frac{k\sigma^2}{p}$$

where $0 < \delta < 2$ *is a constant and* λ_{k+1} *is the* $(k+1)^{th}$ *smallest eigenvalue of* $\mathbf{L}$*.*

$$R_B \leq \text{concentration bound} + \text{bias} + \text{variance}$$

Main Result

Theorem

Let $\mathbf{x}$ *be drawn from the Ising with graph Laplacian* $\mathbf{L}$.

$$R_B := \frac{1}{p}\ \mathbb{E}[\|\widehat{\mathbf{x}}_k - \mathbf{x}\|^2] \leq e^{-p} + \min\left(1, \frac{\delta}{\lambda_{k+1}}\right) + \frac{k\sigma^2}{p}$$

where $0 < \delta < 2$ *is a constant and* λ_{k+1} *is the* $(k+1)^{th}$ *smallest eigenvalue of* $\mathbf{L}$.

$$R_B \leq \text{concentration bound} + \text{bias} + \text{variance}$$

Main Result

Theorem

Let $\mathbf{x}$ be drawn from the Ising with graph Laplacian $\mathbf{L}$.

$$R_B := \frac{1}{p}\ \mathbb{E}[\|\widehat{\mathbf{x}}_k - \mathbf{x}\|^2] \leq e^{-p} + \min\left(1, \frac{\delta}{\lambda_{k+1}}\right) + \frac{k\sigma^2}{p}$$

where $0 < \delta < 2$ is a constant and λ_{k+1} is the $(k+1)^{th}$ smallest eigenvalue of $\mathbf{L}$.

$$R_B \leq \text{concentration bound} + \text{bias} + \text{variance}$$

Main Result

Theorem

Let $\mathbf{x}$ *be drawn from the Ising with graph Laplacian* $\mathbf{L}$*.*

$$R_B := \frac{1}{p}\ \mathbb{E}[\|\widehat{\mathbf{x}}_k - \mathbf{x}\|^2] \leq e^{-p} + \min\left(1, \frac{\delta}{\lambda_{k+1}}\right) + \frac{k\sigma^2}{p}$$

where $0 < \delta < 2$ *is a constant and* λ_{k+1} *is the* $(k+1)^{th}$ *smallest eigenvalue of* $\mathbf{L}$*.*

$$R_B \leq \text{concentration bound} + \text{bias} + \text{variance}$$

Main Result

Theorem

Let $\mathbf{x}$ *be drawn from the Ising with graph Laplacian* $\mathbf{L}$.

$$R_B := \frac{1}{p}\ \mathbb{E}[\|\widehat{\mathbf{x}}_k - \mathbf{x}\|^2] \leq e^{-p} + \min\left(1, \frac{\delta}{\lambda_{k+1}}\right) + \frac{k\sigma^2}{p}$$

where $0 < \delta < 2$ *is a constant and* λ_{k+1} *is the* $(k+1)^{th}$ *smallest eigenvalue of* $\mathbf{L}$.

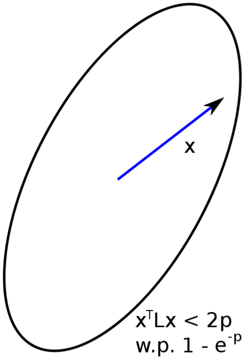
$$R_B \leq \text{concentration bound} + \text{bias} + \text{variance}$$

- Tradeoff between quantile of the eigenvalue distribution (λ_{k+1}) and which quantile it is ($\frac{k}{p}$).

Eigenmaps Geometry

$$\hat{\mathbf{x}} = U_{[k]} U_{[k]}^T \mathbf{y} = U_{[k]} U_{[k]}^T \mathbf{x} + U_{[k]} U_{[k]}^T \zeta$$

$$R_B \leq e^{-p} + \min\left(1, \frac{\delta}{\lambda_{k+1}}\right) + \frac{k\sigma^2}{p}$$

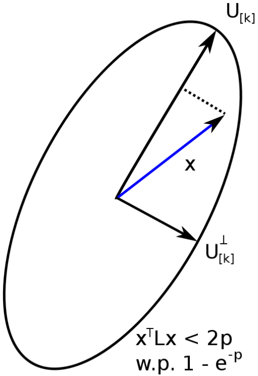


- Chernoff type bound $\Rightarrow$ concentration of prior

Eigenmaps Geometry

$$\hat{\mathbf{x}} = U_{[k]} U_{[k]}^T \mathbf{y} = U_{[k]} U_{[k]}^T \mathbf{x} + U_{[k]} U_{[k]}^T \zeta$$

$$R_B \leq e^{-p} + \min\left(1, \frac{\delta}{\lambda_{k+1}}\right) + \frac{k\sigma^2}{p}$$

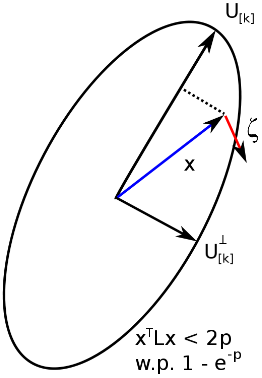


- Chernoff type bound $\Rightarrow$ concentration of prior
- Projection loss at most $\frac{\delta p}{\lambda_{k+1}}$

Eigenmaps Geometry

$$\hat{\mathbf{x}} = U_{[k]} U_{[k]}^T \mathbf{y} = U_{[k]} U_{[k]}^T \mathbf{x} + U_{[k]} U_{[k]}^T \zeta$$

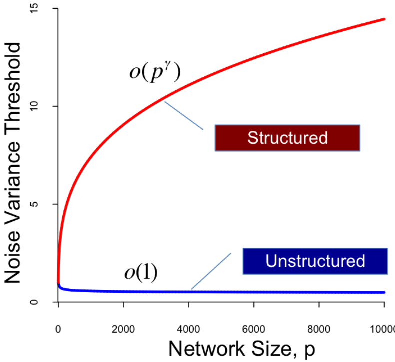
$$R_B \leq e^{-p} + \min\left(1, \frac{\delta}{\lambda_{k+1}}\right) + \frac{k\sigma^2}{p}$$



- Chernoff type bound $\Rightarrow$ concentration of prior
- Projection loss at most $\frac{\delta p}{\lambda_{k+1}}$
- Projection reduces isotropic noise, ζ

Big Picture

$$\text{Recall: } R_B \leq \frac{2}{\lambda_{k+1}} + \frac{k\sigma^2}{p} + e^{-p}$$



Goal: for simple graph models $\mathcal{G}_p$ what is γ?

Hierarchical Structure: Eigenvalue Concentration

Lemma (Ogielski & Stein '85)

For the hierarchical structure with interaction strength, β, and maximum distance between leaves with interaction, $2\ell^*$,

$$\lambda_\ell \geq 2^{\beta\ell^*-1} \text{ is } 2^{\ell-1}\text{-fold degenerate for } \ell \geq \log_2 p - \ell^* + 1$$

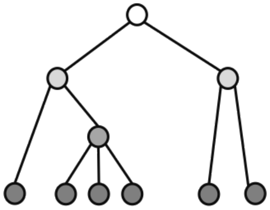


Figure: Hierarchical Graph

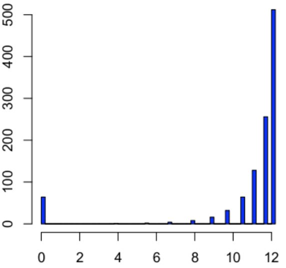


Figure: Eigenvalue Histogram

Hierarchical Structure: Risk Consistency

$$\text{Recall: } R_B \leq \frac{2}{\lambda_{k+1}} + \frac{k\sigma^2}{p} + e^{-p}$$

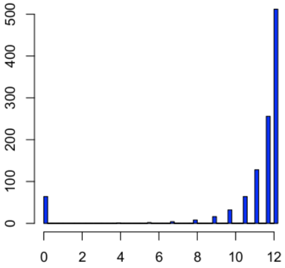


Figure: Eigenvalue Histogram

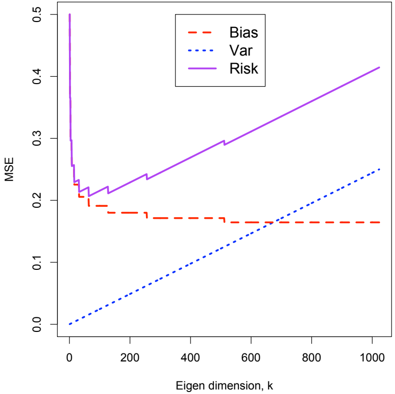


Figure: Bias Var Trade-off

Hierarchical Structure: Risk Consistency

$$\text{Recall: } R_B \leq \frac{2}{\lambda_{k+1}} + \frac{k\sigma^2}{p} + e^{-p}$$

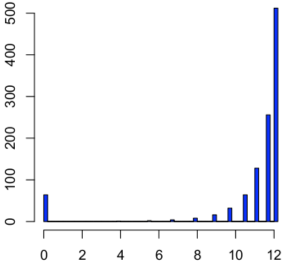


Figure: Eigenvalue Histogram

$$\#\{\lambda_\ell < 2^{\beta\ell^*-1}\} \leq 2^{\log_2 p - \ell^* + 1}$$

$$\ell^* = 1 + \gamma \log_2 p$$

$$\text{Set } k = 2^{\log_2 p - \ell^* + 1} = p^{1-\gamma}$$

$$1/\lambda_{k+1} \leq 2^{1-\beta\ell^*}$$

$$\sigma^2 = o(p^\gamma) \Rightarrow R_B \to 0$$

Lattice: Eigenvalue Concentration

Lemma

For the lattice graph in d dimensions with $p = n^d$ vertices,

$$\frac{\#\{\lambda_i^{\mathbf{L}} \leq d\}}{p} \leq \exp\{-d/8\}$$

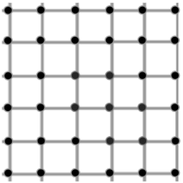


Figure: Lattice Graph

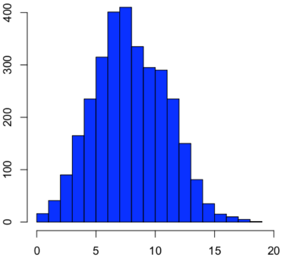


Figure: Eigenvalue Histogram

Lattice: Risk Consistency

$$\text{Recall: } R_B \leq \frac{2}{\lambda_{k+1}} + \frac{k\sigma^2}{p} + e^{-p}$$

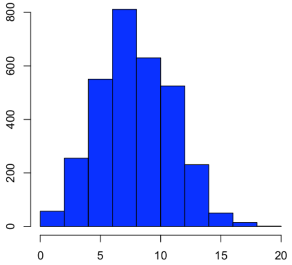


Figure: Eigenvalue Histogram

lattice dimension = 3

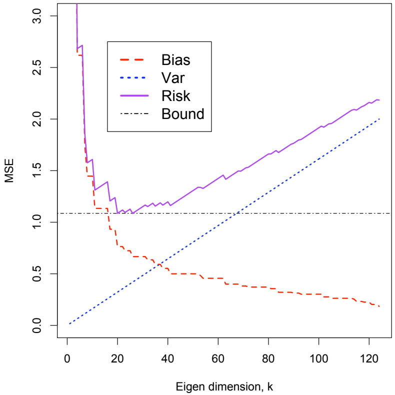


Figure: Bias Var Trade-off

Lattice: Risk Consistency

$$\text{Recall: } R_B \leq \frac{2}{\lambda_{k+1}} + \frac{k\sigma^2}{p} + e^{-p}$$

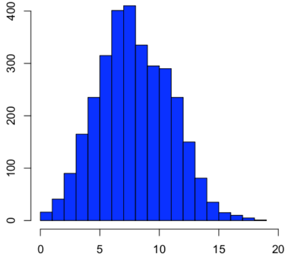


FIGURE: Eigenvalue Histogram

lattice dimension = 4

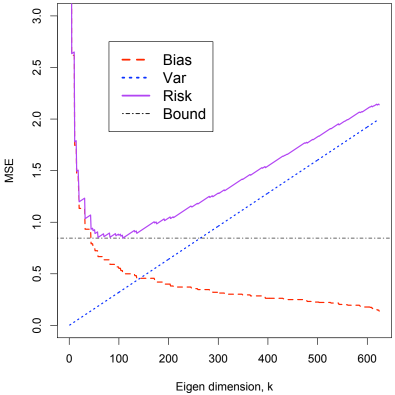


FIGURE: Bias Var Trade-off

Lattice: Risk Consistency

$$\text{Recall: } R_B \leq \frac{2}{\lambda_{k+1}} + \frac{k\sigma^2}{p} + e^{-p}$$

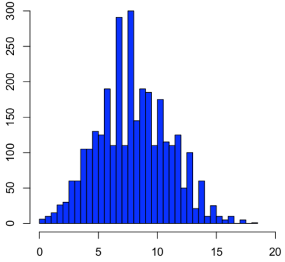


Figure: Eigenvalue Histogram

lattice dimension = 5

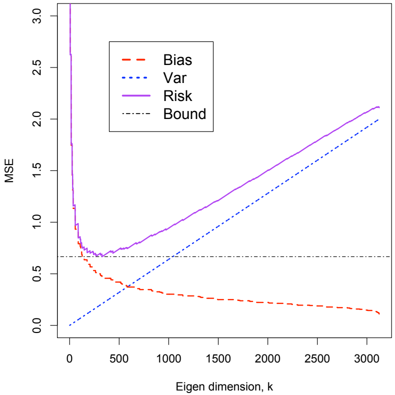


Figure: Bias Var Trade-off

Lattice: Risk Consistency

$$\text{Recall: } R_B \leq \frac{2}{\lambda_{k+1}} + \frac{k\sigma^2}{p} + e^{-p}$$

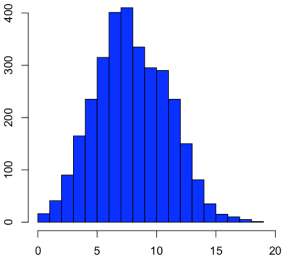


Figure: Eigenvalue Histogram

$$\#\{\lambda_i^{\mathbf{L}} \leq d\} \leq p\exp\{-d/8\}$$

$$d = 8\gamma \ln p$$

$$\text{Set } k = p\exp\{-d/8\} = p^{1-\gamma}$$

$$1/\lambda_{k+1} \leq 1/d$$

$$\sigma^2 = o(p^\gamma) \Rightarrow R_B \to 0$$

Erdös-Rényi Graph

Lemma

Let the probability of an edge be $p^{\gamma-1}$. For any α_p increasing in p, with probability $1 - \mathcal{O}(1/\alpha_p)$,

$$\frac{\#\{\lambda_i \leq p^{\gamma}/2 - p^{\gamma-1}\}}{p} \leq \alpha_p p^{-\gamma} \tag{1}$$

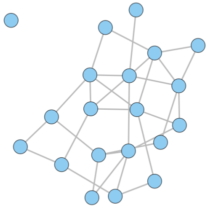


Figure: Erdös-Rényi Graph

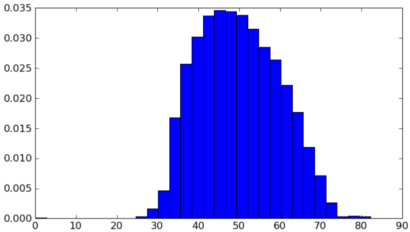
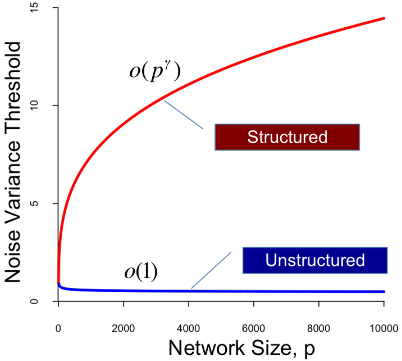


Figure: Eigenvalue Distribution

Big Picture



Tree	Interaction distance: $1 + \gamma \log_2 p$
Lattice	Dimensions: $d = 8\gamma \ln p$
ER	Edge probability: $p^{\gamma-1}$

Estimator Performance: Simulations

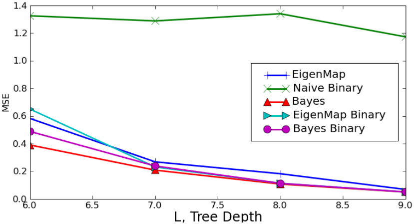


Figure: Tree Graph

Estimator Performance: Simulations

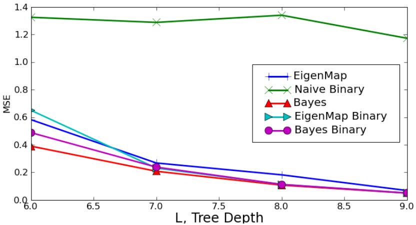


Figure: Tree Graph

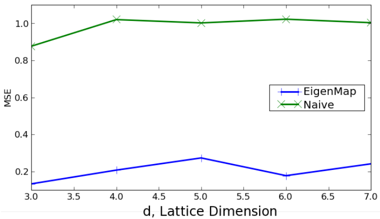


Figure: Lattice Graph

Estimator Performance: Simulations

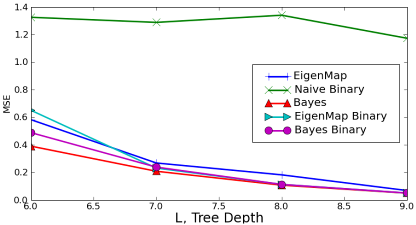


Figure: Tree Graph

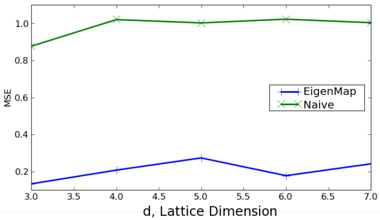


Figure: Lattice Graph

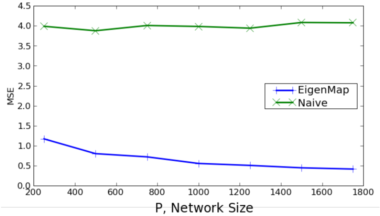


Figure: Erdös-Rényi Graph

Estimator Performance: Simulations

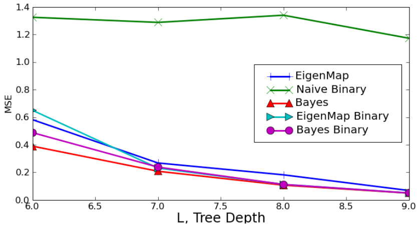


Figure: Tree Graph

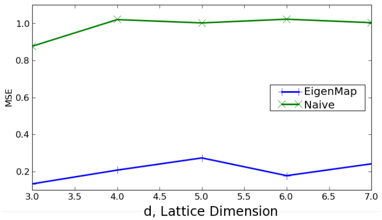


Figure: Lattice Graph

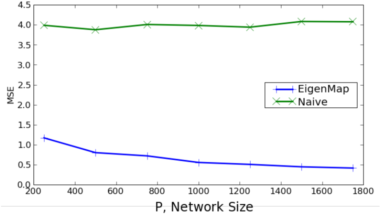


Figure: Erdös-Rényi Graph

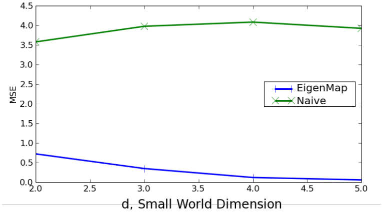


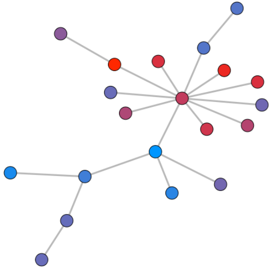
Figure: Small World Graph

Setup

Signal: $\mathbf{x} \sim f_{\mathbf{L}} d\nu$ with

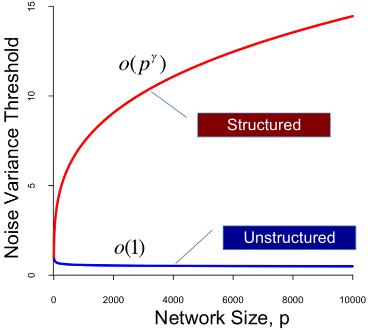
$$f_{\mathbf{L}}(X) \propto e^{-X^T \mathbf{L} X}$$

Observations: $\mathbf{y} = \mathbf{x} + \zeta$ with $\zeta \sim N(0, \sigma^2 I_p)$



Results

Estimator: $\widehat{\mathbf{x}}_k = \mathbf{U}_{[k]} \mathbf{U}_{[k]}^T \mathbf{y}$



- Hierarchical Graph
- Lattice Graph
- Random Graphs

Loss and Bayes Rules

What loss do we use?

Loss and Bayes Rules

What loss do we use?

GGM: Mean Square Error $\mathrm{MSE}(\widehat{\mathbf{x}}) = ||\mathbf{x} - \hat{\mathbf{x}}||^2$

Ising: Hamming, $d_H(\widehat{\mathbf{x}}', \mathbf{x})$, applies to binary estimators

note: $\mathbb{E}[d_H(\widehat{\mathbf{x}}', \mathbf{x})] = \mathrm{MSE}(\widehat{\mathbf{x}}') \leq 4\mathrm{MSE}(\widehat{\mathbf{x}})$ for $\widehat{\mathbf{x}}'_i = I\{\widehat{\mathbf{x}}_i > 1/2\}$

Loss and Bayes Rules

What loss do we use?

GGM: Mean Square Error $\mathrm{MSE}(\widehat{\mathbf{x}}) = ||\mathbf{x} - \hat{\mathbf{x}}||^2$

Ising: Hamming, $d_H(\widehat{\mathbf{x}}', \mathbf{x})$, applies to binary estimators

note: $\mathbb{E}[d_H(\widehat{\mathbf{x}}', \mathbf{x})] = \mathrm{MSE}(\widehat{\mathbf{x}}') \leq 4\mathrm{MSE}(\widehat{\mathbf{x}})$ for $\widehat{\mathbf{x}}'_i = I\{\widehat{\mathbf{x}}_i > 1/2\}$

Can't we calculate a posterior?

Loss and Bayes Rules

What loss do we use?

GGM: Mean Square Error $\mathrm{MSE}(\widehat{\mathbf{x}}) = ||\mathbf{x} - \hat{\mathbf{x}}||^2$

Ising: Hamming, $d_H(\widehat{\mathbf{x}}', \mathbf{x})$, applies to binary estimators

note: $\mathbb{E}[d_H(\widehat{\mathbf{x}}', \mathbf{x})] = \mathrm{MSE}(\widehat{\mathbf{x}}') \leq 4\mathrm{MSE}(\widehat{\mathbf{x}})$ for $\widehat{\mathbf{x}}'_i = I\{\widehat{\mathbf{x}}_i > 1/2\}$

Can't we calculate a posterior? (generalized normal)

$$\mathbf{x}|\mathbf{y} \sim \mathcal{GN}\left((2\sigma^2\mathbf{L} + \mathbf{I})^{-1}\mathbf{y}, \left(2\mathbf{L} + \sigma^{-2}\mathbf{I}\right)^{-1}, d\nu\right)$$

1 Posterior mean for Ising is difficult to calculate

2 No closed form makes asymptotic risk analysis difficult

Loss and Bayes Rules

What about the MAP estimate?

Loss and Bayes Rules

What about the MAP estimate?

- MAP minimizes the $0 - 1$ risk:

$$\widehat{X_{MAP}} = \min_{\hat{X}} \mathbb{E}\delta_{\{\hat{X}=\mathbf{x}\}}$$

- For the Ising model we can solve MAP efficiently with graph cuts.

GGM: MAP estimate = Posterior mean = Bayes optimal rule under MSE

Ising: MAP estimate $\neq$ Posterior mean

Loss and Bayes Rules

What about the MAP estimate?

- MAP minimizes the $0-1$ risk:

$$\widehat{X_{MAP}} = \min_{\hat{X}} \mathbb{E}\delta_{\{\hat{X}=\mathbf{x}\}}$$

- For the Ising model we can solve MAP efficiently with graph cuts.

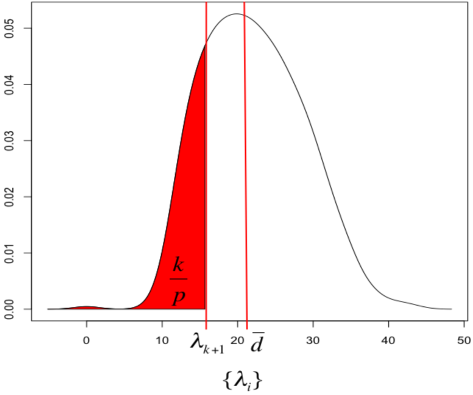
GGM: MAP estimate = Posterior mean = Bayes optimal rule under MSE

Ising: MAP estimate $\neq$ Posterior mean

MAP is not sufficient for the Ising model

The Bulk Spectrum

$$\text{Recall: } R_B \leq \frac{2}{\lambda_{k+1}} + \frac{k\sigma^2}{p} + e^{-p}$$



- Choose from $\{\lambda_i\}$ uniformly at random $\lambda_\bullet$

Modeling and Inference

Dynamics OF Networks

Random Graph Models

- Erdös-Rényi graph [Erdös & Rényi '60, Bollobas '01]
- ERGMs [Rinaldo, Fienberg, Zhou '09, Kolacyzk '09]
- Real-World Graphs [Watts & Strogatz '98, Barabasi & Albert '99]

Community Detection [Bickel & Chen '09, Newman & Girvan '04]

Evolving networks [Durrett '06]

Manifold Sampling [Belkin & Niyogi '08]

Dynamics ON Networks

Graphical Models [Wasserman '03]

- Ising Model [Ising '25], Glauber Dynamics [Martinelli '97]
- Gaussian Graphical Models [Koller & Friedman '09]

Infection Models [Zhou et al '05, Boguna '02]

Signal Estimation

- Estimation [Coifman '06, Lee at al '08]
- Detection [Singh at al '10, Arias-Castro at al '10]

Graph Laplacian

Graph $G = (V, E, W)$ with $D_{i,i} = d_i = \sum_j W_{i,j}$ then

$$\mathbf{L} = D - W$$

Define eigenvalue, eigenvector pairs $\{\lambda_i, u_i\}$ of $\mathbf{L}$ with $\lambda_i \leq \lambda_{i+1}$

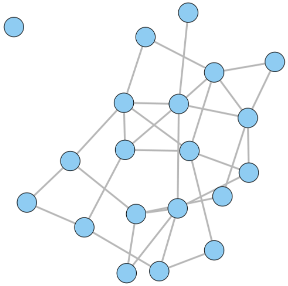


Figure: A Random Graph

- $\mathbf{x}^T L \mathbf{x} = \sum_{i \sim j} W_{i,j} (\mathbf{x}_i - \mathbf{x}_j)^2$
- $\lambda_0 = 0$ and $u_0 = \vec{1}$
- $\sum_i \lambda_i = \sum_i d_i$
- Spectral clustering: thresholding first eigenvector [Shi & Malik '00]
- Dimension reduction: projection to first few generalized eigenvectors [Belkin & Niyogi '02, Ng et al '01]

Proof pt. 1: Chernoff Bound

Lemma

Let $\mathbf{x}$ *be drawn from an Ising model with Laplacian* $\mathbf{L}$ *and* p *nodes.*

$$\mathbb{P}\{\mathbf{x}^T\mathbf{L}\mathbf{x} > \delta p\} \leq e^{-p}$$

for any $\delta \in (1 + \log(2), 2]$

- strategic use of Markov's inequality
- essential that $\mathbf{L}\vec{1} = 0$

Proof pt. 2: Minimax Risk

Lemma

Let $\{\lambda_i\}_{i=1}^{p}$ be eigenvalues of the Laplacian $\mathbf{L}$, with $\lambda_i \le \lambda_{i+1}$. For any $\mathbf{x} \in \mathbb{R}^p$ such that $\mathbf{x}^T \mathbf{L} \mathbf{x} < \delta p$,

$$\mathbb{E}\left(\frac{1}{p}||\widehat{\mathbf{x}}_k - \mathbf{x}||^2 | \mathbf{x}\right) \le \min\left(1, \frac{\delta}{\lambda_{k+1}}\right) + \frac{k\sigma^2}{p}$$

- Set up primal problem of maximizing $||\mathcal{P}_{U_{[p]}^{\perp}}\mathbf{x}||^2$ subject to constraints
- Low dimensional projection reduces variance

Hierarchical Structure: Bulk Spectrum

Lemma (Ogielski & Stein '85)

For the hierarchical structure with L levels, the ℓ^{th} smallest unique eigenvalue ($\ell \in [L]$) is $2^{\ell-1}$-fold degenerate and given as

$$\lambda_\ell = \sum_{i=L-\ell+1}^{L} 2^{i-1}\epsilon_i + 2^{L-\ell}\epsilon_{L-\ell+1}$$

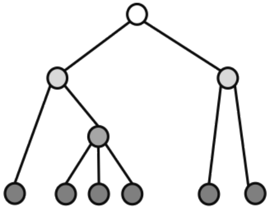


Figure: Hierarchical Graph

Figure: Hierarchical Weight Matrix

See also: Singh at al. *Detecting Weak but Hierarchically-Structured Patterns in Networks*, '10

Hierarchical Structure: Consistency Region

Corollary

If $\epsilon_\ell = 2^{-\ell(1-\beta)}\ \forall \ell \leq \gamma \log_2 p + 1$*, for constants* $\gamma, \beta \in (0,1)$*, and* $\epsilon_\ell = 0$ *otherwise, then the noise threshold for consistent MSE recovery* $(R_B = o(1))$ *is*

$$\sigma^2 = o(p^\gamma).$$

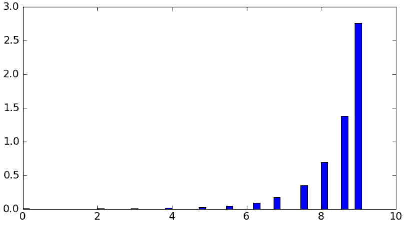


Figure: Eigenvalue Distribution

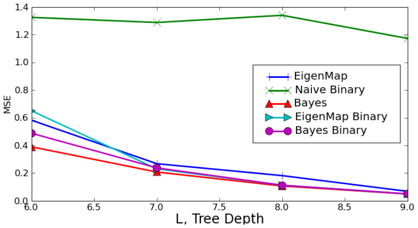


Figure: Estimator Performance

Lattice: Bulk Spectrum

Lemma

Let $\lambda_\bullet^{\mathbf{L}}$ be an eigenvalue of the Laplacian, $\mathbf{L}$, of the lattice graph in d dimensions with $p = n^d$ vertices, chosen uniformly at random. Then

$$\mathbb{P}\{\lambda_\bullet^{\mathbf{L}} \leq d\} \leq \exp\{-d/8\}. \quad (2)$$

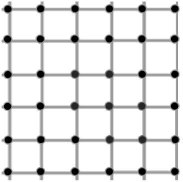


Figure: Lattice Graph

Lattice in d-dimensions:
$i = (i_1, ..., i_d), j = (j_1, ..., j_d) \in [n]^d$

$$\begin{aligned} W_{i,j} &= w_{i_1,j_1}\delta_{i_2,j_2}...\delta_{i_d,j_d} + \\ &... +w_{i_d,j_d}\delta_{i_1,j_1}...\delta_{i_{d-1},j_{d-1}} \end{aligned}$$

- Tensor product of 1-D lattice
- Hoeffding's on eigenvalues

Corollary

If n is a constant, $p = n^d$ and $d = 8\gamma \ln p$, for some constant $\gamma \in (0,1)$, then the noise threshold for consistent MSE recovery ($R_B = o(1)$) is given as:

$$\sigma^2 = o(p^\gamma)$$

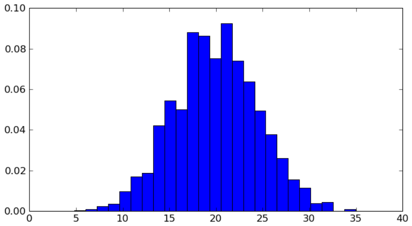


Figure: Eigenvalue Distribution

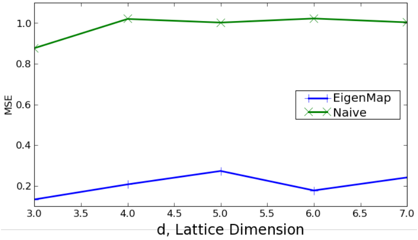


Figure: Estimator Performance

Lemma

For any α_p increasing in p,

$$\mathbb{P}_G\{\mathbb{P}_\bullet\{\lambda_\bullet \leq p^\gamma/2 - p^{\gamma-1}\} \geq \alpha_p p^{-\gamma}\} = \mathcal{O}(1/\alpha_p) \tag{3}$$

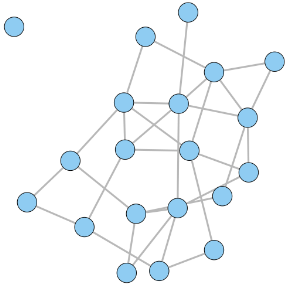


Figure: Erdös-Rényi Graph

- Probability of edge $= p^{\gamma-1}$
- $\mathbb{P}_G$: random graph measure
- $\mathbb{P}_\bullet$: random eigenvalue index
- $\mathbf{L} = (\bar{d}\mathbf{I} - \mathbf{W}) + (\mathbf{D} - \bar{d}\mathbf{I})$ with Wielandt-Hoffman thm.
- $\lambda^{\mathbf{W}}$ semi-circular dist.

Erdös-Rényi Graph: Consistency Region

Corollary

Define consistent MSE recovery to be $R_B = o_{\mathbb{P}_G}(1)$,

$$\sigma^2 = o(p^\gamma).$$

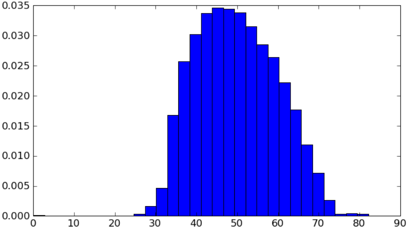


Figure: Eigenvalue Distribution

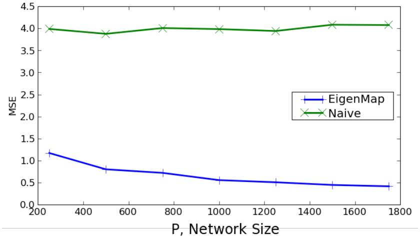


Figure: Estimator Performance

Real-World Graphs

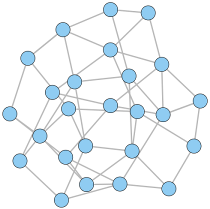


Figure: Small World Graph

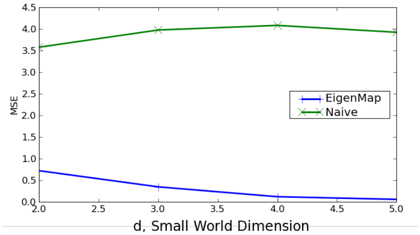


Figure: Estimator Performance

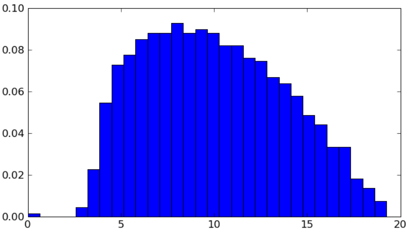


Figure: Eigenvalue Distribution

- Small world graph: proof similar to ER graph
- Scale-free (power law) graph [Chung et al '03]