IDENTIFYING GRAPH-STRUCTURED ACTIVATION PATTERNS IN NETWORKS

James Sharpnack\textsuperscript{1,2}  Aarti Singh\textsuperscript{1}

\textsuperscript{1}Machine Learning Dept.
CMU, Pittsburgh, PA

\textsuperscript{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

\[ x \in \mathbb{R}^p \text{ (or } x \in \{0, 1\}^p) \]
Normal Means Estimation

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

\[ \mathbf{y} = \mathbf{x} + \zeta, \quad \zeta \sim \mathcal{N}(0, \sigma^2 I_p) \]

Task: reconstruct \( \mathbf{x} \) from \( \mathbf{y} \)
Structured Normal Means Estimation

$x \in \mathbb{R}^p$ (or $x \in \{0, 1\}^p$)
**Structured Normal Means Estimation**

\[ x \in \mathbb{R}^p \text{ (or } x \in \{0, 1\}^p \text{ )} \]

Graph: \( G = (V, E, W) \)
Structured Normal Means Estimation

Noisy observations

Noisy observations with structure

Task: reconstruct $x$ from $y$ exploiting dependencies (given by $G$)
Statistical Model

The Model

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

The Model

1. Graph: $G \sim G_p$ with $p$ nodes.

2. Signal: $\mathbf{x} \sim f_L d\nu$ with

\[
 f_L(\mathbf{x}) \propto e^{-\mathbf{x}^T L \mathbf{x}}
\]

GGM: $\nu$ is Lebesgue ($\Sigma^{-1} = L$)
Ising: $\nu$ is Counting
$L = D - W$
$\mathbf{x}^T L \mathbf{x} = \sum_{i \sim j} W_{i,j} (x_i - x_j)^2$
Statistical Model

The Model

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

2. Signal: $x \sim f_L d\nu$ with

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

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

3. Observations: draw iid. noise
$\zeta \sim N(0, \sigma^2 I_p)$
$$y = x + \zeta$$
Estimation of Graph-structured patterns

Bayes Optimal Rules:
Mean square error: posterior mean
Hamming distance: posterior centroid
0/1-loss: posterior max (MAP)

\{ \text{hard to implement} \}

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
- Hamming distance: posterior centroid
- 0/1-loss: posterior max (MAP)  \{ \text{hard to implement} \}

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, \( L \), with \( \lambda_i \leq \lambda_{i+1} \)

Estimator of \( x \) given \( k \in \{1, \ldots, p\} \):

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

1. Easy to analyze asymptotic risk
2. Easy to implement
Laplacian Eigenmaps Estimator

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

Estimator of \( x \) given \( k \in \{1, \ldots, p\} \):

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

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
Laplacian Eigenmaps Estimator

Hierarchical Graph

Hierarchical L

Haar Wavelet

Lattice Graph

Lattice L

Fourier Basis
Laplacian Eigenmaps Estimator

Network activation pattern: $\mathbf{x}$
Noisy observations: $y \left( \sigma^2 = \frac{1}{2} \right)$
Laplacian Eigenmaps Estimator

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

Eigenmaps estimator: \( \hat{x} (k = 3) \)
Large real-world graph ($p = 100$)
Laplacian Eigenmaps Estimator

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

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

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

Thresholded observations: $y > \tau$

Thresholded eigenmaps estimator: $\hat{x} > \tau$
Consistent estimation: $R_B = \mathbb{E}_x \frac{1}{p} \left| \left| \hat{x}_k - x \right| \right|^2 \xrightarrow{p \to \infty} 0$
Consistent estimation: \[ R_B = \mathbb{E}_x \frac{1}{p} \| \hat{x}_k - x \|^2 \to 0 \quad (p \to \infty) \]

Tolerable noise: \[ \sigma^2 = o(p^\gamma) \Rightarrow \text{consistent estimation} \]

\( \gamma \) depends on the network evolution model.
Main Result

Theorem

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

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

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

**Theorem**

Let $x$ be drawn from the Ising with graph Laplacian $L$.

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

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

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

Theorem

Let $x$ be drawn from the Ising with graph Laplacian $L$.

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

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

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

Theorem

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

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

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

\( R_B \leq \) concentration bound + bias + variance
Main Result

Theorem

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

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

where \( 0 < \delta < 2 \) is a constant and \( \lambda_{k+1} \) 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 $\mathbf{L}$.

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

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

$R_B \leq$ concentration bound + bias + variance

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

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

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

- Chernoff type bound $\Rightarrow$
  concentration of prior

$$x^T L x < 2p$$

w.p. $1 - e^{-p}$
Eigenmaps Geometry

$$\hat{x} = U_k U_k^T y = U_k U_k^T x + U_k U_k^T \zeta$$

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

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

\[ \hat{x} = U_k U_k^T y = U_k U_k^T x + U_k U_k^T \zeta \]

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

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

\[ x^T L x < 2p \quad \text{w.p.} \ 1 - e^{-p} \]
Recall: \( R_B \leq \frac{2}{\lambda_{k+1}} + \frac{k\sigma^2}{p} + e^{-p} \)

Goal: for simple graph models \( G_p \) what is \( \gamma \)?
Hierarchical Structure: Eigenvalue Concentration

Lemma (Ogielski & Stein ’85)

For the hierarchical structure with interaction strength, $\beta$, 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

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

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

\[ \#\{\lambda_{\ell} < 2^{\beta\ell^* - 1}\} \leq 2^{\log_2 p - \ell^* + 1} \]
\[ \ell^* = 1 + \gamma \log_2 p \]
Set \( k = 2^{\log_2 p - \ell^* + 1} = p^{1 - \gamma} \)
\[ \frac{1}{\lambda_{k+1}} \leq 2^{1 - \beta \ell^*} \]
\[ \sigma^2 = o(p^\gamma) \Rightarrow R_B \to 0 \]

Figure: Eigenvalue Histogram
Lattice: Eigenvalue Concentration

Lemma

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

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

Figure: Lattice Graph

Figure: Eigenvalue Histogram
LATTICE: RISK CONSISTENCY

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

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

**Figure: Eigenvalue Histogram**

**Figure: Bias Var Trade-off**

lattice dimension = 5
Lattice: Risk Consistency

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

\[ \#\{\lambda_i^L \leq d\} \leq p \exp\{-d/8\} \]

\[ d = 8\gamma \ln p \]

Set \( k = p \exp\{-d/8\} = p^{1-\gamma} \)

\[ 1/\lambda_{k+1} \leq 1/d \]

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

Figure: Eigenvalue Histogram
**Erdös-Rényi Graph**

**Lemma**

Let the probability of an edge be $p^{\gamma-1}$. For any $\alpha_p$ 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: $p^{\gamma-1}$

![Graph showing noise variance threshold vs. network size with structured and unstructured regions.]
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:** Erdős-Rényi Graph

**Figure:** Small World Graph
**Summary**

**Setup**

Signal: \( x \sim f_L d\nu \) with

\[
f_L(X) \propto e^{-X^T L X}
\]

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

![Graphs](image)

**Results**

Estimator: \( \hat{x}_k = U_{[k]} U_{[k]}^T y \)

![Graphs](image)

- 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 \( \text{MSE}(\hat{x}) = ||x - \hat{x}||^2 \)

**Ising:** Hamming, \( d_H(\hat{x}', x) \), applies to binary estimators

**Note:** \( \mathbb{E}[d_H(\hat{x}', x)] = \text{MSE}(\hat{x}') \leq 4\text{MSE}(\hat{x}) \) for \( \hat{x}'_i = I\{\hat{x}_i > 1/2\} \)
Loss and Bayes Rules

What loss do we use?

**GGM:** Mean Square Error \( \text{MSE}(\hat{x}) = ||x - \hat{x}||^2 \)

**Ising:** Hamming, \( d_H(\hat{x}', x) \), applies to binary estimators

**Note:** \( \mathbb{E}[d_H(\hat{x}', x)] = \text{MSE}(\hat{x}') \leq 4\text{MSE}(\hat{x}) \) for \( \hat{x}'_i = l\{\hat{x}_i > 1/2\} \)

Can’t we calculate a posterior?
Loss and Bayes Rules

What loss do we use?

**GGM:** Mean Square Error $\text{MSE}(\hat{x}) = ||x - \hat{x}||^2$

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

**Note:** $\mathbb{E}[d_H(\hat{x}', x)] = \text{MSE}(\hat{x}') \leq 4\text{MSE}(\hat{x})$ for $\hat{x}'_i = I\{\hat{x}_i > 1/2\}$

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

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

1. Posterior mean for Ising is difficult to calculate
2. No closed form makes asymptotic risk analysis difficult
What about the MAP estimate?

- MAP minimizes the $0-1$ risk:
  $$\hat{X}_{MAP} = \min_{\hat{X}} \mathbb{E}_{\delta} \{ \hat{X} = x \}$$

- 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
What about the MAP estimate?

- MAP minimizes the $0 - 1$ risk:
  $$\hat{X}_{\text{MAP}} = \min_{\hat{X}} \mathbb{E}_x \delta\{\hat{X} = x\}$$

- 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
What about the MAP estimate?

- MAP minimizes the $0 - 1$ risk:
  \[ \hat{X}_{MAP} = \min_{\hat{X}} \mathbb{E}_{\delta \{\hat{X} = x\}} \]

- 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

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

- Choose from \( \{\lambda_i\} \) uniformly at random \( \lambda_\bullet \).
Dynamics OF Networks

Random Graph Models
- Erdős-Rényi graph [Erdős & Rényi ’60, Bollobás ’01]
- ERGMs [Rinaldo, Fienberg, Zhou ’09, Kolaczyk ’09]

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

$$L = D - W$$

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

- $x^T L x = \sum_{i \sim j} W_{i,j} (x_i - x_j)^2$
- $\lambda_0 = 0$ 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]

Figure: A Random Graph
**Lemma**

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

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

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

- strategic use of Markov’s inequality
- essential that $\bm{L} \mathbf{1} = 0$
Proof pt. 2: Minimax Risk

**Lemma**

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

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

- 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 $\ell^{th}$ smallest unique eigenvalue ($\ell \in [L]$) is $2^{\ell-1}$-fold degenerate and given as

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

See also: Singh at al. Detecting Weak but Hierarchically-Structured Patterns in Networks, ’10
Hierarchical Structure: Consistency Region

Corollary

If $\ell \epsilon = 2^{-\ell(1-\beta)} \forall \ell \leq \gamma \log_2 p + 1$, 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
Lemma

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

$$\mathbb{P}\{\lambda^L \leq d\} \leq \exp\{-d/8\}.$$  (2)

Lattice in $d$-dimensions:

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

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

- 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 $\alpha_p$ increasing in $p$,

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

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

**Figure:** Erdős-Rényi Graph
Corollary

Define consistent MSE recovery to be $R_B = o_p(1)$,

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

**Figure:** Eigenvalue Distribution

**Figure:** Estimator Performance
Real-World Graphs

**Figure:** Small World Graph

**Figure:** Eigenvalue Distribution

**Figure:** Estimator Performance

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