

# Numerical Methods for Data Science: Spectral Network Analysis, Part II

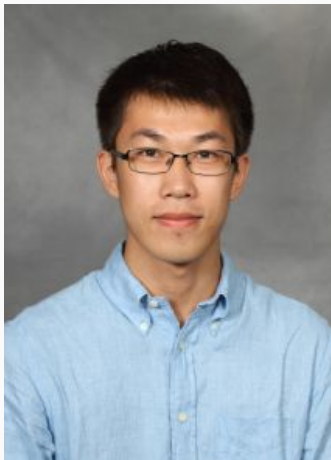
---

David Bindel

21 June 2019

Department of Computer Science  
Cornell University

# Acknowledgements



Thanks **Kun Dong** (Cornell CAM), along with Anna Yesypenko, Moteleolu Onabajo, Jianqiu Wang.

Also: NSF DMS-1620038.

*Mathematicians are like Frenchmen: whatever you say to them they translate into their own language and forthwith it is something entirely different.*

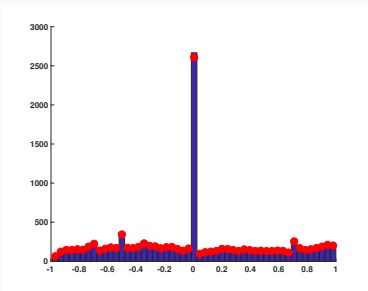
— Goethe

# Stories of Spectral Geometry/Graph Theory

- Dynamics (operator on functions over manifold or graph)
  - Example: Continuous time diffusion + Simon-Ando theory
  - Diffuse according to heat kernel  $\exp(-tLD^{-1})$
  - Rapid mixing + slow equilibration across bottlenecks
  - Regions of rapid mixing via extreme eigenvectors
- Counting and measure (quadratic form)
  - Example: Spectral partitioning
  - Measure cut sizes via  $x^T Lx/4$ ,  $x \in \{\pm 1\}^n$ , relax to  $\mathbb{R}^n$
  - $\lambda_2(L)$  bounds cuts (Cheeger), partition with Fiedler vector
- Geometric embedding (pos semidef form / kernel)
  - Example: Spectral coordinates via  $R = L^\dagger$
  - Diffusion distances:  $d_{ij}^2 = r_{ii}^2 - 2r_{ij} + r_{jj}^2$
  - First few eigenvectors of  $R$  give coordinates that approximate distance

# Eigenvalues Two Ways

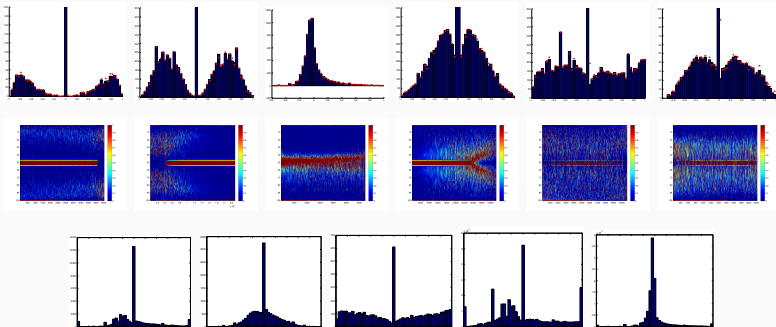
What can we tell from *partial* spectral information (eigenvalues and/or vectors)



Claim: Most spectral analyses involve one of two perspectives:

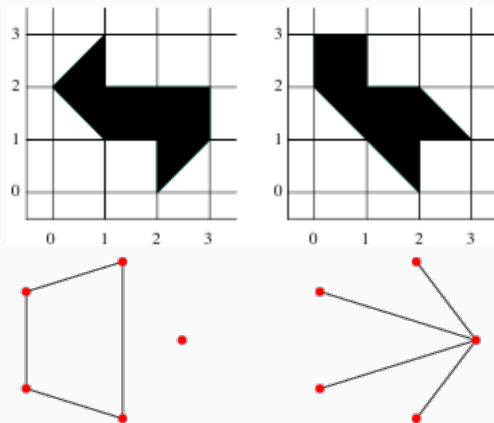
- Approximate something via a *few* (extreme) eigenvalues.
- Look at *all* the eigenvalues (or all in a range).

# Today in Three Acts



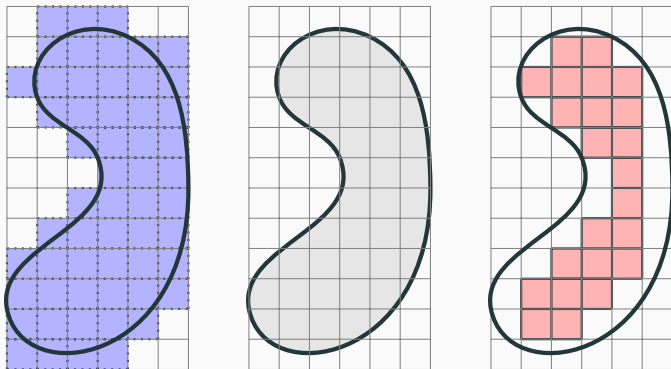
- Act 1: Spectral densities
- Act 2: Algorithms and approximations
- Act 3: Spectral densities of “real world” graphs

# Can One Hear the Shape of a Drum?



*“You mean, if you had perfect pitch could you find the shape of a drum.”* — Mark Kac (quoting Lipmann Bers)  
American Math Monthly, 1966

# What Do You Hear?

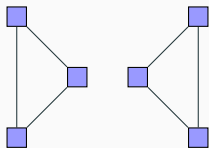
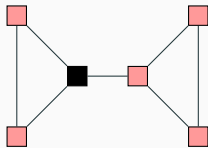
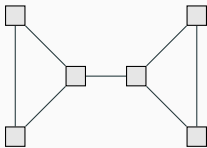
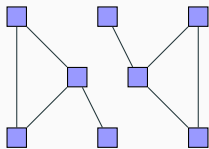


Use  $\mathcal{H}_{lo} \supset \mathcal{H} \supset \mathcal{H}_{hi}$  to get  $\lambda_{k,lo} \leq \lambda_k \leq \lambda_{k,hi}$

$$\lambda_k = \min_{\dim(\mathcal{V})=k, \mathcal{V} \subset \mathcal{H}} \left( \max_{v \in \mathcal{V}} \rho_L(v) \right)$$



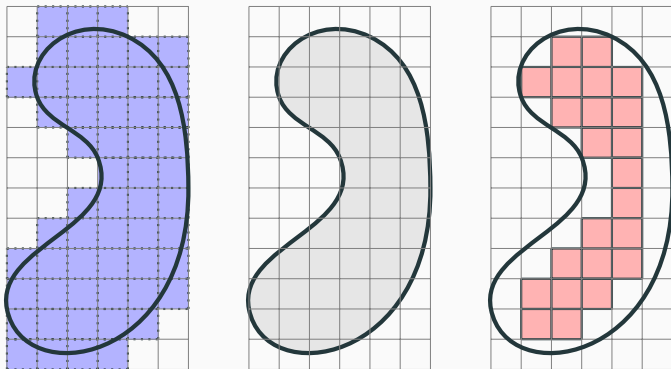
# What Do You Hear?



Use  $\mathcal{H}_{\text{lo}} \supset \mathcal{H} \supset \mathcal{H}_{\text{hi}}$  to get  $\lambda_{k,\text{lo}} \leq \lambda_k \leq \lambda_{k,\text{hi}}$

$$\lambda_k = \min_{\dim(\mathcal{V})=k, \mathcal{V} \subset \mathcal{H}} \left( \max_{v \in \mathcal{V}} \rho_L(v) \right)$$

# What Do You Hear?



Weyl law (asymptotics for  $N(x) = \{\# \text{ eigenvalues } \leq x\}$ ):

$$\lim_{x \rightarrow \infty} \frac{N(x)}{x^{d/2}} = (2\pi)^{-d} \omega_d \text{vol}(\Omega).$$

# What Do You Hear?

What information hides in the eigenvalue distribution?

1. Discretizations of Laplacian: something like Weyl's law
2. Sparse E-R random graphs: Wigner semicircular law
3. Some other random graphs: Wigner semicircle + a bit  
(Farkas *et al*, Phys Rev E (64), 2001)
4. "Real" networks: less well understood

But computing all eigenvalues seems *expensive*!

# A Bestiary of Matrices

- Adjacency matrix:  $A$
- Laplacian matrix:  $L = D - A$
- Unsigned Laplacian:  $L = D + A$
- **Random walk matrix:**  $P = AD^{-1}$  (or  $D^{-1}A$ )
- **Normalized adjacency:**  $\bar{A} = D^{-1/2}AD^{-1/2}$
- **Normalized Laplacian:**  $\bar{L} = I - \bar{A} = D^{-1/2}LD^{-1/2}$
- Modularity matrix:  $B = A - \frac{dd^T}{2n}$
- Motif adjacency:  $W = A^2 \odot A$

All have examples of co-spectral graphs

... through spectrum uniquely identifies *quantum graphs*

## Reminder: Spectral Mapping

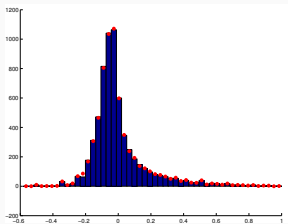
Consider a matrix  $H$ , and let  $f$  be analytic on the spectrum.

Then if  $H = V\Lambda V^{-1}$ ,

$$f(H) = Vf(\Lambda)V^{-1}.$$

(generalizes to non-diagonalizable case)

## Another Perspective: Density of States



Spectra define a *generalized function* (a density):

$$\mathrm{tr}(f(H)) = \int f(\lambda)\mu(\lambda) dx = \sum_{j=k}^N f(\lambda_k)$$

where  $f$  is an analytic test function. Smooth to get a picture: a *spectral histogram* or *kernel density estimate*.

## Example: Estrada Index

Consider

$$\text{tr}(\exp(\alpha A)) = \sum_{k=1}^{\infty} \frac{\alpha^k}{k!} \cdot (\# \text{ closed random walks of length } k).$$

- Global measure of connectivity in a graph.
- Can clearly be computed via DoS.
- Generalizes to other weights.

DoS information equivalent to looking at the *heat kernel trace*:

$$h(s) = \text{tr}(\exp(-sH)) = \mathcal{L}[\mu](s)$$

Use  $H = LD^{-1}$  (continuous time random walk generator)  $\implies$   
 $h(s)/N = P(\text{self-return after time } s \text{ from uniform start}).$



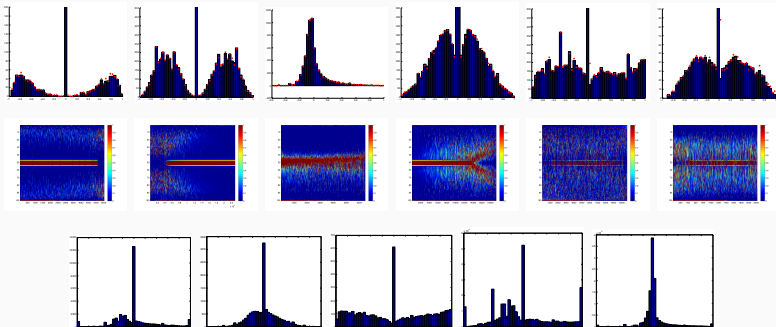
DoS information equivalent to looking at the *power moments*:

$$\text{tr}(H^j).$$

Natural interpretation for matrices associated with graphs:

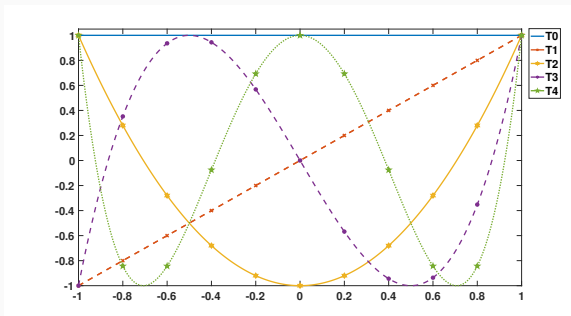
- $A$ : number of length  $k$  cycles.
- $\bar{A}$  or  $P$ : return probability for  $k$ -step random walk (times  $N$ ).
- $L$ : ??

# Today in Three Acts



- Act 1: Spectral densities
- **Act 2: Algorithms and approximations**
- Act 3: Spectral densities of “real world” graphs

# Chebyshev Moments



For numerics, prefer Chebyshev moments to power moments:

$$d_j = T_j(A)$$

where  $T_j(z) = \cos(j \cos^{-1}(z))$  is the  $j$ th Chebyshev polynomial:

$$T_0(z) = 1, \quad T_1(z) = z, \quad T_{k+1}(z) = 2zT_k(z) - T_{k-1}(z).$$

# Exploring Spectral Densities

Kernel polynomial method (see Weisse, Rev. Modern Phys.)

- Spectral distribution on  $[-1, 1]$  is a generalized function:

$$\int_{-1}^1 \mu(x) f(x) dx = \frac{1}{N} \sum_{k=1}^N f(\lambda_k)$$

- Write  $f(x) = \sum_{j=1}^{\infty} c_j T_j(x)$  and  $\mu(x) = \sum_{j=1}^{\infty} d_j \phi_j(x)$ , where  $\int_{-1}^1 \phi_j(x) T_k(x) dx = \delta_{jk}$
- Estimate  $d_j = \text{tr}(T_j(H))$  by stochastic methods
- Truncate series for  $\mu(x)$  and filter (avoid Gibbs)

*Much* cheaper than computing all eigenvalues!

Alternatives: Lanczos (Golub-Meurant), maxent (Röder-Silver)

# Stochastic Trace and Diagonal Estimation

$Z \in \mathbb{R}^n$  with independent entries, mean 0 and variance 1.

$$E[(Z \odot HZ)_i] = \sum_j h_{ij} E[Z_i Z_j] = h_{ii}$$

$$\text{Var}[(Z \odot HZ)_i] = \sum_j h_{ij}^2.$$

Serves as the basis for stochastic estimation of

- Trace (Hutchinson, others; review by Toledo and Avron)
- Diagonal (Bekas, Kokiopoulou, and Saad)

Independent probes  $\implies 1/\sqrt{N}$  convergence (usual MC).

## Beyond Independent Probes

For probes  $Z = [Z_1, \dots, Z_s]$ , have *exact* diagonal

$$d = [(A \odot ZZ^T)e] \oslash [(Z \odot Z)e]$$

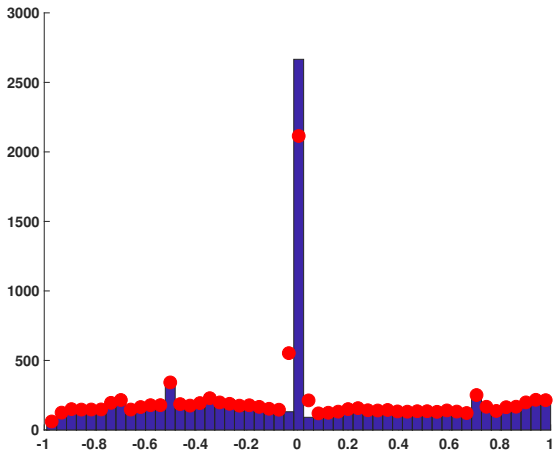
if  $Z_{i,:} Z_{j,:}^T = 0$  whenever  $A_{ij} \neq 0$ .

**Idea:**

- Pick rows  $\{Z_{i,:}\}$  such that  $Z_{i,:} \perp Z_{j,:}$  whenever  $A_{ij} \neq 0$
- $A$  an adjacency matrix  $\implies$  graph coloring.

Combined with randomization, still gives unbiased estimates.

## Example: PGP Network



Spike (non-smoothness) at eigenvalues of 0 leads to inaccurate approximation.

## Motifs and Symmetry

Suppose  $PH = HP$ . Then

$\mathcal{V}$  a max invariant subspace for  $P \implies$

$\mathcal{V}$  a max invariant subspace for  $H$

So *local symmetry*  $\implies$  *localized eigenvectors*.

Simplest example:  $P$  swaps  $(i, j)$

- $e_i - e_j$  an eigenvector of  $P$  with eigenvalue  $-1$
- $e_i - e_j$  an eigenvector of  $\bar{A}$  with eigenvalue

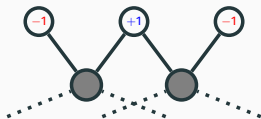
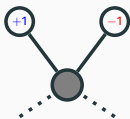
$$\lambda = \rho_{\bar{A}}(e_i - e_j) = \begin{cases} d^{-1}, & (i, j) \in \mathcal{E} \\ 0, & \text{otherwise.} \end{cases}$$

- All other eigenvectors (eigenvalue  $-1$ ) satisfy  $v_i = v_j$

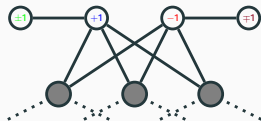
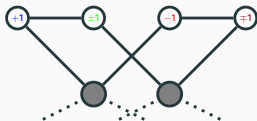


# Motifs in Spectrum

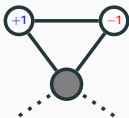
- $\lambda = 0$



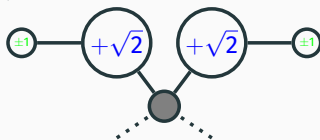
- $\lambda = \pm 1/2$



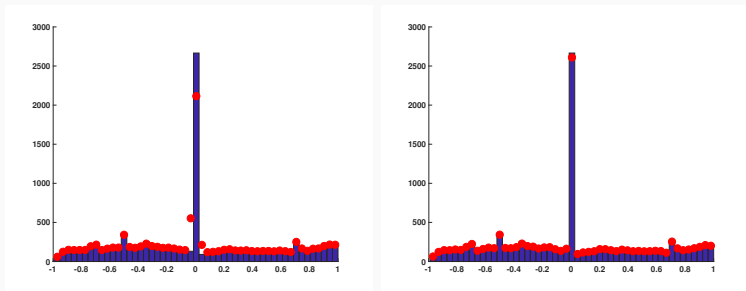
- $\lambda = -1/2$



$$\lambda = \pm 1/\sqrt{2}$$



# Motif Filtering



Motif “spikes” slow convergence – deflate motif eigenvectors!

If  $P \in \mathbb{R}^{n \times m}$  an orthonormal basis for the quotient space,

- Apply estimator to  $P^T \bar{A} P$  to reduce size for  $m \ll n$ .
- or use  $Proj_P(Z)$  to probe the desired subspace.

## Diagonal Estimation and LDoS

Diagonal estimation also useful for *local* DoS  $\nu_k(x)$ ;  
in the symmetric case with  $H = Q\Lambda Q^T$ , have

$$\int f(x)\nu_k(x) dx = f(H)_{kk} = e_k^T Q f(\Lambda) Q^T e_k$$

$$\nu_k(x) = \sum_{j=1}^n q_{kj}^2 \delta(x - \lambda_j)$$

DoS is sum of local densities of states:

$$\mu(x) = \sum_{k=1}^n \nu_k(x)$$

Same game, different moments:

- Estimate  $d_j = [T_j(H)]_{kk}$  by diag estimation
- Truncate series for  $\mu(x)$  and filter (avoid Gibbs)

Diagonal estimator gives moments *for all  $k$  simultaneously!*

Alternatives: Lanczos (Golub-Meurant), maxent (Röder-Silver)

Can compute common *centrality measures* with LDoS

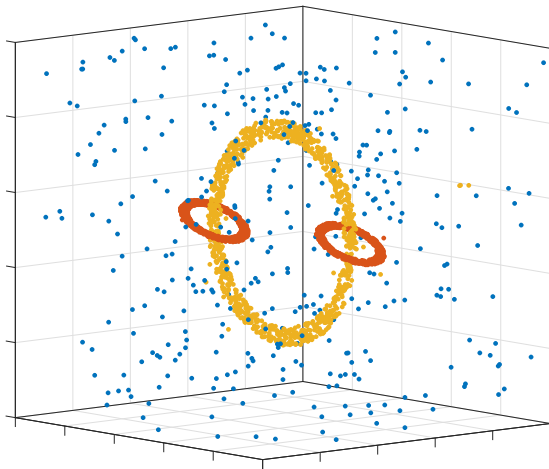
- Estrada centrality:  $\exp(\gamma A)_{kk}$
- Resolvent centrality:  $[(I - \gamma \bar{A})^{-1}]_{kk}$

Some motifs associated with localized eigenvectors:

- Chief example: Null vectors of  $\bar{A}$  supported on leaves.
- Use LDoS + topology to find motifs?

What else?

# LDoS and Clustering



# Phase Retrieval in Graph Reconstruction

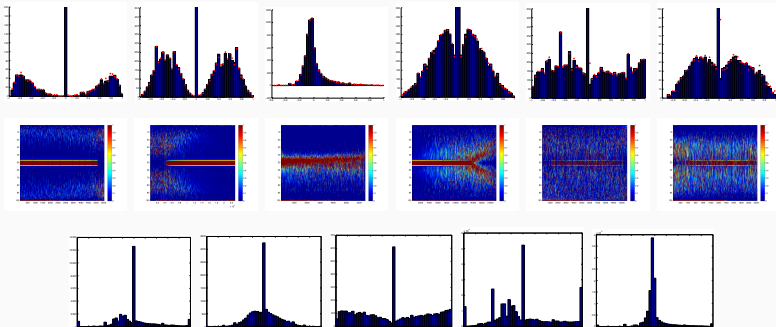
Reconstruct graph from *fully resolved* LDoS at all nodes?

- Assume  $H = Q\Lambda Q^T$
- No multiple eigenvalues  $\implies$  know  $|Q|$  and  $\Lambda$
- Can we recover signs in  $Q$ ?

Feels a little like phase retrieval...

Of course, we usually have noisy LDoS estimates!

# Today in Three Acts



- Act 1: Spectral densities
- Act 2: Algorithms and approximations
- Act 3: Spectral densities of “real world” graphs



## Exploring Spectral Densities (with David Gleich)

- Compute spectrum of normalized Laplacian / RW matrix
- Compare KPM to full eigencomputation

Things we know

- Eigenvalues in  $[-1, 1]$ ; nonsymmetric in general
- Stability: change  $d$  edges, have

$$\lambda_{j-d} \leq \hat{\lambda}_j \leq \lambda_{j+d}$$

- $k$ th moment =  $P$ (return after  $k$ -step random walk)
- Eigenvalue cluster near 1  $\sim$  well-separated clusters
- Eigenvalue cluster near -1  $\sim$  bipartite structure
- Eigenvalue cluster near 0  $\sim$  leaf clusters

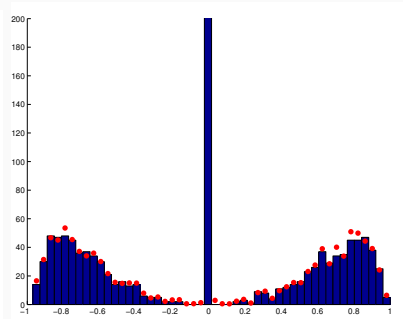
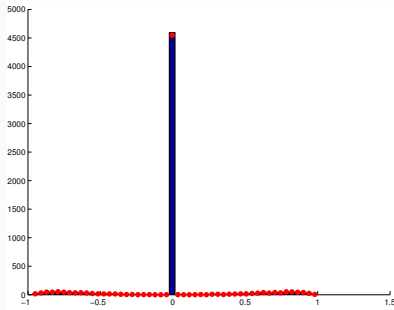
What else can we “hear”?

# Experimental setup

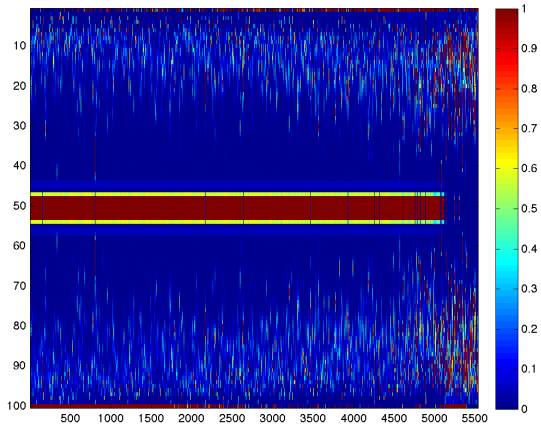
- Global DoS
  - 1000 Chebyshev moments
  - 10 probe vectors (componentwise standard normal)
  - Histogram with 50 bins
- Local DoS
  - 100 Chebyshev moments
  - 10 probe vectors (componentwise standard normal)
  - Plot smoothed density on  $[-1, 1]$
  - Spectrally order nodes by density plot

Suggestions for better pics are welcome!

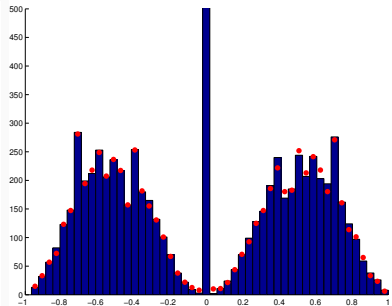
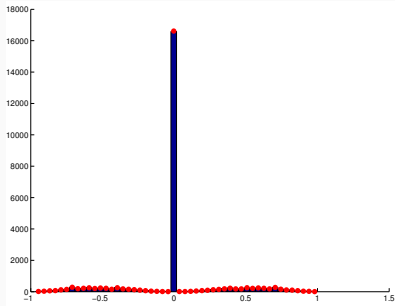
# Erdos



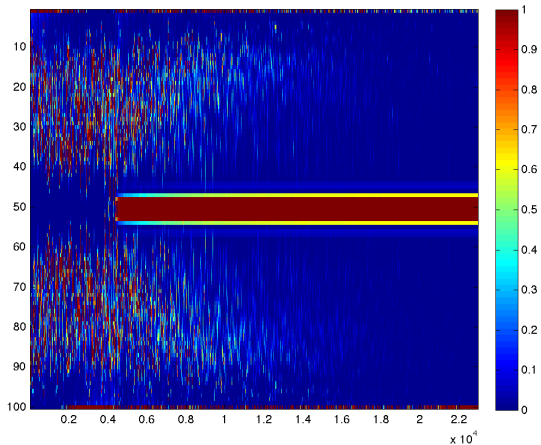
# Erdos (local)



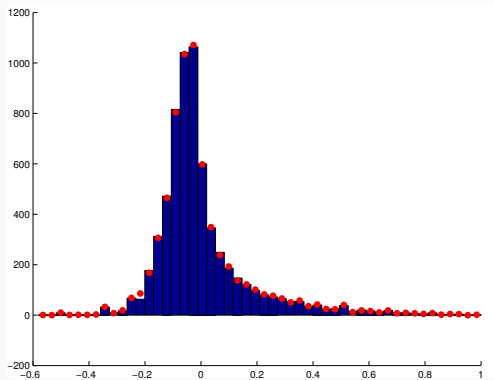
# Internet topology



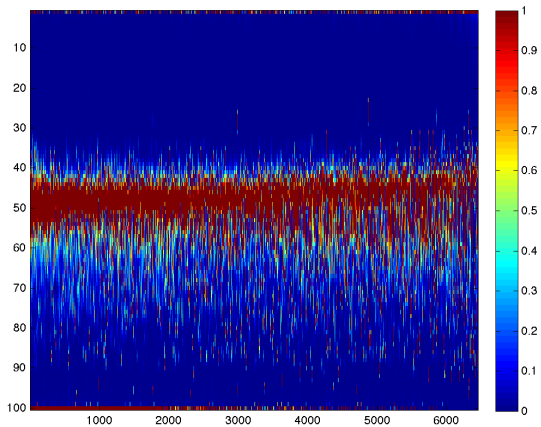
# Internet topology (local)



# Marvel characters

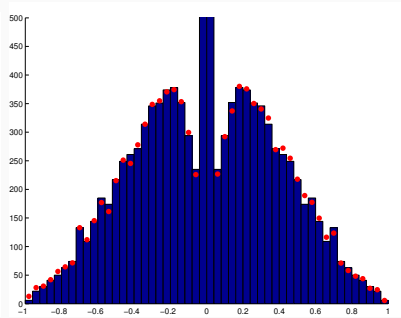
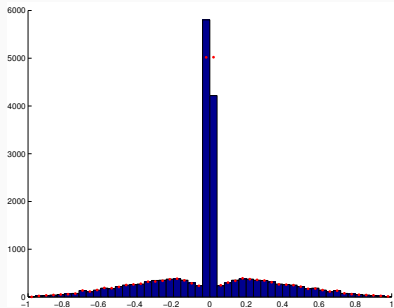


# Marvel characters (local)

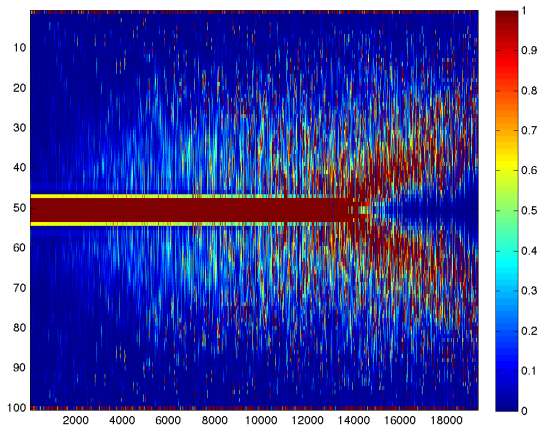


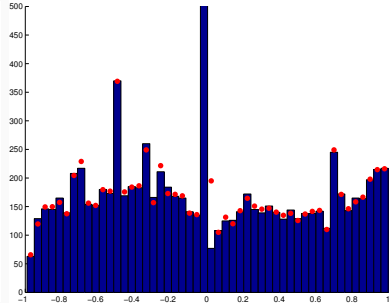
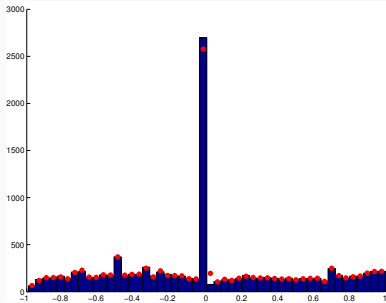


# Marvel comics

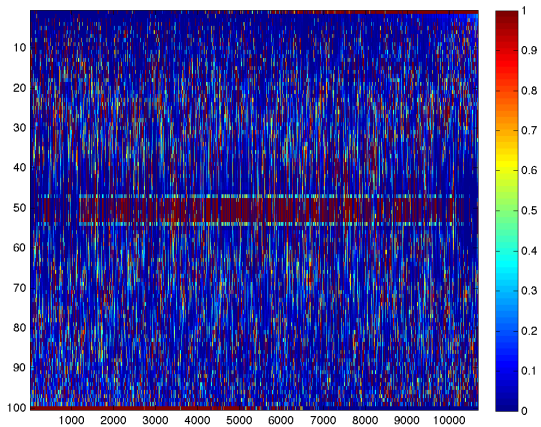


# Marvel comics (local)

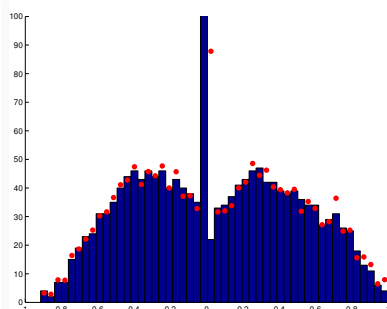
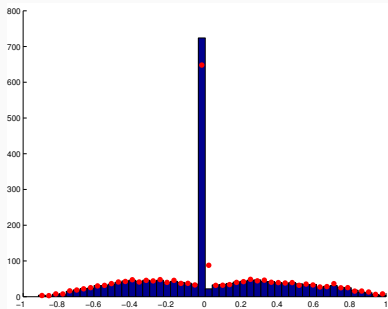




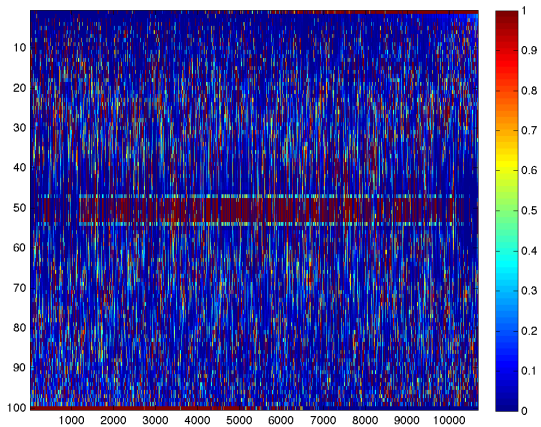
# PGP (local)



# Yeast

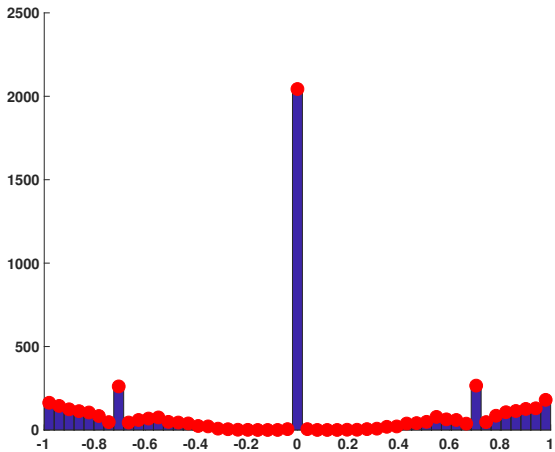


# Yeast (local)



# What about random graph models?

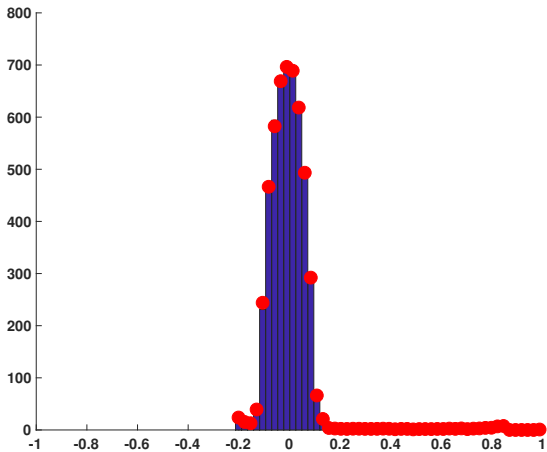
# Barabási–Albert model



Scale-free network (5000 nodes, 4999 edges)

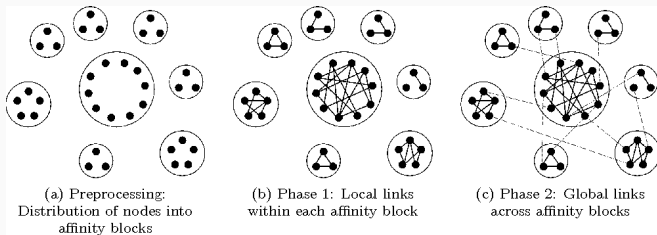


# Watts–Strogatz



Small world network (5000 nodes, 260000 edges)

# Model Verification: BTRE



Kolda et al, SISC (36), 2014

## Block Two-Level Erdős-Rényi model (BTRE)

- First Phase: Erdős-Rényi Blocks
- Second Phase: Using Chung-Lu Model to connect blocks with  $p_{ij} = p(d_i, d_j)$

# Model Verification: BTER

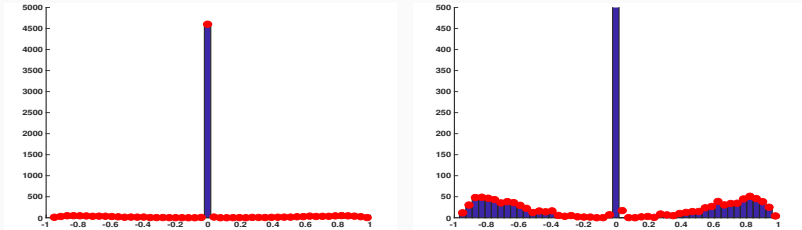


Figure 1: Erdos collaboration network.

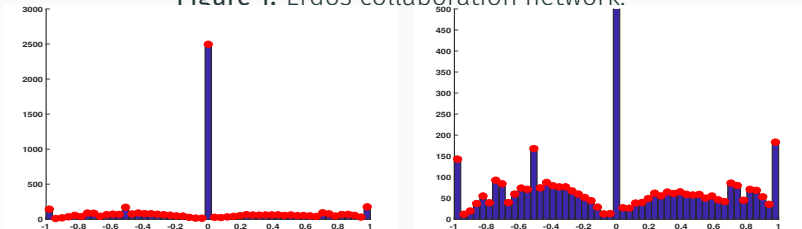
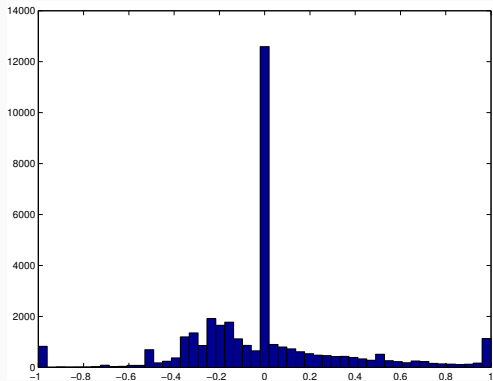


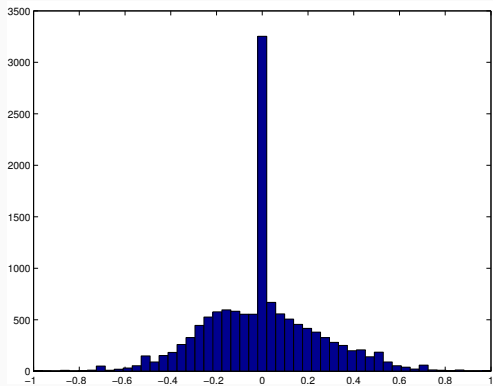
Figure 2: BTER model for Erdos collaboration network.

And a few more...

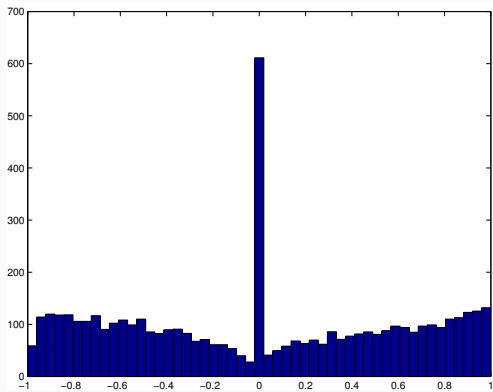
# Enron emails (SNAP)

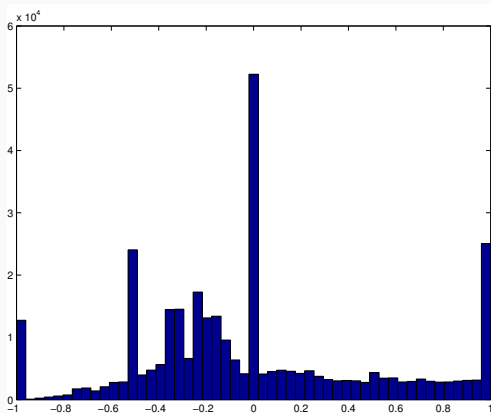


# Reuters911 (Pajek)



# US power grid (Pajek)

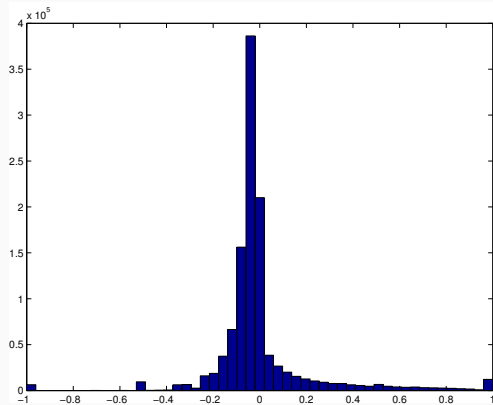




$N = 326186$ ,  $nnz = 1615400$ , 80 s (1000 moments, 10 probes)



# Hollywood 2009 (LAW)



$N = 1139905$ ,  $nnz = 113891327$ , 2093 s (1000 moments, 10 probes)

## Questions for You?

- Any isospectral graphs for multiple matrices?
- Can we recover topology from (exact) LDoS?
- Variance reduction in diagonal estimators?
- Random graphs with spectra that look “real”?
- Compression of moment information for diag estimators?
- More applications?

# What Do You Hear?

