Spectral Network Analysis: Beyond the Spectrum's Edge

David Bindel

18 February 2022

Department of Computer Science
Cornell University

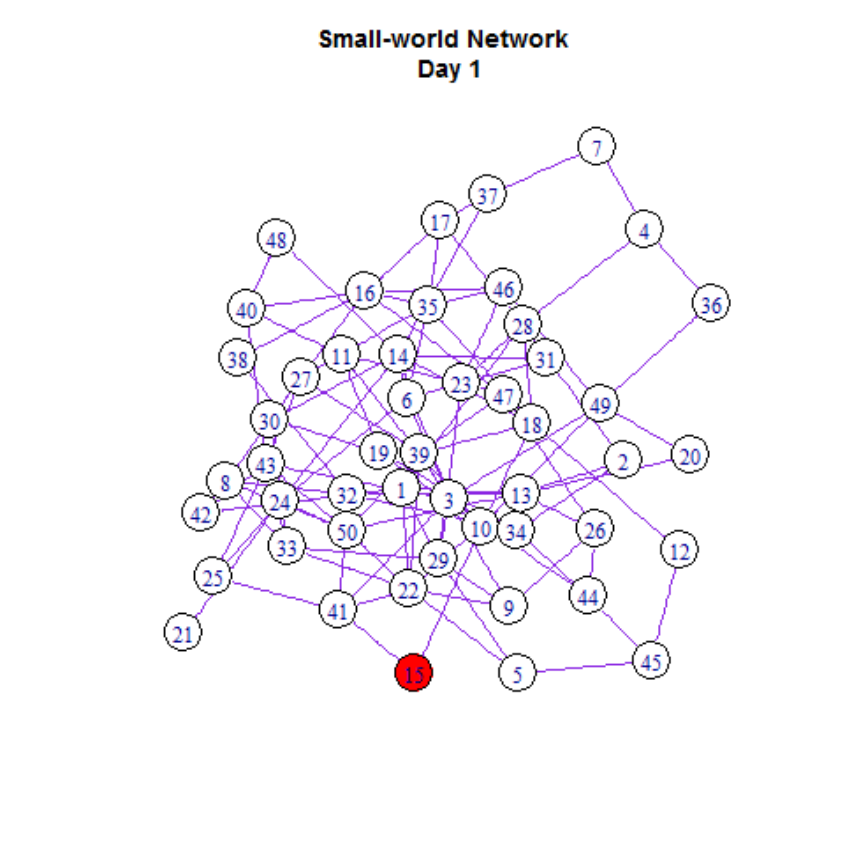
Eigenvalues Two Ways

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.
- Summarize *all* the eigenvalues (or all in a range).

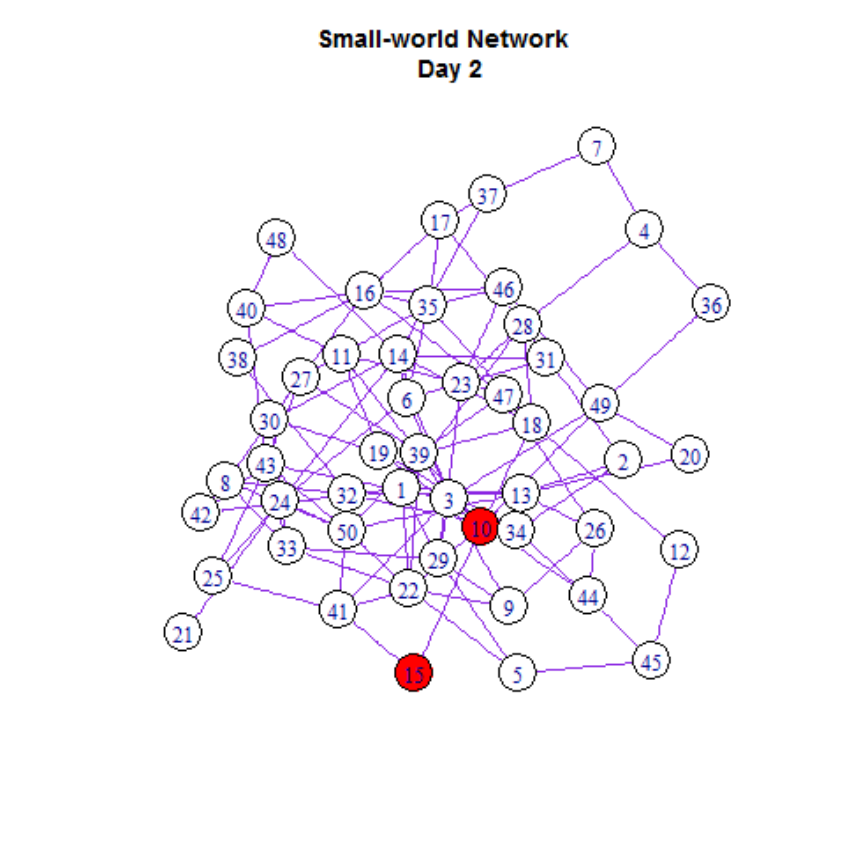
Tales from the Spectrum's Edge



Dynamics (operator on functions over manifold or graph)

- Diffuse according to heat kernel $\exp(-tLD^{-1})$
- Mixing rates and bottlenecks appear through spectrum

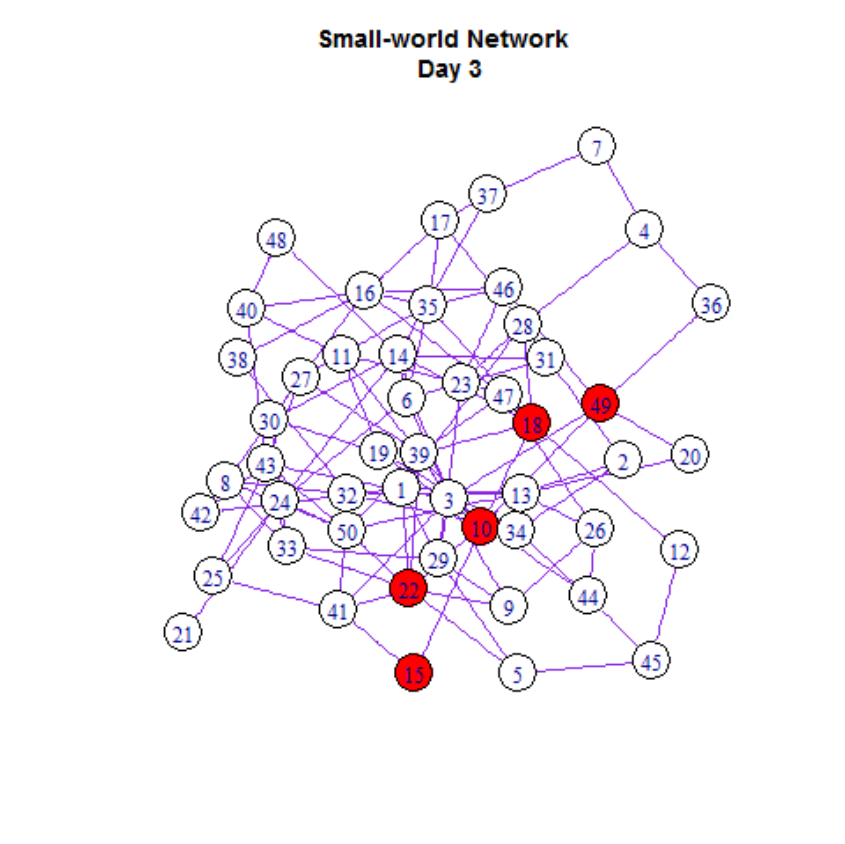
Tales from the Spectrum's Edge



Dynamics (operator on functions over manifold or graph)

- Diffuse according to heat kernel $\exp(-tLD^{-1})$
- Mixing rates and bottlenecks appear through spectrum

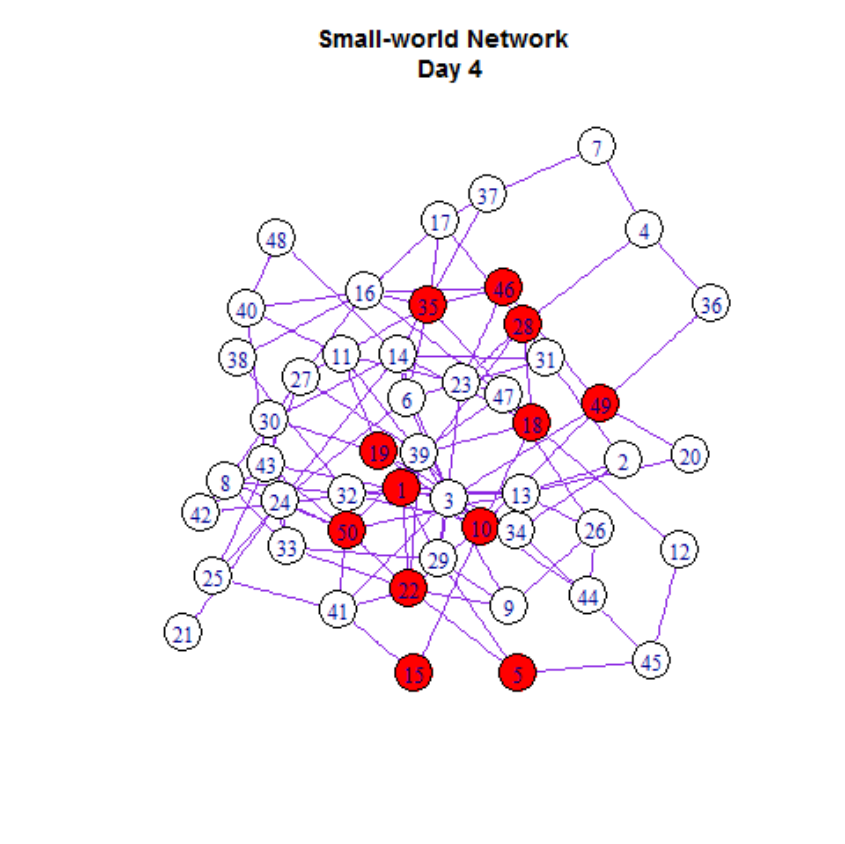
Tales from the Spectrum's Edge



Dynamics (operator on functions over manifold or graph)

- Diffuse according to heat kernel $\exp(-tLD^{-1})$
- Mixing rates and bottlenecks appear through spectrum

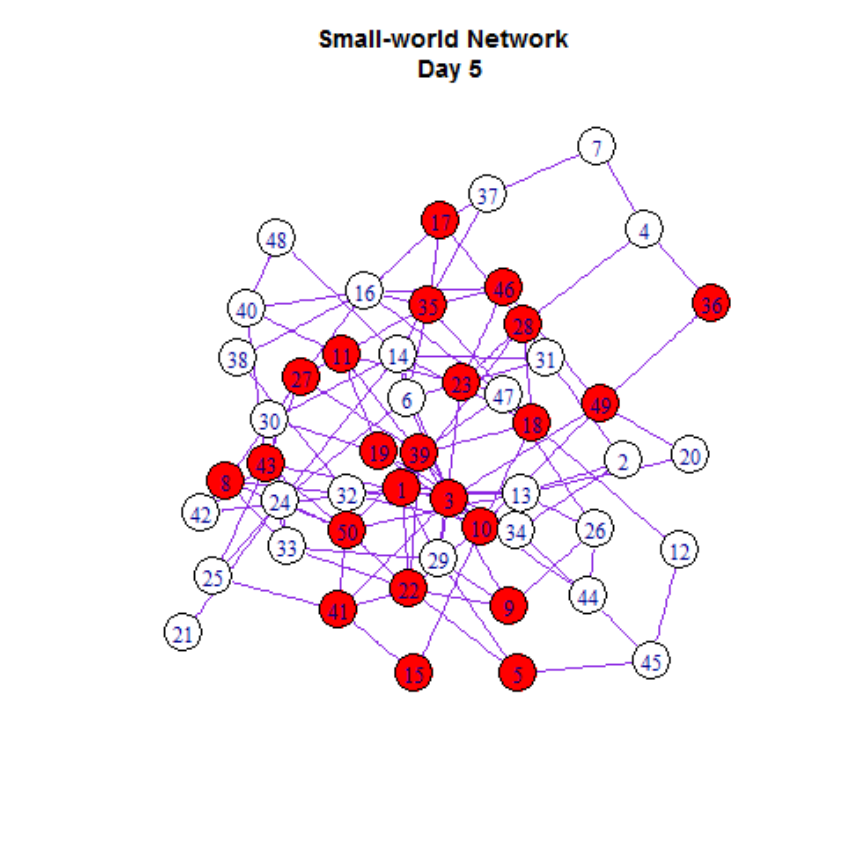
Tales from the Spectrum's Edge



Dynamics (operator on functions over manifold or graph)

- Diffuse according to heat kernel $\exp(-tLD^{-1})$
- Mixing rates and bottlenecks appear through spectrum

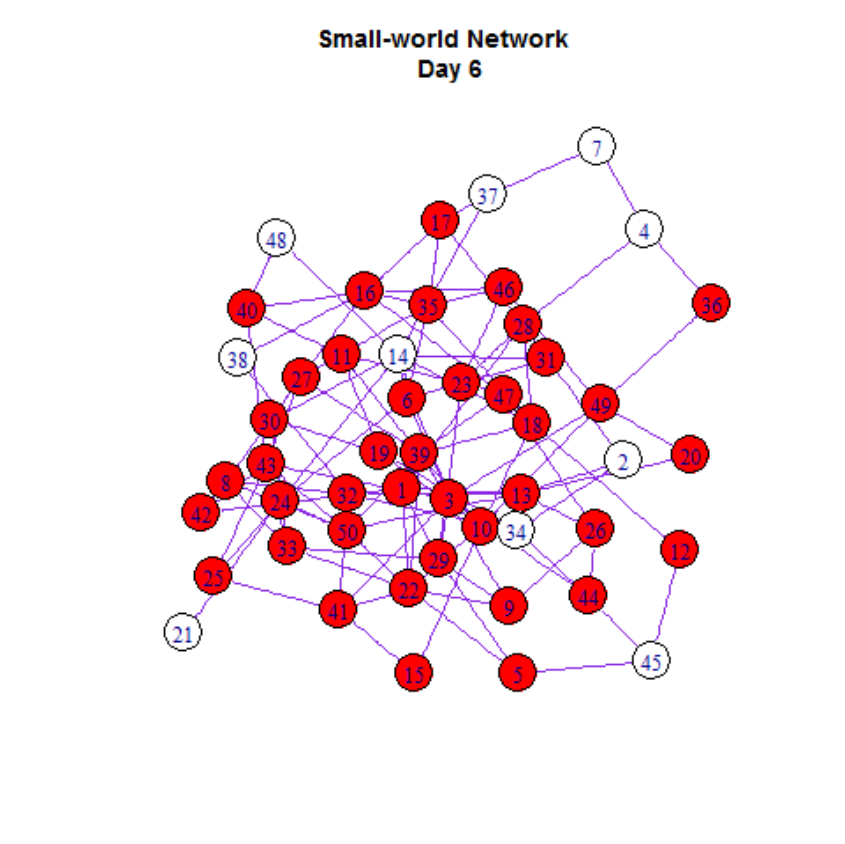
Tales from the Spectrum's Edge



Dynamics (operator on functions over manifold or graph)

- Diffuse according to heat kernel $\exp(-tLD^{-1})$
- Mixing rates and bottlenecks appear through spectrum

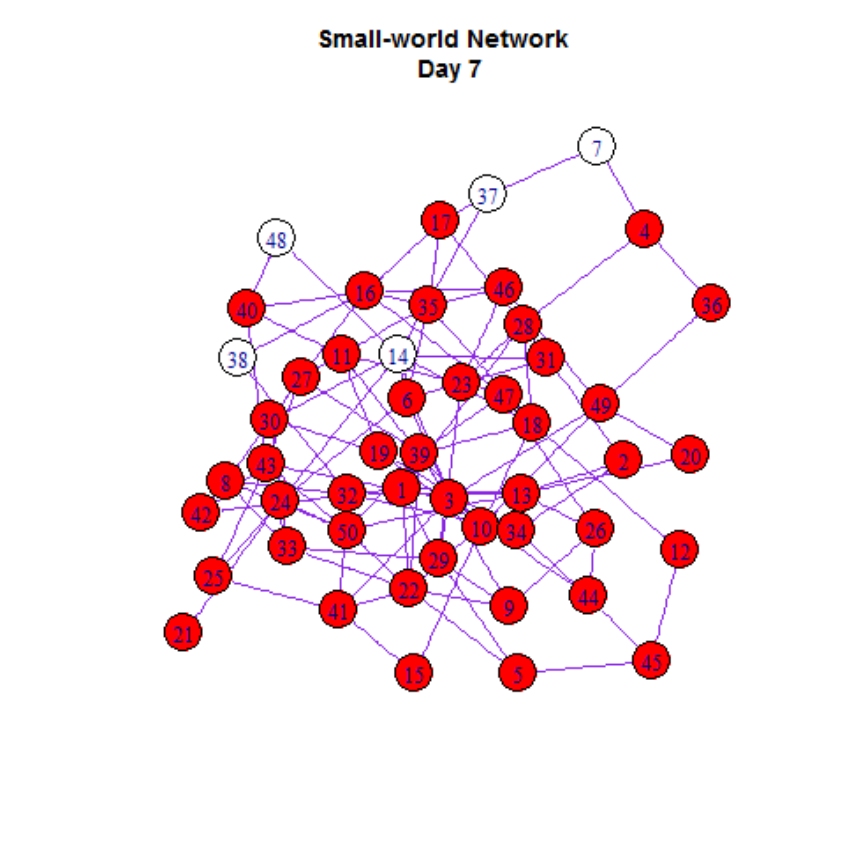
Tales from the Spectrum's Edge



Dynamics (operator on functions over manifold or graph)

- Diffuse according to heat kernel $\exp(-tLD^{-1})$
- Mixing rates and bottlenecks appear through spectrum

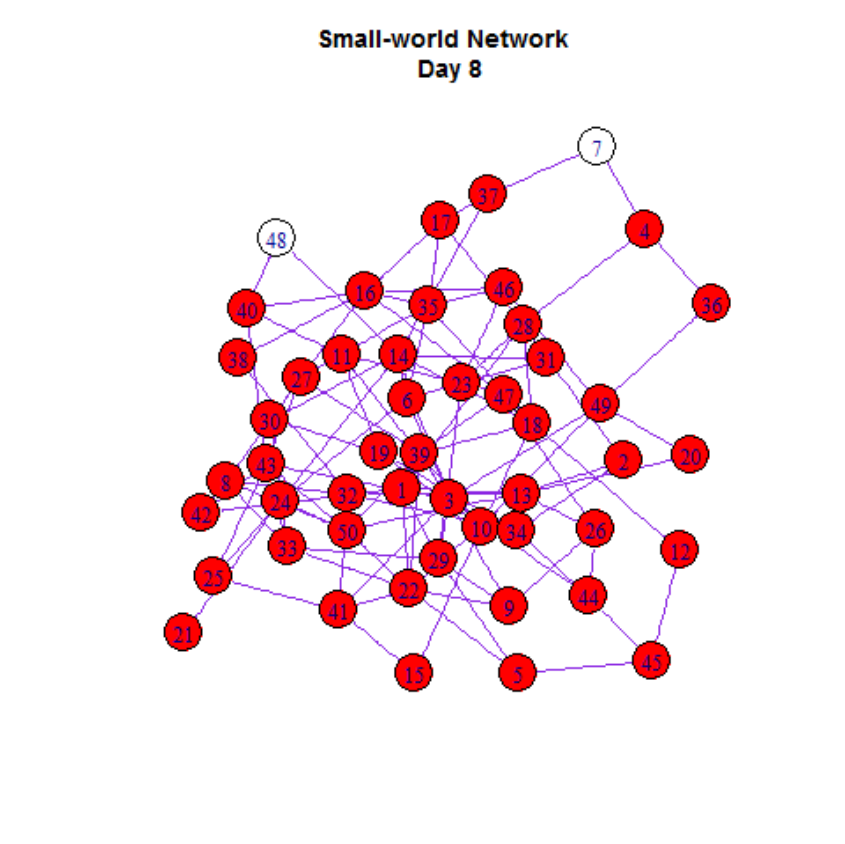
Tales from the Spectrum's Edge



Dynamics (operator on functions over manifold or graph)

- Diffuse according to heat kernel $\exp(-tLD^{-1})$
- Mixing rates and bottlenecks appear through spectrum

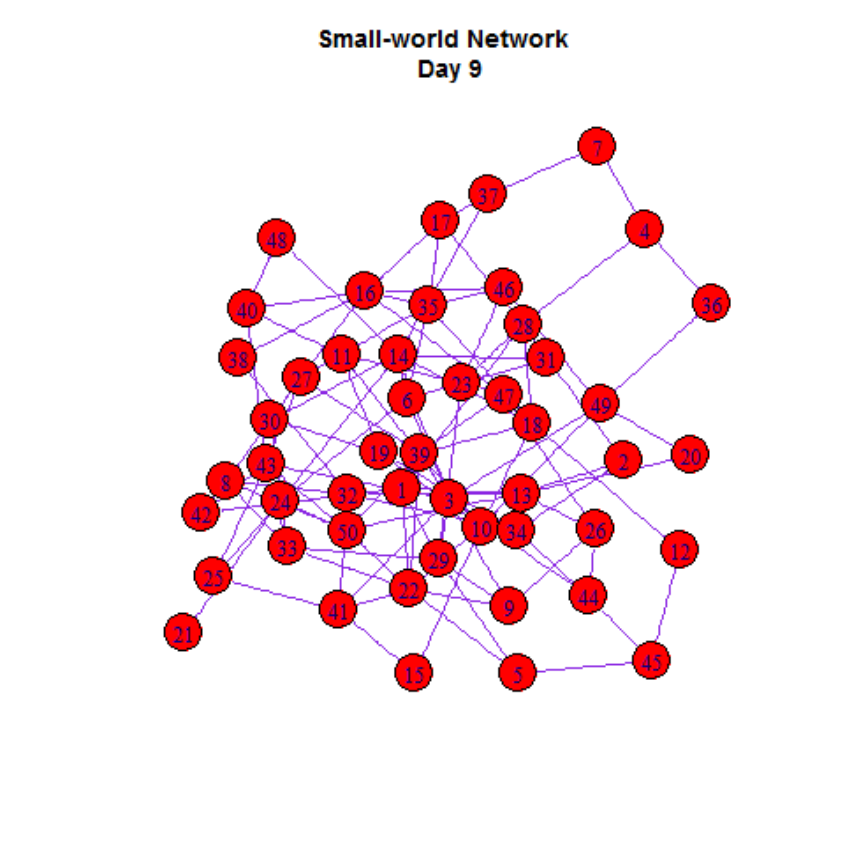
Tales from the Spectrum's Edge



Dynamics (operator on functions over manifold or graph)

- Diffuse according to heat kernel $\exp(-tLD^{-1})$
- Mixing rates and bottlenecks appear through spectrum

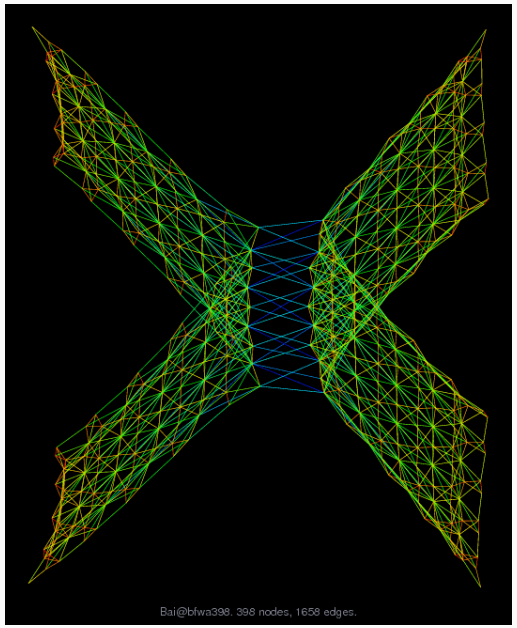
Tales from the Spectrum's Edge



Dynamics (operator on functions over manifold or graph)

- Diffuse according to heat kernel $\exp(-tLD^{-1})$
- Mixing rates and bottlenecks appear through spectrum

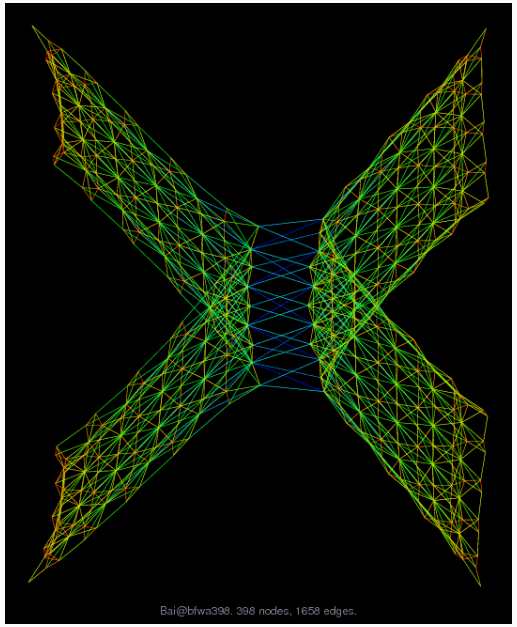
Tales from the Spectrum's Edge: Counting and Measure



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

Tales from the Spectrum's Edge



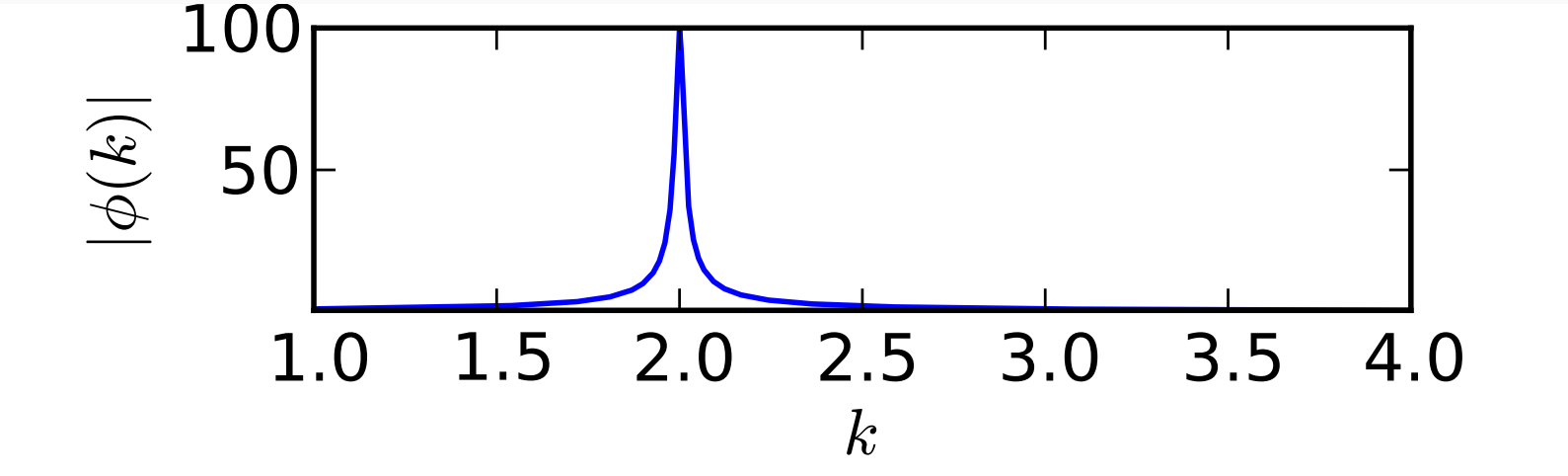
Geometric embedding (pos semidef form / kernel)

- 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

Beyond the Spectrum's Edge

- Edge is great for "coarse" features of a graph/manifold
- What about localized phenomena?
- What about "texture"?
- Inspiration: Interpreting continuous spectra in physics

Scattering and Resonances



For $H = -\delta + V$ with $\mathrm{supp}(V) \subset \Omega$:

$$(H - k^2)\psi = f \text{ on } \Omega$$

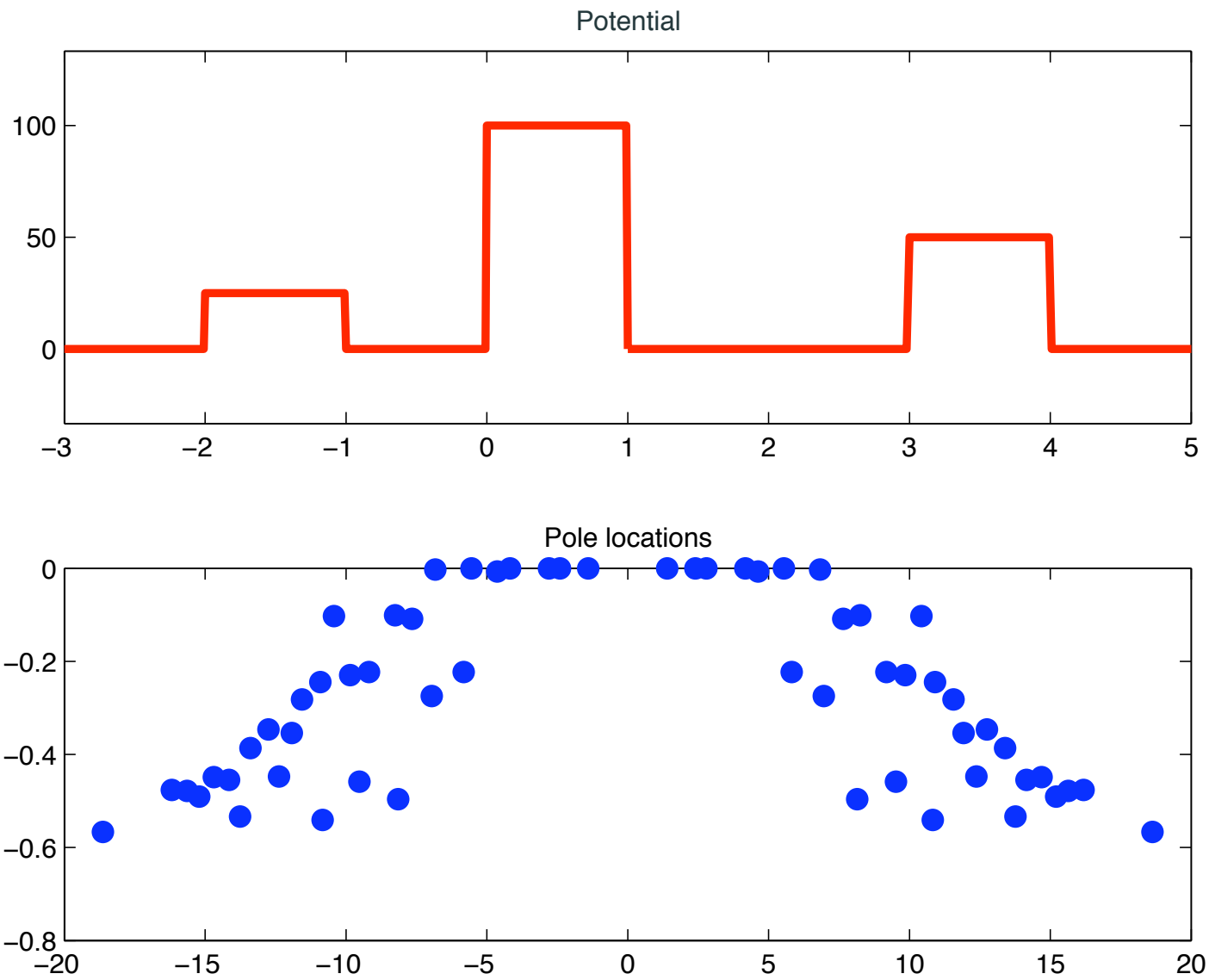
$$(\partial_n - B(k))\psi = 0 \text{ on } \partial\Omega$$

See resonance peaks (Breit-Wigner):

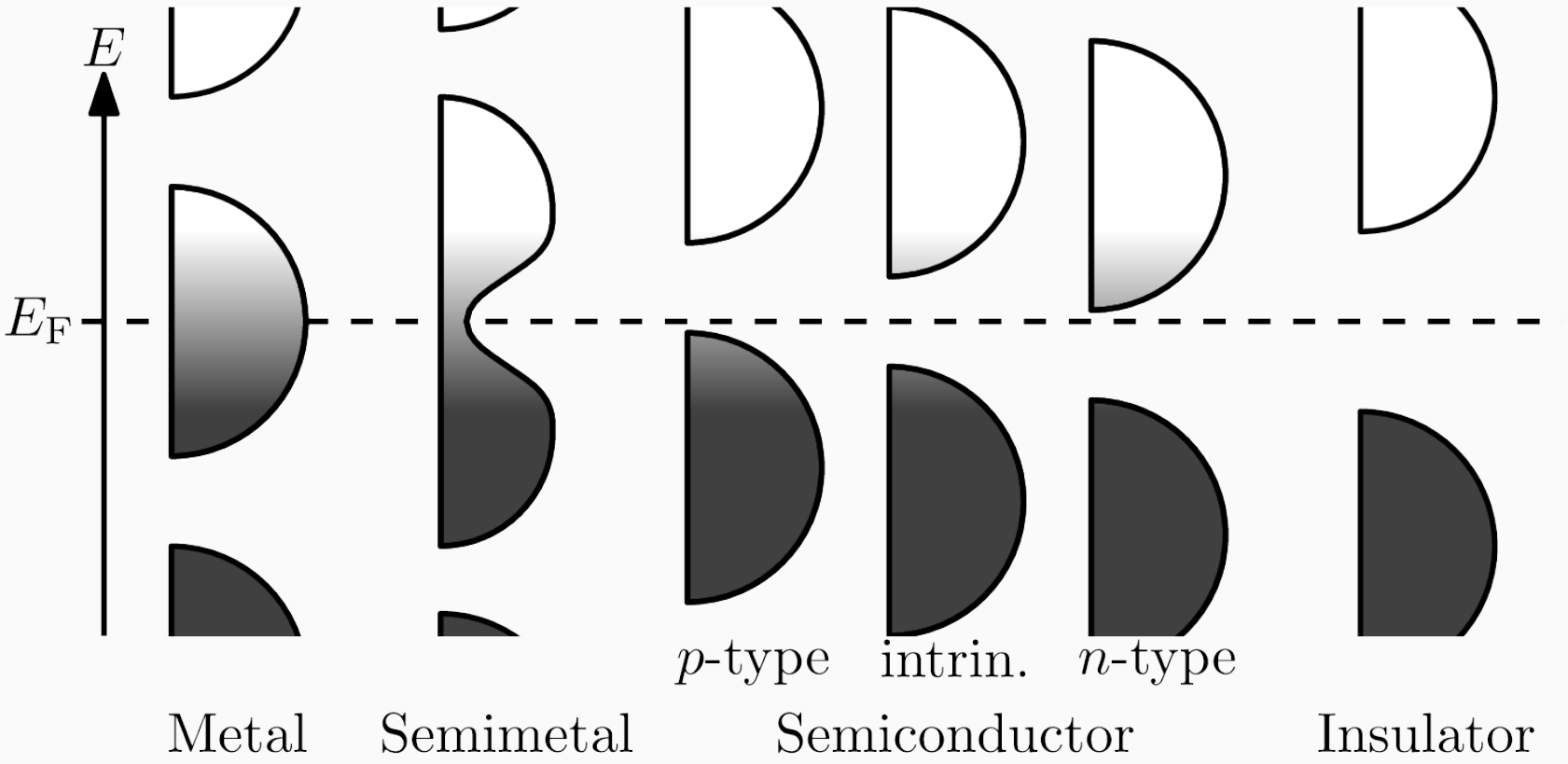
$$\phi(k) \equiv w^* \psi \approx C(k - k^*)^{-1}.$$

Associated with trapped “leaky” local modes.

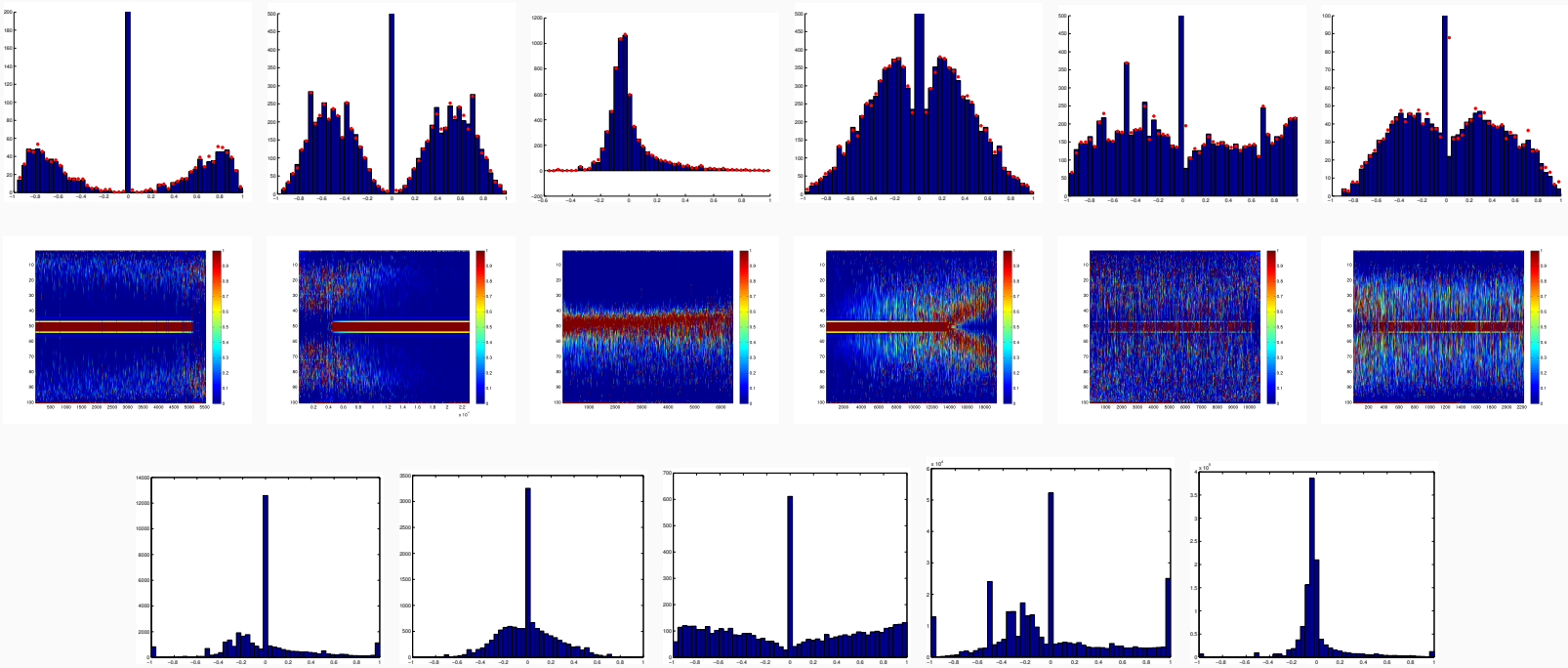
Scattering and Resonances



Density of States in Physics

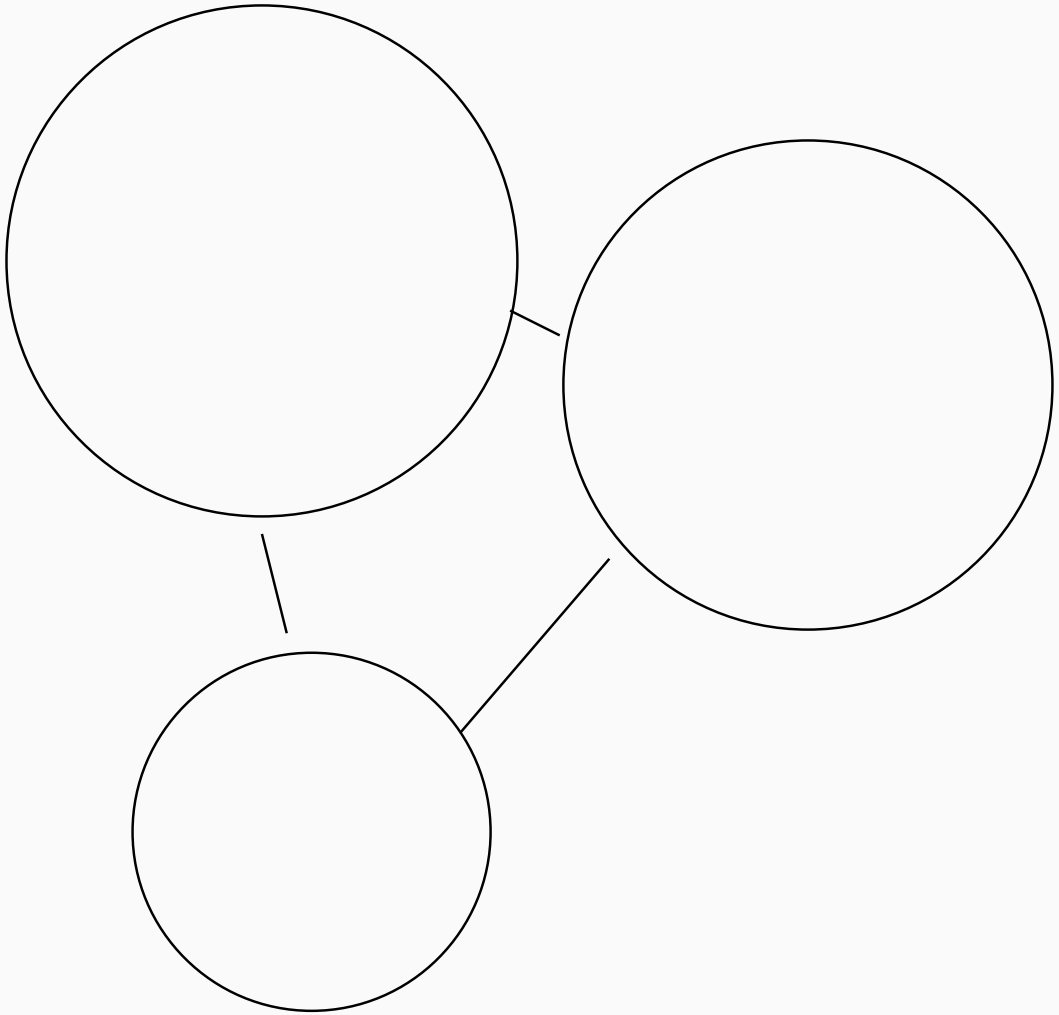


Today in Two Acts



- **Act 1: Trapping, resonance, and local diffusions**
- Act 2: Spectral densities

Random walker



Simon-Ando

Three fast-mixing chains with weak coupling:

$$P = \begin{bmatrix} P_{11} & \epsilon & \epsilon \\ \epsilon & P_{22} & \epsilon \\ \epsilon & \epsilon & P_{33} \end{bmatrix}$$

Random walker (discrete time):

$$\pi^{(k+1)} = P\pi^{(k)}$$

Very quickly get

$$\pi^{(k)} \approx \begin{bmatrix} c_1^{(k)}\pi_1^{(\infty)} \\ c_2^{(k)}\pi_2^{(\infty)} \\ c_3^{(k)}\pi_3^{(\infty)} \end{bmatrix}$$

where $\pi_j^{(\infty)}$ is the stationary distribution for chain j and $c_j^{(k)}$ is converges slowly to stationary behavior.

Simon-Ando and clustering

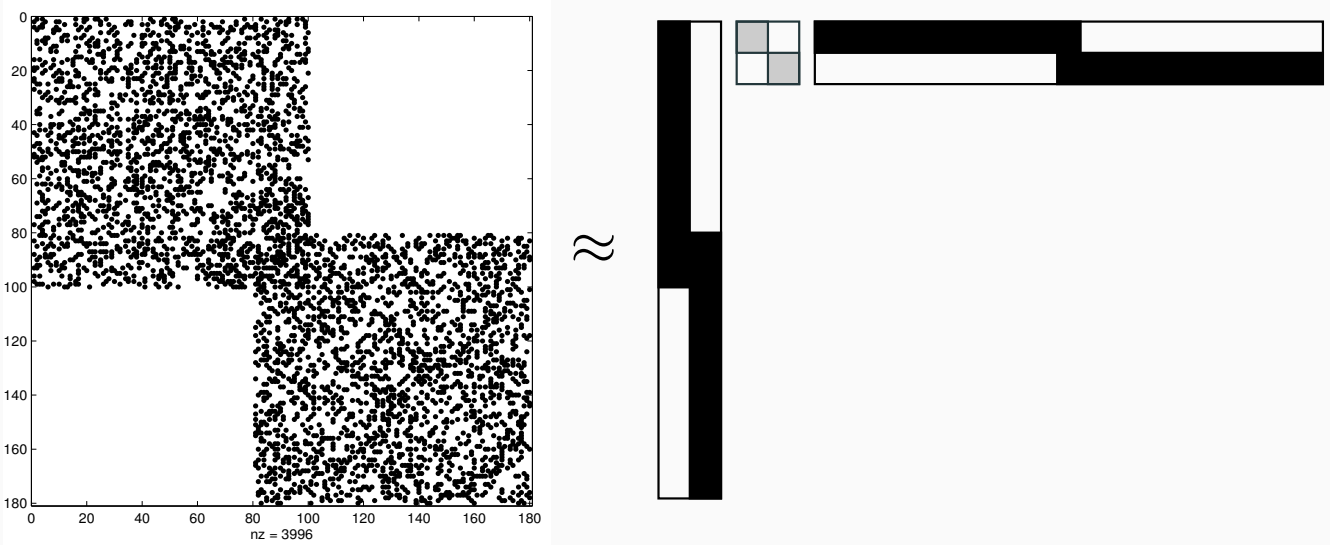
Run a diffusion (or diffusions) into Simon-Ando regime; then

$$\mathrm{sp}\begin{bmatrix} b_1\pi_1^{(\infty)} & c_1\pi_1^{(\infty)} & d_1\pi_1^{(\infty)} \\ b_2\pi_2^{(\infty)} & c_2\pi_2^{(\infty)} & d_2\pi_2^{(\infty)} \\ b_3\pi_3^{(\infty)} & c_3\pi_3^{(\infty)} & d_3\pi_3^{(\infty)} \end{bmatrix} = \mathrm{sp}\begin{bmatrix} \pi_1^{(\infty)} & & \\ & \pi_2^{(\infty)} & \\ & & \pi_3^{(\infty)} \end{bmatrix}$$

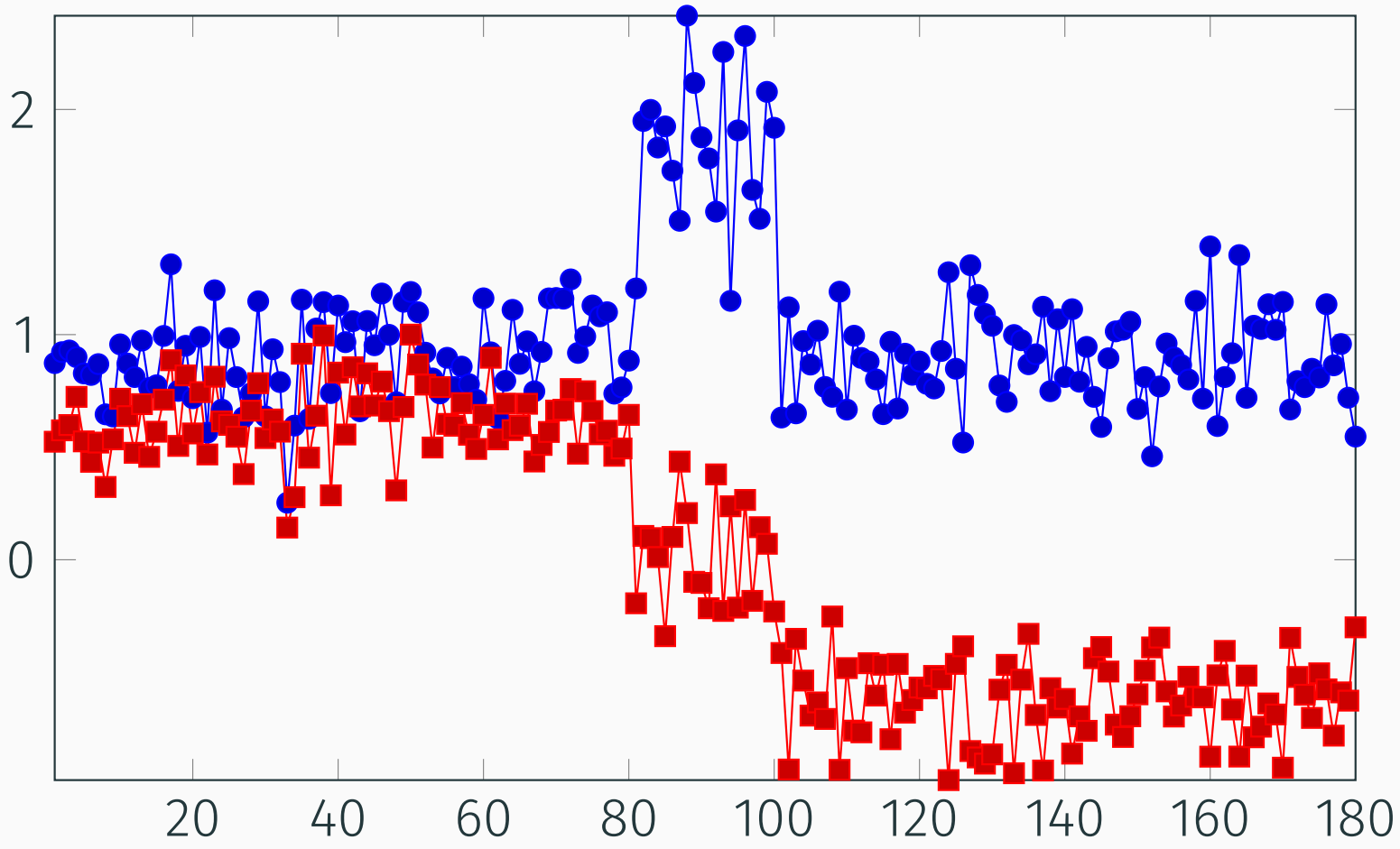
Several ways to find latter (sparse) basis from former:
(k-means clustering, ℓ^1 minimization, ...)

What about very small clusters? Overlap?

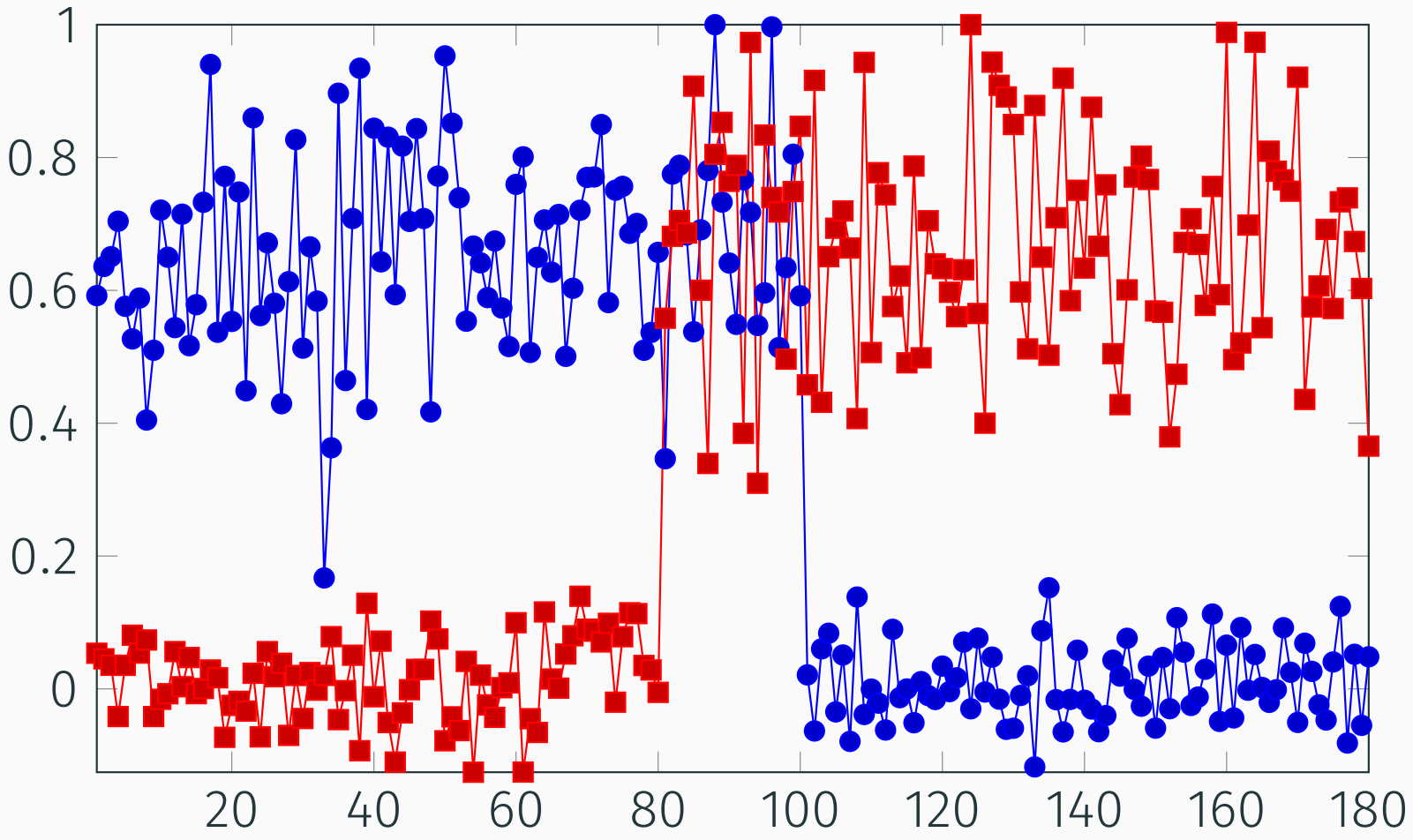
Overlapping case



Dominant vectors for a block model example



Same space, different basis



Local Spectra and Diffusions

Consider $\pi^{(t+1)} = P\pi^{(t)}$ starting from $\pi^{(0)}$.
Full dynamics (ergodic case):

$$\pi^{(t)} = \sum_{j=1}^{n} c_j v_j \lambda_j^t$$

with $1 = \lambda_1 > \lambda_2 \geq \lambda_3 \geq \ldots > -1$.

Transient dynamics for $t < k$:

$$\pi^{(t)} = \sum_{j=1}^{k} d_j u_j \rho_j^t$$

with $1 \geq \rho_1 \geq \rho_2 \geq \ldots > -1$.

Rayleigh-Ritz and Diffusion

Transient dynamics for $t < k$:

$$\pi^{(t)} = \sum_{j=1}^{k} d_j u_j \rho_j^t$$

with $1 \geq \rho_1 \geq \rho_2 \geq \ldots > -1$.

The (u_j, ρ_j) are *Ritz pairs*, approximate eigenpairs drawn from

$$\mathcal{K}_k(P, \pi^{(0)}) = \mathrm{sp}\{\pi^{(0)}, \pi^{(1)}, \ldots, \pi^{(k-1)}\}.$$

Unconverged eigenpairs in a Krylov method are useful "local" objects on their own!

LEMON

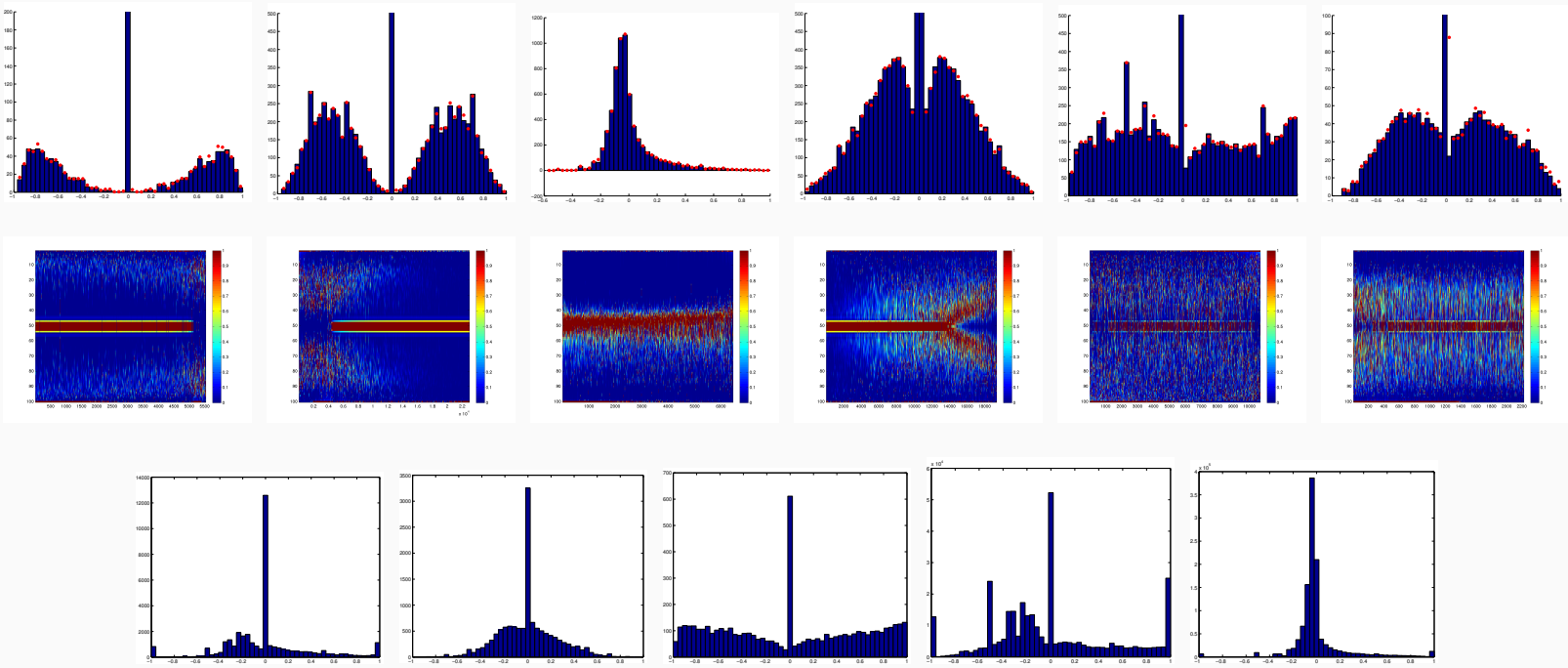
*L*ocal *E*xpansion via *M*inimum *O*ne *Norm*:

- Run diffusion (Lanczos process) from seed node(s)
- Compute Ritz pairs
- Constrained 1-norm minimization over a Ritz space

Works quite well! Line of papers with Kun He, John Hopcroft, several co-authors; see, e.g.

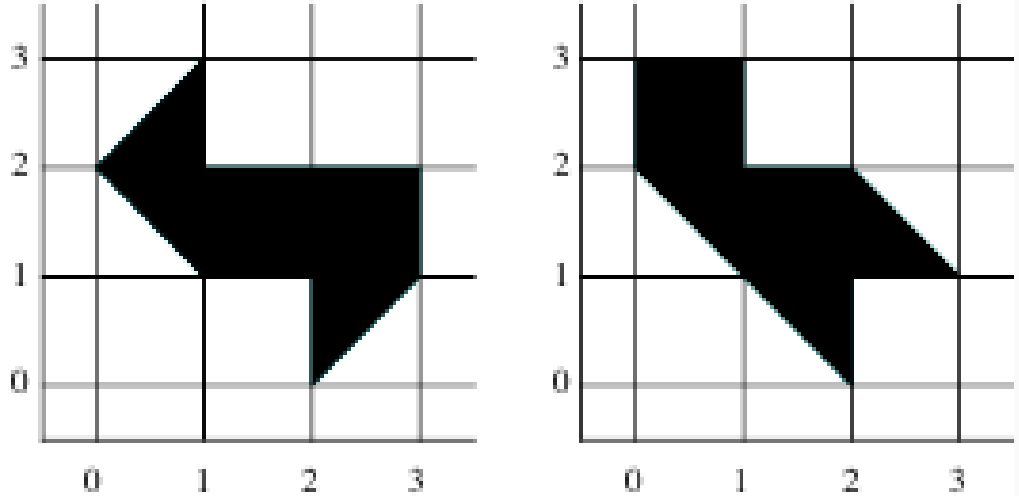
Y. Li, K. He, K. Kloster, D. Bindel, H. Hopcroft, *Local Spectral Clustering for Overlapping Community Detection* in ACM Transactions on Knowledge Discovery from Data, 2018.

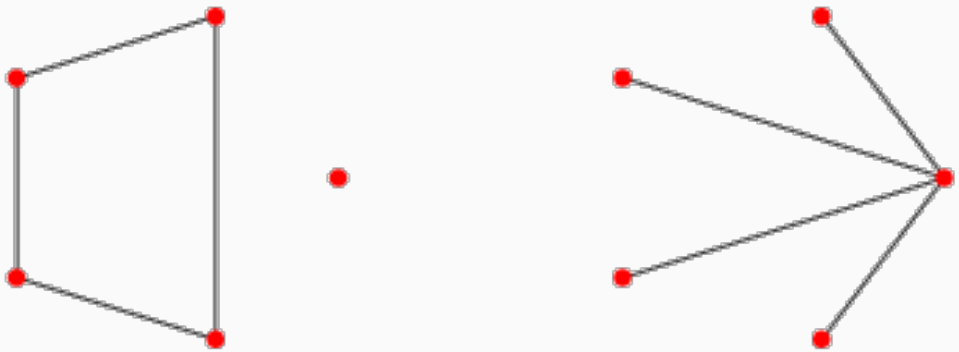
Today in Two Acts



- Act 1: Trapping, resonance, and local diffusions
- **Act 2: Spectral densities**

Can One Hear the Shape of a Drum?

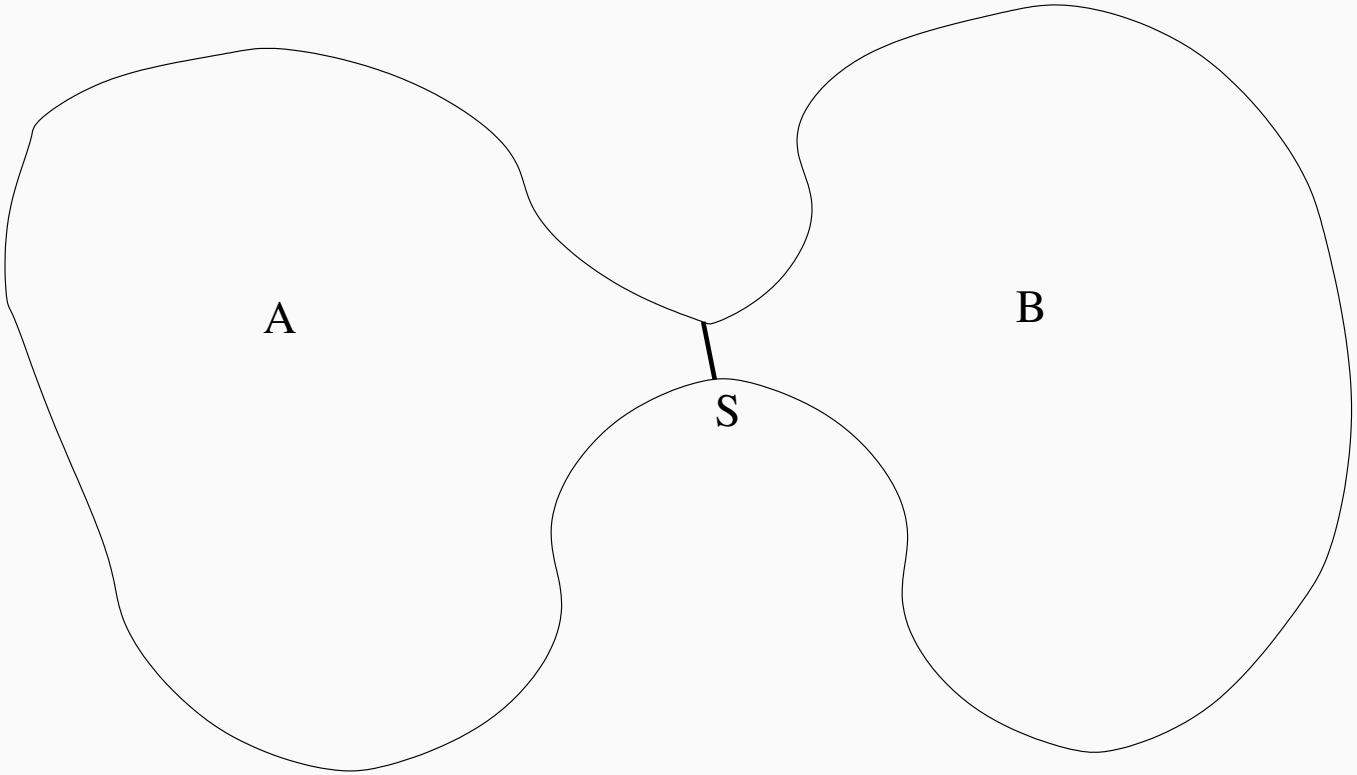




"You mean, if you had perfect pitch could you find the shape of a drum." — Mark Kac (quoting Lipmann Bers)

American Math Monthly, *1966*

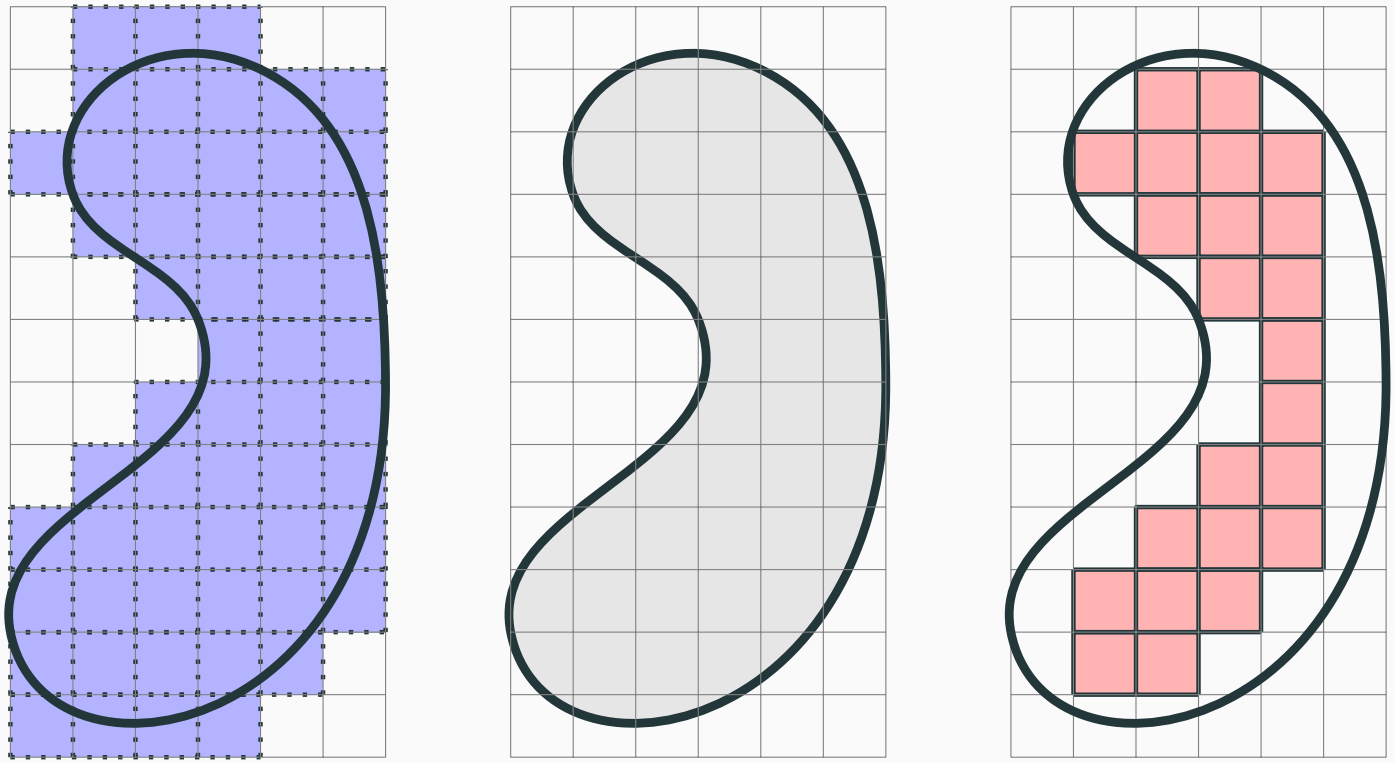
What Do You Hear? (Low frequency)



Size of bottlenecks (Cheeger inequality)

$$h \leq 2\sqrt{\lambda_2}$$

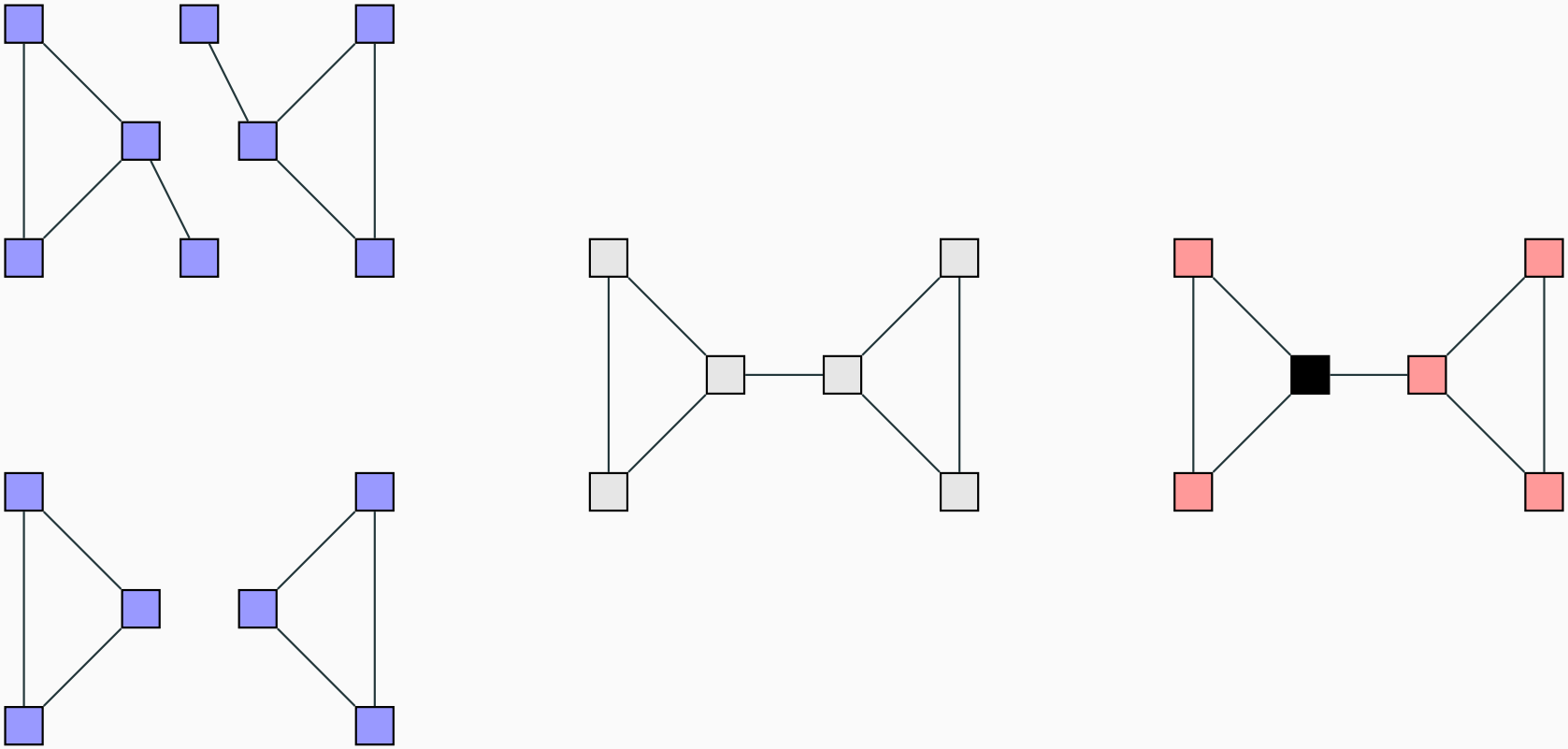
What Do You Hear? (High frequency)



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

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

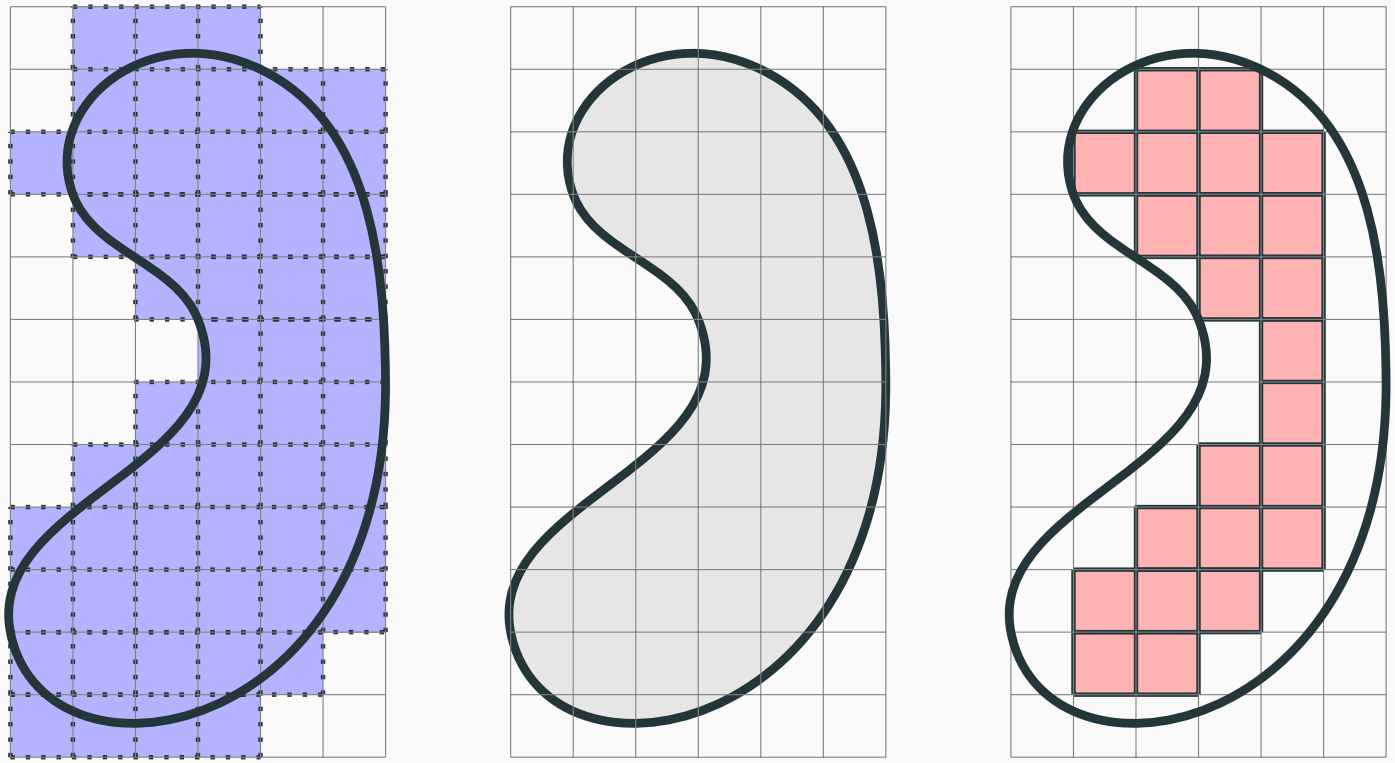
What Do You Hear?



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

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

What Do You Hear?



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

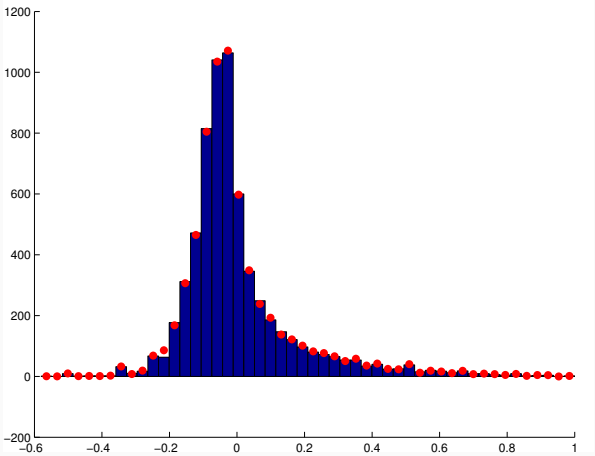
$$\lim_{x\to\infty} \frac{N(x)}{x^{d/2}} = (2\pi)^{-d}\omega_d \,\mathrm{vol}(\Omega).$$

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

Density of States



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

$$\mathrm{tr}(f(H)) = \int f(\lambda)\mu(\lambda)\,dx = \sum_{k=1}^{N} f(\lambda_k)$$

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

Example: Estrada Index

Consider

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

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

Heat Kernels

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

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

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

Power Moments

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

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

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: ??

Exploring Spectral Densities

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

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

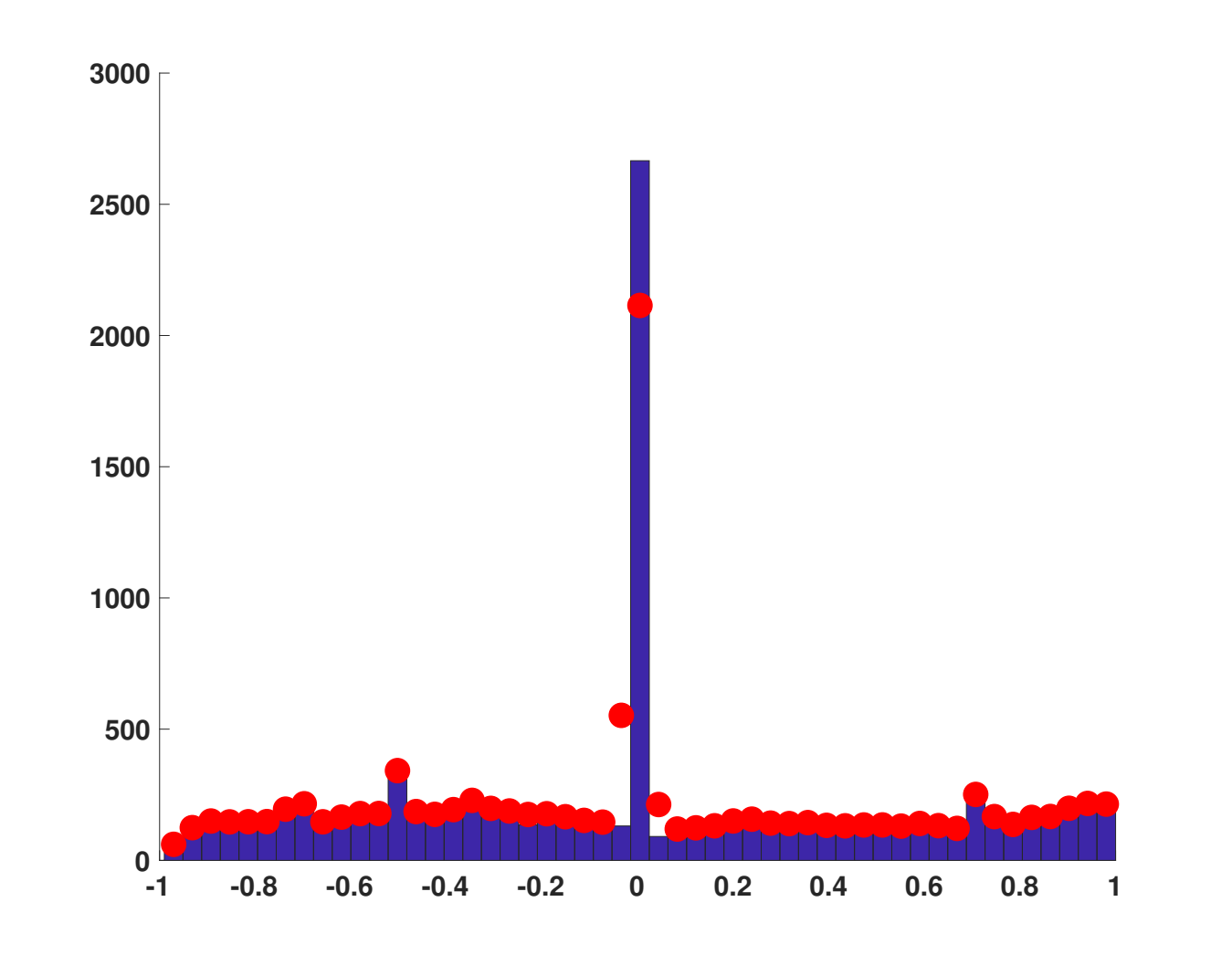
$$\int_{-1}^{1} \mu(x) f(x)\, dx = \frac{1}{N} \sum_{k=1}^{N} f(\lambda_k)$$

- 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 $d_j = \mathbf{tr}(T_j(H))$ 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)

Example: PGP Network



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

Motifs and Symmetry

Suppose $PH = HP$. Then

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

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

Simplest example: P swaps (i, j)

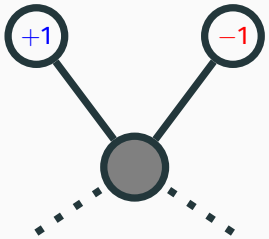
- $e_i - e_j$ an eigenvector of P with eigenvalue -1
- $e_i - e_j$ an eigenvector of $\bar{A}$ with eigenvalue

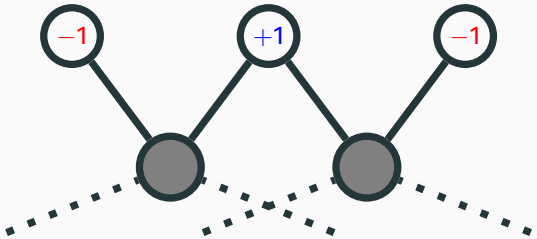
$$\lambda = \rho_{\bar{A}}(e_i - e_j) = \begin{cases} d^{-1}, & (i,j) \in \mathcal{E} \\ 0, & \text{otherwise.} \end{cases}$$

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

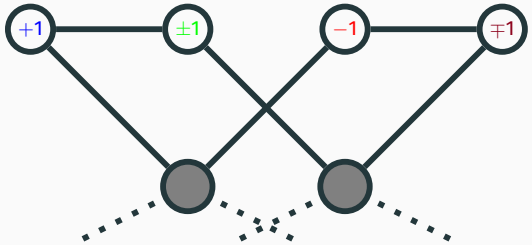
Motifs in Spectrum

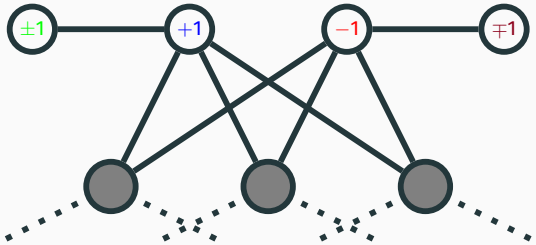
- $\lambda = 0$



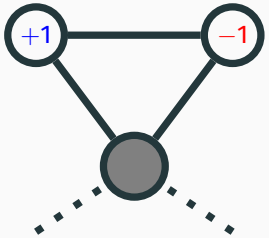


- $\lambda = \pm 1/2$

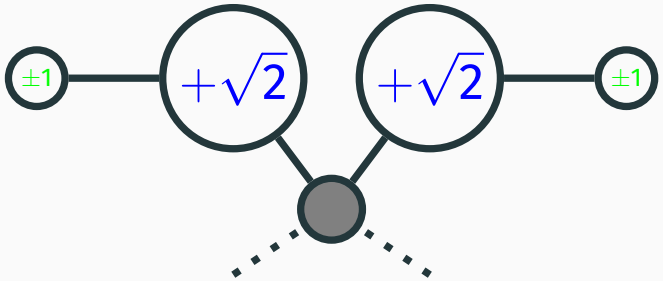




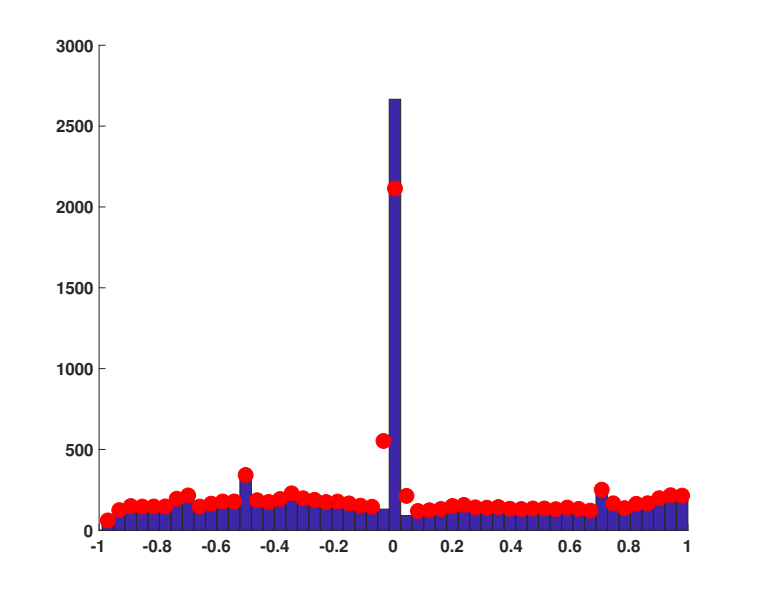
- $\lambda = -1/2$

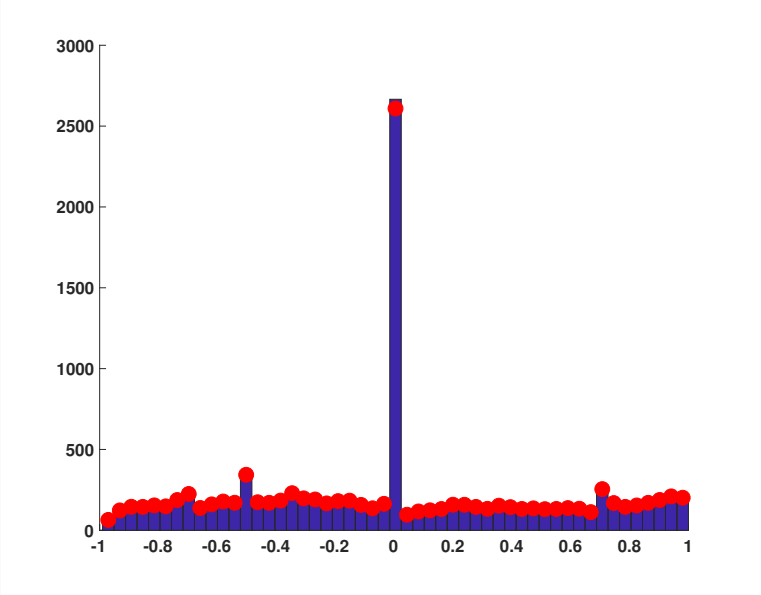


$\lambda = \pm 1/\sqrt{2}$



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

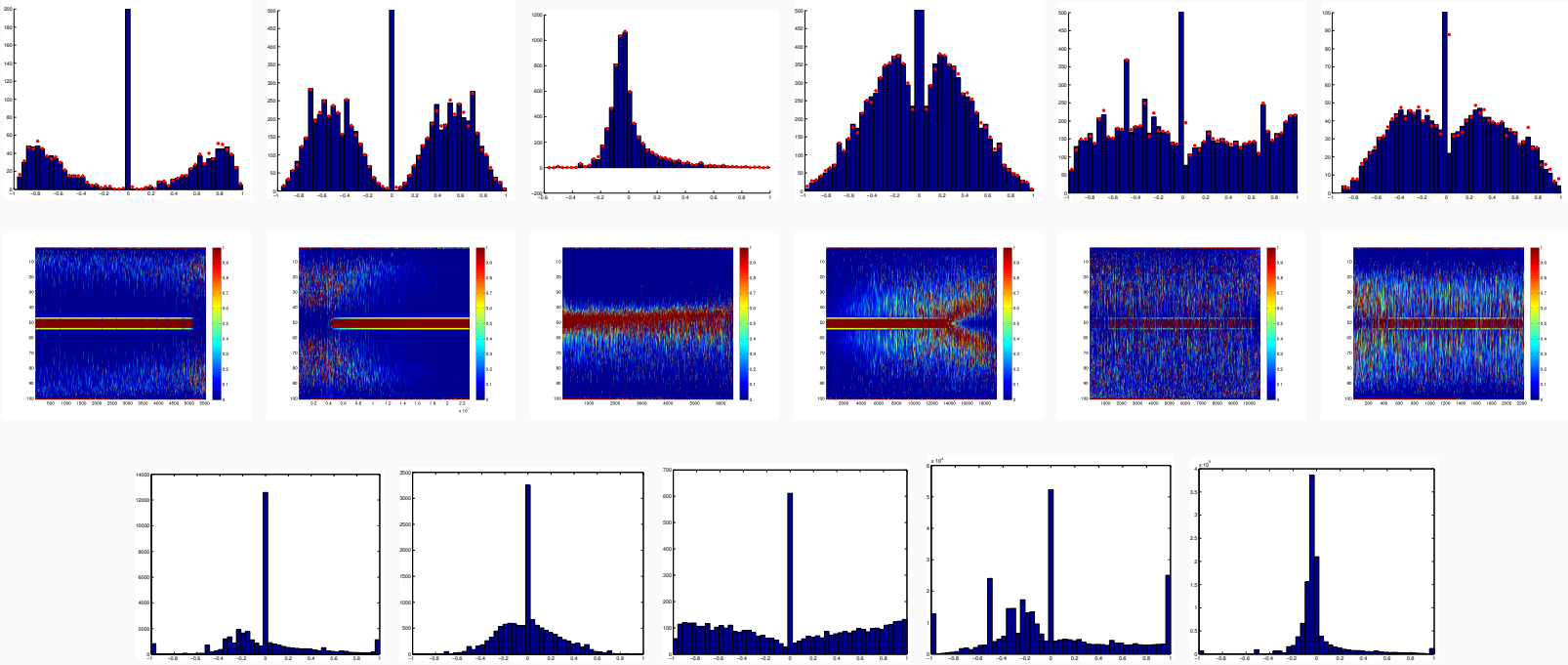
Then combine with standard KPM.

Claim: Network DoS is Useful!

Dong, Benson, Bindel. “Network Density of States.”
Proceedings of KDD 2019.

Huang, Graven, Bindel. “Density of States Graph Kernels.”
Proceedings of SDM 2021.

So What Do You Hear?



Thank you!

Li, He, Kloster, Bindel, Hopcroft, *Local Spectral Clustering for Overlapping Community Detection* in ACM Transactions on Knowledge Discovery from Data, 2018.

Dong, Benson, Bindel. “Network Density of States.” *Proceedings of KDD 2019*.

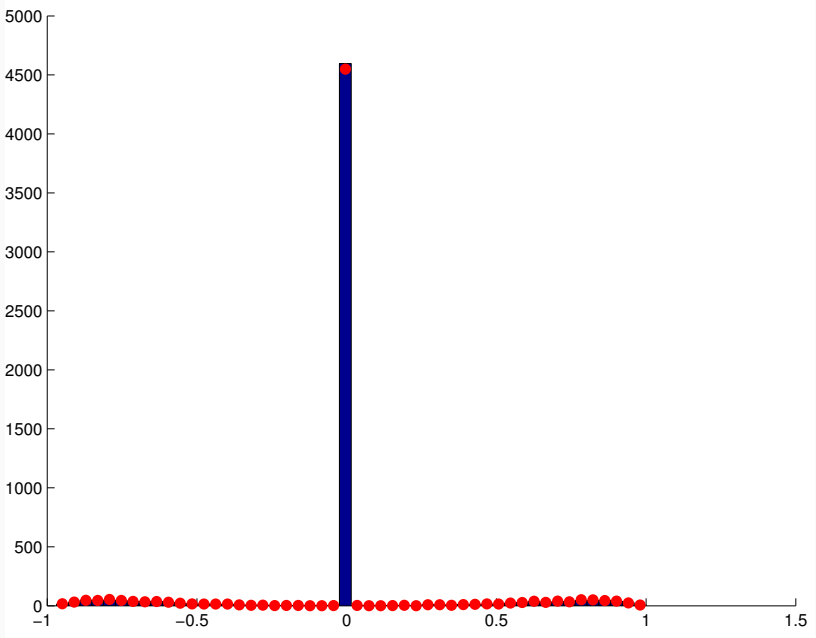
Backup slides

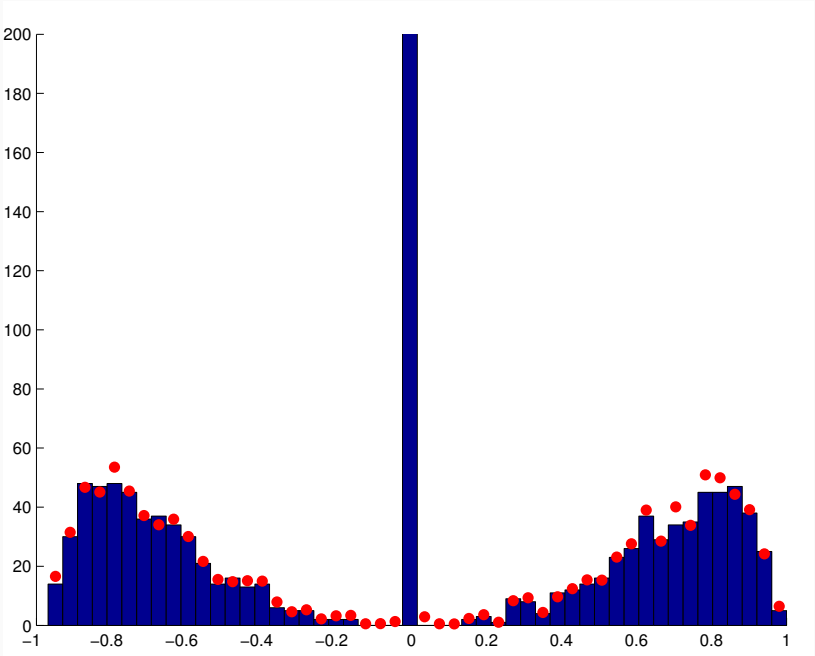
Experimental setup

- 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

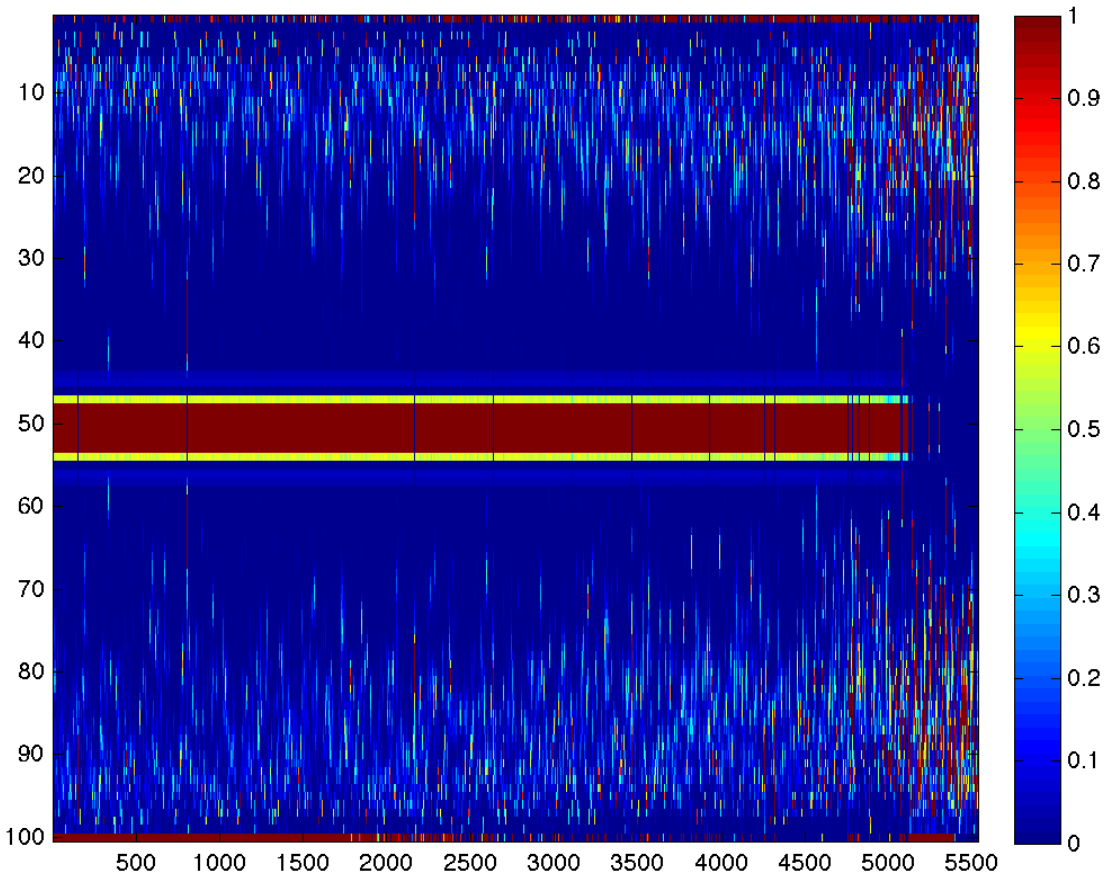
Suggestions for better pics are welcome!

Erdos

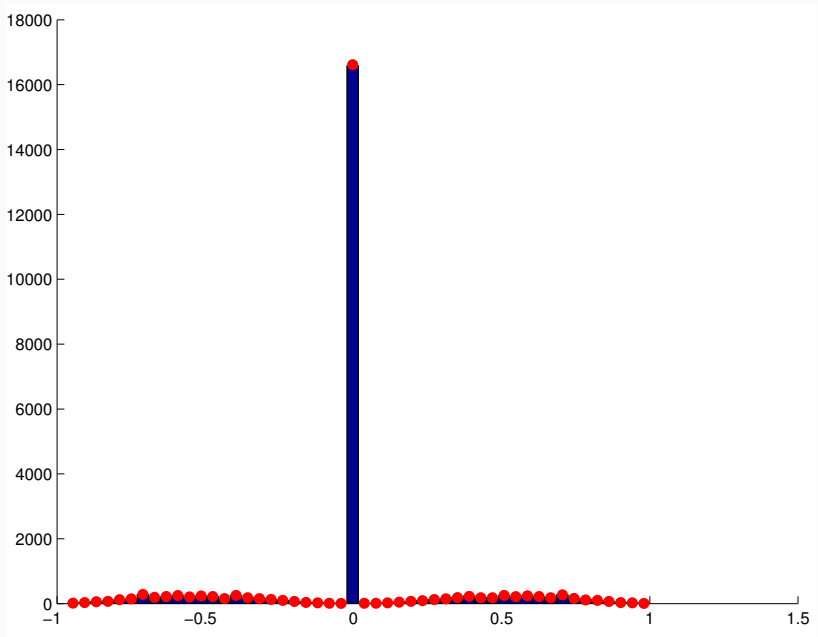

5000
4500
4000
3500
3000
2500
2000
1500
1000
500
0
−1
−0.5
0
0.5
1
1.5

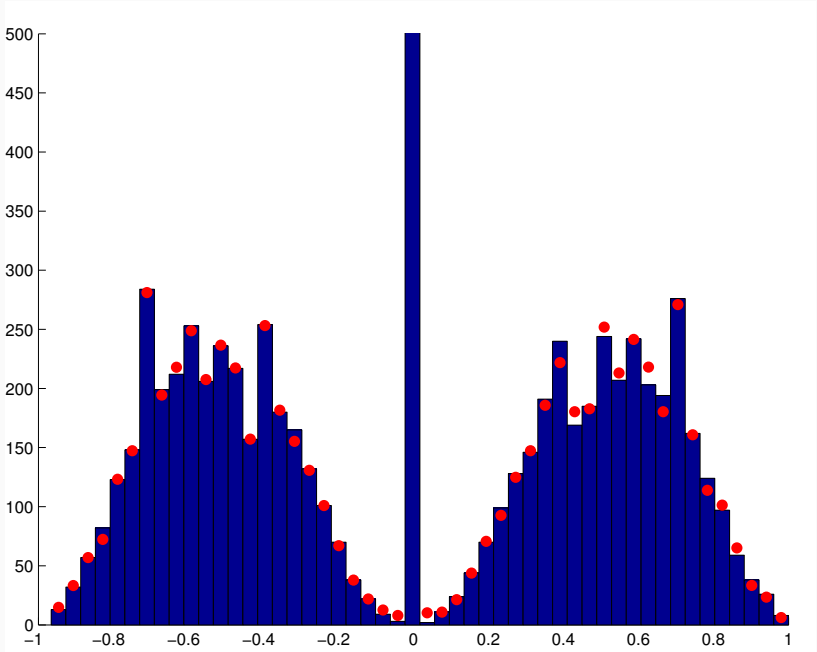

200
180
160
140
120
100
80
60
40
20
0
−1
−0.8
−0.6
−0.4
−0.2
0
0.2
0.4
0.6
0.8
1

Erdos (local)

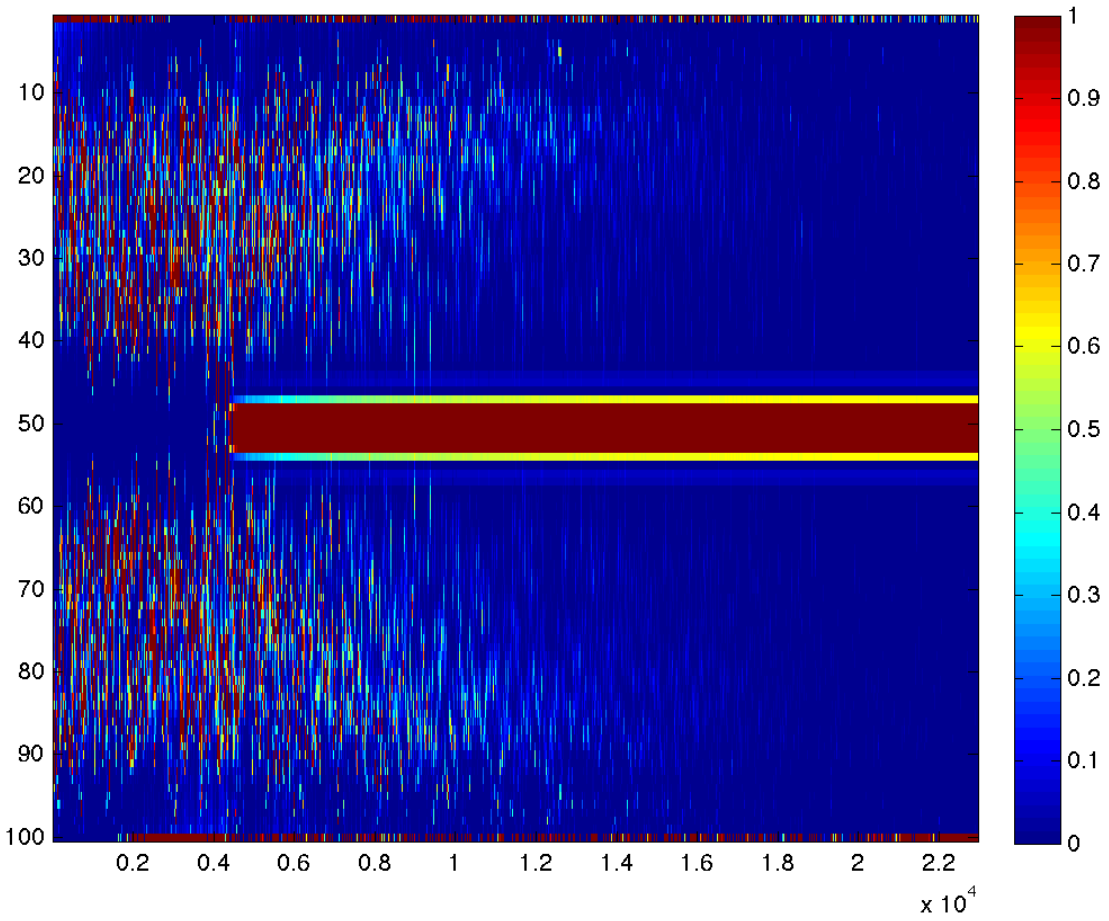

10
20
30
40
50
60
70
80
90
100
500
1000
1500
2000
2500
3000
3500
4000
4500
5000
5500
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0

Internet topology

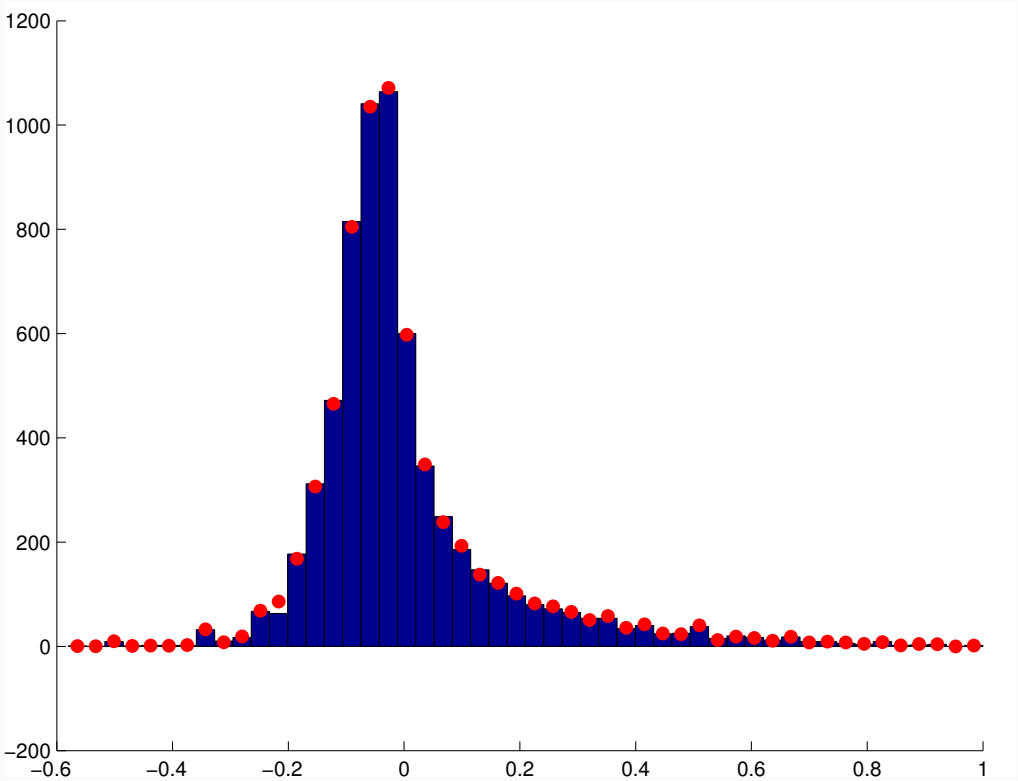




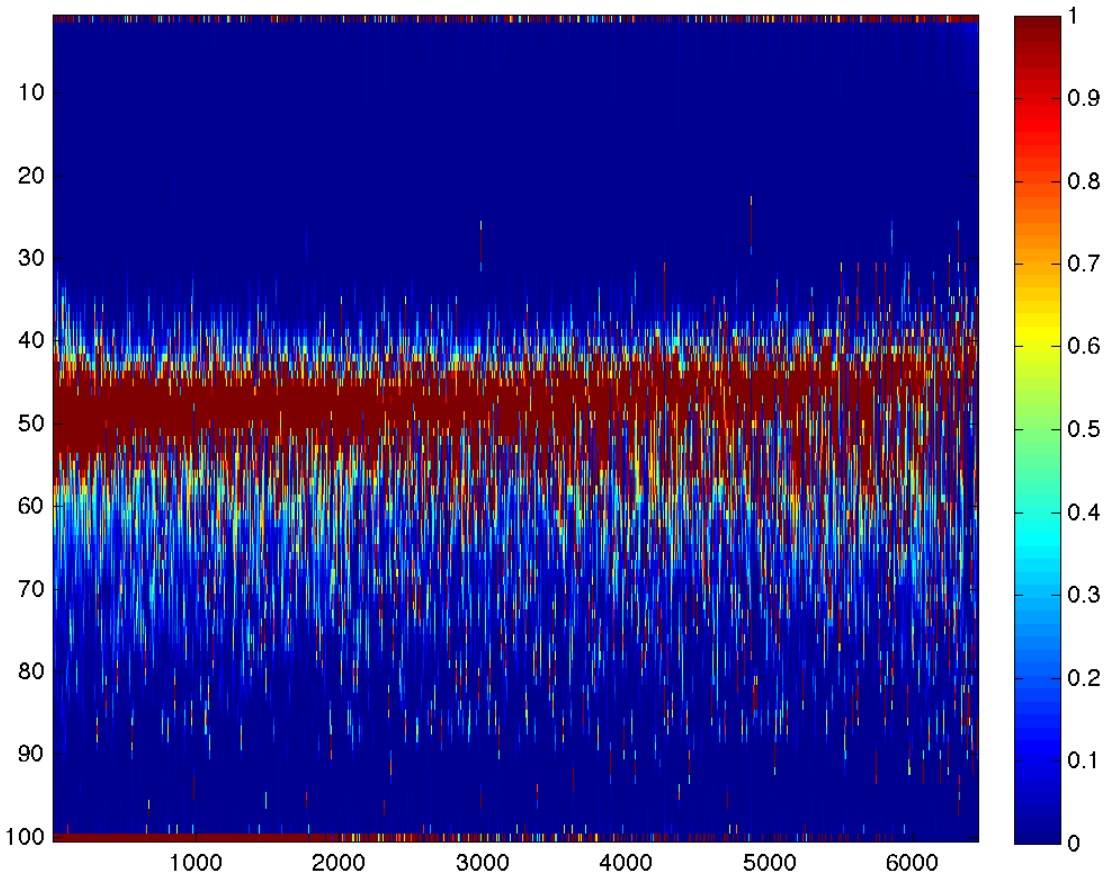
Internet topology (local)



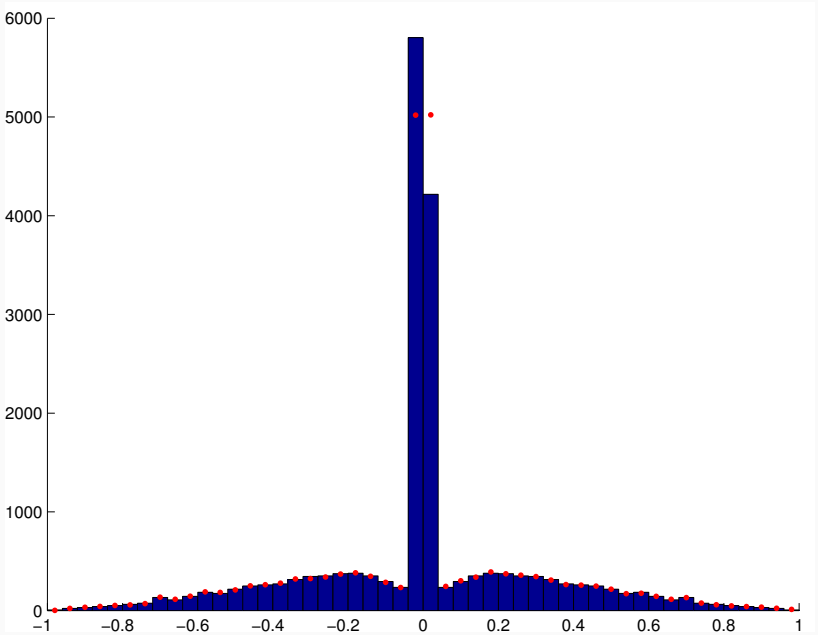
Marvel characters

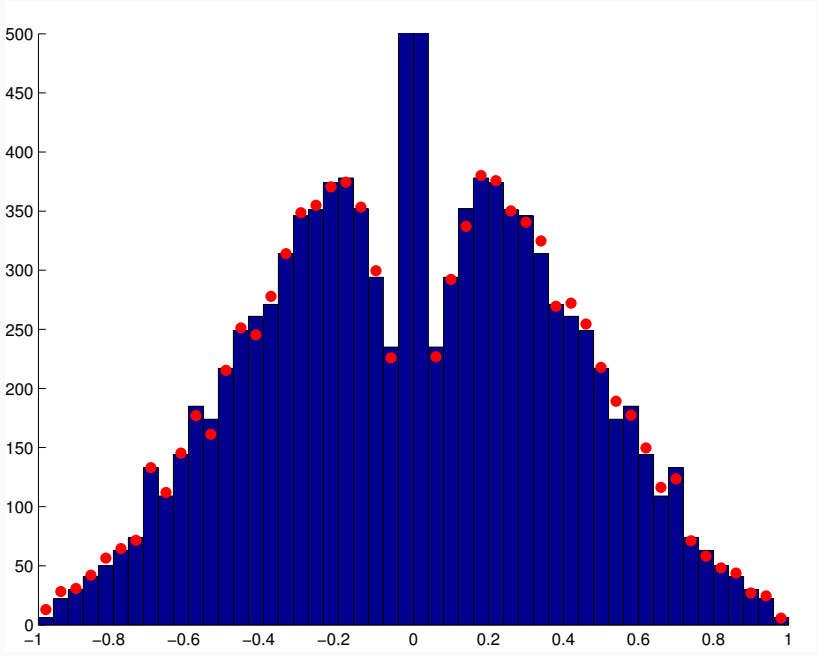


Marvel characters (local)

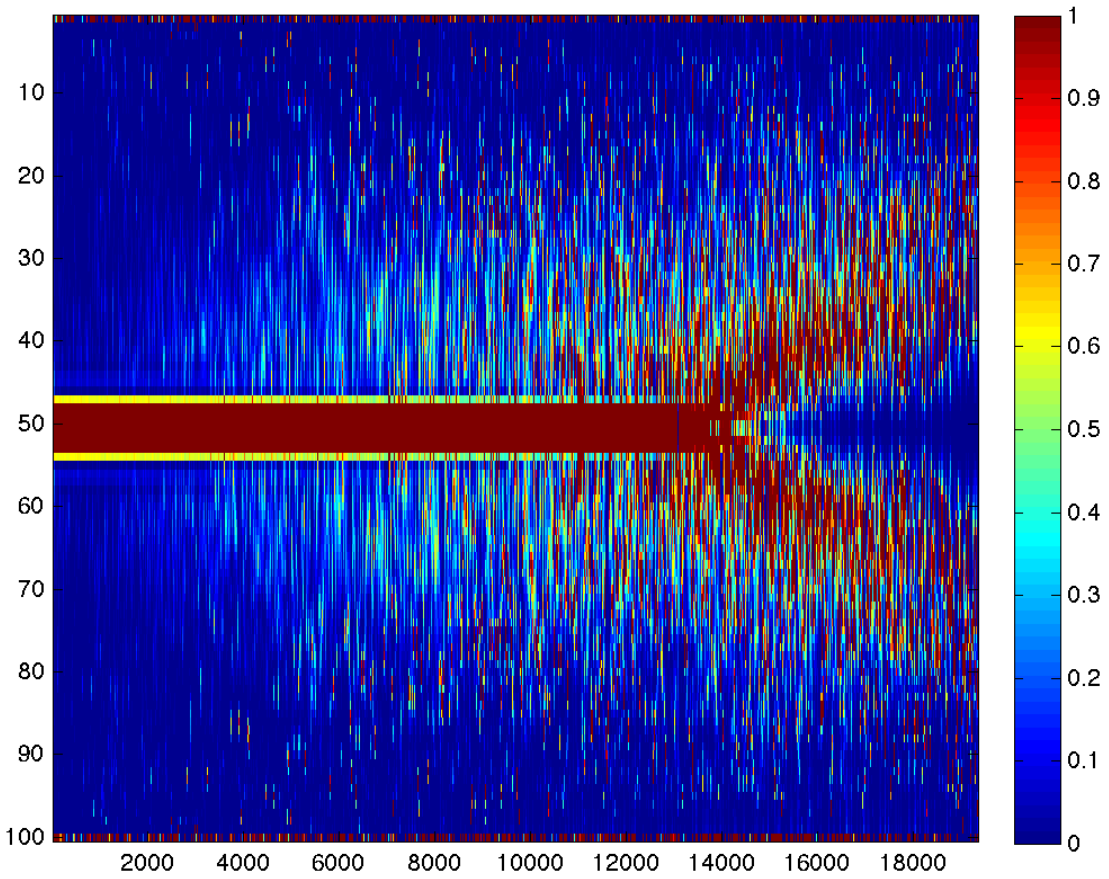


Marvel comics

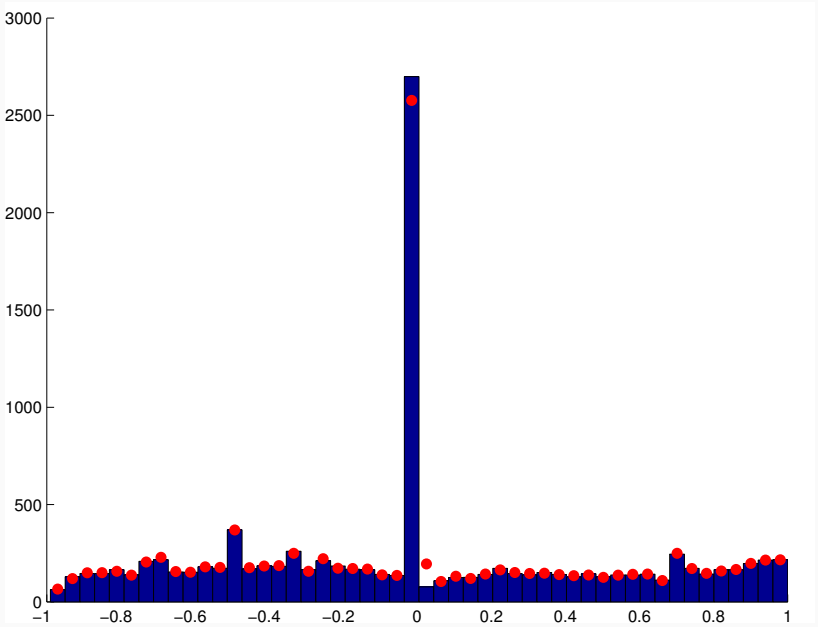


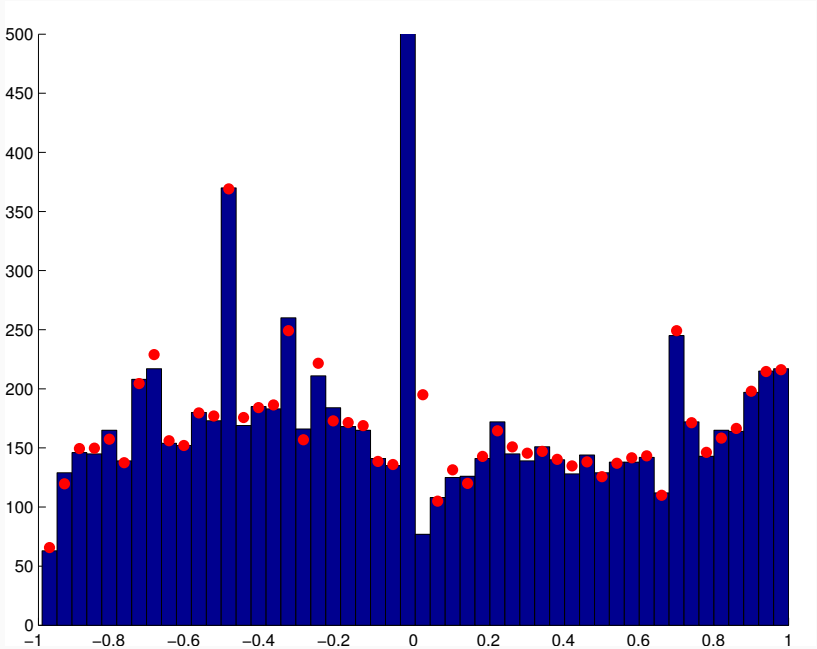


Marvel comics (local)

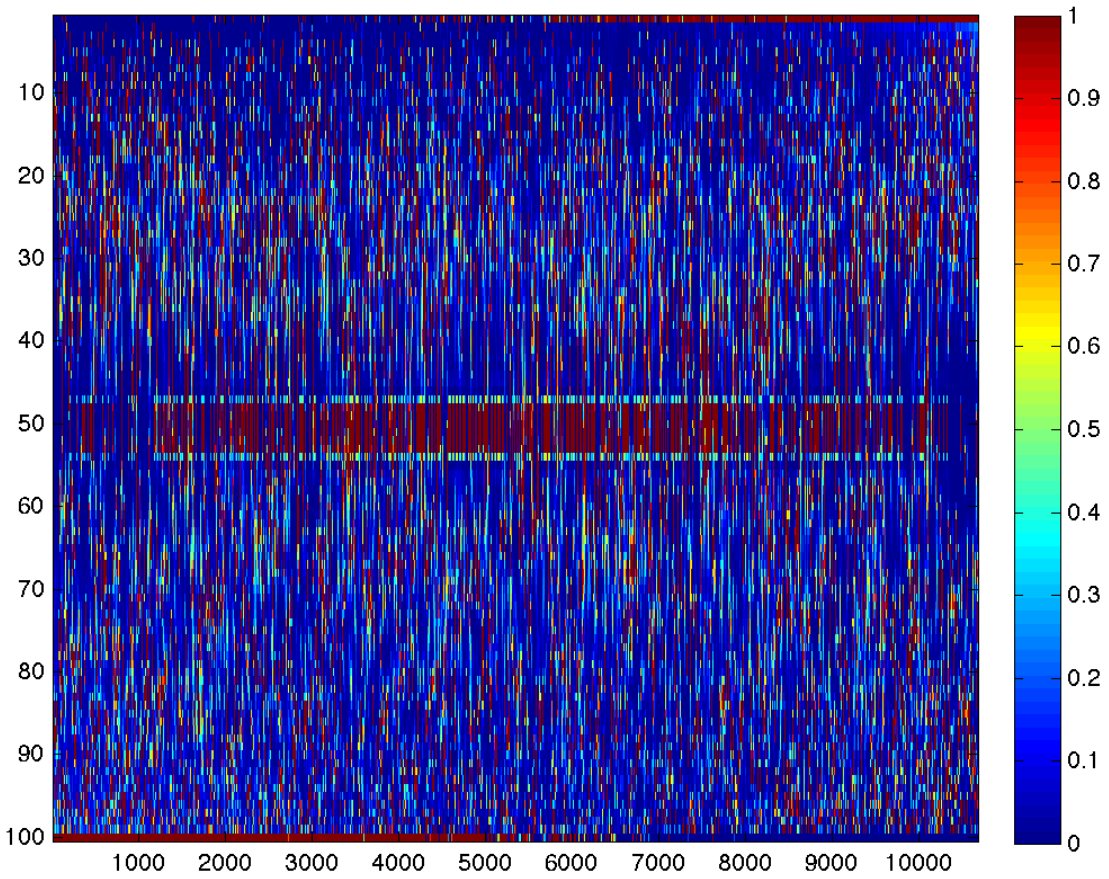


PGP

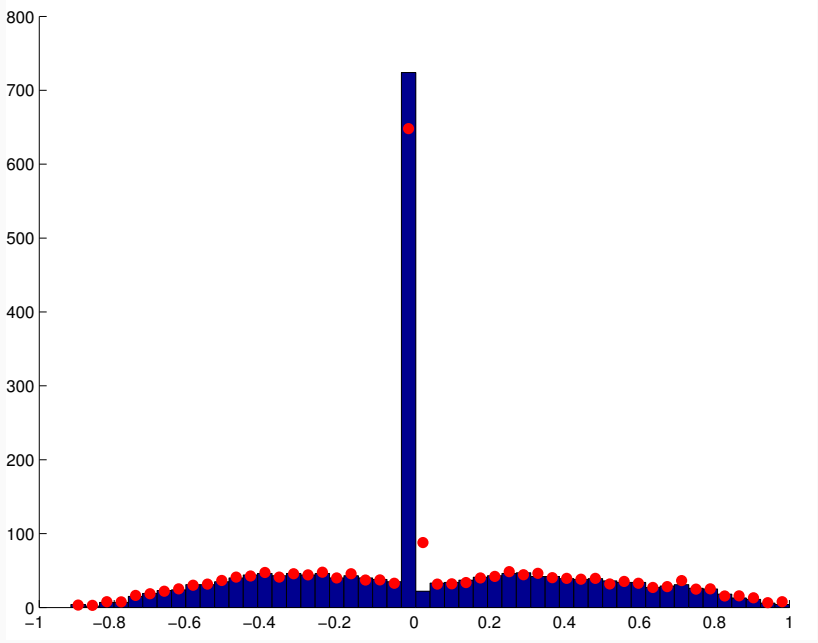

3000
2500
2000
1500
1000
500
0
−1
−0.8
−0.6
−0.4
−0.2
0
0.2
0.4
0.6
0.8
1

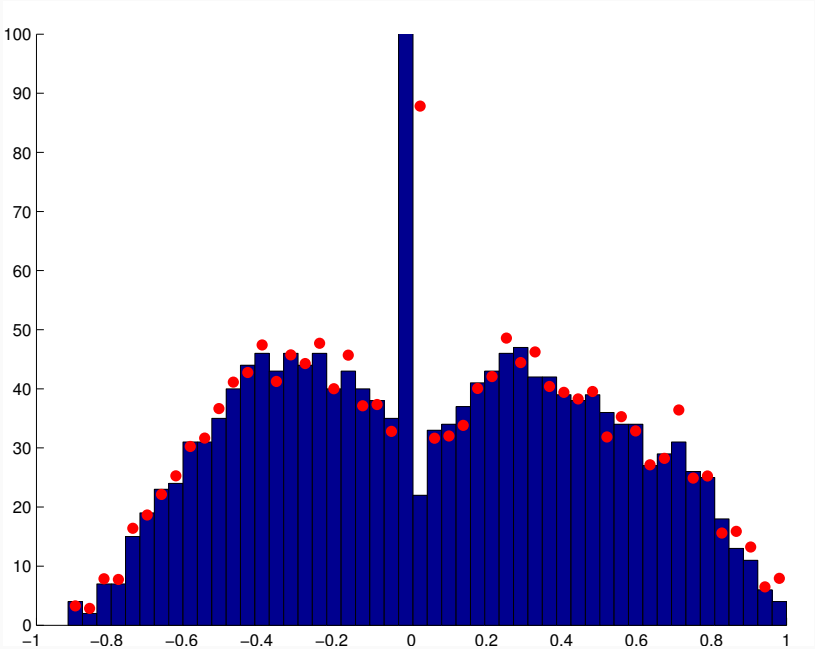

500
450
400
350
300
250
200
150
100
50
0
−1
−0.8
−0.6
−0.4
−0.2
0
0.2
0.4
0.6
0.8
1

PGP (local)

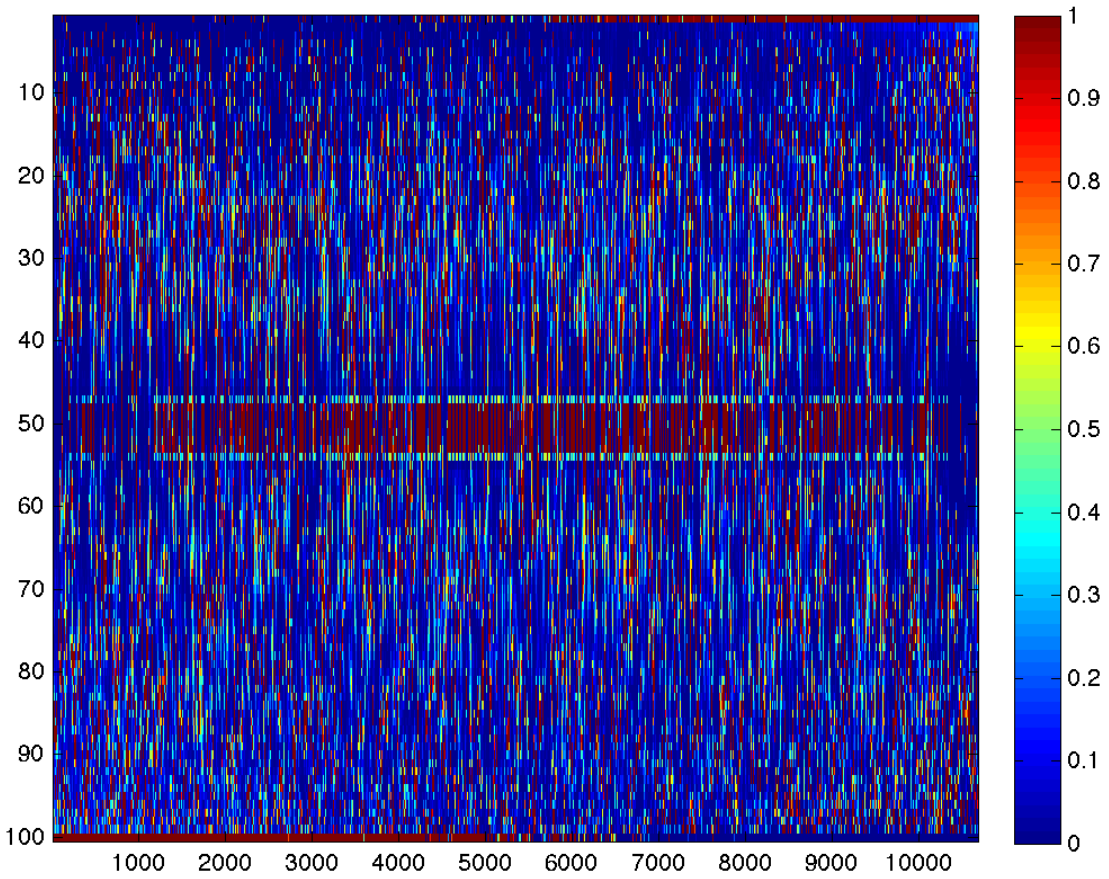

10
20
30
40
50
60
70
80
90
100
1000
2000
3000
4000
5000
6000
7000
8000
9000
10000
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0

Yeast


800
700
600
500
400
300
200
100
0
−1
−0.8
−0.6
−0.4
−0.2
0
0.2
0.4
0.6
0.8
1

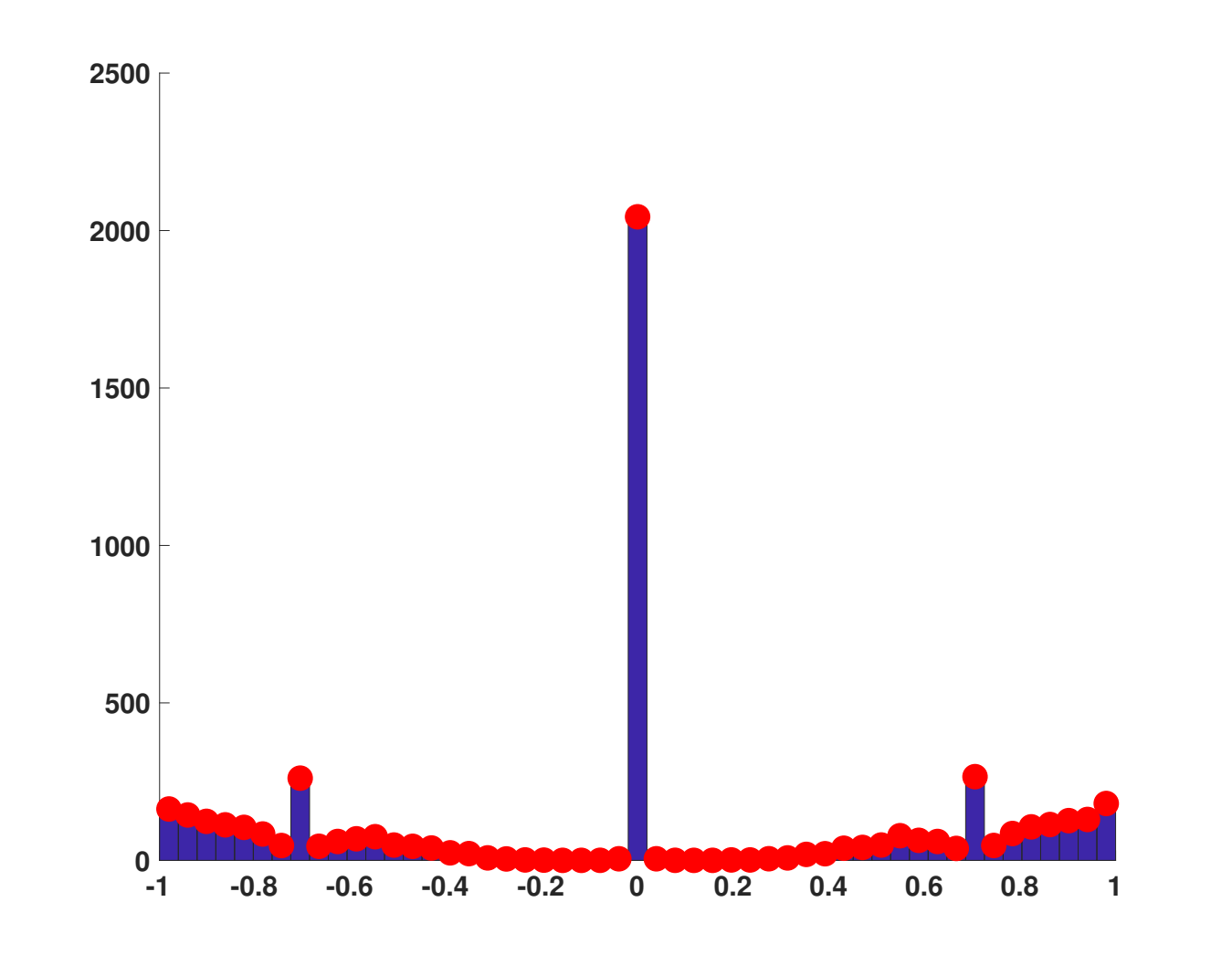

100
90
80
70
60
50
40
30
20
10
0
−1
−0.8
−0.6
−0.4
−0.2
0
0.2
0.4
0.6
0.8
1

Yeast (local)


10
20
30
40
50
60
70
80
90
100
1000
2000
3000
4000
5000
6000
7000
8000
9000
10000
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0

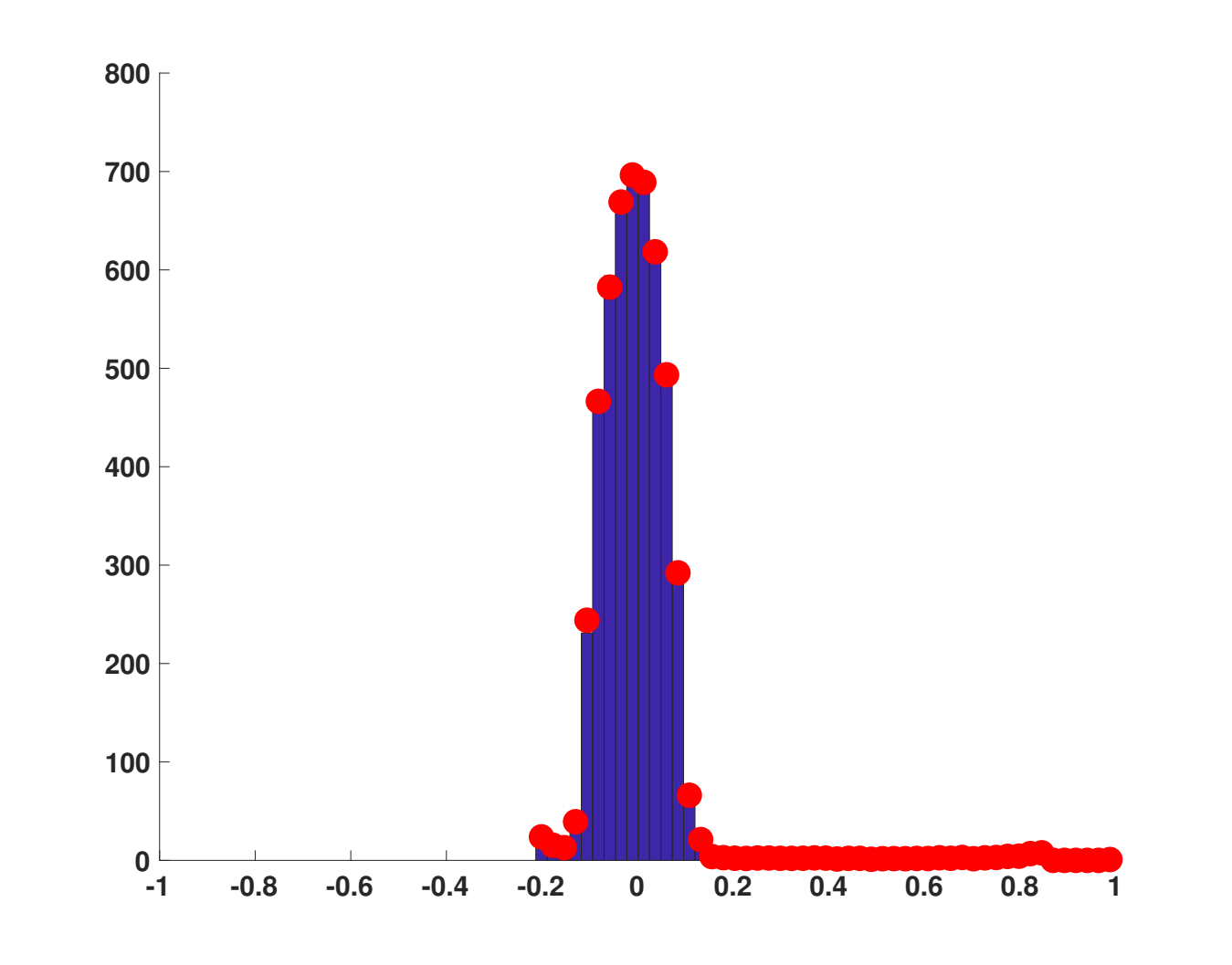
What about random graph models?

Barabási–Albert model



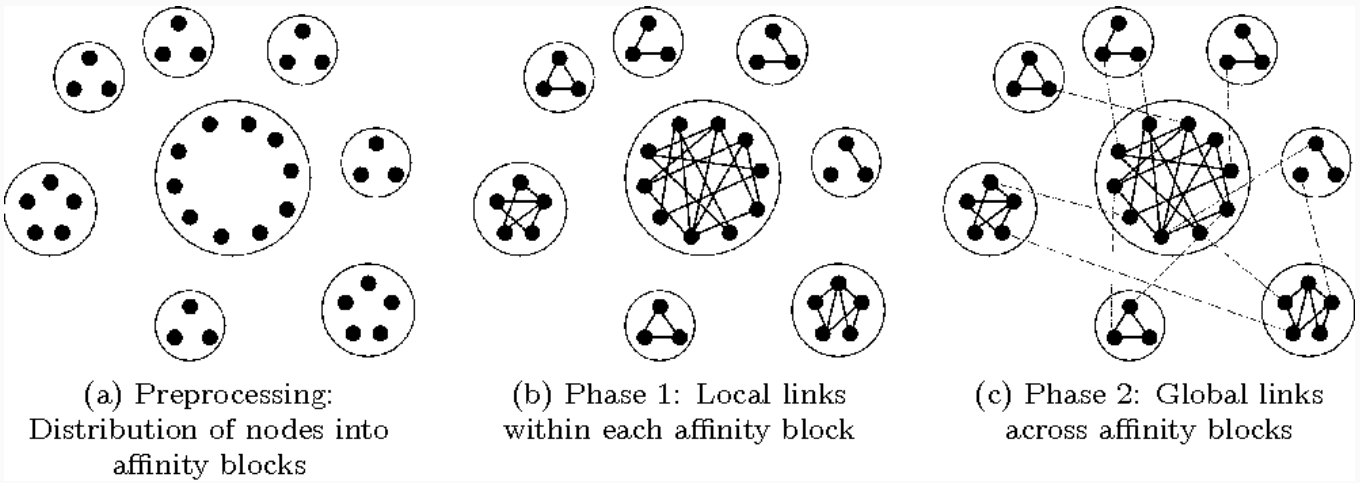
Scale-free network (5000 nodes, 4999 edges)

Watts–Strogatz



Small world network (5000 nodes, 260000 edges)

Model Verification: BTER



Kolda et al, SISC (36), 2014

Block Two-Level Erdős-Rényi model (BTER)

- 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

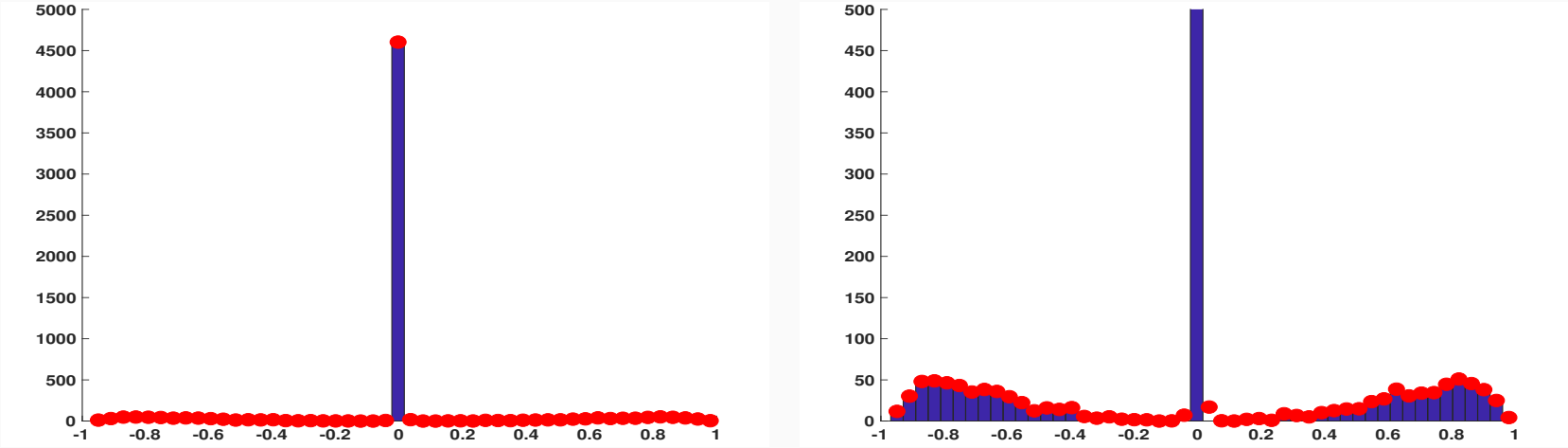


Figure 1: Erdos collaboration network.

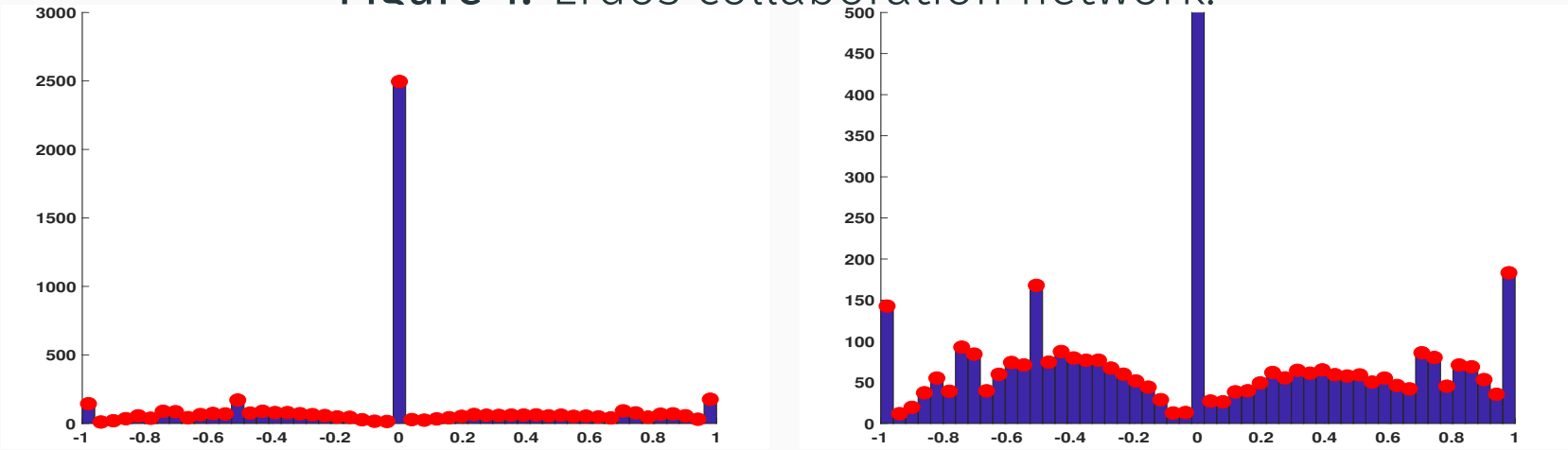
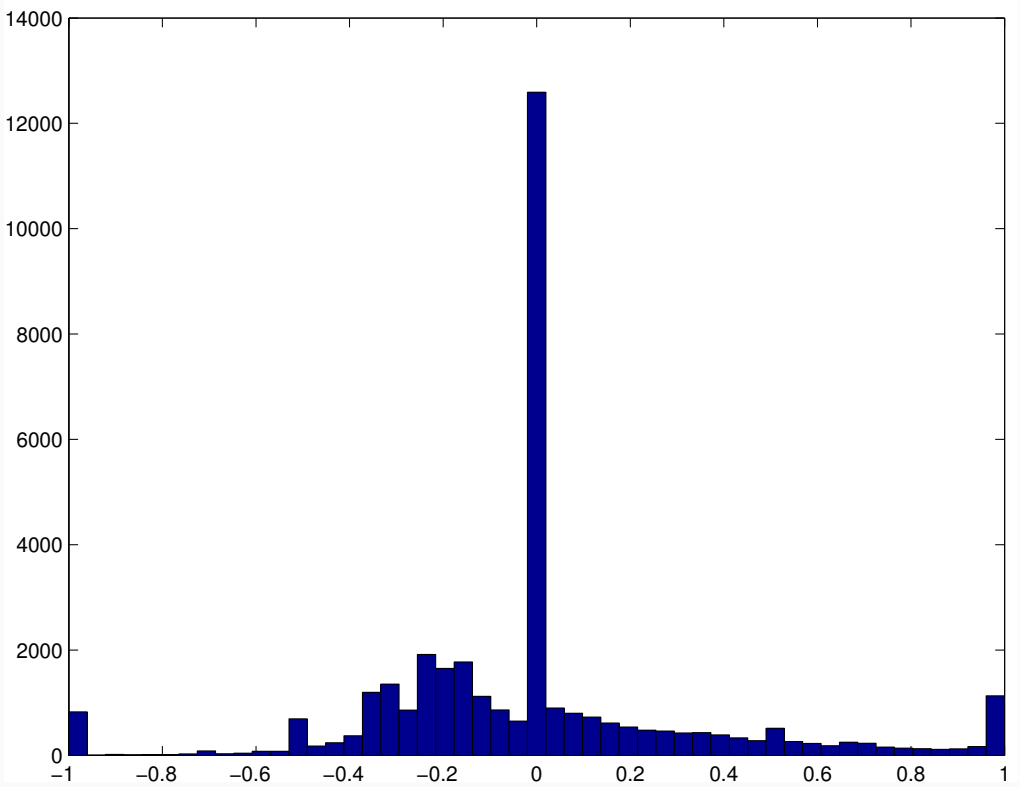


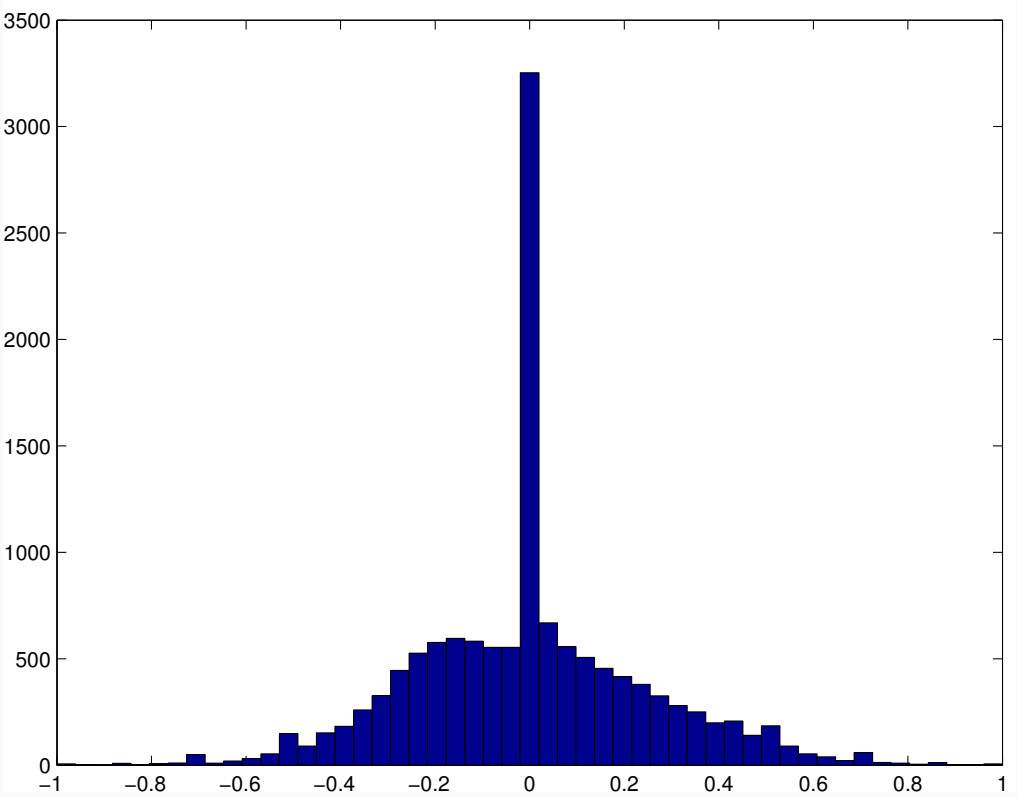
Figure 2: BTER model for Erdos collaboration network.

And a few more...

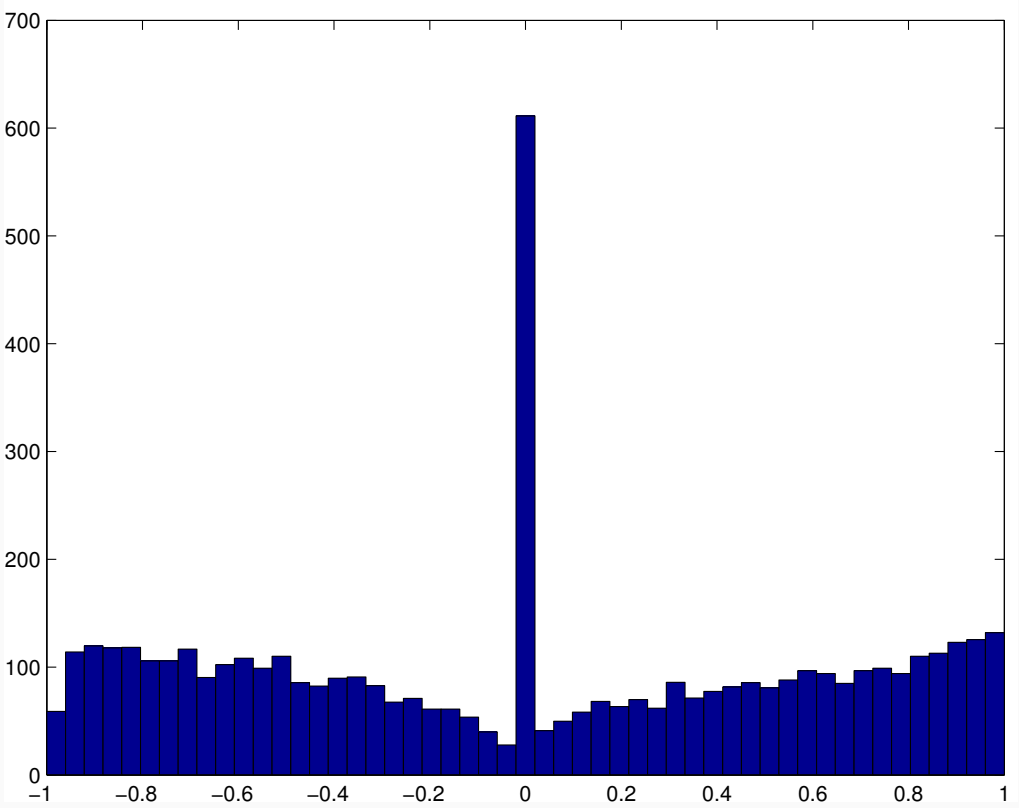
Enron emails (SNAP)



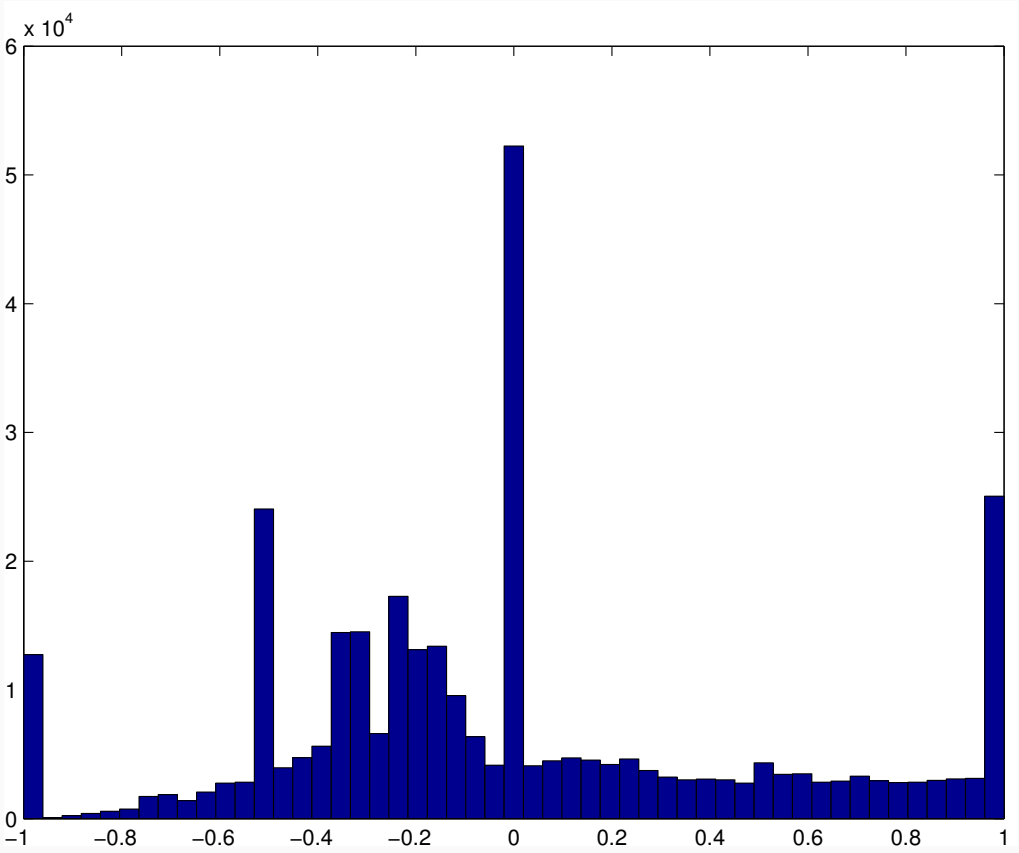
Reuters911 (Pajek)



US power grid (Pajek)

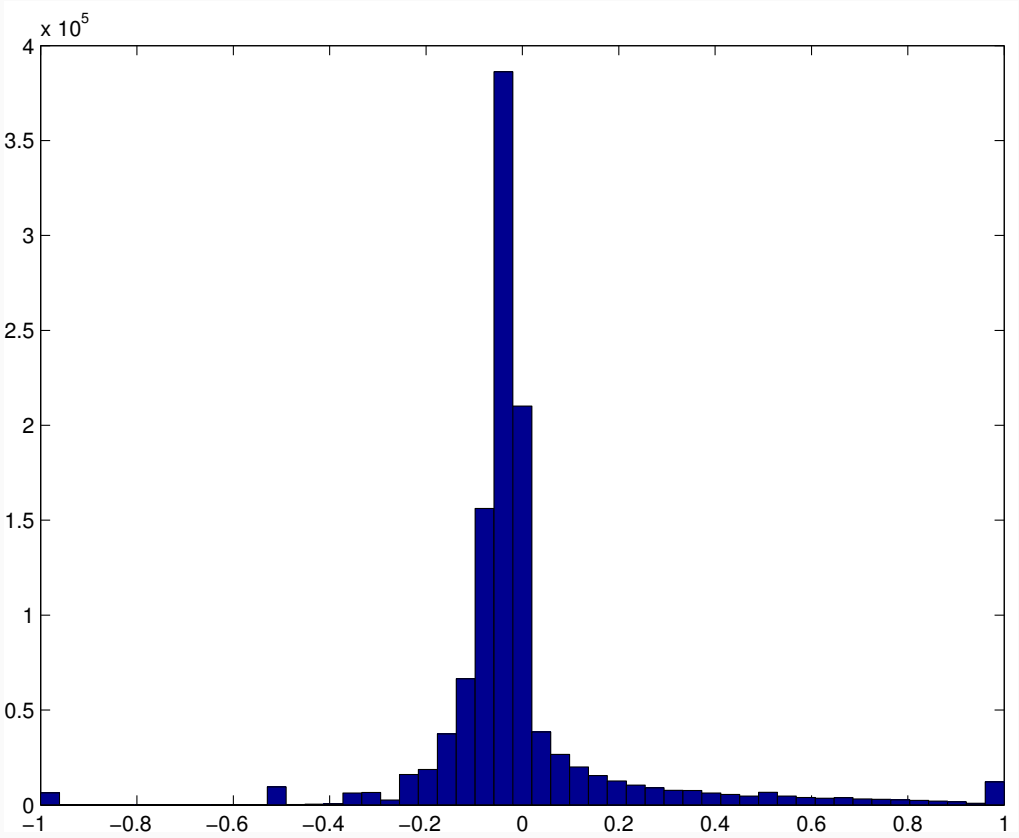


DBLP 2010 (LAW)



$N = 326186$, $nnz = 1615400$, 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?