Spectral Network Analysis: Beyond the Spectrum’s Edge

David Bindel
18 February 2022

Department of Computer Science
Cornell University
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).
Dynamics (operator on functions over manifold or graph)

- Diffuse according to heat kernel $\exp(-tLD^{-1})$
- Mixing rates and bottlenecks appear through spectrum
Dynamics (operator on functions over manifold or graph)

- Diffuse according to heat kernel \( \exp(-tLD^{-1}) \)
- Mixing rates and bottlenecks appear through spectrum
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
Dynamics (operator on functions over manifold or graph)

- Diffuse according to heat kernel $\exp(-tL^{-1}D)$
- Mixing rates and bottlenecks appear through spectrum
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
Dynamics (operator on functions over manifold or graph)

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

- Example: Spectral partitioning
- Measure cut sizes via \( x^T L x / 4, x \in \{ \pm 1 \}^n \), 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
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
For $H = -\delta + V$ with $\text{supp}(V) \subset \Omega$:

$$\begin{align*}
(H - k^2)\psi &= f \text{ on } \Omega \\
(\partial_n - B(k))\psi &= 0 \text{ on } \partial\Omega
\end{align*}$$

See resonance peaks (Breit-Wigner):

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

Associated with trapped “leaky” local modes.
Scattering and Resonances

Potential

Pole locations
Density of States in Physics

$E$

$E_F$

Metal  Semimetal  $p$-type intrin.  Semiconductor  $n$-type  Insulator
Today in Two Acts

• Act 1: Trapping, resonance, and local diffusions
• Act 2: Spectral densities
Random walker
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.
Run a diffusion (or diffusions) into Simon-Ando regime; then

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

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

What about very small clusters? Overlap?
Overlapping case

\[ \approx 14 \]
Dominant vectors for a block model example
Same space, different basis
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 $1 \geq \rho_1 \geq \rho_2 \geq \ldots > -1$. 
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 $(u_j, \rho_j)$ are Ritz pairs, approximate eigenpairs drawn from

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

Unconverged eigenpairs in a Krylov method are useful “local” objects on their own!
Local Expansion via Minimum One 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.

• Act 1: Trapping, resonance, and local diffusions
• Act 2: Spectral densities
“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 $\mathcal{H}_{lo} \supset \mathcal{H} \supset \mathcal{H}_{hi}$ to get $\lambda_{k,lo} \leq \lambda_k \leq \lambda_{k,hi}$

$$\lambda_k = \min_{\dim(\mathcal{V})=k, \mathcal{V} \subset \mathcal{H}} \left( \max_{v \in \mathcal{V}} \rho_L(v) \right)$$
Use $\mathcal{H}_{lo} \supset \mathcal{H} \supset \mathcal{H}_{hi}$ to get $\lambda_{k,lo} \leq \lambda_k \leq \lambda_{k,hi}$

$$\lambda_k = \min_{\dim(\mathcal{V})=k, \mathcal{V} \subset \mathcal{H}} \left( \max_{v \in \mathcal{V}} \rho_L(v) \right)$$
Weyl law (asymptotics for $N(x) = \{\# \text{ eigenvalues } \leq x\}$):

$$\lim_{x \to \infty} \frac{N(x)}{x^{d/2}} = (2\pi)^{-d} \omega_d \text{ vol}(\Omega).$$
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
Spectra define a generalized function (a density):

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

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

Consider

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

- 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 \n\[ h(s)/N = P(\text{self-return after time } s \text{ from uniform start}). \]
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 \( N \)).
- \( L \): ??
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)
Example: PGP Network

Spike (non-smoothness) at eigenvalues of 0 leads to inaccurate approximation.
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 $\nu_i = \nu_j$
Motifs in Spectrum

- $\lambda = 0$
- $\lambda = \pm 1/2$
- $\lambda = -1/2$
- $\lambda = \pm 1/\sqrt{2}$
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!


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.

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 (local)
Internet topology
Marvel characters
PGP
Yeast (local)
What about random graph models?
Barabási–Albert model

Scale-free network (5000 nodes, 4999 edges)
Small world network (5000 nodes, 260000 edges)
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...
US power grid (Pajek)
$N = 326186$, $nnz = 1615400$, 80 s (1000 moments, 10 probes)
$N = 1139905, \text{nnz} = 113891327, 2093 \text{ s (1000 moments, 10 probes)}$
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?