Numerical Methods for Data Science: Spectral Network Analysis, Part I

David Bindel

21 June 2019

Department of Computer Science
Cornell University

Lecture plan

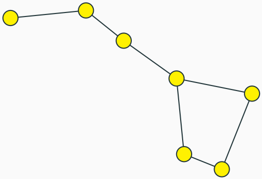
Three threads from "lay of the land" to current research:

- Monday: Latent Factor Models
- Wednesday: Scalable Kernel Methods
- Friday: Spectral Network Analysis
    - 1:30-2:30: Network spectra, optimization, and dynamics
    - 3:00-4:00: Network densities of states

Slides posted on web page (linked from my Cornell page).

Basics of Networks

Networks and Graphs

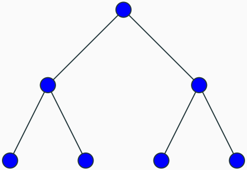


A graph (network) consists of

- Node (or vertex) set $\mathcal{V}$
- Edges $\mathcal{E} \subset \mathcal{V} \times \mathcal{V}$
  - *Undirected* if $(u, v) \in \mathcal{E} \implies (v, u) \in \mathcal{E}$
- Optional *edge weights* $\mathcal{E} \mapsto \mathbb{R}$

Can also add node weights or edge/node attributes.

Example Networks: Classic CS



Often small and/or highly structured:

- Finite state automata
- Search trees and DAGs
- Graphical models (correlated random variables)

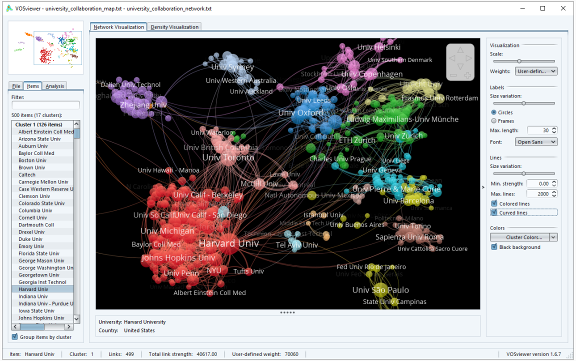
Mostly not the topic for today.

Example Networks: Physical

Connected to physical objects in 2D/3D:

- Rivers, roads, transportation
- Circuits and other electrical networks
- Pipe flow networks
- Computer networks

Example Networks: Citations



http://www.vosviewer.com/

Often directed, some very high-degree nodes, “small world”:

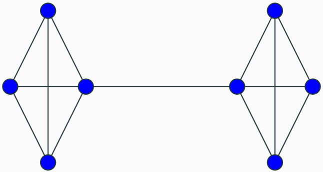
- Web pages, citation networks
- Purchase networks

Other Networks

Lots of others as well!

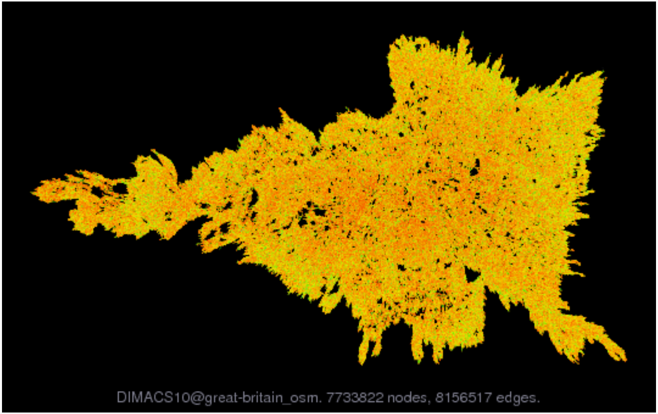
- Friendship networks
- Interaction networks (phone calls, etc)
- Food webs
- Protein interactions
- ...

The Big Questions: Connectedness



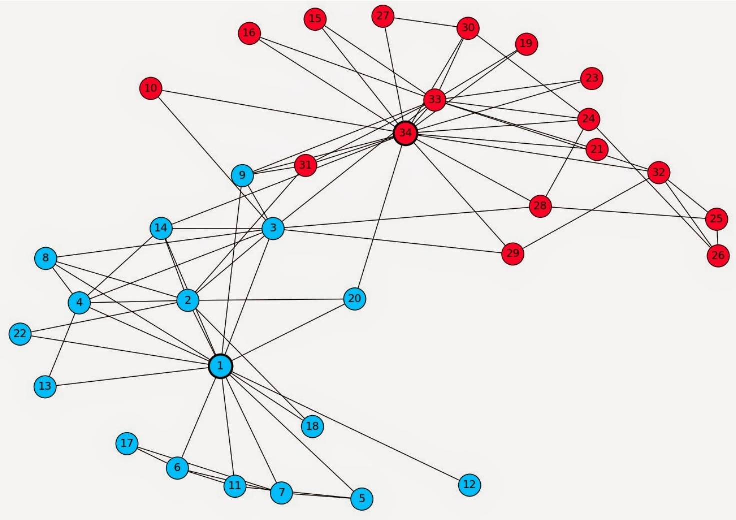
How “well-connected” is the network?

The Big Questions: Geometric Embedding



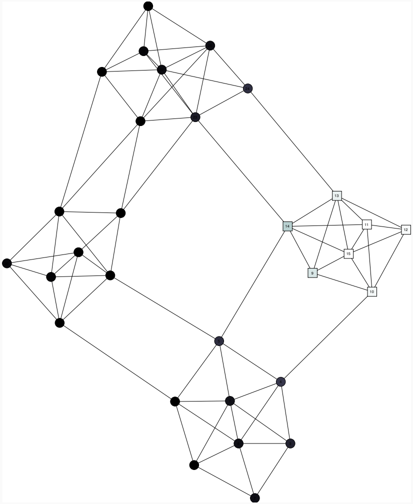
Is there an underlying geometry to the network?

The Big Questions: Centrality and Ranking



Who are important players?

The Big Questions: Clustering and Communities



What are the natural clusters or communities?

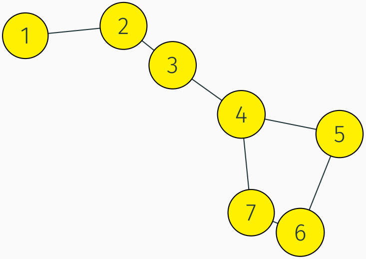
The Big Questions

One might ask many more questions:

- Graph alignment: Can we map between similar structures?
- Link prediction: Can we extrapolate the pattern?
- Cascade analysis: How does information spread?
- ...

Common approach: map to a linear algebra problem!

From Networks to Matrices



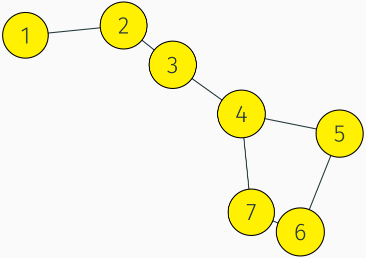
$$\begin{bmatrix} & 1 & & & & & \\ 1 & & 1 & & & & \\ & 1 & & 1 & & & \\ & & 1 & & 1 & & 1 \\ & & & 1 & & 1 & \\ & & & & 1 & & 1 \\ & & & 1 & & 1 & \end{bmatrix}$$

Adjacency A; unweighted is

$$a_{uv} = \begin{cases} 1, & (u, v) \in \mathcal{E} \\ 0, & \text{otherwise} \end{cases}$$

Degree $d_u = \sum_v a_{uv}$ is total adjacent edges (edge weight).
Distinguish in/out in directed case.

From Networks to Matrices

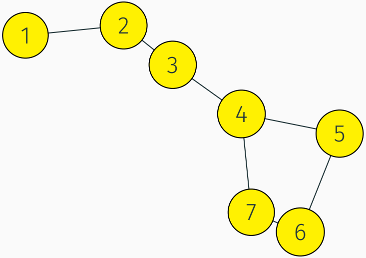


$$\begin{bmatrix} 1 & -1 & & & & & \\ & 1 & -1 & & & & \\ & & 1 & -1 & & & \\ & & & 1 & -1 & & \\ & & & & 1 & -1 & \\ & & & & & 1 & -1 \\ & & & -1 & & & 1 \end{bmatrix}$$

Differencing matrix (unweighted case)

$$g_{ew} = \begin{cases} 1, & e = (w, v) \\ -1, & e = (u, w) \\ 0, & \text{otherwise} \end{cases}$$

From Networks to Matrices



$$\begin{bmatrix} 1 & -1 & & & & & \\ -1 & 2 & -1 & & & & \\ & -1 & 2 & -1 & & & \\ & & -1 & 3 & -1 & & -1 \\ & & & -1 & 2 & -1 & \\ & & & & -1 & 2 & -1 \\ & & & -1 & & -1 & 2 \end{bmatrix}$$

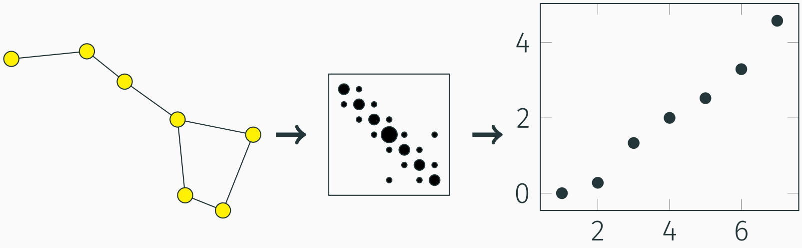
Laplacian $L = G^T G = D - A$; unweighted is

$$l_{uv} = \begin{cases} \text{degree } d_u, & u = v \\ -1, & (u, v) \in \mathcal{E} \\ 0, & \text{otherwise} \end{cases}$$

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: $W = A^2 \odot A$

Spectral Network Analysis



Three stories of why eigenstuff matters:

- Dynamics and diffusion
- Measure and counting
- Kernels and geometry

Story 1: Dynamics and Diffusion

The Random Walker

Graph with adjacency A (a_{ij} denotes edge j to i),

$$p_{ij} = a_{ij}/d_j = \text{probability of step } j \to i$$

Let $w_j(t)$ denote probability a walker is at node j at time t; then

$$w(t+1) = Pw(t) = P^t w(0)$$

Equations for a discrete time Markov chain $\equiv$ power iteration

Suppose P diagonalizable and the walk is ergodic. Then

$$P = V\Lambda V^{-1}, \quad P^t = V\Lambda^t V^{-1}$$

with $\lambda_1 = 1$ and $|\lambda_j| < 1$ for $j \neq 1$.

Let w^∞ denote the stationary distribution; then

$$\|P^t - (w^\infty)e/n^T\| \leq C|\lambda_2|^t$$

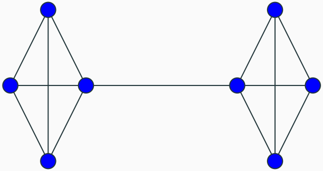
where e is the vector of all ones. Rate of convergence determined by second-largest eigenvalue modulus $|\lambda_2|$.

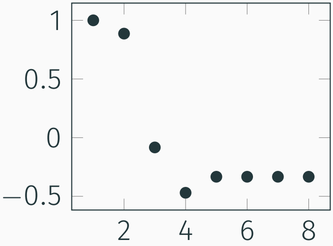
PageRank: Random walk + “teleport” with probability α

$$w(t+1) = (1-\alpha)Pw(t) + \alpha w^{\text{ref}}$$

Get a nice spectral gap $(1-\alpha)$, fast convergence to stationarity.

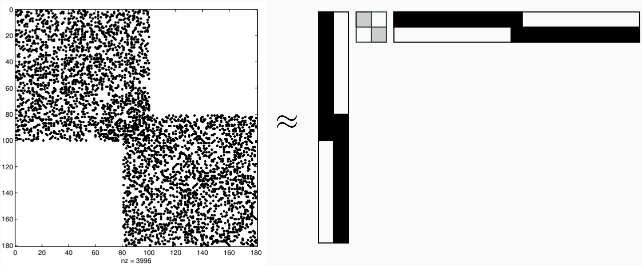
Convergence Fast and Slow





- Eigenvalue at 1: stationary state
- Size of $|\lambda_2|$ determines convergence
- What do the other eigenvalues mean?
- Interpreting two eigenvalues near 1 in dumbbell:
  - Distribution rapidly approaches 2D subspace initially
  - Then slower relaxation to equilibrium
- Corresponds to *metastable states* (c.f. Simon-Ando theory)

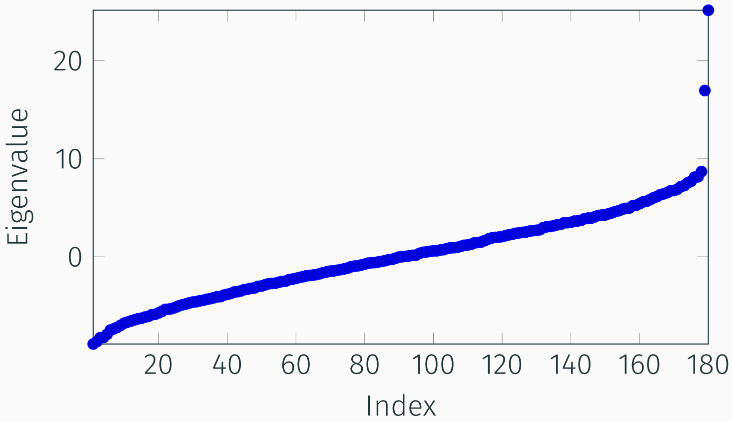
Dynamics in a Block Model



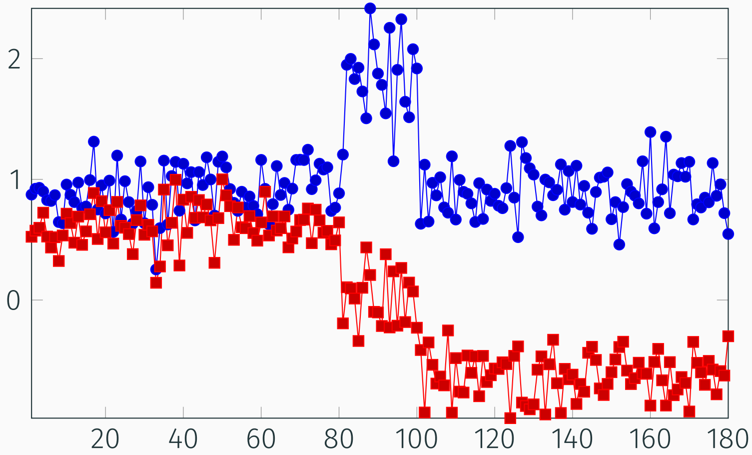
Composite model: $A \approx S\,\mathrm{diag}(\beta)S^T$, $S \in \{0,1\}^{n \times c}$

- Motivation: possibly-overlapping random graphs
- Columns of S are one basis for range space
- Want to go from some general basis back to S

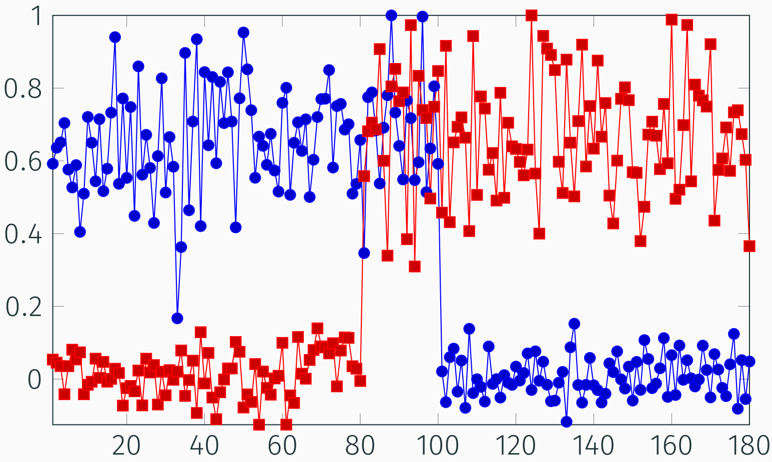
Spectrum for a Block Model Sample



Dominant Vectors



Same Space, Different Basis



And Beyond!

Many variations:

- Path counting and power of A
- Continuous time walks and $\exp(-tL)$
- Hub/authority importance, HITS, and SVD power iteration
- ...

All involve linear time-invariant systems.

Story 2: Measure and Counting

Measurement by Quadratic Forms

Indicate $V' \subseteq V$ by $s \in \{0,1\}^n$. Measure subgraph:

$$s^T A s = |E'| = \text{internal edges}$$
$$s^T D s = \text{edges incident on subgraph}$$
$$s^T L s = \text{edges between } V' \text{ and } \bar{V}'$$
$$s^T B s = \text{“surprising” internal edges}$$

Modularity matrix is $B = A - \frac{dd^T}{2n}$

Graph Bisection

Idea: Find $s \in \{0, 1\}^n$ such that $e^T s = n/2$ to

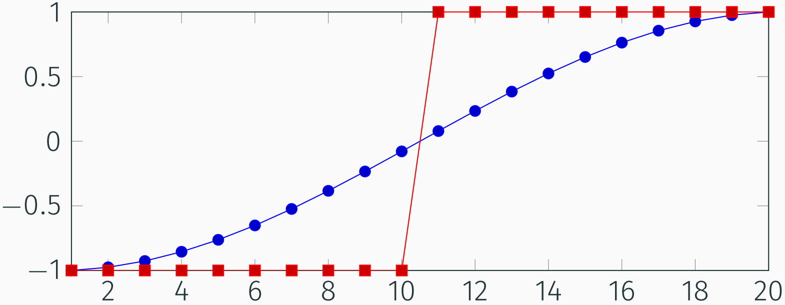
- minimize $s^T L s$ (min cut)
- maximize $s^T B s$ (max modularity)

Equivalently: Find $\bar{s} \in \{\pm 1\}^n$ such that $e^T \bar{s} = 0$ to

- minimize $\bar{s}^T L \bar{s} = s^T L s$ or
- maximize $\bar{s}^T B \bar{s} = s^T B s$

Oops — NP hard!

Relax!



Hard:	$\min \bar{s}^T L \bar{s}$	s.t.	$e^T \bar{s} = 0,$	$\bar{s} \in \{\pm 1\}^n.$
Easy:	$\min v^T L v$	s.t.	$e^T v = 0,$	$v \in \mathbb{R}^n, \|v\|^2 = n.$

Cheeger Inequality

Relaxation gives half of *Cheeger inequality*

$$h(G) \geq \lambda_2 \geq \frac{h^2(G)}{2}$$

where

$$h(G) = \min \left\{ \frac{|\partial A|}{|A|} : A \subset \mathcal{V}, 0 < |A| \leq \frac{1}{2}|\mathcal{V}| \right\}$$

Relation between bottlenecks and λ_2.
Also relevant to mixing picture from dynamics story!

Rayleigh Quotients

$$\frac{s^T A s}{s^T s} = \text{mean internal degree in subgraph}$$
$$\frac{s^T L s}{s^T s} = \text{edges cut between } V' \text{ and } \bar{V}'$$

$$\frac{s^T A s}{s^T D s} = \text{fraction of incident edges internal to } V'$$
$$\frac{s^T L s}{s^T D s} = \text{fraction of incident edges cut}$$

$$\frac{s^T B s}{s^T s} = \text{mean ``surprising'' internal degree in subgraph}$$
$$\frac{s^T B s}{s^T D s} = \text{mean fraction of internal degree that is surprising}$$

Rayleigh Quotients and Eigenvalues

Basic connection (M spd):

$$\frac{x^T K x}{x^T M x} \text{ stationary at x} \qquad \Longleftrightarrow \qquad Kx = \lambda M x$$

Easy despite lack of convexity.

Limits of Rayleigh Quotients

But small variations kill us:

$$\max_{x \neq 0} \frac{x^T A x}{\|x\|_2^2} = \lambda_{\max}(A), \text{ but}$$
$$\max_{x \neq 0} \frac{x^T A x}{\|x\|_1^2} = 1 - \omega^{-1}$$

where ω is the max clique size (Motzkin-Strauss).

Decompose:

$$W^T M W = I \text{ and } W^T K W = \Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n).$$

For any $x \neq 0$,

$$\frac{x^T K x}{x^T M x} = \sum_{j=1}^{n} \lambda_j z_j^2, \text{ where } z = \frac{W^{-1}x}{\|W^{-1}x\|_2}.$$

So

$$\frac{s^T K s}{s^T M s} \approx \lambda_{\max} \implies s \approx \sum_{\lambda_j \approx \lambda_{\max}} w_j z_j.$$

So look at invariant subspaces for extreme eigenvalues.

Story 3: Kernels and Geometry

Classical Multi-Dimensional Scaling

Data: square distances between points:

$$
\begin{aligned}
[D^{(2)}]_{ij} &= d_{ij}^2 = \|u_i - u_j\|^2 \\
&= \|u_i\|^2 - 2\langle u_i, u_j \rangle + \|u_j\|^2 \\
r_i &\equiv \|u_i\|^2 \\
D^{(2)} &= r^{(2)} e^T - 2UU^T + e(r^{(2)})^T
\end{aligned}
$$

Double centering:

$$
B = -\frac{1}{2} J D^{(2)} J = UU^T, \quad J \equiv I - \frac{1}{n} ee^T
$$

Decompose $B = XX^T$ (e.g. via eigs) to get coordinates for data.
Note that B looks like a kernel matrix / Gram matrix.

Resistance Distance

Think of resistors on each edge, consider net flow between source at node *i* and sink at node *j*:

$$d_{ij}^2 = (e_i - e_j)L^\dagger(e_i - e_j)$$

This acts like a squared distance — find coordinates!

$$L = V\Lambda V^T \quad \Longrightarrow \quad L^\dagger = V \begin{bmatrix} 0 & & & \\ & \lambda_2^{-1} & & \\ & & \ddots & \\ & & & \lambda_n^{-1} \end{bmatrix} V^T$$

Truncate to get a low-dimensional embedding.

Leveraging Geometry

General idea: M a kernel matrix / Gram matrix over data

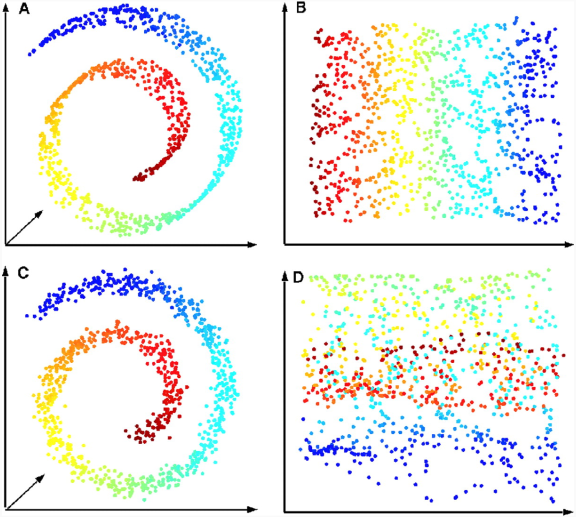
- Eigendecomposition: $M = V \Lambda V^T$
- Coordinates for i: $(v_{ij}\sqrt{\lambda_j})_{j=1}^n$
- Truncate to get low-dimensional embedding

Then do processing over geometry

- Graph layout (use as node coordinates)
- Geometric partitioning (e.g. inertial methods)
- Geometric clustering (k-means)

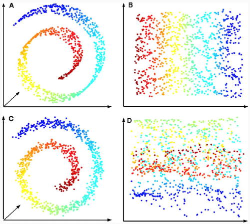
Other than resistance distance, useful metrics?

Beyond Resistance



What if we don't believe global geometry?

Isomap Idea



Distance matrix for points:

- Ordinary pairwise distance nearby
- Graph distance far away

And then apply MDS.

Summary and Preview

The Three Stories

Three stories of why eigenstuff matters:

- Dynamics and diffusion
- Measure and counting
- Kernels and geometry

The Missing Link

So far, focus is the *extreme* eigenpairs:

- Metastable states in dynamics story
- Relaxed solutions in measure/counting story
- Reconstructed coordinates in geometric story

What about the interior of the spectrum?

The Missing Link