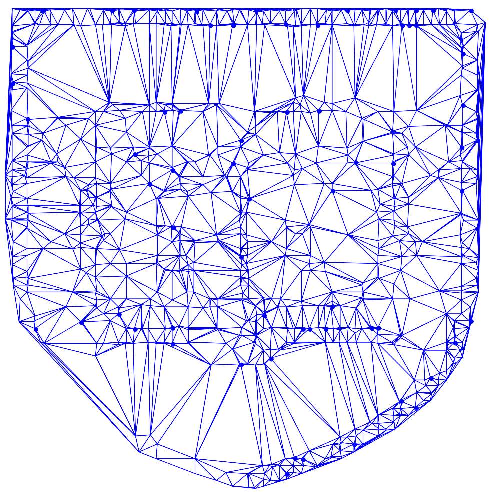
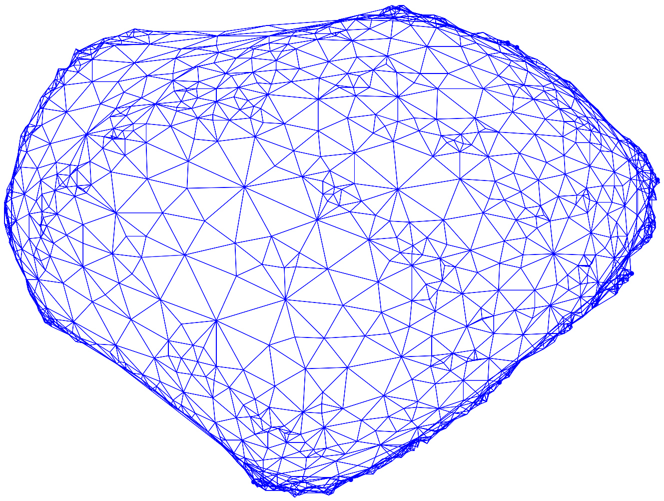
Spectral Graph Theory and its Applications





Daniel A. Spielman
Dept. of Computer Science
Program in Applied Mathematics
Yale Unviersity

Outline

Adjacency matrix and Laplacian

Intuition, spectral graph drawing

Physical intuition

Isomorphism testing

Random walks

Graph Partitioning and clustering

Distributions of eigenvalues and compression

Computation

What I'm Skipping

Matrix-tree theorem.

Most of algebraic graph theory.

Special graphs (e.g. Cayley graphs).

Connections to codes and designs.

Lots of work by theorists.

Expanders.

The Adjacency Matrix

$$A(i,j) = \begin{cases} 1 & \text{if } (i,j) \in E \\ 0 & \text{otherwise} \end{cases}$$

$$\begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

λ is eigenvalue and v is eigenvector if

$$Av = \lambda v$$

Think of $v \in \mathbb{R}^V$, or even better $v : V \to \mathbb{R}$

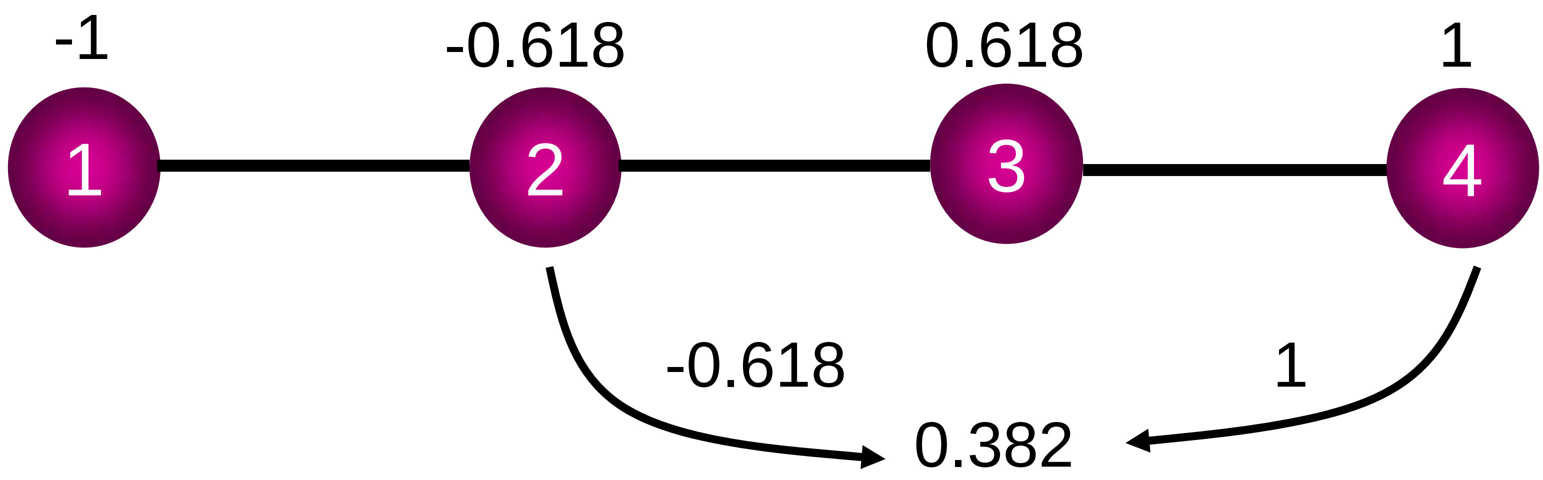
Symmetric -> n real eigenvalues and
real eigenvectors form orthonormal basis

Example

$$\begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} -1 \\ -0.618 \\ 0.618 \\ 1 \end{pmatrix} = 0.618 \begin{pmatrix} -1 \\ -0.618 \\ 0.618 \\ 1 \end{pmatrix}$$

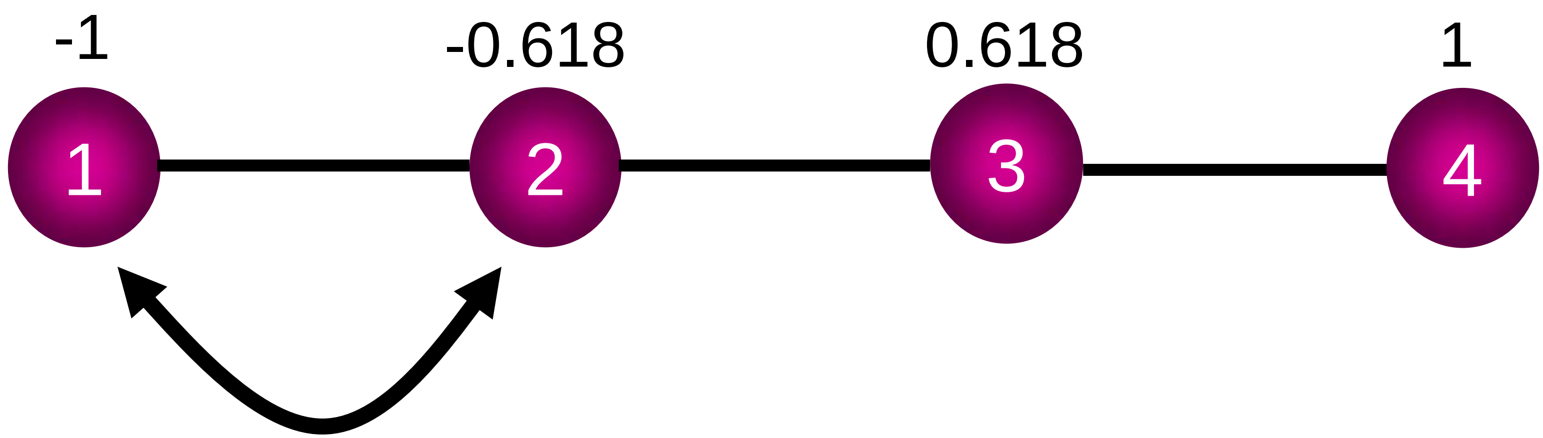
Example

$$\begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} -1 \\ -0.618 \\ 0.618 \\ 1 \end{pmatrix} = 0.618 \begin{pmatrix} -1 \\ -0.618 \\ 0.618 \\ 1 \end{pmatrix}$$



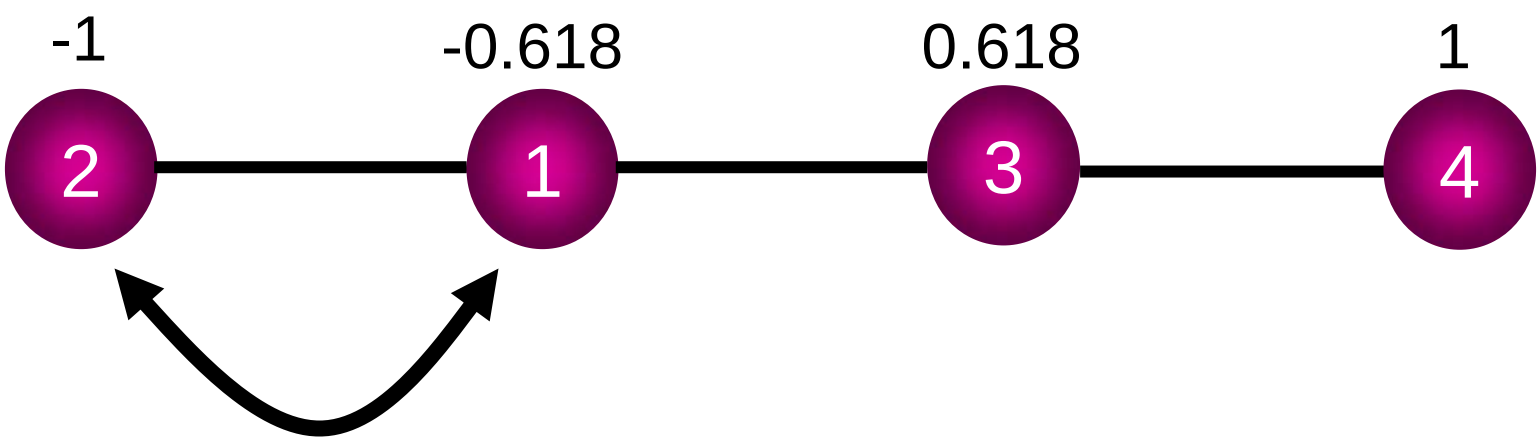
Example: invariant under re-labeling

$$\begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} -1 \\ -0.618 \\ 0.618 \\ 1 \end{pmatrix} = 0.618 \begin{pmatrix} -1 \\ -0.618 \\ 0.618 \\ 1 \end{pmatrix}$$



Example: invariant under re-labeling

$$\begin{pmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} -0.618 \\ -1 \\ 0.618 \\ 1 \end{pmatrix} = 0.618 \begin{pmatrix} -0.618 \\ -1 \\ 0.618 \\ 1 \end{pmatrix}$$



Operators and Quadratic Forms

View of A as an operator:

$$y = Ax \qquad\qquad y(i) = \sum_{j:(i,j)\in E} x(j)$$

View of A as quadratic form:

$$x^T A x = \sum_{(i,j)\in E} x(i)x(j)$$

if $Ax = \lambda x$ and $\|x\| = 1$ then $x^T A x = \lambda$

Laplacian: natural quadratic form on graphs

$$x^T L x = \sum_{(i,j)\in E} (x(i) - x(j))^2$$

$$L = D - A$$ where D is diagonal matrix of degrees

$$\begin{pmatrix} 1 & -1 & 0 & 0 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ 0 & 0 & -1 & 1 \end{pmatrix}$$

Laplacian: fast facts

$L\mathbf{1} = \mathbf{0}$ so, zero is an eigenvalue

$$0 = \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$$

If k connected components, $0 = \lambda_k < \lambda_{k+1}$

Fiedler ('73) called λ_2

"algebraic connectivity of a graph"

The further from 0, the more connected.

Embedding graph in line (Hall '70)

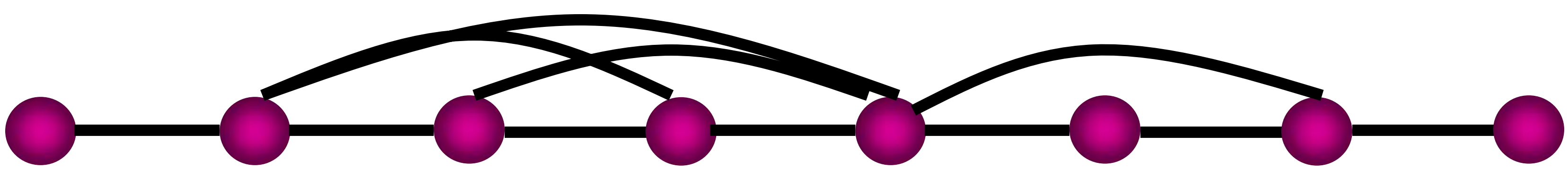
map $V \rightarrow \mathbb{R}$

minimize $\displaystyle\sum_{(i,j)\in E} (x(i) - x(j))^2 = x^T L x$

trivial solution: $x = \mathbf{1}$ So, require $x \perp \mathbf{1}$

Solution $x = v_2$

Atkins, Boman, Hendrickson '97:
Gives correct embedding for graphs like



Courant-Fischer definition of eigvals/vecs

$$\lambda_1 = \min_{x \neq 0} \frac{x^T L x}{x^T x} \qquad v_1 = \arg\min_{x \neq 0} \frac{x^T L x}{x^T x}$$

$$\lambda_2 = \min_{x \perp v_1} \frac{x^T L x}{x^T x} \qquad v_2 = \arg\min_{x \perp v_1} \frac{x^T L x}{x^T x}$$

(here $v_1 = \mathbf{1}$)

$$\boxed{\begin{aligned} \lambda_k &= \min_{S \text{ of dim } k} \ \max_{x \in S} \frac{x^T L x}{x^T x} \\ v_k &= \arg \min_{x \perp v_1, \ldots, v_{k-1}} \frac{x^T L x}{x^T x} \end{aligned}}$$

Embedding graph in plane (Hall '70)

map $V \longrightarrow \mathbb{R}^2$ $\vec{x}(i) \in \mathbb{R}^2$

minimize $\sum_{(i,j)\in E} (\mathrm{dist}(\vec{x}(i), \vec{x}(j))^2$

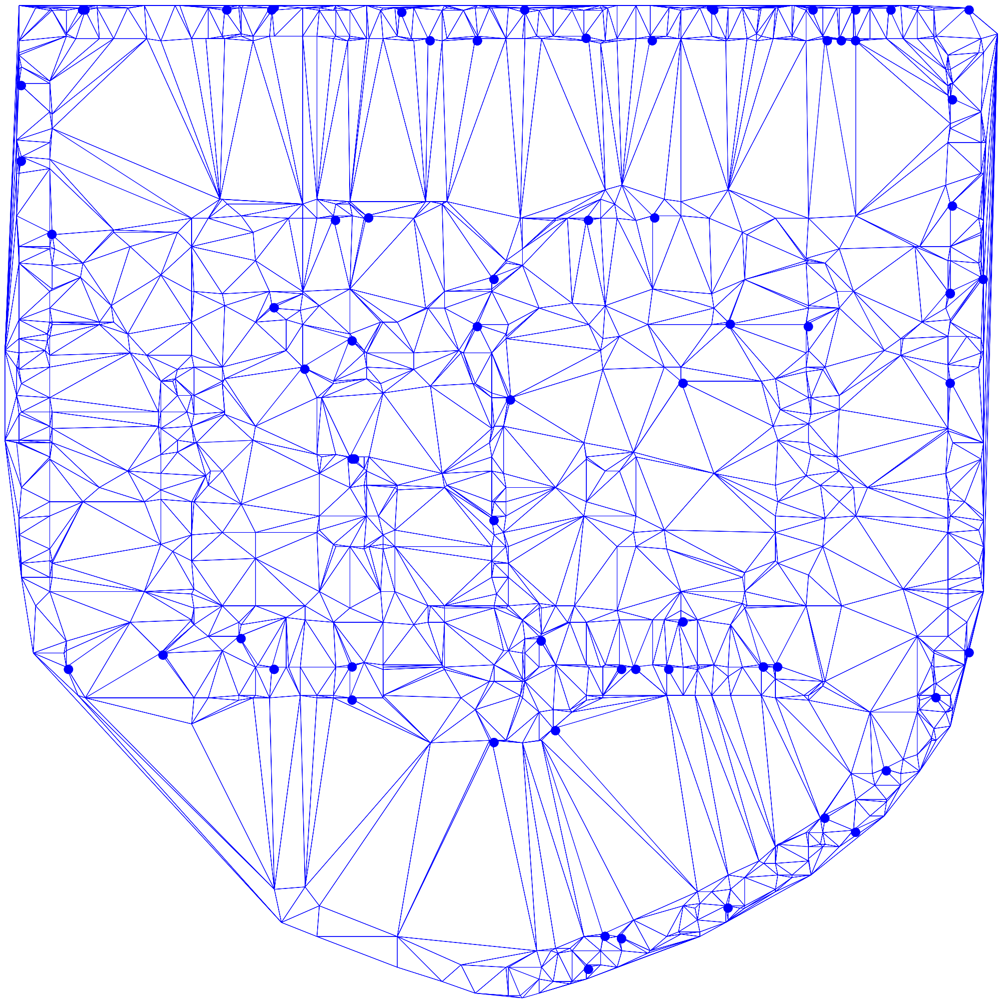
~~trivial solution:~~ $\cancel{\vec{x}(i) = (1,1)}$ So, require $\sum_i \vec{x}(i) = \mathbf{0}$

~~degenerate solution:~~ $\cancel{\vec{x}(i) = (v_2(i), v_2(i))}$

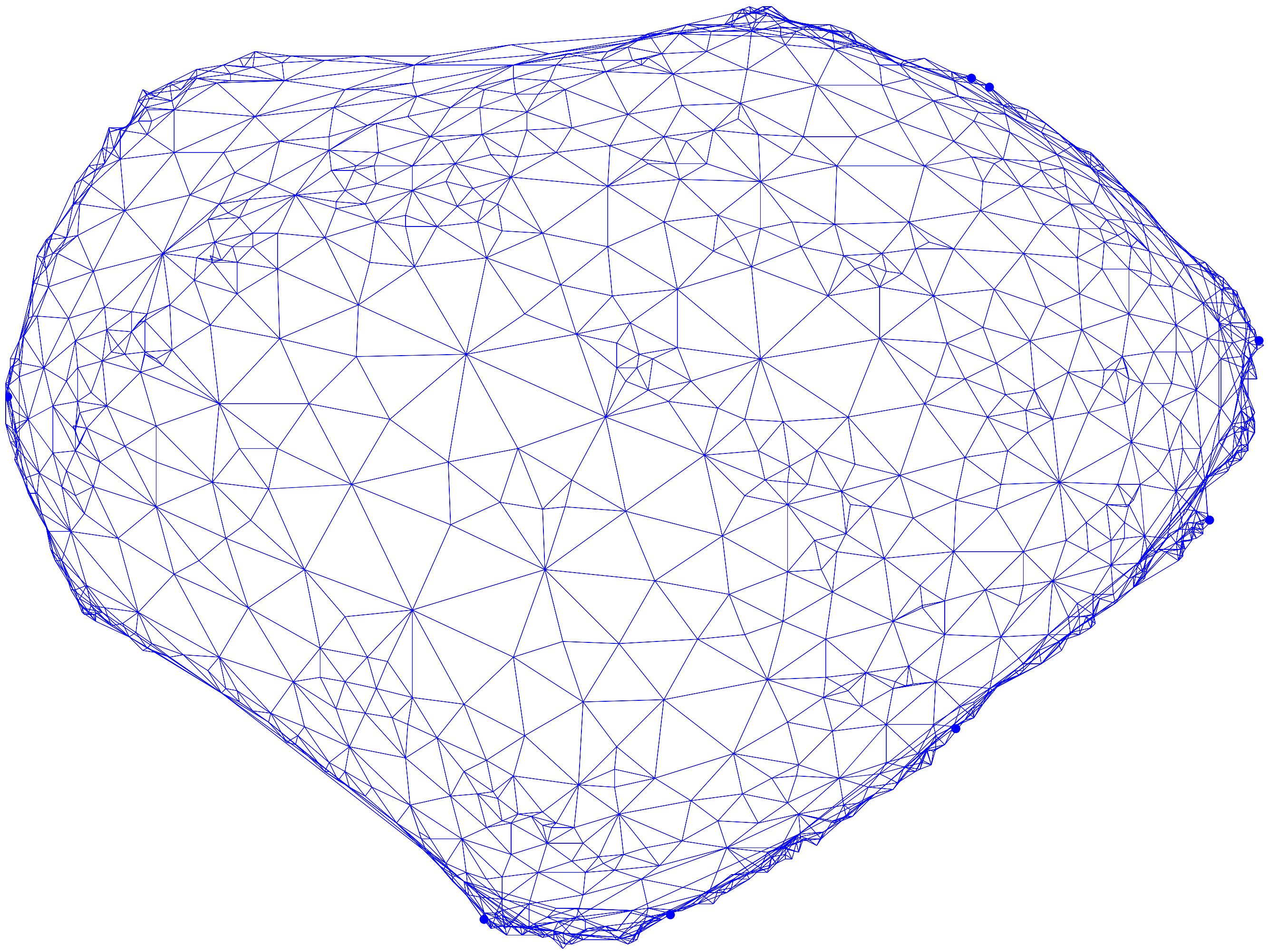
Also require $\sum_i \vec{x}_1(i)\vec{x}_2(i) = 0$

Solution $\vec{x}(i) = (v_2(i), v_3(i))$ up to rotation

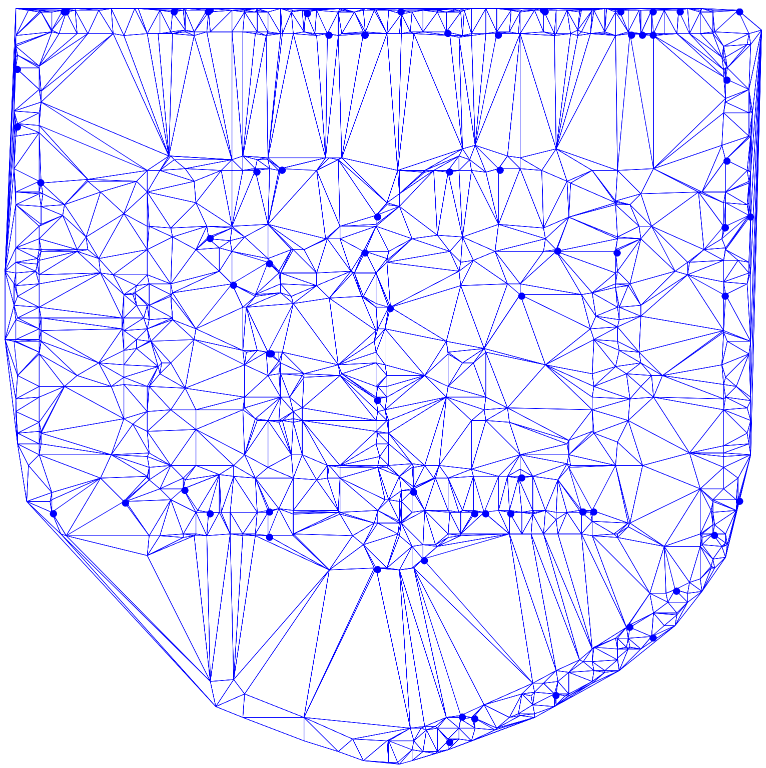
A Graph

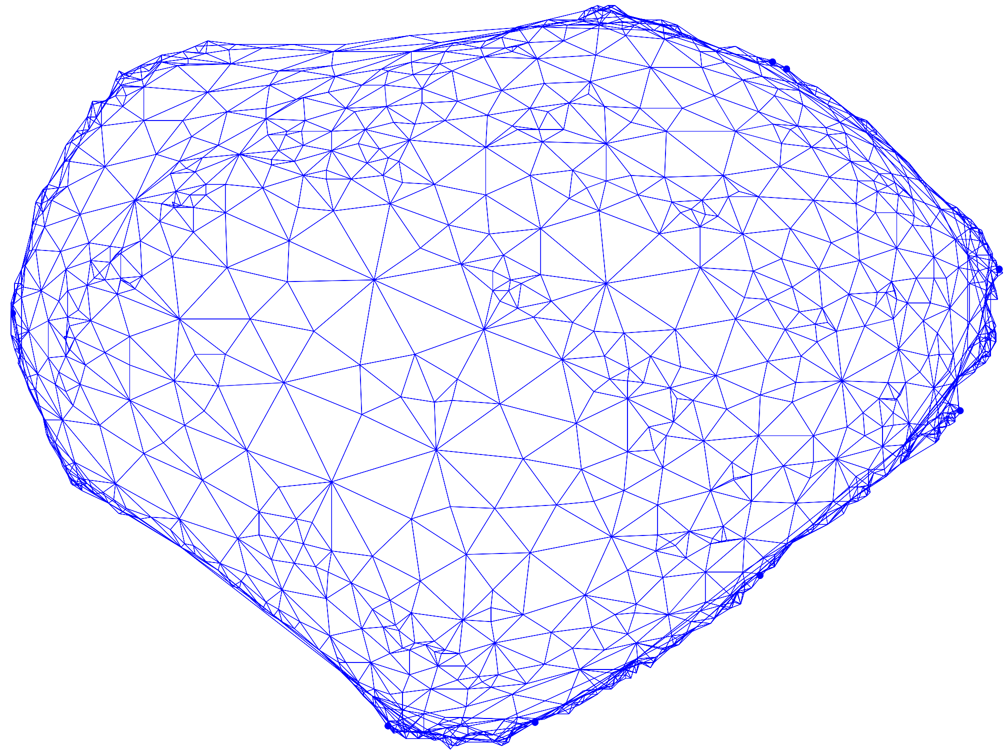


Drawing of the graph using v_2, v_3

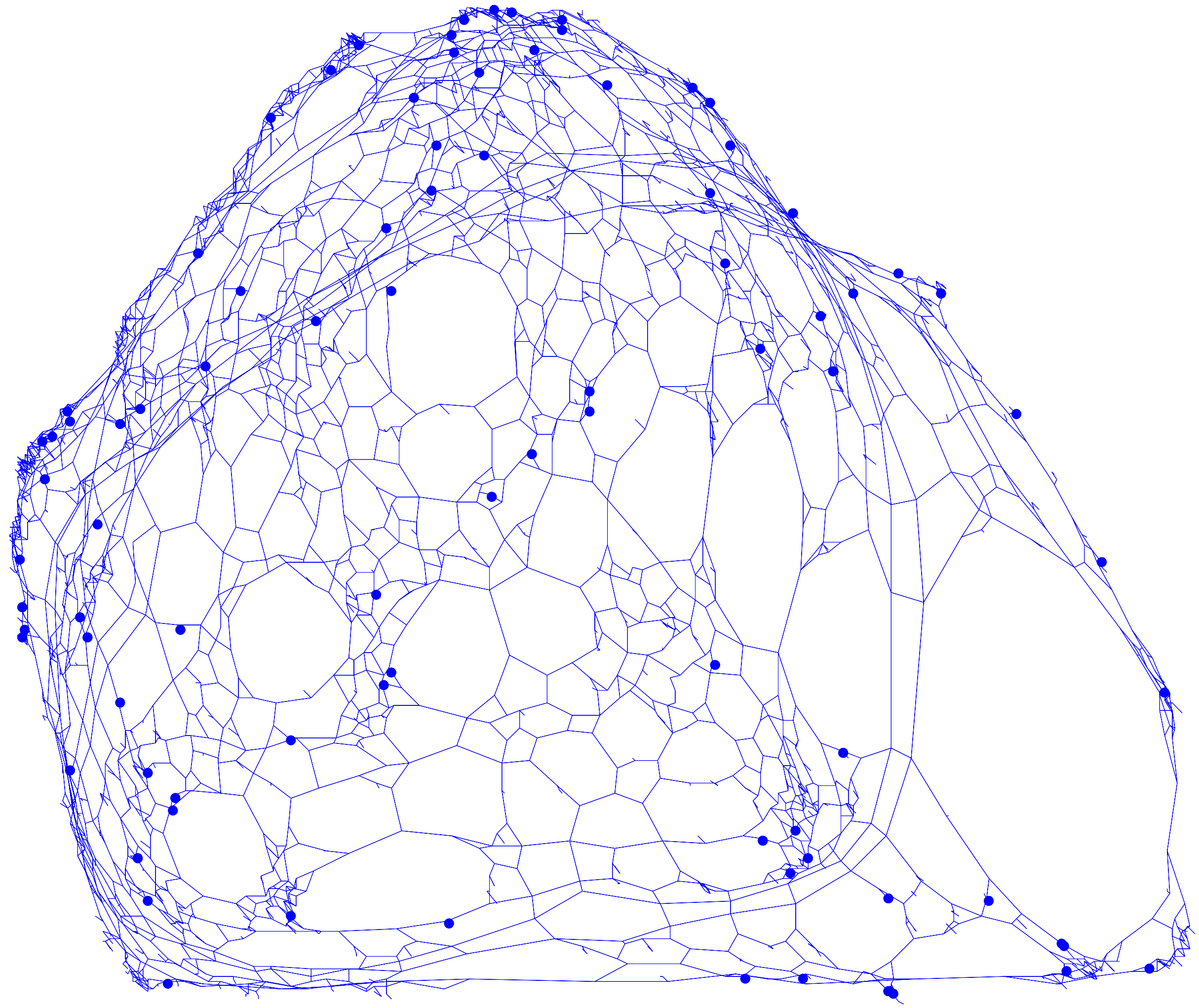


Plot vertex i at $(v_2(i), v_3(i))$

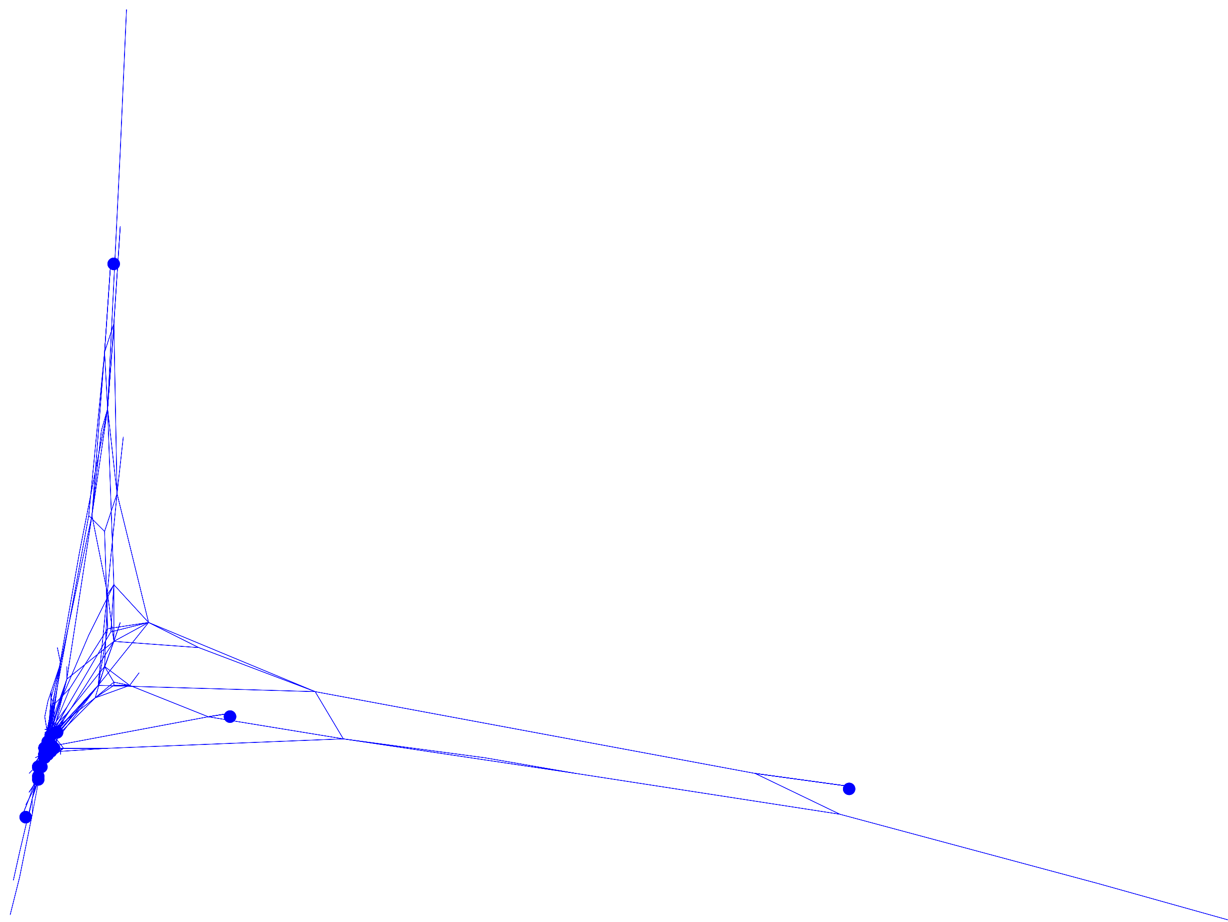




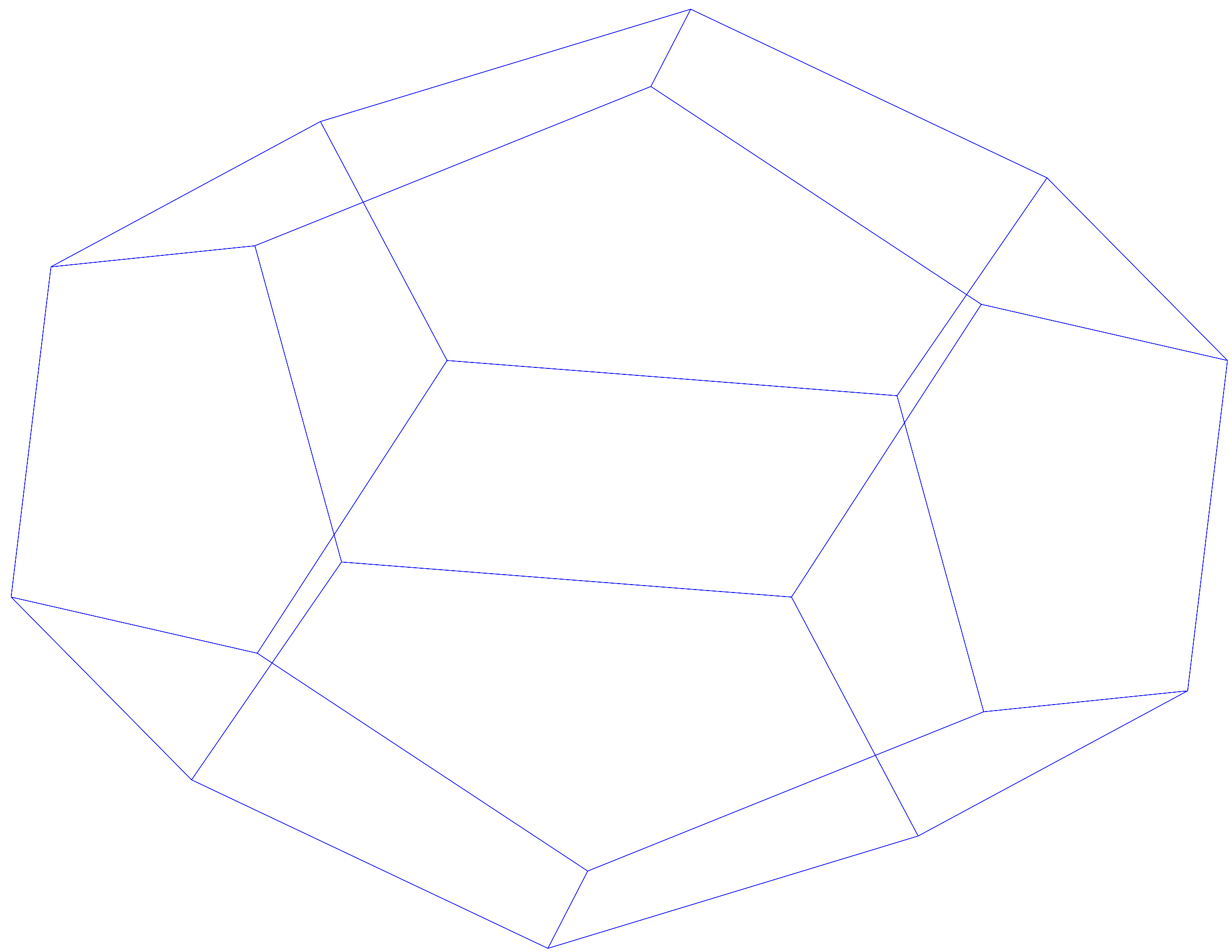
Spectral drawing of Streets in Rome



Spectral drawing of Erdos graph: edge between co-authors of papers



Dodecahedron



Best embedded by first three eigenvectors

Spectral graph drawing: Tutte justification

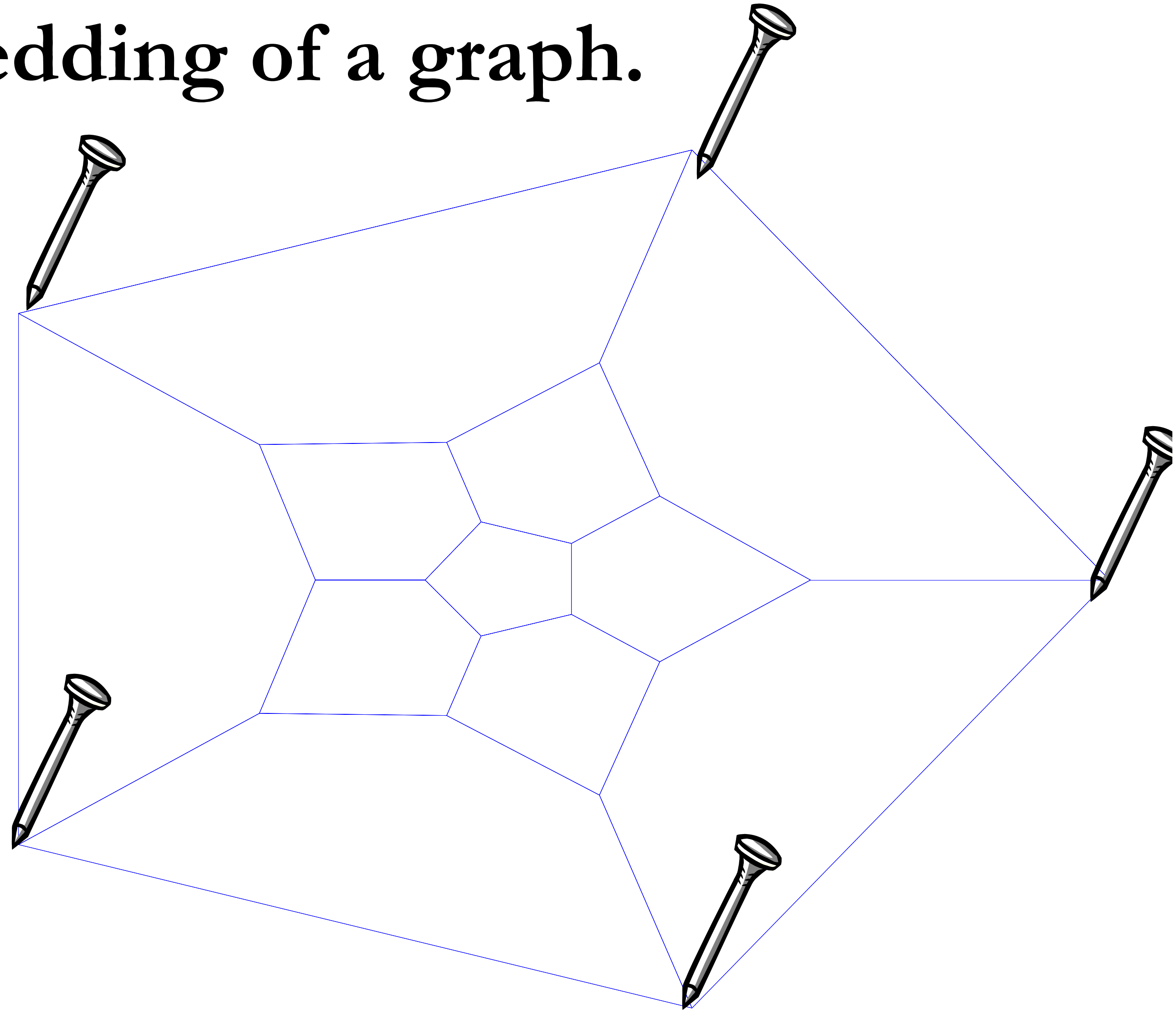
Condition for eigenvector $Lx = \lambda x$

Gives $x(i) = \dfrac{1}{d_i - \lambda} \sum_{j \sim i} x(j)$ for all i

λ small says x(i) near average of neighbors

Tutte '63: If fix outside face, and let every other vertex be average of neighbors, get planar embedding of planar graph.

Tutte '63 embedding of a graph.

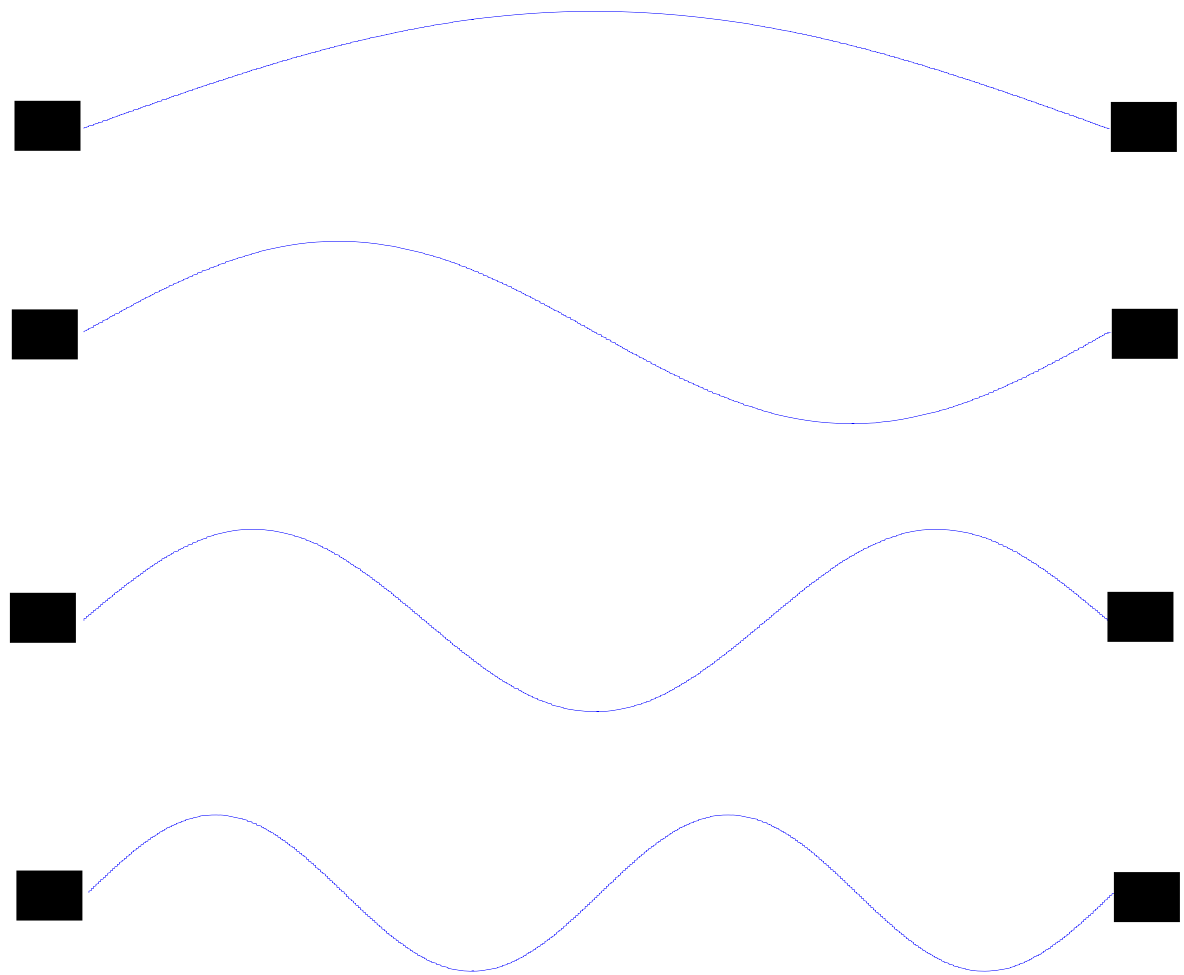


Fix outside face.
Edges -> springs.

Vertex at center
of mass of nbrs.

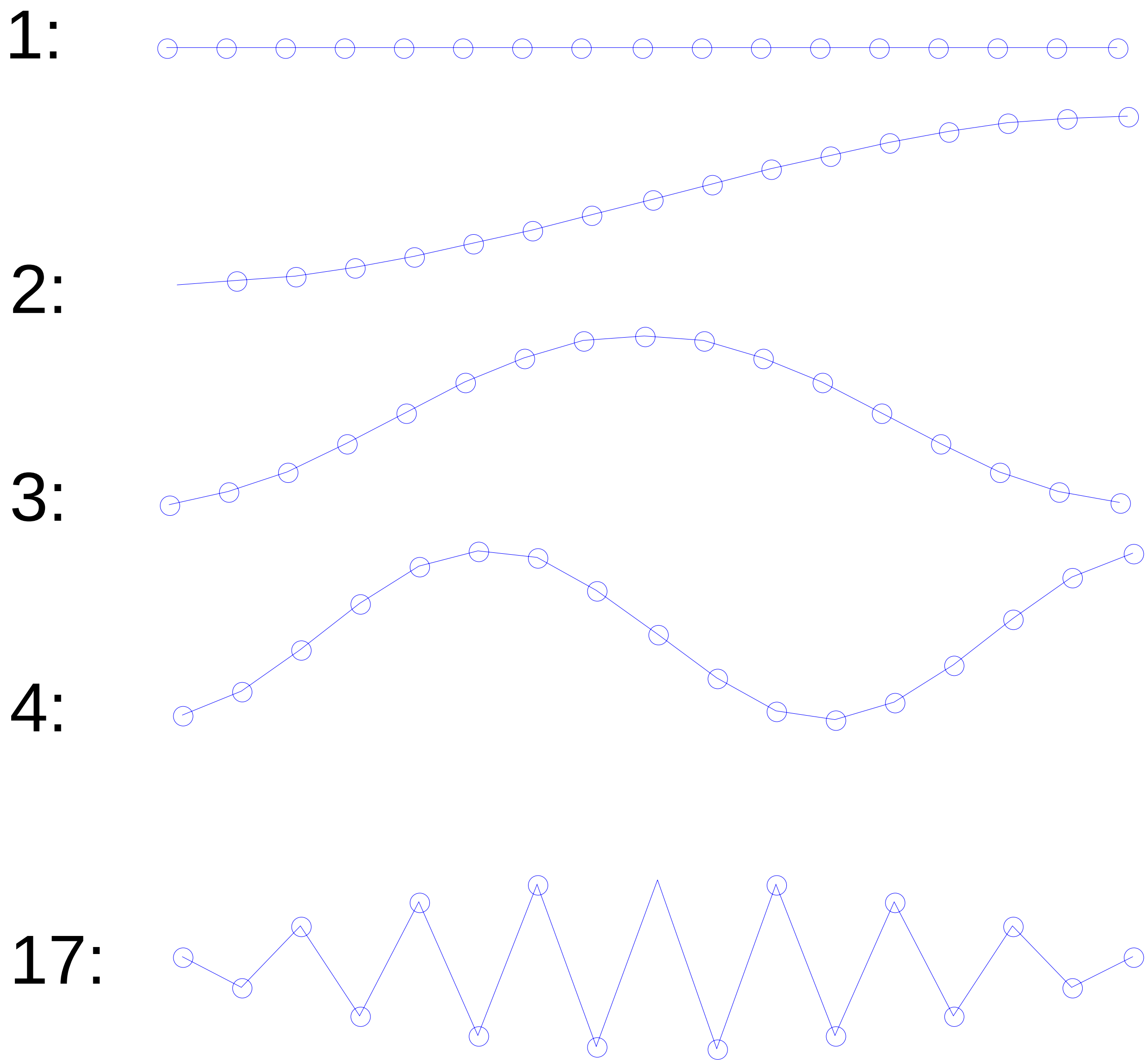
3-connected -> get planar embedding

Fundamental modes: string with fixed ends

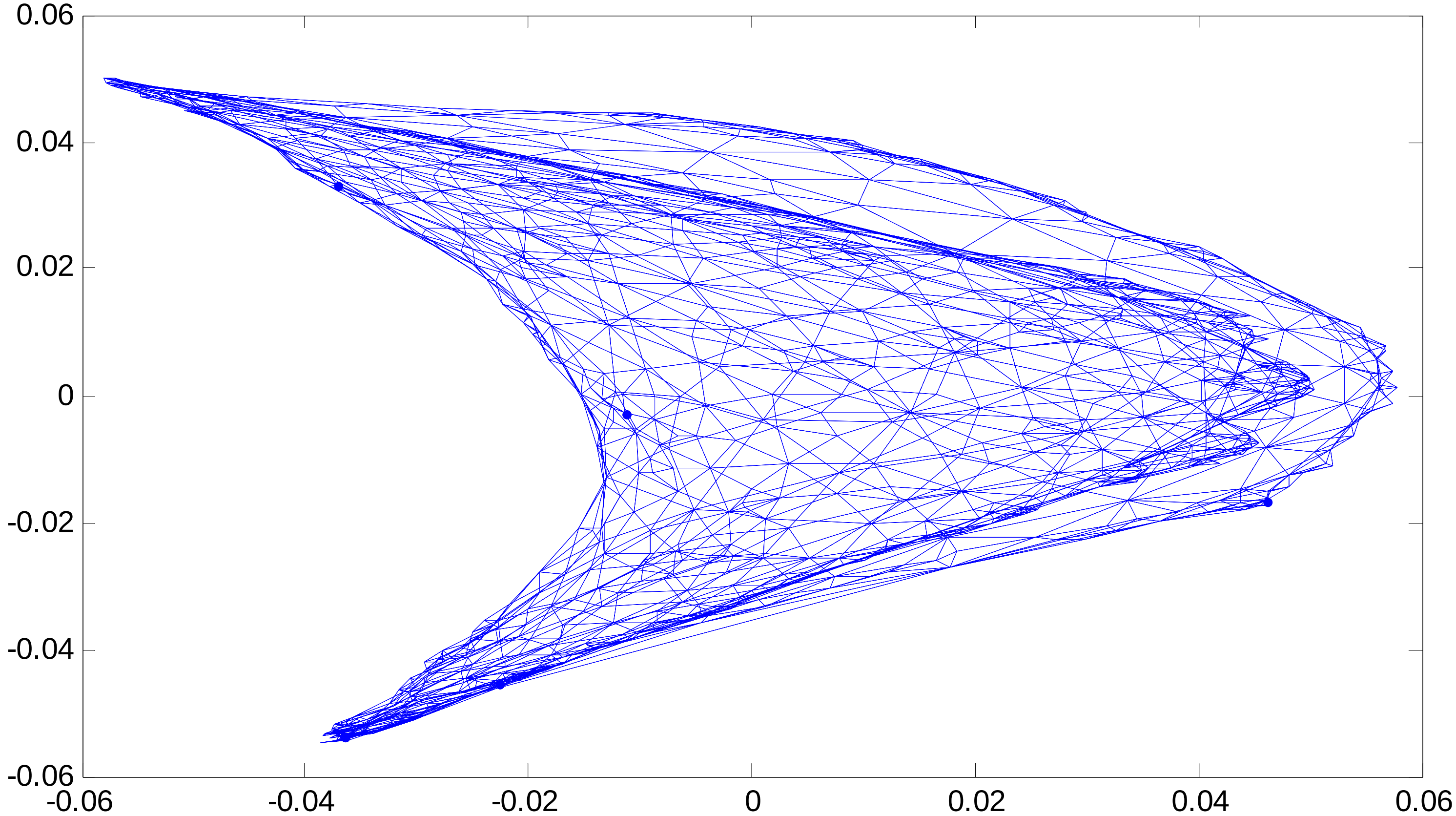


Fundamental modes: string with free ends

Eigenvectors of path graph

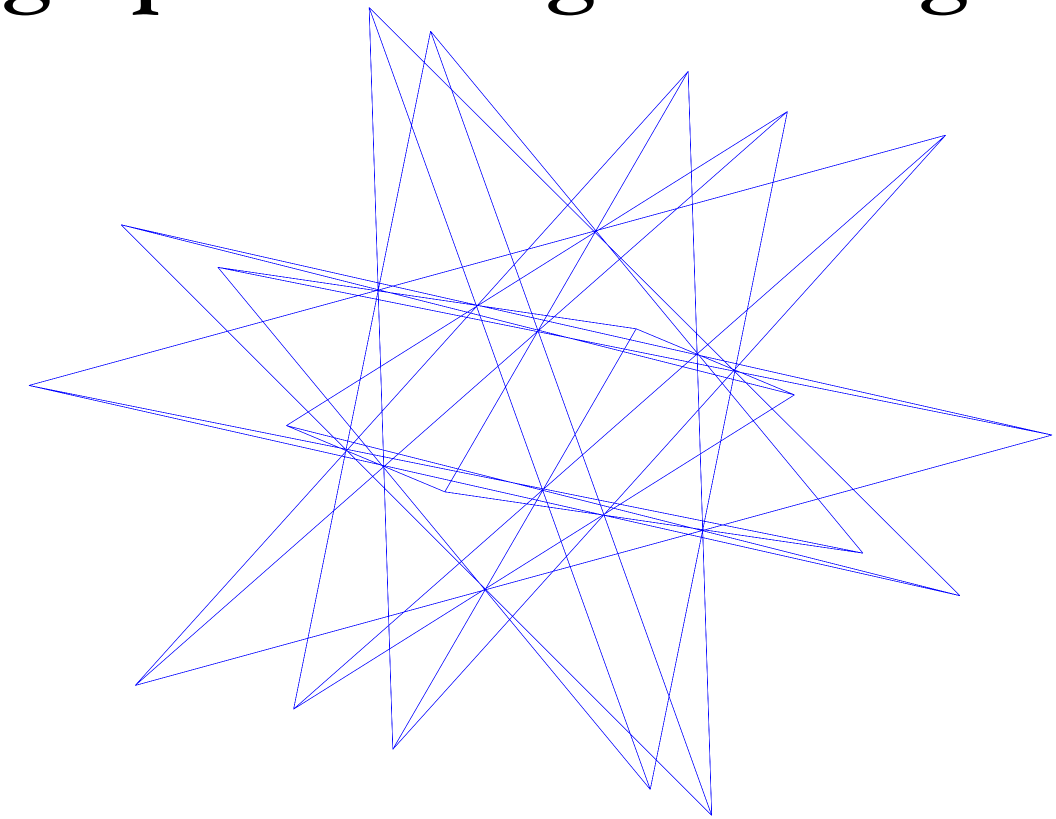


Drawing of the graph using v_3, v_4



Plot vertex i at $(v_4(i), v_3(i))$

Spectral graph coloring from high eigenvectors



Embedding of dodecahedron by 19th and 20th eigvecs.

Coloring 3-colorable random graphs [Alon-Kahale ’97]

Spectral graph drawing: FEM justification

If apply finite element method to solve

Laplace’s equation in the plane

with a Delaunay triangulation

Would get graph Laplacian,

but with some weights on edges

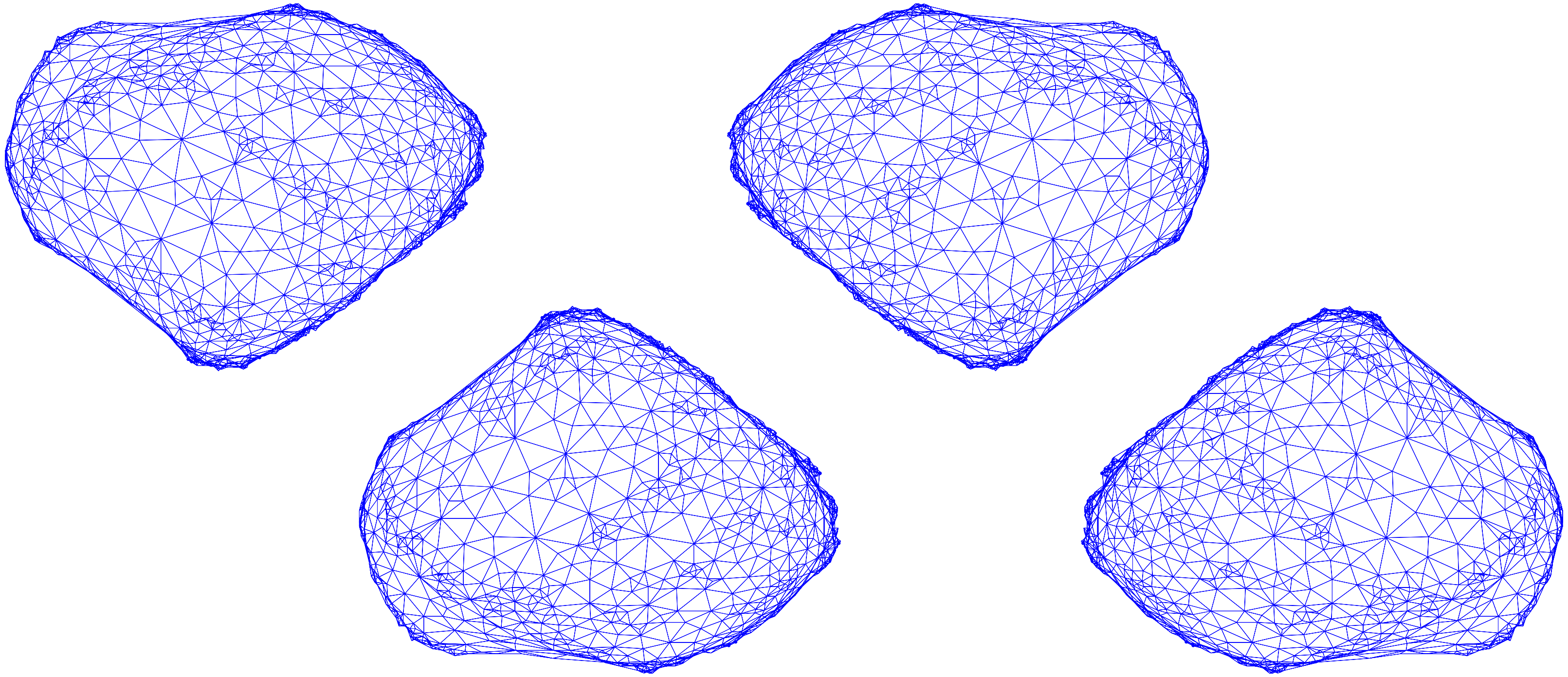
Fundamental solutions are x and y coordinates

(see Strang’s Introduction to Applied Mathematics)

Isomorphism testing

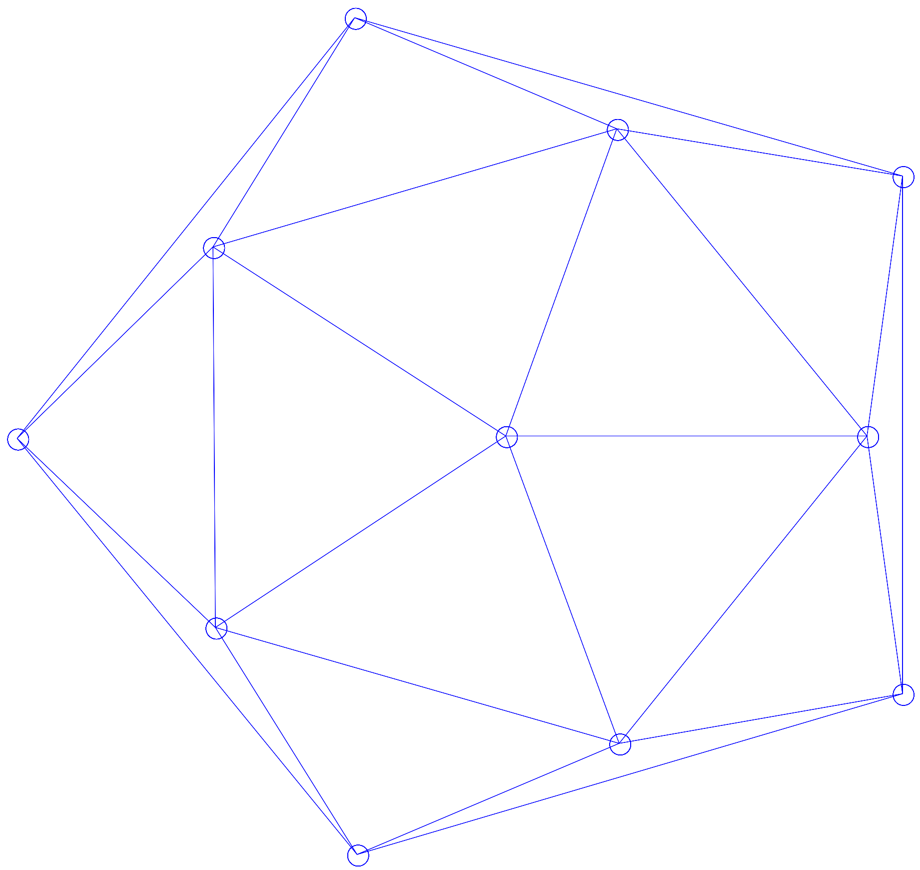
1. different eigenvalues -> non-isomorphic

2. If each vertex distinct in spectral embedding, just need to line up embeddings.



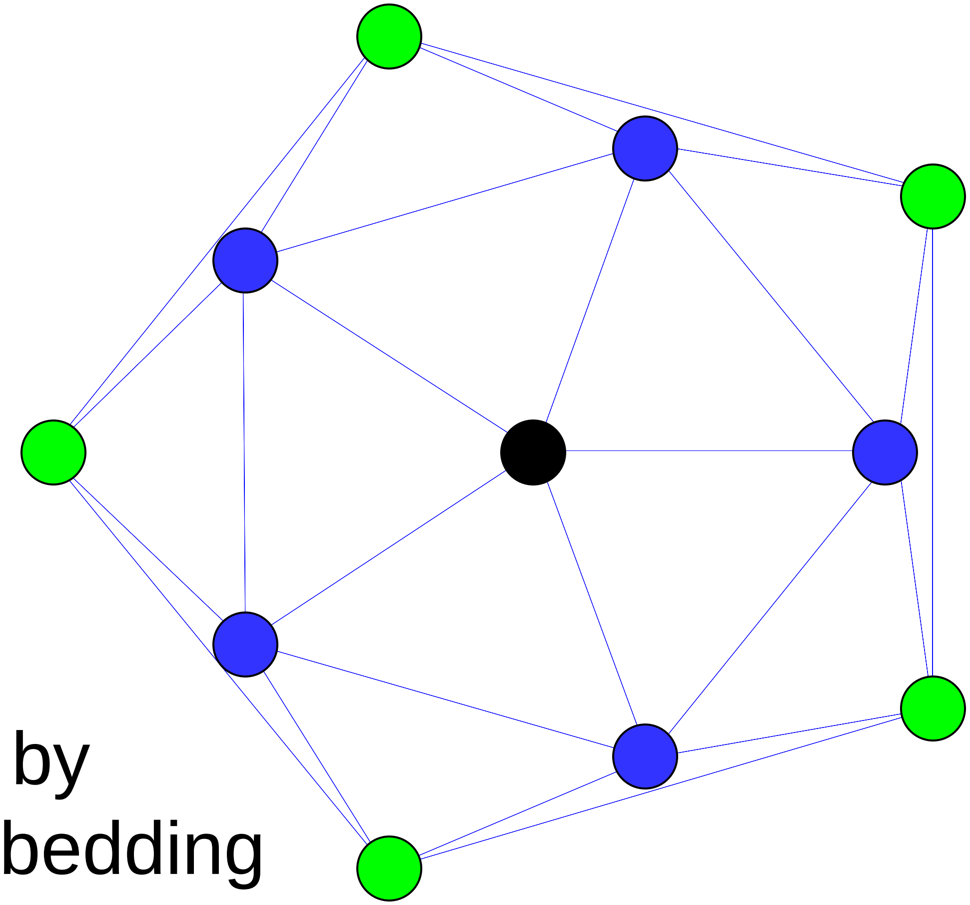
Each eigvec determined up to sign.

Isomorphism testing



$\lambda_2 = \lambda_3$, eigvecs determined up to rotation

Isomorphism testing



Distinguish by
norm in embedding

$\lambda_2 = \lambda_3$, eigvecs determined up to rotation

Isomorphism testing: difficulties

1. Many vertices can map to same place in spectral embedding, if only use few eigenvectors.

2. If λ_i has a high dimensional space, eigvecs only determined up to basis rotations

Ex.: Strongly regular graphs with only 3 eigenvalues, of multiplicities 1, (n-1)/2 and (n-1)/2

3. Some pairs have an exponential number of isomorphisms.

Isomorphism testing: success [Babai-Grigoryev-Mount '82]

If each eigenvalue has multiplicity O(1), can test in polynomial time.

Ideas:

Partition vertices into classes by norms in embeddings.
Refine partitions using other partitions.
Use vertex classes to split eigenspaces.

Use computational group theory to fuse information, and produce description of all isomorphisms.

Random Walks

Random walks and PageRank

Adjacency matrix of directed graph:

$$A(i,j) = \begin{cases} 1 & \text{if } (j,i) \in E \\ 0 & \text{otherwise} \end{cases}$$

Walk transition matrix: $W = AD^{-1}$

Walk distribution at time t: $p_t = W^t p_0$

PageRank vector p: $p = Wp$

Eigenvector of Eigenvalue 1

Random walks and PageRank

PageRank vector p: $p = Wp$

Linear algebra issues:

W is not symmetric, not similar to symmetric, does not necessarily have n eigenvalues

If no nodes of out-degree 0,
Perron-Frobenius Theorem:

Guarantees a unique, positive eigevec p of eigenvalue 1.

Is there a theoretically interesting spectral theory?

Kleinberg and the singular vectors

Consider eigenvectors of largest eigenvalue of

$$AA^T \quad \text{and} \quad A^TA$$

Are left and right singular values of A.

Always exist.

Usually, a more useful theory than eigenvectors, when not symmetric.

(see Strang's Intro. to Linear Algebra)

Random walks on Undirected Graphs

$$W = AD^{-1}$$

Trivial PageRank Vector: $\boldsymbol{d} = W\boldsymbol{d}$ $\qquad$ $\mathbf{1} = \mathbf{1}W$

Not symmetric, but similiar to
symmetrized walk matrix

$$D^{-1/2} W D^{1/2} = D^{-1/2} A D^{-1/2} = S$$

W and S have same eigvals,

$$Sv = \lambda v \quad \Longrightarrow \quad W\left(D^{1/2}v\right) = \lambda\left(D^{1/2}v\right)$$

Random walk converges at rate $1/1\text{-}\lambda_{n\text{-}1}$

For lazy random walk (stay put with prob ½):

$$|p_t(v) - \pi(v)| \leq \sqrt{\frac{d(v)}{\min_u d(u)}}(1 - \lambda_{n-1})^t$$

Where π is the stable distribution

For symmetric S

$$p_0 = \sum_i \left(v_i^T p_0\right) v_i$$

$$p_1 = S p_0 = \sum_i \lambda_i \left(v_i^T p_0\right) v_i$$

$$p_t = S^t p_0 = \sum_i \lambda_i^t \left(v_i^T p_0\right) v_i$$

Normalized Laplacian [Chung]

If consider $1-\lambda_{n-1}$ should look at

$$\mathcal{L} = I - S = D^{-1/2} L D^{-1/2}$$

$$\lambda_2(\mathcal{L}) = 1 - \lambda_{n-1}(W)$$

Relationship to cuts:

$$\lambda_2(\mathcal{L}) = \min_{v \perp d} \frac{v^T L v}{v^T D v} = \min_{v \perp d} \frac{\sum_{(i,j)\in E} (v(i) - v(j))^2}{\sum_i d(i)(v(i))^2}$$

Cheeger's Inequality (Jerrum-Sinclair '89) (Alon-Milman '85, Diaconis-Stroock '91)

$$\lambda_2(\mathcal{L}) = \min_{v \perp d} \frac{v^T L v}{v^T D v} = \min_{v \perp d} \frac{\sum_{(i,j)\in E} (v(i) - v(j))^2}{\sum_i d(i)(v(i))^2}$$

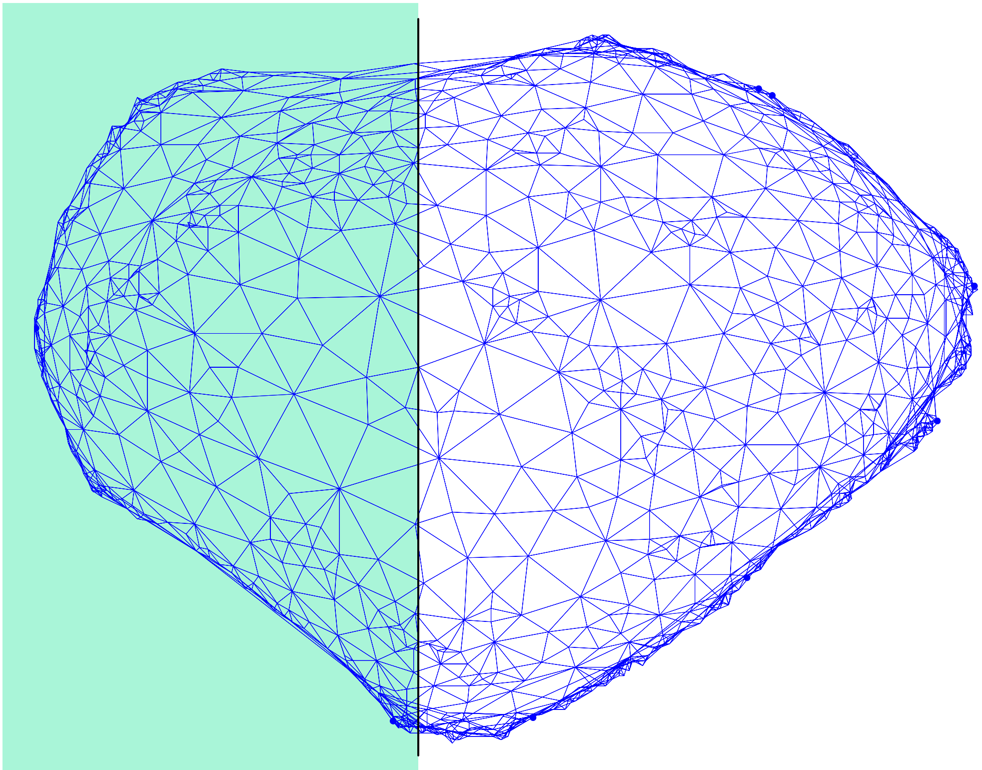
$$\phi(S) \stackrel{\text{def}}{=} \frac{w(\partial(S))}{\min\left(d(S), d(V-S)\right)}$$

$$\lambda_2(\mathcal{L})/2 \leq \min_S \phi(S) \leq \sqrt{2\lambda_2(\mathcal{L})}$$

Cheeger's Inequality (Jerrum-Sinclair '89) (Alon-Milman '85, Diaconis-Stroock '91)

Can find the cut by looking at $w = D^{-1/2} v$

$S = \{i : w(i) < t\}$ for some t



Only need approximate eigenvector (Mihail '89)

Can find the cut by looking at $w = D^{-1/2}v$

$S = \{i : w(i) < t\}$ for some t

Guarantee

$$\phi(S) \leq \sqrt{2\frac{v^T \mathcal{L} v}{v^T v}}$$

Lanczos era.

Normalized Cut

Alternative definition of conductance [Lovasz '96 (?)]

$$\phi(S) \stackrel{\text{def}}{=} \frac{d(V)w(\partial(S))}{d(S)d(V-S)}$$

This way, λ_2 is a relaxation [see Hagen-Kahng '92].

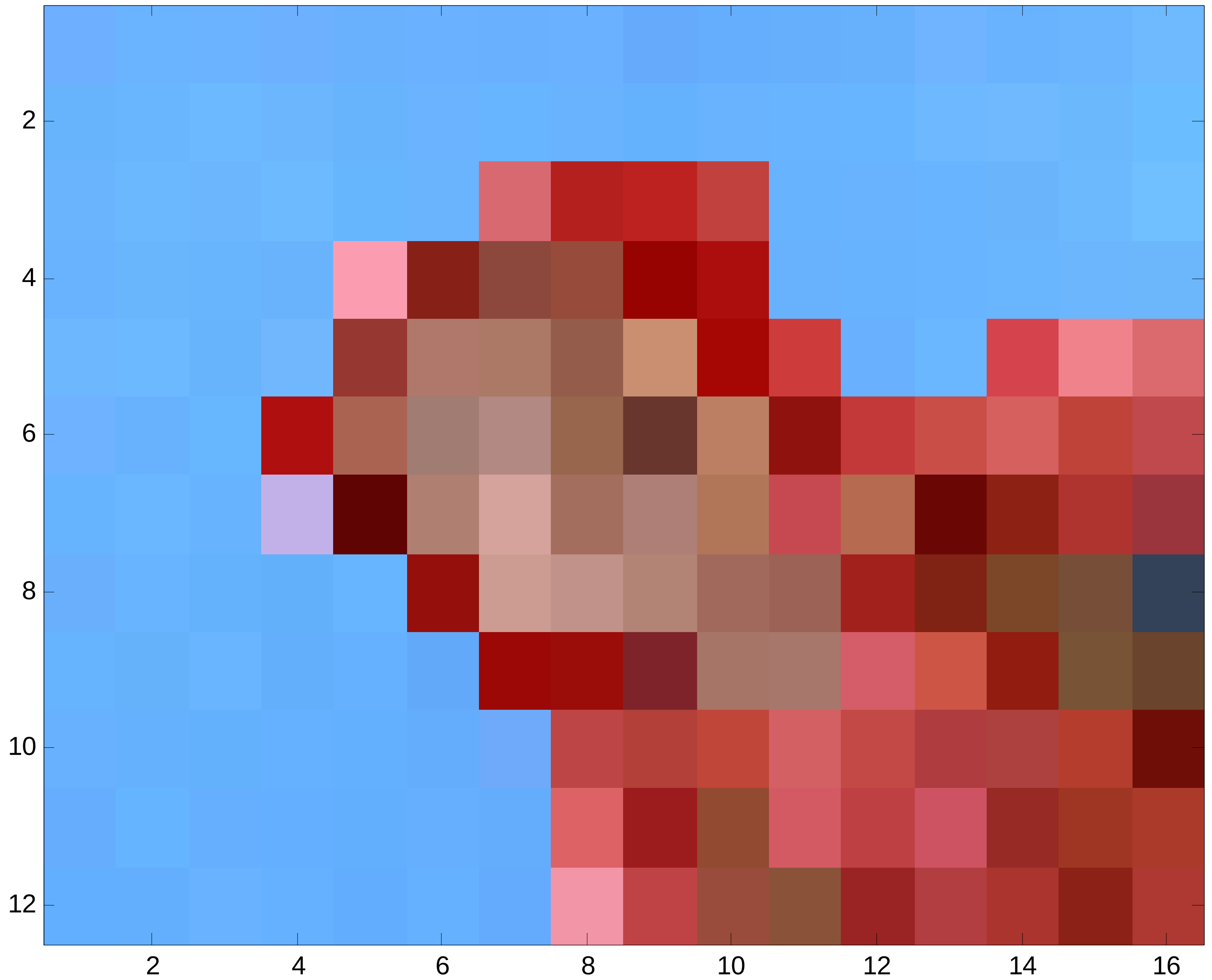
$$\lambda_2(\mathcal{L}) = \min_{v \perp d} \frac{v^T L v}{v^T D v} = \min_{v \perp d} \frac{\sum_{(i,j)\in E}(v(i)-v(j))^2}{\sum_i d(i)(v(i))^2}$$

Equivalent to Normalized Cut [Shi-Malik '00]

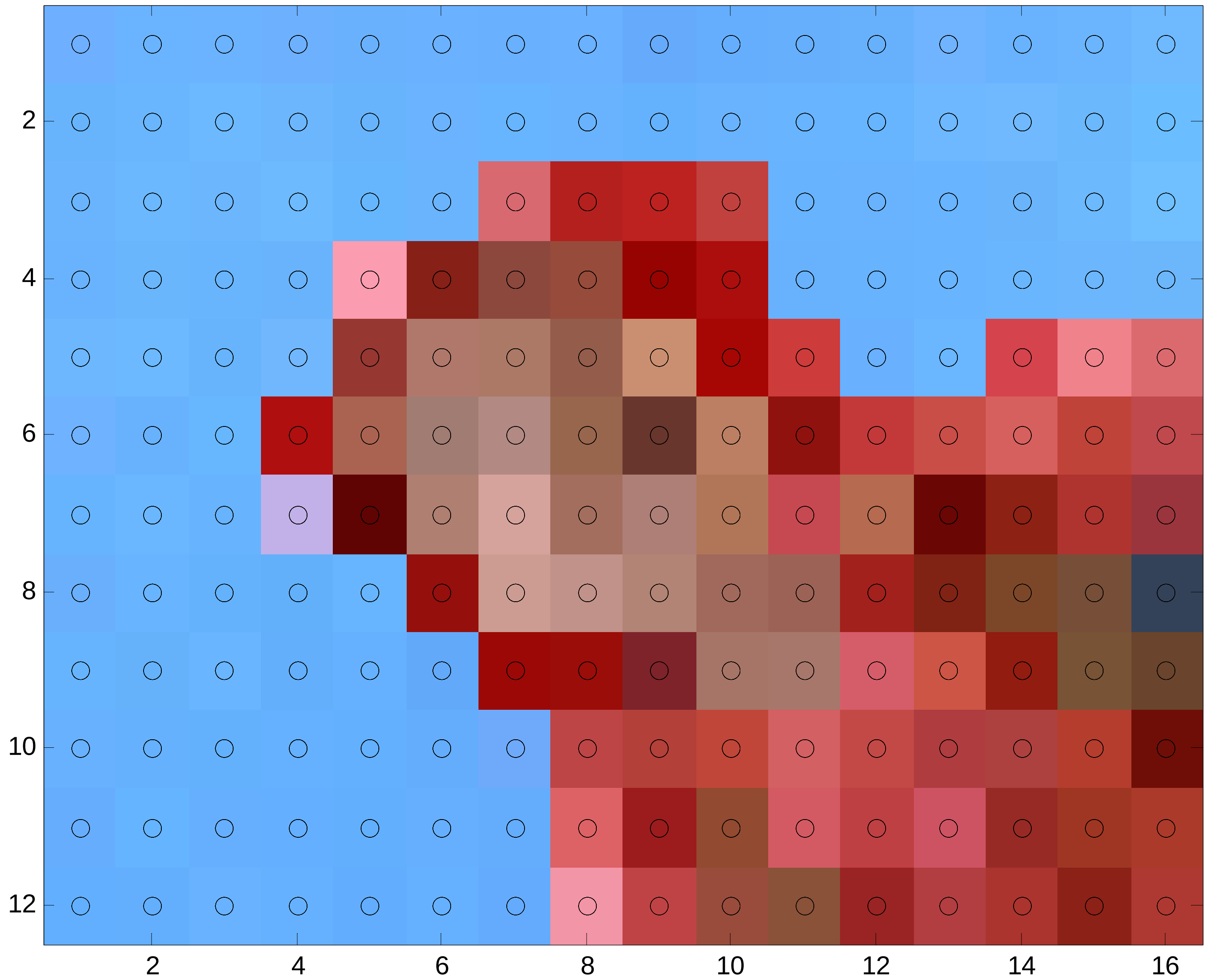
$$\frac{w(\partial(S))}{d(S)} + \frac{w(\partial(S))}{d(V-S)}$$

Spectral Image Segmentation (Shi-Malik ‘00)

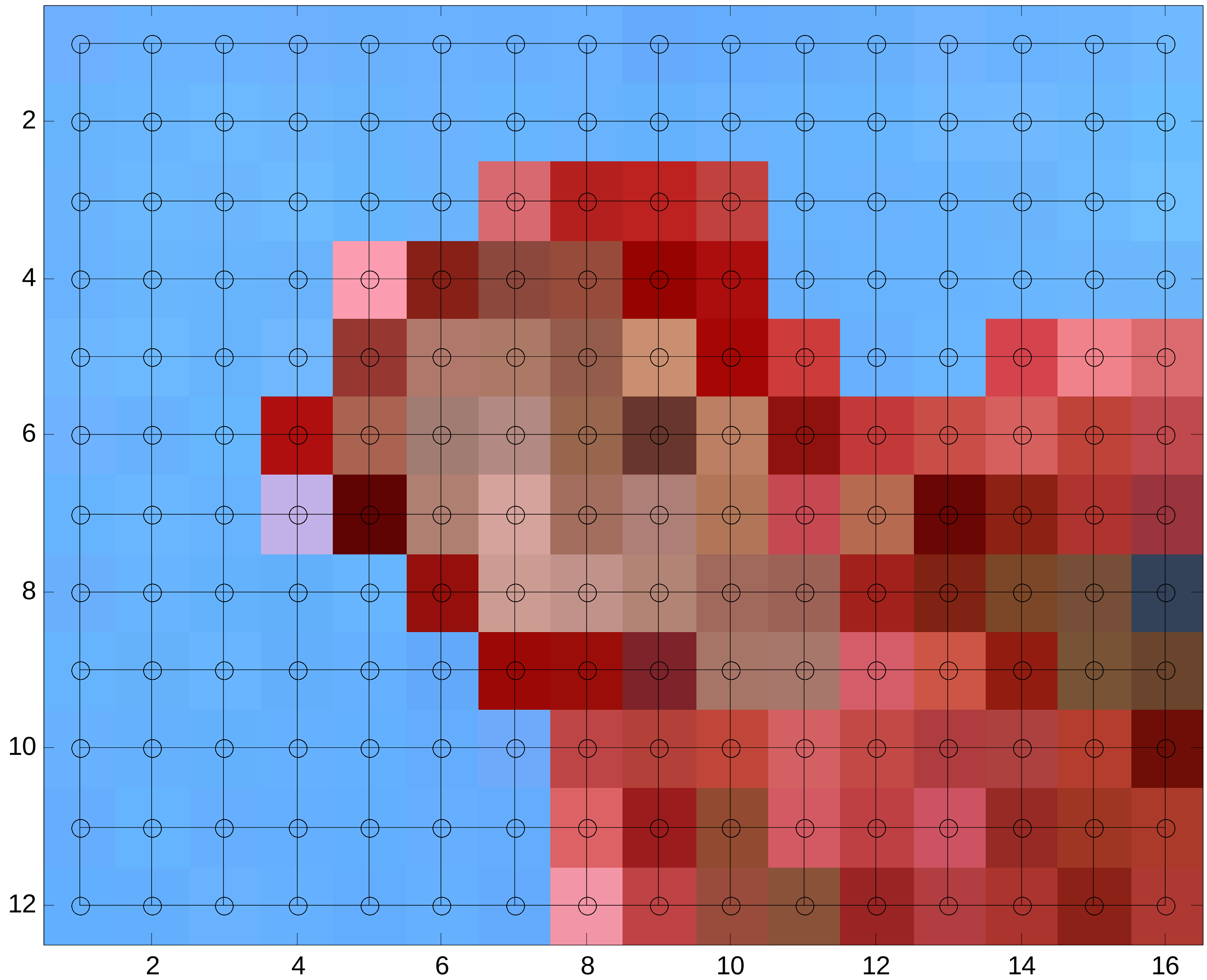
Spectral Image Segmentation (Shi-Malik '00)



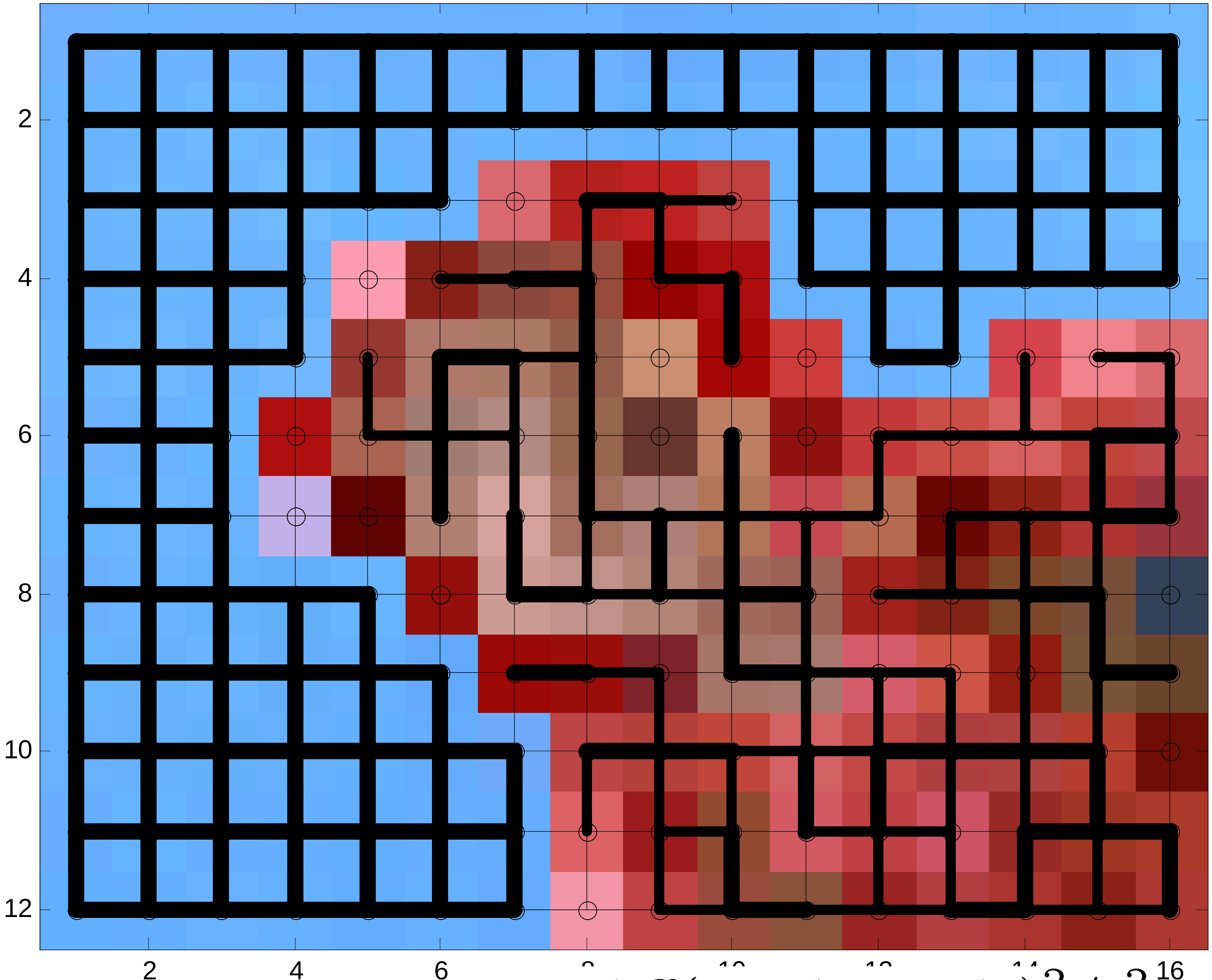
Spectral Image Segmentation (Shi-Malik '00)



Spectral Image Segmentation (Shi-Malik ‘00)

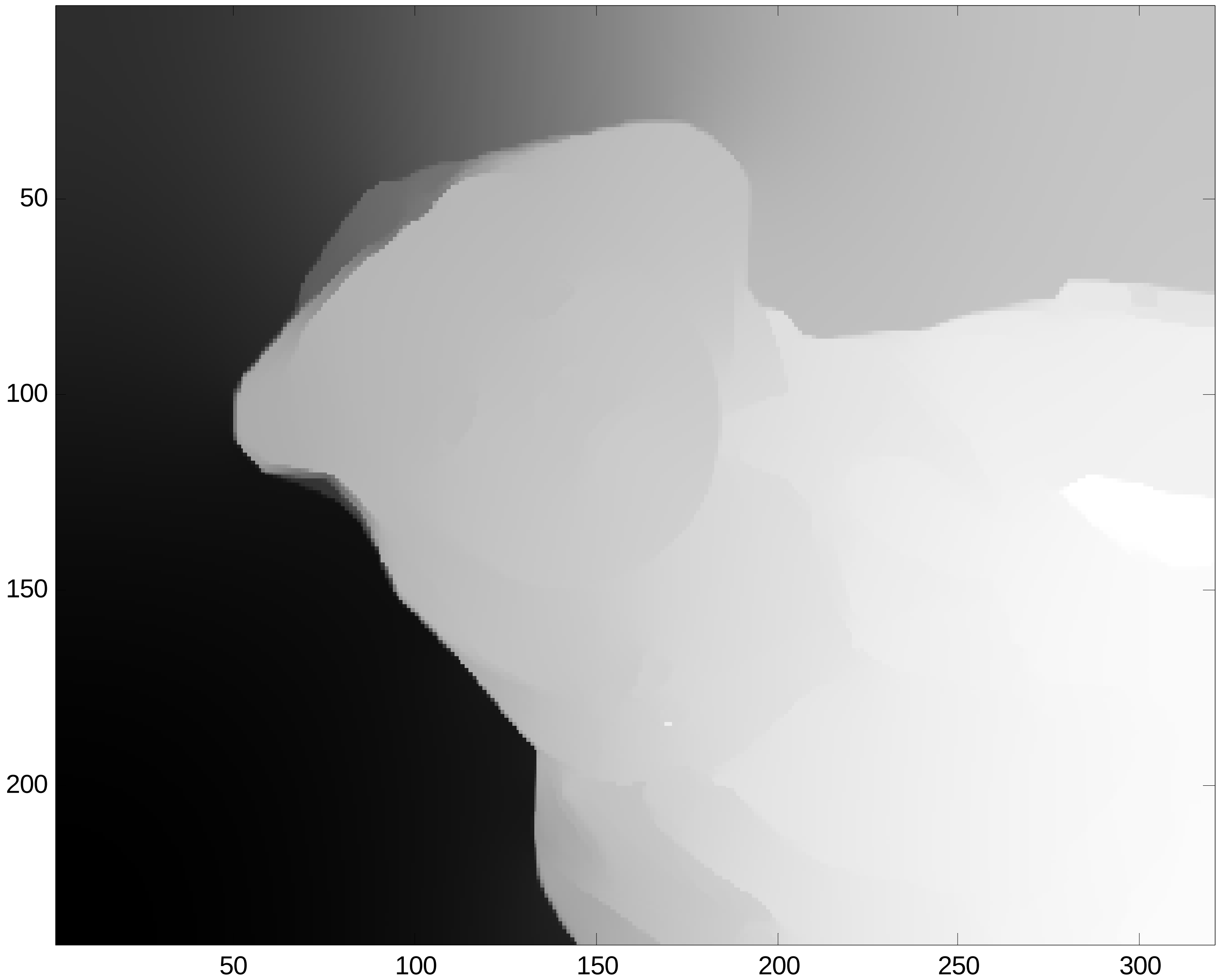


Spectral Image Segmentation (Shi-Malik '00)

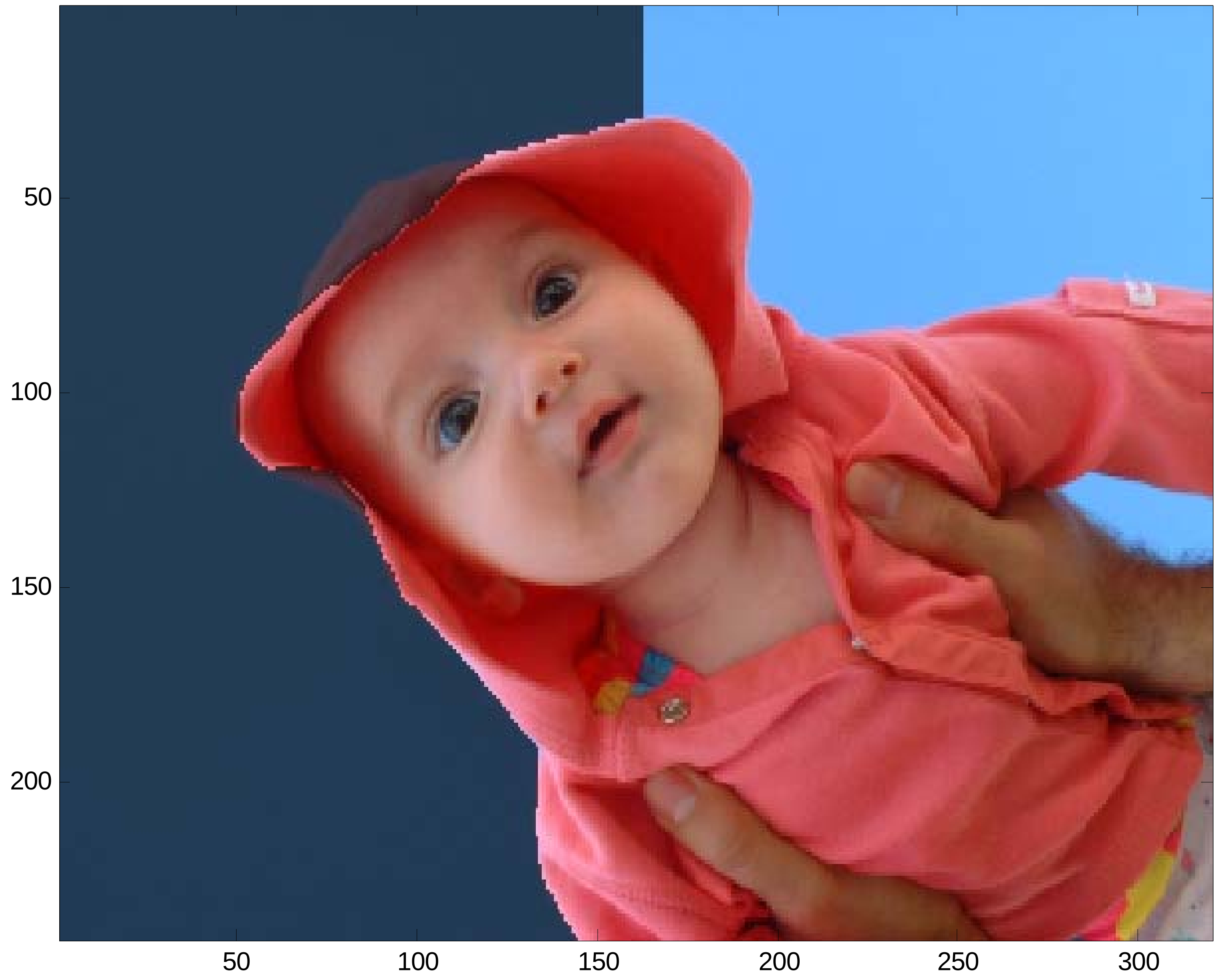


edge weight $e^{-\mathrm{diff}(\mathrm{pixel}\,i,\,\mathrm{pixel}\,j)^2/t^2}$

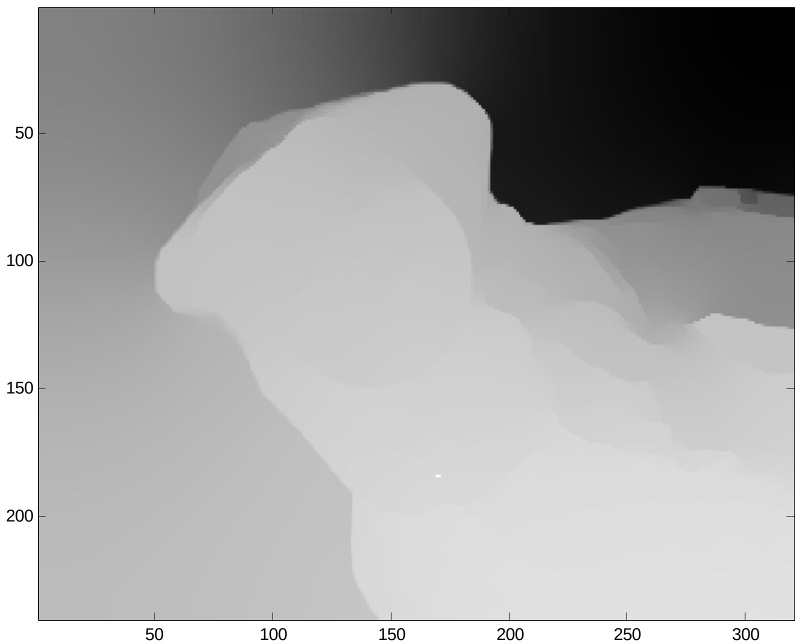
The second eigenvector

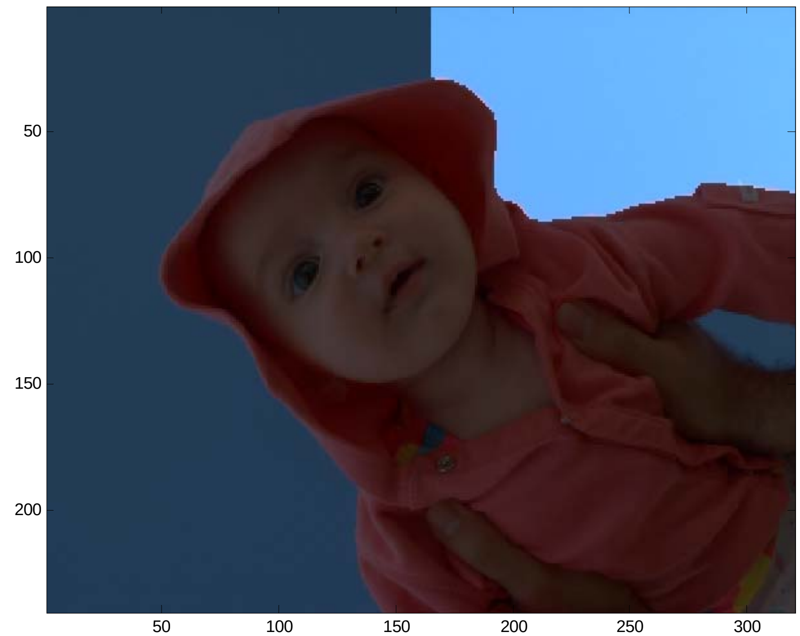


Second Eigenvector's sparsest cut

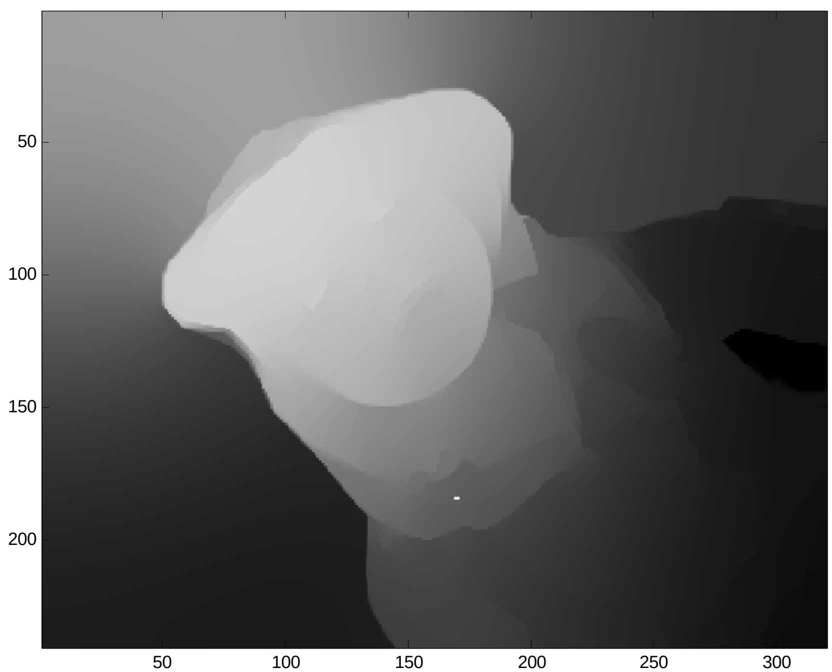


Third Eigenvector





Fourth Eigenvector



Perspective on Spectral Image Segmentation

Ignoring a lot we know about images.

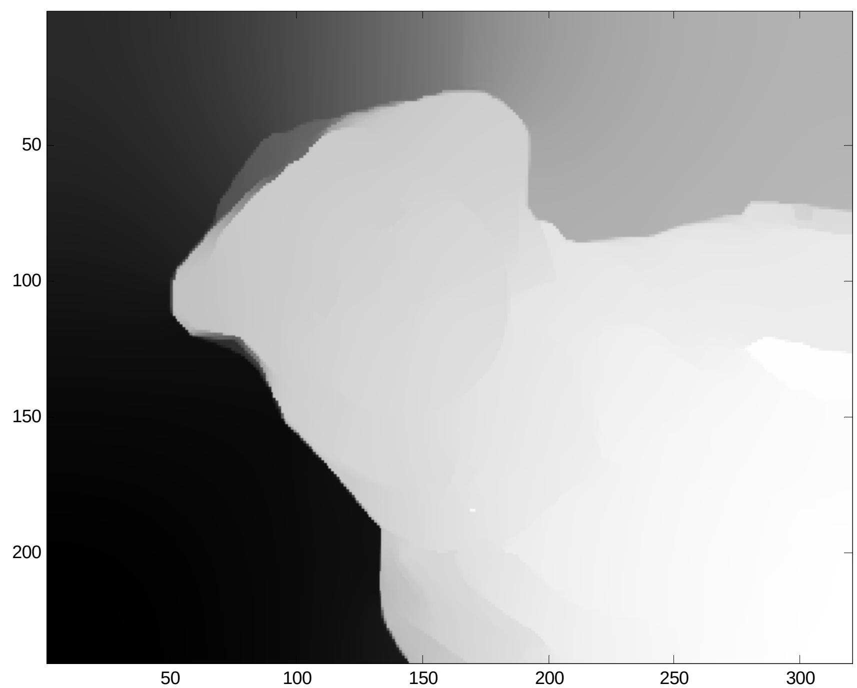
On non-image data, gives good intuition.

Can we fuse with what we know about images?

Generally, can we fuse with other knowledge?

What about better cut algorithms?

Improvement by Miller and Tolliver '06



Improvement by Miller and Tolliver '06

$$\frac{x^T L x}{x^T D x} = \frac{\sum_{(i,j)\in E} (v(i) - v(j))^2 w_{i,j}}{\sum_{(i,j)\in E} (v(i)^2 + v(j)^2) w_{i,j}}$$

Idea: re-weight (i,j) by $\frac{(v(i)^2 + v(j)^2)}{(v(i) - v(j))^2}$

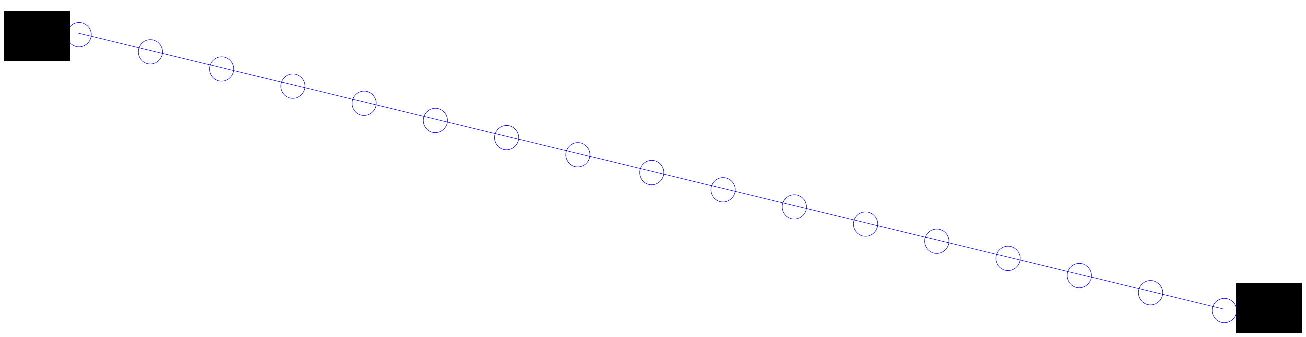
Actually, re-weight by $\Psi\left(\frac{(v(i)^2 + v(j)^2)}{(v(i) - v(j))^2}\right)$

Prove: as iterate $\lambda_2 \rightarrow 0$, get 2 components

One approach to fusing:
Dirichlet Eigenvalues

Fixing boundary values to zero [Chung-Langlands '96]

Fixing boundary values to non-zero. [Grady '06]
Dominant mode by solving linear equation:
computing electrical flow in resistor network



Analysis of Spectral Partitioning

Finite Element Meshes (eigvals right) [S-Teng ’07]

Planted partitions (eigvecs right) [McSherry ‘01]

$$\begin{array}{cc} \left(\begin{array}{c|c} p & q \\ \hline q & p \end{array}\right) & \begin{array}{l} \} A \\ \} B \end{array} \\ \begin{array}{cc} \underbrace{\qquad}_{A} & \underbrace{\quad}_{B} \end{array} & \end{array}$$

Prob

A-A edge = p

B-B edge = p

A-B edge = q

$q < p$

Other planted problems

Finding $cn^{1/2}$ clique in random graph
[Alon-Krivelevich-Sudakov ’98]

Color random sparse k-colorable graph
[Alon-Kahale ’97]

Asymmetric block structure (LSI and HITS)
[Azar-Fiat-Karlin-McSherry-Saia ’01]

Partitioning with widely varying degrees
[Dasgupta-Hopcroft-McSherry ’04]

Planted problem analysis

Sampled A as perturbation of $\left(\begin{array}{c|c} p & q \\ \hline q & p \end{array}\right)$

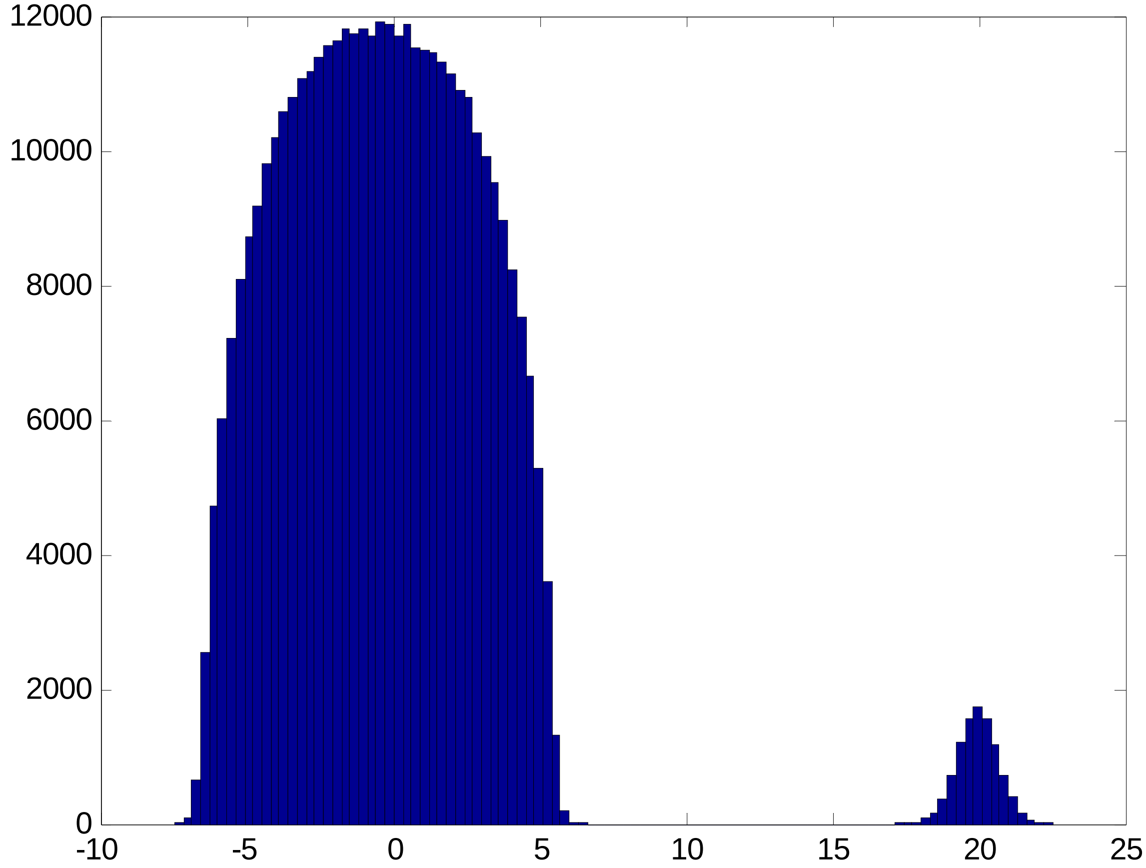
Small perturbations don't change eigenvalues too much.

Eigenvectors stable too, if well-separated from others.

Understand eigenvalues of random matrices
[Furedi-Komlos '81, Alon-Krivelevich-Vu '01, Vu '05]

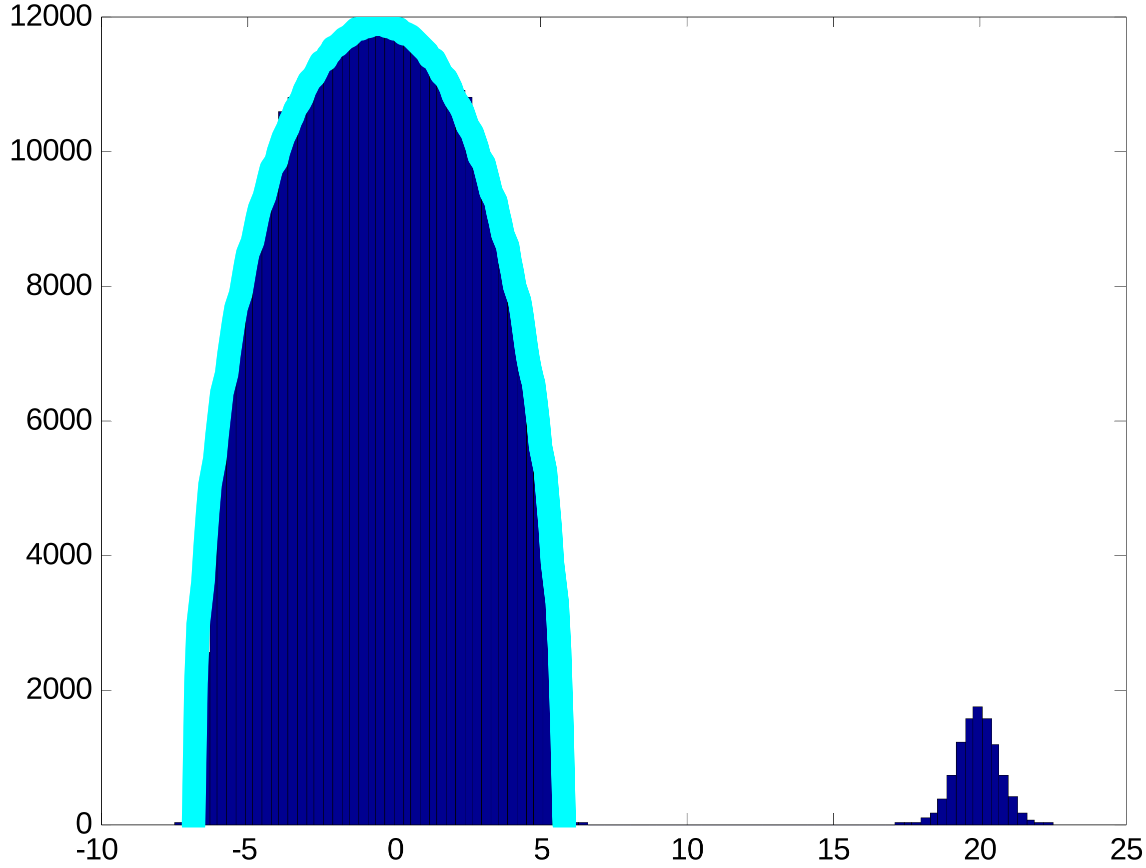
Distribution of eigenvalues of Random Graphs

Histogram of eigvals of random 40-by-40 adjacency matrices



Distribution of eigenvalues of Random Graphs

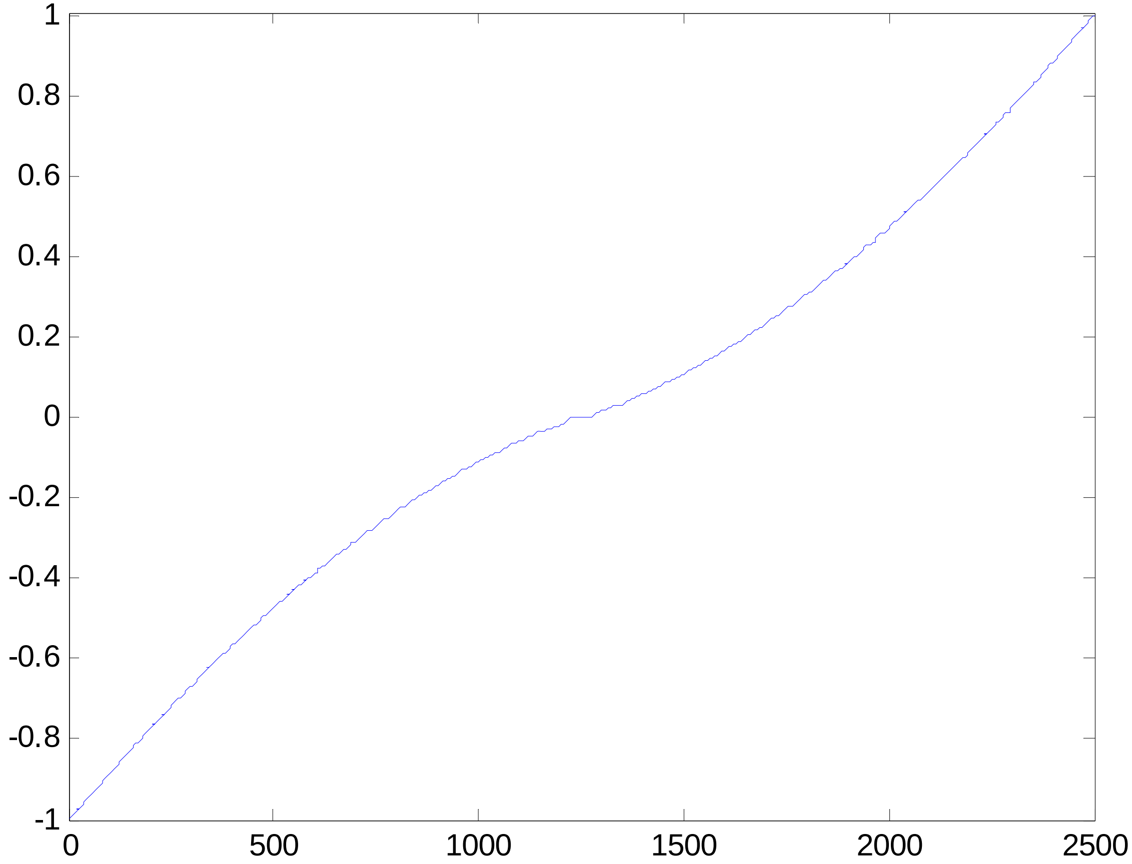
Histogram of eigvals of random 40-by-40 adjacency matrices



Predicted curve: Wigner’s Semi-Circle Law

Eigenvalue distributions

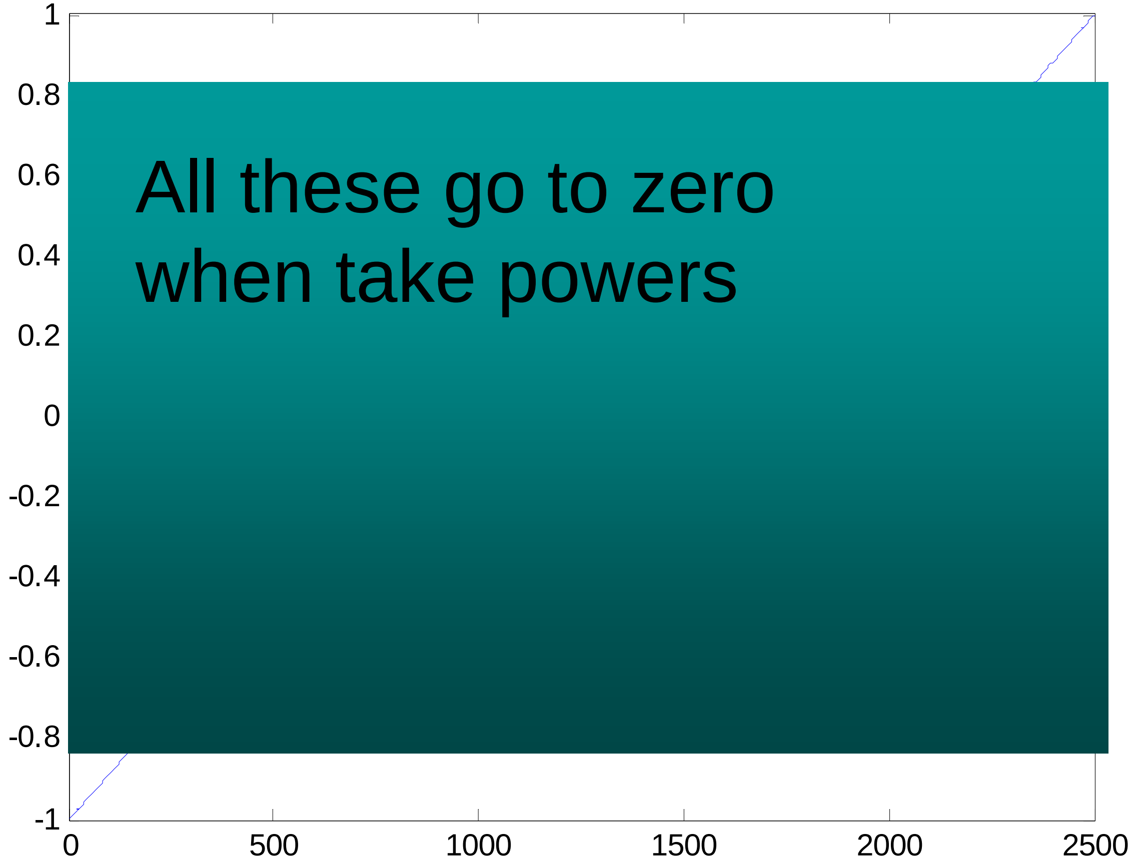
Eigenvalues of walk matrix of 50-by-50 grid graph



Number greater than 1-ε proportional to ε

Eigenvalue distributions

Eigenvalues of walk matrix of 50-by-50 grid graph



Number greater than 1-ε proportional to ε

Compression of powers of graphs

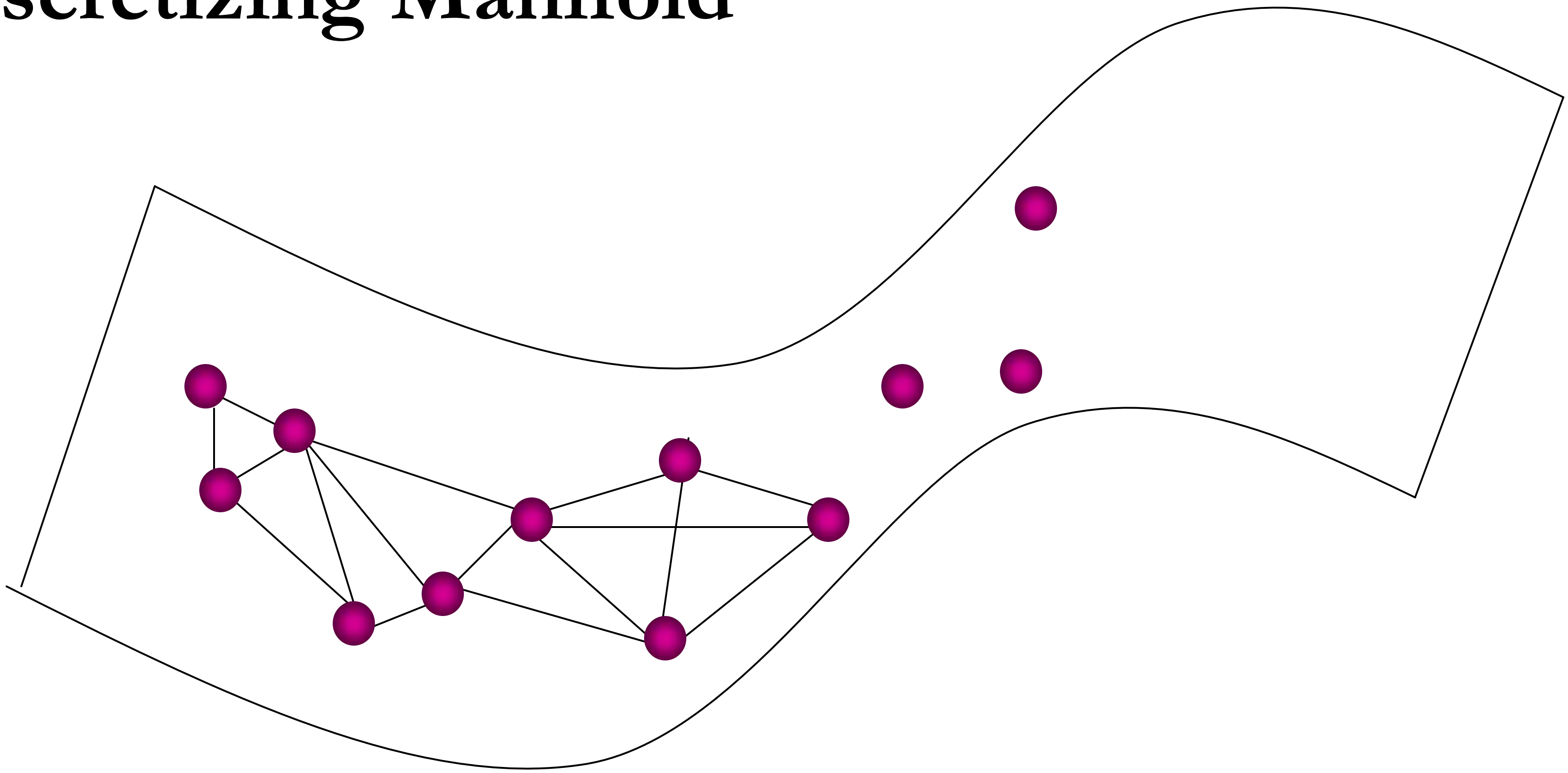
[Coifman, Lafon, Lee, Maggioni, Nadler, Warner, Zucker '05]

If most eigenvalues of A and W bounded from 1.
Most eigenvalues of A^t very small.
Can approximate A^t by low-rank matrix.

Build wavelets bases on graphs.
Solve linear equations and compute eigenvectors.

Make rigorous by taking graph from
discretization of manifold

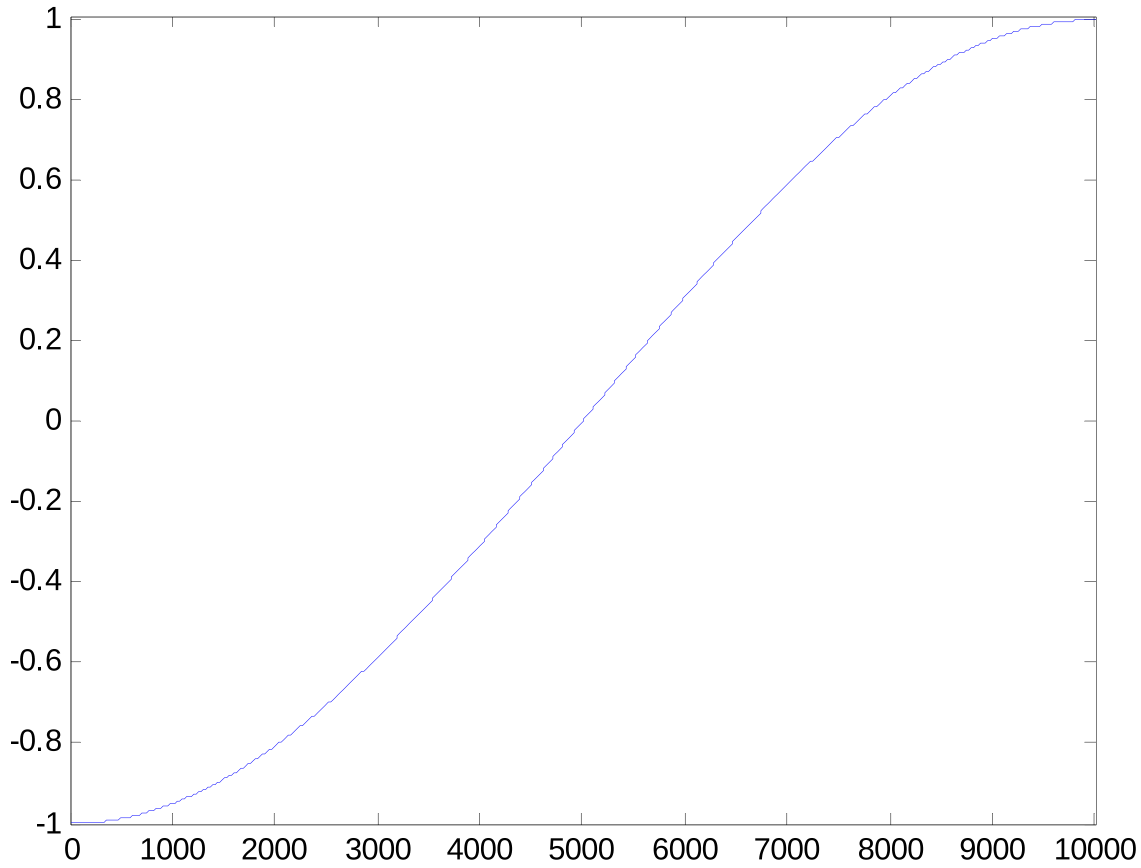
Discretizing Manifold



edge weight $e^{-\mathrm{dist}(x,y)^2/t^2}$

Eigenvalue distributions

Eigenvalues of path graph on 10k nodes



Number greater than 1-ε proportional to $\sqrt{\epsilon}$

Theorem: Eigenvalue distributions

Theorem:
If bounded degree,
number eigenvalues greater than 1-ε is $O(\sqrt{\epsilon})$

Proof: can choose $\sqrt{\epsilon}n$ vertices to collapse so that conductance becomes at least ϵ (like adding an expander on those nodes).

New graph has all eigvals at most 1-ε in abs value.
Is rank $\sqrt{\epsilon}n$ change, so by Courant-Fischer

$$\left|\lambda_{\sqrt{\epsilon}n}\right| \leq 1 - \epsilon$$

Eigenvalue distributions of planar graphs?

For planar graphs,
Colan de Verdiere's results imply

$$\lambda_2 < \lambda_5$$

How big must the gap be?
Must other gaps exist?

Computation

If λ exactly an eigenvalue,
eigvecs = Null($A - \lambda I$)

Not rational, so only approximate

If λ close to just one eigenvalue λ_i

$$(A - \lambda I)^{-1} \approx \frac{1}{\lambda_i - \lambda} v_i v_i^T$$

If λ_i close to λ_{i+1} is like $\lambda_i = \lambda_{i+1}$
v_i and v_{i+1} can rotate with each other

General Symmetric Matrices

Locate any eigval in time $O(n^3 + n \log(1/\epsilon))$

1. Orthogonal similarity transform to tri-diagonal in time $O(n^3)$ by elimination algorithm.

2. Given tri-diagonal matrix,
   count number eigenvalues in any interval
   in time $O(n)$

3. Do binary search to locate eigenvalue

Locate eigenvector: $O(n)$ steps on tri-diagonal,
$O(n^2)$ time to map back to A

Largest eigenvectors by power method

Apply A to random vector r:

$$r = \sum_i \left(v_i^T r\right) v_i$$

$$Ar = \sum_i \lambda_i \left(v_i^T r\right) v_i$$

$$A^t r = \sum_i \lambda_i^t \left(v_i^T r\right) v_i \approx \lambda_n^t \left(v_n^T r\right) v_n$$

In $O((\log n)/\epsilon)$ iters, expect x such that

$$\frac{x^T A x}{x^T x} \geq (1-\epsilon)\lambda_n$$

Using Lanczos, expect $O(\sqrt{(\log n)/\epsilon})$ iters

(better polynomial)

Smallest eigenvectors by inverse power method

Apply L^{-1} to random vector r orthogonal to $\mathbf{1}$

$$L^{-1} r = \sum_{i \geq 2} (1/\lambda_i) \left(v_i^T r \right) v_i$$

$$L^{-t} r = \sum_{i \geq 2} \lambda_i^{-t} \left(v_i^T r \right) v_i \approx \lambda_2^{-t} \left(v_2^T r \right) v_2$$

In $O((\log n)/\epsilon)$ iters, expect x such that

$$\frac{x^T L x}{x^T x} \leq (1+\epsilon)\lambda_2$$

Compute $L^{-1} r$ in time $m \log^{O(1)} n \log(1/\epsilon)$ [STeng04]

if planar, in time $O(m \log(1/\epsilon))$ [Koutis-Miller 06]

Sparsification

Key to fast computation.

Replace A by sparse B for which

$$\max \frac{x^T L_A x}{x^T L_B x} \leq c \qquad\qquad \min \frac{x^T L_A x}{x^T L_B x} \geq 1/c$$

Generalized eigenvalues provide notion of approximation in graphs.

Questions

Cheeger's inequality for other physical problems?

How to incorporate other data into spectral methods?

Make multilevel coarsening rigorous.

What can we do with boundary conditions?

What about generalized eigenvalue problems?

$$\frac{x^T L x}{x^T M x}$$