

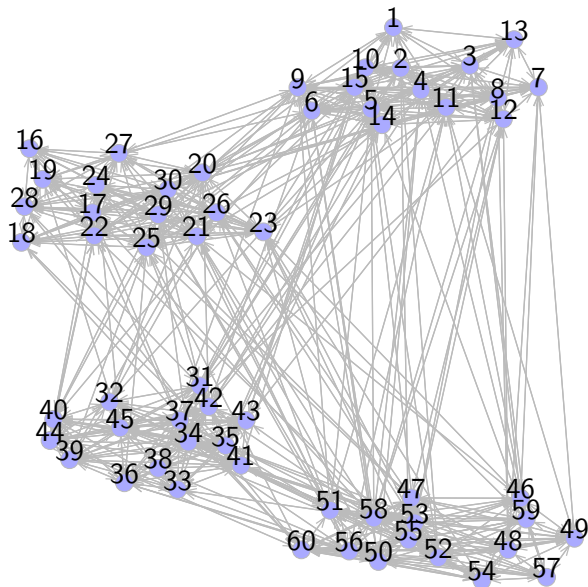
# Eigenresidual Tolerances for Spectral Partitioning

James P Fairbanks

School of Computational Science and Engineering  
Georgia Institute of Technology

March 21 2016

# Graph Partitioning



# Spectral Graph Theory

---

## Graphs and Linear Algebra

---

Adjacency Matrix  $A_{ij} = \begin{cases} 1, & \text{if } v_i \sim v_j \\ 0, & \text{otherwise} \end{cases}$

Degrees  $D_{ii} = d_i = \deg(v_i)$  all other entries 0.

Combinatorial Laplacian  $L = D - A$

Normalized Laplacian  $\hat{L} = I - D^{-1/2}AD^{-1/2}$

Normalized Adjacency  $\hat{A} = D^{-1/2}AD^{-1/2}$

---

# Defining Accuracy for the Numerical Problem [Golub and Van Loan, 2013]

Eigenpair $\lambda_i, \mathbf{v}_i$	$\hat{L}\mathbf{v}_i = \lambda_i\mathbf{v}_i$
Symmetric Positive Semidefinite	$\lambda_1 = 0 \leq \lambda_2 \leq \dots \lambda_n \leq 2 \in \mathbb{R}^+$
Rayleigh Quotient	$\mu = \frac{\mathbf{x}^T \hat{L}\mathbf{x}}{\mathbf{x}^T \mathbf{x}}$
Residual	$r = \left\  \hat{L}\mathbf{x} - \mu\mathbf{x} \right\ $
Error	$\mathbf{z} = \left\  \mathbf{x} - \mathbf{v} \right\ $ where $\hat{L}\mathbf{v} = \lambda\mathbf{v}$

## Define the Quality of Partitioning as Conductance [Chung, 1997]

- ▶ Define  $E(S) = \sum_{i \in S, j \notin S} a_{i,j}$  total weight of edges leaving  $S$ .
- ▶ Define  $\text{Vol}(S) = e_S^T A e_S$  as total weight of edges within  $S$
- ▶ The conductance of a cut  $S$  is:

$$\phi(S) = \frac{E(S)}{\min\{\text{Vol}(S), \text{Vol}(\bar{S})\}}$$

- ▶ When applying spectral methods, accurate eigenpairs are not the direct goal

## Spectral Sweep Cut Algorithm

- ▶ Compute  $\mathbf{x}$  so that  $\|\hat{L}\mathbf{x} - \mu\mathbf{x}\|_2 < \epsilon$
- ▶ Sort  $D^{-\frac{1}{2}}\mathbf{x}$
- ▶ Check all  $n$  sweep cuts of  $D^{-\frac{1}{2}}\mathbf{x}$  for minimum  $\phi$
  
- ▶ You need to choose  $\epsilon$  small enough to get correct cut.
- ▶ We provide the first principled method for choosing  $\epsilon$  for spectral partitioning

# General Cheeger's Inequality [Mihail, 1989]

## Theorem

If  $\mathbf{x}$  is a vector orthogonal to  $D^{\frac{1}{2}}\mathbf{1}$  such that  $\mathbf{x}^T \hat{L}\mathbf{x} = \mu$ , then  $\mathbf{y} = D^{-\frac{1}{2}}\mathbf{x}$  has a sweep cut  $S$  such that  $\phi(S) = \phi(D^{-\frac{1}{2}}\mathbf{x}) \leq \sqrt{2\mu}$ .

- ▶ Smallest possible  $\mu$  is  $\lambda_2$  so approximate  $\mathbf{v}_2$
- ▶ Low energy vectors wrt Laplacian yield good partitions
- ▶ Goal: show that these vectors are easier to compute with high accuracy

## Eigenspace error bounds

We generalize an eigenspace error bound to invariant subspaces.

	[Parlett, 1998]	[F, Sanders 2015]
Target	$\mathbf{x} \approx \mathbf{v}_2$	$\mathbf{x} \approx \sum_2^q \alpha_i \mathbf{v}_i$
Error	$\ \mathbf{x} - \mathbf{P}_2 \mathbf{x}\ $	$\ \mathbf{x} - \mathbf{P}_2^q \mathbf{x}\ $
Gap $\delta$	$ \lambda_2 - \lambda_3 $	$ \mu - \lambda_{q+1} $
<i>Bound</i>	$\ \mathbf{z}\  \delta_2 < \sqrt{8} \ \mathbf{r}\ $	$\ \mathbf{z}\  \delta_q < \sqrt{2} \ \mathbf{r}\ $
Condition	$\mu \leq \frac{\lambda_2 + \lambda_3}{2}$	$\mu \leq \lambda_q$

Our improvement is based on  $|\mu - \lambda_{q+1}| > |\lambda_2 - \lambda_3|$ .



## Blends and Blendspaces

- ▶ Analyze the problem to determine a set of “good” eigenvectors
- ▶ Transition from computing an approximation of a single eigenvector to an approximation of a linear combination
- ▶ Give up control over which linear combination you compute
- ▶ Gain tighter bounds and faster convergence

# Ring of Cliques

$\mathcal{R}_{b,q}$ :  $q$  blocks each of size  $b$ .

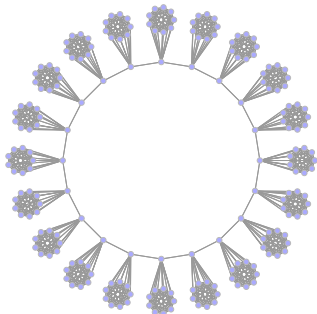


Figure 1: A drawing of  $\mathcal{R}_{b,q}$  laid out to show structure.

$$\begin{bmatrix} J_b & \mathbf{e}_1 \mathbf{e}_1^t & 0 & \dots & & \mathbf{e}_1 \mathbf{e}_1^t \\ \mathbf{e}_1 \mathbf{e}_1^t & J_b & \mathbf{e}_1 \mathbf{e}_1^t & 0 & & \\ 0 & \mathbf{e}_1 \mathbf{e}_1^t & J_b & \mathbf{e}_1 \mathbf{e}_1^t & 0 & \\ \vdots & & \ddots & \ddots & \ddots & \\ & & 0 & \mathbf{e}_1 \mathbf{e}_1^t & J_b & \mathbf{e}_1 \mathbf{e}_1^t \\ \mathbf{e}_1 \mathbf{e}_1^t & & & 0 & \mathbf{e}_1 \mathbf{e}_1^t & J_b \end{bmatrix}$$

where  $J_b = \mathbf{1}_b \mathbf{1}_b^t - I_b$ .

Figure 2: The adjacency matrix of  $\mathcal{R}_{b,q}$  has block circulant structure.

# Ring of Cliques Correct Partitions

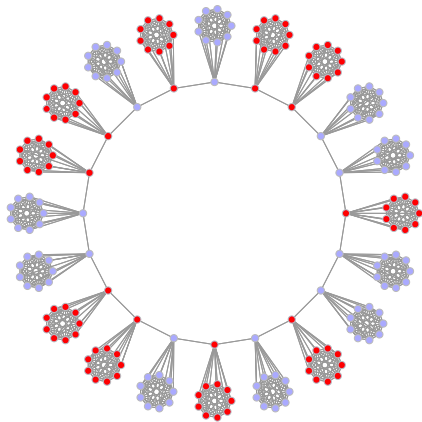


Figure 3: In a “correct” partition of  $\mathcal{R}_{b,q}$  all cliques are entirely red or entirely blue.

## $\mathcal{R}_{b,q}$ Spectrum

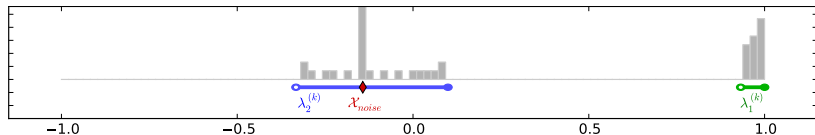


Figure 4: Spectrum of  $\hat{A} = I - \hat{L}$  of  $\mathcal{R}_{34,10}$ .

- ▶ Blend interval  $[\lambda_q, \lambda_p]$  in green
- ▶ Blend space  $\text{Span}\{v_2 \dots v_q\}$
- ▶  $\lambda(\hat{A}) > \frac{1}{2}$  eigenvectors recover partition
- ▶ Spectral gap  $\delta_2 = \mathcal{O}(n^{-2})$
- ▶ Blend gap  $\delta_q = \mathcal{O}(1)$

# Eigenpairs of $\mathcal{R}_{b,q}$ normalized Adjacency Matrix

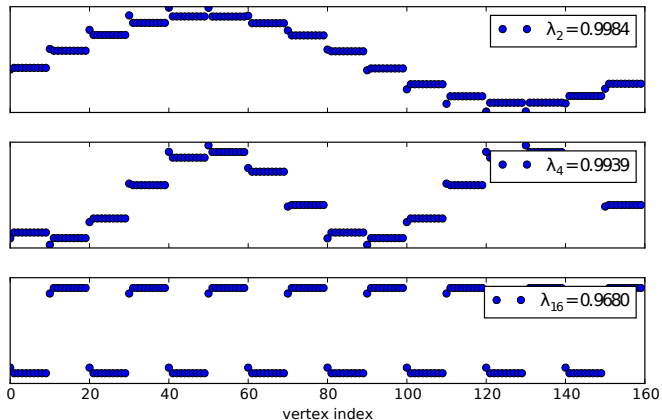


Figure 5: The eigenvectors of  $\hat{A}$  are shown for  $\mathcal{R}_{10,16}$ . The eigenvectors with eigenvalues close to 1 indicate the block structure with differing frequencies.

## Constructing a minimal perturbation $z$

Let  $\alpha_i = \max_{j \in \mathcal{B}_i} z_j$  and  $\beta_i = \min_{j \in \mathcal{B}_i} z_j$  for each  $\mathcal{B}_i$ .

If for all  $i$   $\alpha_i - \beta_{i+1} > \mu_{i+1} - \mu_i$  then there is no good sweep cut.

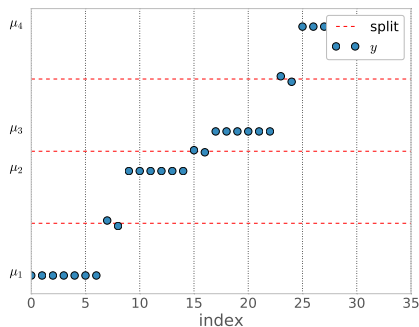


Figure 6: Example of the minimal perturbation construction.

## $\mathcal{R}_{b,q}$ Sufficient Errors

### Theorem

Let  $\mathbf{P}_{\mathcal{W}}$  be the orthogonal projector onto  $\mathcal{W}$ .

For any vector  $\mathbf{y}$ , define  $\mathbf{z} = (\mathbf{I} - \mathbf{P}_{\mathcal{W}})\mathbf{y}$ .

If  $\|\mathbf{z}\|_2 \leq (1 + 2qn)^{-\frac{1}{2}} \|\mathbf{y}\|_2$ , then  $\mathbf{y}$  recovers all the cliques of  $\mathcal{R}_{b,q}$ .

### Sketch of proof.

- ▶ Let  $\alpha_i = \max_{j \in \mathcal{B}_i} z_j$  and  $\beta_i = \min_{j \in \mathcal{B}_i} z_j$  for each  $\mathcal{B}_i$ .
- ▶ Suppose that  $\mathbf{y}$  does not recover all the cliques, then for all  $\mathcal{B}_i$

$$\alpha_i - \beta_{i+1} > \mu_{i+1} - \mu_i.$$

- ▶ Thus we can bound the 1-norm error below as follows:

$$\|\mathbf{z}\|_1 \geq \sum_i (\alpha_i - \beta_{i+1}) \geq \mu_q - \mu_1 \geq n^{-\frac{1}{2}} \|\mathbf{x}\|_2.$$

## $\mathcal{R}_{b,q}$ Necessary Accuracy

### Theorem

For any unit vector  $\mathbf{x} \in \mathcal{W}$  there exists a perturbation  $\mathbf{z}$  where  $\|\mathbf{z}\| < b^{-\frac{1}{2}}$ ,  $\mathbf{P}_{\mathcal{W}}\mathbf{z} = 0$  such that  $\mathbf{y} = \mathbf{x} + \mathbf{z}$  does not recover all the cliques.

Sketch of proof.

$$\|\mathbf{z}\|_2^2 = \sum_{i=0}^{q-1} \alpha_i^2 + \beta_{i+1}^2 = \frac{1}{2} \sum_{i=0}^{q-1} (\mu_{i+1} - \mu_i)^2 < b^{-1} \|\mathbf{x}\|_2^2$$





## $\mathcal{R}_{b,q}$ Minimal Perturbations

Residual tolerances are larger for blends because:

$$\delta_2 = \mathcal{O}(n^{-2}) \text{ but } \delta_q = \mathcal{O}(1)$$

$q$	$b$	sufficient	necessary
$n/25$	25	$\frac{5}{2}n^{-1}$	$\frac{1}{\sqrt{5}}$
25	$n/25$	$\frac{1}{14}n^{-\frac{1}{2}}$	$5n^{-\frac{1}{2}}$
$\sqrt{n}$	$\sqrt{n}$	$\frac{1}{2}n^{-3/4}$	$\frac{1}{2}n^{-1/4}$

Table 1: Error bounds for various parameter regimes.

# $\mathcal{R}_{b,q}$ Minimal Perturbations

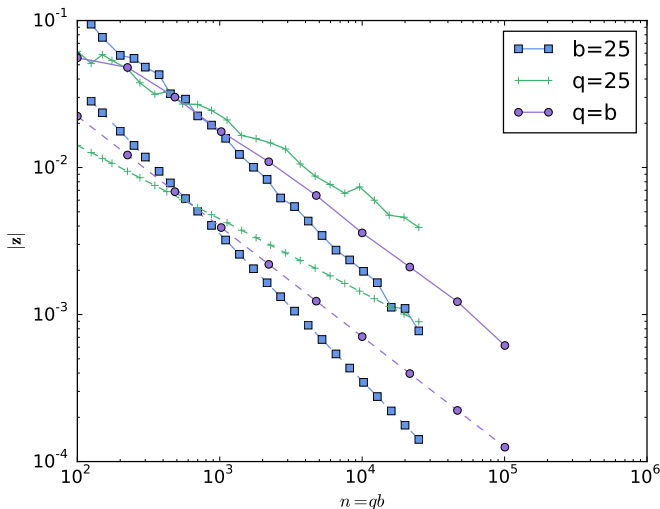


Figure 7: Empirical measurements of minimal error perturbations. Dashed lines indicate lower bounds.

## $\mathcal{R}_{b,q}$ Power Method

### Theorem

Let  $\mathbf{x}_0$  be sampled from  $\mathcal{N}_n(0, 1)$ , with high probability:

The power method  $\mathbf{x}_{k+1} \leftarrow \hat{A}\mathbf{x}_k / \|\mathbf{x}_k\|$  for  $\mathcal{R}_{b,q}$ .





There is a  $k^*$  of  $\mathcal{O}(\log_b q)$  such that for  $k \geq k^*$  a sweep cut of  $\mathbf{x}_k$  recovers the clusters.

For  $q \in \mathcal{O}(b^p)$ , the number of iterations is  $\mathcal{O}(p)$ .

## Conclusions

- ▶ Spectral blends lead to lower accuracy requirements in spectral partitioning
- ▶ Residual tolerance necessary for Fiedler partitions is  $\mathcal{O}(n^{-5/2})$  while data clusters are resolved at  $\mathcal{O}(n^{-1/2})$  for  $\mathcal{R}_{b,q}$
- ▶ Analyzing model problems for data mining is important for determining accuracy requirements for iterative methods
- ▶ Data Analysis structure can be revealed faster than accurate solutions to numerical problem

# References I

-  Chung, F. R. (1997).  
*Spectral graph theory*, volume 92.  
American Mathematical Soc.
-  Golub, G. H. and Van Loan, C. F. (2013).  
*Matrix Computations (4th Ed.)*.  
Johns Hopkins University Press, Baltimore, MD, USA.
-  Mihail, M. (1989).  
Conductance and convergence of markov chains-a  
combinatorial treatment of expanders.  
In *2013 IEEE 54th Annual Symposium on Foundations of  
Computer Science*, pages 526–531. IEEE Comput. Soc. Press.
-  Parlett, B. N. (1998).  
*The Symmetric Eigenvalue Problem*.  
Society for Industrial and Applied Mathematics.

# Block Means Are Spread Out

## Lemma

Define  $\mathcal{B}_i = \{ib + 1 \dots ib + q\}$

Let  $\mathcal{W}$  be the span of  $\{\mathbf{e}_{\mathcal{B}_i} \mid i \in 0 \dots q - 1\}$ .

For any vector  $\mathbf{x}$ , let  $\mu_j = |\mathcal{B}_j|^{-1} \sum_{i \in \mathcal{B}_j} x_i$ .

For any vector  $\mathbf{x} \in \mathcal{W}$ ,  $\|\boldsymbol{\mu}\|_\infty > n^{-1/2} \|\mathbf{x}\|_2$ .

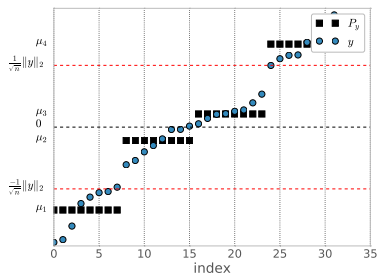


Figure 8: Notation in Lemma