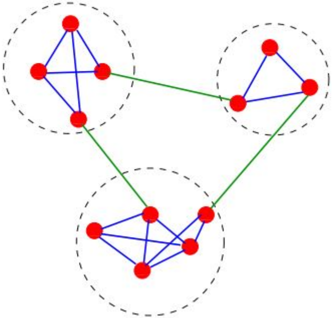
On statistical analysis of spectral graph algorithms for community detection in networks

Ambedkar Dukkipati

Department of Computer Science & Automation
Indian Institute of Science

17 April 2017

The Problem - Graph Clustering



- Partition a graph G into k ‘clusters’.
- Cluster Properties
  - Many edges within clusters
  - Few edges between clusters
- Partitioning Objective
  - Cut across fewest edges possible

Why is this hard?

- Graph partitioning is NP-hard
- Brute force?
  - For a small graph with 100 nodes, the number of different partitions exceeds the number of atoms in the universe!
- Heuristics?
  - Optimality, consistency, efficiency ...

Spectral...Why and What?

Why?

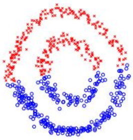
- Nice approximations that give rise to polynomial time algorithms
- with theoretical guarantees, provided by statistical analysis.

What?

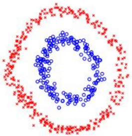
- Underlying objects in a problem can be represented as matrices
- Eigenvalues and eigenvectors of these matrices become a clue to the solution.

Spectral Clustering

- Well studied in literature
- Strong theoretical grounding
  - Spectral Graph Theory
  - Consistency results
- Efficient linear algebraic computations



K-means



Spectral Clustering

Ng et al. *NIPS, 2001*

Graph Coloring

Theorem (Brooks)

Apart from the following cases

1. G is complete
2. G has odd cycles

we have $\chi_G \leq d_{max}$

Theorem (Gershgorin Disk)

Assume A is a nonnegative $n \times n$ real matrix. Then all eigenvalues of A lie in the set

$$\bigcup_{i=1}^{n} \left[A_{ii} - \sum_{j \neq i} A_{ij},\ A_{ii} + \sum_{j \neq i} A_{ij} \right]$$

Graph Coloring

Lemma

Let A be the adjacency matrix of $G = (V, E)$. Let $\mu_1 \geq \mu_2 \geq \ldots \geq \mu_n$ be the eigenvalues of A. Then $\mu_1 \leq d_{max}$.

Proof: By Gershgorin theorem

$$\begin{aligned}
\mu_1 &\leq \max_{1 \leq i \leq n} \left(A_{ii} + \sum_{j \neq i} A_{ij} \right) \\
&= \max_{1 \leq i \leq n} \sum_{j=1}^{n} A_{ij} \\
&= \max_{1 \leq i \leq n} \deg(i) = d_{max}
\end{aligned}$$

Graph Coloring

The previous result can be proved using Rayleigh's principle.

Theorem (Rayleigh's Principle)

Let A be a nonnegative $n \times n$ real matrix and Let μ_1 be the largest eigenvalues of A then

$$\mu_1 = \max_{v \neq 0} \frac{v^T A v}{v^T v}$$

Note: A is a adjacency matrix of graph G and let μ_1 be the largest eigenvalue of A. Then we already have the following:

- $\chi_n \leq d_{max}$
- $\mu_1 \leq d_{max}$

Theorem (Wilf, 1967)

$$\chi_G \leq \lfloor \mu_1 \rfloor + 1$$

Some matrices related to graphs

Let $G = (V, E)$ be a graph. $|V| = n$ and $|E| = e$.

- **Adjacency Matrix:** $A \in \mathbb{R}^{n \times n}$ such that

$$A_{ij} = \begin{cases} 0 & \text{if} \quad i = j, \\ 1 & \text{if} \quad (i,j) \in E, \\ 0 & \text{if} \quad (i,j) \notin E. \end{cases}$$

- **Degree Matrix:** $D \in \mathbb{R}^{n \times n}$ is diagonal matrix such that $D_{ii} = \deg(i)$
- **Incidence Matrix:** $B \in \mathbb{R}^{n \times e}$, where rows indexed by vertices and columns indexed by edges and $B_{ij} = 1$ if vertex i lies on edge j.
- **Laplacian Matrix:** $L \in \mathbb{R}^{n \times n}$ is defined as $L = D - A$
- **Normalized Laplacian:** $L \in \mathbb{R}^{n \times n}$ is defined as $L = I - D^{-1/2} A D^{-1/2}$

Graph Laplacian

Let $G = (V, E)$ be a graph. $|V| = n$ and $|E| = e$. **Laplacian:** $L \in \mathbb{R}^{n \times n}$ such that

$$L_{ij} = \begin{cases} d_i & \text{if} \quad i = j, \\ -1 & \text{if} \quad (i,j) \in E, \\ 0 & \text{if} \quad (i,j) \notin E. \end{cases}$$

Theorem

Let $\lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n$ be eigenvalues of L. Then

1. L is symmetric and positive semidefinite
2. $\lambda_1 = 0$
3. $\lambda_2 > 0$ iff G is connected
4. $\lambda_k = 0$ and $\lambda_{k+1} > 0$ iff G has exactly k-disjoint components

Cuts

Let $G = (V, E)$ be a graph. $|V| = n$ and $|E| = e$. Let $V_1 \subset V$.
Boundary: The boundary of V_1 is defined as

$$\delta V_1 = \{(i, j) \in E : \ i \in V_1 \ \text{and} \ j \notin V_1\}$$

- **Cut:**

$$\mathrm{Cut}(V_1) = |\delta V_1|$$

- **Expansion Cut**

$$\mathrm{ExpansionCut}(V_1, V - V_1) = \frac{|\delta V_1|}{\min\{|V_1|, |V - V_1|\}}$$

- **Ratio Cut:**

$$\mathrm{RatioCut}(V_1, V - V_1) = \frac{|\delta V_1|}{|V_1|} + \frac{|\delta V_1|}{|V - V_1|}$$

Metrics for partitioning

Let $G = (V, E)$ be a graph. $|V| = n$ and $|E| = e$. Let $V_1 \subset V$.
Boundary: The boundary of V_1 is defined as

$$\delta V_1 = \{(i, j) \in E : \ i \in V_1 \ \text{and} \ j \notin V_1\}$$

- **Edge Expansion:**

$$\phi_G = \min_{|V_1| \leq \frac{|V|}{2}} \frac{|\delta V_1|}{|V_1|}$$

- **Ratio Cut:**

$$\eta_G = \min_{|V_1| \leq \frac{|V|}{2}} \frac{|\delta V_1|}{|V_1|} + \frac{|\delta V_1|}{|V - V_1|}$$

A simple calculation of $x^T L x$

$$
\begin{aligned}
x^T L x &= x^T D x - x^T A x \\
&= \sum_{i=1}^{n} d_i x_i^2 - \sum_{i,j=1}^{n} A_{ij} x_i x_j \\
&= \sum_{i=1}^{n} d_i x_i^2 - \sum_{(i,j) \in E} x_i x_j + x_j x_i \\
&= \sum_{(i,j) \in E} (x_i^2 + x_j^2) - \sum_{(i,j) \in E} x_i x_j + x_j x_i \\
&= \sum_{(i,j) \in E} (x_i - x_j)^2
\end{aligned}
$$

Rayleigh Principle or Courant-Fisher Theorem

Theorem

Let M be a symmetric matrix and let $\theta_1 \leq \theta_2 \leq \ldots \leq \theta_n$ be eigenvalues of M. Then

$$\theta_k = \max_{n-k+1 \dim T} \quad \min_{x \in T, x \neq 0} \frac{x^T M x}{x^T x}$$

Theorem

Let L be the Laplacian of a graph $G = (V, E)$. Then

$$\lambda_2 = \min_{x \perp 1} \frac{x^T M x}{x^T x}$$

Cheeger's Inequality

Definition (Cheeger's Constant)

Let $G = (V, E)$ be a graph and consider a graph bisection problem. Then

$$\phi_G = \min_{|V_1| \leq \frac{n}{2}} \frac{|\delta V_1|}{|V_1|}$$

Theorem (Cheeger's Inequality)

Let $d_{\max}$ denote the maximum degree of G and λ_2 be the second smallest eigenvalue of the Laplacian L of G. Then

$$\frac{\lambda_2}{2} \leq \phi_G \leq \sqrt{2\lambda_2 d_{\max}}$$

Note: Look at proofs of Mohar and Spielman

Cheeger's Inequality (Contd...)

Definition (Cheeger's Constant)

Let $G = (V, E)$ be a graph and consider a graph bisection problem. Then

$$\phi_G = \min_{|V_1| \leq \frac{n}{2}} \frac{|\delta V_1|}{|V_1|}$$

Theorem (Cheeger's Inequality)

Let $d_{\max}$ denote the maximum degree of G and λ_2 be the second smallest eigenvalue of the Laplacian L of G. Then

$$2\phi_G \leq \lambda_2 \leq \frac{{\phi_G}^2}{2}$$

Note: Look at proofs of Mohar and Spielman

Graph Bisection

Recall **Ratio Cut:**

$$\mathrm{RCut}(V_1, V_1^c) = \frac{|\delta V_1|}{|V_1|} + \frac{|\delta V_1|}{|V_1^c|}$$

A simple calculation shall give us this:
Define $y \in \mathbb{R}^n$ as

$$y_i = \begin{cases} \sqrt{\frac{|V_1^c|}{|V_1||V|}} & \text{if} \quad i \in V_1, \\ -\sqrt{\frac{|V_1|}{|V_1^c||V|}} & \text{if} \quad i \notin V_1. \end{cases}$$

Then

$$y^T L y = \mathrm{Rcut}(V_1, V_1^c)$$

Let say $\mathcal{Y}^*$ as subset of $\mathbb{R}^n$ denote various y defined as in (*) for various subsets of V_1 of V.

Graph Bisection (contd..)

Objective:

$$\min_{y \in \mathcal{Y}^*} y^T L y$$

Trivial Relaxation:

$$\min_{y \in \mathbb{R}^n} y^T L y$$

Not very useful as $1^T L 1 = 0$

Nice Relaxation:
Since $y^T 1 = \sum_{i \in V} y_i = 0$, y is orthogonal to 1. Also since $y^T y = \sum_{i \in V} y_i^2 = 1$, y is a unit norm vector. Hence the relaxed problem can be

$$\min_{y \perp 1} \frac{y^T L y}{y^T y}$$

Graph k-way partitioning

Ratio Cut:

$$\mathrm{Rcut}(V_1, \ldots, V_k) = \sum_{\ell=1}^{k} \frac{|\delta V_\ell|}{|V_\ell|}$$

Lets define Y: Define $y \in \mathbb{R}^{n \times k}$ such that

$$Y_{i\ell} = \begin{cases} \frac{1}{\sqrt{|V_\ell|}} & \text{if} \quad i \in V_\ell, \\ & \\ 0 & \text{otherwise.} \end{cases} \qquad (**)$$

Claim: $Y^T Y = I$

Claim: $\mathrm{Rcut}(V_1, \ldots, V_k) = \mathrm{Trace}(Y^T L Y)$

Graph k-way partitioning

- Objective

$$\min_{Y \in \mathcal{Y}^{**}} \mathrm{Trace}(Y^T L Y)$$

- Relaxation

$$\min_{\substack{Y \in \mathbb{R}^n \\ Y^T Y = I}} \mathrm{Trace}(Y^T L Y)$$

- Optimal Value

$$Y^{\mathrm{opt}} = [v_1 \ldots v_k]$$

matrix of k leading orthonormal eigenvectors of L

With Normalized Cuts

Normalized Cut:

$$\text{Ncut}(V_1, \ldots, V_k) = \sum_{\ell=1}^{k} \frac{|\delta V_\ell|}{\text{Vol}(V_\ell)}$$

where $\text{Vol}(V_\ell) = \sum_{i \in V_\ell} \deg(i)$

Lets define Y again: Define $y \in \mathbb{R}^{n \times k}$ such that

$$Y_{i\ell} = \begin{cases} \frac{1}{\sqrt{\text{Vol}(V_\ell)}} & \text{if} \quad i \in V_\ell, \\ 0 & \text{otherwise.} \end{cases} \qquad (***)$$

Claim: $Y^T D Y = I$

Claim: $\text{Ncut}(V_1, \ldots, V_k) = \text{Trace}(Y^T L Y)$

With normalized cuts

- Objective

$$\min_{Y \in \mathcal{Y}^{***}} \mathrm{Trace}(Y^T L Y)$$

- Relaxation

$$\min_{\substack{Y \in \mathbb{R}^n \\ Y^T D Y = I}} \mathrm{Trace}(Y^T L Y)$$

- By substituting $\widetilde{Y} = D^{\frac{1}{2}} Y$ the objective translates to

$$\min_{\substack{\widetilde{Y} \in \mathbb{R}^n \\ \widetilde{Y}^T \widetilde{Y} = I}} \mathrm{Trace}(\widetilde{Y}^T D^{-\frac{1}{2}} L D^{-\frac{1}{2}} \widetilde{Y})$$

Spectral Clustering Algorithm

Algorithm

1. Compute graph Laplacian or normalized graph Laplacian
2. Compute k-leading eigenvectors $Y \in \mathbb{R}^{n \times k}$ of L
3. Normalize rows of Y and say it is $\bar{Y}$
4. Run k-means on rows of $\bar{Y}$
5. according to this partition V

K-means Step

$$S^* = \underset{\substack{S \in \mathbb{R}^{n \times k} \\ S \text{has at most } k \text{ distinct rows}}}{\arg\max} ||\bar{Y} - S||_F^2$$

On K-means

Must Look at: Ostrovsky et. al (2012): The Effectiveness of Lloyd-Type Methods for the k-Means Problem

Theorem

Assume that Y satisfies "epsilon-separability", where $\epsilon \leq 0.015$. Then the k-means algorithm of Ostrovsky (2012) returns a solution S^* such that

$$\|Y - S^*\|_F \leq (1+\epsilon) \min_{\substack{S \in \mathbb{R}^{n \times k} \\ S\,\text{has at most } k \text{ distinct rows}}} \|Y - S\|_F$$

with probability $(1 - O(\sqrt{\epsilon}))$ *in time* $O(nrk + rk^3)$. *Here,* $\gamma = \sqrt{\dfrac{1-\epsilon^2}{1-37\epsilon^2}}$.

Error

Let Z be the true membership matrix

$$Z_{i\ell} = \begin{cases} 1 & \text{if} \quad i \in V_\ell, \\ \\ 0 & \text{otherwise.} \end{cases}$$

Let $Z^{'}$ be the membership obtained from the algorithm.
Then the error is

$$\text{Error} = \min_{\substack{\text{Permutation matrices} \\ P \in {0,1}^{k \times k}}} \frac{1}{2} \|Z - Z^{'}P\|_F^2$$

Perturbation Analysis

Let $\tilde{B} \in \mathbb{R}^{n\times n}$ be a symmetric matrix

$H \in \mathbb{R}^{n\times n}$ be a symmetric perturbation matrix

and $B = \tilde{B} + H$

Let $\lambda_1 \leq \cdots \leq \lambda_n$ be the eigenvalues of $\tilde{B}$

$\mu_1 \leq \cdots \leq \mu_n$ be the eigenvalues of B

and $\rho_1 \leq \cdots \leq \rho_n$ be the eigenvalues of H

Matrix Perturbation Theory

Tools of the Trade: Weyl's Inequality

For $i = 1, \cdots, n$

$$\lambda_i + \rho_1 \leq \mu_i \leq \lambda_i + \rho_n$$

Corollary: $|\mu_i - \lambda_i| \leq \max\{|\rho_1|, |\rho_n|\} = \|B - \tilde{B}\|_2$

Tools of the Trade: Davis-Kahan Theorem

Let $\delta = \lambda_{k+1} - \lambda_k$.
Let $\tilde{Y}, Y$ be the k-leading orthonormal eigenvectors of $\tilde{B}, B$ respectively.
If $\delta > 2\|B - \tilde{B}\|_2$, then

$$\|Y - \tilde{Y}Q\|_F \leq \frac{2\sqrt{2k}}{\delta}\|B - \tilde{B}\|_2$$

for some orthonormal $Q \in \mathbb{R}^{k \times k}$.

Perturbation Analysis*

Let $G = (V, E)$ be a graph with Laplacian L. If there exists an "ideal graph" (that has equal sized disjoint components) with Laplacian $\tilde{L}$ such that

$$\|L - \tilde{L}\|_2 < \frac{n}{2k}$$

Then there exists orthonormal Q, $k \times k$ matrix such that

$$\|Y - \sqrt{\frac{k}{n}} ZQ\|_F \leq \frac{2k^{\frac{3}{2}}}{n} \|L - \tilde{L}\|_2$$

Here

$$\|A\|_2 = \max_{x \neq 0} \frac{\|Ax\|_2}{\|x\|_2}$$

and the error of SC is

$$\text{Error} \leq 256 \frac{k^2}{n} \|L - \tilde{L}\|_2$$

*(Ng and Jordan, 2002, NIPS)

Random Graph Models

- Latent Space Model
  - $z_1, \ldots, z_n \in \mathbb{R}^k$ - Latent vectors for each node. IID random variables.
  - The Model: For the random adjacency matrix $W \in \mathbb{R}^{n \times n}$

$$P(W|z_1, \ldots, z_n) = \prod_{i<j} P(W_{ij}|z_i, z_j)$$

  - $\mathscr{W} = \mathbb{E}(W|Z) \in \mathbb{R}^{n \times n}$ completely parametrises the model.
- Stochastic Block Model
  - Special case of Latent Space Model with

$$\mathscr{W} = ZBZ^T$$

  - Membership matrix $Z \in \{0, 1\}^{n \times k}$ has one 1 in each row
  - Block matrix $B \in [0, 1]^{n \times k}$

The Setup

- **Goal:** Prove that Spectral Clustering is weakly consistent over Stochastic Block Model
- All results will be asymptotic in n, the number of graph nodes
- Series of observed matrices $W^{(n)} \in \{0,1\}^{n \times n}$, $L^{(n)}$ and $D^{(n)}$
- Series of population matrices $\mathscr{W}^{(n)} \in [0,1]^{n \times n}$, $\mathscr{L}^{(n)}$ and $\mathscr{D}^{(n)}$

Stochastic Blockmodel Analysis

Question: Can we achieve consistency results if we let the the number of clusters grow with the number of nodes? (Rohe, Chatterjee and Yu, Ann. Stats, 2011)

Block Model: Let $Z \in \{0,1\}^{n \times k}$ and it has exactly one 1 in each row and atleast one 1 in each column. Let $B \in [0,1]^{k \times k}$ be a full rank and symmetric matrix, where diagonal elements of B has larger values than off diagonal. Then the stochastic block model is $\mathcal{W} = ZBZ^T$. ($\mathcal{W}$ is a population version of W)

Strategy:

- Given Z choose B and define $\mathcal{W}$
- Sample W from $\mathcal{W}$ and get Z' from a spectral algorithm. Compute the error by comparing Z' and Z.

Stochastic Blockmodel Analysis

Aim: Let $L^{(n)} \in \{0,1\}^{n\times n}$ and $\mathcal{L}^{(n)} \in [0,1]^{n\times n}$ be sequence of observed and population versions of Laplacians. Then show that under stochastic block model difference between eigenvectors of $L^{(n)}$ and $\mathcal{L}^{(n)}$ can bounded.

Result:(Rohe, Chatterjee and Yu, Ann. Stats, 2011) Spectral clustering algorithm is week consistent.

The Setup

- **Goal:** Prove that Spectral Clustering is weakly consistent over Stochastic Block Model
- All results will be asymptotic in n, the number of graph nodes
- Series of observed matrices $W^{(n)} \in \{0,1\}^{n \times n}$, $L^{(n)}$ and $D^{(n)}$
- Series of population matrices $\mathscr{W}^{(n)} \in [0,1]^{n \times n}$, $\mathscr{L}^{(n)}$ and $\mathscr{D}^{(n)}$

Proof Sketch

1. Bound the eigenvalues of $L^{(n)}$ and $\mathscr{L}^{(n)}$
2. Bound the eigenvectors of $L^{(n)}$ and $\mathscr{L}^{(n)}$
3. Bound the k-means error

Bounding Eigenvalues

Bird's eye view

1. Bound the Frobenius norm

$$\|L^{(n)} - \mathscr{L}^{(n)}\|_F = O(\cdots) \; \textit{almost surely}$$

2. $\|\cdots\|_2 \leq \|\cdots\|_F$
3. Weyl's inequality

$$\|L^{(n)} - \mathscr{L}^{(n)}\|_2 < \epsilon \Rightarrow \|\lambda_i^{(n)} - \tilde{\lambda}_i^{(n)}\| \leq \epsilon \;\; \forall i$$

Bounding Eigenvalues - Obstacle

Bird's eye view

1. Bound the Frobenius norm

$$\|L^{(n)} - \mathscr{L}^{(n)}\|_F = O(\cdots) \textit{ almost surely}$$

Not Possible!

2. $\|\cdots\|_2 \leq \|\cdots\|_F$
3. Weyl's inequality

$$\|L^{(n)} - \mathscr{L}^{(n)}\|_2 < \epsilon \Rightarrow \|\lambda_i^{(n)} - \tilde{\lambda}_i^{(n)}\| \leq \epsilon \quad \forall i$$

Bounding Eigenvalues - Obstacle - Example

Counter Example: $W \in \{0,1\}^{n\times n} \sim Bernoulli(1/2)$

- W/n behaves similar to $L = D^{-1/2} W D^{-1/2}$ as entries of D grow linearly with n.
- $\|W/n - \mathbb{E}(W)/n\|_F = \frac{1}{n}\sqrt{\sum_{i,j}(W_{ij} - \mathbb{E}(W_{ij}))^2} = 1/2$
  Diverges!
- However, $\|WW/n^2 - \mathbb{E}(WW)/n^2\|_F$ converges!

$$\|WW/n^2 - \mathbb{E}(WW)/n^2\|_F = \frac{1}{n^2}\sqrt{\sum_{i,j}([WW]_{ij} - \mathbb{E}[WW]_{ij})^2}$$
$$= o\left(\frac{\log n}{n^{1/2}}\right)$$

where $[WW]_{ij} \sim Binomial(n, 1/4)$

Bounding Eigenvalues - Obstacle - Resolution

- Bound $\|L^{(n)}L^{(n)} - \mathscr{L}^{(n)}\mathscr{L}^{(n)}\|_F$ instead of $\|L^{(n)} - \mathscr{L}^{(n)}\|_F$

Lemma

For a real symmetric matrix $M \in \mathbb{R}^{n\times n}$,

1. λ^2 is an eigenvalue of $MM \Leftrightarrow \lambda$ or $-\lambda$ is an eigenvalue of M.
2. $M\nu = \lambda\nu \Rightarrow MM\nu = \lambda^2\nu$.
3. $MM\nu = \lambda^2\nu \Rightarrow \nu$ can be written as linear combination of eigenvectors corresponding to λ or $-\lambda$.

- Therefore, spectrum of L is implied from that of LL.

Bounding Eigenvalues - Main Theorem

Theorem 1: Convergence in Frobenius Norm

Define

$$\tau_n = \min_i \mathscr{D}_{ii}^{(n)}/n$$

If there exists $N > 0$ such that $\tau_n^2 \log n > 2 \ \forall \ n > N$, then

$$\|L^{(n)}L^{(n)} - \mathscr{L}^{(n)}\mathscr{L}^{(n)}\|_F = o\left(\frac{\log n}{\tau_n^2 n^{1/2}}\right) \textit{ almost surely.}$$

- $\tau_n = \frac{\textit{min expected degree}}{\textit{max possible degree}}$
- τ_n is a measure of sparsity of the graph.

Bounding Eigenvalues - Main Theorem Proof

Tools of the Trade: Borel Cantelli Lemma

Let $E_1, \ldots, E_n$ be a sequence of events in a probability space.

$$\sum_{n=1}^{\infty} \mathbb{P}(E_n) < \infty \Rightarrow \mathbb{P}(\cap_{n=1}^{\infty} \cup_{k=n}^{\infty} E_k) = 0$$

Take E_n to be the event where $\frac{\|L^{(n)}L^{(n)} - \mathscr{L}^{(n)}\mathscr{L}^{(n)}\|_F}{c \log n / (\tau^2 n^{1/2} \epsilon)} \geq \epsilon$.

$$\therefore \|L^{(n)}L^{(n)} - \mathscr{L}^{(n)}\mathscr{L}^{(n)}\|_F = o\left(\frac{\log n}{\tau_n^2 n^{1/2}}\right) \textit{ almost surely.}$$

Bounding Eigenvalues - Proof of Non-Asymptotic Bound

Proof Strategy:

- $L = D^{-1/2} W D^{-1/2}$. D and W are not independent which means the entries of L are not independent.
  - *Independence is an essential ingredient for using concentration of measure inequalities!*
- Introduce an intermediate Laplacian $\tilde{L}$
  - $L = D^{-1/2} W D^{-1/2}$
  - $\tilde{L} = \mathscr{D}^{-1/2} W \mathscr{D}^{-1/2}$
  - $\mathscr{L} = \mathscr{D}^{-1/2} \mathscr{W} \mathscr{D}^{-1/2}$
- Introduce two sets Γ and Λ
  - Γ constrains the matrix D and helps in bounding $\|LL - \tilde{L}\tilde{L}\|_F$
  - Λ constrains $W \mathscr{D}^{-1} W$ and helps in bounding $\|\tilde{L}\tilde{L} - \mathscr{L}\mathscr{L}\|_F$
- Notation: $\mathbb{P}_{\Gamma\Lambda}(B) = \mathbb{P}(B \cap (\Gamma \cap \Lambda))$

Bounding Eigenvalues - Proof of Non-Asymptotic Bound

Define $a = \frac{32\sqrt{2}\log n}{\tau^2 n^{1/2}}$

$$
\begin{aligned}
\mathbb{P}\left(\|LL - \mathscr{L}\mathscr{L}\|_F \geq a\right) &\leq \mathbb{P}_{\Gamma\Lambda}\left(\|LL - \mathscr{L}\mathscr{L}\|_F \geq a\right) + \mathbb{P}\left((\Gamma \cap \Lambda)^c\right) \\
&\leq \mathbb{P}_{\Gamma\Lambda}\left(\sum_{i \neq j}[LL - \mathscr{L}\mathscr{L}]_{ij}^2 \geq a^2/2\right) \textit{ -- term 1} \\
&+ \mathbb{P}_{\Gamma\Lambda}\left(\sum_{i}[LL - \mathscr{L}\mathscr{L}]_{ii}^2 \geq a^2/2\right) \textit{ -- term 2} \\
&+ \mathbb{P}\left((\Gamma \cap \Lambda)^c\right) \textit{ -- term 3}
\end{aligned}
$$

Bounding Eigenvalues - Proof of Non-Asymptotic Bound

$$\mathbb{P}_{\Gamma\Lambda}\Big(\sum_{i\neq j}[LL - \mathscr{L}\mathscr{L}]_{ij}^2 \geq a^2/2\Big) \; - \textit{term } 1$$

$$\leq \sum_{i\neq j}\left[\mathbb{P}_{\Gamma\Lambda}\Big(|LL - \tilde{L}\tilde{L}|_{ij} \geq \frac{a}{\sqrt{8}n}\Big) + \mathbb{P}_{\Gamma\Lambda}\Big(|\tilde{L}\tilde{L} - \mathscr{L}\mathscr{L}|_{ij} \geq \frac{a}{\sqrt{8}n}\Big)\right]$$

$$\underbrace{|\tilde{L}\tilde{L} - \mathscr{L}\mathscr{L}|_{ij}}_{\textit{bound by } \Lambda} = \frac{1}{(\mathscr{D}_{ii}\mathscr{D}_{jj})^{1/2}}\Big|\sum_k (W_{ik}W_{jk} - \mathscr{W}_{ik}\mathscr{W}_{jk})/\mathscr{D}_{kk}\Big|$$

$$\leq \frac{1}{n^2\tau}\Big|\sum_k (W_{ik}W_{jk} - \mathscr{W}_{ik}\mathscr{W}_{jk})/\mathscr{D}_{kk}\Big|$$

$$\underbrace{|LL - \tilde{L}\tilde{L}|_{ij}}_{\textit{bound by } \Gamma} \leq \sum_k \left|\frac{1}{D_{kk}(D_{ii}D_{jj})^{1/2}} - \frac{1}{\mathscr{D}_{kk}(\mathscr{D}_{ii}\mathscr{D}_{jj})^{1/2}}\right|$$

Bounding Eigenvalues - Proof of Non-Asymptotic Bound

Define

$$\Lambda = \bigcap_{i,j} \left\{ \left| \sum_k (W_{ik}W_{jk} - \mathscr{W}_{ik}\mathscr{W}_{jk})/\mathscr{D}_{kk} < n^{1/2}\log n \right| \right\}$$

$$\Gamma = \bigcap_{i,j,k} \left\{ \frac{1}{D_{kk}(D_{ii}D_{jj})^{1/2}} \in \frac{[1 - n^{-1/2}\log n, 1 + n^{-1/2}\log n]}{\mathscr{D}_{kk}(\mathscr{D}_{ii}\mathscr{D}_{jj})^{1/2}} \right\}$$

- With Λ and Γ, $term\ 1 = 0$
- Similarly, we can show that $term\ 2 = 0$
- All that is remaining is to bound $term\ 3$

Bounding Eigenvalues - Proof of Non-Asymptotic Bound

Tools of the Trade: Hoeffding's Inequality

Let $X_1, \ldots, X_n$ be i.i.d. random variables with bounds $X_i \in [a_i, b_i]$. Let $S_n = \sum_{i=1}^n X_i$. Then

$$\mathbb{P}\left(\left|S_n - \mathbb{E}[S_n]\right| > t\right) \leq 2e^{-\frac{2t^2}{\sum_{i=1}^n (b_i - a_i)^2}}$$

- $D_{ii} \in [0, n] \; \forall \; i$ and are i.i.d.
- $W_{ik}W_{jk}/\mathscr{D}_{kk} \in [0, 1/\tau] \; \forall \; k$.

Applying Hoeffding's inequality, we get the required exponential bound on $\mathbb{P}\left((\Gamma \cap \Lambda)^c\right)$.

Bounding Eigenvectors

- The next step is to bound the eigenvectors of L and $\mathscr{L}$.
- Notation:
  - For symmetric matrix M, $\lambda(M)$ is the set of eigenvalues of M.
  - For a real interval $S \subset \mathbb{R}$, $\lambda_S(M) = \{\lambda(M) \cap S\}$

Bounding Eigenvectors (Contd...)

Tools of the Trade: Davis-Kahan Theorem

Let $S \subset \mathbb{R}$ be an interval. Denote $\mathscr{X}$ as an orthonormal matrix whose column space is the eigenspace of $\mathscr{L}\mathscr{L}$ corresponding to the eigenvalues in $\lambda_S(\mathscr{L}\mathscr{L})$. Denote by X the analogous matrix for LL. Define the distance between S and the spectrum of $\mathscr{L}\mathscr{L}$ outside of S as

$$\delta = \min\{|\ell - s|; \ell \in \lambda(\mathscr{L}\mathscr{L}), \ell \notin S, s \in S\}$$

If $\mathscr{X}$ and X are of the same dimension, then there is an orthonormal matrix O such that

$$\frac{1}{2}\|X - \mathscr{X}O\|_F^2 \leq \frac{\|LL - \mathscr{L}\mathscr{L}\|_F^2}{\delta^2}$$

Bounding Eigenvectors

Tools of the Trade: Weyl's Inequality

Define $\bar{\lambda}_1 \geq \ldots \geq \bar{\lambda}_n$ to be the elements of $\lambda(\mathscr{L}\mathscr{L})$ and $\lambda_1 \geq \ldots \geq \lambda_n$ to be the elements of $\lambda(LL)$. Then the eigenvalues of $\mathscr{L}\mathscr{L}$ and LL converge in the following sense.

$$\max_i |\lambda_i - \bar{\lambda}_i| \leq \|LL - \mathscr{L}\mathscr{L}\|_2$$

- Weyl's inequality bounds the eigenvalues of $\mathscr{L}\mathscr{L}$ and LL.
- Davis-Kahan theory bounds the eigenvectors of $\mathscr{L}\mathscr{L}$ and LL.

....and Then

- and then convergence of eigenvalues and eigenvectors...
- and then K-means...
- and then then result.

Bounding Eigenvectors - Main Theorem

Theorem 2: Convergence of Eigenvalues and Eigenvectors

Define sequences of intervals $S_n \in \mathbb{R}$ and $S'_n = \{\ell : \ell^2 \in S_n\}$. Define

$$\delta_n = inf\{|\ell - s|; \ell \in \lambda(\mathscr{L}^{(n)}\mathscr{L}^{(n)}), \ell \notin S_n, s \in S_n\}$$
$$\delta'_n = inf\{|\ell - s|; \ell \in \lambda_{S_n}(\mathscr{L}^{(n)}\mathscr{L}^{(n)}), s \notin S_n\}$$

Let k_n be the size of $\lambda_{S'_n}(L^{(n)})$ and $\mathscr{K}_n$ be the size of $\lambda_{S'_n}(\mathscr{L}^{(n)})$. Let $X_n \in \mathbb{R}^{n \times k_n}$ and $\mathscr{X}_n \in \mathbb{R}^{n \times \mathscr{K}_n}$ be the matrices whose orthonormal columns are eigenvectors corresponding to eigenvalues in $\lambda_{S'_n}(L^{(n)})$ and $\lambda_{S'_n}(\mathscr{L}^{(n)})$ respectively.

Bounding Eigenvectors - Main Theorem

Theorem 2: Convergence of Eigenvalues and Eigenvectors (Contd...)

Assumptions:

1. (Sparsity) $\tau_n^2 > 2/\log n$
2. (Eigen-gap) $n^{-1/2}(\log n)^2 = O(\min\{\delta_n, \delta_n'\})$

Then eventually, $k_n = \mathscr{K}_n$. Afterward,

$$\frac{1}{2}\|X_n - \mathscr{X}_n O_n\|_F = o\Big(\frac{\log n}{\delta_n \tau_n^2 n^{1/2}}\Big)$$

Bounding Eigenvectors - Main Theorem Proof

$$
\begin{aligned}
\max_i |\lambda_i^{(n)} - \bar{\lambda}_i^{(n)}| &\leq \|L^{(n)}L^{(n)} - \mathscr{L}^{(n)}\mathscr{L}^{(n)}\|_2 && \color{blue}{Weyl's\ Inequality} \\
&\leq \|L^{(n)}L^{(n)} - \mathscr{L}^{(n)}\mathscr{L}^{(n)}\|_F && \color{blue}{\|.\|_2 \leq \|.\|_F} \\
&= o\Big(\frac{\log n}{\tau_n^2 n^{1/2}}\Big) && \color{blue}{Main\ Theorem\ 1} \\
&= o\Big(n^{-1/2}(\log n)^2\Big) && \color{blue}{Assumption\ 1} \\
&= O(\min\{\delta_n, \delta_n'\}) && \color{blue}{Assumption\ 2}
\end{aligned}
$$

$$
\begin{aligned}
\frac{1}{2}\|X_n - \mathscr{X}_n O_n\|_F &\leq \frac{\|L^{(n)}L^{(n)} - \mathscr{L}^{(n)}\mathscr{L}^{(n)}\|_F^2}{\delta_n^2} && \color{blue}{Davis-Kahan} \\
&= o\Big(\frac{\log n}{\delta_n \tau_n^2 n^{1/2}}\Big) && \color{blue}{Main\ Theorem\ 1} \quad \square
\end{aligned}
$$

Bounding the k-means error

k-means

- **Input:** Data Points - $\{x_1, \ldots, x_n\} \in \mathbb{R}^k$ which are the n rows of the matrix $X \in \mathbb{R}^{n \times k}$
- **Output:** Centroids - $\{c_1, \ldots, c_k\} \in \mathbb{R}^k$ which are the k *unique* rows of the matrix $C \in \mathscr{R}^{n \times k}$ where $\mathscr{R}^{n \times k} = \{M \in \mathbb{R}^{n \times k} : M \textit{ has no more than } k \textit{ unique rows}\}$
- **Objective:**

$$\min_{\{m_1, \ldots, m_k\} \in \mathbb{R}^k} \sum_i \min_g \|x_i - m_g\|_2^2 = \min_{M \in \mathscr{R}^{n \times k}} \|X - M\|_F^2$$

Bounding the k-means error - Couple of Lemmas

Lemma 1

Consider SBM: $W = ZBZ^T \in \mathbb{R}^{n \times n}$ for $B \in \mathbb{R}^{k \times k}$ and $Z \in \{0, 1\}^{n \times k}$.

1. There exists $\mu \in \mathbb{R}^{k \times k}$ such that $Z\mu = \mathscr{X} \in \mathbb{R}^{n \times k}$ whose columns are eigenvectors of $\mathscr{L}$ corresponding to non-zero eigenvalues.
2. $z_i \mu = z_j \mu \Leftrightarrow z_i = z_j$ where z_i is the ith row of Z.

- Lemma 1 shows that applying k-means on the rows of $\mathscr{X} = Z\mu$ can reveal the block structure in the expected Laplacian $\mathscr{L}$.

Bounding the k-means error - Couple of Lemmas

Lemma 2

Define P to be the population of the largest block in Z.

$$P = \max_{j=1,\ldots,k} (Z^T Z)_{jj}$$

For the orthonormal matrix $O \in \mathbb{R}^{k \times k}$ in Theorem 2,

$$\|c_i - z_i \mu O\|_2 < 1/\sqrt{2P} \Rightarrow \|c_i - z_i \mu O\|_2 < \|c_i - z_j \mu O\|_2 \; for \; z_j \neq z_i.$$

- Lemma 2 lays down the sufficient condition for correct k-means clustering.
- Motivated by Lemma 2, we define the set of misclustered nodes as:

$$\mathscr{M} = \{i : \|c_i - z_i \mu O\|_2 \geq 1/\sqrt{2P}\}.$$

Bounding the k-means error - Main Theorem

Theorem 3: Bound on the misclustered nodes

Under the assumptions:

1. (Sparsity) $\tau_n^2 > 2/\log n$
2. (Eigen-gap) $n^{-1/2}(\log n)^2 = O(\lambda_{k_n}^2)$

The number of misclustered nodes is bounded by

$$|\mathscr{M}| = o\Big(\frac{P_n(\log n)^2}{\lambda_{k_n}^4 \tau_n^4 n}\Big) \textit{ almost surely.}$$

Bounding the k-means error - Main Theorem Proof

$$C = \underset{M \in \mathscr{R}^{n \times k}}{arg \min} \|X - M\|_F^2 \Rightarrow \|X - C\|_2 \leq \|X - Z\mu O\|_2 \tag{1}$$

$$\begin{aligned} \|C - Z\mu O\|_2 &\leq \|C - X\|_2 + \|X - Z\mu O\|_2 && \textit{Triangle Inequality} \\ &\leq 2\|X - Z\mu O\|_2 && \textit{Equation (1)} \end{aligned} \tag{2}$$

In Theorem 2, define $S_n = [\lambda_{k_n}^2/2, 1]$ and $\delta_n = \delta_n' = \lambda_{k_n}^2/2$. By assumption, $n^{-1/2}(\log n)^2 = O(\lambda_{k_n}^2) = O(\min(\delta_n, \delta_n'))$.

Bounding the k-means error - Main Theorem Proof (Contd...)

$$
\begin{aligned}
\therefore |\mathscr{M}| = \sum_{i \in \mathscr{M}} 1 &\leq 2P_n \sum_{i \in \mathscr{M}} \|c_i - z_i \mu O\|_2^2 \\
&\leq 2P_n \|C - Z\mu O\|_F^2 \\
&\leq 2P_n \|X - Z\mu O\|_F^2 && \textit{Equation (2)} \\
&= o\Big(\frac{P_n (\log n)^2}{\lambda_{k_n}^4 \tau_n^4 n}\Big) \textit{ almost surely.} \quad \square
\end{aligned}
$$

Consistency in Special Cases

The four-parameter Stochastic Block Model: $SBM(k, s, r, p)$

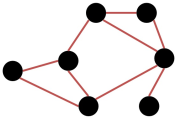
- k blocks each containing s nodes
- Probability of edge between nodes from same cluster is $r \in [0, 1]$ and from different clusters is $p + r \in [0, 1]$
- $B = p\mathbb{I}_{k \times k} + r\mathbb{1}\mathbb{1}^T$, $\lambda_k = 1/(k(r/p) + 1)$ and $P_n = n/k$.

Consistency under $SBM(k, s, r, p)$

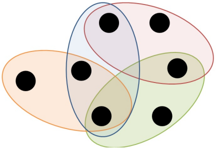
- $|\mathscr{M}| = o(k^3(\log n)^2)$ almost surely.
- For $k = O(n^{1/4}/\log n)$, $\frac{|\mathscr{M}|}{n} = o(n^{-1/4})$ almost surely.

Hypergraphs

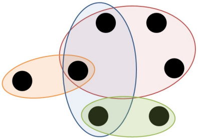
- Collection of sets / Generalization of graphs
- Each edge can connect more than two nodes



Graph
(2-uniform)



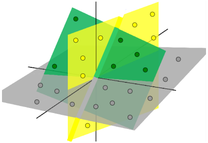
3-uniform
hypergraph



Hypergraph

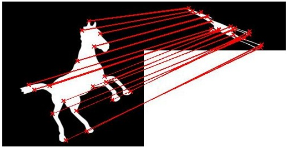
Hypergraphs in Computer Vision

Subspace clustering



Motion segmentation

Matching / Image Registration



Involves 3-way / 4-way similarities (uniform hypergraph)

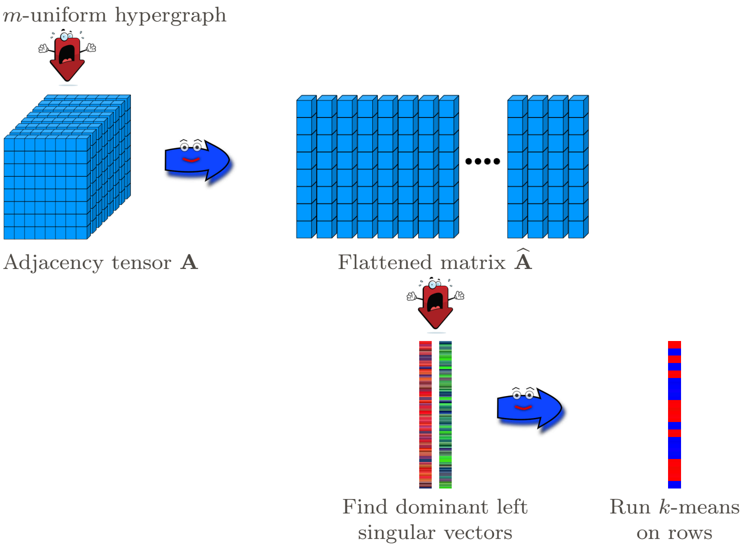
Hypergraph Partitioning Methods

- Partitioning circuits [Schweikert & Kernighan '79]
- Graph approximation for hypergraphs [Hadley '95]
- Spectral hypergraph partitioning [Zien et al. '99]
- hMETIS for VLSI design [Karypis & Kumar '00]

- Uniform hypergraph in databases [Gibson et al. '00]
- Uniform hypergraph in vision [Agarwal et al. '05]
- Tensor based algorithms [Govindu '05; Chen & Lerman '09]

- Learning with non-uniform hypergraph [Zhou et al. '07]
- Higher order learning [Duchenne et al. '11; Rota Bulo & Pellilo '13; etc.]

Spectral Uniform Hypergraph Partitioning†



† (Govindu 2005)

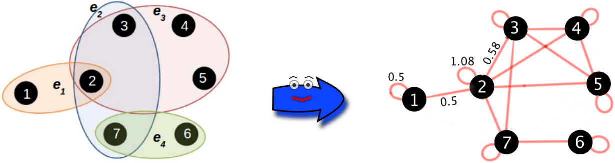
Normalized Hypergraph Cut

Approach: [Zhou, Huang & Schölkopf '07]

- Solve spectral relaxation of minimizing normalized hypergraph cut

Reduction to graph:

- $A, D \in \mathbb{R}^{n \times n}$ so that $A_{ij} = \sum_{e \ni i,j} \frac{1}{|e|}$, $D_{ii} = \text{degree}(i)$



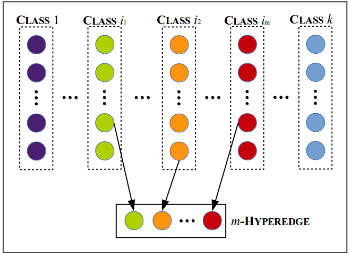
Spectral clustering:

- Normalized Laplacian, $L = I - D^{-1/2} A D^{-1/2}$
- Compute k leading orthonormal eigenvectors of L
- k-means on normalized rows of eigenvector matrix

Planted Partition Model (non-uniform hypergraph)[‡]

Model:

- Given n nodes, and k (hidden) classes
- Maximum edge cardinality M
- Unknown m^{th}-order tensors $B^{(m)} \in [0,1]^{k\times k\times\ldots\times k}$
- Unknown sparsity factors $\alpha_{m,n}$, $m = 2, 3, \ldots, M$
- Independent edges with label-dependent distribution



$$\text{Prob}(m\text{-edge}) = \alpha_{m,n} B^{(m)}_{i_1 i_2 \ldots i_m}$$

[‡]Ghoshdastidar & Dukkipati (2017), Annals of Statistics

Consistency of NH-Cut[§]

Define:

- $\mathcal{A} = \mathsf{E}[A]$, $\mathcal{D} = \mathsf{E}[D]$ and $\mathcal{L} = I - \mathcal{D}^{-1/2}\mathcal{A}\mathcal{D}^{-1/2}$
- $d = \min_i \mathsf{E}[\text{degree}(i)]$
- $\delta = k^{th}$ eigen-gap of $\mathcal{L}$

Theorem

There exists constant $C > 0$, such that, if

$$\delta > 0 \quad \text{and} \quad d > C\frac{kn_{\max}(\log n)^2}{n_{\min}\delta^2}$$

then with probability $(1 - o(1))$

$$\text{Error}(\psi, \psi') = O\left(\frac{kn_{\max}\log n}{\delta^2 d}\right) = o(n).$$

[§] Ghoshdastidar & Dukkipati (2017), Annals of Statistics

Concluding Remarks

Spectral approaches offer:

- nice approximations for problems of community detection in networks with
- theoretical guarantees (Still lot to do!) to establish which one would indulge in
  - results from numerical linear algebra (Davis-Kahan theorems),
  - concentration inequalities from random matrix theory.

Must read:

- Speilman's lecture notes on spectral graph theory
- Luxburg's review on spectral clustering

Acknowledgements: Some of the figures in this presentation have been borrowed from Debarghya.