Certifying the Global Optimality of Graph Cuts via Semidefinite Programming: A Theoretic Guarantee for Spectral Clustering

Shuyang Ling

Courant Institute of Mathematical Sciences, NYU

Aug 13, 2018

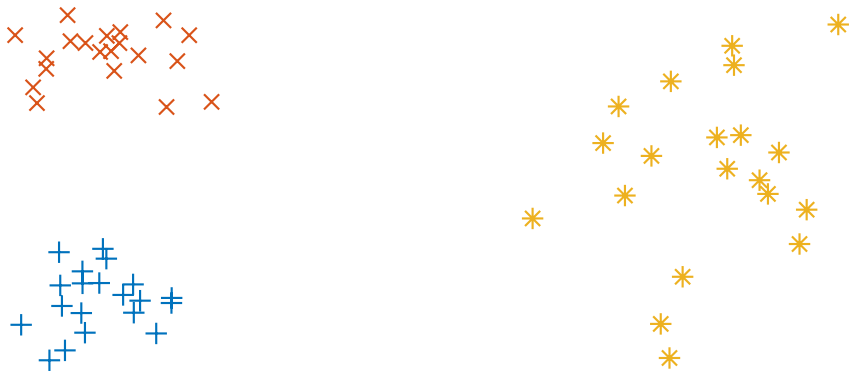
Joint work with Prof. Thomas Strohmer at UC Davis

Outline

- Motivation: data clustering
- K-means and spectral clustering
- A graph cut perspective of spectral clustering
- Convex relaxation of ratio cuts and normalized cuts
- Theory and applications

Data clustering and unsupervised learning

Question: Given a set of N data points in $\mathbb{R}^d$, how to partition them into k clusters based on the similarity?



K-means clustering

K-means

Cluster the data by minimizing the k-means objective function:

$$\min_{\{\Gamma_l\}_{l=1}^k} \underbrace{\sum_{l=1}^{k} \sum_{i \in \Gamma_l} \left\| x_i - \overbrace{\frac{1}{|\Gamma_l|} \sum_{i \in \Gamma_l} x_i}^{\text{centroid}} \right\|^2}_{\text{within-cluster sum of squares}}$$

where $\{\Gamma_l\}_{l=1}^k$ is a partition of $\{1, \cdots, N\}$.

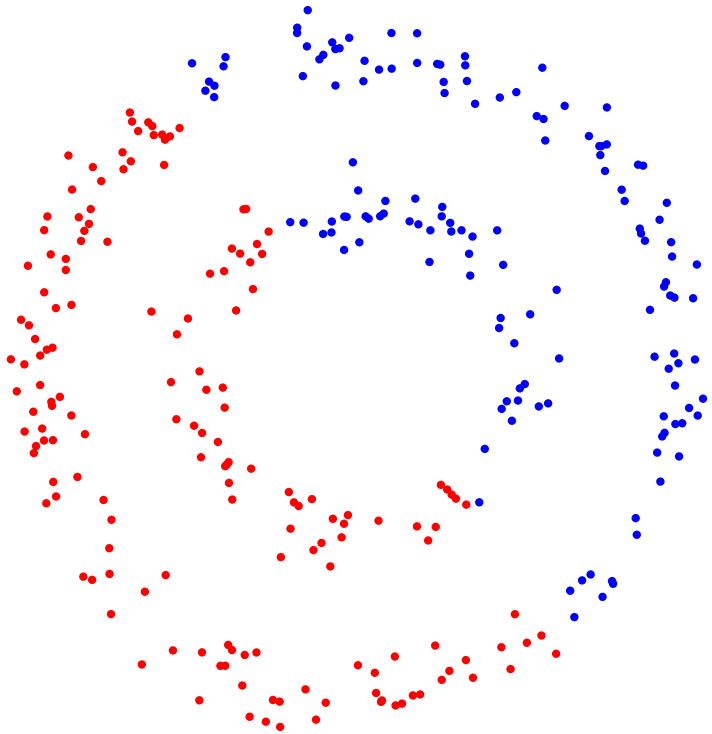
- Widely used in vector quantization, unsupervised learning, Voronoi tessellation, etc.
- An NP-hard problem, even if $d = 2$. [Mahajan, etc 09]
- Heuristic method: Lloyd's algorithm [Lloyd 82]

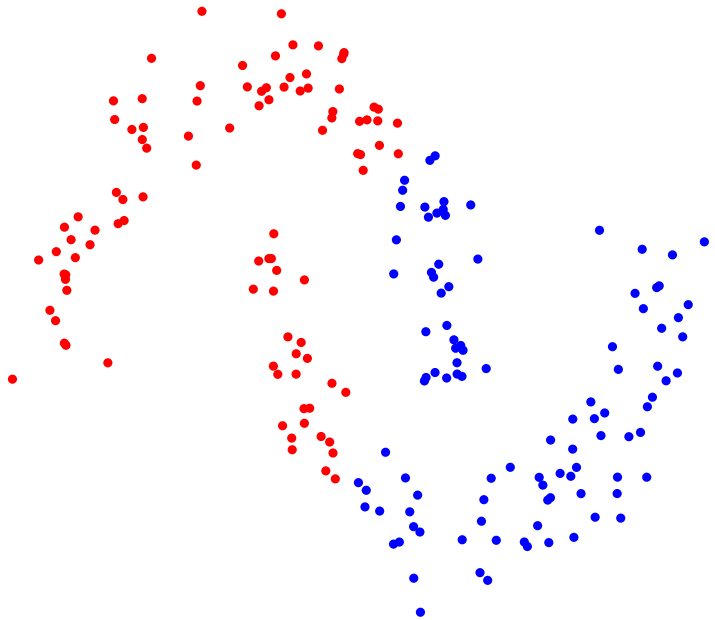
Limitation of k-means

Limitation of k-means

K-means only works for datasets with individual clusters:

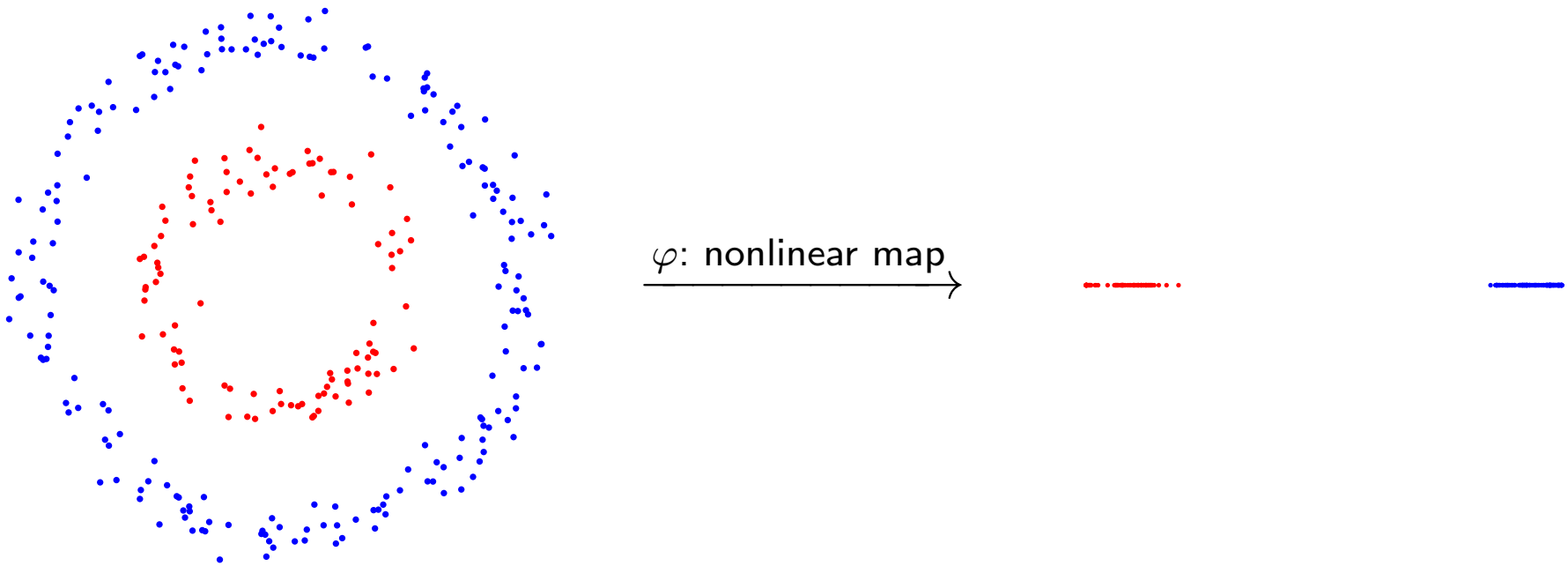
- isotropic and within convex boundaries
- well-separated





Kernel k-means and nonlinear embedding

Goal: map the data into a feature space so that they are well-separated and k-means would work.



How: locally-linear embedding, isomap, multidimensional scaling, Laplacian eigenmaps, diffusion maps, etc.

Focus: We will focus on Laplacian eigenmaps. Spectral clustering consists of Laplacian eigenmaps followed by k-means clustering.

Graph Laplacian

Suppose $\{x_i\}_{i=1}^N \in \mathbb{R}^d$ and construct a similarity (weight) matrix W via

$$w_{ij} := \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right), \quad W \in \mathbb{R}^{N\times N},$$

where σ controls the size of neighborhood. In fact, W represents a weighted undirected graph.

Definition of graph Laplacian

The (unnormalized) graph Laplacian associated to W is

$$L = D - W$$

where

$$D = \text{diag}(W1_N)$$

is the degree matrix.

Properties of graph Laplacian

The (unnormalized) graph Laplacian associated to W is

$$L = D - W, \qquad D = \mathrm{diag}(W 1_N).$$

Properties

- L is positive semidefinite,

$$z^\top L z = \sum_{i<j} w_{ij}(z_i - z_j)^2.$$

- 1_N is in the null space of L, i.e., $\lambda_1(L) = 0$.
- $\lambda_2(L) > 0$ if and only if the graph is connected.
- The dimension of null space equals the number of connected components.

Laplacian eigenmaps and k-means

Laplacian eigenmaps

For the graph Laplacian L, we let the Laplacian eigenmap be

$$\begin{bmatrix} \varphi(x_1) \\ \vdots \\ \varphi(x_N) \end{bmatrix} := \underbrace{[u_1, \cdots, u_k]}_{U} \in \mathbb{R}^{N \times k}$$

where $\{u_l\}_{l=1}^k$ are the eigenvectors w.r.t. the k *smallest* eigenvalues.

In other words, φ maps data in $\mathbb{R}^d$ to $\mathbb{R}^k$, a coordinate in terms of eigenvectors:

$$\varphi : \underbrace{x_i}_{\mathbb{R}^d} \longrightarrow \underbrace{\varphi(x_i)}_{\mathbb{R}^k}.$$

Then we apply k-means to $\{\varphi(x_i)\}_{i=1}^N$ to perform clustering.

Algorithm of spectral clustering based on graph Laplacian

Unnormalized spectral clustering[1]

- **Input:** Given the number of clusters k and a dataset $\{x_i\}_{i=1}^N$, construct the similarity matrix W from $\{x_i\}_{i=1}^N$.
- Compute the unnormalized graph Laplacian

$$L = D - W$$

- Compute the eigenvectors $\{u_l\}_{l=1}^k$ of L w.r.t. the smallest k eigenvalues.
- Let $U = [u_1, u_2, \cdots, u_k] \in \mathbb{R}^{N \times k}$. Perform k-means clustering on the rows of U by using Lloyd's algorithm.
- Obtain the partition based on the outcome of k-means.

[1] For more details, see an excellent review by [Von Luxburg, 2007]

Another variant of spectral clustering

Normalized spectral clustering

- **Input:** Given the number of clusters k and a dataset $\{x_i\}_{i=1}^N$, construct the similarity matrix W from $\{x_i\}_{i=1}^N$.
- Compute the normalized graph Laplacian

$$L_{\text{sym}} = I_N - D^{-\frac{1}{2}} W D^{-\frac{1}{2}} = D^{-\frac{1}{2}} L D^{-\frac{1}{2}}$$

- Compute the eigenvectors $\{u_l\}_{l=1}^k$ of L_{sym} w.r.t. the smallest k eigenvalues.
- Let $U = [u_1, u_2, \cdots, u_k] \in \mathbb{R}^{N \times k}$. Perform k-means clustering on the rows of $D^{-\frac{1}{2}} U$ by using Lloyd's algorithm.
- Obtain the partition based on the outcome of k-means.

Comments on spectral clustering

Pros and Cons of spectral clustering

Pros:

- Spectral clustering enjoys high popularity and conveniently applies to various settings.
- Rich connections to random walk on graph, spectral graph theory, and differential geometry.

Cons:

- Rigorous justifications of spectral clustering are still lacking.
- The two-step procedures complicate the analysis, e.g. how to analyze the performance of Laplacian eigenmaps and the convergence analysis of k-means?

Our goal: we take a different route by looking at the **convex relaxation** of spectral clustering to understand its performance better.

A graph cut perspective

Key observation:

- The matrix W is viewed as a weight matrix of a graph with N vertices.
- Partitioning the dataset into k clusters is equivalent to finding a k-way graph cut such that any pair of induced subgraphs is not well-connected.

Graph cut

The cut is defined as the weight sum of edges whose two ends are in different subsets,

$$\text{cut}(\Gamma, \Gamma^c) := \sum_{i \in \Gamma, j \in \Gamma^c} w_{ij}$$

where Γ is a subset of vertices and Γ^c is its complement.

A graph cut perspective

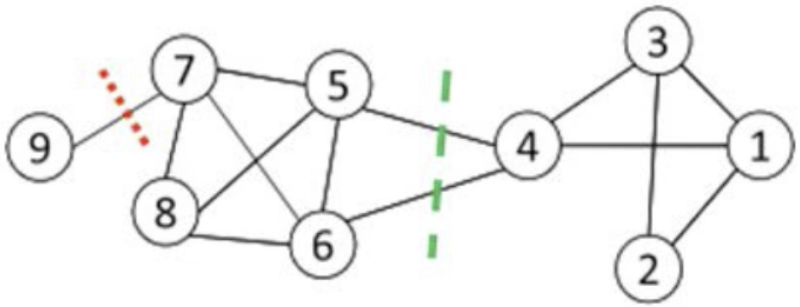
Graph cut

The cut is defined as the weight sum of edges whose two ends are in different subsets,

$$\text{cut}(\Gamma, \Gamma^c) := \sum_{i \in \Gamma, j \in \Gamma^c} w_{ij}$$

where Γ is a subset of vertices and Γ^c is its complement.

However, minimizing $\text{cut}(\Gamma, \Gamma^c)$ may not lead to satisfactory results since it is more likely to get an imbalanced cut.



RatioCut

RatioCut

The ratio cut of $\{\Gamma_a\}_{a=1}^k$ is given by

$$\text{RatioCut}(\{\Gamma_a\}_{a=1}^k) = \sum_{a=1}^{k} \frac{\text{cut}(\Gamma_a, \Gamma_a^c)}{|\Gamma_a|}.$$

In particular, if $k = 2$,

$$\text{RatioCut}(\Gamma, \Gamma^c) = \frac{\text{cut}(\Gamma, \Gamma^c)}{|\Gamma|} + \frac{\text{cut}(\Gamma, \Gamma^c)}{|\Gamma^c|}.$$

- But, it is worth noting minimizing RatioCut is NP-hard.
- A possible solution is to relax!!

RatioCut and graph Laplacian

Let $1_{\Gamma_a}(\cdot)$ be an indicator vector which maps a vertex to a vector in $\mathbb{R}^N$ via

$$1_{\Gamma_a}(I) = \begin{cases} 1, & I \in \Gamma_a, \\ 0, & I \notin \Gamma_a. \end{cases}$$

Relating RatioCut to graph Laplacian

There holds

$$\mathrm{cut}(\Gamma_a, \Gamma_a^c) = \left\langle L,\ 1_{\Gamma_a} 1_{\Gamma_a}^\top \right\rangle = 1_{\Gamma_a}^\top L 1_{\Gamma_a}$$

$$\mathrm{RatioCut}(\{\Gamma_a\}_{a=1}^k) = \sum_{a=1}^k \frac{1}{|\Gamma_a|} \left\langle L,\ 1_{\Gamma_a} 1_{\Gamma_a}^\top \right\rangle = \langle L,\ X_{\mathrm{rcut}} \rangle ,$$

where

$$X_{\mathrm{rcut}} := \sum_{a=1}^k \frac{1}{|\Gamma_a|} 1_{\Gamma_a} 1_{\Gamma_a}^\top \ \longleftarrow\ \text{a block-diagonal matrix}$$

Spectral relaxation of RatioCut

Spectral clustering is a relaxation by these two properties,

$$X_{\text{rcut}} = UU^\top, \quad U^\top U = I_k, \quad U \in \mathbb{R}^{N\times k}.$$

Spectral relaxation of RatioCut

Substituting $X_{\text{rcut}} = UU^\top$ results in

$$\min_{U\in\mathbb{R}^{N\times k}} \left\langle L, UU^\top \right\rangle \quad \text{s.t.} \quad U^\top U = I_k,$$

whose global minimizer is easily found via computing the eigenvectors w.r.t. the k smallest eigenvalues of the graph Laplacian L.

The spectral relaxation gives exactly the first step of unnormalized spectral clustering.

Normalized cut

For normalized spectral clustering, we consider

$$\mathrm{NCut}(\{\Gamma_a\}_{a=1}^k) := \sum_{a=1}^k \frac{\mathrm{cut}(\Gamma_a, \Gamma_a^c)}{\mathrm{vol}(\Gamma_a)}$$

where $\mathrm{vol}(\Gamma_a) = 1_{\Gamma_a}^\top D 1_{\Gamma_a}$. Therefore,

$$\mathrm{NCut}(\{\Gamma_a\}_{a=1}^k) = \langle L_{\mathrm{sym}}, X_{\mathrm{ncut}} \rangle .$$

Here $L_{\mathrm{sym}} = D^{-\frac{1}{2}} L D^{-\frac{1}{2}}$ is the normalized Laplacian and

$$X_{\mathrm{ncut}} := \sum_{a=1}^k \frac{1}{1_{\Gamma_a}^\top D 1_{\Gamma_a}} D^{\frac{1}{2}} 1_{\Gamma_a} 1_{\Gamma_a}^\top D^{\frac{1}{2}}.$$

By relaxing $X_{\mathrm{ncut}} = UU^\top$, it gives the spectral relaxation of normalized graph Laplacian.

Performance bound via matrix perturbation argument

Let $W^{(a,a)}$ be the weight matrix of the a-th cluster and define

$$W_{\text{iso}} := \begin{bmatrix} W^{(1,1)} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & W^{(k,k)} \end{bmatrix}, \quad W_\delta = W - W_{\text{iso}}$$

Decomposition of W

We decompose the original graph into two subgraphs:

$$W = \overbrace{W_{\text{iso}}}^{\text{diagonal blocks}} + \overbrace{W_\delta}^{\text{off-diagonal}}$$

$$\underbrace{\qquad\qquad}_{k \text{ connected components}} \qquad \underbrace{\qquad\qquad}_{k\text{-partite graph}}$$

The corresponding degree matrix

$$D_{\text{iso}} := \text{diag}(W_{\text{iso}} 1_N), \quad D_\delta := \text{diag}(W_\delta 1_N)$$

where D_{iso}: inner-cluster degree matrix; D_δ: outer-cluster degree matrix.

Spectral clustering

The unnormalized graph Laplacian associated to W_{iso} is

$$L_{\text{iso}} := D_{\text{iso}} - W_{\text{iso}} \longleftarrow \text{a block-diagonal matrix.}$$

There holds $\lambda_l(L_{\text{iso}}) = 0, 1 \le l \le k, \quad \lambda_{k+1}(L_{\text{iso}}) = \min \lambda_2(L_{\text{iso}}^{(a,a)}) > 0.$

What happens if the graph has k connected components?

The nullspace of L_{iso} is spanned by k indicator vectors in $\mathbb{R}^N$, i.e., the columns of U_{iso},

$$U_{\text{iso}} := \begin{bmatrix} \frac{1}{\sqrt{n_1}} 1_{n_1} & 0 & \cdots & 0 \\ 0 & \frac{1}{\sqrt{n_2}} 1_{n_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\sqrt{n_k}} 1_{n_k} \end{bmatrix} \in \mathbb{R}^{N \times k}, \quad U_{\text{iso}}^\top U_{\text{iso}} = I_k.$$

However, this is not the case if the graph is fully connected.

Matrix perturbation argument

Hope: If L is close to L_{iso}, so is U to U_{iso}.

Davis-Kahan $\sin\theta$ theorem

Davis-Kahan perturbation:

$$\min_{R \in O(3)} \|U - U_{\text{iso}} R\| \leq \frac{\sqrt{2}\|L - L_{\text{iso}}\|}{\lambda_{k+1}(L_{\text{iso}})}$$

where

$$\|L - L_{\text{iso}}\| \leq 2\|D_\delta\|, \quad \lambda_{k+1}(L_{\text{iso}}) = \min \lambda_2(L_{\text{iso}}^{(a,a)}) > 0.$$

In other words, if $\|D_\delta\|$ is small, i.e., the difference between L and L_{iso} is small, and $\lambda_{k+1}(L_{\text{iso}}) > 0$, the eigenspace U should be close to U_{iso}. But it does not imply the underlying membership.

Convex relaxation of ratio cuts

Let's review the RatioCut,

$$\text{RatioCut}(\{\Gamma_a\}_{a=1}^k) = \sum_{a=1}^k \frac{1}{|\Gamma_a|} \left\langle L, 1_{|\Gamma_a|} 1_{|\Gamma_a|}^\top \right\rangle = \langle L, X_{\text{rcut}} \rangle ,$$

where $X_{\text{rcut}} := \sum_{a=1}^k \frac{1}{|\Gamma_a|} 1_{\Gamma_a} 1_{\Gamma_a}^\top$.

Question: What constraints does X_{rcut} satisfy for any given $\{\Gamma_a\}_{a=1}^k$?

Convex sets

Given a partition $\{\Gamma_a\}_{a=1}^k$, the corresponding X_{rcut} satisfies

- X_{rcut} is positive semidefinite, $X_{\text{rcut}} \succeq 0$;
- X_{rcut} is nonnegative, $X_{\text{rcut}} \geq 0$ entrywisely;
- the constant vector is an eigenvector of Z: $X_{\text{rcut}} 1_N = 1_N$;
- the trace of X_{rcut} equals k, i.e., $\text{Tr}(X_{\text{rcut}}) = k$.

Convex relaxation of normalized cuts

RatioCut-SDP - SDP relaxation of RatioCut

We relax the originally combinatorial optimization by

$$\min \langle L, Z \rangle \ \text{ s.t. } \ Z \succeq 0, \ Z \geq 0, \ \mathrm{Tr}(Z) = k, \ Z 1_N = 1_N.$$

The major difference from spectral relaxation is the *nonnegativity* constraint.

NCut-SDP - SDP relaxation of normalized cut

The corresponding convex relaxation of normalized cut[a] is

$$\min \langle L_{\mathrm{sym}}, Z \rangle \ \text{ s.t. } \ Z \succeq 0, \ Z \geq 0, \ \mathrm{Tr}(Z) = k, \ Z D^{\frac{1}{2}} 1_N = D^{\frac{1}{2}} 1_N.$$

[a] A similar relaxation was proposed by [Xing, Jordan, 2003]

Question: How well do these two convex relaxations work?

Intuitions

From the plain perturbation argument,

$$\min_{R \in O(3)} \|U - U_{\text{iso}} R\| \leq \frac{2\sqrt{2}\|D_\delta\|}{\lambda_{k+1}(L_{\text{iso}})}$$

where $\lambda_{k+1}(L_{\text{iso}}) = \min \lambda_2(L_{\text{iso}}^{(a,a)}) > 0$ if each cluster is connected.

What the perturbation argument tells us:

The success of spectral clustering depends on two ingredients:

- Within-cluster connectivity: $\lambda_2(L_{\text{iso}}^{(a,a)})$), a.k.a. algebraic connectivity or Fiedler eigenvalue, or equivalently, $\lambda_{k+1}(L_{\text{iso}})$.
- Inter-cluster connectivity: the operator norm of D_δ quantifies the maximal outer-cluster degree.

Main theorem for RatioCut-SDP

Theorem (Ling, Strohmer, 2018)

The SDP relaxation gives X_{rcut} as the unique global minimizer if

$$\|D_\delta\| < \frac{\lambda_{k+1}(L_{\text{iso}})}{4},$$

where $\lambda_{k+1}(L_{\text{iso}})$ is the $(k+1)$-th smallest eigenvalue of the graph Laplacian L_{iso}.

Here $\lambda_{k+1}(L_{\text{iso}})$ satisfies

$$\lambda_{k+1}(L_{\text{iso}}) = \min_{1\le a\le k} \lambda_2(L_{\text{iso}}^{(a,a)})$$

where $\lambda_2(L_{\text{iso}}^{(a,a)})$ is the second smallest eigenvalue of graph Laplacian w.r.t. the a-th cluster.

A random walk perspective of normalized cut

For normalized cut, we first consider the Markov transition matrices on the whole dataset and each individual cluster:

$$P := D^{-1}W, \qquad P_{\text{iso}}^{(a,a)} = (D^{(a,a)})^{-1}W^{(a,a)}.$$

Two factors

- Within-cluster connectivity: the smaller the second largest eigenvalue[a] of $P_{\text{iso}}^{(a,a)}$ is, the stronger connectivity of a-th cluster is.
- Inter-cluster connectivity: let P_δ be the off-diagonal parts of P. $\|P_\delta\|_\infty$ is the largest probability of a random walker moving out of its own cluster after one step.

[a] This eigenvalue governs the mixing time of Markov chain.

Main theorem for NCut-SDP

Theorem (Ling, Strohmer, 2018)

The SDP gives X_{ncut} as the unique global minimizer if

$$\frac{\|P_\delta\|_\infty}{1-\|P_\delta\|_\infty} < \frac{\min \lambda_2(I_{n_a} - P_{\text{iso}}^{(a,a)})}{4}$$

- Note that

$$\frac{\|P_\delta\|_\infty}{1-\|P_\delta\|_\infty}$$

  is small if a random walker starting from *any* node is *more* likely to stay in its own cluster than leave it after one step, and vice versa.

- $\min \lambda_2(I_{n_a} - P_{\text{iso}}^{(a,a)})$ is the smallest eigengap of Markov transition matrix defined across all the clusters.

Main contribution of our result

Our contribution

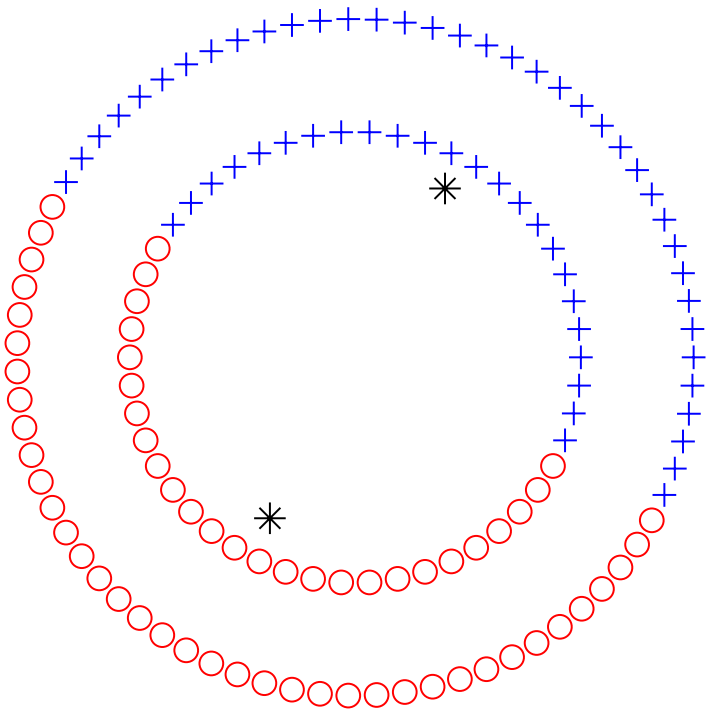
- Purely deterministic. Exact recovery under natural and mild condition. No assumptions needed on the random data generative model.
- General and applicable to various settings.
- Easily verifiable criteria for the global optimality of a graph cut under ratio cut and normalized cut.
- Not only applies to spectral clustering but also to community detection under stochastic block model.

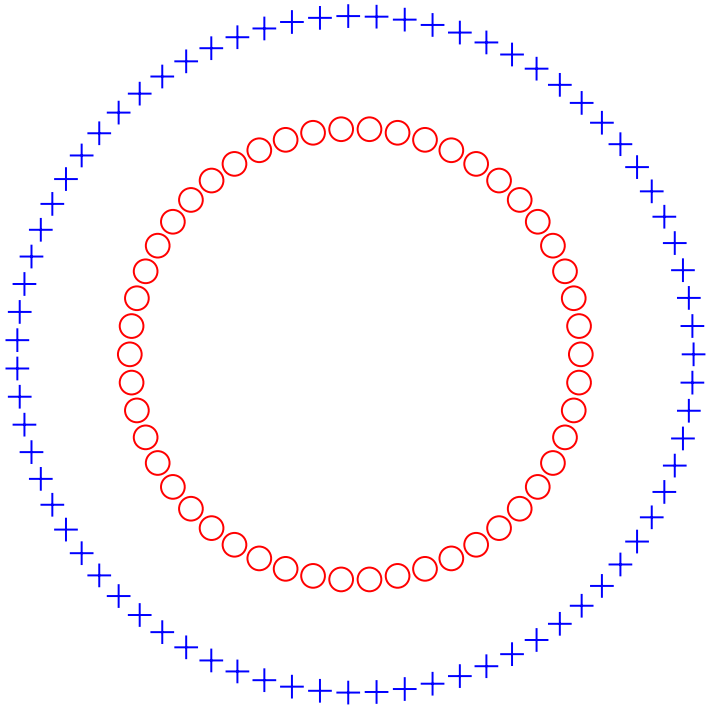
Example 1: two concentric circles

We consider

$$x_{1,i} = \begin{bmatrix} \cos(\frac{2\pi i}{n}) \\ \sin(\frac{2\pi i}{n}) \end{bmatrix}, \ 1 \leq i \leq n; \qquad x_{2,j} = \kappa \begin{bmatrix} \cos(\frac{2\pi j}{m}) \\ \sin(\frac{2\pi j}{m}) \end{bmatrix}, \ 1 \leq j \leq m$$

where $m \approx \kappa n$ and $\kappa > 1$.





Example 1: two concentric circles

Theorem

Let the data be given by $\{x_{1,i}\}_{i=1}^n \cup \{x_{2,i}\}_{i=1}^m$. *The bandwidth* σ *is chosen as*

$$\sigma = \frac{4\gamma}{n\sqrt{\log(\frac{m}{2\pi})}}.$$

Then SDP recovers the underlying two clusters exactly if

$$\underbrace{\kappa - 1}_{\text{minimal separation}} \geq \mathcal{O}\left(\frac{\gamma}{n}\right).$$

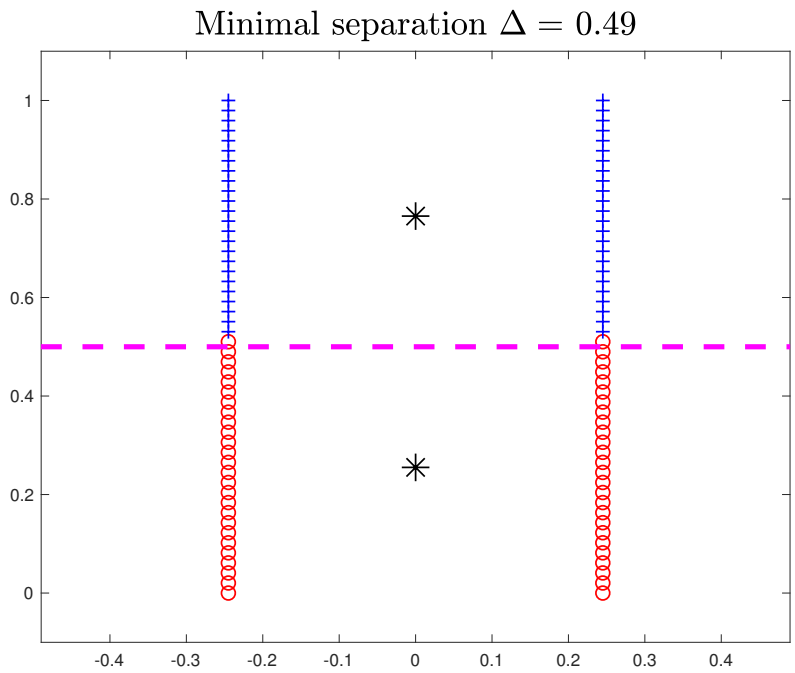
The bound is near-optimal since two adjacent points on one each is $\mathcal{O}(n^{-1})$ apart. One cannot hope to recover the two clusters if the minimal separation is smaller than $\mathcal{O}(n^{-1})$.

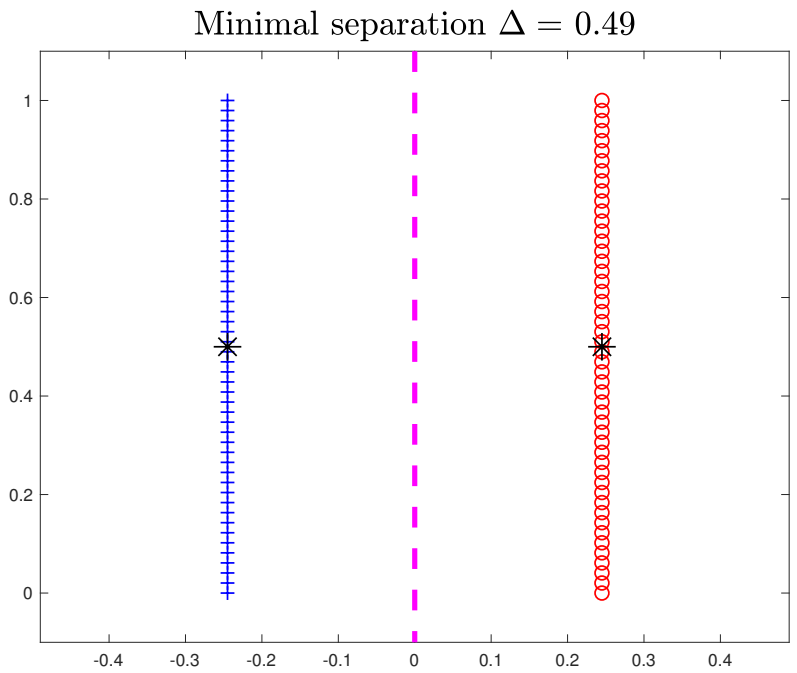
Example 2: two parallel lines

Suppose the data points are distributed on two lines with separation Δ,

$$x_{1,i} = \begin{bmatrix} -\frac{\Delta}{2} \\ \frac{i-1}{n-1} \end{bmatrix}, \quad x_{2,i} = \begin{bmatrix} \frac{\Delta}{2} \\ \frac{i-1}{n-1} \end{bmatrix}, \quad 1 \le i \le n.$$

Which one partition is given by k-means objective function?





Each cluster is highly anisotropic.

Example 2: two parallel lines

K-means does not work if $\Delta < 0.5$.

Theorem

Let the data be given by $\{x_{1,i}\}_{i=1}^n \cup \{x_{2,i}\}_{i=1}^n$. *Use the heat kernel with bandwidth*

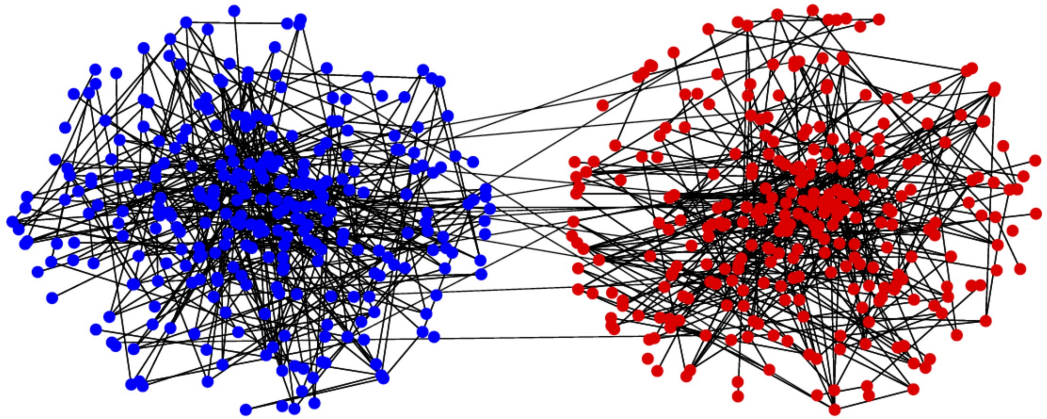
$$\sigma = \frac{\gamma}{(n-1)\sqrt{\log(\frac{n}{\pi})}}, \qquad \gamma > 0.$$

Assume the separation Δ *satisfies*

$$\Delta \geq \mathcal{O}\left(\frac{\gamma}{n}\right).$$

Then SDP recovers the underlying two clusters exactly.

Example 3: community detection under stochastic block model



The adjacency matrix is a binary random matrix:

1. if member i and j are in the same community, $\Pr(w_{ij} = 1) = p$ and $\Pr(w_{ij} = 0) = 1 - p$;
2. if member i and j are in different communities, $\Pr(w_{ij} = 1) = q$ and $\Pr(w_{ij} = 0) = 1 - q$.

Let $p > q$ and we are interested in when SDP relaxation is able to recover the underlying membership.

Graph cuts and stochastic block model

Theorem

Let $p = \frac{\alpha \log n}{n}$ and $q = \frac{\beta \log n}{n}$. The RatioCut-SDP recovers the underlying communities exactly if

$$\alpha > 26\left(\frac{2}{3} + \frac{\beta}{2} + \sqrt{\beta}\right)$$

with high probability.

Compared with the state-of-the-art results where

$$\alpha > 2 + \beta + 2\sqrt{2\beta}$$

is needed for exact recovery[2], our performance bound is slightly looser by a constant factor.

[2] [Abbe, Bandeira, Hall, 2016]

Numerics - random instances

So far, the two toy examples are purely deterministic and it is of course more interesting to look at random data. However, this task is nontrivial which leads to a few open problems.

Therefore we turn to simulation to see how many random instances satisfy the conditions in our theorem and how they depends on the choice of σ and minimal separation Δ.

- RatioCut-SDP:

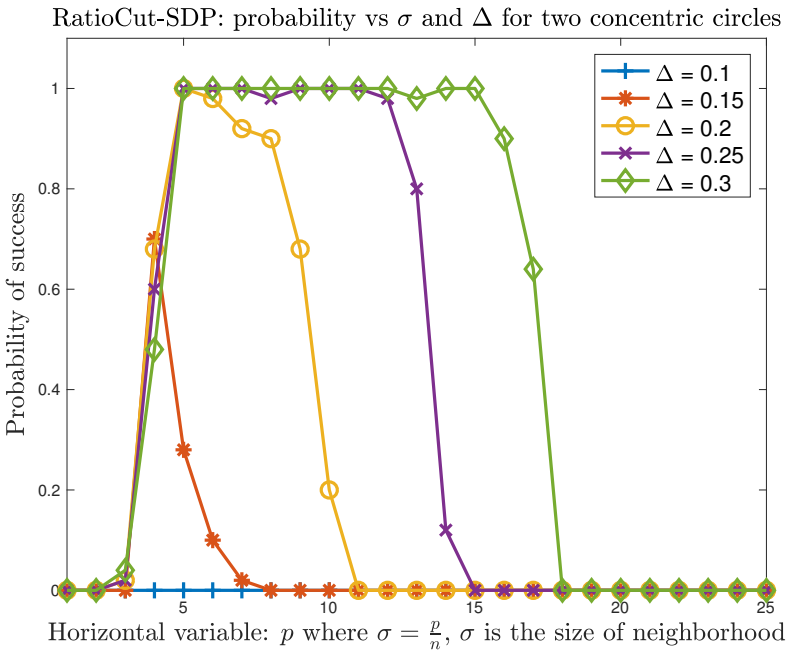
$$\|D_\delta\| \leq \frac{\lambda_{k+1}(L_{\text{iso}})}{4},$$

- NCut-SDP:

$$\frac{\|P_\delta\|_\infty}{1-\|P_\delta\|_\infty} \leq \frac{\min \lambda_2(I_{n_a} - P_{\text{iso}}^{(a,a)})}{4}.$$

Numerics for two concentric circles

Exact recovery is guaranteed empirically if $\Delta \geq 0.2$, $5 \leq p \leq 75\Delta - 8$, and $\sigma = n^{-1}p$.



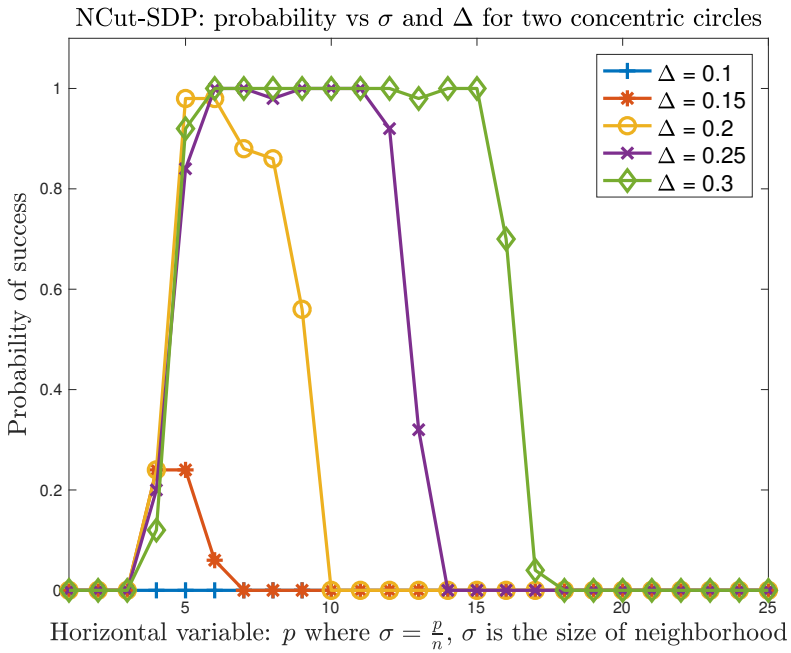
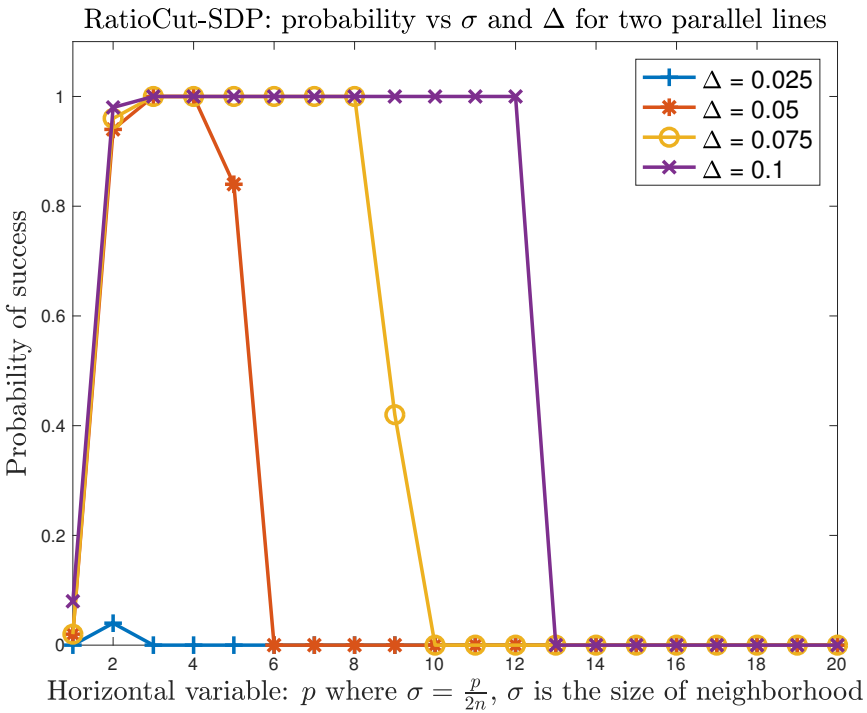


Figure: Two concentric circles with radii $r_1 = 1$ and $r_2 = 1 + \Delta$. The smaller circle has $n = 250$ uniformly distributed points and the larger one has $m = \lfloor 250(1 + \Delta) \rfloor$ where $\Delta = r_2 - r_1$.

Numerics for two lines

Exact recovery is guaranteed empirically if $\Delta \geq 0.05$ and $2 \leq p \leq 150\Delta - 4$.



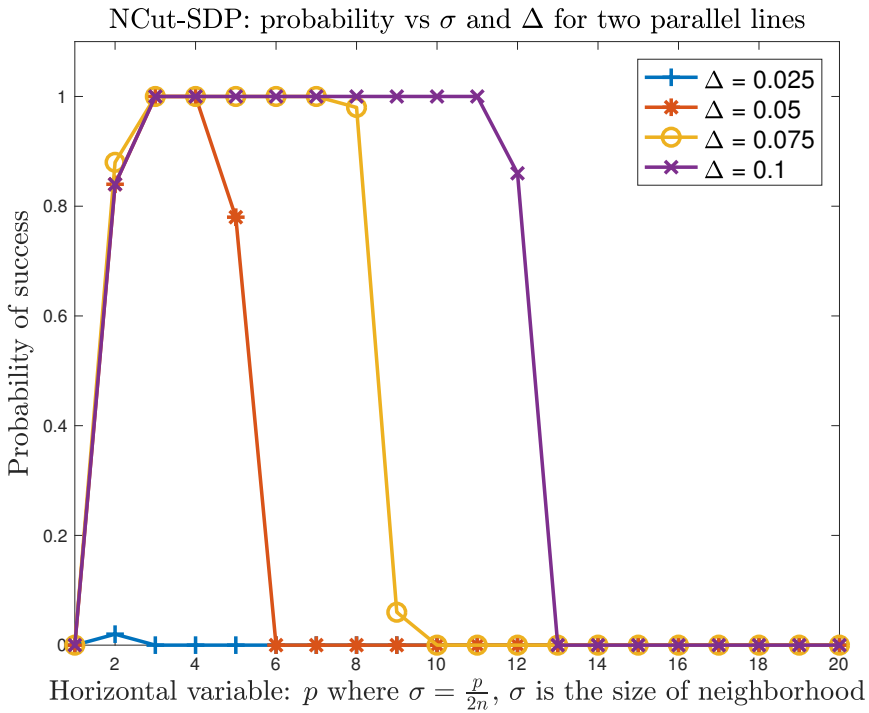
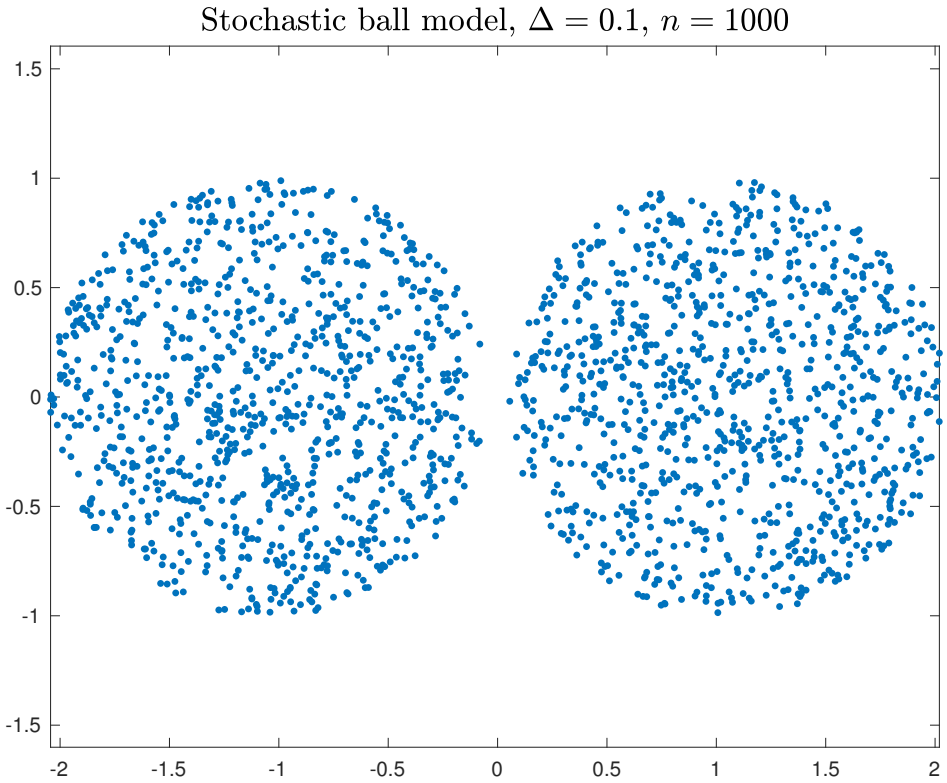


Figure: Two parallel line segments of unit length with separation Δ and 250 points are sampled uniformly on each line.

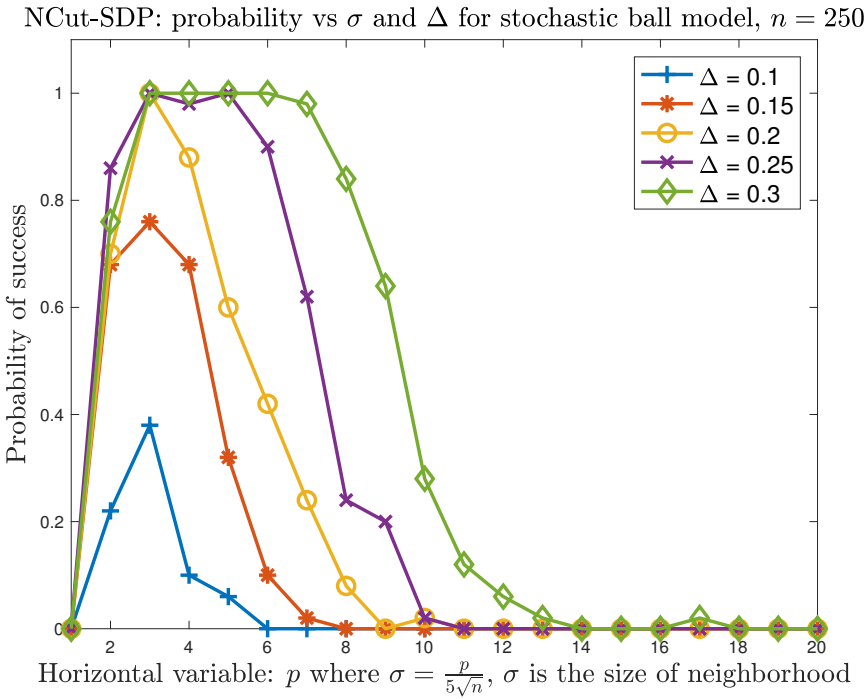
Numerics - stochastic ball model

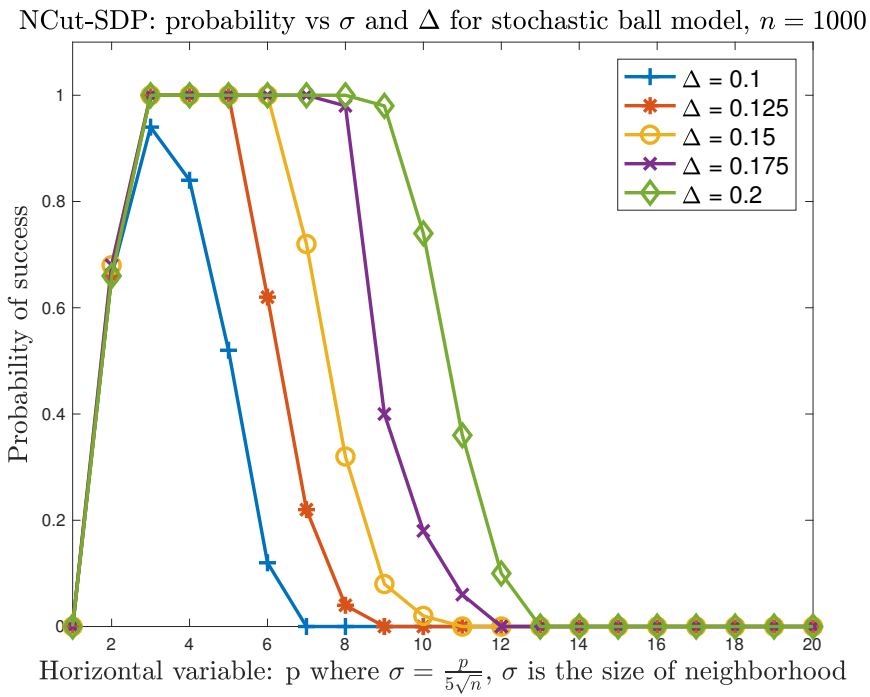


Stochastic ball model with two clusters and the separation between the centers is $2 + \Delta$. Each cluster has 1000 uniformly distributed points

Numerics: stochastic ball model

1. if $n = 250$, we require $\Delta \geq 0.2$, $\sigma = \frac{p}{5\sqrt{n}}$ and $2 \leq p \leq 40\Delta - 4$;
2. if $n = 1000$, we require $\Delta \geq 0.1$, $\sigma = \frac{p}{5\sqrt{n}}$, and $2 \leq p \leq 60\Delta - 2$.





Performance of NCut-SDP for 2D stochastic ball model. Left: each ball has $n = 250$ points; Right: each ball contains $n = 1000$ points.

Comparison with k-means

For stochastic ball model, k-means should be the best choice since each cluster is isotropic and separated. However, we have shown that exact recovery via k-means SDP is impossible[3] if

$$\Delta \leq \sqrt{\frac{3}{2}} - 1 \approx 0.2247.$$

On the other hand, the SDP relaxation of RatioCut and NCut still works

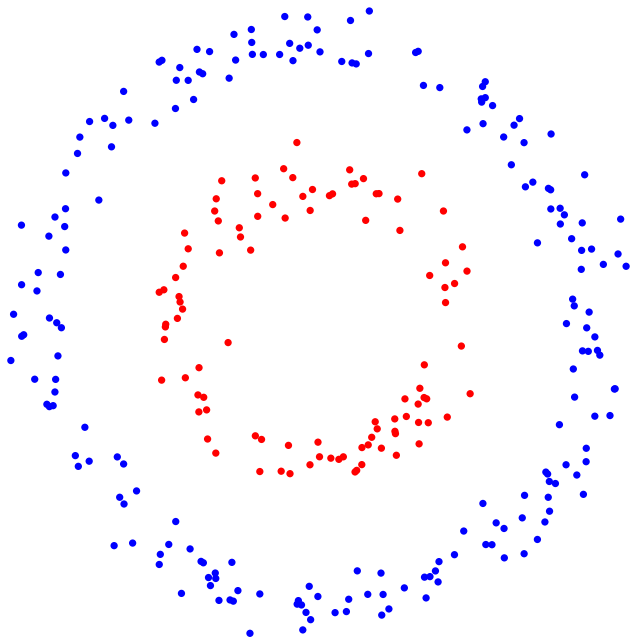
1. if $n = 250$, we require $\Delta \geq 0.2$, $\sigma = \frac{p}{5\sqrt{n}}$ and $2 \leq p \leq 40\Delta - 4$;
2. if $n = 1000$, we require $\Delta \geq 0.1$, $\sigma = \frac{p}{5\sqrt{n}}$, and $2 \leq p \leq 60\Delta - 2$.

[3] See arXiv:1710.06008, by Li, etc.

Outlook for random data model

Open problem

Suppose there are n data points drawn from a probability density function $p(x)$ supported on a manifold $\mathcal{M}$. How can we estimate the second smallest eigenvalue of the graph Laplacian (either normalized or unnormalized) given the kernel function Φ and σ?



A possible solution to the open problem

- It is well known that the graph Laplacian will converge to the Laplace-Beltrami operator on the manifold (pointwisely and spectral convergence).
- In other words, if we know the second smallest eigenvalue of Laplace-Beltrami operator manifold and convergence rate from graph Laplacian to its continuum limit, we can have a lower bound of the second smallest eigenvalue of graph Laplacian.
- It require tools from empirical process, differential geometry, ...

Conclusion and future works

We establish a systematic framework to certify the global optimality of a graph cut under ratio cut and normalized cut. The performance guarantee is purely algebraic and deterministic, only depending on the algebraic properties of graph Laplacian. It may lead to a novel way to look at spectral graph theory.

- How to estimate the Fiedler eigenvalue of graph Laplacian associated to a random data set?
- Can we derive an analogue theory for directed graph?
- Find a faster and scalable solver for SDP and how to analyze the classical spectral clustering algorithm?
- **Preprint:** Certifying global optimality of graph cuts via semidefinite relaxation: a performance guarantee for spectral clustering, *arXiv:1806.11429*.

Thank you!