A Tutorial on Spectral Clustering
Part 2: Advanced/related Topics

Chris Ding

Computational Research Division
Lawrence Berkeley National Laboratory
University of California

Advanced/related Topics

- Spectral embedding: simplex cluster structure
- Perturbation analysis
- K-means clustering in embedded space
- Equivalence of K-means clustering and PCA
- Connectivity networks: scaled PCA & Green's function
- Extension to bipartite graphs: Correspondence analysis
- Random talks and spectral clustering
- Semi-definite programming and spectral clustering
- Spectral ordering (distance-sensitive ordering)
- Webpage spectral ranking: Page-Rank and HITS

Spectral Embedding: Simplex Cluster Structure

- Compute K eigenvectors of the Laplacian.
- Embed objects in the K-dim eigenspace

What is the structure of the clusters?

Simplex Embedding Theorem.

Assume objects are well characterized by spectral clustering objective functions. In the embedded space, objects aggregate to K distinct centroids:

- Centroids locate on K corners of a simplex
  - Simplex consists K basis vectors + coordinate origin
  - Simplex is rotated by an orthogonal transformation T
  - Columns of T are eigenvectors of a $K \times K$ embedding matrix Γ

(Ding, 2004)

3

K-way Clustering Objectives

$$J = \sum_{1 \le p < q \le K} \frac{s(C_p, C_q)}{\rho(C_p)} + \frac{s(C_p, C_q)}{\rho(C_q)} = \sum_k \frac{s(C_p, G - C_q)}{\rho(C_p)}$$

$$\rho(C_k) = \begin{cases} n_k = |C_k| & \textbf{for Ratio Cut} \\ d(C_k) = \sum_{i \in C_k} d_i & \textbf{for Normalized Cut} \\ s(C_k, C_k) = \sum_{i \in C_k, j \in C_k} w_{ij} & \textbf{for MinMaxCut} \end{cases}$$

G - C_k is the graph complement of C_k

4

Simplex Spectral Embedding Theorem

Simplex Orthogonal Transform Matrix $T = (\mathbf{t}_1, \cdots, \mathbf{t}_K)$

T are determined by: $\Gamma \mathbf{t}_k = \lambda_k \mathbf{t}_k$

Spectral Perturbation Matrix $\Gamma = \Omega^{-\frac{1}{2}} \overline{\Gamma} \Omega^{-\frac{1}{2}}$

$$\overline{\Gamma} = \begin{bmatrix} h_{11} & -s_{12} & \cdots & -s_{1K} \\ -s_{21} & h_{22} & \cdots & -s_{2K} \\ \vdots & \vdots & \cdots & \vdots \\ -s_{K1} & -s_{K2} & \cdots & h_{KK} \end{bmatrix}$$

$$s_{pq} = s(C_p, C_q)$$

$$h_{kk} = \sum_{p|p \neq k} s_{kp}$$

$$\Omega = \mathbf{diag}[\rho(C_1), \cdots, \rho(C_k)]$$

Properties of Spectral Embedding

- Original basis vectors:

$$h_k = (0 \cdots 0, \overbrace{1 \cdots 1}^{n_k}, 0 \cdots 0) / n_k$$

- Dimension of embedding is K-1: (q_2, ... q_K)
  - $q_1 = (1, \ldots, 1)^T$ is constant & trivial
  - Eigenvalues of Γ (=eigenvalues of D-W)
  - Eigenvalues determine how well clustering objective function characterize the data
- Exact solution for K=2

2-way Spectral Embedding (Exact Solution)

Eigenvalues

$$\lambda_{\mathrm{Rcut}} = \frac{s(C_1,C_2)}{n(C_1)} + \frac{s(C_1,C_2)}{n(C_2)},\ \lambda_{\mathrm{Ncut}} = \frac{s(C_1,C_2)}{d(C_1)} + \frac{s(C_1,C_2)}{d(C_2)},\ \lambda_{\mathrm{MMC}} = \frac{s(C_1,C_2)}{s(C_1,C_1)} + \frac{s(C_1,C_2)}{s(C_2,C_2)}$$

Recover the original 2-way clustering objectives

For Normalized Cut, orthogonal transform T rotates

$$h_1 = (1\cdots 1,0\cdots 0)^T\ ,\quad h_2 = (0\cdots 0,1\cdots 1)^T$$

into

$$q_1 = (1\cdots 1)^T,\quad q_2 = (a,\cdots,a,-b,\cdots,-b)^T$$

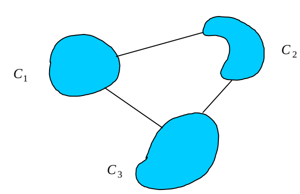
Spectral clustering inherently consistent! (Ding et al, KDD'01)

Perturbation Analysis

$$Wq = \lambda Dq \qquad \hat{W}z = (D^{-1/2}WD^{-1/2})z = \lambda z \quad q = D^{-1/2}z$$

Assume data has 3 dense clusters sparsely connected.

$$W = \begin{bmatrix} W_{11} & W_{12} & W_{13} \\ W_{21} & W_{22} & W_{23} \\ W_{31} & W_{32} & W_{33} \end{bmatrix}$$



Off-diagonal blocks are between-cluster connections, assumed small and are treated as a perturbation

(Ding et al, KDD'01)

Perturbation Analysis

0th order:

$$W^{(0)} = \begin{bmatrix} W_{11} & & \\ & W_{22} & \\ & & W_{33} \end{bmatrix} \qquad \hat{W}^{(0)} = \begin{bmatrix} \hat{W}_{11}^{(0)} & & \\ & \hat{W}_{22}^{(0)} & \\ & & \hat{W}_{33}^{(0)} \end{bmatrix}$$

1st order:

$$W^{(1)} = \begin{bmatrix} & W_{12} & W_{13} \\ W_{21} & & W_{23} \\ W_{31} & W_{32} & \end{bmatrix} \qquad \hat{W}^{(1)} = \begin{bmatrix} \hat{W}_{11} - \hat{W}_{11}^{(0)} & \hat{W}_{12} & \hat{W}_{13} \\ \hat{W}_{21} & \hat{W}_{22} - \hat{W}_{22}^{(0)} & \hat{W}_{23} \\ \hat{W}_{31} & \hat{W}_{32} & \hat{W}_{33} - \hat{W}_{33}^{(0)} \end{bmatrix}$$

$$\hat{W}_{pq}^{(0)} = D_{pp}^{-1/2} W_{pq} D_{qq}^{-1/2}$$

$$\hat{W}_{pq} = (D_{p1} + D_{p2} + D_{p3})^{-1/2} W_{pq} (D_{q1} + D_{q2} + D_{q3})^{-1/2}$$

K-means clustering

- Developed in 1960's (Lloyd, MacQueen, etc)
- Computationally Efficient (order-*mN*)
- Widely used in practice
  – Benchmark to evaluate other algorithms

Given *n* points in *m*-dim: $X = (x_1, x_2, \cdots, x_n)$

K-means $\min J_K = \sum_{k=1}^{K} \sum_{i \in C_k} || x_i - c_k ||^2$

K-means Clustering in Spectral Embedded Space

Simplex spectral embedding theorem provides theoretical basis for K-means clustering in the embedded eigenspace

- Cluster centroids are well separated (corners of the simplex)
- *K*-means clustering is invariant under (i) coordinate rotation $x \rightarrow Tx$, and (ii) shift $x \rightarrow x + a$
- Thus orthogonal transform T in simplex embedding unnecessary

• Many variants of K-means (Ng et al, Bach & Jordan, Zha et al, Shi & Xu, etc)

We have proved

Spectral embedding + *K*-means clustering is the appropriate method

We now show :

K-means itself is solved by PCA

Equivalence of *K*-means Clustering and Principal Component Analysis

- Cluster indicators specify the solution of *K*-means clustering
- Principal components are eigenvectors of the Gram (Kernel) matrix = data projections in the principal directions of the covariance matrix
- Optimal solution of K-means clustering: continuous solution of the discrete cluster indicators of *K*-means are given by Principal components

(Zha et al, NIPS'01; Ding & He, 2003)

13

Principal Component Analysis (PCA)

- Widely used in large number of different fields
  - Best low-rank approximation (SVD Theorem, Eckart-Young, 1930) : Noise reduction
  - Unsupervised dimension reduction
  - Many generalizations
- Conventional perspective is inadequate to explain the effectiveness of PCA
- New results: Principal components are cluster indicators for well-motivated clustering objective

14

Principal Component Analysis

n points in m-dim: $X = (x_1, x_2, \cdots, x_n)$

Principal directions: u_k

Covariance $S = XX^T$ $\quad XX^T u_k = \lambda_k u_k$

Principal components: v_k

Gram (Kernel) matrix $X^T X$ $\quad X^T X v_k = \lambda_k v_k$

Singular Value Decomposition: $X = \sum_{k=1}^{m} \lambda_k u_k v_k^T$

15

2-way K-means Clustering

Cluster membership indicator:

$$q(i) = \begin{cases} +\sqrt{n_2 / n_1 n} & \text{if } i \in C_1 \\ -\sqrt{n_1 / n_2 n} & \text{if } i \in C_2 \end{cases}$$

$$J_K = n\langle x^2 \rangle - J_D, \quad J_D = \frac{n_1 n_2}{n}\left[2\frac{d(C_1,C_2)}{n_1 n_2} - \frac{d(C_1,C_1)}{n_1^2} - \frac{d(C_2,C_2)}{n_2^2} \right]$$

Define distance matrix: $D = (d_{ij}),\ d_{ij} = |x_i - x_j|^2$

$$J_D = -q^T D q = -q^T \tilde{D} q = 2 q^T X^T X q$$

$\tilde{D}$ is the centered distance matrix

$$\min J_K \Rightarrow \max J_D$$

16

2-way K-means Clustering

Cluster indicator satisfy: $\sum_i q(i) = 0, \sum_i q^2(i) = 1$

Relax the restriction q (i) take discrete values. Let it take continuous values in [-1,1]. Solution for q is the eigenvector of the Gram matrix.

Theorem: The (continuous) optimal solution of q is given by the principal component v_1 .

Clusters C_1, C_2 are determined by:

$$C_1 = \{i \mid v_1(i) < 0\}, C_2 = \{i \mid v_1(i) \geq 0\}$$

Once C_1, C_2 are computed, iterate ***K***-mean to convergence

17

Multi-way K-means Clustering

Unsigned Cluster membership indicators $h_1, \cdots, h_K$:

$$\begin{array}{ccc} C_1 & C_2 & C_3 \end{array}$$

$$\begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = (h_1, h_2, h_3)$$

18

Multi-way K-means Clustering

For $K \geq 2$, $J_K = \sum_i x_i^2 - \sum_{k=1}^{K} \frac{1}{n_k} \sum_{i,j \in C_k} x_i^T x_j$

(Unsigned) Cluster membership indicators $h_1, \cdots, h_K$:

$$h_k = (0 \cdots 0, \overbrace{1 \cdots 1}^{n_k}, 0 \cdots 0)^T / n_k$$

$$J_K = \sum_i x_i^2 - \sum_{k=1}^{K} h_k^T X^T X h_k$$

Let $H = (h_1, \cdots, h_K)$

$$J_K = \sum_i x_i^2 - \mathrm{Tr}(H_k^T X^T X H_k)$$

19

Multi-way K-means Clustering

Regularized Relaxation of *K*-means Clustering

Redundancy in $h_1, \cdots, h_K$: $\sum_{k=1}^{K} n_k^{1/2} h_k = e = (11 \cdots 1)^T$

Transform to signed indicator vectors q_1 - q_k via the k x k orthogonal matrix T:

$$(q_1, ..., q_k) = (h_1, \cdots, h_k) T \qquad Q_k = H_k T$$

Require 1st column of T = $(n_1^{1/2}, \cdots, n_k^{1/2})^T / n^{1/2}$

Thus $q_1 = e / n^{1/2} = \text{const}$

20

Regularized Relaxation of *K*-means Clustering

$$J_K = Tr(Y^T Y) - Tr(Q_{k-1}^T Y^T Y Q_{k-1})$$

(Regularized relaxation) $$Q_{k-1} = (q_2, \ldots, q_k)$$

Theorem: The optimal solutions of $q_2 \cdots q_k$ are given by the principal components $v_2 \cdots v_k$. J_K is bounded below by total variance minus sum of K eigenvalues of covariance:

$$n\overline{y^2} - \sum_{k=1}^{K-1} \lambda_k < \min J_K < n\overline{y^2}$$

Scaled PCA

similarity matrix $S=(s_{ij})$ (generated from XX^T)

$$D = \mathrm{diag}(d_1, \cdots, d_n) \qquad d_i = s_{i.}$$

Nonlinear re-scaling: $\tilde{S} = D^{-\frac{1}{2}} S D^{-\frac{1}{2}}, \tilde{s}_{ij} = s_{ij} / (s_{i.} s_{j.})^{1/2}$

Apply SVD on $\tilde{S} \Rightarrow$

$$S = D^{\frac{1}{2}} \tilde{S} D^{\frac{1}{2}} = D^{\frac{1}{2}} \sum_k z_k \lambda_k z_k^T D^{\frac{1}{2}} = D \left[\sum_k q_k \lambda_k q_k^T \right] D$$

$q_k = D^{-1/2} z_k$ is the scaled principal component

Subtract trivial component $\lambda_0 = 1,\ z_0 = d^{1/2}/s..,\ q_0 = 1$

$$\Rightarrow \quad S - dd^T/s.. = D \sum_{k=1} q_k \lambda_k q_k^T D$$

(Ding, et al, 2002)

Optimality Properties of Scaled PCA

Scaled principal components have optimality properties:

Ordering

– Adjacent objects along the order are similar
– Far-away objects along the order are dissimilar
– Optimal solution for the permutation index are given by scaled PCA.

Clustering

– Maximize within-cluster similarity
– Minimize between-cluster similarity
– Optimal solution for cluster membership indicators given by scaled PCA.

Difficulty of *K*-way clustering

- 2-way clustering uses a single eigenvector
- *K*-way clustering uses several eigenvectors
- How to recover 0-1 cluster indicators *H*?

eigenvectors : $Q = (q_1, ..., q_k)$

has both positive and negative entries

indicators : $H = (h_1, \cdots, h_k)$ $\quad Q = HT$

Avoid computing the transformation *T*:

- Do *K*-means, which is invariant under *T*
- Compute connectivity network QQ^T, which cancels *T*

Connectivity Network

$$C_{ij} = \begin{cases} 1 & \text{if } i, j \text{ belong to same cluster} \\ 0 & \text{otherwise} \end{cases}$$

SPCA provides $C \cong D \sum_{k=1}^{K} q_k \lambda_k q_k^T D$

Green's function : $C \approx G = \sum_{k=2}^{K} q_k \frac{1}{1-\lambda_k} q_k^T$

Projection matrix: $C \approx P \equiv \sum_{k=1}^{K} q_k q_k^T$

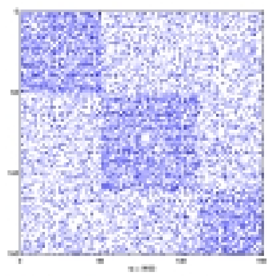
(Ding et al, 2002)

Connectivity network

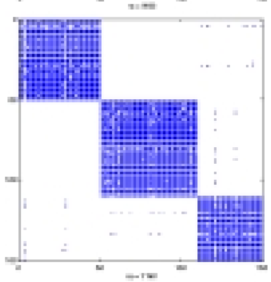
- Similar to Hopfield network
- Mathematical basis: projection matrix
- Show self-aggregation clearly
- Drawback: how to recover clusters
  - Apply *K*-means directly on *C*
  - Use linearized assignment with cluster crossing and spectral ordering (ICML'04)

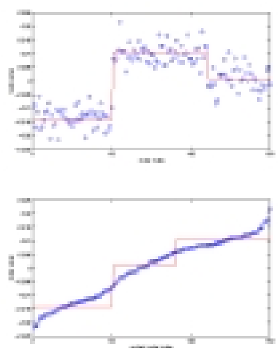
Connectivity network: Example 1

Similarity matrix W



Connectivity matrix





$\lambda_2 = 0.300, \bar{\lambda}_2 = 0.268$

Between-cluster connections suppressed

Within-cluster connections enhanced

Effects of self-aggregation

Connectivity of Internet Newsgroups

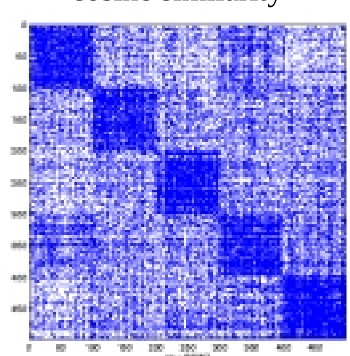
NG2: comp.graphics
NG9: rec.motorcycles
NG10: rec.sport.baseball
NG15:sci.space
NG18:talk.politics.mideast

100 articles from each group.
1000 words
Tf.idf weight. Cosine similarity

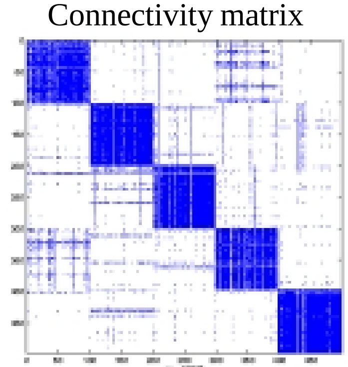
Spectral Clustering 89%

Direct *K*-means 66%

cosine similarity

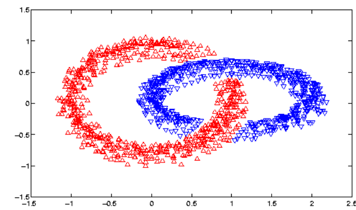


Connectivity matrix

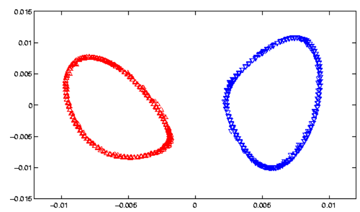


Spectral embedding is not topology preserving

700 3-D data points form 2 interlock rings



In eigenspace, they shrink and separate



Correspondence Analysis (CA)

- Mainly used in graphical display of data
- Popular in France (Benzécri, 1969)
- Long history
  - Simultaneous row and column regression (Hirschfeld, 1935)
  - Reciprocal averaging (Richardson & Kuder, 1933; Horst, 1935; Fisher, 1940; Hill, 1974)
  - Canonical correlations, dual scaling, etc.
- Formulation is a bit complicated ("convoluted" Jolliffe, 2002, p.342)
- "A neglected method", (Hill, 1974)

Scaled PCA on a Contingency Table ⇒ Correspondence Analysis

Nonlinear re-scaling: $\tilde{P} = D_r^{-\frac{1}{2}} P D_c^{-\frac{1}{2}}, \ \tilde{p}_{ij} = p_{ij} / (p_{i.} p_{j.})^{1/2}$

Apply SVD on $\tilde{P}$ Subtract trivial component

$$P - rc^T / p_{..} = D_r \sum_{k=1} f_k \lambda_k g_k^T D_c$$

$$r = (p_{1.}, \cdots, p_{n.})^T$$

$$c = (p_{.1}, \cdots, p_{.n})^T$$

$$f_k = D_r^{-\frac{1}{2}} u_k, \ g_k = D_c^{-\frac{1}{2}} v_k$$

are the scaled row and column principal component (standard coordinates in CA)

31

Tutorial on Spectral Clustering, ICML 2004, Chris Ding © University of California

Information Retrieval

Bell Lab tech memos

5 comp-sci and 4 applied-math memo titles:

C1: Human machine interface for lab ABC computer applications

C2: A survey of user opinion of computer system response time

C3: The EPS user interface management system

C4: System and human system engineering testing of EPS

C5: Relation of user-perceived response time to error management

M1: The generation of random, binary, unordered trees

M2: The intersection graph of paths in trees

M3: Graph minors IV: widths of trees and well-quasi-ordering

M4: Graph minors: A survey

32

Tutorial on Spectral Clustering, ICML 2004, Chris Ding © University of California

Word-document matrix: row/col clustering

words \ docs	c4	c1	c3	c5	c2	m4	m3	m2	m1
human	1	1							
EPS	1		1						
interface		1	1						
system	2		1		1				
computer					1				
user				1	1				
response				1	1				
time				1	1				
survey					1	1			
minors						1	1		
graph						1	1	1	
tree							1	1	1

Bipartite Graph: 3 types of Connectivity networks

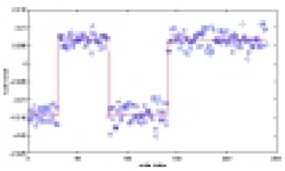
$$Q_K Q_K^T = \begin{bmatrix} F_K F_K^T & F_K G_K^T \\ G_K F_K^T & G_K G_K^T \end{bmatrix} \qquad Q_K = (\mathbf{q}_1, \cdots, \mathbf{q}_K) = \begin{bmatrix} F_K \\ G_K \end{bmatrix}$$

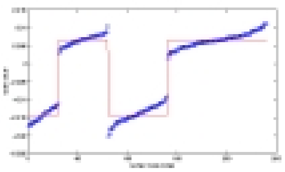
$$F_K F_K^T = \begin{bmatrix} \mathbf{e}_{r_1}\mathbf{e}_{r_1}^T / 2s_{11} & 0 \\ 0 & \mathbf{e}_{r_2}\mathbf{e}_{r_2}^T / 2s_{22} \end{bmatrix}$$ row-row clustering

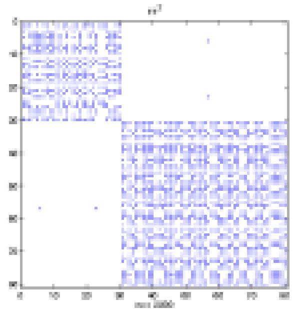
$$G_K G_K^T = \begin{bmatrix} \mathbf{e}_{c_1}\mathbf{e}_{c_1}^T / 2s_{11} & 0 \\ 0 & \mathbf{e}_{c_2}\mathbf{e}_{c_2}^T / 2s_{22} \end{bmatrix}$$ Column-column clustering

$$F_K G_K^T = \begin{bmatrix} \mathbf{e}_{r_1}\mathbf{e}_{c_1}^T / 2s_{11} & 0 \\ 0 & \mathbf{e}_{r_2}\mathbf{e}_{c_2}^T / 2s_{22} \end{bmatrix}$$ Row-column association

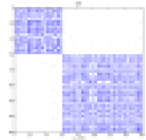
Example



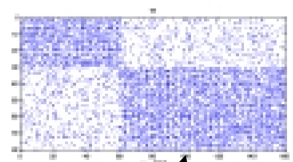




column-column: GG^T

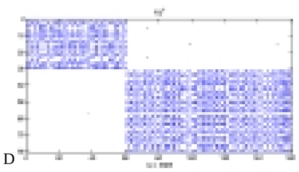


row-row: FF^T



Original data matrix

$\lambda_2 = 0.456, \bar{\lambda}_2 = 0.477$

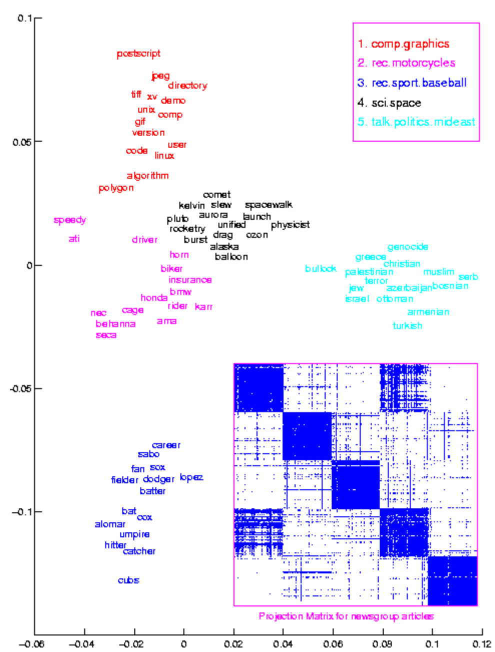


row-column: FG^T

35

Internet Newsgroups

Simultaneous clustering of documents and words



Random Walks and Normalized Cut

Similarity matrix W, $\quad P = D^{-1}W$ Stochastic matrix

$\pi^T P = \pi^T \quad \Rightarrow$ equilibrium distribution: $\quad \pi = d$

$$Px = \lambda x \;\Rightarrow\; Wx = \lambda Dx \;\Rightarrow\; (D-W)x = (1-\lambda)Dx$$

Random walks between *A,B:*

$$J_{NormCut} = \frac{P(A \to B)}{\pi(A)} + \frac{P(B \to A)}{\pi(B)}$$

(Meila & Shi, 2001)

PageRank: $\quad P = \alpha L D_{out}^{-1} + (1-\alpha)ee^T$

37

Semi-definite Programming for Normalized Cut

Normalized Cut :

$$y_k = D^{1/2}(0\cdots0,\overbrace{1\cdots1}^{n_k},0\cdots0)^T / \| D^{1/2}h_k \|$$

$$\tilde{W} = D^{-1/2}WD^{-1/2}$$

Optimize : $\min\limits_{Y} \mathbf{Tr}(Y^T(I-\tilde{W})Y)$, **subject to** $\; Y^TY = I$

$\Rightarrow \mathrm{Tr}[(I-\tilde{W})YY^T] \qquad \Rightarrow \min\limits_{Z} \mathrm{Tr}[(I-\tilde{W})Z] \qquad \boxed{Z = YY^T}$

$$\boxed{s.t.\, Z \ge 0,\;\; Z \succ 0, Zd = d, \mathrm{Tr}\, Z = K,\;\; Z = Z^T}$$

Compute Z via SDP. $Z=Y'Y'^T$.
$Y''=D^{-1/2}Y'$. K-means on Y''.

(Xing & Jordan, 2003)

Z = connectivity network

38

Spectral Ordering

Hill, 1970, spectral embedding

Find coordinate x to minimize

$$J = \sum_{ij} (x_i - x_j)^2 w_{ij} = x^T (D - W) x$$

Solution are eigenvectors of Laplacian

Barnard, Pothen, Simon, 1993, envelop reduction of sparse matrix: find ordering such that the envelop is minimized

$$\min \sum_{ij} (i-j)^2 w_{ij} \quad \Rightarrow \quad \min \sum_{ij} (\pi_i - \pi_j)^2 w_{ij}$$

39

Tutorial on Spectral Clustering, ICML 2004, Chris Ding © University of California

Distance-sensitive ordering

Ordering is determined by permutation indexes

4-variable. For a given ordering, there are 3 *distance*=1 pairs, two *d*=2 pairs, one *d*=3 pair.

$$\begin{bmatrix} 0 & 1 & 2 & 3 \\ & 0 & 1 & 2 \\ & & 0 & 1 \\ & & & 0 \end{bmatrix}$$

$$\pi(1,\cdots,n) = (\pi_1,\cdots,\pi_n)$$

$$J_d(\pi) = \sum_{i=1}^{n-d} s_{\pi_i,\pi_{i+d}}$$

$$\min_{\pi} J, \; J(\pi) = \sum_{d=1}^{n-1} d^2 J_d(\pi)$$

The larger distance, the larger weights. Large distance similarities reduced more than small distance similarities

(Ding & He, ICML'04) 40

Tutorial on Spectral Clustering, ICML 2004, Chris Ding © University of California

Distance-sensitive ordering

Theorem. The continuous optimal solution for the discrete inverse permutation indexes are given by the scaled principal component q_1.

The shifted and scaled inverse permutation indexes

$$q_i = \frac{\pi_i^{-1} - (n+1)/2}{n/2} = \{\frac{1-n}{n}, \frac{3-n}{n}, \cdots, \frac{n-1}{n}\}$$

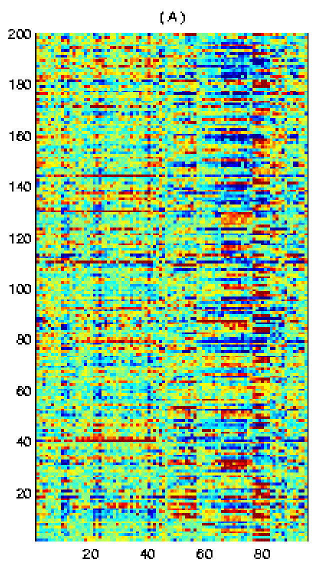
Relax the restriction on q. Allow it be continuous.

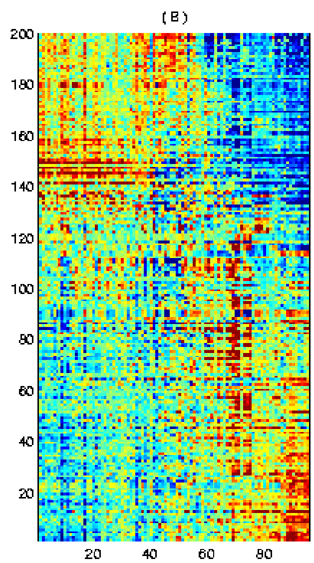
Solution for q becomes the eigenvector of

$$(D-S)q = \lambda Dq$$

41

Re-ordering of Genes and Tissues





$$r = \frac{J(\pi)}{J(\text{random})}$$

$$r = 0.18$$

$$r_{d=1} = \frac{J_{d=1}(\pi)}{J_{d=1}(\text{random})}$$

$$r_{d=1} = 3.39$$

42

C. Ding, RECOMB 2002

Webpage Spectral Ranking

Rank webpages from the hyperlink topology.

L : adjacency matrix of the web subgraph

PageRank (Page & Brin): rank according to principal eigenvector π (equilibrium distribution)

$$\pi T = \pi, \; T = 0.8 D_{out}^{-1} L + 0.2 e e^T$$

HITS (Kleinberg): rank according to principal eigenvector of authority matrix

$$(L^T L) q = \lambda q$$

Eigenvectors can be obtained in closed-form

Webpage Spectral Ranking

HITS (Kleinberg) ranking algorithm

Assume web graph is fixed degree sequence random graph (Aiello, Chung, Lu, 2000)

Theorem. Eigenvalues of $L^T L$

$$\lambda_1 > h_1 > \lambda_2 > h_2 > \cdots\cdots, \; h_i = d_i - \frac{d_i^2}{n-1}$$

Eigenvectors: $$u_k = \left(\frac{d_1}{\lambda_k - h_1}, \frac{d_2}{\lambda_k - h_2}, \cdots, \frac{d_n}{\lambda_k - h_n}\right)^T$$

Principal eigenvector u_1 is monotonic decreasing if $d_1 > d_2 > d_3 > \cdots\cdots$

$\Rightarrow$ HITS ranking is identical to indegree ranking

(Ding, et al, SIAM Review '04)

Webpage Spectral Ranking

PageRank: weight normalization

HITS : mutual reinforcement

Combine PageRank and HITS. Generalize. ⇒

Ranking based on a similarity graph $S = L^T D_{out}^{-1} L$

Random walks on this similarity graph has the equilibrium distribution: $(d_1, d_2, \cdots, d_n)^T / 2E$

PageRank ranking is identical to indegree ranking

(1st order approximation, due to combination of PageRank & HITS)

(Ding, et al, SIGIR'02)

45

PCA: a Unified Framework for clustering and ordering

- PCA is equivalent to *K*-means Clustering
- Scaled PCA has two optimality properties
  - Distance sensitive ordering
  - Min-max principle Clustering
- SPCA on contingency table ⇒ Correspondence Analysis
  - Simultaneous ordering of rows and columns
  - Simultaneous clustering of rows and columns
- Resolve open problems
  - Relationship between Correspondence Analysis and PCA (open problem since 1940s)
  - Relationship between PCA and *K*-means clustering (open problem since 1960s)

46

Spectral Clustering:

a rich spectrum of topics

a comprehensive framework for learning

A tutorial & review of spectral clustering

Tutorial website will post all related papers (send your papers)

Acknowledgment

Hongyuan Zha, Penn State
Horst Simon, Lawrence Berkeley Lab
Ming Gu, UC Berkeley
Xiaofeng He, Lawrence Berkeley Lab
Michael Jordan, UC Berkeley
Michael Berry, U. Tennessee, Knoxville
Inderjit Dhillon, UT Austin
George Karypis, U. Minnesota
Haesen Park, U. Minnesota

Work supported by Office of Science, Dept. of Energy