Normalized Cuts and Spectral Clustering

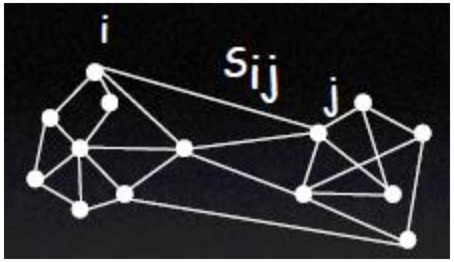
J. Shi & J. Malik, TPAMI2000
Presenter: Qi Wang

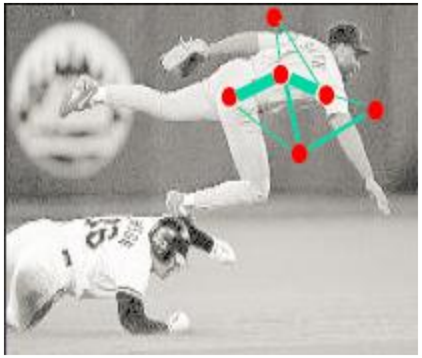
Contents

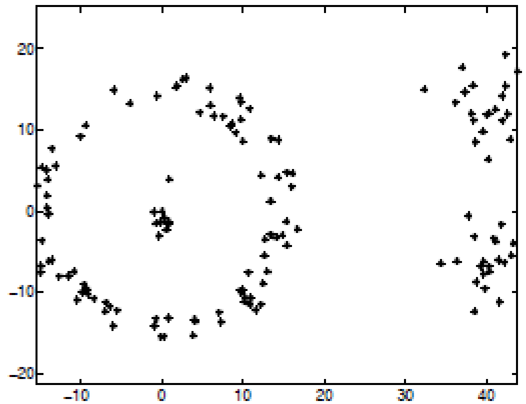
- Graph cuts
- Ncuts formulation
- Graph cuts methods comparison
- Practical issues
  - Constructing similarity graph
  - Continuous vs. discrete solutions
  - Multi-scale
- Other interpretations
- Pros and cons

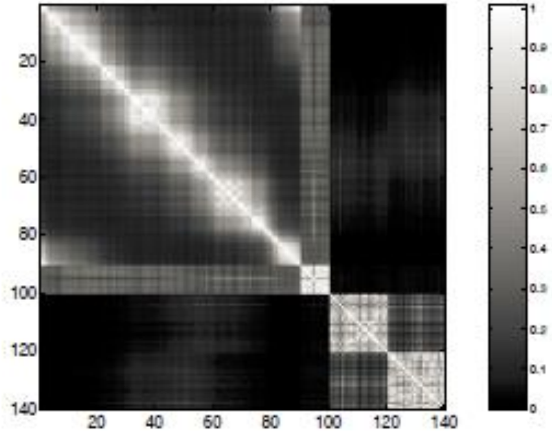
Graph cuts

- Similarity graph



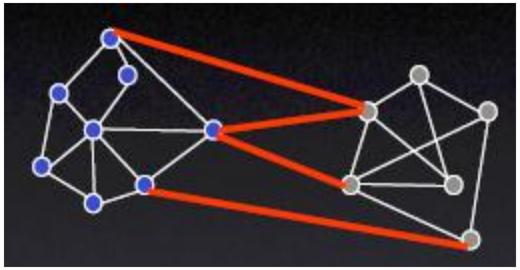


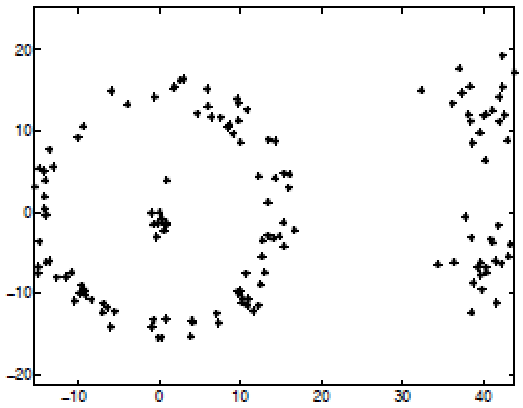


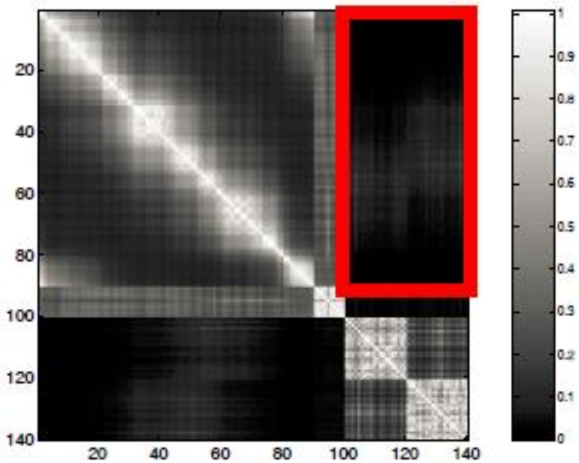


Graph cuts

- $cut(A,B) = \sum_{i \in A, j \in \bar{A}} s_{ij}$
- Objective: Find the cuts that best groups the vertices according to similarity.
- Two major concerns of Graph cuts:
  - A good optimization criterion
  - Computational tractability







Graph cuts

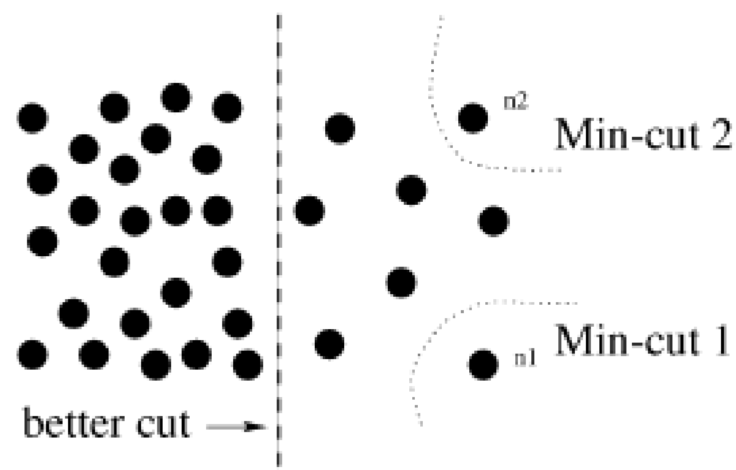
- Min cut

$$\min cut(\mathrm{A}, \overline{\mathrm{A}})$$

- Have efficient algorithm for exact discrete solution

- Problem: Tend to cut small sets of isolated nodes

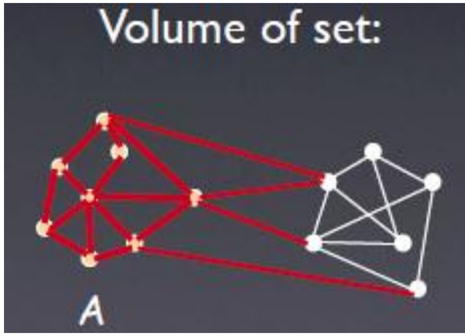
- Solution: Normalize the cuts in terms of subsets volumes(Ncuts)



Normalized cuts

- Objective

$$\min Ncut(A, \overline{A}) = \frac{cut(A, \overline{A})}{vol(A)} + \frac{cut(A, \overline{A})}{vol(\overline{A})}$$



- Graph Laplacian

$$L = D - W$$

 where $D = diag(d_1, d_2, \dots, d_N)$ and $W_{ij} = s_{ij}$

- Define cluster indicator vector

$$f_i = \begin{cases} \sqrt{\dfrac{vol(\overline{A})}{vol(A)}} & if\ v_i \in A \\ \sqrt{\dfrac{vol(A)}{vol(\overline{A})}} & if\ v_i \in \overline{A} \end{cases}$$

- We have

$$f'Lf = vol(V)Ncut(A, \overline{A})$$

Normalized cuts

- The Ncuts equivalent problem

$$\begin{aligned} &\min f'Lf\ , \\ \text{s.t.}\quad & f \text{ discrete as defined above;} \\ &(Df)'\mathbf{1} = 0; f'Df\ = vol(V). \end{aligned}$$

- Relax

$$\begin{aligned} &\min f'Lf\ , \\ \text{s.t.}\quad &(Df)'\mathbf{1} = 0; f'Df\ = vol(V). \end{aligned}$$

- Substitute $g \coloneqq D^{1/2}f$

$$\begin{aligned} &\min_{g\in\mathbb{R}^N} g'D^{-1/2}LD^{-1/2}g\ \ , \\ \text{s.t.}\quad &(g)'D^{1/2}\mathbf{1} = 0;\ \|g\|^2 = vol(V). \end{aligned}$$

- According to Rayleigh-Ritz theorem, the solution is given by second eigenvector of $D^{-1/2}LD^{-1/2}$, and thus f is second generalized eigenvector of

$$Lu = \lambda Du$$

Normalized cuts

- Generalized to $k > 2$ clusters:

$$H_{ij} = \begin{cases} \dfrac{1}{\sqrt{vol(A_j)}} & if\ v_i \in A_j \\ 0 & otherwise \end{cases}$$

$$\min tr(H'LH) \quad s.t.\ H'DH = I$$

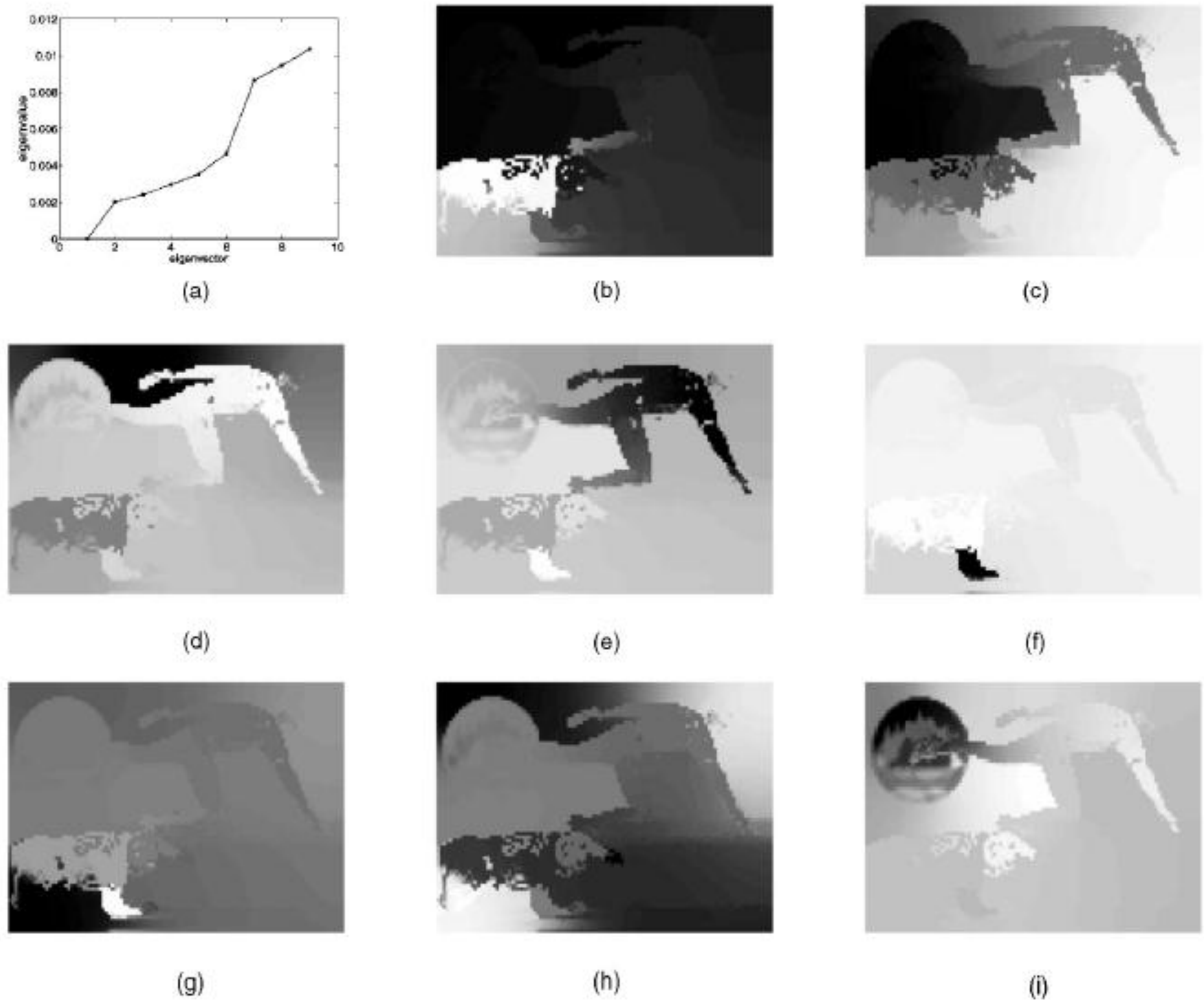
- or let $H = D^{-1/2}T$,

$$\min_{T \in \mathbb{R}^{n \times k}} tr\left(D^{-1/2} L D^{-1/2}\right) \ \ s.t. T'T = I$$

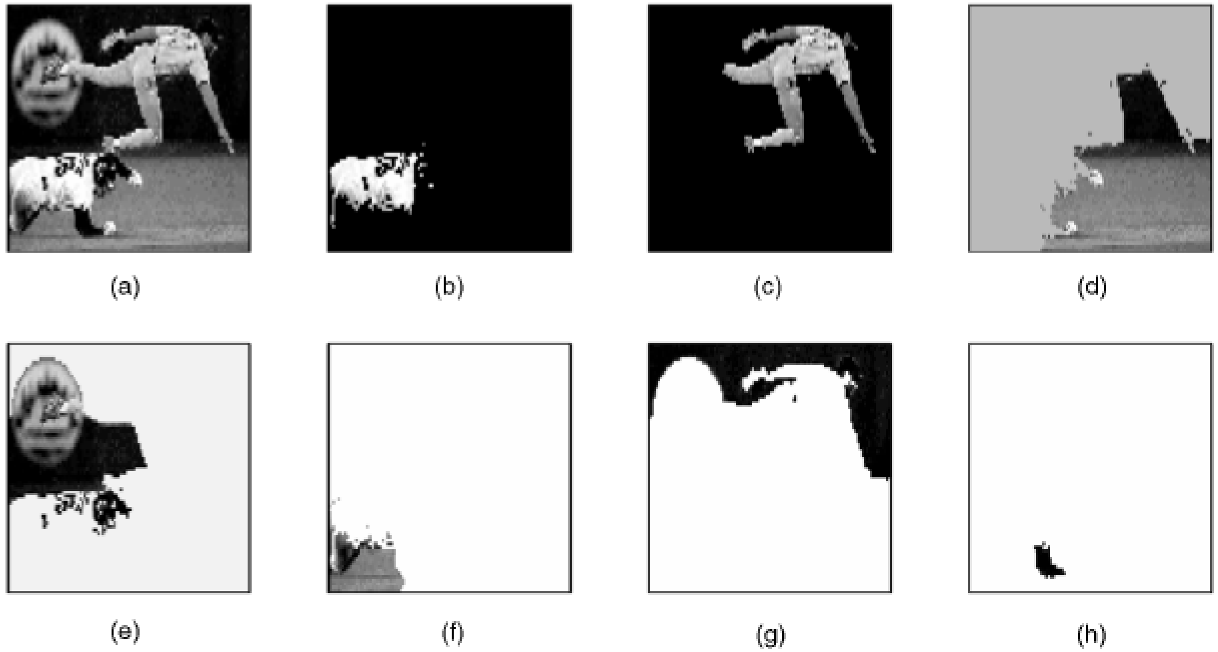
- T is given by first k eigenvectors of $D^{-1/2} L D^{-1/2}$

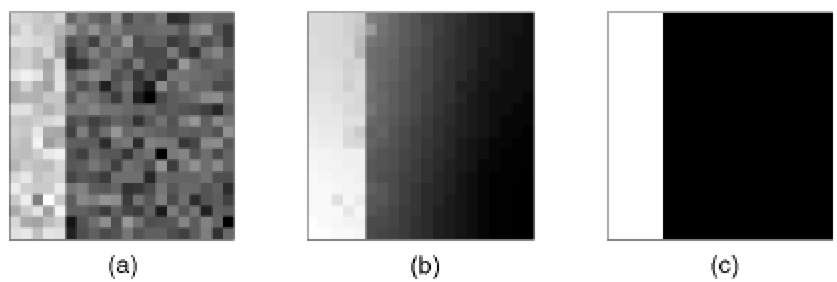
Normalized cuts

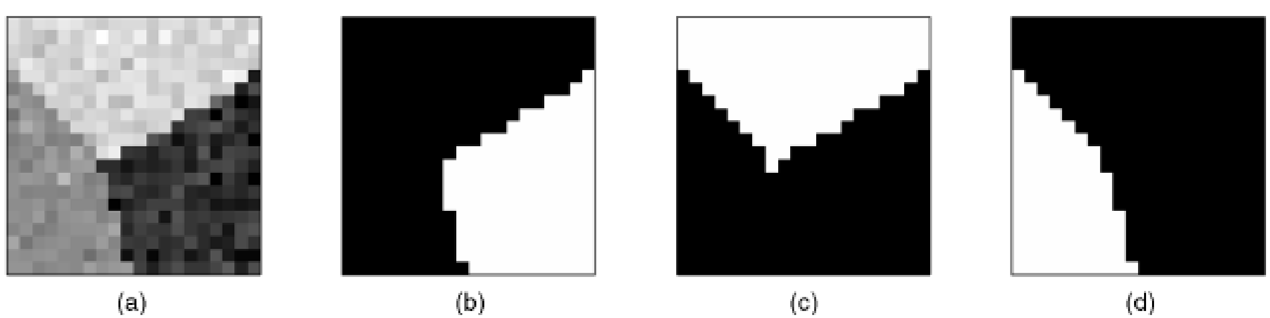
- Computation of eigenvector: Lanczos method
- Post processing: From relaxed to discrete
  - For 2-way cut, simply threshold. View k-way cut as recursive 2-way cut.
  - Simultaneous k-way cut, apply k-means to the relaxed solution
  - Other theories related with the approximation(Yu and Shi, ICCV'03; Bach and Jordan, NIPS'04)

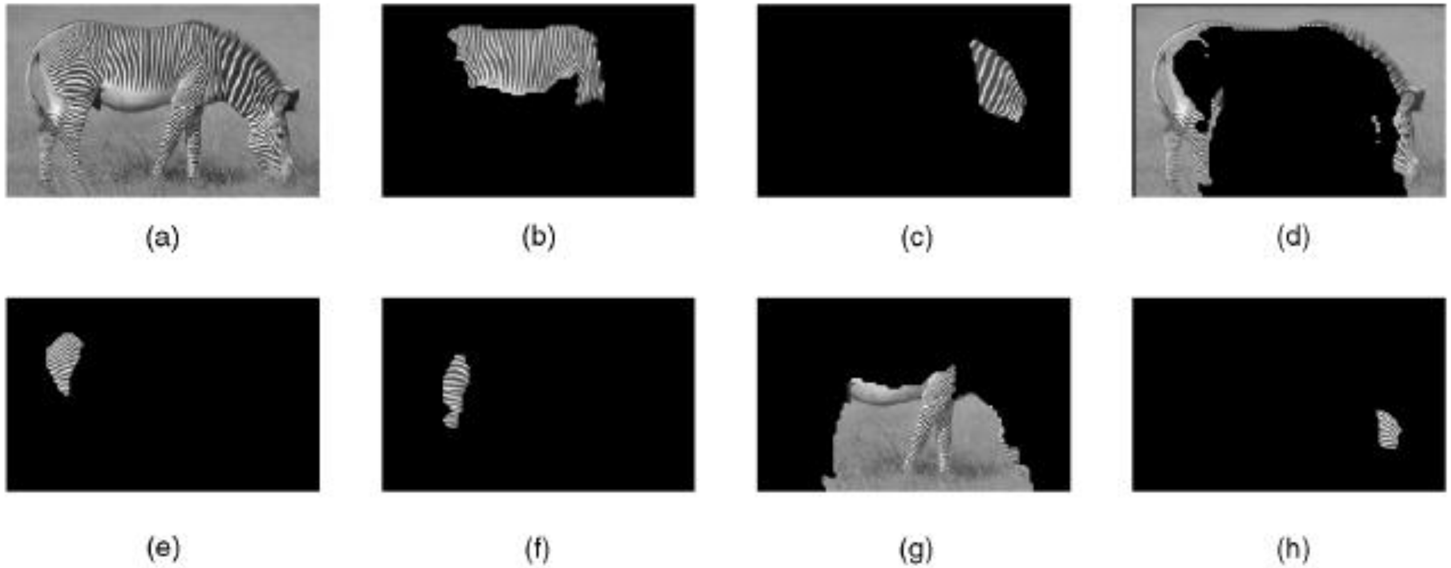


0.012
0.01
0.008
0.006
0.004
0.002
0
eigenvalue
0
2
4
6
8
10
eigenvector
(a)
(b)
(c)
(d)
(e)
(f)
(g)
(h)
(i)


(a)
(b)
(c)
(d)
(e)
(f)
(g)
(h)

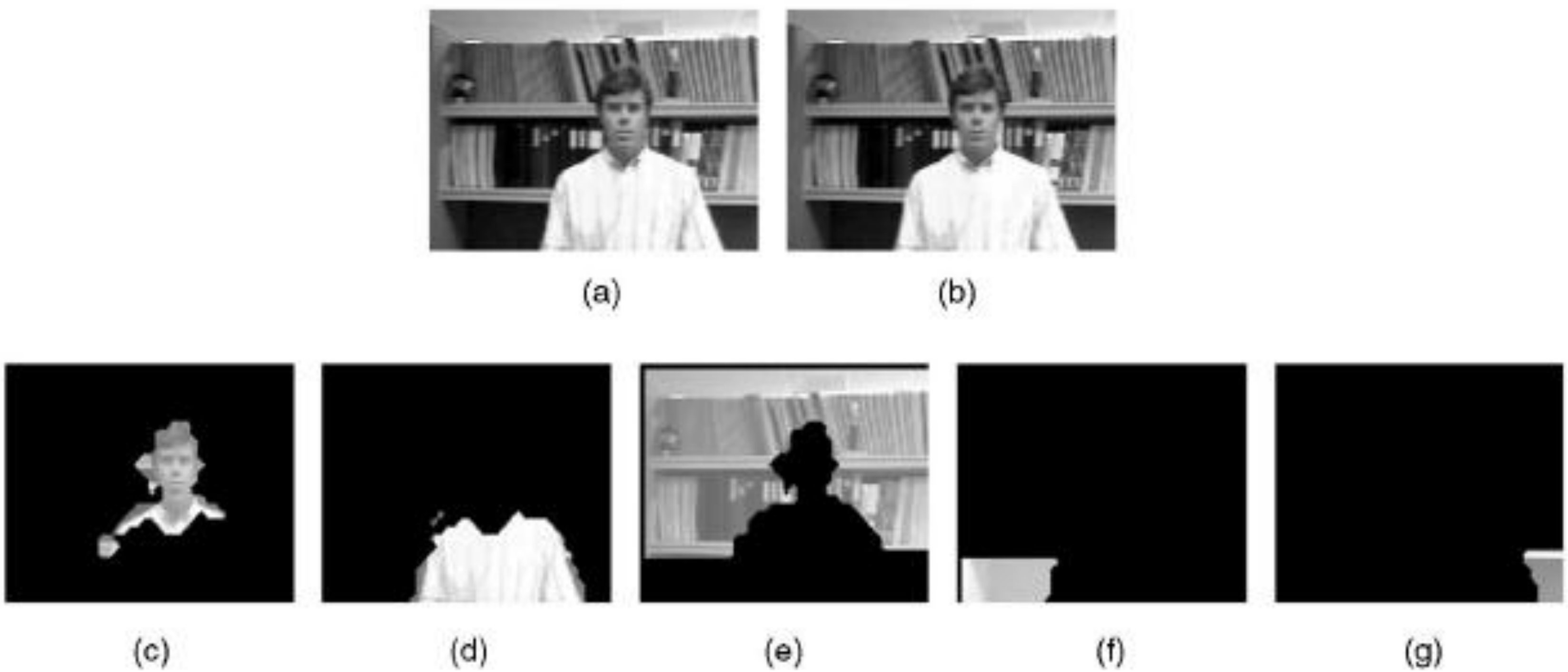

(a)
(b)
(c)


(a)
(b)
(c)
(d)



(a) (b) (c) (d)

(e) (f) (g) (h)



(a) (b)

(c) (d) (e) (f) (g)

Comparison of different graph cuts

- Ncut

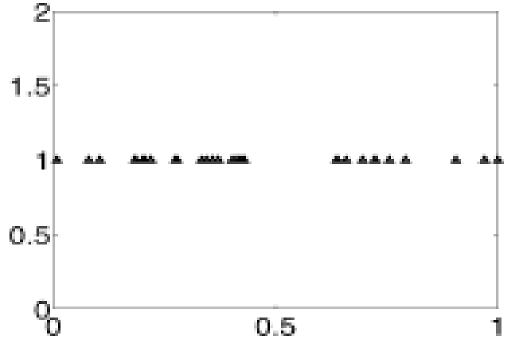
$$\min \frac{cut(A,\bar{A})}{vol(A)} + \frac{cut(A,\bar{A})}{vol(\bar{A})}$$

- Average cut

$$\min \frac{cut(A,\bar{A})}{|A|} + \frac{cut(A,\bar{A})}{|\bar{A}|}$$

- Average association

$$\max \frac{assoc(A,A)}{|A|} + \frac{assoc(\bar{A},\bar{A})}{|\bar{A}|}$$

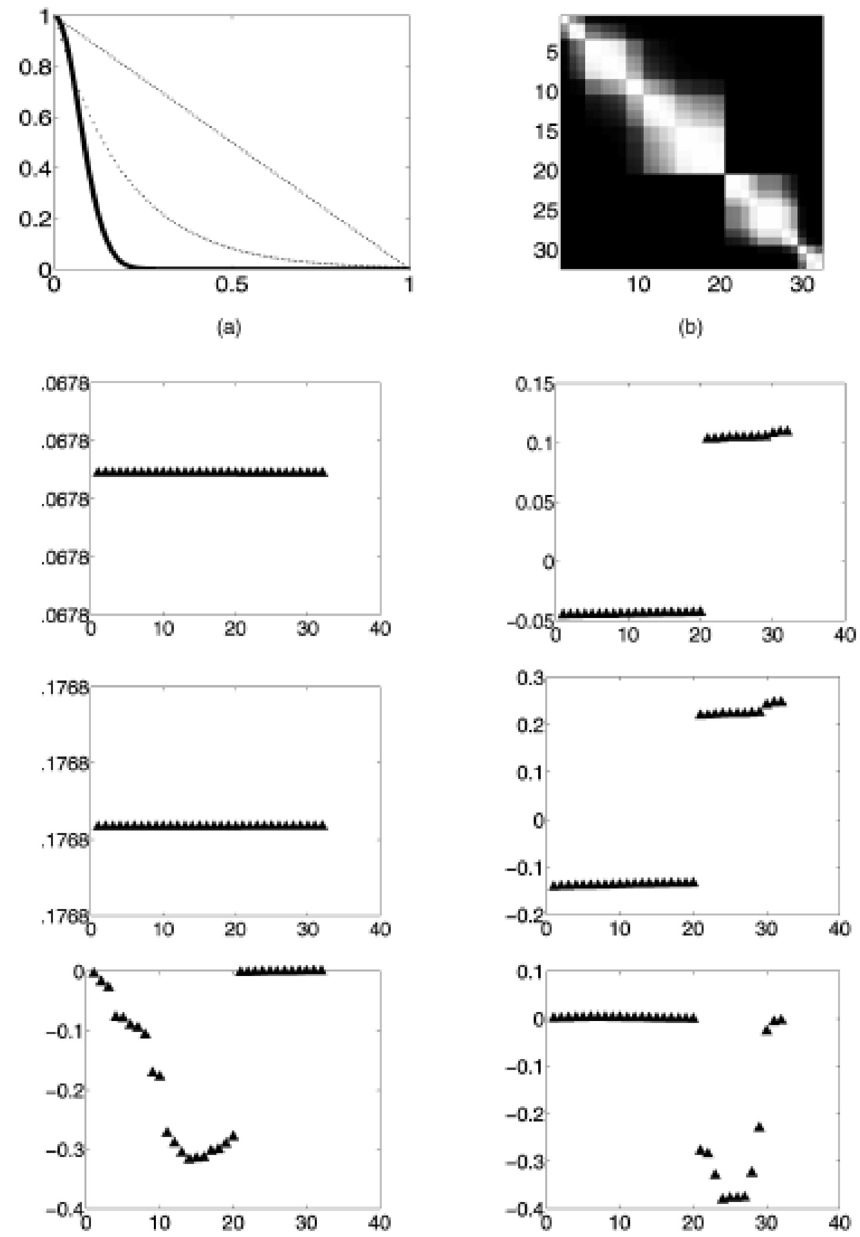


20 points~U(0,0.5)

12 points~U(0.65,1)

Weight function

Weight matrix



(a) (b)

Normalized Cut:

$$(D - W)x = \lambda D x$$

$$W x = (1 - \lambda) D x$$

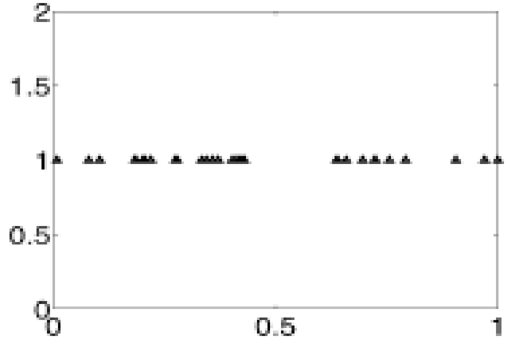
Average Cut:

$$(D - W)x = \lambda x$$

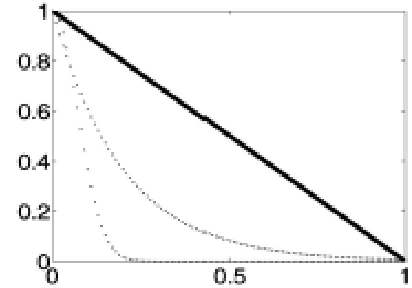
Average Association:

$$W x = \lambda x$$

(c) (d)

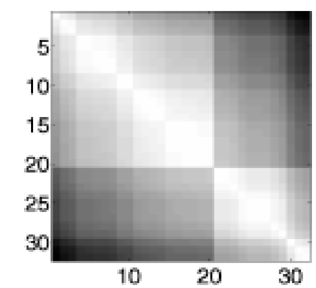


Weight function



(a)

Weight matrix



(b)

Normalized Cut:

$(D - W)x = \lambda D x$

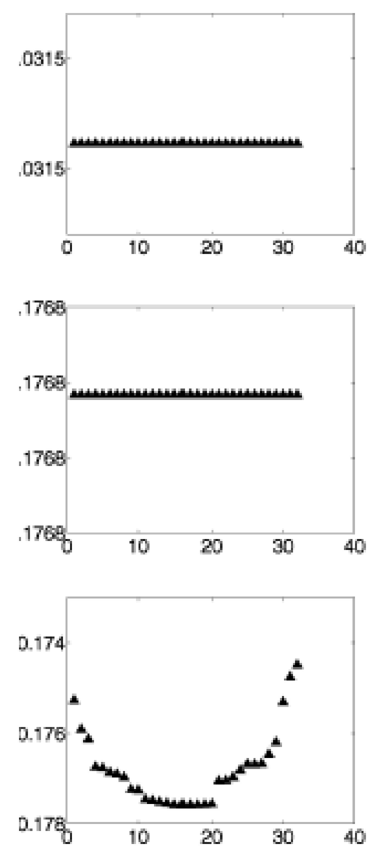
$W x = (1 - \lambda) D x$

Average Cut:

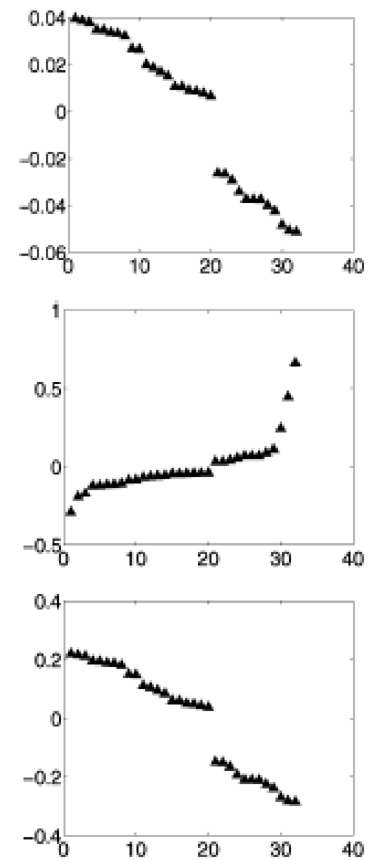
$(D - W)x = \lambda x$

Average Association:

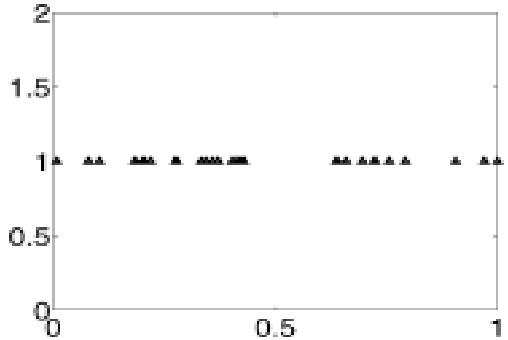
$W x = \lambda x$



(c)

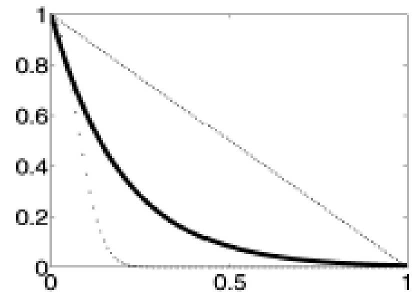


(d)

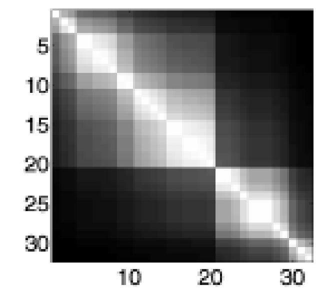


Weight function

Weight matrix



(a)



(b)

Normalized Cut:

$$(D - W)x = \lambda D x$$

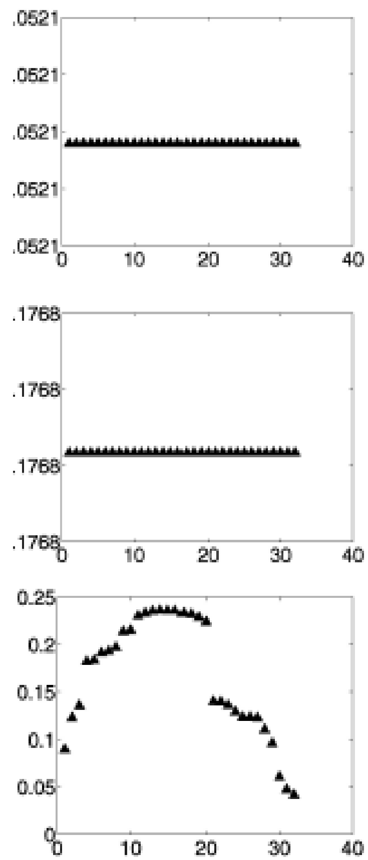
$$Wx = (1 - \lambda) D x$$

Average Cut:

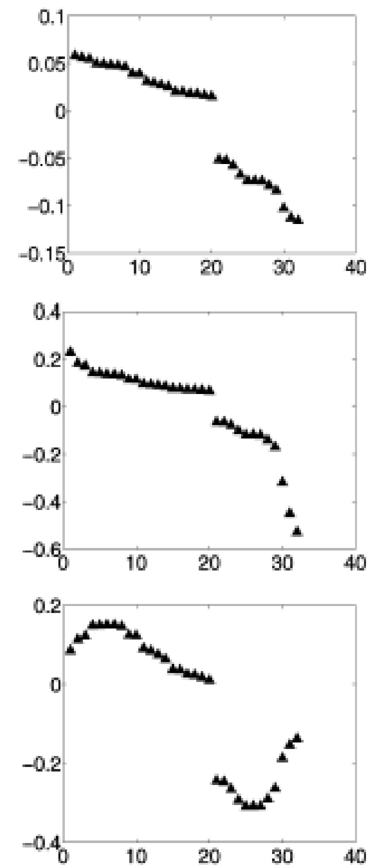
$$(D - W)x = \lambda x$$

Average Association:

$$Wx = \lambda x$$



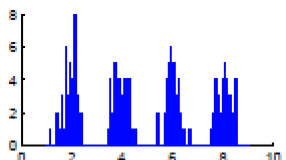
(c)



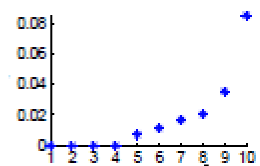
(d)

Practical issues

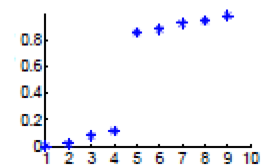
- Constructing similarity matrix
  - Gaussian distance $s(x_i, x_j) = \exp\left(-\|x_i - x_j\|^2 / (2\sigma^2)\right)$
  - Full connection vs. partial connection



Sample histogram

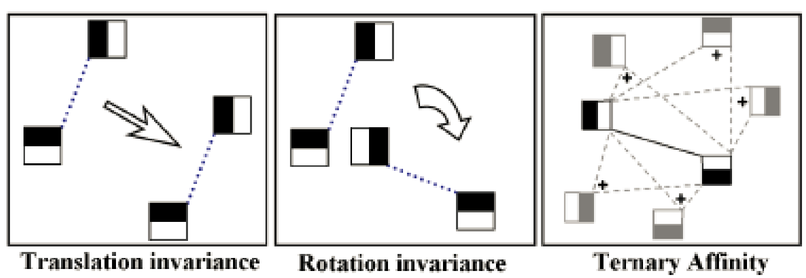


eigenvalue partial connection



eigenvalue full connection

  - Should be tuned for specific tasks(Ex. Cour, Gogin and Shi, AISTAT'05)



$$f(x_1, y_1, \theta_1; x_2, y_2, \theta_2) = \tilde{f}(x_2 - x_1;\ y_2 - y_1;\ \theta_2 - \theta_1)$$

Similarity matrix can be learned

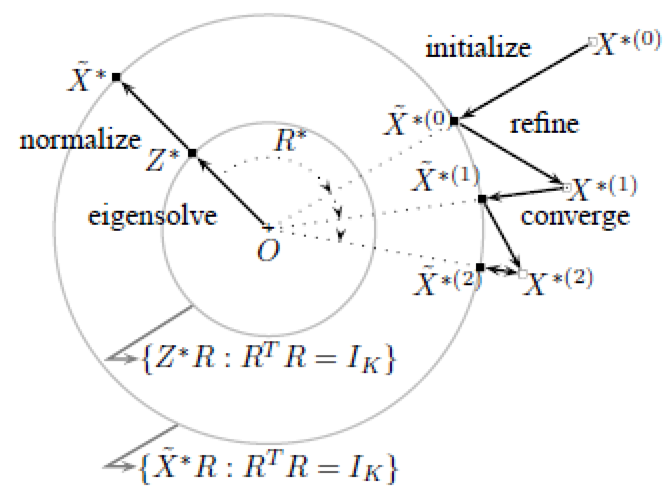
- Cour, Gogin and Shi, AISTAT'05
- Given: Labeled set I with known indicator vector $X^*(I)$, the cost function

$$\varepsilon(W, I) = \frac{1}{2} \| X_2[W(I)] - X^*(I) \|^2$$

- Find the optimum W using gradient method
- Bach and Jordan, NIPS'04

Relaxed vs. discrete

- How to obtain more accurate discrete clustering result from its relaxed solution?
- Orthogonal invariance: if Z is a solution to the relaxed k-way Ncut problem, so is ZR, R is the rotation matrix of $\mathbb{R}^k$ (Yu and Shi, ICCV'03; Bach and Jordan, NIPS'04)
- An iterative method to find a discrete solution X^* according to Z^*



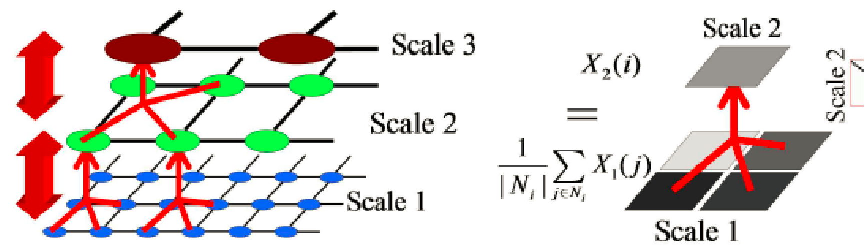
Multi-scale

- Cour, Benezit and Shi, CVPR'05
- Main concern: propagation of local grouping clues across multiple ranges
- Add another constrain across scales:

$X_s - partitioning\ vector\ of\ scale\ s$;

$\mathcal{N}_i - the\ neighbourhood\ of\ vertex\ i$;

$$X_{s+1}(i) = \frac{1}{|\mathcal{N}_i|} \sum_{j \in \mathcal{N}_i} X_s(j)$$



Other interpretations

- Random walks(Maila and Shi, NIPS'01)

$$Ncut(A,\overline{A}) = P(A \to \overline{A}) + P(\overline{A} \to A)$$

- Perturbation theory(Luxburg Tutorial'07)
  - The eigenvector deviation due to the matrix perturbation is bounded
  - View the actual W as a perturbated version of an ideal $\widetilde{W}$ with clusters disjoint, the eigenvectors should also approximate the ideal indicator vector

Pros and cons

- Pros
  - Ncuts does not make strong assumptions on the cluster forms(unlike k-mean, which assumes the clusters to be convex)
  - It's computational efficient

- Cons
  - Very sensitive to the selection of similarity matrix and parameters

Sources of materials

- Timothee Cour, Jianbo Shi, Nicolas Gogin, Learning Spectral Graph Segmentation, AISTAT 2005.
- Timothee Cour, Florence Benezit, Jianbo Shi, Spectral Segmentation with Multi-Scale Graph Decomposition , CVPR 2005.
- Jianbo Shi and Jitendra Malik, Normalized Cuts and Image Segmentation, TPAMI 2000.
- Stella X. Yu, Jianbo Shi, Multiclass Spectral Clustering, ICCV 2003.
- Ulrike von Luxburg, A Tutorial on Spectral Clustering
- Francis R. Bach, Michael I. Jordan, Learning Spectral Clustering