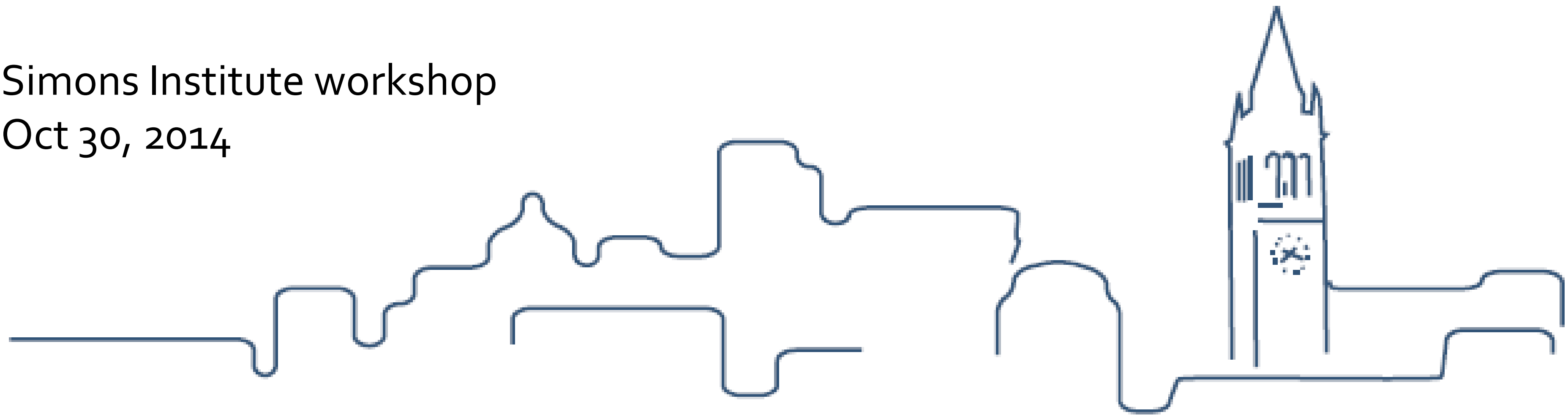
Comparing the Theory and Practice of Spectral Algorithms to Combinatorial Flow Algorithms for Expander Ratio, Normalized Cut, Clustering and Conductance

Dorit S. Hochbaum
University of California, Berkeley

Simons Institute workshop
Oct 30, 2014



Notations and Preliminaries

- An undirected graph G=(V,E) $n = |V|$ $m = |E|$
- Edges' weights $w_{ij} \quad \forall [i,j] \in E$
- Nodes' weights $q_i \quad \forall i \in V$
- Capacity of a Cut $C(A,B) = \sum_{i \in A, j \in B} w_{ij}$
- Weighted degree $d_i = \sum_{j | [i,j] \in E} w_{ij}$
- Degree Volume $d(A) = \sum_{i \in A} d_i = 2C(A,A) + C(A,\bar{A})$
- Node Volume $q(A) = \sum_{i \in A} q_i$

The graph expander problem

- The graph expander problem:
  Expander graphs are used to generate good error correcting codes, and in cryptography.

$$\min_{|S|\le\frac{n}{2}} \frac{C(S,\bar{S})}{|S|} = \min_{S\subset V} \frac{C(S,\bar{S})}{\min\{|S|,|\bar{S}|\}}$$

The Cheeger problem

- The Cheeger problem, normalized cut:
  for effective segmentation of images.

$$\min_{d(S)\le\frac{d(V)}{2}} \frac{C(S,\bar{S})}{d(S)} = \min_{S\subset V} \frac{C(S,\bar{S})}{\min\{d(S),d(\bar{S})\}}$$

Also called Conductance, when the underlying graph is directed and used to assess the convergence rate of Markov chain processes.

A generalization: quantity normalized cut

- The q-normalized cut of a graph:
  Useful in clustering where q_i is a "characteristic" (e.g. texture) of node i

$$\min_{q(S)\leq\frac{q(V)}{2}} \frac{C(S,\bar{S})}{q(S)} = \min_{S\subset V} \frac{C(S,\bar{S})}{\min\{q(S),q(\bar{S})\}}$$

Formulations summary

- The graph expander problem

$$\min_{|S|\le\frac{n}{2}} \frac{C(S,\bar{S})}{|S|}$$

- The Cheeger problem, Normalized cut, Conductance

$$h_G = \min_{d(S)\le\frac{d(V)}{2}} \frac{C(S,\bar{S})}{d(S)}$$

- The q-normalized cut of a graph

$$\min_{q(S)\le\frac{q(V)}{2}} \frac{C(S,\bar{S})}{q(S)}$$

An intuitive clustering criterion

Find a cluster that combines two objectives:
One, is to have large similarity within the cluster, and
to have small similarity between the cluster its complement.

The combination of the two objectives can be expressed as:

$$\min_{S \subset V} \frac{C(S,\bar{S})}{C(S,S)} \quad \text{or}$$

$$\min_{S \subset V} C(S,\bar{S}) - \lambda C(S,S) \quad \text{or}$$

We call this problem **normalized-cut-prime,** or **NC’**.

$$\min_{S \subset V} C_1(S,\bar{S}) - \lambda C_2(S,S)$$

Normalized Cut and NC'

- Shi and Malik 2001:

  Normalized cut: NP-hard

$$\min_{S \subset V} \frac{C(S,\bar{S})}{d(S)} + \frac{C(S,\bar{S})}{d(\bar{S})}$$

- Sharon et al. 2007 called this problem normalized cut:

  Normalized cut': NP-hard?

$$\min_{S \subset V} \frac{C(S,\bar{S})}{C(S,S)}$$

How do NC and NC' compare [H10]

- Shi & Malik

$$\min_{S\subset V} \frac{C(S,\bar{S})}{d(S)} + \frac{C(S,\bar{S})}{d(\bar{S})}$$

- Sharon et. al

$$\min_{S\subset V} \frac{C(S,\bar{S})}{C(S,S)} = \frac{C(S,\bar{S})}{1/2[d(S)-C(S,\bar{S})]} =$$

$$= \frac{1}{\frac{d(S)}{2C(S,\bar{S})}-1} \Rightarrow \max_{S\subset V} \frac{d(S)}{2C(S,\bar{S})} \Rightarrow \min_{S\subset V} \frac{C(S,\bar{S})}{d(S)}$$

Matrix Representation

$$W = \begin{bmatrix} 0 & & & \\ & \ddots & w_{ji} & \\ & w_{ij} & \ddots & \\ & & & 0 \end{bmatrix}$$

$$Q = \begin{bmatrix} q_1 & 0 & \cdots & 0 \\ 0 & \ddots & 0 & \vdots \\ \vdots & 0 & q_i & 0 \\ 0 & \cdots & 0 & q_n \end{bmatrix}$$

$$D = \begin{bmatrix} \sum_{\{i=1|[i,j]\in E\}} w_{ij} & 0 & \cdots & 0 \\ 0 & \ddots & 0 & \vdots \\ \vdots & 0 & \sum_{\{i|[i,j]\in E\}} w_{ij} & 0 \\ 0 & \cdots & 0 & \sum_{\{i=n|[i,j]\in E\}} w_{ij} \end{bmatrix}$$

The Laplacian Matrix $\mathcal{L} = D - W$

Two-terms forms of the problems:

Expander
s-normalized

$$\min_{\substack{\emptyset \neq S \subset V \\ S \neq V}} \frac{C(S,\bar{S})}{|S|} + \frac{C(S,\bar{S})}{|\bar{S}|}$$

Cheeger constant
Normalized Cut

$$\min_{\substack{\emptyset \neq S \subset V \\ S \neq V}} \frac{C(S,\bar{S})}{d(S)} + \frac{C(S,\bar{S})}{d(\bar{S})}$$

Half-q-normalized
q-normalized

$$\min_{\substack{\emptyset \neq S \subset V \\ S \neq V}} \frac{C(S,\bar{S})}{q(S)} + \frac{C(S,\bar{S})}{q(\bar{S})}$$

Single and two-term forms are within a factor of 2:

Expander $\leftrightarrow$ S-normalized

$$\frac{1}{2}(S-normalized) \le Expander \le (S-normalized)$$

Cheeger $\leftrightarrow$ Normalized Cut

$$\frac{1}{2}(NC) \le Cheeger \le (NC)$$

Half-q-normalized $\leftrightarrow$ q-normalized

$$\frac{1}{2}(q-normalized) \le half-q-normalized \le (q-normalized)$$

Two terms expressions and the Rayleigh ratio (Lemma 3.1, [H13])

Lemma 1

$$\min_{\emptyset \neq S \subset V} \frac{C(S,\bar{S})}{q(S)} + \frac{C(S,\bar{S})}{q(\bar{S})} = \min_{\substack{y^T Q\vec{1}=0 \\ y_i \in \{-b,1\}}} \frac{y^T \overbrace{(D-W)}^{\mathcal{L}} y}{y^T Q y}$$

A special case of this was shown by Shi and Malik, for q(S)=d(S).

The combinatorial versus the spectral continuous relaxations

$$(RRP) \quad \min \frac{y^T \mathcal{L} y}{y^T Q y}$$

~~$Y^T Q \vec{1} = 0$~~

~~$y_i \in \{-b, 1\}$~~

Combinatorial relaxation of Raleigh ratio Problem

The spectral method

An optimal solution is achieved for $\mathcal{L}y = \lambda Q y$

Where λ is the smallest non-zero eigenvalue (Fiedler Eigenvalue). We solve for the eigenvector z:

$$\left(Q^{-1/2}\mathcal{L}Q^{-1/2}\right)z = \lambda z$$

and set $y = Q^{-1/2}z$ which solves the continuous relaxation.

Solving the combinatorial relaxation

$$y_i = \begin{cases} 1 & i \in S \\ -b & otherwise \end{cases}$$

The combinatorial relaxation Rayleigh problem

Lemma 2:

$$\alpha(b) = \min_{y\in\{-b,1\}} \frac{y^T(D-W)y}{y^TQy} = \min_{\emptyset\neq S\subset V} \frac{(1+b)^2 C(S,\bar{S})}{q(S)+b^2 q(\bar{S})}$$

For all b,

Two - term $\geq \alpha$(b)

Single - term $\geq \dfrac{\alpha(\mathrm{b})}{2}$

$$(S-normalized),(NC),(q-normalized)\geq \alpha(b)$$

$$Expander, Cheeger, \text{half - q - normalized} \geq \frac{\alpha(b)}{2}$$

Recall Lemma 1:

$$\min_{\substack{y^TQ\vec{1}=0,\\ y\in\{-b,1\}}} \frac{y^T(D-W)y}{y^TQy} = \min_{\emptyset\neq S\subset V} \frac{C(S,\bar{S})}{q(S)} + \frac{C(S,\bar{S})}{q(\bar{S})}$$

Solving the combinatorial Rayleigh problem optimally

- The problem is a ratio problem

General technique for ratio Problems: **The λ-question**

$$\min_{x \in F} \frac{f(x)}{g(x)} < \lambda\ ?$$

can be solved if one can solve the following λ-question:

$$f(x) - \lambda g(x) < 0\ ?$$

*This λ is unrelated to an eigenvector –just a parameter

Solving the λ-question

- The λ-question of whether the value of RRP is less than λ is equivalent to determining whether:

$$\min_{y_i \in \{-b.1\}} y^T(D-W)y - \lambda y^T Q y < 0\,?$$

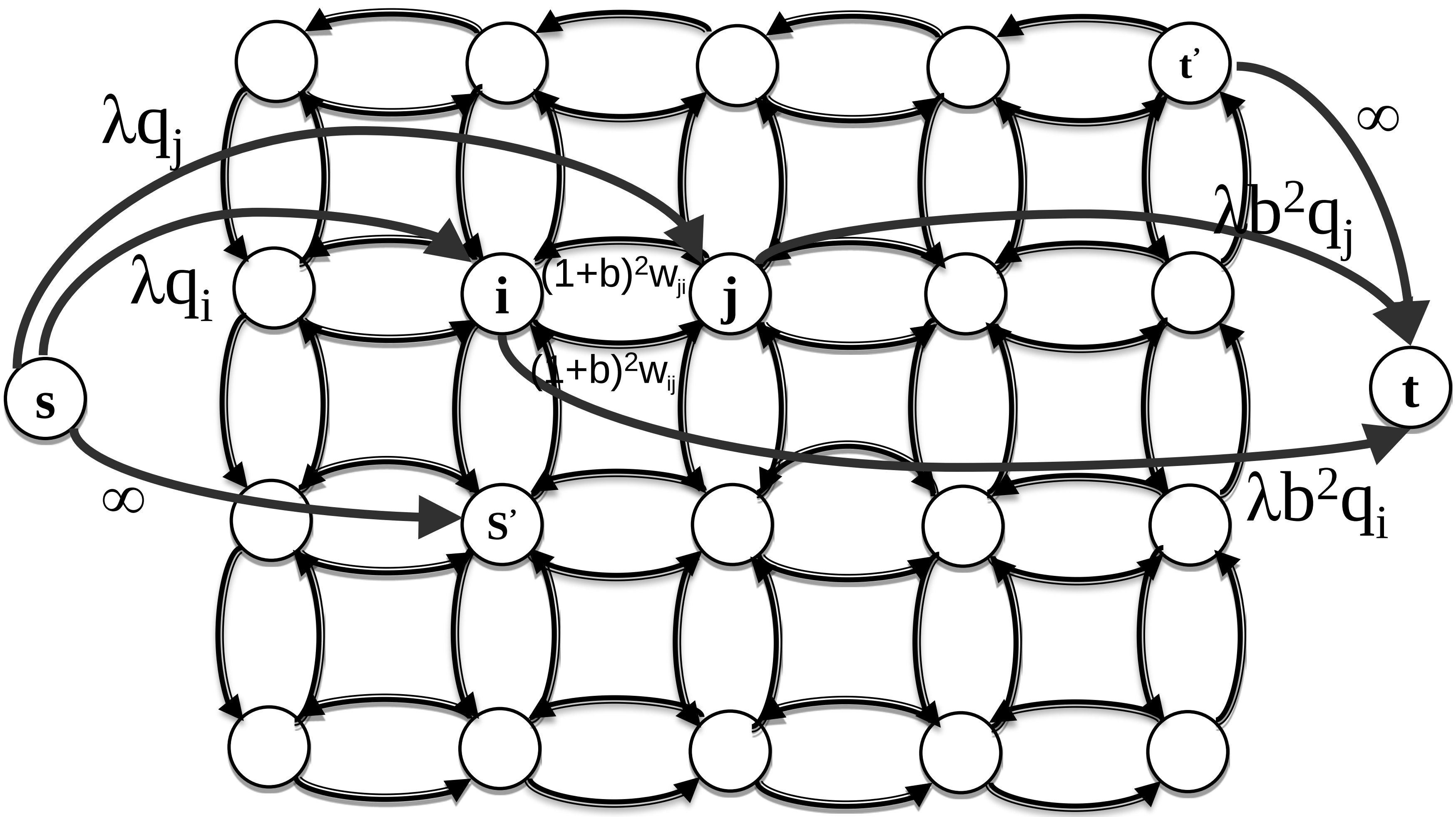
$$\Downarrow \text{OR (from Lemma 1)} \Downarrow$$

Linearized Rayleigh ratio problem (RRP)

$$\left\{\min_{s \subset V} (1+b)^2 C(S,\bar{S}) - \lambda\left[q(S) + b^2 q(\bar{S})\right]\right\} < 0\,?$$

The graph G_{st} for testing the λ-question

(looks arbitrary, but not to worry - it works, as shown next)



It works because the problem can be formulated as "monotone integer program"

Theorem

The source set of a minimum cut in the graph G_{ST} is an optimal solution to the linearized (RRP) (here $T = \bar{S}$)

Proof

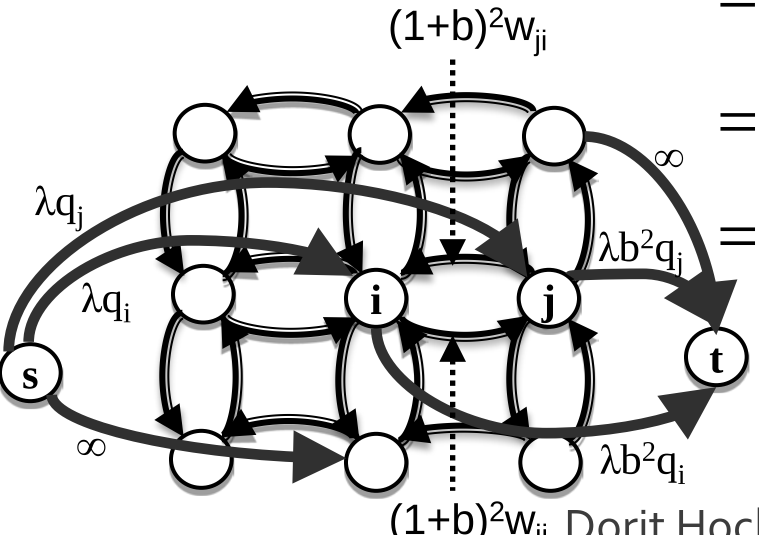
$$C(S \cup \{s\}, T \cup \{t\}) = \lambda q(T) + \lambda b^2 q(S) + C(S,T) =$$

$$= \lambda(1+b^2)q(V) - \lambda q(S) - \lambda b^2 q(T) + C(S,T) =$$

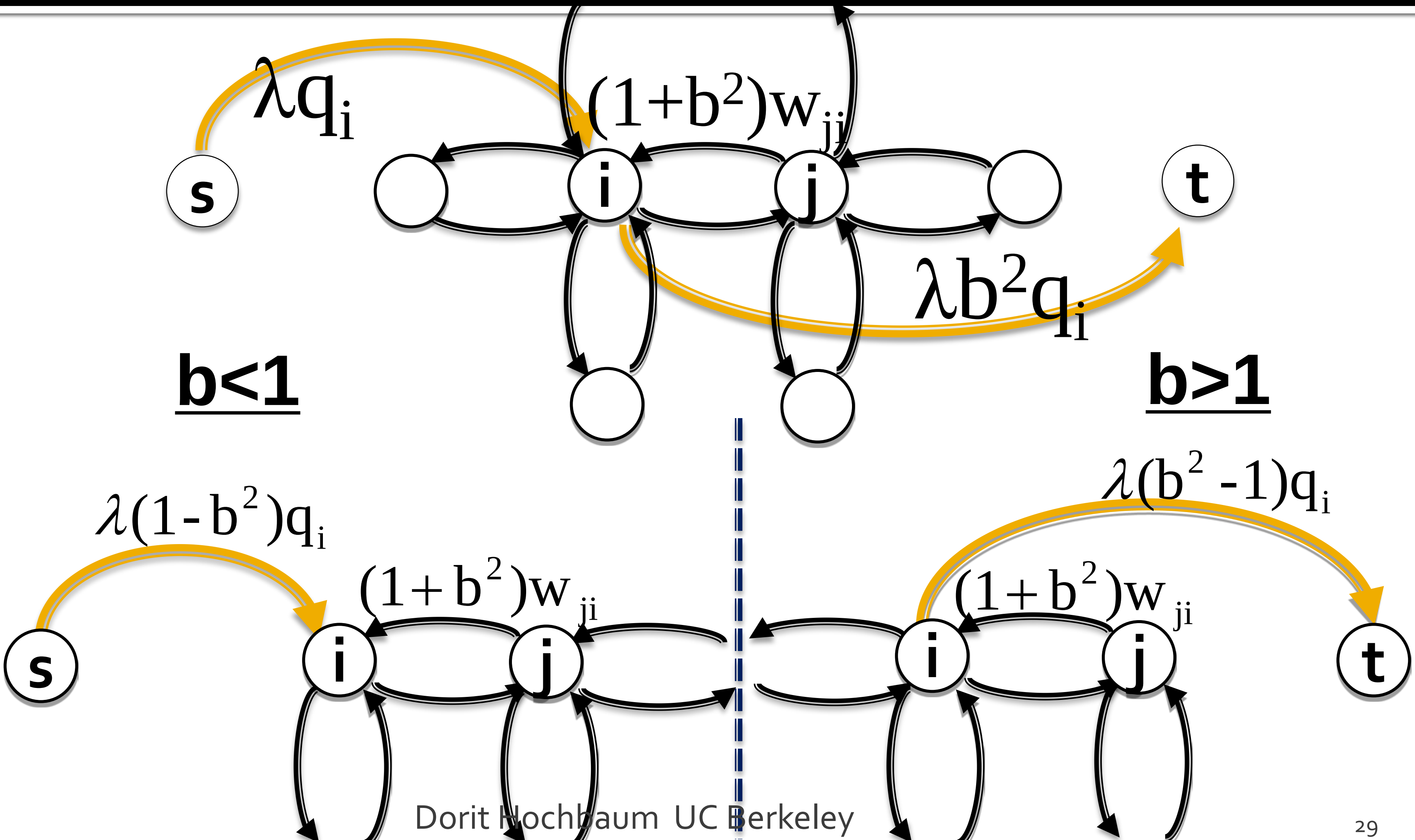
$$= \text{const} - \lambda q(S) - \lambda b^2 q(T) + C(S,T) =$$

$$= \text{const} + C(S,T) - \lambda[q(S) + b^2 q(T)] =$$

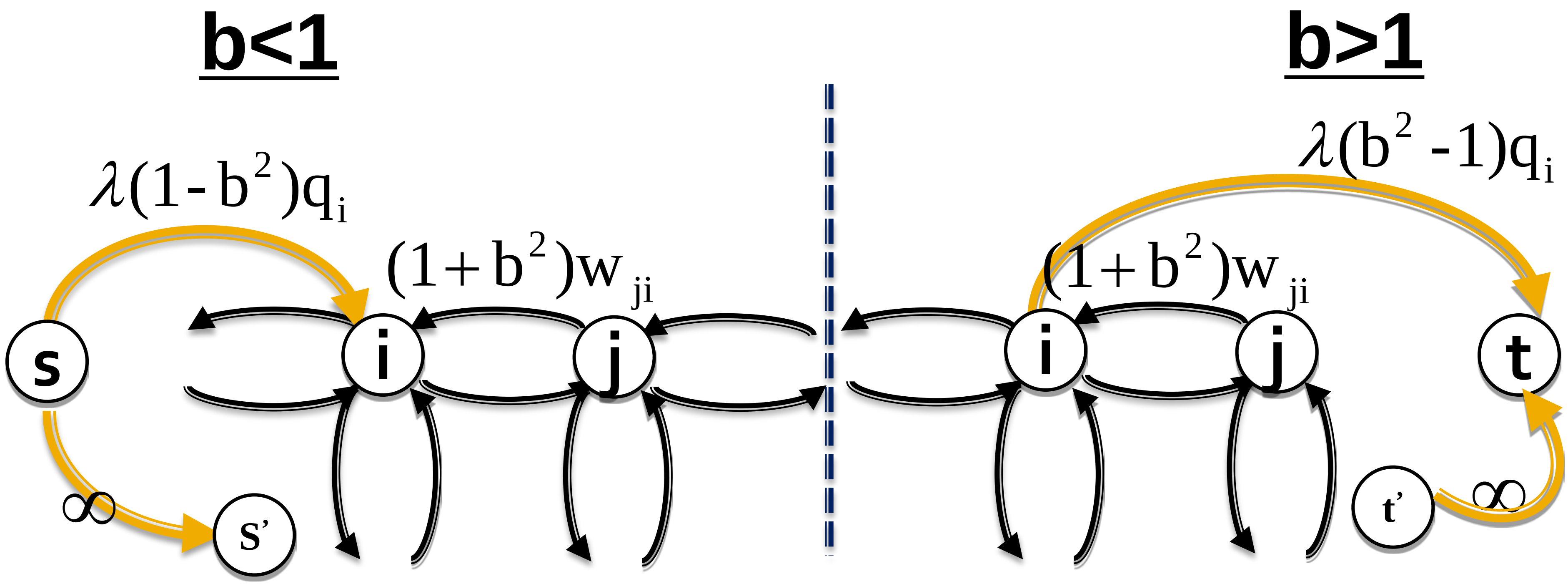
$$= \text{const} + (RRP)$$



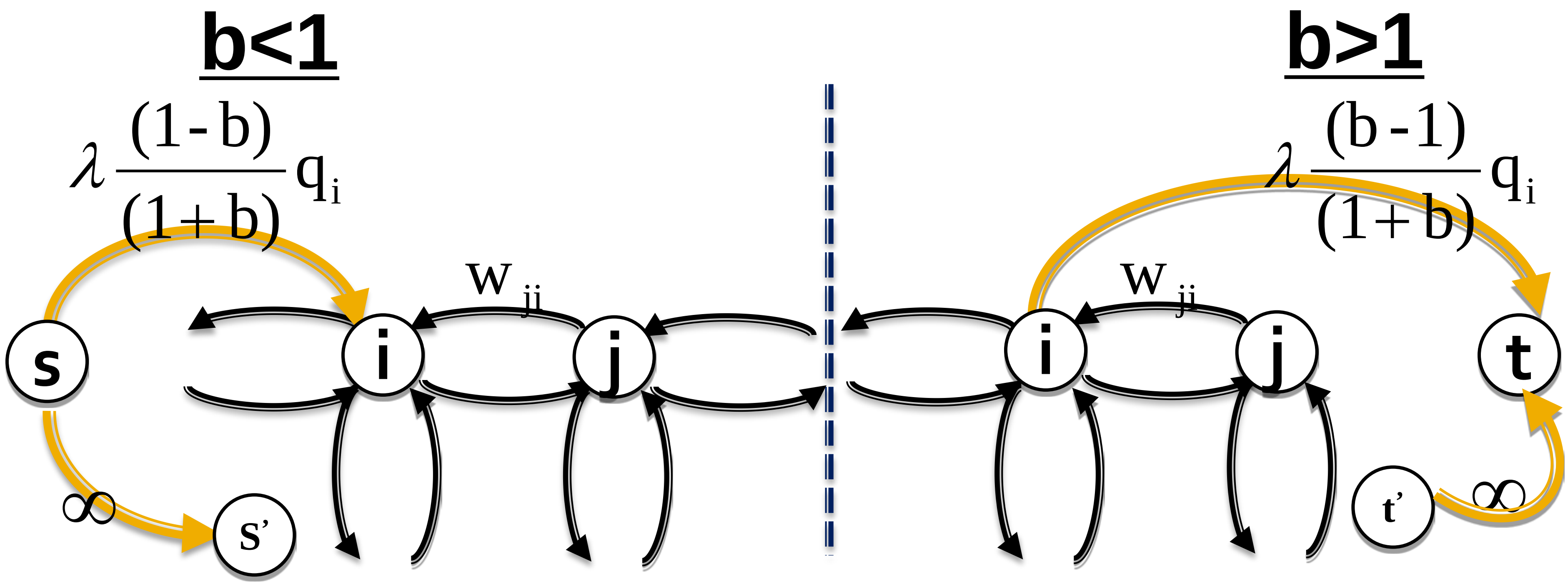
Simplifying the graph



Scaling arcs weights

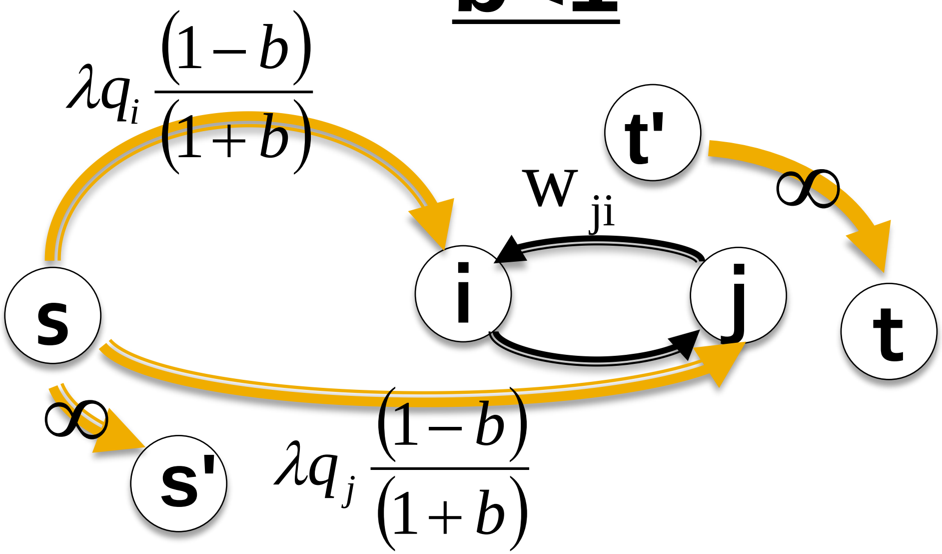


Scaling arcs weights

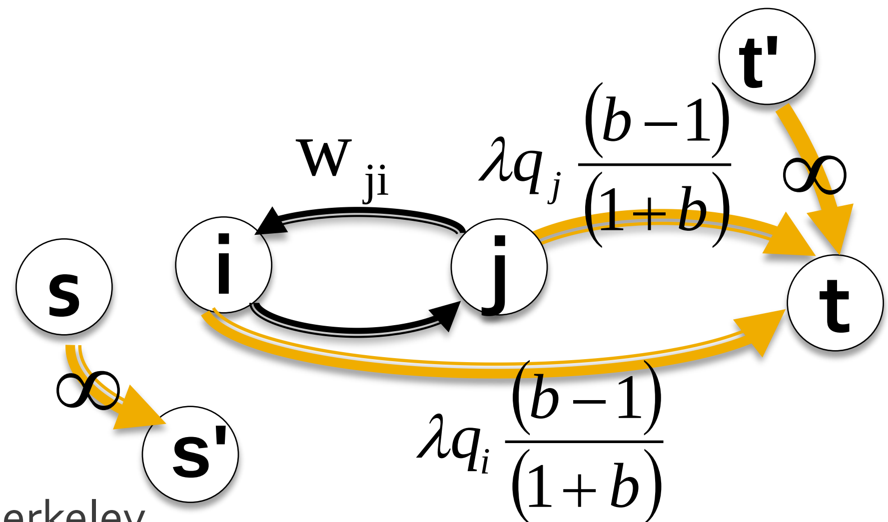


The Simplified equivalent graph

b<1



b>1



Solving the parametric min st cut

- The problem is a ***parametric*** cut problem: This is a graph setup when source adjacent arcs are monotone nondecreasing and sink adjacent are monotone nonincreasing (for b<1) with the parameter.
- A parametric cut problem can be solved in the complexity of a single minimum cut (plus finding the zero of n monotone functions) [GGT89], [H08].
- Here we let the parameter be β

$$\beta = \begin{cases} \lambda \dfrac{1-b}{1+b} & b < 1 \\ \lambda \dfrac{b-1}{1+b} & b \geq 1 \end{cases}$$

In G_{st}

- The cut problem in the graph G_{st}, as a function of β is parametric (the capacities are linear in the parameter on one side and independent of it on the other).
- In a parametric graph the sequence of source sets of cuts for increasing source-adjacent capacities is ***nested***.
- There are no more than n breakpoints for β.
- There are k≤n nested source sets of minimum cuts.

Solving for all values of b efficiently

- For

$$\beta = \begin{cases} \lambda \dfrac{1-b}{1+b} & b<1 \\ \lambda \dfrac{b-1}{1+b} & b \geq 1 \end{cases}$$

- Given the values of β at the breakpoints, we can generate, for each value of b, *all* the breakpoints.
- Consequently, by solving once the parametric problem for β we obtain simultaneously, *all the breakpoint solutions for all b*, in the complexity of a single minimum cut.
- To solve for the minimum ratio: For each b we find the last (largest value) breakpoint where the objective value <0.

Recall problem NC'

$$\min_{S \subset V} \frac{C(S, \bar{S})}{C(S, S)}$$

It has the same solution as

$$\min_{S \subset V} \frac{C(S, \bar{S})}{d(S)} = \alpha(0)$$

Comparison between NC' and the spectral method

$$NC = 35 \cdot 10^{-4}$$

$$NC = 1.702 \cdot 10^{-4}$$

Original image **Eigenvector result** **NC' result**

Another comparison

$NC = 127 \cdot 10^{-4}$

$NC = 1.466 \cdot 10^{-4}$

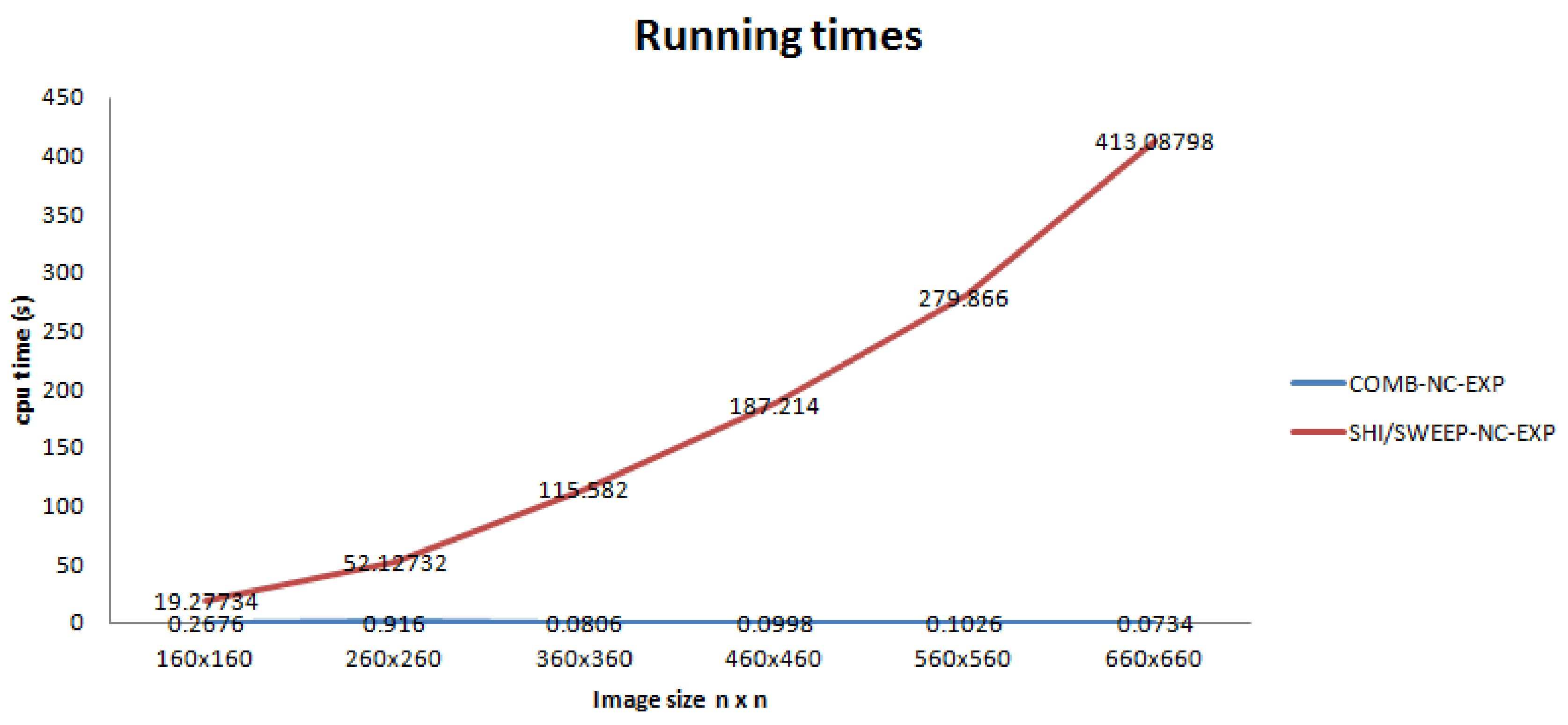
Original image **Eigenvector result** **NC' result**

Empirical testing for the general problems

- For normalized cut d_i is the sum of similarity weights.
- For q-normalized cut, there are, in addition to similarity weights defining the Laplacian, also node weights determined by entropy.
- Exponential similarity weights are applied.
- Total of 20 cases tested.
- Size of images is small due to spectral method software limitations.

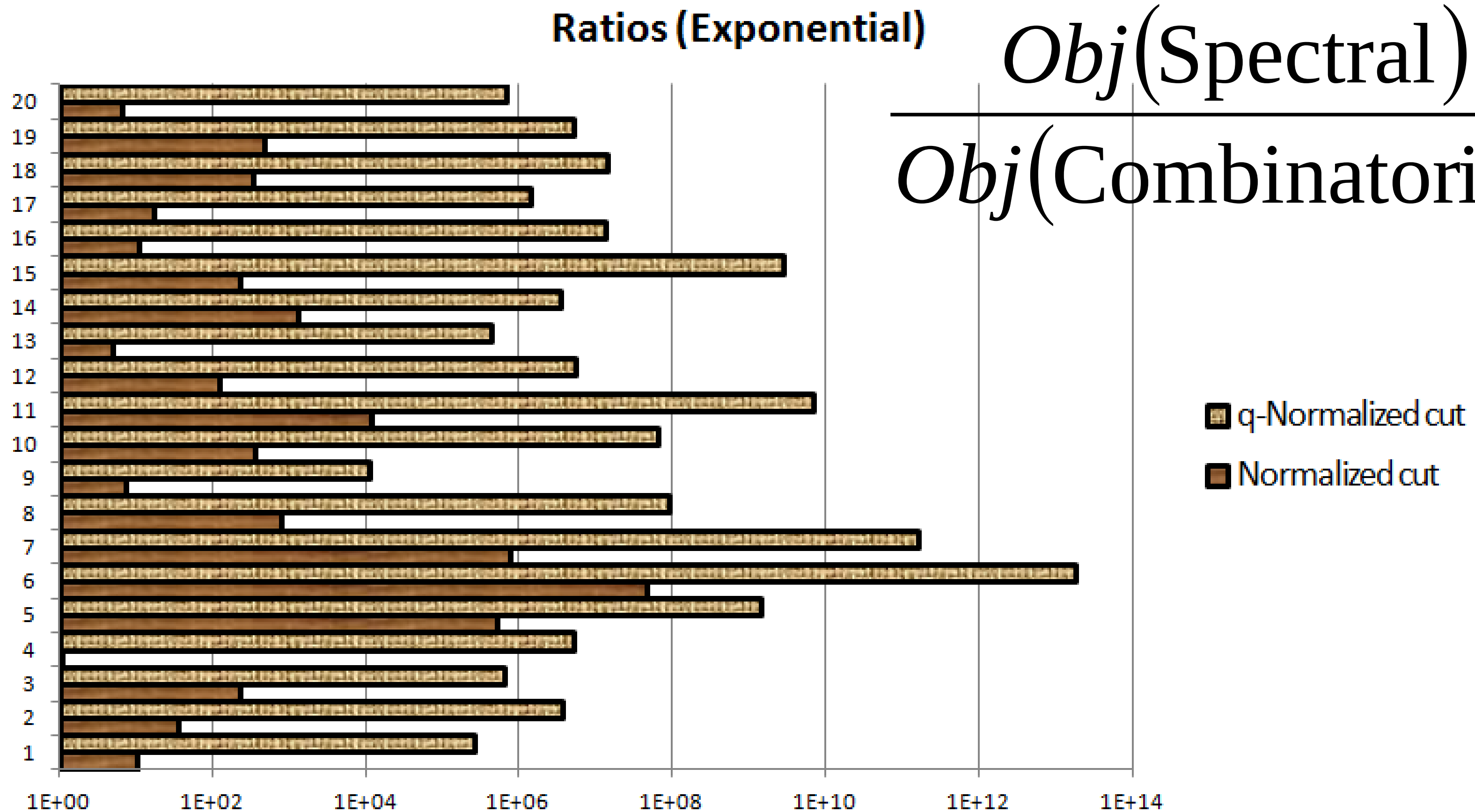
The 20 images

Scalability of NC' versus the spectral algorithm (Shi)



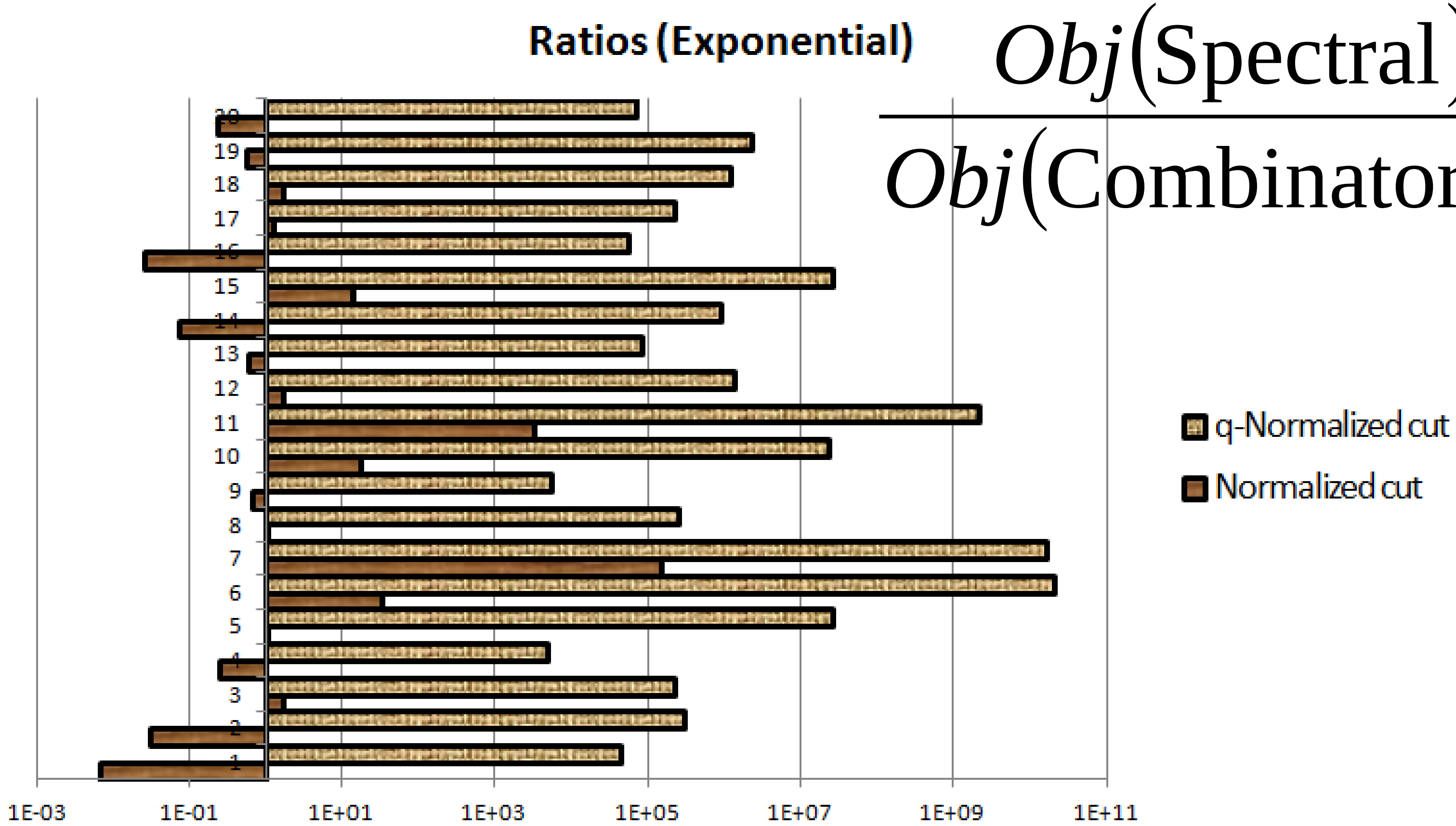
A comparison of NC values of NC' with the spectral algorithm

$$\frac{Obj(\text{Spectral})}{Obj(\text{Combinatorial})}$$



A comparison of NC values of NC' with the sweep spectral algorithm

$$\frac{Obj(\text{Spectral})}{Obj(\text{Combinatorial})}$$



The performance for spectral sweep [H,Cheng, Bertelli13]

For h_G the Cheeger Constant, λ_1 the Fielder eigenvalue,

$$\frac{\lambda_1}{2} \leq h_G \leq \sqrt{2\lambda_1}. \tag{1}$$

The proof of the second inequality of the above bound, introduces a bipartition generated by applying the spectral sweep technique to the Fiedler eigenvector to find a lowest value bipartition for the Cheeger constant's objective. Let the Cheeger constant objective value for this bipartition be denoted by h_{SWEEP}, then at best the sweep solution has the same upper bound as the optimal solution:

$$h_G \leq h_{SWEEP} \leq \sqrt{2\lambda_1} \leq 2\sqrt{h_G}. \tag{2}$$

For NC_{SWEEP} be the lowest value of a bipartition for the normalized cut objective, generated by the spectral sweep technique on the Fiedler eigenvector in the spectral method. (Note: NC_{SWEEP} and h_{SWEEP} may not correspond to the same bipartition.) Let $NC(h_{SWEEP})$ be the objective value of normalized cut for the bipartition that generates the value of h_{SWEEP}, then the following inequality holds

$$NC_{SWEEP} \leq NC(h_{SWEEP}) \leq 2h_{SWEEP}, \quad (3)$$

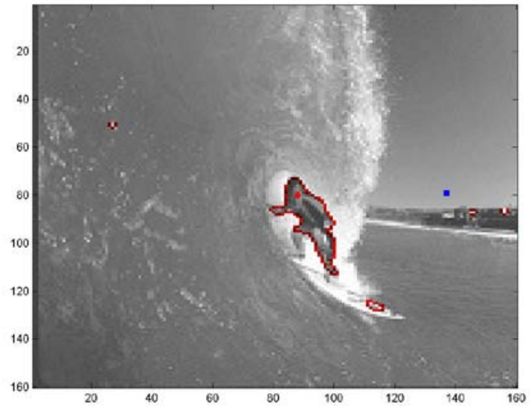
Combining $h_{SWEEP} \leq 2\sqrt{h_G}$ from (2) with (3),

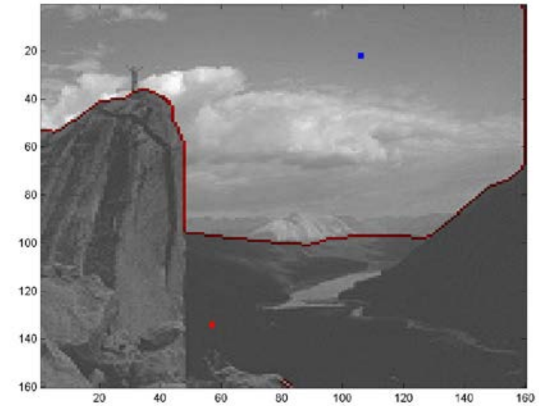
$$NC_{SWEEP} \leq 2h_{SWEEP} \leq 4\sqrt{h_G} \leq 4\sqrt{NC_G} \leq 4\sqrt{NC_{NC'}}.$$

Subjective Visual Segmentation Quality Comparison

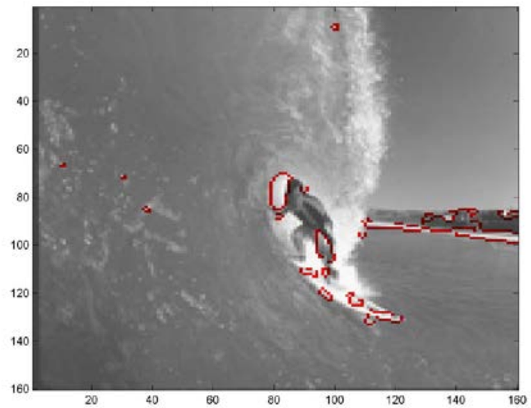
- Normalized Cut

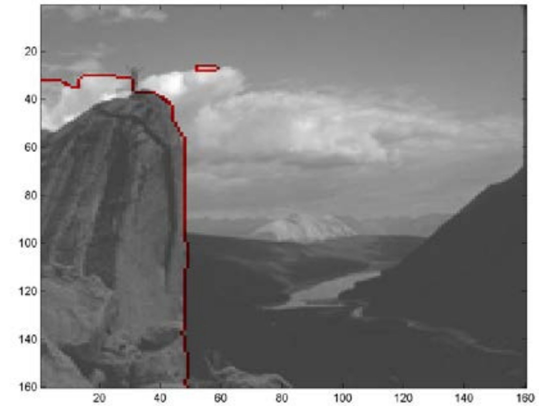
Combinatorial





Spectral

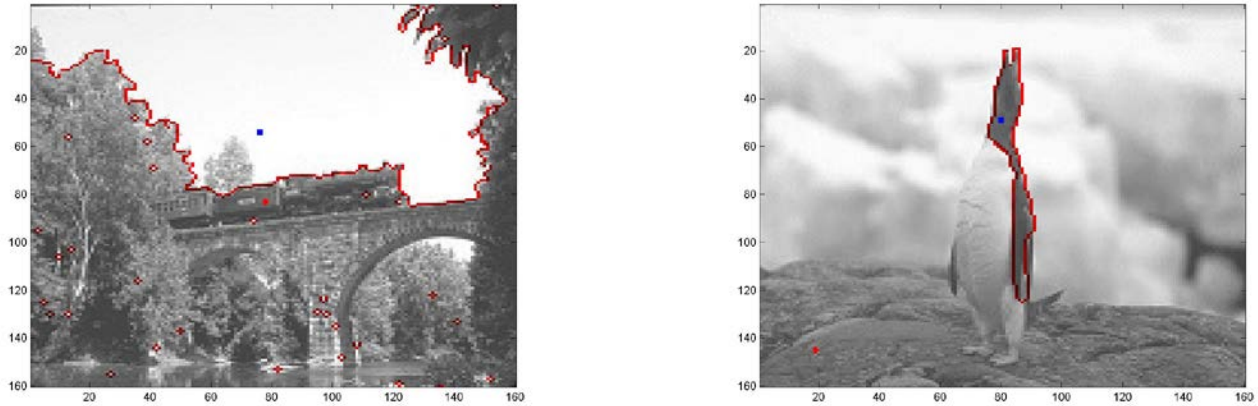




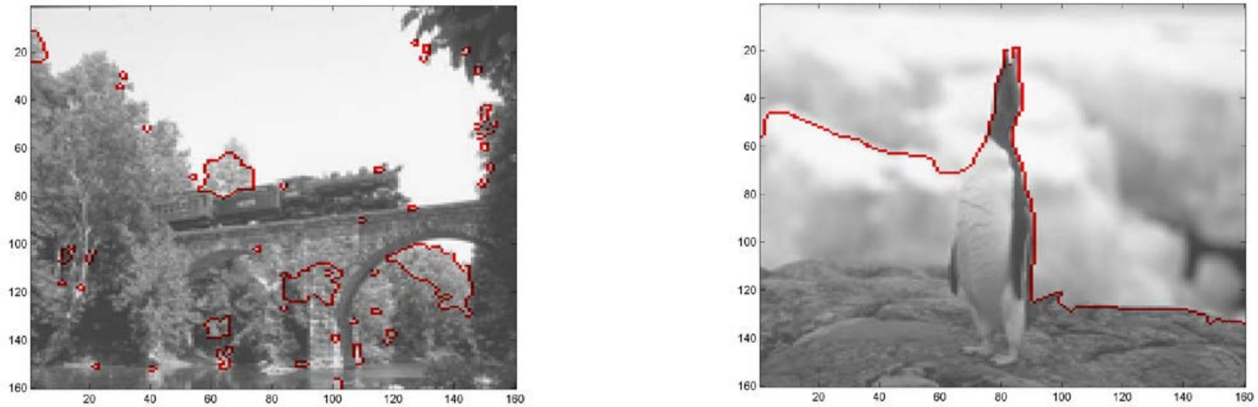
Subjective Visual Segmentation Quality Comparison (cont.)

- Normalized Cut (cont.)

Combinatorial



Spectral

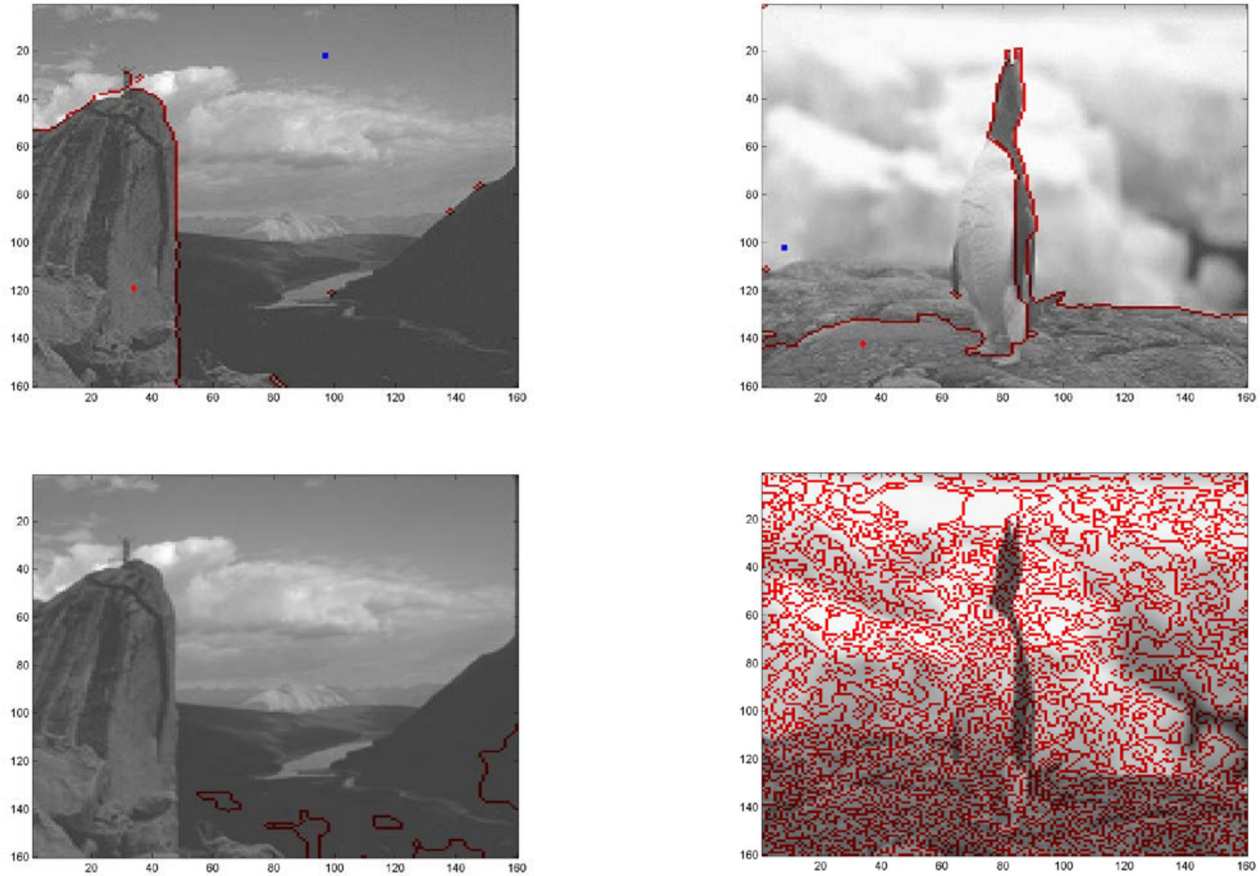


Subjective Visual Segmentation Quality Comparison (cont.)

- q-Normalized Cut (Entropy)

Combinatorial

Spectral

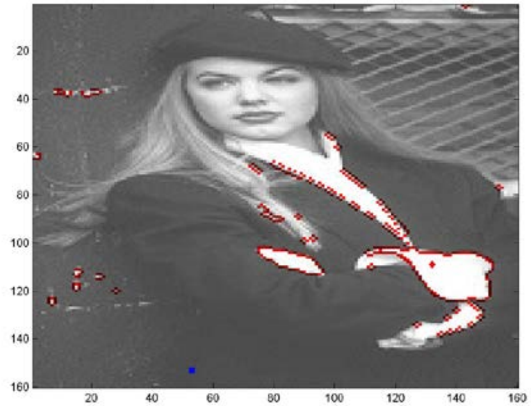


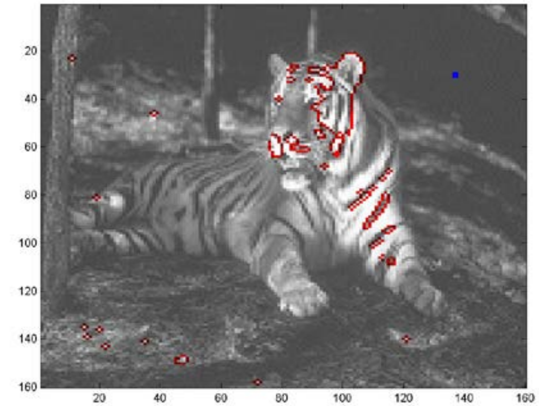
Dorit Hochbaum UC Berkeley

Subjective Visual Segmentation Quality Comparison (cont.)

- q-Normalized Cut (Entropy) (cont.)

Combinatorial



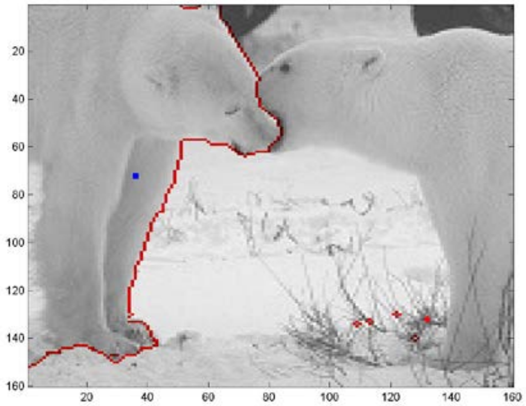


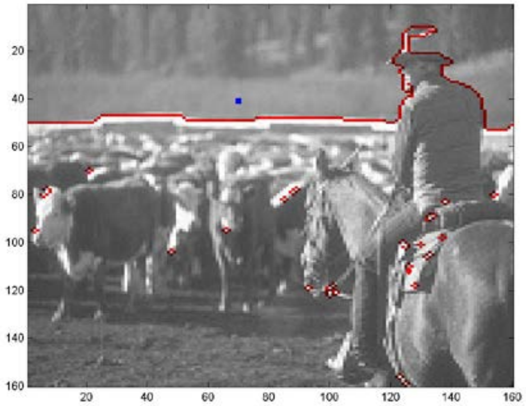
Spectral

The benefit of nested cuts in providing better segmentation quality

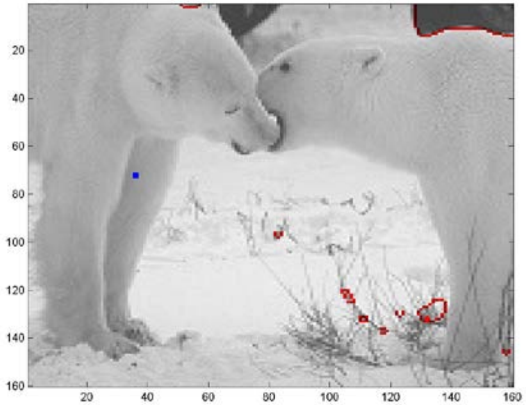
- Normalized Cut

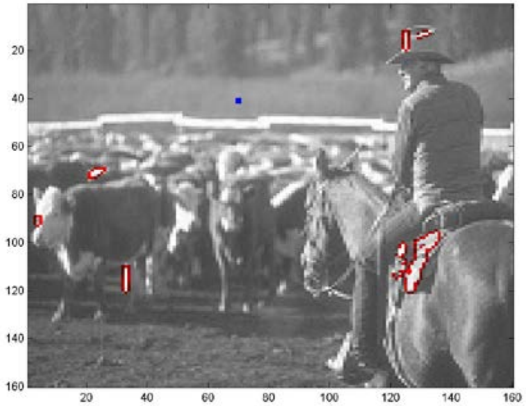
Cut presenting subjectively better visual segmentation





Cut minimizing objective function value

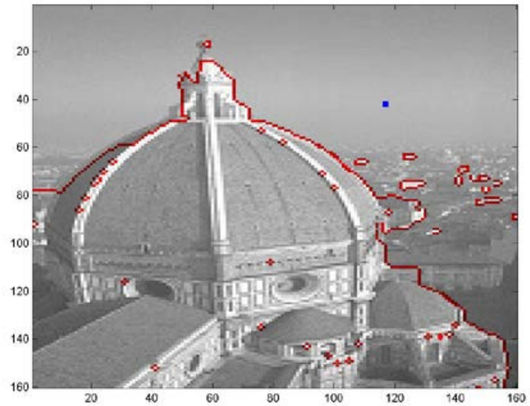


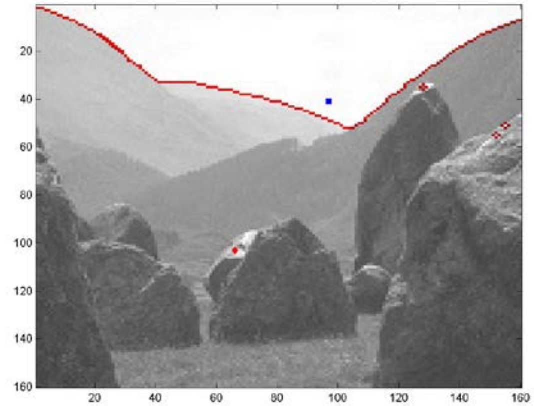


The benefit of nested cuts in providing better segmentation quality (cont.)

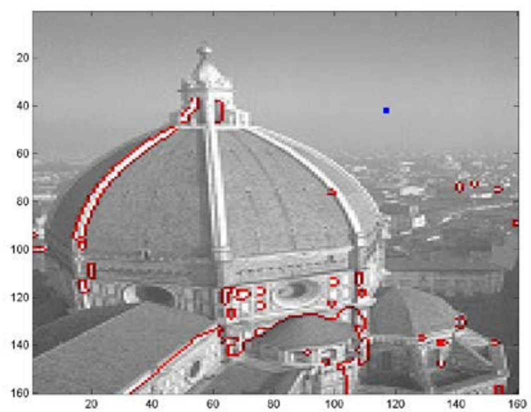
- Normalized Cut (cont.)

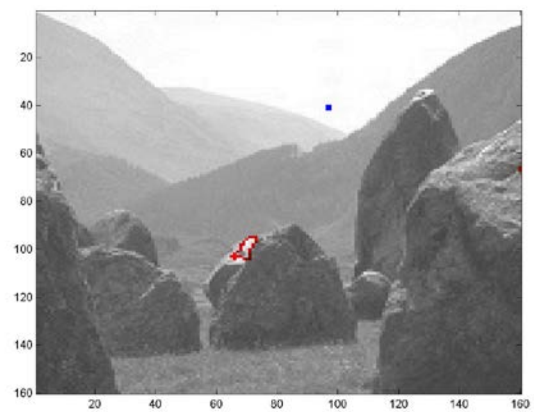
Cut presenting subjectively better visual segmentation





Cut minimizing objective function value



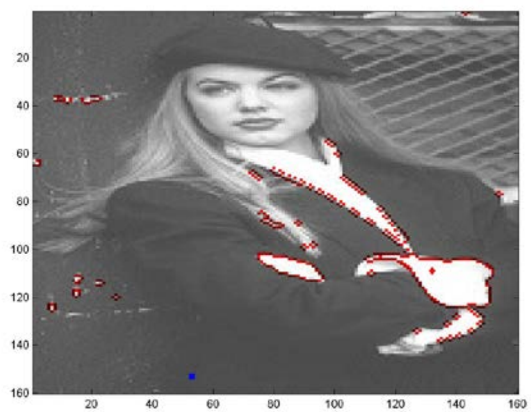


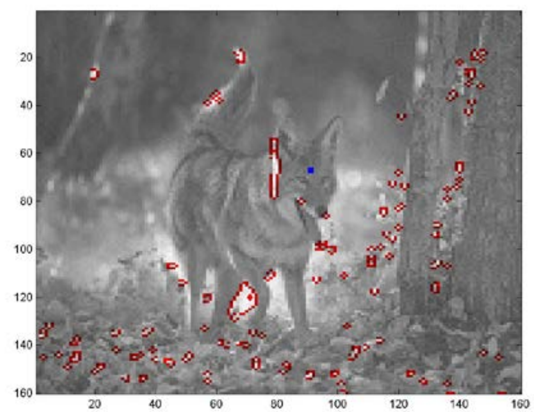
The benefit of nested cuts in providing better segmentation quality (cont.)

- q-Normalized Cut (Entropy)

Cut presenting subjectively better visual segmentation

Cut minimizing objective function value

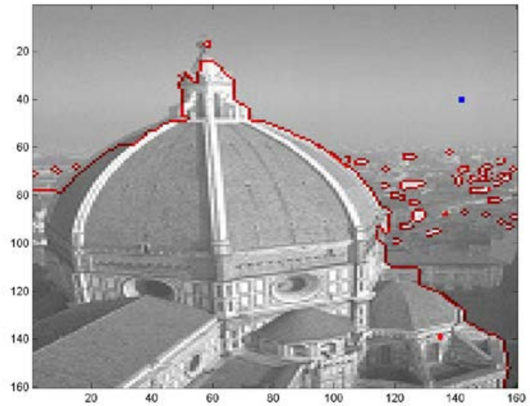


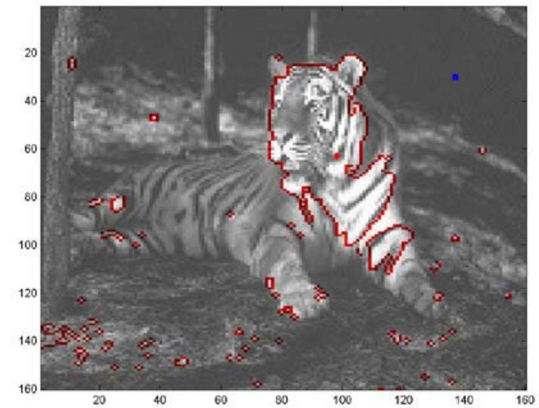


The benefit of nested cut in providing better segmentation quality (cont.)

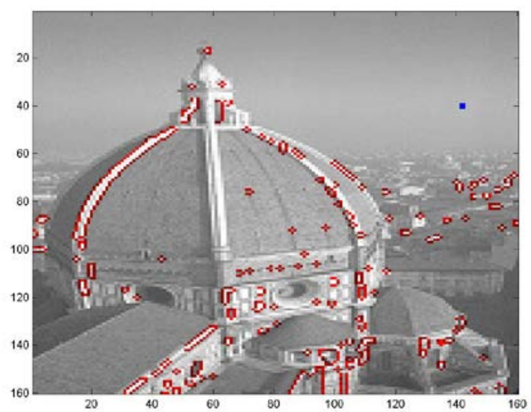
- q-Normalized Cut (Entropy) (cont.)

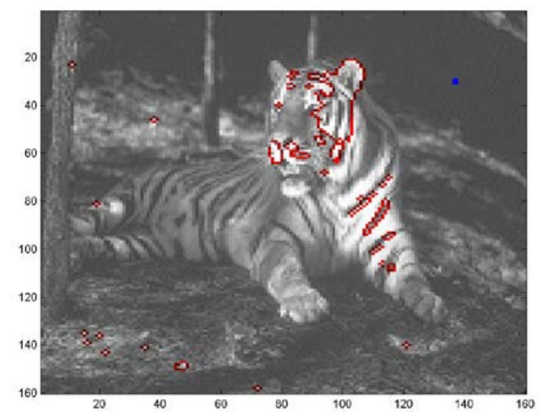
Cut presenting subjectively better visual segmentation





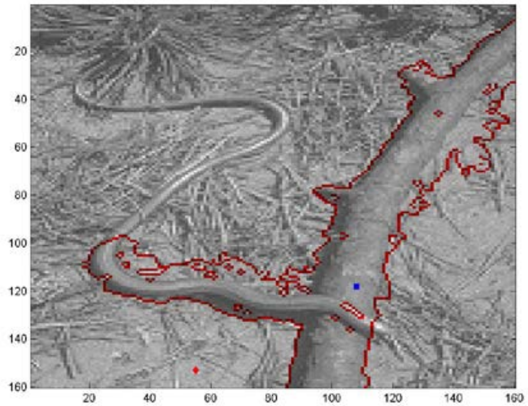
Cut minimizing objective function value

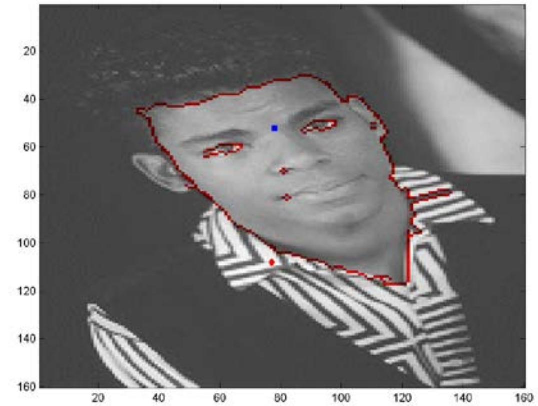




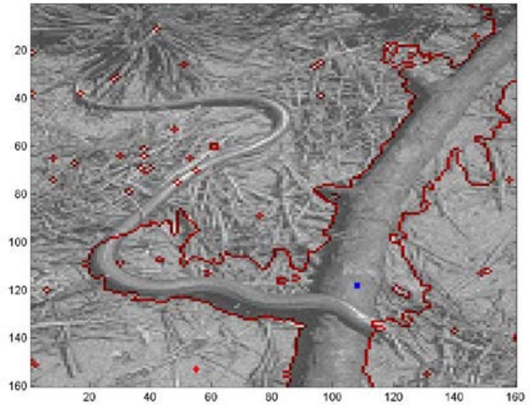
The benefit of defining node weights as entropy

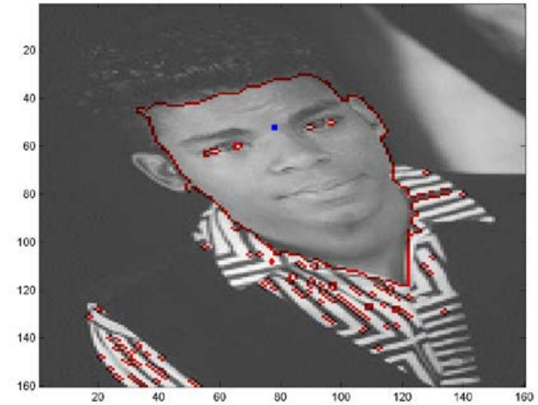
Cut presenting subjectively best visual segmentation using q-normalized cut





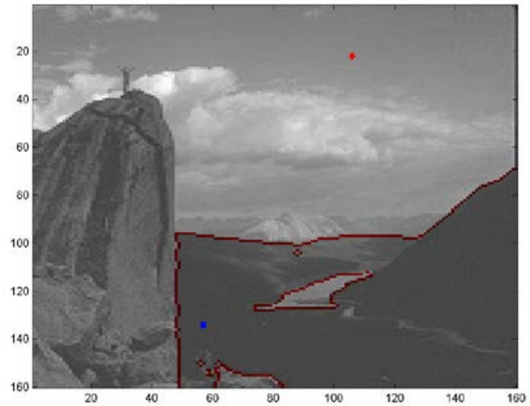
Cut presenting subjectively best visual segmentation using normalized cut



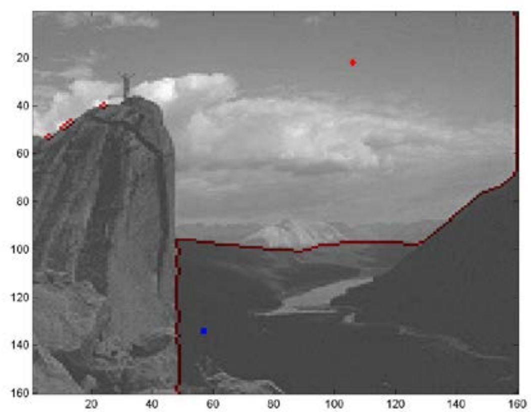


The benefit of defining node weights as entropy (cont.)

Cut presenting subjectively best visual segmentation using q-normalized cut



Cut presenting subjectively best visual segmentation using normalized cut



Conclusions

- The combinatorial technique provides better visual results in image segmentation.
- The combinatorial technique is faster than the spectral method (and requires substantially less storage)
- The combinatorial technique gives, on average, better quality solutions to several clustering problems.
- We used $$\min_{S \subset V} \frac{C_1(S,\bar{S})}{C_2(S,S)}$$ for: gene expression; knee cartilage volume computation (OA); pattern recognition; video tracking; enhancing nuclear detectors capabilities; drug efficacy studies, and general data mining.

References

1. Dorit S. Hochbaum.
Polynomial time algorithms for ratio regions and a variant of normalized cut.
IEEE Transactions on Pattern Analysis and Machine Intelligence, 32 (5),
pp. 889-898, May 2010.
2. Dorit S. Hochbaum.
A polynomial time algorithm for Rayleigh ratio on discrete variables:
Replacing spectral techniques for expander ratio, normalized cut and Cheeger constant.
Operations Research , Vol. 61, No. 1, January-February 2013, pp. 184-198.
3. Dorit S. Hochbaum, Cheng Lu, Erik Bertelli.
Evaluating Performance of Image Segmentation Criteria and Techniques.
EURO Journal on Computational Optimization, Vol 1:1-2, May 2013, pp. 155-180.

Questions

Prof. Dorit S. Hochbaum
hochbaum@ieor.berkeley.edu
http://www.ieor.berkeley.edu/~hochbaum/

Lemma 1's Proof

- Proof:

$$y_i = \begin{cases} 1 & \text{if } i \in S \\ -b & \text{if } i \in T = \bar{S} \end{cases}$$

$$y^T Q y = q(S) + b^2 q(T) \qquad y^T Q \mathbf{1} = 0 \Leftrightarrow b = \frac{q(S)}{q(T)}$$

$$y^T \mathcal{L} y = y^T D y - y^T W y$$

$$= \sum_{i \in S} d_i + b^2 \sum_{i \in \bar{S}} d_i - [C(S,S) - 2bC(S,\bar{S}) + b^2 C(\bar{S},\bar{S})]$$

$$= C(S,S) + C(S,\bar{S}) + b^2 C(S,\bar{S}) + b^2 C(\bar{S},\bar{S})$$

$$-[C(S,S) - 2bC(S,\bar{S}) + b^2 C(\bar{S},\bar{S})]$$

$$= (1 + b^2 + 2b) C(S,\bar{S}) = (1+b)^2 C(S,\bar{S})$$

Lemma 1's Proof

$$y^T \mathcal{L} y = (1+b)^2 C(S,\bar{S}) \qquad y^T Q y = q(S) + b^2 q(T)$$

$$\min_{\substack{y^T Q \vec{1}=0 \\ y\in\{-b,1\}}} \frac{y^T \mathcal{L} y}{y^T Q y} = \min_{\substack{y^T Q \vec{1}=0 \\ S\subset V}} \frac{(1+b)^2 C(S,\bar{S})}{q(S)+b^2 q(\bar{S})}$$

$$y^T Q \mathbf{1} = 0 \Leftrightarrow b = \frac{q(S)}{q(T)}$$

$$\frac{(1+b)^2 C(S,\bar{S})}{q(S)+b^2 q(\bar{S})} = \frac{(1+\frac{q(S)}{q(\bar{S})})^2 C(S,\bar{S})}{q(S)+(\frac{q(S)}{q(\bar{S})})^2 q(\bar{S})} = \frac{(1+\frac{q(S)}{q(\bar{S})})^2 C(S,\bar{S})}{q(S)(1+\frac{q(S)}{q(\bar{S})})}$$

$$= \frac{(1+\frac{q(S)}{q(\bar{S})}) C(S,\bar{S})}{q(S)}.$$

Lemma 1's proof

$$\min_{\substack{y^T Q \vec{1}=0 \\ y\in\{-b,1\}}} \frac{y^T \mathcal{L} y}{y^T Q y} = \min_{S\subset V} C(S,\bar{S})[\frac{1}{q(S)} + \frac{1}{q(\bar{S})}]$$

$\square$