Spectral Clustering based on the graph p-Laplacian

Thomas Bühler and Matthias Hein

Saarland University, Germany
Machine Learning Group

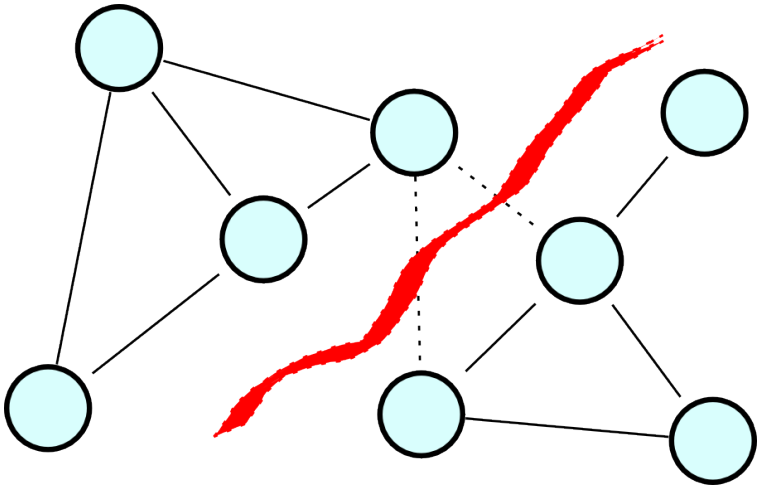
June 15, 2009

Clustering as graph partitioning

Given: any data with some similarity measure
Goal: divide data into subsets that optimize some clustering objective

Clustering as graph partitioning

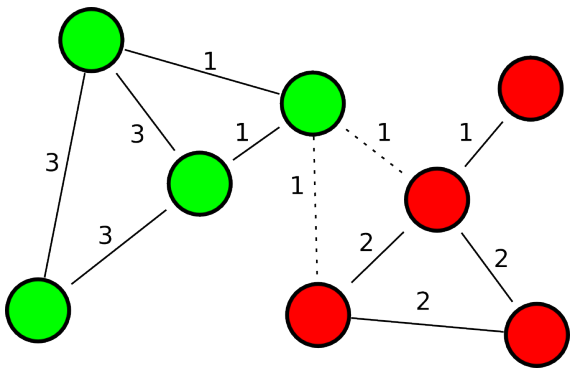
Given: any data with some similarity measure
Goal: divide data into subsets that optimize some clustering objective



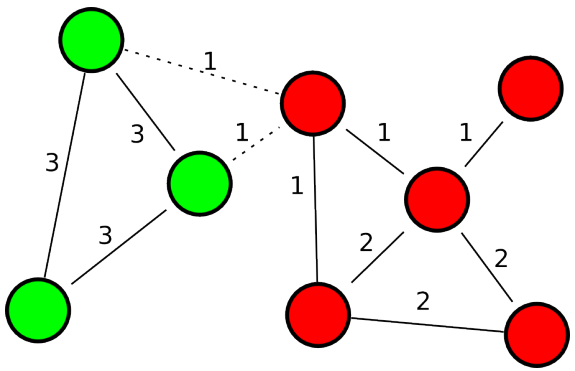
- Representation of data as undirected, weighted Graph $G(V, E)$ with edges E and vertices V
- Weight matrix W encodes similarity
- Clustering problem can be formulated as graph partitioning problem

Balanced graph cut criteria

Balancing of cardinality:



Balancing of volume:



Ratio cut (Hagen & Kahng, 91)

$$\mathrm{RCut}(C,\overline{C}) = \frac{\mathrm{cut}(C,\overline{C})}{|C|} + \frac{\mathrm{cut}(C,\overline{C})}{|\overline{C}|}$$

Normalized cut (Shi & Malik, 00)

$$\mathrm{NCut}(C,\overline{C}) = \frac{\mathrm{cut}(C,\overline{C})}{\mathrm{vol}(C)} + \frac{\mathrm{cut}(C,\overline{C})}{\mathrm{vol}(\overline{C})}$$

Ratio Cheeger cut

$$\mathrm{RCC}(C,\overline{C}) = \frac{\mathrm{cut}(C,\overline{C})}{\min\{|C|,|\overline{C}|\}}$$

Normalized Cheeger cut

$$\mathrm{NCC}(C,\overline{C}) = \frac{\mathrm{cut}(C,\overline{C})}{\min\{\mathrm{vol}(C),\mathrm{vol}(\overline{C})\}}$$

where $\mathrm{vol}(C) = \sum_{i\in C} d_i$ and $d_i = \sum_{j\in V} w_{ij}$.

Spectral Clustering as relaxation of balanced graph cuts

Reformulation of RCut:
For any partition $C, \overline{C}$ define the function f_C as

$$(f_C)_i = \begin{cases} \frac{1}{|C|} & , i \in C \\ -\frac{1}{|\overline{C}|} & , i \notin C \end{cases} .$$

It holds for the well-known **unnormalized graph Laplacian** $\Delta_2 = D - W$:

$$\mathrm{RCut}(C, \overline{C}) = F_2(f_C) \qquad \text{where } F_2(f) = \frac{\langle f, \Delta_2 f \rangle}{\|f\|_2^2} .$$

Spectral Clustering as relaxation of balanced graph cuts

Reformulation of RCut:
For any partition $C, \overline{C}$ define the function f_C as

$$(f_C)_i = \begin{cases} \frac{1}{|C|} & , i \in C \\ -\frac{1}{|\overline{C}|} & , i \notin C \end{cases} .$$

It holds for the well-known **unnormalized graph Laplacian** $\Delta_2 = D - W$:

$$\mathrm{RCut}(C, \overline{C}) = F_2(f_C) \qquad \text{where } F_2(f) = \frac{\langle f, \Delta_2 f \rangle}{\|f\|_2^2} .$$

Relaxation of RCut:

$$v^{(2)} = \arg\min_{f \in \mathbb{R}^V} \left\{ \frac{\langle f, \Delta_2 f \rangle}{\|f\|_2^2} \mid \langle f, \mathbf{1} \rangle = 0 \right\} .$$

Quality guarantees on the standard relaxation

From the eigenvector $v^{(2)}$ to a partition $C, \overline{C}$:

- Thresholding: $C = \underset{C_t = \{i \in V \mid v^{(2)}(i) > t\}}{\arg\min} \mathrm{RCC}(C_t, \overline{C_t})$.

Quality guarantees on the standard relaxation

From the eigenvector $v^{(2)}$ to a partition $C, \overline{C}$:

- Thresholding: $C = \underset{C_t = \{i \in V \mid v^{(2)}(i) > t\}}{\arg\min} \mathrm{RCC}(C_t, \overline{C_t})$.

How good is this partition?

Quality guarantees on the standard relaxation

From the eigenvector $v^{(2)}$ to a partition $C, \overline{C}$:

- Thresholding: $C = \underset{C_t = \{i \in V \mid v^{(2)}(i) > t\}}{\arg\min} \mathrm{RCC}(C_t, \overline{C_t})$.

How good is this partition?

- Standard isoperimetric inequality (unnormalized case):

$$\frac{h_{\mathrm{RCC}}^2}{2 \max_i d_i} \leq \lambda_2^{(2)} \leq 2\, h_{\mathrm{RCC}} \qquad \text{(see Chung, 97)},$$

where $h_{\mathrm{RCC}} = \inf_C \left\{ \mathrm{RCC}(C, \overline{C}) \right\}$.

Quality guarantees on the standard relaxation

From the eigenvector $v^{(2)}$ to a partition $C, \overline{C}$:

- Thresholding: $C = \underset{C_t = \{i \in V \mid v^{(2)}(i) > t\}}{\arg\min} \mathrm{RCC}(C_t, \overline{C_t})$.

How good is this partition?

- Standard isoperimetric inequality (unnormalized case):

$$\frac{h_{\mathrm{RCC}}^2}{2 \max_i d_i} \leq \lambda_2^{(2)} \leq 2\, h_{\mathrm{RCC}} \qquad \text{(see Chung, 97)},$$

where $h_{\mathrm{RCC}} = \inf_C \left\{ \mathrm{RCC}(C, \overline{C}) \right\}$.

- Relation to Cheeger cut h^*_{RCC} after thresholding:

$$\frac{h_{\mathrm{RCC}}}{\max_{i \in V} d_i} \leq \frac{h^*_{\mathrm{RCC}}}{\max_{i \in V} d_i} \leq 2 \left(\frac{h_{\mathrm{RCC}}}{\max_{i \in V} d_i} \right)^{\frac{1}{2}} .$$

Quality guarantees on the standard relaxation

From the eigenvector $v^{(2)}$ to a partition $C, \overline{C}$:

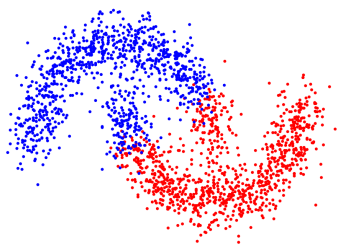
- Thresholding: $C = \underset{C_t = \{i \in V \mid v^{(2)}(i) > t\}}{\arg\min} \mathrm{RCC}(C_t, \overline{C_t})$.

How good is this partition?

- Standard isoperimetric inequality (unnormalized case):

$$\frac{h_{\mathrm{RCC}}^2}{2 \max_i d_i} \leq \lambda_2^{(2)} \leq 2\, h_{\mathrm{RCC}} \qquad \text{(see Chung, 97)},$$

where $h_{\mathrm{RCC}} = \inf_C \left\{ \mathrm{RCC}(C, \overline{C}) \right\}$.



- Relation to Cheeger cut h^*_{RCC} after thresholding:

$$\frac{h_{\mathrm{RCC}}}{\max_{i \in V} d_i} \leq \frac{h^*_{\mathrm{RCC}}}{\max_{i \in V} d_i} \leq 2 \left(\frac{h_{\mathrm{RCC}}}{\max_{i \in V} d_i} \right)^{\frac{1}{2}} .$$

Quality guarantees on the standard relaxation

From the eigenvector $v^{(2)}$ to a partition $C, \overline{C}$:

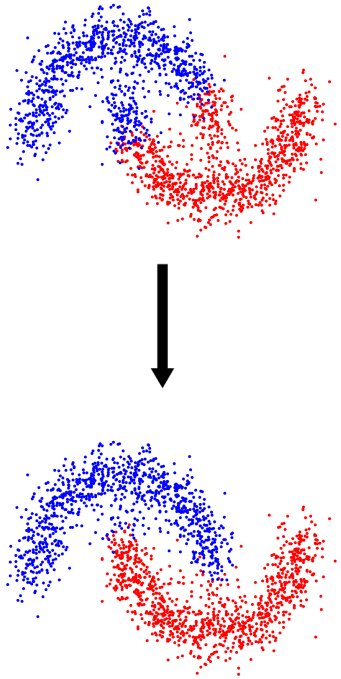
- Thresholding: $C = \underset{C_t = \{i \in V \mid v^{(2)}(i) > t\}}{\arg\min} \mathrm{RCC}(C_t, \overline{C_t})$.

How good is this partition?

- Standard isoperimetric inequality (unnormalized case):

$$\frac{h_{\mathrm{RCC}}^2}{2 \max_i d_i} \leq \lambda_2^{(2)} \leq 2\, h_{\mathrm{RCC}} \qquad \text{(see Chung, 97)},$$

where $h_{\mathrm{RCC}} = \inf_C \left\{ \mathrm{RCC}(C, \overline{C}) \right\}$.



- Relation to Cheeger cut h^*_{RCC} after thresholding:

$$\frac{h_{\mathrm{RCC}}}{\max_{i \in V} d_i} \leq \frac{h^*_{\mathrm{RCC}}}{\max_{i \in V} d_i} \leq 2 \left(\frac{h_{\mathrm{RCC}}}{\max_{i \in V} d_i} \right)^{\frac{1}{2}} .$$

The graph p-Laplacian

Quadratic form induced by the standard graph Laplacian:

$$\langle f, \Delta_2 f \rangle = \frac{1}{2} \sum_{i,j \in V} w_{ij} (f_i - f_j)^2.$$

The graph p-Laplacian

Quadratic form induced by the standard graph Laplacian:

$$\langle f, \Delta_2 f \rangle = \frac{1}{2} \sum_{i,j \in V} w_{ij} (f_i - f_j)^2.$$

What is the operator Δ_p that for $p \geq 1$ induces the general form

$$\langle f, \Delta_p f \rangle = \frac{1}{2} \sum_{i,j \in V} w_{ij} |f_i - f_j|^p \ ?$$

The graph p-Laplacian

Quadratic form induced by the standard graph Laplacian:

$$\langle f, \Delta_2 f \rangle = \frac{1}{2} \sum_{i,j \in V} w_{ij} (f_i - f_j)^2.$$

What is the operator Δ_p that for $p \geq 1$ induces the general form

$$\langle f, \Delta_p f \rangle = \frac{1}{2} \sum_{i,j \in V} w_{ij} |f_i - f_j|^p \ ?$$

The unnormalized graph p-Laplacian

$$(\Delta_p f)_i = \sum_{j \in V} w_{ij} \, \phi_p \left(f_i - f_j \right)$$

where $\phi_p(x) = |x|^{p-1} \operatorname{sign}(x)$

The graph p-Laplacian

Quadratic form induced by the standard graph Laplacian:

$$\langle f, \Delta_2 f \rangle = \frac{1}{2} \sum_{i,j \in V} w_{ij} (f_i - f_j)^2.$$

What is the operator Δ_p that for $p \geq 1$ induces the general form

$$\langle f, \Delta_p f \rangle = \frac{1}{2} \sum_{i,j \in V} w_{ij} |f_i - f_j|^p \ ?$$

The unnormalized graph p-Laplacian

$$(\Delta_p f)_i = \sum_{j \in V} w_{ij} \, \phi_p \left(f_i - f_j \right)$$

where $\phi_p(x) = |x|^{p-1} \operatorname{sign}(x)$

Case $p = 2$:

$$(\Delta_2 f)_i = \sum_{j \in V} w_{ij} \, \left(f_i - f_j \right)$$

Variational characterization of second eigenvalue

Eigenvalue $\lambda_p \in \mathbb{R}$ and **eigenvector** $v_p \in \mathbb{R}^V$ of the p-Laplacian Δ_p:

$$\forall i \in V: \qquad (\Delta_p v_p)_i = \lambda_p \, \phi_p\left((v_p)_i\right) , \qquad \phi_p(x) = |x|^{p-1} \operatorname{sign}(x) .$$

Motivation: Eigenvectors as critical points of $F_p(f) = \frac{\left\langle f, \Delta_p f \right\rangle}{\|f\|_p^p}$

Variational characterization of second eigenvalue

Eigenvalue $\lambda_p \in \mathbb{R}$ and **eigenvector** $v_p \in \mathbb{R}^V$ of the p-Laplacian Δ_p:

$$\forall i \in V: \qquad (\Delta_p v_p)_i = \lambda_p \, \phi_p\left((v_p)_i\right) , \qquad \phi_p(x) = |x|^{p-1} \operatorname{sign}(x) .$$

Motivation: Eigenvectors as critical points of $F_p(f) = \frac{\langle f, \Delta_p f \rangle}{\|f\|_p^p}$

Variational characterization of second eigenvalue (Amghibech, 2003)

$$F_p^{(2)}(f) = \frac{\langle f, \Delta_p f \rangle}{\min_{c \in \mathbb{R}} \|f - c\mathbf{1}\|_p^p} = \frac{\frac{1}{2} \sum_{i,j \in V} w_{ij} |f_i - f_j|^p}{\min_{c \in \mathbb{R}} \|f - c\mathbf{1}\|_p^p}$$

Second eigenvalue:
$\lambda_p^{(2)} = \min_{f \in \mathbb{R}^n} F_p^{(2)}(f)$

Second eigenvector:
computed from $\arg\min_{f \in \mathbb{R}^n} F_p^{(2)}(f)$

p-Spectral Relaxation of balanced graph cuts

Analogously to the case $p = 2$, for each partition $C, \overline{C}$, there exists a function $f_{p,C}$ such that

$$F_p^{(2)}(f_{p,C}) = \mathrm{cut}(C, \overline{C}) \left| \frac{1}{|C|^{\frac{1}{p-1}}} + \frac{1}{|\overline{C}|^{\frac{1}{p-1}}} \right|^{p-1}$$

p-Spectral Relaxation of balanced graph cuts

Analogously to the case $p = 2$, for each partition $C, \overline{C}$, there exists a function $f_{p,C}$ such that

$$F_p^{(2)}(f_{p,C}) = \mathrm{cut}(C, \overline{C}) \left| \frac{1}{|C|^{\frac{1}{p-1}}} + \frac{1}{|\overline{C}|^{\frac{1}{p-1}}} \right|^{p-1}$$

Special cases:

$$F_2^{(2)}(f_{2,C}) = \mathrm{RCut}(C, \overline{C}),$$
$$\lim_{p \to 1} F_p^{(2)}(f_{p,C}) = \mathrm{RCC}(C, \overline{C}).$$

p-Spectral Relaxation of balanced graph cuts

Analogously to the case $p = 2$, for each partition $C, \overline{C}$, there exists a function $f_{p,C}$ such that

$$F_p^{(2)}(f_{p,C}) = \mathrm{cut}(C, \overline{C}) \left| \frac{1}{|C|^{\frac{1}{p-1}}} + \frac{1}{|\overline{C}|^{\frac{1}{p-1}}} \right|^{p-1}$$

Special cases:

$$F_2^{(2)}(f_{2,C}) = \mathrm{RCut}(C, \overline{C}),$$
$$\lim_{p \to 1} F_p^{(2)}(f_{p,C}) = \mathrm{RCC}(C, \overline{C}).$$

The second p-eigenvector is a relaxation of the above problem.

Bound in terms of the optimal cut

From the p-eigenvector $v_p^{(2)}$ to a partition $C, \overline{C}$:

- Thresholding: $C = \underset{C_t = \{i \in V \mid v_p^{(2)}(i) > t\}}{\arg\min} \mathrm{RCC}(C_t, \overline{C_t})$.

Bound in terms of the optimal cut

From the p-eigenvector $v_p^{(2)}$ to a partition $C, \overline{C}$:

- Thresholding: $C = \underset{C_t = \{i \in V \mid v_p^{(2)}(i) > t\}}{\arg\min} \mathrm{RCC}(C_t, \overline{C_t})$.

Found Cheeger cut h^*_{RCC} vs. optimal Cheeger cut h_{RCC}

$$\frac{h_{\mathrm{RCC}}}{\max_{i \in V} d_i} \leq \frac{h^*_{\mathrm{RCC}}}{\max_{i \in V} d_i} \leq p \left(\frac{h_{\mathrm{RCC}}}{\max_{i \in V} d_i} \right)^{\frac{1}{p}}$$

Bound in terms of the optimal cut

From the p-eigenvector $v_p^{(2)}$ to a partition $C, \overline{C}$:

- Thresholding: $C = \underset{C_t = \{i \in V \mid v_p^{(2)}(i) > t\}}{\arg\min} \mathrm{RCC}(C_t, \overline{C_t})$.

Found Cheeger cut h^*_{RCC} vs. optimal Cheeger cut h_{RCC}

$$\frac{h_{\mathrm{RCC}}}{\max_{i \in V} d_i} \leq \frac{h^*_{\mathrm{RCC}}}{\max_{i \in V} d_i} \leq p \left(\frac{h_{\mathrm{RCC}}}{\max_{i \in V} d_i} \right)^{\frac{1}{p}}$$

The cut found by p-Spectral Clustering converges to the optimal Cheeger cut as $p \to 1$

Consecutive minimization of the functional $F_p^{(2)}$

Problem: Direct minimization of $F_p^{(2)}$ leads often to fast convergence to non-optimal local minimum.

Consecutive minimization of the functional $F_p^{(2)}$

Problem: Direct minimization of $F_p^{(2)}$ leads often to fast convergence to non-optimal local minimum.

Idea:

- Solve sequence of minimization problems

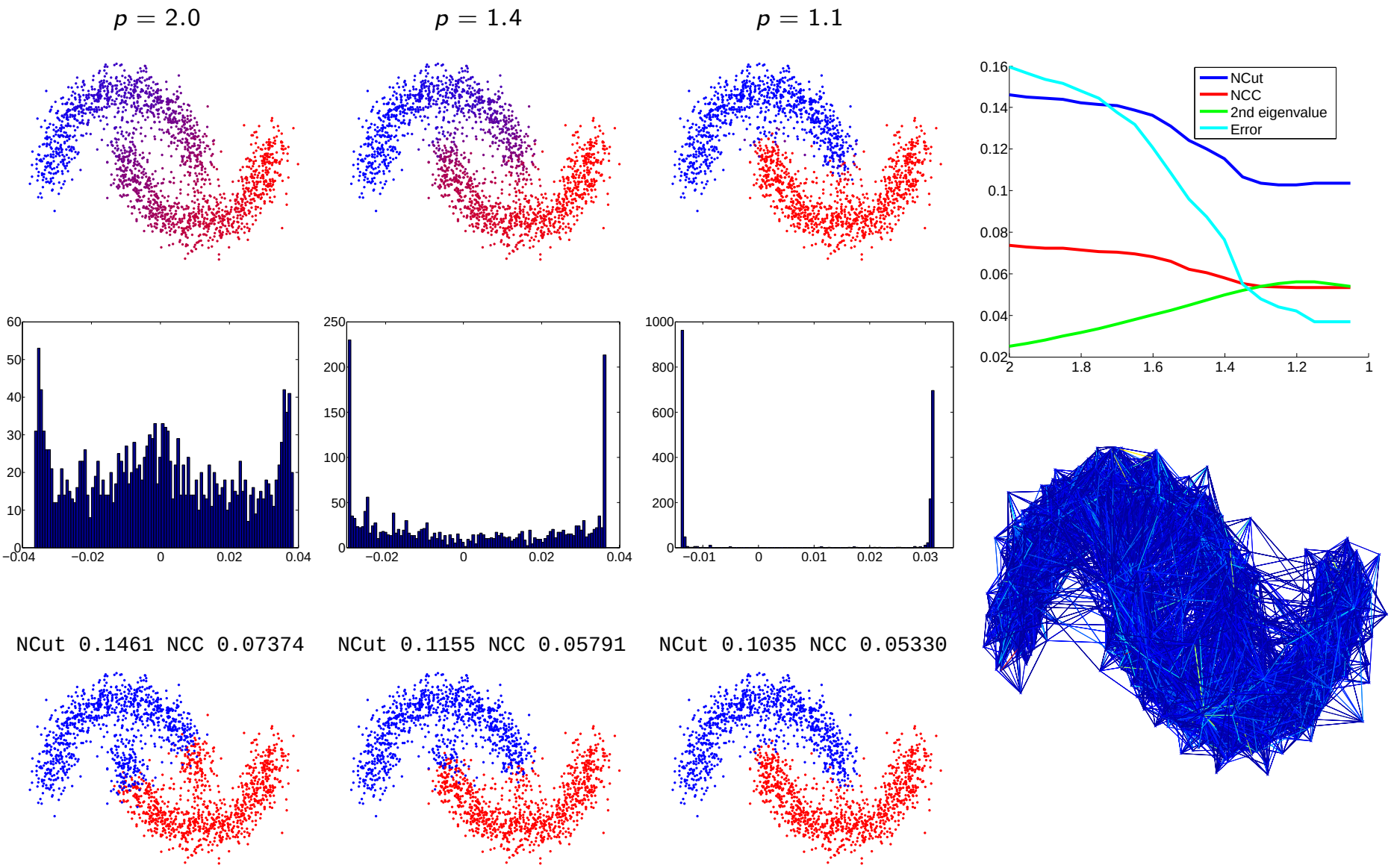
$$F_{p_0}^{(2)}, F_{p_1}^{(2)}, ..., F_p^{(2)}, \quad \text{with } p_0 = 2 > p_1 > ... > p \ ,$$

  and each step is initialized with the solution of the previous step.
- The subproblems are minimized via approximate Newton steps.

Motivation:

- Global minimizer for $p = 2$ is second eigenvector of standard graph Laplacian.
- As $F_p^{(2)}$ continuous in p: Local minimizer of $F_{p_1}^{(2)}$ should be close to local minimizer of $F_{p_2}^{(2)}$ if p_1 close to p_2.
- Superlinear convergence of Newton-like methods close to a local optimum.

High dimensional noisy two moons



Multiway clustering: USPS/MNIST

- More than two clusters are obtained via recursive clustering scheme
- Multi-partition criterion: $\mathrm{RCut}(C_1, \ldots, C_k) = \sum_{i=1}^{k} \frac{\mathrm{cut}(C_i, \overline{C_i})}{|C_i|}$

Multiway clustering: USPS/MNIST

- More than two clusters are obtained via recursive clustering scheme
- Multi-partition criterion: $\mathrm{RCut}(C_1, \ldots, C_k) = \sum_{i=1}^{k} \frac{\mathrm{cut}(C_i, \overline{C_i})}{|C_i|}$

Results for the full USPS and MNIST datasets:

	USPS		MNIST	
p	RCut	Error	RCut	Error
2.0	0.819	0.233	0.225	0.189
1.9	0.741	0.142	0.209	0.172
1.8	0.718	0.141	0.186	0.170
1.7	0.698	0.139	0.170	0.169
1.6	0.684	0.134	0.164	0.170
1.5	0.676	0.133	0.161	0.133
1.4	0.693	0.141	0.158	0.132
1.3	0.684	0.138	0.155	0.131
1.2	0.679	0.137	0.153	0.129

Multiway clustering: USPS/MNIST

- More than two clusters are obtained via recursive clustering scheme
- Multi-partition criterion: $\mathrm{RCut}(C_1, \ldots, C_k) = \sum_{i=1}^{k} \frac{\mathrm{cut}(C_i, \overline{C_i})}{|C_i|}$

Results for the full USPS and MNIST datasets:

	USPS		MNIST	
p	RCut	Error	RCut	Error
2.0	0.819	0.233	0.225	0.189
1.9	0.741	0.142	0.209	0.172
1.8	0.718	0.141	0.186	0.170
1.7	0.698	0.139	0.170	0.169
1.6	0.684	0.134	0.164	0.170
1.5	0.676	0.133	0.161	0.133
1.4	0.693	0.141	0.158	0.132
1.3	0.684	0.138	0.155	0.131
1.2	0.679	0.137	0.153	0.129

The price you pay for better cuts:

Runtime USPS	
p	t / sec
2.0	10
1.8	99
1.6	224
1.4	1147
1.2	4660

Conclusion

The graph p-Laplacian and its eigenvectors

- Variational characterization of the eigenvectors of the p-Laplacian
- Efficient numerical scheme to compute the second eigenvector

Conclusion

The graph p-Laplacian and its eigenvectors

- Variational characterization of the eigenvectors of the p-Laplacian
- Efficient numerical scheme to compute the second eigenvector

Theoretical justification of p-Spectral Clustering

- p-Spectral Clustering as relaxation of balanced graph cuts
- Convergence of cut found by p-Spectral Clustering towards the optimal Cheeger cut as $p \to 1$

Conclusion

The graph p-Laplacian and its eigenvectors

- Variational characterization of the eigenvectors of the p-Laplacian
- Efficient numerical scheme to compute the second eigenvector

Theoretical justification of p-Spectral Clustering

- p-Spectral Clustering as relaxation of balanced graph cuts
- Convergence of cut found by p-Spectral Clustering towards the optimal Cheeger cut as $p \to 1$

Experimental Evaluation

- Strong improvement in the clustering result for decreasing values of p