Multiclass Spectral Clustering

Stella X. Yu Carnegie Mellon University
Jianbo Shi University of Pennsylvania

A principled account on finding discrete near-global optima for spectral clustering methods.
**$K$-Way Normalized Cuts**

\[
\max \quad \text{knassoc}(\Gamma^K_V) = \frac{1}{K} \sum_{l=1}^{K} \text{linkratio}(\mathcal{V}_l, \mathcal{V}_l)
\]

\[
\min \quad \text{kncuts}(\Gamma^K_V) = \frac{1}{K} \sum_{l=1}^{K} \text{linkratio}(\mathcal{V}_l, \mathcal{V} \setminus \mathcal{V}_l)
\]
A Principled Solution to Normalized Cuts

\[
\max \; \text{knassoc}(\Gamma^K_V) = \frac{1}{K} \sum_{l=1}^{K} \text{linkratio}(V_l, V_l)
\]

NP complete even for \( K = 2 \) and planar graphs

Fast solution to find near-global optima:

1. Find global optima in the relaxed continuous domain
   optima = eigenvectors \( \times \) orthonormal transforms

2. Find a discrete solution closest to continuous optima
   closeness = measured in \( L_2 \) norm between solutions
Solution Diagram

Final solution: \((X^*(2), \tilde{X}^*(2))\)
Representation

- Partition matrix

$$X = [X_1, \ldots, X_K]$$

- Maximize

$$\varepsilon(X) = \frac{1}{K} \sum_{l=1}^{K} \frac{\text{links}(\mathcal{V}_l, \mathcal{V}_l)}{\text{degree}(\mathcal{V}_l)} = \frac{1}{K} \sum_{l=1}^{K} \frac{X_l^T W X_l}{X_l^T D X_l}$$

- Subject to

  binary \hspace{1cm} X \in \{0, 1\}^{N \times K}

  exclusion \hspace{1cm} X 1_K = 1_N
Step 1: Find Continuous Global Optima

- Eigensolution \((V, S)\) that optimizes:

\[
\text{maximize} \quad \varepsilon(Z) = \frac{1}{K} \text{tr}(Z^T W Z)
\]

subject to \(Z^T D Z = I_K\)

- Scaled partition matrix \(Z\):

\[
Z = f(X) = X \left( X^T D X \right)^{-\frac{1}{2}}
\]

\[
X = f^{-1}(Z) = \text{Diag}(\text{diag}^{-\frac{1}{2}}(ZZ^T))Z
\]

- Set of all continuous optima:

\[
\{ \tilde{X}^* R : \tilde{X}^* = \text{Diag}(\text{diag}^{-\frac{1}{2}}(VV^T))V, \quad R^T R = I_K \}
\]
Step 2: Discretize Continuous Optima

- Find a partitioning closest to continuous optima

\[
\begin{align*}
\text{minimize} & \quad \phi(X, R) = \|X - \tilde{X}^* R\|^2 \\
\text{subject to} & \quad R^T R = I_K, \quad X \in \{0, 1\}^{N \times K}, \quad X 1_K = 1_N.
\end{align*}
\]

- This bilinear program can be solved iteratively:
  1. Given a continuous solution \( \tilde{X} = \tilde{X}^* R^* \), solve \( X^* \) by:

\[
X^*(i, l) = \text{istrue}(l = \arg \max_k \tilde{X}(i, k)), \quad i \in \mathbb{V}.
\]

  2. Given a discrete solution \( X^* \), solve \( R^* \) by:

\[
R^* = \tilde{U} U^T, \quad X^T \tilde{X}^* = U \Omega \tilde{U}^T, \quad \Omega = \text{Diag}(\omega).
\]
Bipartitioning of A Point Set

\[ Z^* = [V_1, V_2] \]

\[ \tilde{X}^* = f^{-1}(Z^*) \]

\[ \tilde{X}^* R^*(1) \]

\[ \tilde{X}^* R^*(0) \]
Pixel Similarity based on Intensity Edges

image oriented filter pairs edge magnitudes
Discrete Near-Global Optima

\( K = 4 : 0.9901 \quad 0.9899 \quad 0.9881 \)

a few;
all good
Multiclass Real Image Segmentation
Summary

$K$-way normalized cuts have:
- eigendecomposition for continuous global optima
- bilinear iterations for discrete near-global optima.

New understanding on the eigenvectors:
- a basis for generating all optima
- the first eigenvector is as important
- approximating \emph{scaled} partition matrices
- $K$ eigenvectors for optimal $K$-way partitioning.

New understanding on discretization:
- continuous and discrete optima in a pair
- a bilinear program solved by alternating SVD and NMS
- fast, robust, and guaranteed near-global optimality.