Flexible Constrained Spectral Clustering

Xiang Wang and Ian Davidson
Knowledge Discovery and Data Mining Lab
Department of Computer Science
University of California, Davis
Spectral Clustering and Constraints

- Spectral clustering is an important clustering algorithm with wide applications.
- Spectral clustering is unsupervised and does not guarantee meaningful/actionable results.
- *Must-Link* and *Cannot-Link* constraints allow human experts to inject supervision into spectral clustering.

![Input image](image1.png) ![Unconstrained result](image2.png) ![Constrained result](image3.png)
Limitations of Previous Work

• They only encode hard constraints
• They try to satisfy each and every constraint
• However, many real-world applications give rise to soft constraints
  – Real-valued constraints as a confidence score
  – Inconsistent constraints from multiple sources
Our Contributions

• We propose a flexible framework for constrained spectral clustering
  – It can incorporate both hard and soft constraints
  – It uses a user-specified threshold to lower-bound how well the constraints should be satisfied

• We interpret our approach as finding the normalized min-cut of a labeled graph

• We empirically justify the effectiveness of our method with comparison to existing methods
Background: (Unconstrained) Spectral Clustering

| \( G \) | A graph with \( N \) nodes |
| \( A \) | The affinity matrix |
| \( D \) | The degree matrix |
| \( I \) | The identity matrix |
| \( L(L) \) | The (normalized) graph Laplacian |
| \( Q(Q) \) | The (normalized) constraint matrix |
| \( u(v) \) | The (normalized) cluster indicator vector |

Objective function (normalized min-cut in relaxed form):

\[
\arg \min_{v \in \mathbb{R}^N} v^T \bar{L} v,
\]

s.t. \( v^T v = \text{vol}(G), \ v \perp D^{1/2} 1 \).

Solution: The second smallest eigenvector of the graph Laplacian.
Encoding Constraints

We encode pairwise constraints in an $N \times N$ matrix $Q$:

$$Q_{ij} = Q_{ji} = \begin{cases} +1 & \text{if } ML(i, j) \\ -1 & \text{if } CL(i, j) \\ 0 & \text{no supervision available} \end{cases}.$$ 

Let $u$ be the cluster indicator vector:

$$u^T Q u = \sum_{i=1}^{N} \sum_{j=1}^{N} u_i u_j Q_{ij}$$

is a measure of how well the constraints in $Q$ are satisfied by $u$.

We relax $u$ and $Q$ to encode both hard and soft constraints:

$$u \in \mathbb{R}^N, Q \in \mathbb{R}^{N \times N}.$$
Our Objective Function

We append a new term to explicitly encode constraints, and turn unconstrained spectral clustering to constrained spectral clustering.

\[
\arg\min_{v \in \mathbb{R}^N} \mathbf{v}^T \bar{L} \mathbf{v}, \quad \text{s.t.} \quad \mathbf{v}^T \bar{Q} \mathbf{v} \geq \alpha, \quad \mathbf{v}^T \mathbf{v} = \text{vol}(G), \quad \mathbf{v} \neq D^{1/2} \mathbf{1}.
\]

- Finding normalized min-cut
- Normalizing \( v \)
- Ruling out the trivial solution

Lower-bounding how well the constraints in \( Q \) are satisfied
Solving the Objective Function (1)

Objective: \[
\arg \min_{v \in \mathbb{R}^N} v^T \bar{L}v, \text{ s.t. } v^T \bar{Q}v \geq \alpha, \ v^T v = \text{vol}(G),
\]

Introducing Lagrange multipliers:

\[\Lambda(v, \lambda, \mu) = v^T \bar{L}v - \lambda(v^T \bar{Q}v - \alpha) - \mu(v^T v - \text{vol}(G)).\]

Introducing Karush-Kuhn-Tucker (KKT) conditions:

(Stationarity) \[\bar{L}v - \lambda \bar{Q}v - \mu v = 0,\]

(Primal feasibility) \[v^T \bar{Q}v \geq \alpha, \ v^T v = \text{vol}(G),\]

(Dual feasibility) \[\lambda \geq 0,\]

(Complementary slackness) \[\lambda(v^T \bar{Q}v - \alpha) = 0.\]

Case 1: \(\lambda = 0\). This case is trivial.
Solving the Objective Function (2)

Case 2: $\lambda > 0$.

These equations have infinite number of solutions. Thus we introduce an auxiliary parameter.

\[
\begin{align*}
\bar{L}v - \lambda \bar{Q}v - \mu v &= 0, \\
v^T \bar{Q}v &\geq \alpha, \quad v^T v = \text{vol}(G), \\
\lambda &\geq 0, \\
\lambda(v^T \bar{Q}v - \alpha) &= 0. \\
\end{align*}
\]

Then we solve instead:

\[
\begin{align*}
\bar{L}v &= \lambda(\bar{Q} - \frac{\beta}{\text{vol}(G)} I)v \\
v^T v &= \text{vol}(G),
\end{align*}
\]

This is a generalized eigenvalue problem.

And we show the result will guarantee that:

\[
\begin{align*}
v^T \bar{Q}v &= \alpha > \beta.
\end{align*}
\]

$\beta$ lower-bounds $\alpha$, and thus indirectly lower-bounds how well the constraints in $Q$ are satisfied.
A Summary of Our Algorithm

1. Solving the generalized eigenvalue problem with a user-specified $\beta$.

$$\bar{L}v = \lambda (\bar{Q} - \frac{\beta}{\text{vol}(G)} I) v$$

2. Find eigenvectors associated with positive eigenvalues.

3. The eigenvector that minimizes $v^T \bar{L}v$ is the optimal solution.

A sufficient condition for the existence of solutions:

$$\beta < \lambda_{max} \text{vol}(G),$$

$\lambda_{max}$ to be the largest eigenvalue of $\bar{Q}$. 
Our formulation can be interpreted as finding the normalized min-cut of a *labeled* graph.

A illustrative example where the affinity structure suggests \{1,2,3\mid 4,5,6\} but the labeling suggests \{1,2,3,4\mid 5,6\}.
Results on Image Segmentation (1)

The pixels in the bounded areas are labeled. Then pairwise constraints are generated based on known labels.
Results on Image Segmentation (2)
Results on UCI Benchmarks

CSP: Our method
ModAff: A previous method that modifies the affinity matrix directly
GrBias: A previous method that only encodes ML constraints
Conclusions

• We propose a constrained spectral clustering framework that can incorporate both hard and soft clustering.
• Our method uses a user-specified parameter to lower-bound how well the given constraints must be satisfied by the resultant clustering.
• Our formulation can be interpreted as a min-cut problem on labeled graph.
• We empirically justified the advantage of our method on several benchmark data sets.
Acknowledgments

• The authors gratefully acknowledge support of this research from NSF (IIS-0801528) and ONR (N000140910712 P00001).
• The first author thanks the support from KDD’10 Student Travel Award.