Normalized Cuts and Cheeger's Inequality

Ian Chen

University of Illinois Urbana-Champaign

June 7, 2026

Image Segmentation

- Given an image, identify relevant regions, large and internally coherent
- Clustering is an unsupervised problem, and very useful.

Figure: **Image segmentation:** Image segmentation identifies important regions of the image. (a) the input image (grayscale). (b-d) shows the mask of three segments detected by an image segmenter. We see the head, the body of the subject , and the background.

Jianbo Shi and J. Malik, "Normalized cuts and image segmentation," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 22, no. 8, pp. 888-905, Aug. 2000, doi: 10.1109/34.868688.

Primer on Laplacians

- For a graph G, (edge-)incidence matrix $K \in \mathbb{R}^{|E|\times|V|}$

$$K_{eu} = \begin{cases} \sqrt{w_e} & e = (u,v), u < v \\ -\sqrt{w_e} & e = (u,v), u > v \\ 0 & \text{otherwise} \end{cases}$$

- K is the gradient operator on vertex-valued functions

$$(K\mathbf{f})_{u\to v} = \sqrt{w_{uv}}(f_u - f_v)$$

- K^T is the divergence operator on edge-valued vector fields

$$(K^T\mathbf{g})_u = \sum_{u\to v\in E} \sqrt{w_{u\to v}}g_{u\to v} - \sum_{v\to u\in E} \sqrt{w_{v\to u}}g_{v\to u}$$

Primer on Laplacians (cont.)

- For the *characteristic vector* of a set S, $(\mathcal{X}_S)_u = 1_{u \in S}$, then

$$\|K\mathcal{X}_S\|_2^2 = \sum_{e \in E(S)} (0)^2 + \sum_{e \in E(\bar{S})} (0)^2 + \sum_{e \in \partial S} (\pm\sqrt{w_e})^2 = w(\partial S)$$

- Laplacian is $L = K^T K = D - A \in \mathbb{R}^{|V| \times |V|}$, $D = \mathrm{diag}(A\mathbf{1}_n) = \mathrm{diag}(\mathbf{d})$
- Normalized adjacency and Laplacian

$$\tilde{A} = D^{-1/2} A D^{-1/2} \qquad \tilde{L} = D^{-1/2} L D^{-1/2} = I - \tilde{A}$$

- $\mathbf{x}^T L \mathbf{x} = \|K\mathbf{x}\|_2^2 = \sum_{ij \in E} w_{ij} (x_i - x_j)^2$

Spectral clustering

Figure: **Normalized versus standard mincut segmentation**: (a) a toy image with four circles (background masked out). (b) using spectral clustering on the four circles correctly identifies circles. (c) using mincut clustering on the four circles finds 3 segments of size 1. Trivial mincuts suggest a need to penalization for small segments.

Spectral Clustering (cont.)

- Normalized cuts penalizes large boundaries and small segments:

$$\mathrm{vol}(S) = \sum_{u \in S} \sum_{v \in V} w_{uv} \qquad \sigma(S_1, \ldots, S_k) = \sum_{\ell=1}^{k} \frac{w(\partial S_\ell)}{\mathrm{vol}(S_\ell)}$$

- Rewrite σ in terms of L and D as Rayleight-quotient:

$$\sigma(S_1, \ldots, S_k) = \sum_{\ell=1}^{k} \frac{\mathcal{X}_{S_\ell}^T L \mathcal{X}_{S_\ell}}{\mathcal{X}_{S_\ell}^T D \mathcal{X}_{S_\ell}}$$

Spectral Clustering (cont.)

- Assignment weighted matrix $H \in \mathbb{R}^{n\times k}$, row i is assignment of vertex i

$$h_{ij} = 1_{\{i\in S_j\}} \frac{1}{\sqrt{\mathrm{vol}(S_j)}}$$

- Write $\sigma(G)$ in terms of a discrete matrix optimization problem

$$\mathbf{h}_j^T L \mathbf{h}_j = \frac{\mathcal{X}_{S_j}^T L \mathcal{X}_{S_j}}{\mathrm{vol}(S_j)}, \qquad \mathbf{h}_j^T D \mathbf{h}_j = \frac{\mathcal{X}_{S_j}^T D \mathcal{X}_{S_j}}{\mathrm{vol}(S_j)} = 1$$

$$\implies \sigma(S_1,\ldots,S_k) = \mathrm{Tr}\left(H^T L H\right), H^T D H = I_k$$

- Change of variables to normalized Laplacian

$$J = D^{1/2} H \implies \mathrm{Tr}\left(H^T L H\right) = \mathrm{Tr}\left(J^T \tilde{L} J\right), J^T J = I_k$$

- SVD of $\tilde{L}$ implies J columns are k smallest eigenvectors of $\tilde{L}$ (singular K).
- Rounding typically done via k-means on spectral embedding.

Spectral Clustering (cont.)

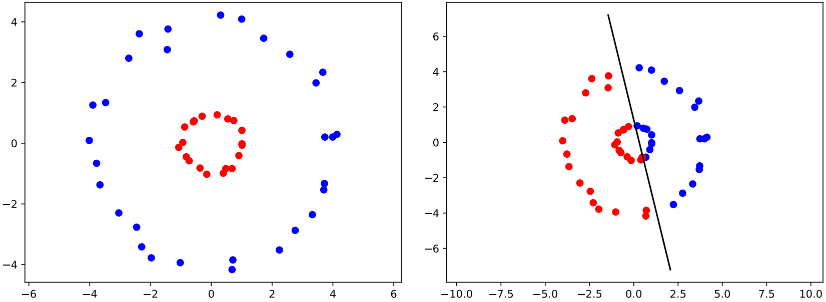

Figure: **Spectral clustering**: Spectral embedding is able to find clusters based on connectivity. (a), the k-means clustering of Spectral embedding. (b), the k-means clustering using Euclidean embedding.

Cheeger's Inequality (Easy Direction)

- $k = 2$, finding S is a 1-D problem ($S_2 = \bar{S}_1$); $\sigma(G) = \min_S \sigma(S, \bar{S})$.
- Write $\lambda_1 \leq \ldots \leq \lambda_n$ eigenvalues, $\mathbf{v}_1, \ldots, \mathbf{v}_n$ eigenvectors of $\tilde{L}$
- $\lambda_1 = 0$, as $L\mathbf{1}_n = (D - A)\mathbf{1}_n = \mathbf{d} - \mathbf{d} = \mathbf{0}$
- $\mathbf{v}_1 = \frac{1}{\sqrt{n}}\mathbf{1}_n$ independent of G, and we can discard it
- Relaxed objectives gives inequality in terms of λ_2:

$$\sigma(G) = \min_{J \text{ assignment}, J^T J = I_k} \operatorname{Tr}\left(J^T L J\right) \geq \min_{J^T J = I_k} \operatorname{Tr}\left(J^T L J\right) = \lambda_2$$

Cheeger's Inequality (Hard Direction)

- Rounding a solution via sweep-cuts of $\mathbf{v}_2$:
  - Compute $\mathbf{x}$ 2nd eigenvector of $\tilde{L}$.
  - Return best cut among sweeps $\{ S_\tau = \{ v \in V : \operatorname{sign}(\mathbf{x}(v))(\mathbf{x}(v))^2 \leq \tau \} \}$
- WLOG, assume that $\mathbf{x}$ is ascending and $x_1^2 + x_n^2 = 1$.
- For $Y \neq 0$, then considering sign of $Z = X - \operatorname{Ev}(X)Y/\operatorname{Ev}(Y)$,

$$\Pr\left\{ \frac{X}{Y} \leq \frac{\operatorname{Ev}(X)}{\operatorname{Ev}(Y)} \right\} > 0$$

- Suffices to show that (for any distribution on τ)

$$\operatorname{Ev}(\min(\operatorname{vol}(S_\tau), \operatorname{vol}(\bar{S_\tau}))) = \mathbf{x}^T D\mathbf{x}$$

$$\operatorname{Ev}(|\partial S_\tau|) \leq \sqrt{\mathbf{x}^T \tilde{L}\mathbf{x}}\sqrt{2\mathbf{x}^T D\mathbf{x}} = \sqrt{2\lambda_2}\mathbf{x}^T D\mathbf{x}$$

Cheeger's Inequality (Hard Direction) (cont.)

- First bound by noting smaller S determined by zero point of $\mathbf{x}$ ($\mathbf{x} \perp \mathbf{d}$)

$$\Pr\left\{u \in \arg\min(\mathrm{vol}(S), \mathrm{vol}(\bar{S}))\right\} = \begin{cases} \tau \in [x(u), 0] & x(u) < 0 \\ \tau \in [0, x(u)] & x(u) > 0 \end{cases}$$

$$\implies \mathrm{Ev}(\min(\mathrm{vol}(S), \mathrm{vol}(\bar{S}))) = \sum_u \deg(u) x(u)^2 = \mathbf{x}^T D \mathbf{x}$$

Cheeger's Inequality (Hard Direction) (cont.)

- Second bound with Cauchy-Schwartz to connect ℓ_1 and ℓ_2 norms,

$$
\begin{aligned}
\Pr\{uv \in E(S, \bar{(S)})\} &= \begin{cases} \left|x(u)^2 - x(v)^2\right| & \text{different signs} \\ x(u)^2 + x(v)^2 & \text{same signs} \end{cases} \\
&\leq |x(u) - x(v)|(|x(u)| + |x(v)|) \\
\implies \mathrm{Ev}(|\partial S|) &\leq \sum_{uv \in E} |x(u) - x(v)|(|x(u)| + |x(v)|) \\
&\leq \sqrt{\sum_{uv \in E} w(uv)(x(u) - x(v))^2} \sqrt{\sum_{uv \in E} w(uv)(|x(u)| + |x(v)|)^2} \\
&\leq \sqrt{\mathbf{x}^T L \mathbf{x}} \sqrt{2\mathbf{x}^T D \mathbf{x}}
\end{aligned}
$$

Cheeger's Inequality

- We showed that $\lambda_2 \leq \sigma(G) \leq 2\sqrt{2\lambda_2}$
- In terms of the conductance of a graph, $\varphi(S) = \frac{w(\partial S)}{\min(\mathrm{vol}(S), \mathrm{vol}(V-S))}$

$$\frac{\lambda_2}{2} \leq \varphi(G) \leq \sqrt{2\lambda_2}$$

- Establishes equivalence between spectral expanders and edge expanders
- Useful in bounding mixing times of random walks, $(I - \tilde{A})$ is the deviation from stationary distribution