Normalized Cuts and Cheeger's Inequality

Ian Chen

May 31, 2026

1 Introduction

We motivate this discussion through image segmentation: given an array of pixels, we want to segment the image into large, internally-coherent pieces. The GrabCut algorithm [3] assigns a probability z_i of whether pixel i is in the foreground or background (i.e. binary segmentation). To ensure cohesive segments, we formulate the optimization through a graph (typically grid), so that we want to minimize

$$L_{\text{grabcut}}(\boldsymbol{z}) = \sum_{ij \in E} \left(\underbrace{z_i z_j E_{ij,11}}_{\text{both foreground}} + \underbrace{(1-z_i) z_j E_{ij,01}}_{\text{boundary}} + \underbrace{z_i (1-z_j) E_{ij,10}}_{\text{boundary}} + \underbrace{(1-z_i)(1-z_j) E_{ij,00}}_{\text{both background}} \right)$$

and can be solved exactly in the case of $E_{ij,00} = 0 = E_{ij,11}$ and $E_{ij,01} = E_{ij,10} > 0$, e.g. via Gaussian kernel,

$$E_{ij,01} = E_{ij,10} = e^{\frac{-1}{2\sigma^2}(\boldsymbol{f}_i - \boldsymbol{f}_j)^T (\boldsymbol{f}_i - \boldsymbol{f}_j)}$$

where $\boldsymbol{f}_i$ is a feature associated with the image, such as $(\boldsymbol{p}, \boldsymbol{c})$ position and color. More generally, when segmenting into k pieces, we build an affinity matrix and find a cut. In an effort to avoid small trivial segments, the *normalized cut* objective is given as

$$\sigma(S_1, \ldots, S_k) = \sum_{\ell=1}^{k} \frac{w(\partial S_\ell)}{\text{vol}(S_\ell)}$$

where $w(\partial S)$ denotes the sum of weights in the boundary, i.e. all edges with exactly one endpoint in S, and the volume is the sum of weights of all half-edges from S. We penalize small segments (i.e. with little volume), and also segments with poor separation (i.e. with large boundaries). Eigenvalues are natural – they are permutationally covariant – and indeed we will see their use.

2 Laplacians

We start with a graph $G = (V, E)$, with $|V| = n, |E| = m$ and non-negative weights $w(\cdot)$, e.g. the grid-graph with Gaussian kernel for images. The edge-incidence matrix $K \in \mathbb{R}^{m \times n}$ indexes rows/cols by edges/nodes, with

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

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

$$\|K \boldsymbol{\mathcal{X}}_S\|_2^2 = \sum_{uv \in E(S,S)} (\sqrt{w_e} - \sqrt{w_e})^2 + \sum_{uv \in E(V-S, V-S)} (0-0)^2 + \sum_{uv \in \partial S} (\pm\sqrt{w_e} - 0)^2 = w(\partial S)$$

K acts as the gradient operator of a vertex-valued function, and K^T acts as the divergence operator of edge-valued vector-field. K^TK is the Laplacian L, and is easy to rewrite as $L = D - A$, for A the adjacency and $D = \mathrm{diag}(A\mathbf{1}_n) = \mathrm{diag}(\boldsymbol{d})$. The normalized Laplacian is denoted $\tilde{L} = D^{-1/2}LD^{-1/2} = I - D^{-1/2}AD^{-1/2} = I - \tilde{A}$. The Laplacian is positive semi-definite, with $\boldsymbol{x}^TL\boldsymbol{x} = \|K\boldsymbol{x}\|_2^2 = \sum_{ij\in E} w_{ij}(x_i - x_j)^2$.

3 Spectral clustering

Fix $k > 0$. We can re-write the normalized cut objective as a Rayleigh quotient,

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

Solving this problem exactly is NP-Hard, so we use this quadratic form to formulate an abstraction. First, we consider a weighted assignment matrix $H \in \mathbb{R}^{n\times k}$, where $\boldsymbol{h}_i$, the i-th row indicates the assignment of vertex i:

$$h_{ij} = \begin{cases} 1/\sqrt{\mathrm{vol}(S_j)} & i \in S_j \\ 0 & \text{otherwise} \end{cases}$$

so that $\sigma(S_1, \ldots, S_k) = \mathbf{Tr}(H^TLH)$, since $\boldsymbol{h}_j^T L\boldsymbol{h}_j = \frac{\mathcal{X}_{S_j}^T L\mathcal{X}_{S_j}}{\mathrm{vol}(S_j)}$. Note also that $\boldsymbol{h}_j^T D\boldsymbol{h}_j = \frac{\mathcal{X}_{S_j}^T D\mathcal{X}_{S_j}}{\mathrm{vol}(S_j)} = 1$, so we want to optimize $\mathbf{Tr}(H^TLH)$ subject to $H^TDH = 1$ and H has the discrete entries as above. Relaxing discreteness,

$$\hat{H}^{\text{opt}} = \underset{H^TDH=I_k}{\arg\min}\ \mathbf{Tr}(H^TLH) = \underset{J=D^{1/2}H, J^TJ=I_k}{\arg\min}\ \mathbf{Tr}(J^T\tilde{L}J) \tag{1}$$

Thus, by a standard exercise in SVD, $\hat{J}^{\text{OPT}}$ is has columns of the k-minimum eigenvectors of $\tilde{L}$. A standard rounding is to treat the rows of $\hat{H}^{\text{OPT}}$ as an embedding of the points, and run k-means clustering, which has elegant theory via the Davis-Kahan theorem.

4 Cheeger's Inequality

To finish off Cheeger's inequality, we interpret the above by specializing $k = 2$. We know that $\mathbf{1}_n$ is always in the null-space of L, as $(D - A)\mathbf{1}_n = \boldsymbol{d} - \boldsymbol{d} = \mathbf{0}$. Thus, $\boldsymbol{d}^{1/2}$ is in the null-space of $\tilde{L}$, as $\tilde{L}\boldsymbol{d}^{1/2} = D^{-1/2}LD^{-1/2}\boldsymbol{d}^{1/2} = D^{-1/2}L\mathbf{1}_n = \mathbf{0}$. The first column of $\hat{H}^{\text{OPT}}$ is thus $D^{-1/2}\boldsymbol{d}^{1/2} = \mathbf{1}_n$, and can be ignored (it is independent of G). Indeed, with its 0 contribution, the optimal *relaxed* objective value is the second-smallest eigenvalue of $\tilde{L}$.

Theorem 1. *Let $\lambda_1 \leq \ldots \leq \lambda_n$ be the eigenvalues of $\tilde{L}$, $\sigma(G) = \min_{S\subset V} \sigma(S, V - S)$. Then, $\lambda_2 \leq \sigma(G)$.*

Now, let us address the problem of rounding the solution. We proceed via sweep-cuts:

- Compute $\boldsymbol{x}$ 2nd eigenvector of the normalized Laplacian $D^{-1/2}LD^{-1/2} = \tilde{L}$.
- Compute $\left\{S_\tau = \left\{v \in V : \mathrm{sign}(\boldsymbol{x}(v))(\boldsymbol{x}(v))^2 \leq \tau\right\}\right\}$, and take the minimum normalized cut among all τ.

WLOG, assume that $\boldsymbol{x}$ is ascending and $x_1^2 + x_n^2 = 1$. Suppose we sample τ uniformly among $[-(\boldsymbol{x}(1))^2, (\boldsymbol{x}(n))^2]$, and we show that $\mathbf{Pr}_\tau\left(\sigma(S_\tau) \leq \sqrt{2\lambda_2}\right) > 0$. Indeed, it suffices that[1]

$$\mathbf{E}(\min(\mathrm{vol}(S), \mathrm{vol}(\bar{S}))) = \boldsymbol{x}^TD\boldsymbol{x}, \qquad \mathbf{E}(|\partial S|) \leq \sqrt{\boldsymbol{x}^T\tilde{L}\boldsymbol{x}}\sqrt{2\boldsymbol{x}^TD\boldsymbol{x}} = \sqrt{2\lambda_2}\boldsymbol{x}^TD\boldsymbol{x}$$

[1]For $Y \neq 0$, then $\mathbf{Pr}\left(\frac{X}{Y} \leq \frac{\mathbf{E}(X)}{\mathbf{E}(Y)}\right) > 0$, since $Z = X - \mathbf{E}(X)Y/\mathbf{E}(Y)$ is zero everywhere, else has positive probability of both signs.

For the first bound, for both signs of $x(u)$, indeed (as the minimum is determined by the 0 point, since $\boldsymbol{x} \perp \boldsymbol{d}$)

$$\begin{aligned}
\mathbf{Pr}\big(u \in \arg\min(\mathrm{vol}(S), \mathrm{vol}(\bar{S}))\big) &= \begin{cases} \tau \in [-x(u)^2, 0] = x(u)^2 & x(u) < 0 \\ \tau \in [0, +x(u)^2] = x(u)^2 & x(u) > 0 \end{cases} \\
\implies \mathbf{E}(\min(\mathrm{vol}(S), \mathrm{vol}(\bar{S}))) &= \sum_u \deg(u) x(u)^2 = \boldsymbol{x}^T D \boldsymbol{x}
\end{aligned}$$

For the second, we apply Cauchy-Schwartz to connect ℓ_1 and ℓ_2 norms,

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

Theorem 2. *Let $\lambda_1 \leq \ldots \leq \lambda_n$ be the eigenvalues of $\tilde{L}$, $\sigma(G) = \min_{S \subset V} \sigma(S, V - S)$. Then, $\sigma(G) \leq 2\sqrt{2\lambda_2}$.*

Putting this together, we find Cheeger's inequality, typically stated in terms of the conductance

$$\varphi(G) = \min_{S \subset V} \frac{w(\partial S)}{\min(\mathrm{vol}(S), \mathrm{vol}(V - S))} \qquad \frac{\lambda_2}{2} \leq \varphi(G) \leq \sqrt{2\lambda_2}$$

5 Bibliographic Notes

The background of image segmentation is adapted from Professor Forsyth's unreleased textbook[2][1]. The section of Laplacians is mostly standard but is adapted from the Fall 2009 instance of MIT 18-409[3]. The spectral cuts is adapted from a tutorial in Luxburg [2], as well as Daniel Spielman's book on spectral graph theory. The proof of Cheeger's inequality is described in Trevisan's 2011 offering of CS359[4].

References

[1] David Forsyth. *Computer Vision*. TBD, 2026.

[2] Ulrike von Luxburg. "A tutorial on spectral clustering". In: *Statistics and Computing* 17.4 (Aug. 2007), pp. 395–416. ISSN: 1573-1375. DOI: `10.1007/s11222-007-9033-z`. URL: `http://dx.doi.org/10.1007/s11222-007-9033-z`.

[3] Carsten Rother, Vladimir Kolmogorov, and Andrew Blake. "GrabCut: interactive foreground extraction using iterated graph cuts". In: *ACM Transactions on Graphics* 23.3 (Aug. 2004), pp. 309–314. ISSN: 1557-7368. DOI: `10.1145/1015706.1015720`. URL: `http://dx.doi.org/10.1145/1015706.1015720`.

[2] Draft at Spring 2026 CS 543 offering, `http://luthuli.cs.uiuc.edu/~daf/Courses/CV2026/classwebpage.html`

[3] `https://ocw.mit.edu/courses/18-409-topics-in-theoretical-computer-science-an-algorithmists-toolkit-fall-2009/`

[4] `https://theory.stanford.edu/~trevisan/cs359g/`