Multilevel k-Way Spectral Clustering with Degree-Free Laplacian

Shijie Xu
School of Computer and Technology, University of Science and Technology of China, China
xushijie@mail.ustc.edu.cn

Abstract

...We propose a general degree-free Laplacian methods which can be used in the multilevel graph partitioning...

Introduction

Graph partitioning (also called graph clustering) is an important problem and has various applications in many areas. Spectral graph partitioning has been studied well over decades (Chung 1997). Spectral methods have been used effectively for solving a number of graph partitioning objectives, including ratio cut (Chan, Schlag, and Zien 1992) and normalized cut (Shi and Malik 2000), and typically perform well because they compute the global optimal solution of a relaxation to the original partitioning object. Since the eigenvector computation is prohibitive in many cased (e.g. if many eigenvectors of a very large matrix are required), the effective methods are desirable.

The multilevel methods for graph partitioning produce partitions at a low cost with three phases. First, they construct a sequence of increasing coarse approximations to the graph. Second, they partition the smallest graph in the sequence with base clustering approaches. Third, they project the coarse partition back through the sequence of graphs, periodically improving it with a local refinement algorithm. Different approaches has been performed in the second phase such as spectral bisection (Hendrickson and Leland 1995), region-growing (Dhillon, Guan, and Kulis 2007). Since spectral clustering methods are typically the best algorithm for optimizing weighted graph clustering objectives (Dhillon, Guan, and Kulis 2007), a multilevel spectral clustering is immediately desirable.

The rest of the paper is organized as follows. Section gives formal definitions of degree-free Laplacian and related results...

Related Work

Preliminaries

Definition 1 (Degree-Free Laplacian). *Let $G = (V, M, W)$ be a connected undirected graph with vertex-weight matrix M and edge-weight matrix W where $V = \{1, 2, \dots, n\}$ is the vertex set of G. Let D be the degree matrix of the graph, i.e. $D = diag\{d_1, d_2, \dots, d_n\} = diag\{\sum_j W_{ij}\}$, and $M = diag\{m_1, m_2, \dots, m_n\}$. The degree-free laplacian of the graph is a matrix $L_M := M^{-1}(D - W)$. Furthermore the symmetric Laplacian is defined as $L_{sym} = M^{1/2}(D - W)M^{-1/2}$.*

Definition 2. *Suppose A, B are two disjoint subsets of V, the cut of A and B on the graph G is $\mathrm{Cut}(A, B) := \sum_{i \in A, j \in B} W_{ij}$. Suppose $\pi = \{C_1, C_2, \dots, C_k\}$ is a k-partition of the graph G, i.e. $V = C_1 \cup C_2 \cup \ldots \cup C_k$ and $C_i \cap C_j = \emptyset$ for any $i \neq j$, then we say $i \sim j$ with partition π if they are in the same part. We define nomalized cut function of parition π as*

$$\mathrm{NCut}(\pi) := \sum_i \frac{\mathrm{Cut}(C_i, \bar{C}_i)}{\mathrm{vol}(C_i)}$$

where $\mathrm{vol}(C_i) = \sum_{x \in C_i} \mathrm{d}_\omega(x)$.

Definition 3. *Denote $\mathcal{G}$ as the space of all functions $f : V \to \mathbb{R}$, $\mathrm{d}_\omega \in \mathcal{G}$ is called vertex-weight. For all $f \in \mathcal{G}$, the gradient of f with weight function d_ω is a vector $\nabla f := ((f(y) - f(x))\sqrt{\frac{W(x,y)}{\mathrm{d}_\omega(x)}})_{y \in V}$, the laplacian operator on f with weight d_ω is defined as $\Delta f := \sum_{y \in V}(f(y) - f(x))\frac{W(x,y)}{\mathrm{d}_\omega(x)}$, and the integration of f with weight d_ω on the graph G is $\int_G f \mathrm{d}_\omega := \sum_{x \in V} f(x)\mathrm{d}_\omega(x)$, omitted as $\int f$. The inner product $\langle , \rangle$ is represented as $\langle f, g \rangle = \int fg$.*

Lemma 1. *Suppose $\mathbb{1}_C \in \mathcal{G}$ be the characteristic function of C, i.e. $\mathbb{1}_C(x) = 1$ if $x \in C$, and 0 otherwise. We have*

$$\mathrm{Cut}(C, \bar{C}) = \int |\nabla \mathbb{1}_C|^2, \qquad \mathrm{vol}(C) = \int \mathbb{1}_C^2. \tag{1}$$

We notice that the first equation is independent of weight d_ω.

Definition 4. *Suppose $\tilde{G} = (\tilde{V}, \tilde{M}, \tilde{W}), G = (V, M, W)$ are two graphs and $|V| = n$, we say $\tilde{G}$ is a 1-coarsening of G with i, j if $\tilde{V} \subset V$ and there exists $i \neq j$ and*

$$T = \begin{pmatrix} I_{i-1} & v_1' & 0 \\ 0 & v_2' & I_{n-i} \end{pmatrix} \in \mathbb{R}^{(n-1) \times n}$$

where $[v_1 \; v_2] = \delta_j$, or

$$T = \begin{pmatrix} I_{j-1} & v_1' & 0 \\ 0 & v_2' & I_{n-j} \end{pmatrix} \in \mathbb{R}^{(n-1) \times n}$$

where $[v_1\ v_2] = \delta_i$ s.t.

$$\tilde{W} = T'WT, \tilde{M} = T'MT,$$

which can be denoted as $\tilde{G} \leq G$. Furthermore we say $\hat{G}$ is a k-coarsening of G for $k \geq 2$ if there exists $G_1, C_2, \ldots, C_{k-1}$ s.t. $\hat{G} \leq G_1 \leq G_2 \leq \ldots \leq G_k \leq G$. Also we say $\tilde{\pi} \leq \pi$ if $\tilde{\pi}$ is a coarsened partition of π.

Lemma 2. *The Laplacian operator Δ in space $(\mathcal{S}, \langle,\rangle)$ is equivalent to $L_M = M^{-1}(D - W)$ in $(\mathrm{R}^{n\times n}, \langle,\rangle)$ where D is the degree matrix of W, i.e. $D_{ii} = \sum_j W_{ij}$ and $M_{ii} = \mathrm{d}_\omega(i)$.*

Lemma 3. *L_M is symmetric and positive semi-definite, has the smallest eigenvalues $\lambda_1 \geq 0$, the definite is strictly held whene $M_{ii} > D_{ii}$ for some i, i.e. L has eigenvalues $0 \leq \lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n$.*

Lemma 4. *Suppose $K : \mathcal{G}^k \to \mathbb{R}$ is a functional on $\mathcal{G}^k$. Consider the constrained optimization problem: $f_i \in \mathcal{G}$ are k orthogonal functions minimize the following equation:*

$$K(f_1, f_2, \ldots, f_k) = \sum_{i=1}^{k} \frac{\int |\nabla f_i|^2}{\int f_i^2},$$

then f_i which are the first smallest eigenfunctions of Δ minimize the functional.

Lemma 5. *Suppose π is a partition of G and $i \sim j$, if $\tilde{G} \leq G$ with i, j, and $\tilde{\pi} \leq \pi$, then $\mathrm{NCut}(\pi) = \mathrm{NCut}(\tilde{\pi})$.*

Algorithm

In this section we describe our multilevel spectral clustering algorithm, which consists of three parts, coarsening phase, clustering phase, and refining phase.

Coarsening Phase

Clustering Phase

Refining Phase

Experiments

Conclusion

Acknowledgements

Appendices

Proof of lemma 2

Proof. We have

$$L_M(i,j) = \begin{cases} \frac{D_{ii}}{\mathrm{d}_\omega(i)} - \frac{W(i,i)}{\mathrm{d}_\omega(i)} & j = i, \\ -\frac{W(i,j)}{\mathrm{d}_\omega(i)} & j \neq i, \\ 0 & \text{otherwise.} \end{cases}$$

therefore $L_M f = \sum_{y\in V}(f(y) - f(x))\frac{W(x,y)}{\mathrm{d}_\omega(x)}$. □

Proof of lemma 3:

Proof. We would show $f'L_M f \geq 0$ for all $f \in \mathbb{R}^n$. Assume u is an eigen value of L_M, i.e. $L_M u = \lambda u$ for some constant λ, let $u = M^{-1/2}v$ we have $L_M M^{-1/2}v = \lambda M^{1/2}v$, hence $L_{\mathrm{sym}} v = \lambda v$, i.e. L_{sym} has the same eigenvalues with L_M. We would show L_{sym} is positive semi-definite, i.e. $f' L_{\mathrm{sym}} f \geq 0$ for all $f \in \mathbb{R}^n$. Denote $g := M^{-1/2}f$, then we know $f'L_{\mathrm{sym}} f = g'(D - W)g$, we have

$$\begin{aligned} f'L_{\mathrm{sym}} f &= f'M^{-1/2}(D - W)M^{-1/2}f \\ &= g'(D - W)g \\ &= \sum_{i\neq j} W_{ij}(g_i - g_j)^2 \geq 0, \end{aligned}$$

and the equality is held only for $g = 0$. Therefore we proved the positive semi-definite. Next, let λ is an eigenvalue of L_{sym} and x the corresponding normalized eigenvector, we have

$$\begin{aligned} \lambda = \langle x, \lambda x\rangle &= \langle x, L_M x\rangle \\ &= \langle L_M' x, x\rangle = \langle L_M x, x\rangle \\ &= \langle \lambda x, x\rangle = \bar{\lambda}. \end{aligned}$$

Therefore $\lambda \in \mathbb{R}$. □

Proof of lemma 4:

Proof. Suppose $f_1, f_2, \ldots, f_k$ achieve the minimum of K. We first show that

$$K(f_1, f_2, \ldots, f_k) = \sum_i \frac{\int |\nabla f_i|^2}{\int f_i^2} = \sum_i \frac{\langle f_i, \Delta f_i\rangle}{\langle f_i, f_i\rangle}. \quad (2)$$

Suppose u_i are all normalized eigenfunctions of Δ. and $f_i = \sum_j a_{ij} u_j$, we have

$$K(f_1, f_2, \ldots, f_k) = \sum_i \frac{\sum_j a_{ij}^2 \lambda_j}{\sum_j a_{ij}^2}. \quad (3)$$

Denote $A = \{a_{ij}\} \in \mathbb{R}^{n\times k}, U = \{u_{ij}\}, \Lambda = diag\{\lambda_1, \lambda_2, \ldots, \lambda_n\} \in \mathbb{R}^{n\times n}$, we have

$$K(f_1, f_2, \ldots, f_k) = \mathrm{tr} A'\Lambda A, \qquad s.t.\ A'A = I_k. \quad (4)$$

From Rayleigh-Ritz theorem we know that $\lambda_1 + \lambda_2 + \ldots + \lambda_k \leq K \leq \lambda_{n-k+1} + \lambda_{n-k+2} + \ldots + \lambda_n$. Therefore $f_i = u_i$ achieve this minimum. □

References

Chan, P. K.; Schlag, M.; and Zien, J. 1992. Spectral K -Way Ratio-Cut Partitioning Part I : Preliminary Results. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems* 13(9):1088 – 1096.

Chung, F. R. K. 1997. Spectral Graph Theory.

Dhillon, I. S.; Guan, Y.; and Kulis, B. 2007. Weighted graph cuts without eigenvectors a multilevel approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 29(11):1944–1957.

Hendrickson, B., and Leland, R. 1995. A multilevel algorithm for partitioning graphs. 28–es.

Shi, J., and Malik, J. 2000. Normalized Cuts and Image Segmentation. 22(8):1–18.