Machine Learning with Graphs: Spectral Graph Theory 3/3 - Partitioning, Spectral Clustering, Graph Signal Processing

Arlei Silva

Spring 2022

Graph Partitioning

A graph cut $(S, \overline{S})$ divides a weighted graph $G = (V, E, W)$ into partitions $S \subseteq V$ and $\overline{S} = V - S$. Let $|(S, \overline{S})| = |\{(u, v) \in E | u \in S \wedge v \in \overline{S}\}|$ and $|S|$ be the size of the partition S. The *edge expansion* of S is defined as follows:

$$\sigma(S, \overline{S}) = \frac{|(S, \overline{S}))|}{\min\{|S|, |\overline{S}|\}}$$

The *sparsest cut problem* asks for a cut $(S, \overline{S})$ with minimum value of $\sigma(S, \overline{S})$. Notice that sparsest cut is an optimization problem. We can show that this problem is NP-hard using a reduction from *max-cut*, which is also NP-hard [3].

We will show an interesting connection between the sparsest cut and the spectrum of the Laplacian matrix. This will generalize our earlier observation that λ_2 is 0 iff G is disconnected. Let's assume that $|S| \leq |\overline{S}| \leq |V|/2$. Moreover, let $\mathbf{1}_S \in \{0, 1\}^n$ be the indicator vector of S:

$$[\mathbf{1}_S]_v = \begin{cases} 1 & \text{if } v \in S \\ 0 & \text{otherwise} \end{cases}$$

It follows that:

$$\mathbf{1}_S^T L \mathbf{1}_S = \sum_{(u,v) \in E} ([\mathbf{1}_S]_u - [\mathbf{1}_S]_v)^2 = |(S, \overline{S})|$$

The quadratic form of $\mathbf{1}_S$ seems useful to connect the Laplacian and the sparsest cut problem, but $\mathbf{1}_S$ is clearly not an eigenvector of the Laplacian—it is not orthogonal to a constant vector. This can be fixed by instead considering $\mathbf{x} = \mathbf{1}_S - (|S|/|V|)\mathbf{1}_n$. The entries of $\mathbf{x}$ are as follows:

$$\mathbf{x}_v = \begin{cases} 1 - |S|/|V| & \text{if } v \in S \\ -|S|/|V| & \text{otherwise} \end{cases}$$

Now, we can show that $\mathbf{x}$ is orthogonal to the constant vector, $\sum_v \mathbf{x}_v = \sum_v [\mathbf{1}_S]_v - (|S|/|V|)\sum_v [\mathbf{1}_n]_v = |S| - |S| = 0$. Moreover, similar to $\mathbf{1}_S$, $\mathbf{x}$ also captures the size of the cut:

$$\mathbf{x}^T L\mathbf{x} = \sum_{(u,v)\in E} ([\mathbf{1}_S]_u - |S|/|V| - [\mathbf{1}_S]_v + |S|/|V|)^2 = |(S, \overline{S})|$$

Based on the Courant-Fischer theorem, we also need the squared norm of $\mathbf{x}$ to associate it to eigenvectors of L:

$$\mathbf{x}^T\mathbf{x} = \sum_{v\in S}\left(1 - \frac{|S|}{|V|}\right)^2 + \sum_{v\in \overline{S}}\left(-\frac{|S|}{|V|}\right)^2 = \frac{|S||\overline{S}|}{|V|}$$

Putting everything together:

$$\lambda_2 \leq \min_{\mathbf{x}\perp \mathbf{1}_n} \frac{\mathbf{x}^T L\mathbf{x}}{\mathbf{x}^T\mathbf{x}} = \min_S |V|\frac{|(S,\overline{S})|}{|S||\overline{S}|} \leq 2\min_S \sigma(S,\overline{S}) \tag{1}$$

Similar to the edge expansion, we define the *conductance* of S as follows:

$$\phi(S,\overline{S}) = \frac{|(S,\overline{S})|}{\min\{vol(S), vol(\overline{S})\}}$$

where $vol(S) = \sum_{v\in S} deg(v)$ is the volume of S.

Notice that the conductance and the edge expansion are quite similar. However, we need a different denominator to associate the conductance to the spectrum of the Laplacian:

$$\frac{\mathbf{y}^T L\mathbf{y}}{\mathbf{y}^T D\mathbf{y}} = \frac{\mathbf{x}^T D^{-1/2} L D^{-1/2}\mathbf{x}}{\mathbf{x}^T\mathbf{x}} = \frac{\mathbf{x}^T \mathcal{L}\mathbf{x}}{\mathbf{x}^T\mathbf{x}}$$

where we have used the transformation $\mathbf{y} = D^{1/2}\mathbf{x}$ and $\mathcal{L} = D^{-1/2}LD^{-1/2}$ is the *normalized Laplacian*. The entries of $\mathcal{L}$ are as follows:

$$\mathcal{L}_{uv} = \begin{cases} 1 & \text{if } u = v \\ -1/\sqrt{d_u d_v} & \text{if} (u,v)\in E \\ 0 & \text{otherwise} \end{cases}$$

We can show that $\mathbf{d}^{1/2} = [\sqrt{deg(v_1)}, \ldots \sqrt{deg(v_n)^{1/2}}]$ is the smallest eigenvector of $\mathcal{L}$, with eigenvalue 0:

$$D^{-1/2}LD^{-1/2}\mathbf{d}^{1/2} = D^{-1/2}L\mathbf{1}_n = 0$$

We can use $\mathbf{d}^{1/2}$ to compute the second eigenvector of the *generalized eigenvalue problem*. Notice, that we cannot apply the Courant-Fischer theorem here directly because of the denominator.

$$\lambda_2 = \min_{\mathbf{x}\perp \mathbf{d}^{1/2}} \frac{\mathbf{x}^T\mathcal{L}\mathbf{x}}{\mathbf{x}^T\mathbf{x}} = \min_{\mathbf{y}\perp \mathbf{d}} \frac{\mathbf{y}^T L\mathbf{y}}{\mathbf{y}^T D\mathbf{y}}$$

We could also have applied a generalized version of Courant-Fischer to achieve the same result. Similar edge expansion, we will apply a correction to the indicator vector for S, $\mathbf{y} = \mathbf{1}_S - (vol(S)/vol(V))\mathbf{1}_n$. It follows that $\mathbf{y}^T\mathbf{d} = 0$ and $\mathbf{y}^T L\mathbf{y} = |(S, \overline{S})|$. We still have to compute the denominator:

$$\begin{aligned}\mathbf{y}^T D\mathbf{y} &= \sum_{v\in S} deg(v)\left(1 - \frac{vol(S)}{vol(V)}\right)^2 + \sum_{v\in\overline{S}} deg(v)\left(-\frac{vol(S)}{vol(V)}\right)^2 \\ &= \frac{vol(S)vol(\overline{S})}{vol(V)}\end{aligned}$$

Putting everything together:

$$\lambda_2 \leq \min_{\mathbf{x}\perp\mathbf{d}} \frac{\mathbf{x}^T L\mathbf{x}}{\mathbf{x}^T D\mathbf{x}} = \min_S vol(V)\frac{|(S, \overline{S})|}{vol(S)vol(\overline{S})} \leq 2\min_S \phi(S, \overline{S}) \qquad (2)$$

Equations 1 and 2 show that edge expansion and conductance have lower-bounds related to the second eigenvalue of the Laplacian and normalized Laplacian, respectively. However, they do not guarantee that there are partitions with expansion and sparsity near λ_2. This would require upper-bounds as a function of λ_2. For the case of conductance, such an upper-bound is given by the *Cheeger's inequality* [1, 5].

$$\phi(S, \overline{S}) \leq \sqrt{2\lambda_2}$$

The proof of the Cheeger's inequality is constructive, which means that we can use it to find cuts $(S, \overline{S})$ with such a conductance. The algorithm is quite simple: (1) compute generalized eigenvector of $\mathbf{y}* = \arg\min_{\mathbf{y}} \mathbf{y}^T L\mathbf{y}/\mathbf{y}^T D\mathbf{y}$; (2) sort vertices in V as $v_1, v_2, \ldots v_n$ in increasing order according to $\mathbf{y}*$; (3) sweep sorted vertices with $1 \leq i \leq n$, letting $S = \{v_i | i <= i\}$, and pick partition with minimum conductance. This gives us a poly-time approximation algorithm for minimizing conductance, which is also NP-hard.

Other Applications

For the remainder of this lecture, we will briefly cover two applications of spectral graph theory.

Spectral Clustering

Spectral clustering is a popular clustering algorithm based on the spectrum of the (standard or normalized) Laplacian matrix [6]. Given a dataset $\mathcal{D} = \{\mathbf{x}_1, \ldots \mathbf{x}_n\}$, our goal is to partition $\mathcal{D}$ into k non-overlapping clusters $C_1, C_2, \ldots C_k$. First, we build a complete weighted graph, with $n(n-1)/2$ edges, where $W_{ij} = exp(-\gamma||\mathbf{x}_i - \mathbf{x}_j||)$ and γ is a parameter. The matrix W captures pairwise similarities using the so-called *Radial Basis Function (RBF) Kernel*—there are other alternatives. We can use W to cluster the data as follows:

1. Compute (standard or normalized) Laplacian $L = D - W$;
2. Compute first r eigenvectors $\mathbf{u}_1, \dots \mathbf{u}_r$ of L;
3. Create new feature matrix from eigenvectors, where $\mathbf{x}'_i = [\mathbf{u}_1[i], \dots \mathbf{u}_r[i]]$;
4. Cluster data using new features $\mathbf{x}'_1, \dots \mathbf{x}'_n$ (e.g. applying k-means).

Notice that we can also apply spectral clustering directly to a weighted or unweighted graph.

Signal Processing on Graphs

Signal Processing on Graphs (SPG) is a recent effort to generalize traditional signal processing to graph data [4]. A graph signal $\mathbf{f} \in \mathbb{R}^n$ contains a value $\mathbf{f}[v]$ for node $v \in V$. SPG enables the application of several operations defined for traditional signals (e.g. compression, sampling, filtering) to graph signals. For instance, the *graph Fourier transform* generalizes the traditional Fourier transform as follows:

$$\hat{f}(\lambda_\ell) = \langle \mathbf{f}, \mathbf{u}_\ell \rangle$$

where $\mathbf{u}_\ell$ is an eigenvector of the Laplacian of the graph and $\hat{f}(\lambda_\ell)$ is a Fourier coefficient.

We can also define an inverse operation:

$$\mathbf{f}[v] = \sum_{\ell=1}^{n} \hat{f}(\lambda_\ell)\mathbf{u}_\ell[v]$$

Because the eigenvectors of L form an orthonormal basis for signals $\mathbf{f}$, the inverse gives a perfect reconstruction of the input signal. Moreover, based on the intuition have built in the past lectures, smooth signals should project well on the smallest eigenvectors of L. Thus, we can approximate such signals with few graph Fourier coefficients.

References

[1] Fan RK Chung and Fan Chung Graham. *Spectral graph theory*. American Mathematical Soc., 1997.

[2] Jiaqi Jiang. An introduction to spectral graph theory. `http://math.uchicago.edu/~may/REU2012/REUPapers/JiangJ.pdf`, 2012.

[3] Richard M Karp. Reducibility among combinatorial problems. In *Complexity of computer computations*, pages 85–103. Springer, 1972.

[4] David I Shuman, Sunil K Narang, Pascal Frossard, Antonio Ortega, and Pierre Vandergheynst. The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains. *IEEE signal processing magazine*, 30(3):83–98, 2013.

[5] Daniel Spielman. Spectral graph theory. *Combinatorial scientific computing*, 2012.

[6] Ulrike Von Luxburg. A tutorial on spectral clustering. *Statistics and computing*, 17(4):395–416, 2007.