COMS W6998-5: Algorithms through Geometric Lens (Fall'17) October 31, 2017

Lecture 16 – Multi-Way Cuts, Metric Embeddings I

Instructors: *Alex Andoni, Ilya Razenshteyn* Scribe: *Sam Buchanan*

1 Recap and Wrapup

Let us recall some definitions from the previous lecture. Let $G = (V, E)$, $S \subseteq V$, and see the previous scribe for definitions of the set boundary functional and the volume functional.

Definition 1 (Conductance)**.**

$$\phi(G) = \min_{\substack{S \subseteq V,\, S \neq \varnothing \\ \operatorname{vol}(S) \leq (1/2) \operatorname{vol}(G)}} \phi(S),$$

$$\phi(S) = \frac{\partial(S)}{\operatorname{vol}(S)}.$$

It remains to prove the upper bound in the Cheeger inequality:

$$\mu_2 \leq 2\phi(G),$$

where

$$\mu_2 = \min_{\boldsymbol{x} \perp \boldsymbol{v}_1} \frac{\boldsymbol{x}^T \hat{\boldsymbol{L}} \boldsymbol{x}}{\boldsymbol{x}^T \boldsymbol{x}}, \quad v_1 = \left(\sqrt{d_1}, \dots, \sqrt{d_n}\right).$$

Claim 1. *$\mu_2 \leq 2\phi(G)$.*

Proof. It is enough to prove that for any nonempty $S \subseteq V$ satisfying $\operatorname{vol}(S) \leq (1/2) \operatorname{vol}(G)$, there exists $\boldsymbol{x}$ such that

$$\frac{\boldsymbol{x}^T \hat{\boldsymbol{L}} \boldsymbol{x}}{\|\boldsymbol{x}\|_2^2} \leq 2\phi(S).$$

Making the substitution $\boldsymbol{y} = \boldsymbol{D}^{-\frac{1}{2}} x$, we have to show

$$\frac{\boldsymbol{y}^T \boldsymbol{L} \boldsymbol{y}}{\boldsymbol{y}^T \boldsymbol{D} \boldsymbol{y}} \leq 2\phi(S).$$

It is natural to try to take $\boldsymbol{y}$ to be the characteristic vector of the set S, but as we saw previously, $\boldsymbol{D}^{\frac{1}{2}} \boldsymbol{1}_S \not\perp \boldsymbol{v}_1$. This motivates studying instead the orthogonal projection $\boldsymbol{x} = \boldsymbol{D}^{\frac{1}{2}} \boldsymbol{1}_S - \sigma \boldsymbol{v}_1$ where $\sigma \in \mathbb{R}$ is chosen so that $\langle \boldsymbol{x}, \boldsymbol{v}_1 \rangle = 0$. Enforcing this condition determines the value of σ:

$$\sigma = \frac{\boldsymbol{1}_S^T \boldsymbol{D}^{\frac{1}{2}} \boldsymbol{v}_1}{\|\boldsymbol{v}_1\|_2^2},$$

and, simplifying, we get

$$\sigma = \frac{\sum_{i \in S} d_i}{\sum_{i \in G} d_i} = \frac{\mathrm{vol}(S)}{\mathrm{vol}(G)}.$$

In turn, this yields $\boldsymbol{y} = \boldsymbol{D}^{-\frac{1}{2}}(\boldsymbol{D}^{\frac{1}{2}}\mathbf{1}_S - \sigma \boldsymbol{v}_1) = \mathbf{1}_S - \sigma \mathbf{1}_G$. Using the property proved in the last lecture:

$$\boldsymbol{y}^T \boldsymbol{L} \boldsymbol{y} = \sum_{\substack{(i,j) \in E \\ i<j}} (y_i - y_j)^2 = \partial(S).$$

We now turn our attention to the denominator. We compute

$$\begin{aligned}
\boldsymbol{y}^T \boldsymbol{D} \boldsymbol{y} &= (\mathbf{1}_S - \sigma \mathbf{1}_G)^T D (\mathbf{1}_S - \sigma \mathbf{1}_G) \\
&= \mathbf{1}_S^T \boldsymbol{D} \mathbf{1}_S + \sigma^2 \mathbf{1}_G^T \boldsymbol{D} \mathbf{1}_G - 2\sigma \mathbf{1}_G^T \boldsymbol{D} \mathbf{1}_S \\
&= \mathrm{vol}(S) + \sigma^2 \mathrm{vol}(G) - 2\sigma \mathrm{vol}(S) \\
&= \mathrm{vol}(S) + \frac{\mathrm{vol}(S)^2}{\mathrm{vol}(G)} - 2 \mathrm{vol}(S) \frac{\mathrm{vol}(S)}{\mathrm{vol}(G)} \\
&= \mathrm{vol}(S) \left(1 - \frac{\mathrm{vol}(S)}{\mathrm{vol}(G)}\right).
\end{aligned}$$

This yields

$$\begin{aligned}
\frac{\boldsymbol{y}^T \boldsymbol{L} \boldsymbol{y}}{\boldsymbol{y}^T \boldsymbol{D} \boldsymbol{y}} &= \frac{\partial(S)}{\mathrm{vol}(S)\left(1 - \frac{\mathrm{vol}(S)}{\mathrm{vol}(G)}\right)} \\
&= \phi(S) \frac{1}{1 - \frac{\mathrm{vol}(S)}{\mathrm{vol}(G)}} \\
&\leq 2\phi(S),
\end{aligned}$$

where we have used the constraint $\mathrm{vol}(S) \leq (1/2)\mathrm{vol}(G)$ from the definition of conductance to obtain the inequality. □

We will sketch the proof of the other side of the inequality.

Claim 2. $\phi(G) \leq \sqrt{2\mu_2}$.

Proof. The proof is algorithmic, and builds off the idea that if μ_2 is small, then there exists S with small $\phi(S)$.

1. Denote the second eigenvector as v_2.
2. For $\boldsymbol{x}$ defined as $\boldsymbol{D}^{\frac{1}{2}}\mathbf{1}_S - \sigma \boldsymbol{v}_1$, we found that $x_i > 0$ if $i \in S$ and $x_i < 0$ otherwise.
3. Building off this, we show that there exists $\theta \in \mathbb{R}$ such that $S := \{i : v_i \geq \theta\}$ satisfies $\phi(S) \leq \sqrt{2\mu_2}$.

□

This gives a recipe for performing spectral partitioning:

1. Build $\hat{\boldsymbol{L}}$
2. Compute $\boldsymbol{v}_2$
3. Test all θ, $S = \{i : v_i \geq \theta\}$
4. Choose the set S with smallest $\phi(S)$
5. Output S satisfying $\phi(S) \leq \sqrt{2\mu_2}$

It is clear that this can be done in polynomial time.

2 Multi-Way Cuts

We now generalize the approach we have developed for approximating the set of vertices of minimum conductance in a graph G to partitions consisting of more than two sets.

Fix some $k \in \mathbb{N}$ such that $k > 2$.

Definition 2 (k-way conductance)**.**

$$\phi_k(G) = \min_{\substack{S_1,\ldots,S_k \neq \varnothing \\ \textit{disjoint}}} \max_{i \in [k]} \phi(S_i)$$

Theorem 1.

$$\frac{\mu_k}{2} \leq \phi_k(G) \leq O(k^2\sqrt{\mu_k}) \leq O(\sqrt{\mu_{2k}\lg(k)})$$

We sketch an algorithmic proof for this theorem.

Proof. 1. $\boldsymbol{v}_1 \ldots, \boldsymbol{v}_k$ are k eigenvectors corresponding to the smallest k eigenvalues.

2. Write

$$\begin{bmatrix} \boldsymbol{u}_1 \\ \hline \boldsymbol{u}_2 \\ \hline \vdots \\ \hline \boldsymbol{u}_n \end{bmatrix} = \left[\ \boldsymbol{v}_1 \ \middle|\ \boldsymbol{v}_2 \ \middle|\ \cdots \ \middle|\ \boldsymbol{v}_n\ \right]$$

   for $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n \in \mathbb{R}^k$.

3. Use a variant of ball carving:
   (a) Instantiate $\hat{\boldsymbol{u}}_i = \boldsymbol{u}_i / \|\boldsymbol{u}_i\|_2$, $i \in [n]$
   (b) Pick k standard Gaussian vectors $\boldsymbol{g}_1, \ldots, \boldsymbol{g}_k \in \mathbb{R}^k$
   (c) For each $j \in [k]$:
   $$h_j(i) = \begin{cases} \langle \hat{\boldsymbol{u}}_i, \boldsymbol{g}_j \rangle & \text{if } j = \arg\max_j \langle \hat{\boldsymbol{u}}_i, \boldsymbol{g}_j \rangle \\ 0 & \text{otherwise} \end{cases}$$
   (d) For each $j \in [k]$:

i. Sort h_j, and try all thresholds θ:

$$S_j^\theta = \{i : h_j(i) \geq \theta\}$$

(e) Extract set S_j if there exists θ such that $\phi(S_j^\theta) \leq O(\sqrt{\mu_k \lg(k)})$.

4. It can be shown that, at the end, we will extract at least $k/2$ sets.

□

Remark 1. The above is of course just a sketch; one must do some massaging of the sets S_j^θ in order to get sets that satisfy the constraints in the definition of $\phi_k(G)$.

And now for something completely different.

3 Metric Embeddings

Definition 3 (Metric space)**.**
A metric space is a pair (X, d) where X is a set and $d : X \times X \to \mathbb{R}_+$ is a function satisfying (for $x_1, x_2, x_3 \in X$):

1. $d(x_1, x_2) = 0$ *if and only if* $x_1 = x_2$
2. $d(x_1, x_2) = d(x_2, x_1)$
3. $d(x_1, x_3) \leq d(x_1, x_2) + d(x_2, x_3)$.

Definition 4 (Metric embedding)**.**
Let (X, d_X) and (Y, d_Y) be metric spaces. We say $f : X \to Y$ is an embedding with distortion $D > 1$ if there exists $\lambda > 0$ such that, for all $x_1, x_2 \in X$:

$$(\lambda)d_X(x_1, x_2) \leq d_Y(f(x_1), f(x_2)) \leq (\lambda D)d_X(x_1, x_2)$$

We say the embedding is isometric if $D = 1$.

This property is indispensable. There are many cases, for example when the original metric is $\|\cdot\|_1$, where it may be unclear how to solve a problem in the original space, but after embedding in a different space the problem becomes tractable by more familiar methods.

Example 1. The following are metrics.

1. ℓ_p distances for $p \geq 1$ (Minkowski inequality)
2. Edit distance on binary strings in $\{0, 1\}^d$ (how many replacements are required to obtain one string from another)
3. Shortest path metric on an undirected weighted graph

Note that any finite metric space can be reduced to the case of the shortest path metric: just set the weights in a complete graph to be the distances between the points in the metric space.

Here is our first theorem on this topic.

Theorem 2. *Let X be a set of n points equipped with some metric. Then there exists an embedding $f : X \to \mathbb{R}^d$ where ℓ^∞ is the metric for $\mathbb{R}^d$ with distortion 1.*

This seems quite useful at first glance, but it turns out it is perhaps not so:

1. The fact that such a result exists means that ℓ^∞ is "too general": it "contains" every metric space within itself.
2. One can get $d = n$ as the embedding dimension, and this is approximately tight. (bad dependence)

The following is a more useful result.

Theorem 3 (Bourgain '85, London-Linial-Rabinovich '91)**.** *Let X be a set of n points equipped with a metric. Then there exists an embedding $f : X \to (\mathbb{R}^d,\ \ell_1/\ell_2)$ with parameters $D = O(\log(n))$ and $d = O(\log(n))$.*

Remark 2. We will see that the distortion we get above cannot be improved.

4 Application: Sparsest Cut

Given an undirected and unweighted graph $G = (V, E)$, we are interested in finding the "best" cut in this graph. Some possibilities for the meaning of "best" are:

1. Minimum cut. This can be computed efficiently, but sometimes the cut we obtain is "not useful" (Ilya's example: in a social network, we might find a guy with one friend.)
2. "Balanced" small cut.

One possibility for "balanced":

$$\arg\min_{S \subseteq V} \frac{E(S, S^c)}{|S||S^c|}$$

The minimand here is related to the conductance we defined previously when the graph G is d-regular (every vertex has degree d). In this case it is equal to $(d/n)\Theta(1)\phi(G)$.

Towards computing this quantity, recall that Cheeger gives $\Omega(\mu_2(\hat{L})) \le \phi(G) \le O(\sqrt{\mu_2(\hat{L})})$, and as a corollary there exists a polynomial time algorithm that outputs $S \subseteq V$ with $\phi(S) \le O(\sqrt{\phi(G)})$. We ask now whether this approximation is good or bad.

Case 1. $\phi(G) = \Omega(1)$. Then the approximation is very good.

Case 2. $\phi(G) \approx \frac{1}{n}$. Then the approximation yields a cut with $\phi(S) \approx 1/\sqrt{n}$. (terrible)

This motivates us to try to do better.

Theorem 4. *There exists a polynomial time algorithm outputting $S \subseteq V$ such that*

$$\frac{E(S, S^c)}{|S||S^c|} \le O(\log(n)) \cdot \min_{S^\star} \frac{E(S^\star, (S^\star)^c)}{|S^\star||(S^\star)^c|}.$$

We very briefly sketch the idea behind the algorithm in the following diagram:

$$G \xrightarrow[\text{programming}]{\text{linear}} \text{metric on the vertices} \xrightarrow{\text{Bourgain}} \ell_1 \to \text{cut}.$$

Remark 3. The current best result is due to Arora, Rao, and Vazirani, and obtains $O(\log(n)) \to O(\sqrt{\log(n)})$ in the approximation factor.

Next time we will fully describe the algorithm, assuming the theorem of Bourgain, and after that prove the theorem of Bourgain.