Advanced Approximation Algorithms **(CMU 15-854B, Spring 2008)**

Lecture 19: Sparsest Cut and L_1 Embeddings

25 March, 2008

Lecturer: Anupam Gupta *Scribe: Amitabh Basu*

1 Sparsest Cut: Problem Definition

We will be studying the Sparsest cut problem in this lecture. In this context we will see how metric methods help in the design of approximation algorithms. We proceed to define the problem and briefly give some motivation for studying the problem.

The input to the *Sparsest Cut* problem is

- A weighted graph $G = (V, E)$ with positive edge weights (or costs or capacities, as they are called in this context) c_e for every edge $e \in E$. As is usual, $n = |V|$.
- A set of pairs of vertices $\{(s_1, t_1), (s_2, t_2) \dots (s_k, t_k)\}$, with associated demands D_i between s_i and t_i.

Given such a graph, we define *sparsity* of a cut $S \subseteq V$ to be

$$\Phi(S) = \frac{c(S, \bar{S})}{D(S, \bar{S})}$$

where

$$c(S, \bar{S}) = \sum_{e \text{ s.t. } e \text{ crosses the cut } S, \bar{S}} c(e)$$

and

$$D(S, \bar{S}) = \sum_{i \text{ s.t. } s_i, t_i \text{ are separated by } S, \bar{S}} D_i$$

The objective of the Sparsest Cut problem is to find a cut S^* which minimizes; let is define $\Phi^* = \Phi(S^*) = \min_{S \subseteq V} \Phi(S)$. As an example, see Figure 1. The dashed edges are the demand edges of demand value 1. The solid edges are edges of the graph with capacity 1. The sparsest cut value for this graph is 1.

2 Motivation

Consider the special case of the Sparsest Cut problem when we have unit demand between every pair of vertices: i.e., the demands consist of all pairs $\binom{V}{2}$ and $D_{xy} = 1$ for all $x \neq y \in V$. Then we

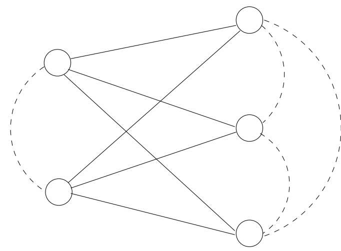

Figure 1: Example illustrating the Sparsest Cut problem

are minimizing the following quantity:

$$\Phi_{unit} = \min_{S \subseteq V} \frac{c(S, \bar{S})}{|S||\bar{S}|}. \tag{1}$$

If $|S| \leq n/2$, then $n/2 \leq |\bar{S}| \leq n$, then the expression (1) is the same (upto a factor of $\{n/2, n\}$) as

$$\min_{|S| \leq \frac{n}{2}} c(S, \bar{S})/|S|. \tag{2}$$

The above quantity—the cost of the cut edges divided by the size of the separated set S—is called the "expansion" of the set S in a weighted graph. (In unit weight graphs, this corresponds exactly to the notion of edge-expansion of graphs). Having a good approximation for this problem enables us to find good balanced separators, which are extremely useful for "divide-and-conquer" type of algorithms on graphs. See, e.g. [5] for a survey on how to find balanced separators using the sparsest cut problem as a subroutine, and for applications to approximation algorithms.

2.1 Expander Graphs

As mentioned above, for unit demands and unit edge-weights, the sparsity is approximately the same as the edge-expansion of a graph. Graphs with high (edge-)expansion are often called *expanders*.

Examples of expanders are

- For example, K_n has $\Theta(n)$ expansion; however, this is not surprising, since the degree of the graph is also large.
- Infinite 3-regular tree: this has finite degree, but is an infinite graph. See Figure 2.

The challenge is in constructing families of finite expanders with bounded-degree—i.e., we want to construct a family of graphs $\{G_n\}$ where G_n has n vertices, for infinitely many values of n, such that the expansion of graphs in this family is bounded below by a constant, and the degree is bounded above by a constant. It is not difficult to show that random bounded-degree graphs are

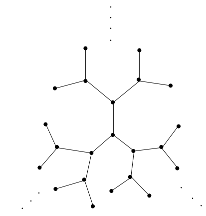

Figure 2: Infinite 3-regular tree

expanders with high probability, but how does one construct explicit expanders? By now there are many constructions known (see, e.g., the survey by Hoory, Linial, and Wigderson [3]). Here is one such construction due to Margluis.

- Consider the $n \times n$ grid graph. For every vertex (i, j), add edges to four other vertices : $(i, i+2j), (i, i+2j+1), (i+2j, j), (i+2j+1, j)$. The proof that this is an expander is quite non-trivial, but the intuition comes from considering the uniform unit square in $\Re^2$, considering two mappings $T(x, y) \Rightarrow (x+2y, y)$ and $T'(x, y) \Rightarrow (x, x+2y)$, and showing that if we take any region R, such that $area(R) \leq \frac{1}{2}$, then $T(R) \cup T'(R)$ is large.

3 Algorithms for Sparsest Cut

Here is a simple claim for sparsest cuts.

Claim 3.1. *There exists a sparsest cut* $(S, \bar{S})$*, such that the graphs* $G[S]$ *and* $G[\bar{S}]$ *are connected. (We assume that the graph was connected to begin with.)*

Proof. WLOG, assume $G[S]$ is not connected. Say it has components $C_1 \dots C_t$. Let the total capacity of edges from C_i to $\bar{C}_i$ be c_i and the demand be d_i. The sparsity of cut S is $\frac{c_1 + \dots + c_t}{d_1 + \dots d_t}$.

Now since all the quantities c_i, d_i are non-negative, by simple arithmetic, there exists i such that $c_i/d_i \leq \frac{c_1 + \dots + c_t}{d_1 + \dots d_t}$. This implies that the cut C_i is at least as good as S. □

3.1 Sparsest Cut on Trees

Using claim 3.1, we know that the sparsest cut on trees will be exactly one edge. Therefore, the sparsest cut problem on trees becomes easy to solve in polynomial time.

3.2 Metric Spaces - A Digression

A metric space on a set V is defined as a distance measure $d : V \times V \rightarrow \Re$, with 3 properties:

1. $d(x, y) = 0$ iff $x = y$

2. $d(x, y) = d(y, x)$

3. $d(x, y) + d(y, z) \geq d(x, z)$

Typical examples are $\Re^d$, equipped with the ℓ_p norms. Recall that the ℓ_p norm is the following - $d(x, y) = (\sum_{i=1}^{d} |x_i - y_i|^p)^{\frac{1}{p}}$.

3.2.1 Cut Metrics

Another example is that associated with "cuts" (i.e., subsets of V): given a cut $S \subseteq V$, the "cut metric" associated with S is δ_S, where

$$\delta_S(x, y) = \begin{cases} 0 & \text{if } x, y \in S \text{ or } x, y \in \bar{S} \\ 1 & \text{otherwise} \end{cases}$$

3.2.2 Viewing Metrics as Vectors in $\Re^{\binom{n}{2}}$

Any n-point metric can be associated with vectors in $\Re^{\binom{n}{2}}$, with each coordinate corresponding to a pair of vertices from the metric space. Given the metric d, we refer to the corresponding vector as $\bar{d}$. In this setting, the following facts are trivial :

1. $\alpha\bar{d} + (1 - \alpha)\bar{d}$ is a metric for all $0 \leq \alpha \leq 1$

2. For all $\alpha \geq 0$ and metrics $\bar{d}$, $\alpha\bar{d}$ is a metric.

Therefore, the set of all metrics forms a convex cone in $\Re^{\binom{n}{2}}$. In this setting the Sparsest Cut problem can be restated as

$$\min_{\text{all cut metrics } S} \frac{\bar{c} \cdot \bar{\delta_S}}{\bar{D} \cdot \bar{\delta_S}} \tag{3}$$

where $\bar{c}$ is the vector in $\Re^{\binom{n}{2}}$ with $\bar{c}_{ij}$ being the capacity of the edge between vertex i and j, and $\bar{D}_{ij}$ being the demand between vertex i and j; of course, $\bar{c} \cdot \bar{\delta_S}$ denotes the dot product of the two corresponding vectors.

Let us denote the positive cone generated by all cut metrics by CUT_n, i.e.,

$$CUT_n = \{\bar{d} | d = \sum_{S \subseteq V} \alpha_S \delta_S, \qquad \alpha \geq 0\}$$

Since the optimum of (3) will be achieved at an extreme point (or in this case, an *extreme ray*), we have

$$\Phi^* = \min_{d \in CUT_n} \frac{\bar{c} \cdot \bar{\delta_S}}{\bar{D} \cdot \bar{\delta_S}} \tag{4}$$

3.2.3 What is CUT_n?

It turns out that the set CUT_n is the same as a natural class of metric spaces.

Fact 3.2. *CUT_n = all n-point subsets of $\Re^t$ under the ℓ_1 norm.*

Proof. Consider any metric in CUT_n. For every S with $\alpha_S > 0$, we have a dimension and in that dimension we put value 0 for $x \in S$ and α_S for $x \in \bar{S}$. This shows $CUT_n \subseteq \ell_1$ metrics.

For the other direction, consider a set of n-points from $\Re^n$. Take one dimension d and sort the points in increasing value along that dimension. Say we get $v_1, \ldots, v_k$ as the set of distinct values. Define $k-1$ cut metrics $S_i = \{x | x_d \leq v_{i+1}\}$. Also let $\alpha_i = v_{i+1} - v_i$. Now along this dimension, $|x_d - y_d| = \sum_{i=1}^{k} \alpha_i \delta_{S_i}$. We can construct cut metrics for every dimension. Thus we have a metric in CUT_n for every n-point metric in ℓ_1. $\square$

Note that the above proof can easily be made algorithmic:

Lemma 3.3. *Given a metric $\mu \in \ell_1$ with D dimensions, there is a procedure taking time $\mathrm{poly}(n, D)$ that outputs a set of at most nD values $\alpha_S \geq 0$ for $S \subseteq V$ such that*

$$\mu = \sum \alpha_S \delta_S.$$

3.2.4 Rewriting Φ^* as optimizing over CUT_n

So we can rewrite (4) as

$$\Phi^* = \min_{d \in \ell_1} \frac{\bar{c} \cdot \bar{\delta_S}}{\bar{D} \cdot \bar{\delta_S}} \tag{5}$$

Since the sparsest cut problem is NP-hard, we can't hope to solve the above optimization problem over metrics in ℓ_1. Hence we consider a relaxation of this problem. We relax the domain of d to the set of all metrics

$$\lambda^* := \min_{d \text{ metric}} \frac{\bar{c} \cdot \bar{\delta_S}}{\bar{D} \cdot \bar{\delta_S}} \tag{6}$$

Clearly, this quantity $\lambda^* \leq \Phi^*$. Moreover, we can compute it via a linear program.

$$\begin{array}{rcl} \min c_{ij} d_{ij} & & \\ \text{subject to} \quad d_{ij} & \leq & d_{ik} + d_{kj} \\ D_{ij} d_{ij} & = & 1 \\ d_{ij} & \geq & 0 \end{array} \tag{7}$$

3.3 Metric Embeddings

Suppose we solve the LP to find the metric d that achieves the optimal value λ^* in (6): how can we obtain a cut $(S, \bar{S})$ such that $\Phi(S)$ is close to Φ^*?

The plan is to embed the metric returned by the linear program above into an ℓ_1 metric so that the distances are not changed by too much. Then if we can somehow recover a cut metric from the ℓ_1 metric with the same objective, then we would be done (modulo what we lose because of the embedding). More formally, we have the following theorem:

Theorem 3.4. *Suppose for each metric (V, d), there exists a metric $\mu = \mu(d) \in \ell_1$ such that*

$$d(x, y) \leq \mu(x, y) \leq \alpha d(x, y)$$

for all $x, y \in V$. Then the Sparsest cut LP (7) has an integrality gap of at most α.

Proof. Take the metric d returned by the linear program and consider the metric $\mu \in \ell_1$, such that $d \leq \mu \leq \alpha d$. Then

$$\bar{c} \cdot \bar{\mu} \leq \alpha \bar{c} \cdot \bar{d} \tag{8}$$

using the fact that $\mu \leq \alpha\, d$. Moreover

$$\bar{D} \cdot \bar{\mu} \geq \bar{D} \cdot \bar{d} \tag{9}$$

since $\mu \geq d$. Hence

$$\Phi(\mu) := \frac{\bar{D} \cdot \bar{\mu}}{\bar{c} \cdot \bar{\mu}} \leq \frac{\alpha \bar{D} \cdot \bar{d}}{\bar{c} \cdot \bar{d}} = \alpha \lambda^* \leq \alpha \Phi^*. \tag{10}$$

In other words, we have found a solution to (5) with value at most α times the (optimal) sparsest cut, implying that the integrality gap of the LP is at most α. □

We have the following embedding theorem due to Bourgain [2] (with the claim about the dimensions due to Linial et al. [4]):

Theorem 3.5. *For all metrics d, there exists $\mu \in \ell_1$ such that $\alpha = O(\log n)$. Moreover, the number of dimensions needed is at most $O(\log^2 n)$.*

Corollary 3.6. *The LP relaxation of Sparsest Cut has integrality gap of $O(\log n)$.*

3.4 How do we find the Sparsest Cut?

We have shown that the integrality gap between (5) and (6) is small by taking the metric d obtained by (6) and embedding it into ℓ_1. However, *we want a cut $(S, \bar{S})$ with small $\Phi(S)$ — how do we obtain that?* To put in this final piece of the solution, we need to construct a cut $(S, \bar{S})$ from the ℓ_1 metric μ such that

$$\Phi(S) \leq \Phi(\mu).$$

To do this, recall Fact 3.2 which says that any metric in ℓ_1 can be written as a positive linear combination of cuts. I.e., the metric μ can be written as $\sum \alpha_S \delta_S$ with $\alpha_S \geq$. Given this representation, we get,

$$\Phi(\mu) = \frac{\bar{c} \cdot \bar{\mu}}{\bar{D} \cdot \bar{\mu}} = \frac{\bar{c} \cdot (\sum \alpha_S \delta_S)}{\bar{D} \cdot (\sum \alpha_S \delta_S)} \tag{11}$$

$$= \frac{\sum \alpha_S (\bar{c} \cdot \delta_S)}{\sum \alpha_S (\bar{D} \cdot \delta_S)} \tag{12}$$

$$\geq \min_{S, \alpha_S > 0} \frac{\alpha_S (\bar{c} \cdot \delta_S)}{\alpha_S (\bar{D} \cdot \delta_S)} = \min_{S, \alpha_S > 0} \Phi(S). \tag{13}$$

So, we can simply pick the best cut S amongst the ones with non-zero α_S in the cut-decomposition of μ.

Finally, we use Lemma 3.3 and Theorem 3.5 to argue that the representation of μ as a positive linear combination of cut metrics can be found in $\mathrm{poly}(n)$ time, with at most $O(n \log^2 n)$ cuts. This finally proves the result that not only is the integrality gap small (Corollary 3.6), but

Theorem 3.7. *Given a metric d that is a solution to the LP (6), we can efficiently find a cut $(S, \bar{S})$ such that $\Phi(S) \leq O(\log n) \times$ LP-value.*

References

[1] Avis D. and Deza M., The cut cone, ℓ_1 -embeddability, complexity and multicommodity flows. Networks 21 (1991) 595-617.

[2] J. Bourgain. On Lipschitz embeddings of finite metric spaces in {Hilbert} space. Israel J. of Math. 1985.

[3] Shmolo Hoory, Nathan Linial, and Avi Wigderson. Expander graphs and their applications *Bulletin of the AMS*, 43(2006) 439–561.

[4] Nathan Linial, Eran London, and Yuri Rabinovich. The geometry of graphs and some of its algorithmic applications. *Combinatorica*, 15(2):215–245, 1995. (Preliminary version in *35th FOCS*, 1994).

[5] David B. Shmoys. Cut problems and their application to divide-and-conquer. In Dorit S. Hochbaum, editor, *Approximation Algorithms for NP-hard Problems*, pages 192–235. PWS Publishing, 1997.