College of Computer & Information Science | Fall 2007
Northeastern University | 20 November 2007
CS G399: Algorithmic Power Tools I | Scribe: Tao Jin

Lecture Outline:

- Review of Sparsest-Cut problem
- Frechet's-Embedding
- Bourgain's Theorem
  - Construction of Bourgain's Embedding
  - Proof of Bourgain's Theorem

In this lecture we will first review the sparsest cut problem and its relation with metric embeddings. Then we will show that for any metric space, there is an isometric Frechet's embedding. Last, we will prove establish a celebrated theorem due to Bourgain on embedding arbitrary metrics into the l_1 metric using Frechet's embedding, which will guarantee us an $O(\log n)$ approximation factor to the optimal solution of the sparsest-cut problem.

1 Sparsest-Cut Review

In the last lecture we defined the sparsest-cut problem.

Sparsest-Cut Problem:

Given a graph $G = (V, E)$ and $w : E \to Z^+$. Determine cut $(S, \bar{S})$ such that $\frac{E(S,\bar{S})}{|S||\bar{S}|}$ is minimized where $E(S, \bar{S})$ is number of edges crossing S and $\bar{S}$.

For any cut $(S, \bar{S})$, we defined the distance between two vertices in the graph as follow,

$$d_S(i,j) = \begin{cases} 1 & (i,j) \text{ } \textit{is cross cut edge} \\ 0 & \textit{otherwise} \end{cases}$$

We claim that any cut induces a so-called cut metric d_S. Also, we have proved that we could relax any metric cut to l_1 metric. Then we have,

$$\begin{aligned} \text{Sparsest Cut} &= \min_S \frac{E(S,\bar{S})}{|S||\bar{S}|} \\ &= \min_{d_S} \frac{\sum_{i,j\in E} d_S(i,j)}{\sum_{i,j} d_S(i,j)} \\ &= \min_{d_{l_1}} \frac{\sum_{i,j\in E} d_{l_1}(i,j)}{\sum_{i,j} d_{l_1}(i,j)} \qquad (1) \end{aligned}$$

Suppose there is a metric d which embeds in l_1 with distortion D. According to the definition of distortion,

$$\frac{d(i,j)}{D} \le d_{l_1}(i,j) \le d(i,j)$$

Then we have,

$$d(i,j) \geq d_{l_1}(i,j) \Rightarrow \sum_{i,j \in E} d(i,j) \geq \sum_{i,j \in E} d_{l_1}(i,j) \tag{2}$$

$$d(i,j) \leq D * d(i,j) \Rightarrow \sum_{i,j} d(i,j) \leq D * \sum_{i,j} d_{l_1}(i,j) \tag{3}$$

So by inequalities 2 and 3, for any metric d

$$\frac{\sum_{i,j \in E} d(i,j)}{\sum_{i,j} d(i,j)} \geq \frac{1}{D} \frac{\sum_{i,j \in E} d_{l_1}(i,j)}{\sum_{i,j} d_{l_1}(i,j)}$$

Since we have proved the equality (1), it is easy to see that for any metric d which embeds in l_1 with distortion D, the optimal solution to minimize $\frac{\sum_{i,j \in E} d(i,j)}{\sum_{i,j} d(i,j)}$ has approximation factor of D to the optimal solution of the sparsest-cut problem.

By Bourgain's Theorem, we can choose d to have $\text{dist}(d) = O(\log n)$, thus we can obtain an $O(\log n)$-approximate solution by solving the linear program below,

$$\min \sum_{i,j \in E} d_l(i,j)$$

$$\text{such that } \sum_{i,j} d_l(i,j) = 1$$

$$\forall i,j,k d_l(i,j) + d_l(j,k) \geq d_l(i,k)$$

2 Frechet embedding

Theorem 1 (Frechet's Theorem). *Let(V,d) be an arbitrary n-point metric space. Then there is an isometric embedding* Φ:$V \rightarrow l_\infty^n$.

Proof: Let's consider the following Frechet's embedding. Given a metric (V,d), $|V| = n$. Define $\Phi : v_i \rightarrow < d(v_i, v_1), d(v_i, v_2), ..., d(v_i, v_n) >$.

By triangle inequality,

$$\begin{aligned} \|\Phi(v_i) - \Phi(v_j)\|_\infty &= \max_t |d(v_i, v_t) - d(v_j, v_t)| \\ &\leq d(v_i, v_j) \end{aligned}$$

so the embedding is nonexpanding. On the other hand,

for all $u \in V$:

$$\|\Phi(v_i) - \Phi(v_j)\|_\infty \geq |\Phi_j(v_i) - \Phi_j(v_j)| = d(v_i, v_j)$$

where $\Phi_j(v_i)$ denotes the *j-th* coordinate of v in the new metric space. By the above two inequalities, it is easy to see that the embedding is isometric. □

Generally, it is rare to find isometric embeddings between two spaces of interst, so we relax and allow the embedding to alter the distances. By Bourgain's Theorem, any metric embeds in l_1 with $O(\log n)$ distortion. In the rest of the lecture, we will prove Bourgain's Theorem using Frechet embeddings.

3 Bourgain's Theorem

Theorem 2. *Let (V, d) be a metric space, and let n denote $|V|$. There exists an embedding ϕ from (V, d) into l_1^h, where $h = O(\log n)$, such that the distortion is $O(\log n)$. Moreover, ϕ can be computed in polynomial time by a randomized algorithm.*

3.1 Construction of Bourgain's Embedding

Given a metric (V, d) on n points, we will randomly pick k sets of nodes: $S_1, ..., S_k \subseteq V$ and define the embedding as

$$\Phi : v_i \rightarrow < d(v_i, S_1), d(v_i, S_2), ..., d(v_i, S_k) >,$$

where $d(v_i, S) = \min_{s \in S} d(v_i, s)$. Let $k = O(\log^2 n)$. For simplicity, we assume that n is power of two.

How to pick the subsets:

pick $\log n$ random sets of size 1
pick $\log n$ random sets of size 2
pick $\log n$ random sets of size 4
...
pick $\log n$ random sets of size n/2

for each $j = 0, 1, 2, ..., \log n - 1$, we randomly and independently choose $\log n$ subsets of V, each of cardinality $n/2^j$, i.e. each of the $\begin{pmatrix} n \\ n/2^j \end{pmatrix}$ substes of V of cardinality $n/2^j$ is equally likely to be chosen.

3.2 Proof of Bourgain's Theorem

Proof: Based on the embedding construction above, we claim that

$$\Omega(\log n) * d(v_i, v_j) \leq \|\phi(v_i), \phi(v_j)\|_1 \leq k * d(v_i, v_j) \tag{4}$$

Lemma 1. *For any $S \subseteq V, |d(v_i, S) - d(v_j, S)| \leq d(v_i, v_j)$.*

As shown in Figure 1, for any $S \subseteq V$, assume $d(v_i, S) = d(v_i, v_m), d(v_j, S) = d(v_j, v_n)$, where n and m could be the same,

$$\begin{aligned} |d(v_i, S) - d(v_j, S)| &= |d(v_i, v_m) - d(v_j, v_n)| \\ &\leq |d(v_i, v_m) - d(v_j, v_m)| \quad (5) \\ &\leq d(v_i, v_j) \quad (6) \end{aligned}$$

The inequality 5 follows from the definition of $d(v, S)$, and the inequality 6 follows from triangle inequality.

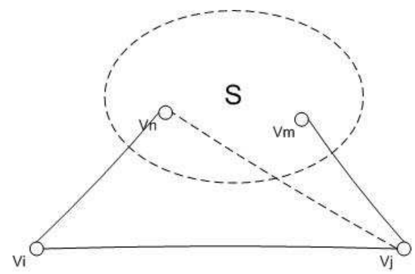

Figure 1: $|d(v_i, S) - d(v_j, S)| \leq d(v_i, v_j)$

Upper Bound:

$$\begin{aligned} \|\phi(v_i), \phi(v_j)\|_1 &= \sum_{t=1..\log^2 n} |d(v_i, S_t) - d(v_j, S_t)| \\ &\leq \log^2 n * d(v_i, v_j) \end{aligned} \tag{7}$$

The inequality 7 follows from Lemma 1. Thus, we have proved the upper bound for our claim of inequality 4.

Lower Bound:

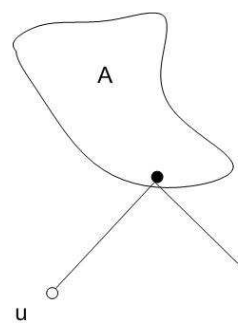

Figure 2: $d(u, A) = d(v, A)$

First we consider a simple example shown in Figure 2. Let A_{ij} denotes the jth set selected with cardinality 2^{i-1}. As shown in Figure 2, if the distance between u and set A_{ij} is the same as the one between v and A_{ij}, that means $|d(u, A_{ij}) - d(v, A_{ij})| = 0$, which doesn't contribute to the metric distance $\|\phi(u), \phi(v)\|_1$. In order to obtain the desired lower bound, we are more interested in the case that how the selected sets could contribute to $\|\phi(u), \phi(v)\|_1$.

Definition 1. $B(x, r) = \{y \in V | d(x, y) \leq r\}$. The ball includes all nodes which are at most r away from x.

Definition 2. For fixed u and v, $R_t = \min\{r \colon |B(u, r)| \geq 2^t \& |B(v, r)| \geq 2^t\}$. Further, R_t is defined if and only if $R_t \leq d(u, v)/2$, i.e. we require that $B(u, r)$ and $B(v, r)$ are disjoint, but note that they could "touch" each other.

Now we consider the case shown in Figure 3. The set A of cardinality 2^i intersects the ball $B(u, R_{i-1})$ and $B(v, R_i)$, then it is easy to see that $d(u, A) \leq R_{i-1}$ and $d(v, A) \geq R_i$. Thus, $|d(u, A) - d(v, A)| \geq R_i - R_{i-1}$.

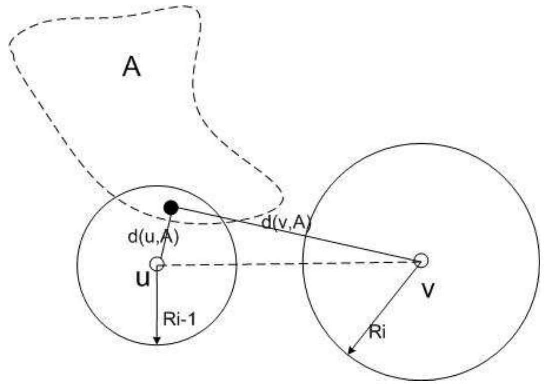

Figure 3: good ball example

Claim 1. *When we select set A of cardinality 2^{i-1}, for any i, the probability that we chose a set that intersects $B(u, R_{i-1})$ and disjoints from $B(v, R_i)$ is constant.*

Proof: By Definition 2, $|B(u, R_{i-1})| \leq 2^i$ and $|B(v, R_i)| \geq 2^i$. So the probability of choosing a set A which intersects $|B(u, R_{i-1})|$ but misses $|B(v, R_i)|$ is

$$\begin{aligned}
Pr[A \cap B(u, R_{i-1}) \neq \emptyset \& A \cap B(v, R_i) = \emptyset] &= Pr[A \cap B(u, R_{i-1}) \neq \emptyset] * Pr[A \cap B(v, R_i) = \emptyset] \\
&\geq (1 - (1 - \frac{1}{2^i})^{2^{i-1}}) * (1 - \frac{1}{2^i})^{2^i} \\
&\geq (1 - (1 - \frac{1}{2^i})^{2^i/2}) * (1 - \frac{1}{2^i})^{2^i} \\
&\geq (1 - e^{-1/2}) * \frac{1}{4}
\end{aligned}$$

□

Therefore, for a set of subsets A_{ij} of cardinality 2^{i-1},

$$\begin{aligned}
E[\sum_{j=1..\log n} |d(u, A_{ij}) - d(v, A_{ij})|] &\geq \log n \Omega(1)(R_i - R_{i-1}) \\
&\geq \Omega(\log n)(R_i - R_{i-1})
\end{aligned}$$

$$\begin{aligned}
E[\|\phi(u), \phi(v)\|_1] &\geq E[\sum_{R_i \ is \ defined} \sum_{j=1..\log n} |d(u, A_{ij}) - d(v, A_{ij})|] \\
&\geq \sum_{R_i \ is \ defined} \Omega(\log n)(R_i - R_i - 1) \qquad (8) \\
&\geq \Omega(\log n) * (R_m - R_0) \qquad (9) \\
&\geq \Omega(\log n) d(u, v) \qquad (10)
\end{aligned}$$

It is easy to see that the inequality 8 telescopes, and the inequality 9 is the result of reducing 8. By the definition of R_m, we know that the inequality 10 holds for certain.

Now we have proved $\Omega(\log n) * d(v_i, v_j) \leq \|\phi(v_i), \phi(v_j)\|_1 \leq k * d(v_i, v_j)$. Then according to the Linearity of a norm under scalar multiplication, i.e.$\|\lambda\vec{v}\| = |\lambda|\|\vec{v}\|$, we have

$$\frac{d(v_i, v_j)}{\Omega(\log n)} \leq \|\phi'(v_i), \phi'(v_j)\|_1 \leq d(v_i, v_j),$$

where $\phi'(v_i) = \phi(v_i)/log^2 n$. Thus, this embedding has distortion of $O(logn)$ in $O(log^2 n)$ dimension. $\square$