COMS W6998-5: Algorithms through Geometric Lens (Fall'17) Nov 1, 2017

Lecture 17 – Metric Embeddings II: Sparsest Cut

Instructors: *Alex Andoni, Ilya Razenshteyn* Scribes: *Bihao Zhang*

1 Main Problem: Sparsest Cut

Problem Given a graph: $G = (V, E)$.Our goal is to minimize the **sparsity** of the graph:

$$\Phi(G) := \frac{E(S, \bar{S})}{|S| \cdot |\bar{S}|}$$

where $S \subset V$ is a non-empty proper subset of vertex set V. And $E(S, \bar{S})$ denotes the sum of the weights of the edges across the cut as:

$$E(S, \bar{S}) = \sum_{e, \textbf{edge e crosses the cut}} w(e)$$

Without mentioned,we will set the weights of the edges as $w(e) = 1$ everywhere.

We'll prove the following theorem:

Theorem 1. *There exists a polynomial time algorithm outputting $S \subset V$ such that:*

$$\frac{E(S, \bar{S})}{|S| \cdot |\bar{S}|} \leq O(\log n) \min_{S^* \subset V} \frac{E(S^*, \bar{S^*})}{|S^*| \cdot |\bar{S^*}|}$$

We denote $OPT := \min_{S^* \subset V} \frac{E(S^*, \bar{S^*})}{|S^*| \cdot |\bar{S^*}|}$ as the minimum sparsity of the graph.

Here's the blueprint for proving this theorem:

$$\text{Graph} \xrightarrow[\textit{Programming}]{\textit{Linear}} \text{Metric Space} \xrightarrow{\text{Bourgain's Thm}} l_1 \xrightarrow{\text{simple}} \textit{cut}$$

2 Preparation1: Semimetrics

Firstly,we will define the concept of "semimetric" (or pseudometric):

Definition 2. *A **semimetric** on X is a function: $d : X \times X \rightarrow R$ such that:*

$$1)d(x,y) \geq 0;$$
$$2)d(x,x) = 0; (\textit{Replace the identity of indiscernibles for metrics})$$
$$3)d(x,y) = d(y,x);$$
$$4)d(x,z) \leq d(x,y) + d(y,z);$$

Now,for a cut presented by vertex set S,we'll have the following function is a semimetric (by easily checking each condition):

Definition 3. *For a graph $G = (V, E)$ and a cut presented by a vertex set S,function $d_S : V \times V \to R$ is defined as follows.*

$$d_S(u,v) = \begin{cases} 1 & \textit{u,v are separated by cut } S \\ 0 & \textbf{\textit{otherwise}} \end{cases}$$

And the function d_S is called ***the semimetric of the cut*** *S.*

Now based on the concept of semimetrics on the graph,we define a more general "optimal"

Definition 4. *We denote OPT^* as the minimum of the following function (d is chosen from the semimetrics on the graph):*

$$OPT^* := \min_{d-semimetric\ on\ V} \frac{\sum_{(u,v)=e\in E} d(u,v)}{\sum_{u,v} d(u,v)}$$

And:

$$d^* := \arg\min_{d-semimetric\ on\ V} \frac{\sum_{(u,v)=e\in E} d(u,v)}{\sum_{u,v} d(u,v)}$$

Here,d^* may not be unique.

Then,we'll have two claims:

Claim 5. *$OPT \geq OPT^*$*

This claim can be deduced from the statement that for every cut $V = S \cup \bar{S}$,d_S is a semimetric on the graph G.

Claim 6. *Given a graph $G = (V, E)$,we can compute OPT^* and a d^* in polynomial time.*

Proof. We can prove this claim by showing the problem can be turned into a linear programming problem:

That is,for $n(n-1)/2, (n = |V|)$ variables $d(u,v), (u \in V, v \in V, u \neq v)$.We are going to calculate :

$$\min \sum_{(u,v)=e\in E} d(u,v)$$

under the constraints:

$$\begin{aligned} &1) d(u,v) \geq 0 \\ &2) d(u,v) + d(v,w) \geq d(u,w) \\ &3) \sum_{u,v} d(u,v) = 1 \end{aligned}$$

And using the result from linear programming we'll get we can solve this problem in polynomial time. □

3 Preparation 2: What can we get after embedding into l_1?

We claimed in our blueprint that after we embed the (semi)metric on the graph into l_1 we can simply get the cut we want. We'll show this fact in this section.

From section 2,we have shown that we can get d^* and OPT^* in polynomial time. And start by this result.We'll prove the following claims:

Claim 7 (Claims for embedding without distortion)**.** *Assume that* $d^*(u,v) = ||f(u) - f(v)||_1$ *for some* $f : V \to R^d$*.Then there exists some* $S \subset V$*,such that:*

$$\frac{E(S,\bar{S})}{|S| \cdot |\bar{S}|} \leq OPT^* \leq OPT$$

Also,we can find an polynomial time algorithm outputting such S*.*

Proof. We'll use the following two results:

1) There exists a set of weights $\omega_S \geq 0$,such that:

$$||f(u) - f(v)||_1 = \sum_{S \subset V} \omega_S d_S(u,v)$$

holds for every $u, v \in V$

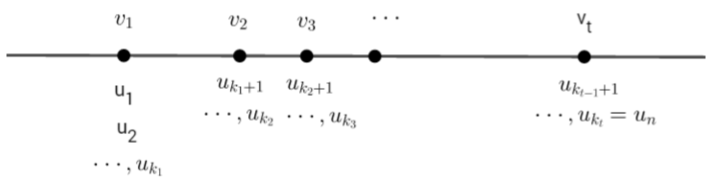
We can prove this statement in two steps:

1.1)It suffices to show that the statement holds for $d = 1$. It is because we can easily decompose R^d space into the straight sum of d R^1 space and then summing up the result as we are using the l_1 norm for $f(u)$.

1.2)Now for the graph G.We sort the vertex of the graph in the way that their corresponding $f(u)$ are listed in a ascending order.That is:

$$\begin{aligned} &f(u_1) = f(u_2) = \cdots f(u_{k_1}) = v_1 \\ &< f(u_{k_1+1}) = \cdots f(u_{k_2}) = v_2 \\ &< \cdots \\ &< f(u_{k_p}) = \cdots f(u_n) = v_t, n = |V|, k_t = n \end{aligned}$$

We can show the list using the figure below.



Now,we claim that for any $u, v \in V$,we'll have:

$$||f(u) - f(v)||_1 = \sum_{r=1}^{t-1} |v_{r+1} - v_r| \cdot d_{\{u_1,...u_{k_r}\}}(u, v)$$

Follows from for any $p, q \in \{1, ..., t\}$,we assume $p < q$ then we can decompose $||v_q - v_p|| = \sum_{r=p}^{q-1} |v_{r+1} - v_r|$,and noticing that $d_{\{u_1,...,u_{k_r}\}}(u, v) = 1$,for $f(u) \leq f(u_{k_r}) \leq f(v), f(u) \neq f(v)$.

2) For any series of number: $\omega_i \geq 0, A_i, B_i \geq 0, i \in \{1, ..., n\}$.We have such an $k \in \{1, ..., n\}$ exists,that k satisfies:

$$\frac{A_k}{B_k} \leq \frac{\sum_i \omega_i A_i}{\sum_i \omega_i B_i}$$

Now from the first statement,we can represent OPT^* as following:

$$\begin{aligned} OPT^* &= \frac{\sum_{(u,v)\in E} ||f(u) - f(v)||}{\sum_{u,v} ||f(u) - f(v)||} \\ &= \frac{\sum_S \omega_S \sum_{(u,v)\in E} d_S(u, v)}{\sum_S \omega_S \sum_{u,v} d_S(u, v)} \end{aligned}$$

And applying the second statement to this equation we will have:

$$\exists S, \frac{\sum_{(u,v)\in E} d_S(u, v)}{\sum_{(u,v)} d_S(u, v)} \leq OPT^*$$

Thus we've accomplished the proof.

□

Remark:We can easily deduce an polynomial time algorithm outputting the S from the prove.Simply speaking,the algorithm takes three steps:

i)Calculating all the coordinate of function f;

ii)Sorting the value of every vertex by the coordinate of function f;

iii)Try all "prefixes cuttings".

If we embed the semimetric on the graph into l_1 with distortion,we can fix the claim above into:

Claim 8 (Claims for embedding with distortion)**.** *Assume that $d^*(u,v) \leq ||f(u) - f(v)||_1 \leq D \cdot d^*(u,v)$ for some $f : V \to R^d$ and a constant D (that is embedding with distortion D).Then there exists some $S \subset V$,such that:*

$$\frac{E(S,\bar{S})}{|S| \cdot |\bar{S}|} \leq D \cdot OPT^* \leq D \cdot OPT$$

Also,we can find an polynomial time algorithm outputting such S.

Proof. We can just follow the similar steps for proving **Claim 7**.Just using the following property:

$$\frac{\sum_{(u,v)\in E} ||f(u) - f(v)||_1}{\sum_{u,v} ||f(u) - f(v)||_1} \leq \frac{D \cdot \sum_{(u,v)\in E} d^*(u,v)}{\sum_{u,v} d^*(u,v)} = D \cdot OPT^*$$

□

4 Metric Embedding: Bourgain's Theorem

Hence now the key point is that how to embed the semimetric on the graph into l_1 norm with as small distortion as possible.Actually,we have the following theorem:

Theorem 9 (Bourgain's Theorem)**.** *For every X is a n points (semi)metric space,then there exists a function: $f : X \to R^{O(\log^2 n)}$ s.t. f is an embedding with distortion $O(\log n)$.*

And we can calculate this mapping in polynomial time.

*Remark:*We will see that the distortion $O(\log n)$ is tight.

Now we will introduce Frechet Embeddings

Definition 10 (Frechet Embeddings)**.** *For d **nonempty** subsets of X: $S_1 \subset X$, $S_2 \subset X$,...,$S_d \subset X$. We define a function:$f : X \to R^d$ as* Frechet Embeddings*:*

$$f(x) = (f_{S_1}(x), \cdots, f_{S_d}(x))$$

Where $f_{S_i}(x) := dist(x, S_i) = \min_{\bar{x}\in S} d_X(x, \bar{x})$.

Here,we have two claims:

Claim 11.

$$\forall S, |f_S(u) - f_S(v)| \leq d_X(u,v)$$

This can be proved by using the triangle inequality for semimetrics.And applying this claim for each dimension of $f(x)$,we'll have:

Claim 12.

$$||f(u) - f(v)||_1 \leq d \cdot d_X(u,v)$$

Now we construct a series of sets as following:

For $1 \leq i \leq \lfloor \log_2 n \rfloor, 1 \leq j \leq O(\log n)$.We decide whether $x \in S_{ij}$ by applying an independent "pick-up" with the possibility $Pr[x \in S_{ij}] = 2^{-i}$

Then applying **Claim 12**,we will have:

Lemma 13. *if* $g(x) = (f_{S_{ij}(x)})_{i,j}$*,then:*

$$||g(u) - g(v)||_1 \leq O(\log^2 n) d_X(u, v)$$

As this is the result for the upper bound for mapping g,we also need the estimate for lower bound of g.

We notice that if we want $|f_S(u) - f_S(v)| \geq \delta$,we must think of the r satisfies $f_S(u) \leq r$ and $f_S(v) \geq r + \delta$.Thus we have:

Lemma 14. *If there exists* r *satisfies* $|B(u, r)| = K$ *(*$|B(u, r)| =$ *number of points in* X *with distance from* u *less than* r*),and* $|B(v, r + \delta)| \in [k/100, 100k]$*.We will have:*

$$Pr[|f_{S_{\log k,j}(u)} - f_{S_{\log k,j}(v)}| \geq \delta] \geq \Omega(1)$$

5 Reference

[Bou85] J. Bourgain. On Lipschitz embeddingof nite metric spaces in Hilbert space. Israel J. Math., 52(1-2):4652, 1985.

[LLR94] Nathan Linial, Eran London, and Yuri Rabinovich. The geometry of graphs and some of its algorithmic applications. In 35th Annual Symposium on Foundations of Computer Science, pages 577591, Santa Fe, New Mexico, 2022 November 1994. IEEE.

[LR88] Tom Leighton and Satish Rao. An approximate max-ow min-cut theorem for uniform multicommodity ow problems with applications to approximation algorithms. In 29th AnnualSymposium on Foundationsof Computer Science, pages 422431, White Plains, New York, 2426 October 1988. IEEE.