Stanford University — Graph Partitioning and Expanders Handout 4
Luca Trevisan May 15, 2013

Lecture 4

In which we introduce the Leighton-Rao linear programming relaxation of the non-uniform sparsest cut problem, and we show how to round its solutions via a theorem of Bourgain about embedding arbitrary metric spaces into ℓ_1.[1]

1 The Non-uniform sparsest cut problem

Let $G = (V, E)$ be an undirected graph. Unlike past lectures, we will not need to assume that G is regular. Recall that, for a subset $S \subseteq V$, we defined the *sparsity* of the partition $(S, V - S)$ as

$$\sigma_G(S) := \frac{\frac{1}{|E|} \cdot E(S, V - S)}{\frac{2}{V^2} \cdot |S| \cdot |V - S|}$$

which is the ratio between the fraction of edges that are cut by $(S, V - S)$ and the fraction of pairs of vertices that are disconnected by the removal of those edges.

More generally, give two (possibly weighted) undirected graphs $G = (V, E_G)$ and $H = (V, E_H)$ over the same set of vertices, we define the non-uniform sparsity of $S \subseteq V$ as

$$\sigma_{G,H}(S) := \frac{|E_H|}{|E_G|} \cdot \frac{E_G(S, V - S)}{E_H(S, V - S)}$$

where we denote by $|E_G|$ the total weight of the edges of G, and by $E_G(S, V - S)$ the total weight of the edges of G that have one endpoint in S and one endpoint in $V - S$.

For graphs G, H with the same vertex set V, the non-uniform sparsest cut problem is to find

$$\sigma(G,H) := \min_{S \subseteq V} \quad \sigma_{G,H}(S)$$

Notice that $\sigma_G(S)$ is the same as $\sigma_{G,H}$ where H is a clique in which every vertex has a self-loop of weight $1/2$.

If d_v is the degree of v in G, then define H to be the graph in which the edge $\{u,v\}$ has weight $d_u \cdot d_v$. Then, $|E_G| = vol_G(V)/2$, $|E_H| = \frac{1}{2}\sum_u \cdot (\sum_v d_u d_v) = \frac{1}{2}(\sum_v d_v)^2$, $E_H(S, V-S) = vol_G(S) \cdot vol_G(V-S)$, and the sparsity of S is

$$\sigma_{G,H}(S) = \left(\sum_v d_v \right) \cdot \frac{E(S,\bar{S})}{vol_G(S) \cdot vol_G(V-S)}$$

and if S is such that $vol(S) \leq vol(V-S)$, then $\phi(G) \leq \sigma_{G,H}(S) \leq 2\phi(G)$.

Thus the non-uniform sparsest cut problem generalizes the (uniform) sparsest cut problem that we described before, and, for a proper choice of H, is a 2-approximation of the conductance of G.

Notice also that if H is a graph that has only the one edge $\{s,t\}$, then $\sigma(G,H)$ is the (s,t)-min-cut problem for the graph G.

2 A Linear Programming relaxation

Another way to formulate the sparsest cut problem is

$$\sigma(G,H) := \frac{|E_H|}{|E_G|} \cdot \min_{\mathbf{x} \in \{0,1\}^n} \frac{\sum_{\{u,v\}} G_{u,v} |x_u - x_v|}{\sum_{\{u,v\}} H_{u,v} |x_u - x_v|}$$

where $G_{u,v}$ is the weight of the edge $\{u,v\}$ in G and $H_{u,v}$ is the weight of the edge $\{u,v\}$ in H.

The observation that led us to see λ_2 as the optimum of a continuous relaxation of $\sigma(G)$ was to observe that, for a boolean vector $\mathbf{x}$, $|x_u - x_v| = |x_u - x_v|^2$, and then relax the problem by allowing arbitrary vectors $\mathbf{x}$ instead of just boolean vectors.

The Leighton-Rao relaxation of sparsest cut is obtained using, instead, the following observation: if, for a set S, $\mathbf{x}$ is the boolean indicator vector of S and we define $d_S(u,v) := |x_u - x_v|$, then $d_S(\cdot,\cdot)$ defines a semi-metric over the set V, because d_S is symmetric, $d_S(v,v) = 0$, and the triangle inequality holds. So we could think about allowing *arbitrary semi-metrics* in the expression for σ, and define

$$LR(G,H) := \min_{\substack{d : V \times V \to \mathbb{R} \\ d \text{ semi-metric}}} \frac{|E_H|}{|E_G|} \cdot \frac{\sum_{\{u,v\}} G_{u,v} d(u,v)}{\sum_{\{u,v\}} H_{u,v} d(u,v)} \tag{1}$$

This might seem like such a broad relaxation that there could be graphs on which $LR(G,H)$ bears no connection to $\sigma(G,H)$. Instead, we will prove the fairly good estimate

$$LR(G,H) \leq \phi(G,H) \leq O(\log |V|) \cdot LR(G,H) \tag{2}$$

Furthermore, we will show that $LR(G,H)$, and an optimal solution $d(\cdot,\cdot)$ can be computed in polynomial time, and the second inequality above has a constructive proof, from which we derive a polynomial time $O(\log |V|)$-approximate algorithm for sparsest cut.

3 Formulating the Leighton-Rao Relaxation as a Linear Program

The value $LR(G,H)$ and an optimal $d(\cdot,\cdot)$ can be computed in polynomial time by solving the following linear program

$$\begin{array}{lll} \text{minimize} & \sum_{\{u,v\}} G_{u,v} d_{u,v} & \\ \text{subject to} & & \\ & \sum_{\{u,v\}} H_{u,v} d_{u,v} = \frac{|E_H|}{|E_G|} & \\ & d_{u,w} \leq d_{u,w} + d_{w,v} & \forall u,v,w \in V \\ & d_{u,v} \geq 0 & \forall u,v \in V \end{array} \tag{3}$$

that has a variable $d_{u,v}$ for every unordered pair of distinct vertices $\{u,v\}$. Clearly, every solution to the linear program (3) is also a solution to the right-hand side of the definition (1) of the Leighton-Rao parameter, with the same cost. Also every semi-metric can be normalized so that $\sum_{\{u,v\}} H_{u,v} d(u,v) = \frac{|E_H|}{|E_G|}$ by multiplying every distance by a fixed constant, and the normalization does not change the value of the right-hand side of (1); after the normalization, the semimetric is a feasible solution to the linear program (3), with the same cost.

In the rest of this lecture, we will show how to round a solution to (3) into a cut, achieving the logarithmic approximation promised in (2).

4 An L1 Relaxation of Sparsest Cut

In the Leighton-Rao relaxation, we relax distance functions of the form $d(u,v) = |x_u - x_v|$, where $\mathbf{x}$ is a boolean vector, to completely arbitrary distance functions. Let us consider an intermediate relaxation, in which we allow distance functions that can be realized by *an embedding of the vertices in an ℓ_1 space.*

Recall that, for a vector $\mathbf{x} \in \mathbb{R}^n$, its ℓ_1 norm is defined as $||\mathbf{x}||_1 := \sum_i |x_i|$, and that this norm makes $\mathbb{R}^n$ into a metric space with the ℓ_1 distance function

$$||\mathbf{x} - \mathbf{y}||_1 = \sum_i |x_i - y_i|$$

The distance function $d(u,v) = |x_u - x_v|$ is an example of a distance function that can be realized by mapping each vertex to a real vector, and then defining the distance between two vertices as the ℓ_1 norm of the respective vectors. Of course it is an extremely restrictive special case, in which the dimension of the vectors is one, and in which every vertex is actually mapping to either zero or one. Let us consider the relaxation of sparsest cut to arbitrary ℓ_1 mappings, and define

$$\sigma'(G,H) := \inf_{m, f: V \to \mathbb{R}^m} \quad \frac{|E_H|}{|E_G|} \cdot \frac{\sum_{\{u,v\}} G_{u,v} \cdot ||f(u) - f(v)||_1}{\sum_{\{u,v\}} H_{u,v} \cdot ||f(u) - f(v)||_1}$$

This may seem like another very broad relaxation of sparsest cut, whose optimum might bear no correlation with the sparsest cut optimum. The following theorem shows that this is not the case.

Theorem 1 *For every graphs G, H, $\sigma(G,H) = \sigma'(G,H)$.*

Furthermore, there is a polynomial time algorithm that, given a mapping $f : V \to \mathbb{R}^m$, finds a cut S such that

$$\sigma_{G,H}(S) \le \frac{|E_H|}{|E_G|} \cdot \frac{\sum_{\{u,v\}} G_{u,v} ||f(u) - f(v)||_1}{\sum_{u,v} H_{u,v} ||f(u) - f(v)||_1} \tag{4}$$

PROOF: We use ideas that have already come up in the proof the difficult direction of Cheeger's inequality. First, we note that for every nonnegative reals $a_1, \ldots, a_m$ and positive reals $b_1, \ldots, b_m$ we have

$$\frac{a_1 + \cdots a_m}{b_1 + \cdots b_m} \ge \min_i \frac{a_i}{b_i} \tag{5}$$

as can be seen by noting that

$$\sum_j a_j = \sum_j b_j \cdot \frac{a_j}{b_j} \geq \left(\min_i \frac{a_i}{b_i} \right) \cdot \sum_j b_j$$

Let $f_i(v)$ be the i-th coordinate of the vector $f(v)$, thus $f(v) = (f_1(v), \ldots, f_m(v))$. Then we can decompose the right-hand side of (4) by coordinates, and write

$$\begin{aligned}
&\frac{\sum_{\{u,v\}} G_{u,v} ||f(u) - f(v)||_1}{\sum_{\{u,v\}} H_{u,v} ||f(u) - f(v)||_1} \\
&= \frac{\sum_i \sum_{\{u,v\}} G_{u,v} |f_i(u) - f_i(v)|}{\sum_i \sum_{\{u,v\}} H_{u,v} |f_i(u) - f_i(v)|} \\
&\geq \min_i \frac{\sum_{\{u,v\}} G_{u,v} |f_i(u) - f_i(v)|}{\sum_{\{u,v\}} H_{u,v} |f_i(u) - f_i(v)|}
\end{aligned}$$

This already shows that, in the definition of ϕ', we can map, with no loss of generality, to 1-dimensional ℓ_1 spaces.

Let i^* be the coordinate that achieves the minimum above. Because the cost function is invariant under the shifts and scalings (that is, the cost of a function $x \to f(x)$ is the same as the cost of $x \to af(x) + b$ for every two constants $a \neq 0$ and b) there is a function $g : V \to \mathbb{R}$ such that g has the same cost function as f_{i*} and its range is such that

$$\max_v g(v) - \min_v g(v) = 1$$

Let us now pick a threshold t uniformly at random from the interval $[\min_v g(v), \max_v g(v)]$, and define the random variables

$$S_t := \{v : g(v) \leq t\}$$

We observe that for every pairs of vertices u, v we have

$$\mathbb{E} \, |1_{S_t}(u) - 1_{S_t}(v)| = |g(u) - g(v)|$$

and so we get

$$\begin{aligned}
&\frac{\sum_{\{u,v\}} G_{u,v} ||f(u) - f(v)||_1}{\sum_{\{u,v\}} H_{u,v} ||f(u) - f(v)||_1} \\
&\geq \frac{\sum_{\{u,v\}} G_{u,v} |g(u) - g(v)|}{\sum_{\{u,v\}} H_{u,v} |g(u) - g(v)|}
\end{aligned}$$

$$= \frac{\mathbb{E}\sum_{\{u,v\}} G_{u,v}|1_{S_t}(u) - 1_{S_t}(v)|}{\mathbb{E}\sum_{\{u,v\}} H_{u,v}|1_{S_t}(u) - 1_{S_t}(v)|}$$

Finally, by an application of (5), we see that there must be a set S among the possible values of S_t such that (4) holds.

Notice that the proof was completely constructive: we simply took the coordinate f_{i^*} of f with the lowest cost function, and then the "threshold cut" given by f_{i^*} with the smallest sparsity. $\square$

5 A Theorem of Bourgain

We will derive our main result (2) from the L1 "rounding" process of the previous section, and from the following theorem of Bourgain (the efficiency considerations are due to Linial, London and Rabinovich).

Theorem 2 (Bourgain) *Let $d : V \times V \to \mathbb{R}$ be a semimetric defined over a finite set V. Then there exists a mapping $f : V \to \mathbb{R}^m$ such that, for every two elements $u, v \in R$,*

$$||f(u) - f(v)||_1 \leq d(u,v) \leq ||f(u) - f(v)||_1 \cdot c \cdot \log |V|$$

where c is an absolute constant. Given d, the mapping f can be found with high probability in randomized polynomial time in $|V|$.

To see that the above theorem of Bourgain implies (2), consider graphs G, H, and let d be the optimal solution of the Leighton-Rao relaxation of the sparsest cut problem on G, H, and let $f : V \to \mathbb{R}$ be a mapping as in Bourgain's theorem applied to d. Then

$$LR(G,H) = \frac{|E_H|}{|E_G|} \cdot \frac{\sum_{\{u,v\}} G_{u,v} d(u,v)}{\sum_{\{u,v\}} H_{u,v} d(u,v)}$$

$$\geq \frac{|E_H|}{|E_G|} \cdot \frac{\sum_{\{u,v\}} G_{u,v}||f(u) - f(v)||_1}{c \cdot \log |V| \cdot \sum_{\{u,v\}} H_{u,v}||f(u) - f(v)||_1}$$

$$\geq \frac{1}{c \cdot \log |V|} \cdot \sigma(G,H)$$

The theorem has a rather short proof, but there is an element of "magic" to it. We will discuss several examples and we will see what approaches are suggested by the examples. At the end of the discussion, we will see the final proof as, hopefully, the "natural" outcome of the study of such examples and failed attempts.

6 Preliminary and Motivating Examples

A first observation is that embeddings of finite sets of points into L1 can be equivalently characterized as probabilistic embeddings into the real line.

Fact 3 *For every finite set V, dimension m, and mapping $F : V \to \mathbb{R}^m$, there is a finitely-supported probability distribution D over functions $f : V \to \mathbb{R}$ such that for every two points $u, v \in V$:*

$$\mathop{\mathbb{E}}_{f \sim D} |f(u) - f(v)| = ||F(u) - F(v)||_1$$

Conversely, for every finite set V and finitely supported distribution D over functions $f : V \to \mathbb{R}$, there is a dimension m and a mapping $F : V \to \mathbb{R}^m$ such that

$$\mathop{\mathbb{E}}_{f \sim D} |f(u) - f(v)| = ||F(u) - F(v)||_1$$

PROOF: For the first claim, we write $F_i(v)$ for the i-th coordinate of $F(v)$, that is $F(v) = (F_1(v), \ldots, F_m(v))$, and we define D to be the uniform distribution over the m functions of the form $x \to m \cdot F_i(x)$.

For the second claim, if the support of D is the set of functions $\{f_1, \ldots, f_m\}$, where function f_i has probability p_i, then we define $F(v) := (p_1 f_1(v), \ldots, p_m f_m(v))$. $\square$

It will be easier to reason about probabilistic mappings into the line, so we will switch to the latter setting from now on.

Our task is to associate a number to every point v, and the information that we have about v is the list of distances $\{d(u, v)\}$. Probably the first idea that comes to mind is to pick a random reference vertex $r \in V$, and work with the mapping $v \to d(r, v)$, possibly scaled by a multiplicative constant. (Equivalently, we can think about the deterministic mapping $V \to \mathbb{R}^{|V|}$, in which the vertex v is mapped to the sequence $(d(u_1, v), \ldots, d(u_n, v)$, for some enumeration $u_1, \ldots, u_n$ of the elements of V.)

This works in certain simple cases.

Example 4 (Cycle) *Suppose that $d(\cdot, \cdot)$ is the shortest-path metric on a cycle, we can see that, for every two points on the cycle, $\mathbb{E}_{r \in V} |d(r, u) - d(r, v)|$ is within a constant factor of their distance $d(u, v)$. (Try proving it rigorously!)*

Example 5 (Simplex) *Suppose that $d(u, v) = 1$ for every $u \neq v$, and $d(u, u) = 0$. Then, for every $u \neq v$, we have $\mathbb{E}_{r \in V} |d(r, u) - d(r, v)| = \mathbb{P}[r = u \vee r = v] = 2/n$, so, up to scaling, the mapping incurs no error at all.*

But there are also simple examples in which this works very badly.

Example 6 (1-2 Metric) *Suppose that for every $u \neq v$ we have $d(u,v) \in \{1,2\}$ (any distance function satisfying this property is always a metric) and that, in particular, there is a special vertex z at distance 2 from all other vertices, while all other vertices are at distance 1 from each other. Then, for vertices u,v both different from z we have, as before*

$$\mathbb{E}[|d(r,u) - d(r,v)|] = \frac{2}{n}$$

but for every v different from z we have

$$\mathbb{E}[|d(r,z) - d(r,v)|] = \frac{n-2}{n} \cdot |2-1| + \frac{1}{n} \cdot |2-0| + \frac{1}{n} \cdot |0-2| = 1 + \frac{2}{n}$$

and so our error is going to be $\Omega(n)$ instead of the $O(\log n)$ that we are trying to establish.

Maybe the next simplest idea is that we should pick at random several reference points $r_1, \ldots, r_k$. But how do we combine the information $d(r_1,u), \ldots, d(r_k,u)$ into a single number to associate to u? If we just take the sum of the distances, we are back to the case of sampling a single reference point. (We are just scaling up the expectation by a factor of k.)

The next simplest way to combine the information is to take either the maximum or the minimum. If we take the minimum, we see that we have the very nice property that we immediately guarantee that our distances in the L1 embedding are no bigger than the original distances, so that it "only" remains to prove that the distances don't get compressed too much.

Fact 7 *Let $d : V \times V \to \mathbb{R}$ be a semimetric and $A \subseteq V$ be a non-empty subset of points. Define $f_A : V \to \mathbb{R}$ as*

$$f_A(v) := \min_{r \in A} d(r,v)$$

Then, for every two points u,v we have

$$|f_A(u) - f_A(v)| \leq d(u,v)$$

PROOF: Let a be the point such that $d(a,u) = f_A(u)$ and b be the point such that $d(b,v) = f_A(v)$. (It's possible that $a = b$.) Then

$$f_A(u) = d(a,u) \geq d(v,a) - d(u,v) \geq d(v,b) - d(u,v) = f_A(v) - d(u,v)$$

and, similarly,

$$f_A(v) = d(b,v) \geq d(u,b) - d(u,v) \geq d(u,a) - d(u,v) = f_A(u) - d(u,v)$$

□

Is there a way to sample a set $A = \{r_1, \ldots, r_k\}$ such that, for every two points u, v, the expectation $\mathbb{E}\,|f_A(u) - f_A(v)|$ is not too much smaller than $d(u,v)$? How large should the set A be?

Example 8 (1-2 Metric Again) *Suppose that for every $u \neq v$ we have $d(u,v) \in \{1,2\}$, and that we pick a subset $A \subseteq V$ uniformly at random, that is, each event $r \in A$ has probability $1/2$ and the events are mutually independent.*

Then for every $u \neq v$:

$$\frac{1}{4} \cdot d(u,v) \leq |\,\mathbb{E}\,|f_A(u) - f_A(v)| \leq d(u,v)$$

because with probability $1/2$ the set A contains exactly one of the elements u, v, and conditioned on that event we have $|f_A(u) - f_A(v)| \geq 1$ (because one of $f_A(u), f_A(v)$ is zero and the other is at least one), which is at least $d(u,v)/2$.

If we pick A uniformly at random, however, we incur an $\Omega(n)$ distortion in the case of the shortest path metric on the cycle. In all the examples seen so far, we can achieve constant distortion if we "mix" the distribution in which A is a random set of size 1 and the one in which A is a chosen uniformly at random among all sets, say by sampling from the former probability with probability $1/2$ and from the latter with probability $1/2$.

Example 9 (Far-Away Clusters) *Suppose now that $d(\cdot,\cdot)$ has the following structure: V is partitioned into clusters $B_1, \ldots, B_k$, where $|B_i| = i$ (so $k \approx \sqrt{2n}$), and we have $d(u,v) = 1$ for vertices in the same cluster, and $d(u,v) = n$ for vertices in different clusters.*

If u, v are in the same cluster, then $d(u,v) = 1$ and

$$\mathbb{E}\,|f_A(u) - f_A(v)| = \mathbb{P}[A \textit{ contains exactly one of } u, v]$$

If u, v are in different clusters B_i, B_j, then $d(u,v) = n$ and

$$\mathbb{E}\,|f_A(u) - f_A(v)| \approx n\,\mathbb{P}[A \textit{ intersects exactly one of } B_i, B_j]$$

If we want to stick to this approach of picking a set A of reference elements according to a certain distribution, and then defining the map $f_A(v) := \min_{r \in A} d(r, v)$, then the set A must have the property that for every two sets S, T, there is at least a probability p that A intersects exactly one of S, T, and we would like p to be as large as possible, because the distortion caused by the mapping will be at least $1/p$.

This suggest the following distribution D:

1. Sample a power of two t uniformly at random in $\{1, 2, 4, \ldots, 2^{\lfloor \log_2 n \rfloor}\}$
2. Sample $A \subseteq V$ by selecting each $v \in V$, independently, to be in A with probability $1/t$ and to be in $V - A$ with probability $1 - 1/t$.

This distribution guarantees the above property with $p = 1/O(\log n)$.

Indeed, the above distribution guarantees a distortion at most $O(\log n)$ in all the examples encountered so far, including the tricky example of the clusters of different size. In each example, in fact, we can prove the following claim: for every two vertices u, v, there is a scale t, such that conditioned on that scale being chosen, the expectation of $|f_A(u), f_A(v)|$ is at least a constant times $d(u, v)$. We could try to prove Bourgain's theorem by showing that this is true in every semimetric.

Let us call D_t the conditional distribution of D conditioned on the choice of a scale t. We would like to prove that for every semimetric $d(\cdot, \cdot)$ and every two points u, v there is a scale t such that

$$\mathop{\mathbb{E}}_{A \sim D_t} |f_A(u) - f_A(v)| \geq \Omega(d(u, v))$$

which, recalling that $|f_A(u) - f_A(v)| \leq d(u, v)$ for every set A, is equivalent to arguing that

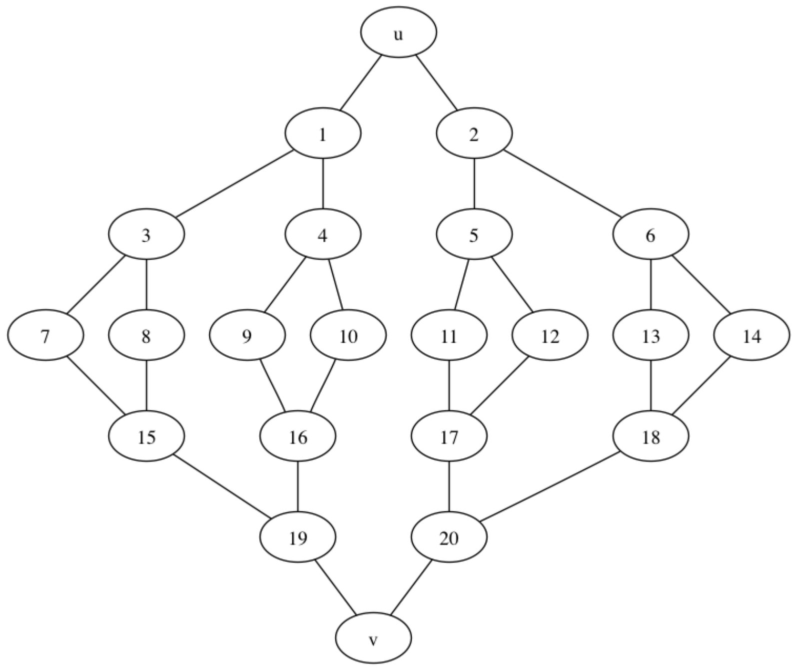
$$\mathop{\mathbb{P}}_{A \sim D_t} [|f_A(u) - f_A(v)| \geq \Omega(d(u, v))] \geq \Omega(1)$$

For this to be true, there must be distances d_1, d_2 such that $d_1 - d_2 \geq \Omega(d(u, v))$ and such that, with constant probability according to D_t, we have $f_A(u) \geq d_1$ and $f_A(v) \leq d_2$ (or vice-versa). For this to happen, there must be a constant probability that A avoids the set $\{r : d(u, r) < d_1\}$ and intersects the set $\{r : d(v, r) \leq d_2\}$. For this to happen, both sets must have size $\approx t$.

This means that if we want to make this "at least one good scale for every pair of points" argument work, we need to show that for every two vertices u, v there is a "large" distance d_1 and a "small" distance d_2 (whose difference is a constant times $d(u, v)$) such that a large-radius ball around one of the vertices and a small-radius ball around the other vertex contain roughly the same number of elements of V.

Consider, however, the following example.

Example 10 (Joined Trees) *Consider the graph obtained by taking two complete binary trees of the same size and identifying their leaves, as in the picture below.*



Consider the shortest-path metric $d(\cdot,\cdot)$ in the above graph. Consider the "root" vertices u and v. Their distance $d(u,v)$ is $\approx \log n$, but, at every scale t, both $f_A(u)$ and $f_A(v)$ are highly concentrated around t and, it can be calculated that, at every scale t, we have

$$\mathop{\mathbb{E}}_{A\sim D_t}[|f_A(u) - f_A(v)|] = \Theta(1)$$

This is still good, because averaging over all scales we still get

$$\mathop{\mathbb{E}}_{A\sim D}[|f_A(u) - f_A(v)|] \geq \Omega(1) = \frac{1}{O(\log n)} \cdot d(u,v)$$

but this example shows that the analysis cannot be restricted to one good scale but has, in some cases, to take into account the contribution to the expectation coming from all the scales.

In the above example, the only way to get a ball around u and a ball around v with approximately the same number of points is to get balls of roughly the same radius. No scale could then give a large contribution to the expectation $\mathbb{E}_{A\sim D}[|f_A(u) - f_A(v)|]$; every scale, however, gave a noticeable contribution, and adding them up we had a bounded distortion. The above example will be the template for the full proof, which will do an "amortized analysis" of the contribution to the expectation coming from each scale t, by looking at the radii that define a ball around u and a ball around v with approximately t elements.

7 The Proof of Bourgain's Theorem

Given Fact 3 and Fact 7, proving Bourgain's theorem reduces to proving the following theorem.

Theorem 11 *For a finite set of points V, consider the distribution D over subsets of V sampled by uniformly picking a scale $t \in \{1, \ldots, 2^{\lfloor \log_2 |V| \rfloor}\}$ and then picking independently each $v \in V$ to be in A with probability $1/t$. Let $d : V \times V \to \mathbb{R}$ be a semimetric. Then for every $u, v \in V$,*

$$\mathop{\mathbb{E}}_{A \sim D}[|f_A(u) - f_A(v)|] \geq \frac{1}{c \log_2 |V|} \cdot d(u, v)$$

where c is an absolute constant.

Proof: Fix two vertices u and v

For each t, let ru_t be the distance from u to the t-th closest point to u (counting u), or $d(u, v)/3$, whichever is smaller, and define rv_t similarly. By definition, we have

$$|\{w : d(u, w) < ru_t\}| < t$$

Call t^* the minimum of $|\{w : d(u, w) < d(u, v)/3\}|$ and $|\{w : d(v, w) < d(u, v)/3\}|$. Then, for $t \leq t^*$ we have that both ru_t and rv_t are $< d(u, v)/3$, but for $t \geq t^* + 1$ we have that at least one of ru_t or rv_t (possibly, both) equals $d(u, v)/3$. Note also that for $t \leq t^*$ we have

$$|\{w : d(u, w) \leq ru_t\}| \geq t$$

and similarly for v.

We claim that there is an absolute constant c such that for every scale $t \leq t^*$, we have

$$\mathop{\mathbb{E}}_{A \sim D_t} |f_A(u) - f_A(v)| \geq c \cdot (ru_{2t} + rv_{2t} - ru_t - rv_t) \tag{6}$$

We prove the claim by showing that there are two disjoint events, each happening with probability $\geq c$, such that in one event $|f_A(u) - f_A(v)| \geq ru_{2t} - rv_t$, and in the other event $|f_A(u) - f_A(v)| \geq rv_{t2t} - ru_t$.

1. The first event is that A avoids the set $\{z : d(u,z) < ru_{2t}\}$ and intersects the set $\{z : d(v,z) \leq rv_t\}$. The former set has size $< 2t$, and the latter set has size $\leq t$; the sets are disjoint because we are looking at balls or radius $\leq d(u,v)/3$ around u and v; so the event happens with a probability that is at least an absolute constant. When the event happens,

$$|f_A(u) - f_A(v)| \geq f_A(u) - f_A(v) \geq ru_{2t} - rv_t$$

2. The second event is that A avoids the set $\{z : d(v,z) < rv_{2t}\}$ and intersects the set $\{z : d(u,z) \leq ru_t\}$. The former set has size $< 2t$, and the latter set has size $\leq t$; the sets are disjoint because we are looking at balls or radius $\leq d(u,v)/3$ around u and v; so the event happens with a probability that is at least an absolute constant. When the event happens,

$$|f_A(u) - f_A(v)| \geq f_A(v) - f_A(u) \geq rv_{2t} - ru_t$$

So we have established (6). Summing over all scales up to the largest power of two $t' \leq t^*$, we have

$$\begin{aligned} &\mathop{\mathbb{E}}_{A \sim D} |f_A(u) - f_A(v)| \\ &\geq \frac{c}{1 + \log_2 n} \cdot (ru_{2t'} + rv_{2t'} - ru_1 - rv_1) \\ &\geq \frac{c}{1 + \log_2 n} \cdot \frac{d(u,v)}{3} \end{aligned}$$

□

There is one remaining point to address. In Fact 3, we proved that a distribution over embeddings on the line can be turned into an L1 embeddings, in which the number of dimensions is equal to the size of the support of the distribution. In our proof, we have used a distribution that ranges over $2^{|V|}$ possible functions, so this would give rise to an embedding that uses a superpolynomial number of dimensions.

To fix this remaining problem, we sample $m = O(\log^3 |V|)$ sets $A_1, \ldots, A_m$ and we define the embedding $f(u) := (m^{-1} \cdot f_{A_1}(u), \ldots, m^{-1} \cdot f_{A_m}(u))$. It remains to prove that this randomized mapping has low distortion with high probability, which is an immediate consequence of the Chernoff bounds. Specifically, we use the following form of the Chernoff bound:

Lemma 12 *Let $Z_1, \ldots, Z_m$ be independent nonnegative random variables such that, with probability 1, $0 \leq Z_i \leq M$. Let $Z := \frac{1}{m}(Z_1 + \cdots + Z_m)$. Then*

$$\mathbb{P}[\mathbb{E}\, Z - Z \geq t] \leq e^{-2mt^2/M^2}$$

Let us look at any two vertices u, v. Clearly, for every choice of $A_1, \ldots, A_m$, we have $||f(u) - f(v)||_1 \leq d(u, v)$ so it remains to prove a lower bound to their L1 distance. Let us call Z the random variable denoting their L1 distance, that is

$$Z := ||f(u) - f(v)|| = \sum_{i=1}^{m} \frac{1}{m} |f_{A_i}(u) - f_{A_i}(v)|$$

We can write $Z = \frac{1}{m} \cdot (Z_1 + \cdots + Z_m)$ where $Z_i := |f_{A_i}(u) - f_{A_i}(v)|$, so that Z is the sum of identically distributed nonnegative random variables, such that

$$Z_i \leq d(u, v)$$
$$\mathbb{E}\, Z_i \geq \frac{c}{\log |V|} d(u, v)$$

Applying the Chernoff bound with $M = d(u, v)$ and $t = \frac{c}{2 \log |V|} d(u, v)$, we have

$$\mathbb{P}\left[Z \leq \frac{c}{2 \log |V|} d(u, v) \right]$$
$$\leq \mathbb{P}\left[Z \leq \mathbb{E}\, Z - \frac{c}{2 \log |V|} d(u, v) \right]$$
$$\leq 2^{-2mc^2/(2 \log |V|)^2}$$

which is, say, $\leq 1/|V|^3$ if we choose $m = c' \log^3 |V|$ for an absolute constant c'.

By taking a union bound over all pairs of vertices,

$$\mathbb{P}\left[\forall u, v.\ ||f(u) - f(v)||_1 \geq \frac{c}{2 \log |V|} \cdot d(u, v) \right] \geq 1 - \frac{1}{|V|}$$

8 Tightness of the Analysis of the Leighton-Rao Relaxation

If (X, d) and (X', d') are metric spaces, we say that a mapping $f : X \to X'$ is an *embedding of* (X, d) *into* (X', d) *with distortion at most* c if there are parameters c_1, c_2, with $c = c_1 c_2$ such that, for every $u, v \in X$, we have

$$\frac{1}{c_1} \cdot d'(u, v) \leq d(u, v) \leq c_2 \cdot d'(u, v)$$

The metric space $\mathbb{R}^m$ with distance $||u-v|| = \sqrt{\sum_i (u_i - v_i)^2}$ is denoted by ℓ^2_m, and the metric space $\mathbb{R}^m$ with distance $||u-v||_1 = \sum_i |u_i - v_i|$ is denoted by ℓ^1_m. We just proved the following result.

Theorem 13 (Bourgain) *There is an absolute constant c such that every finite metric space (V,d) embeds into ℓ^1_m with distortion at most $c \log |V|$, where $m = O(\log^3 |V|)$.*

If we solve the Leighton-Rao linear programming relaxation to approximate the sparsest cut of a graph $G = (V,E)$, and we let $d(\cdot,\cdot)$ be an optimal solution, we note that, if we weigh each edge $(u,v) \in E$ by $d(u,v)$, and then compute shortest paths in this weighted graph, then, for every two vertices x,y, the distance $d(x,y)$ is precisely the length of the shortest path from x to y. In particular, if we are using the Leighton-Rao relaxation in order to approximate the sparsest cut in a given *planar* graph, for example, then the solution $d(\cdot,\cdot)$ that we need to round is not an arbitrary metric space, but it is the shortest path metric of a weighted planar graph. It is conjectured that, in this case, the Leighton-Rao relaxation could deliver a constant-factor approximation.

Question 1 *Is there an absolute constant c such that every metric space (X,d) constructed as the shortest-path metric over the vertices of a planar graph can be embedded into ℓ^1_m with distortion at most c, where $m = |V|^{O(1)}$?*

So far, it is known that *k-outerplanar* graphs, for constant k, embed in ℓ^1 with constant distortion.

This is just an example of a large family of questions that can be asked about the embeddability of various types of metric spaces into each other.

For general finite metric spaces, the logarithmic distortion of Bougain's theorem is best possible.

In order to prove the optimality of Bourgain's theorem, we will state without proof the existence of constant degree families of expanders. In a later part of the course we will prove their existence and give efficient constructions.

Theorem 14 (Existence of Expanders) *There are absolute constants d and c such that, for infinitely many n, there is an n-vertex d-regular graph G_n such that $\phi(G_n) \geq c$.*

On such graphs, the Leighton-Rao relaxation is $LR(G_n) \leq O(1/\log n)$, showing that our proof that $LR(G) \geq \phi(G) O(logn)$ is tight.

For every two vertices u,v, define $d(u,v)$ as the length of (that is, the number of edges in) a shortest path from u to v in G_n.

Then

$$\sum_{u,v} A_{u,v} d(u,v) = 2|E|$$

Because each graph G_n is d-regular, it follows that for every vertex v there are $\leq 1 + d + \cdots + d^k < d^{k+1}$ vertices at distance $\leq k$ from v. In particular, at least half of the vertices have distance $\geq t$ from v, where $t = \lfloor \log_d n/2 \rfloor - 1$, which implies that

$$\sum_{u,v} d(u,v) \geq n \cdot \frac{n}{2} \cdot t = \Omega(n^2 \log n)$$

Recall that

$$LR(G) = \min_{d \text{ semimetric}} \frac{|V|^2}{2|E|} \frac{\sum_{u,v} A_{u,v} d(u,v)}{\sum_{u,v} d(u,v)}$$

and so

$$LR(G_n) \leq O\left(\frac{1}{\log n}\right)$$

even though

$$\phi(G_n) \geq \Omega(1)$$

Note that we have also shown that every embedding of the shortest-path metric $d(\cdot,\cdot)$ on G_n into ℓ^1 requires distortion $\Omega(\log n)$, and so we have proved the tightness of Bourgain's theorem.

Exercises

1. Let $G = (V, E_G), H = (V, E_H)$ be an instance of the non-uniform sparsest cut problem. Let $d(u,v)$ a feasible solution to the Leighton-Rao relaxation

$$LR(G,H) := \min_{\substack{d : V \times V \to \mathbb{R} \\ d \text{ semi-metric}}} \frac{|E_H|}{|E_G|} \cdot \frac{\sum_{\{u,v\}} G_{u,v} d(u,v)}{\sum_{\{u,v\}} H_{u,v} d(u,v)}$$

Let $d'(u,v)$ be the length of the shortest path from u to v in the graph that has the edges of G, and each edge $(u,v) \in E_G$ is weighted by $d(u,v)$. Show that $d'(u,v)$ is a feasible solution whose cost is smaller than or equal to the cost $d(u,v)$.

2. Using the above fact, show that if G is a cycle and H is a clique, then the solution in which $d(u,v)$ is the length of the shortest path from u to v in G is an optimal solution.

   [Hint: start from an optimal solution, derive from it another solution of the same cost in which $d(u,v)$ is the same for every u, v that are adjacent in G, then apply the fact proved in the previous exercise.]