Sparsest cut

Sparsest cut problem:
Given graph $G=(V, E)$, $c: E \to \mathcal{R}^+$

Pairs of nodes $s_1t_1, s_2t_2, \ldots, s_kt_k$

Each pair s_it_i has a demand $dem(i) > 0$

For $E' \subseteq E$, let $c(E') = \sum_{e \in E'} c(e)$

let $dem(E') = \sum_{i,\ s_it_i \text{ separated by } E'} dem(i)$ where s_it_i is separated by E' if they are not connected in $G[E \setminus E']$

let $sparsity(E') = c(E')/dem(E')$

Goal: find a cut E' of *minimum* sparsity (sparsest cut)

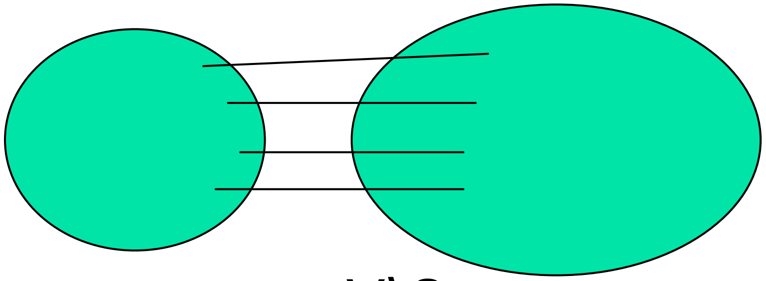
(problem is NP-hard)

Sparsest cut

Sparsest cut has many applications that we will discuss later on

Observation: if G is connected there always exists a sparsest cut E' where G[E\E'] consists of two connected components S, V\S

Proof: exercise



S

V\S

For this reason sometimes sparsest cut is defined as
find $S \subset V$ to minimize $c(\delta(S))/dem(\delta(S))$

Uniform vs Non-uniform

A special case of the sparsest cut problem is the following:

$k = n(n-1)/2$ and every pair of vertices uv is a commodity with $dem(uv) = 1$

Most interesting applications of sparsest cut are for this special case. Sometimes this is called the *uniform* case of the sparsest cut problem. The *non-uniform* case refers to the general problem.

Maximum concurrent flow

For most cut problems there is usually a flow problem that is dual to it.

For sparsest cut, it is the *maximum concurrent flow* problem: given

$G=(V,E)$, $c: E \to \mathcal{R}^+$ (now interpret c as edge capacities)

Pairs of nodes $s_1t_1, s_2t_2, \ldots, s_kt_k$

Each pair s_it_i has a demand $dem(i) > 0$

Goal: maximize λ s.t each pair s_it_i can *concurrently* send flow of $\lambda\ dem(i)$

Maximum concurrent flow

Let λ^* be the optimum value for the maximum concurrent flow problem

First we observe that for each $S \subset V$

$\lambda^* \leq c(\delta(S))/dem(\delta(S))$

because the demand crossing S (λ^* dem(δ(S)) cannot exceed the capacity of the cut $\delta(S)$

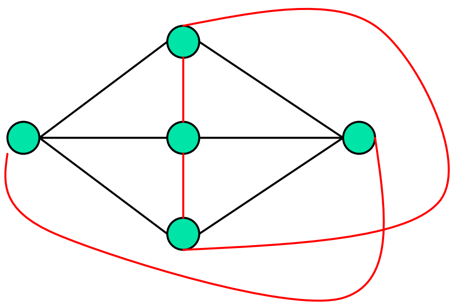
Therefore $\lambda^* \leq \min_S c(\delta(S))/dem(\delta(S))$ and hence λ^* is a lower bound on the *minimum sparsity*

Note that for $k=1$, (single pair) λ^* = min sparsity from maxflow-mincut theorem (do you see why?)

Flow-cut gap

For k = 2, λ^* = min sparsity (this Hu's two-commodity flow theorem)

However for k > 3 we can have λ^* < min sparsity. Here is an example



Graph is in black edges. Red edges are the demand pairs. Capacities/demands are all 1

Natural question is whether $\lambda^* \geq \alpha$ (min sparsity) for some $\alpha < 1$. Would also allow us to get an $1/\alpha$ approximation for min sparsity since λ^* can be computed via an LP

LP for λ^*

We can write a straight forward LP for computing λ^*

We use exponential # of variables but a compact formulation can easily be derived

P_i: set of paths from s_i to $t_{i,}$ $P = \cup_i P_i$

$f(p)$ variable for flow on path p

$\max \lambda$

s.t

$$\sum_{p \in P_i} f(p) \geq \lambda \,\text{dem}(i) \quad 1 \leq i \leq k$$

$$\sum_{p:\, e \in p} f(p) \leq c(e) \quad e \in E$$

$$f(p) \geq 0$$

Dual of LP

The dual of the LP can be seen as a meaningful relaxation for sparsest cut

Variables for dual:

d_e for each $e \in E$ (interpret as distance/length of e)

d_i for $1 \leq i \leq k$ (interpret as distance from s_i to t_i)

$\min \sum_{e \in E} c(e)\, d(e)$

s.t

$\sum_i \text{dem}(i)\, d_i \geq 1$

$\sum_{e \in p} d_e \geq d_i$ for all $p \in P_i$

$d_e \geq 0$

$d_i \geq 0$

Interpretation of the dual

The dual assigns distances to edges which induce shortest path distances on all vertices

The dual is nothing but the following (why?)

$\min_{d \text{ is a semi-metric}} \sum_{uv \in E} c(uv)\, d(uv) \,/\, \sum_{i=1}^{k} dem(i)\, d(s_i t_i)$

where $d(uv)$ is the distance between u and v

Since we cannot use ratios in LPs the denominator is normalized to a constraint which says

$\sum_{\iota=1}^{k} dem(i)\, d(s_i t_i) \geq 1$

and the numerator is minimized. Note that scaling does not affect the ratio

Interpretation of the dual

We can interpret the dual directly as a relaxation of the sparsest cut problem.

Note that each cut $E' \subseteq E$ induces a semi-metric $d_{E'}$ on the vertices where $d_{E'}(uv) = 1$ if u,v are separated by E' and $d_{E'}(uv) = 0$ otherwise

Thus the sparsest cut problem is asking precisely for the following:

$\min_{E' \subseteq E} \sum_{uv \in E} c(uv)\, d_{E'}(uv) / \sum_{i=1}^{k} dem(i)\, d_{E'}(s_i t_i)$

We cannot solve above so instead of minimizing over cut-metrics we minimize over all metrics which turns out to be a linear program and hence solvable

Rounding the dual

We give two ways to round the dual.

The first uses a relatively simple reduction to the multicut problem but illustrates the relationship between the two cut problems and a general technique. The ratio one obtains is not optimal.

The second uses a sophisticated connection to embedding metric spaces into real normed spaces and how that leads to an optimum ratio

Rounding via multicut relationship

Recall the minimum multicut problem.

We are given graph G and pairs $s_1t_1, s_2t_2, \ldots, s_kt_k$ but the pairs had no demands

The goal was to separate *all* pairs at minimum cost

In sparsest cut we want to separate only a subset of the pairs but the measure is the cost of cut to demand that is separated. If somehow we knew which pairs to separate, then we could use the multicut algorithm to separate those pairs!

We will see that we can use the LP solution to guide us in this process.

Rounding via multicut relationship

Recall the LP for sparsest cut

$\min \sum_{e \in E} c(e)\, d(e)$

s.t

$\sum_i \text{dem}(i)\, d_i \geq 1$

$\sum_{e \in p} d_e \geq d_i$ for all $p \in P_i$

$d_e \geq 0$

$d_i \geq 0$

Let $d\text{max} = \max_{i=1}^{k} d_i$

Rounding via multicut relationship

Let $dmax = \max_{i=1}^{k} d_i$

For $l \geq 0$, let $A_l = \{ i \mid dmax/2^{l+1} < d_i \leq dmax/2^l \}$

let $D = \sum_{i=1}^{k} dem(i)$ where $dem(i)$ are integers
let $dem(A_l) = \sum_{i \in A_l} dem(i)$

Lemma: There exists h such that
$dem(A_h)\, dmax/2^{h+1} \geq 1/(8 \log D)$

Note that $\sum_{i=1}^{k} dem(i)\, d_i \geq 1$

We derive the lemma from this.

Rounding via multicut relationship

Note that $A_0, A_1, \ldots,$ are disjoint

Therefore

$\sum_{i=1}^{k} \text{dem}(i)\, d_i = \sum_{l \geq 0} \sum_{i \in A_l} \text{dem}(i)\, d\_i \geq 1$

$\sum_{i \in A_l} \text{dem}(i)\, d\_i \leq (1/2) \sum_{i \in A_l} \text{dmax}/2^{l+1}$

since $i \in A_l$ implies $d_i \in (\text{dmax}/2^{l+1}, \text{dmax}/2^{l}]$

therefore $\sum_{l \geq 0} \text{dem}(A_l)\, \text{dmax}/2^{l+1} \geq 1/2$

let $t = 2 \log D-1$

$\sum_{l \geq 0} \text{dem}(A_l)\, \text{dmax}/2^{l+1}$

$= \sum_{l \leq t} \text{dem}(A_l)\, \text{dmax}/2^{l+1} + \sum_{l > t} \text{dem}(A_l)\, \text{dmax}/2^{l+1}$

Rounding via multicut relationship

$\sum_{l > t}$ dem(A_l) dmax/2^{l+1} \leq \sum_{l > t} dem(A_l)/D^2
since $2^t \geq D^2$ and dmax ≤ 1

therefore
$\sum_{l > t}$ dem(A_l) dmax/2^{l+1} $\leq$ ($\sum_{l > t}$ dem(A_l))/D^2 $\leq$ 1/D since
$\sum_{l > t}$ dem(A_l) $\leq$ D

We can assume wlog that $D \geq 4$ for otherwise we can get a simple D approximation
therefore $\sum_{l \leq t}$ dem(A_l) dmax/2^{l+1} $\geq$ 1/2 - 1/D $\geq$ 1/4

since the lhs is a sum of 2log D terms, one of them must be at least 1/(8log D) which proves the lemma

Rounding via multicut relationship

Lemma: There exists h such that $\sum_{i \in A_h} d_i \geq 1/(8 \log D)$

We solve a multicut problem for the set A_h,
that is we separate all pairs $s_i t_i$ with $i \in A_h$

How do we argue that this is would lead to a good solution?

Let us write down the LP for multicut problem on A_h

$\min \sum_e c(e)\, l(e)$

s.t

$\sum_{e \in p} l(e) \geq 1$ for all $i \in P_i$, $i \in A_h$

$l(e) \geq 0$

Rounding via multicut relationship

Let us write down the LP for multicut problem on A_h

$\min \sum_e c(e)\, l(e)$

s.t

$\sum_{e \in p} l(e) \geq 1$ for all $i \in P_i$, $i \in A_h$

$l(e) \geq 0$

Recall that we showed that if l is a feasible solution to above LP then we can find a cut that separates all pairs in A_h with cost $O(\log k) \sum_e c(e)\, l(e)$

Rounding via multicut relationship

We obtain a feasible solution for the LP using the values from the sparsest cut LP.

Let $\alpha = 2^{h+1}/dmax$

Set $l'(e) = \alpha\, d(e)$

We claim that l' is feasible for the multicut LP on A_h (recall that $d_i \geq dmax/2^{h+1}$ for $i \in A_h$)

Note that for i in A_h and $p \in P_i$

$\sum_{e \in p} d(e) \geq d_i \geq dmax/2^{h+1}$

therefore
$\sum_{e \in p} l'(e) \geq 1$ for $p \in P_i$

Rounding via multicut relationship

Therefore, we can find a multicut $E' \subseteq E$ of cost

$O(\log k) \sum_e c(e)\, l'(e)$ that separates all pairs in A_h

What is sparsity of E'?

$\text{sparisty}(E') \leq c(E')/\sum_{i \in A_h} \text{dem}(i)$

$\leq O(\log k) \sum_e c(e)\, l'(e) \;\; / \text{dem}(A_h)$

$\leq O(\log k)\, \alpha \sum_e c(e)\, d(e) \;/ \text{dem}(A_h)$

$\leq O(\log k) \sum_e c(e)\, d(e) / \; (\text{dem}(A_h) \;/\alpha)$

By lemma, $\text{dem}(A_h)/\alpha = \text{dem}(A_h)\text{dmax}/2^{h+1} \geq 1/(8\log D)$

hence $\text{sparsity}(E') \leq O(\log k \log D) \sum_e c(e)d(e)$

Rounding via multicut relationship

hence sparsity(E') $\leq$ O(log k log D) $\sum_e c(e)d(e)$

Note that $OPT_{LP} = \sum_e c(e)d(e) \leq$ min sparsity

therefore
sparsity(E') $\leq$ O(log k log D) (min sparsity)

Thus we obtain an O(log k log D) approximation.

The dependence of the ratio on D is in general undesirable and in fact a sophisticated argument can be used to reduce the ratio to $O(\log^2 k)$

Rounding via l_1 embeddings

We now present a sophisticated rounding method that yields an O(log k) approximation via metric embeddings

Metric embeddings are a powerful tool in a variety of settings and they got their impetus in computer science with the application to sparsest cut

Metric embeddings

In metric embeddings we study when one metric space can be embedded (mapped) into another metric space such that distances of the points are *distorted* as little as possible.

Formally let (V, d) and (V', d') be two metric spaces.
An embedding of (V, d) into (V', d') is a 1-1 map $f: V \to V'$

f is an *expansion* if forall $u,v \in V$, $d'(f(u),f(v)) \geq d(uv)$

f is a *contraction* if forall $u,v \in V$, $d'(f(u),f(v)) \leq d(uv)$

Metric embeddings

The *distortion* of f , dist(f) is defined to be

$\max_{u,v \in V} \max \{ d'(f(u),f(v))/d(u,v), \; d(u,v)/d'(f(u),f(v)) \}$

The above is a bit messy because f in general need not be an expansion or a contraction

If f is an expansion then

$dist(f) = \max_{u,v \in V} d'(f(u),f(v))/d(u,v)$

If f is a contraction then

$dist(f) = \max_{u,v \in V} d(u,v)/d'(f(u),f(v))$

Note that $dist(f) \geq 1$

If $dist(f) = 1$ then f is called an *isometric embedding*

Embeddings into normed spaces

Of particular interest to us are embeddings of finite metric spaces (generated by graphs) into normed Euclidean spaces, R^h (for some dimension h) equipped with some l_p norm, $p \geq 1$

For two points $x, y \in R^h$, the distance defined by

$d(x,y) = |x-y|_p = (\sum_{i=1}^{h} |x_i - y_i|^p)^{1/p}$ is a metric for $p \geq 1$

In particular the norms l_1, l_2 are of much interest in applications

l_1 embeddings

We focus on l_1 embeddings of finite metrics for their application to sparsest cut. That is, we wish to embed a finite metric (V, d) into R^h for some h to minimize distortion.

We will prove Bourgain's theorem

Theorem (Bourgain): A finite metric on n points can be embedded into $R^{O(\log^2 n)}$ with distortion $O(\log n)$

and apply the theorem to get an $O(\log k)$ approximation for sparsest cut

Cut-metrics and l_1 embeddings

The connection between sparsest cut and l_1 embeddings is seen from the characterization of l_1 embeddings

Given a set V and a set $S \subseteq V$, the cut-semi-metric d_S on V induced by S is given by

$d_S(u,v) = 1$ if $|S \text{ Å } \{u,v\}| = 1$

$d_S(u,v) = 0$ otherwise

Note that d_S is an l_1 metric in R^1. The embedding is given by $f(u) = 0$ if $u \in S$ and $f(u) = 1$ if $u \notin S$

Cut-metrics and l_1 embeddings

Theorem: A metric (V, d) is isometrically embeddable in l_1 (dimension can be arbitrary) iff there exists $\lambda:2^V \to \mathcal{R}^+$ such that $d(uv) = \sum_S \lambda(S)\, d_S(uv)$ for all $u,v \in V$

Proof:

if $d(uv) = \sum_S \lambda(S)\, d_S(uv)$ then we can embed d into l_1 in R^h where h is the number of S with $\lambda(S) > 0$ as follows:

Let $S_1, S_2, \ldots, S_h$ be the sets. Then the embedding is given by

$f(u) = (\lambda(S_1) I_{S_1}(u), \lambda(S_1) I_{S_1}(u), \ldots, \lambda(S_h) I_{S_h}(u))$
where $I_S(u) = 0$ if $u \in S$ and $I_S(u) = 1$ if $u \notin S$

Cut-metrics and l_1 embeddings

only if:

suppose f is a mapping of V into R^h such that

$d(u,v) = |f(u) - f(v)|_1$ for each u,v

Let u(i) be the i'th coordinate of f(u)

Then $|f(u) - f(v)|_1 = \sum_i |u(i) - v(i)|$

Define metrics $d_1, d_2, \ldots, d_h$ on V where

$d_i(u,v) = |u(i) - v(i)|$

To prove that $d = \sum_S \lambda(S)\, d_S$ it is sufficient to prove that each $d_i = \sum_S \lambda_i(S)\, d_S$

Cut-metrics and l_1 embeddings

consider d_i

Let $V = v_1, v_2, \ldots, v_n$

Wlog assume that $v_1(i) \leq v_2(i) \leq \ldots \leq v_n(i)$

For $j = 1$ to $n-1$, let $S_j = \{v_1, v_2, \ldots, v_j\}$

let $\lambda_i(S_j) = v_{j+1}(i) - v_j(i)$

and let $\lambda_i(S) = 0$ if S is not one of $S_1, \ldots, S_{n-1}$

It is easy to check that $d_i = \sum_S \lambda_i(S)\, d_S$

Cut-metrics and l_1 embeddings

Exercise: Prove that any tree metric is l_1 embeddable

Exercise: Prove that any ring metric is l_1 embeddable

l_1 embeddings and sparsest cut

Suppose every n point metric is embeddable into l_1 with distortion $\alpha(n)$

Then we will show that the integrality gap of the LP relaxation we studied is at most $\alpha(n)$

This is based on the characterization of l_1 metrics as those expressible as positive sum of cut-metrics

Note however that it does not immediately give a polynomial time algorithm. We will later use the specific embeddings of Bourgain to derive a randomized polynomial time algorithm

l_1 embeddings and sparsest cut

Recall the LP relaxation was equivalent to

$\min_{d \text{ semi-metric}} \ \sum_{uv} c(uv)\, d(uv) / \sum_i dem(i)\, d(s_i t_i)$

Let d^* be an optimum solution to above relaxation

By definition d^* is embeddable into l_1 with distortion $\alpha(n)$

Since l_1 embeddings are positive sums of semi-metrics it implies that there is a $\lambda: 2^V \rightarrow \mathcal{R}^+$ s.t forall $u,v \in V$

$\sum_S \lambda(S)\, d_S(uv) \leq d^*(u,v) \leq \alpha(n) \sum_S \lambda(S)\, d_S(uv)$

we assume wlog that the embedding is a constraction

l_1 embeddings and sparsest cut

Now we claim that there is a cut of sparsity at most $\alpha(n)$ OPT_{LP}

Note that

$OPT_{LP} = \sum_{uv} c(uv)\, d^*(uv) / \sum_i dem(i)\, d^*(s_i t_i)$

Let $A = \sum_{uv} c(uv)\, d^*(uv)$ and $B = \sum_i dem(i)\, d^*(s_i t_i)$

$A \geq \sum_{uv} c(uv) \sum_S \lambda(S)\, d_S(uv)$

$\geq \sum_S \lambda(S) \sum_{uv \in \delta(S)} c(uv) \geq \sum_S \lambda(S)\, c(\delta(S))$

where we interchanged the order of summation and used the fact that $d_S(uv) = 1$ if $uv \in \delta(S)$ and 0 otherwise

l_1 embeddings and sparsest cut

$B = \sum_i \text{dem}(i)\, d^*(s_i t_i)$
$\leq \sum_i \text{dem}(i)\, \alpha(n) \sum_S \lambda(S)\, d_S(s_i t_i)$

(interchanging order of summation)
$\leq \alpha(n) \sum_S \lambda(S) \sum_{s_i t_i \in \delta(S)} \text{dem}(i)$
$\leq \alpha(n) \sum_S \lambda(S)\, \text{dem}(\delta(S))$

Therefore
$OPT_{LP} = A/B \geq \sum_S \lambda(S)\, c(\delta(S)) / \sum_S \alpha(n)\, \lambda(S)\, \text{dem}(\delta(S))$
or $\lambda(S)\, c(\delta(S)) / \sum_S \lambda(S)\, \text{dem}(\delta(S)) \leq \alpha(n)\, OPT_{LP}$

l_1 embeddings and sparsest cut

Therefore

$OPT_{LP} = A/B \geq \sum_S \lambda(S)\, c(\delta(S)) / \sum_S \alpha(n)\, \lambda(S)\, dem(\delta(S))$

or $\lambda(S)\, c(\delta(S)) / \sum_S \lambda(S)\, dem(\delta(S)) \leq \alpha(n)\, OPT_{LP}$

Since $\lambda(S) \geq 0$ for all S, it follows that there exists a set S^* such that

$c(\delta(S^*))/dem(\delta(S^*)) \leq \alpha(n)\, OPT_{LP}$

This proves the existence of a set of sparsity at most $\alpha(n)$ times OPT_{LP}

This also shows that the flow-cut gap is at most $\alpha(n)$