CS880: Approximation and Online Algorithms	**Scribe:** Taylor Kemp
Lecture 10: Metric Embeddings, Multi-Cut	**Date:** Oct. 9, 2019

10.1 Overview

- Cut Problems and their connections to metrics
- Minimum S-T Cut
- Multi-way Cut
- Multi Cut

10.2 Minimum s-t Cut

10.2.1 Problem Statement

In the minimum s-t cut problem we are given a graph $G = (V, E)$ with edge costs c_e as well as a source node $s \in V$, and a sink node $t \in V$. The goal is then to find the minimum cost set of edges that forms a cut $C \subset G$ such that s and t lie on separate sides of the cut. e.g.

$$\min_C \sum_{e \in C} c_e \tag{10.2.1}$$

10.2.2 LP Formulations

let us define x_e to be the indicator variable that edge $e \in E$ lies in our cut C. Let us also define P to be some choice of path in G from s to t. We can summarize the cut constraint as each path from s to t must contain at least one edge cut. From this we can then solve for the minimum cut using the following LP

$$min \sum_e x_e c_e \tag{10.2.2}$$

$$s.t. \sum_{e \in P} x_e \geq 1 \forall P \tag{10.2.3}$$

$$x_e \geq 0 \forall e \in E \tag{10.2.4}$$

Alternatively, if we consider edge lengths x_e, and denote $d_x(u, v)$ as the distance from u to v on these lengths x_e, we can obtain the following equivalent formulation as the distance $d_x(s, t)$ is just the sum of edge lengths x_e of the shortest path from s to t.

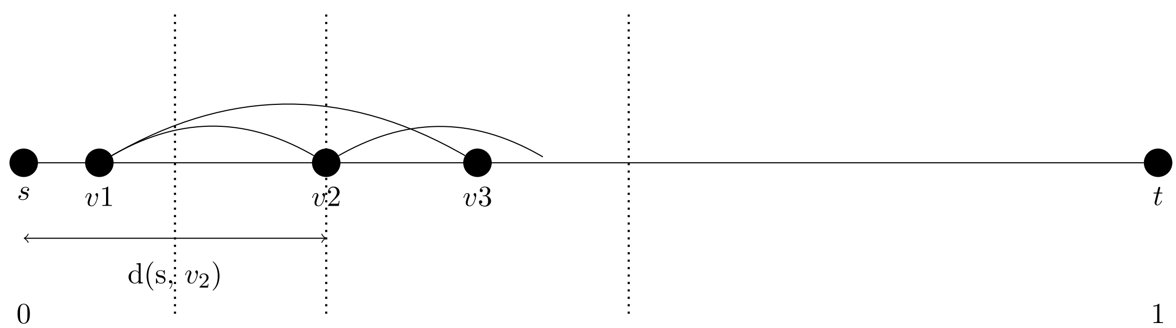
$$min \sum_e d_x(e)c_e$$
$$s.t.\ d_x(s,t) \geq 1$$
$$x_e \geq 0 \forall e \in E$$

Definition 10.2.1 *We say a function $d : S \to \mathbb{R}$ on some set S is a metric if it satisfies the following:*

- $d(x, y) > 0 \forall x \neq y$
- $d(x, x) = 0$
- $d(x, y) \leq d(x, z) + d(z, y)$

From this definition and the previous formulation of the LP. We can reformulate the LP once again as the following.

$$min \sum_e d_e c_e$$
$$s.t.\ d_x(s,t) \geq 1$$
$$\text{d is a metric}$$



From 10.4.1 we can observe that choosing a cut is equivalent to choosing some radius r, and looking at the set of edges that crosses the surface of the ball

$$B_d(s, r) = \{v : d(s, b) \leq r\} \tag{10.2.5}$$

Let us consider picking a random $r \sim U[0, 1]$.

Claim 10.2.2

$$\exp c(\delta(B_d(s, r))) \leq \sum_e c_e x_e$$

Proof:

$$\begin{aligned} & Pr[e \in \delta(B_d(s,r))] \\ = & Pr[d(s,u) \le r < d(s,v)] \\ = & d(s,v) - d(s,u) \le x_e \end{aligned}$$

From this we can conclude the following

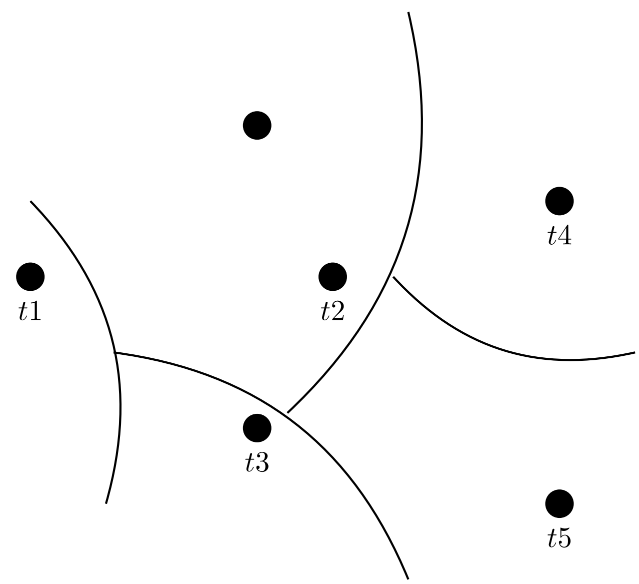
$$\begin{aligned} \exp c(\delta(B_d(s,r))) = & \sum_e c_e Pr[e \in \delta(B_d(s,r))] \\ & \le \sum_e c_e x_e \end{aligned}$$

This also implies that the LP has an integral solution. ∎

10.3 Multiway Cut

10.3.1 Problem Statement

Similar to the minimum s-t cut problem, we are given a graph $G = (V, E)$ with edge costs c_e. However, now we are given a set of k terminals $t_1, t_2, \cdots t_k$. The goal is then to find the minimum cost cut F, s.t. $\forall i, j \in [k]$, there is no path between t_i, t_j in the graph $(V, E \backslash F)$.



It is worth mentioning that we can obtain a $2-$approximation for this problem by taking the min cut to isolate each t_i. This is because if we look at the optimal solution to the multiway cut problem, each edge in the solution defines a boundary for at most two terminals, t_i, t_j. If we take the boundary formed around t_i in the multiway cut solution, the cost of this boundary must be at least the cost of the min cut to isolate t_i. Since this edge appears at most twice over all such boundaries, we have that the cost of taking the min cut to isolate each t_i is no greater than twice the optimal solution to multiway cut.

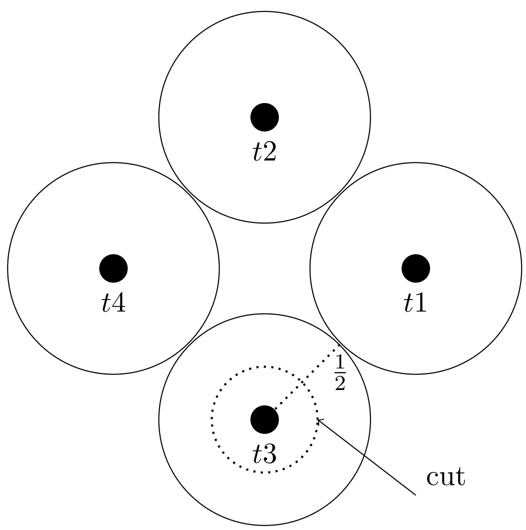
10.3.2 LP Formulation

We can formulate this similarly to the previous example, except now, we consider all distances $d_x(t_i, t_j), \forall i, j \in [k]$ instead of just the distance $d_x(s, t)$.

$$min \sum_e c_e d_e$$
$$s.t. \text{ d is a metric}$$
$$d_x(t_i, t_j) \geq 1 \ \forall i, j \in [k], i \neq j$$

Algorithm 1 Randomized Rounding Algorithm

for i$\in$ [k] **do**
  pick r$\in$ U[0, $\frac{1}{2}$]
  let $F_i = \delta(B(t_i, r))$
end for
return $\cup_i F_i$



Let us define the volume of a ball centered at terminal t with radius r as follows

$$Vol(B(t,r)) = \sum_{(u,v):u\in B(t,r), v\in B(t,r), (u,v)\in E} c_e d_e + \sum_{(u,v):u\in B(t,r), v\notin B(t,r), (u,v)\in E} c_e d_e \frac{r - d(t,u)}{d(t,v) - d(t,u)}$$

This is just the volume of all edges contained completely within the ball as well as the volume of the edges with $u \in B(t, r)$ and $v \notin B(t, r)$ that lies within the ball.

Claim 10.3.1 $\sum_i Vol(t_i, \frac{1}{2}) \leq \sum_e c_e d_e$

Proof: This follows from the fact that $\sum_i \text{Vol}(t_i, \frac{1}{2}) \leq Vol(E) = \sum_e c_e d_e$ ■

Claim 10.3.2 $\mathbf{E}[c(F_i)] \leq 2Vol(t_i, \frac{1}{2}) \forall i$

Proof:

$$Pr[(u,v) \in F_i] = Pr[d(t_i,u) \leq r < d(t_i,v)] \tag{10.3.6}$$

$$= \frac{d(t_i,v) - d(t_i,u)}{\frac{1}{2}} \leq 2d(u,v) \tag{10.3.7}$$

$$\leq 2d_e + 2d_e \frac{\frac{1}{2} - d(t_i,u)}{d(t_i,v)d(t_i,u)} \tag{10.3.8}$$

Taking the expectation we obtain 10.3.2. Putting everything together now, we can write the following in order to show the randomized rounding algorithm achieves a 2-approximation.

$$\sum_i \mathbf{E}[c(F_i)] \leq 2 \sum_i Vol(t_i, \frac{1}{2}) \leq 2 \sum_e c_e d_e$$

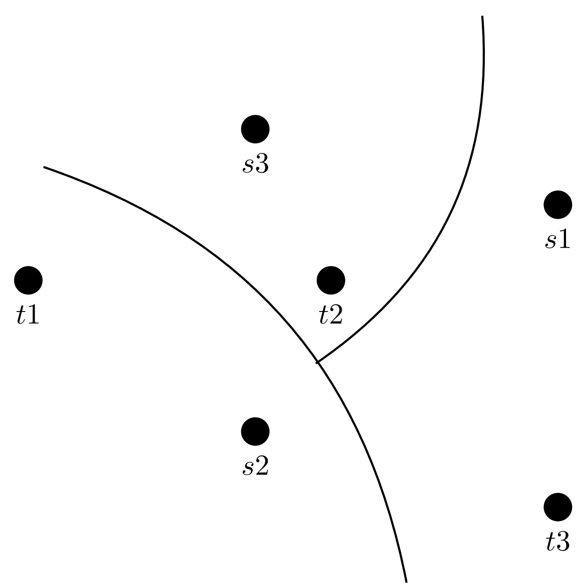
■

It is worth mentioning that by constructing the k-1 largest of these cuts in order we are able to tighten the analysis to a $2(1 - \frac{1}{k})-$approximation.

10.4 Multi Cut

10.4.1 Problem Statement

Let us consider a new problem on the graph $G = (V, E)$ with edge weights c_e, and k pairs (s_i, t_i). Our goal now is to find some F s.t. $(V, E \backslash F)$ contains no $s_i \rightarrow t_i$ paths $\forall i \in [k]$



10.4.2 LP Formulation

Similar to the previous example, we can construct the following LP

$$\min \sum_e c_e d_e$$
$$\text{d is a metric}$$
$$d(s_i, t_i) \geq 1 \forall i \in [k]$$

Let us also consider adding a point mass at each terminal of $\frac{V}{k}$ where V is the volume of the whole graph. From this we would then compute the volume as

$$Vol(B(t,r)) = \frac{V}{k} + \sum_{(u,v):u \in B(t,r), v \in B(t,r), (u,v) \in E} c_e d_e + \sum_{(u,v):u \in B(t,r), v \notin B(t,r), (u,v) \in E} c_e d_e \frac{r - d(t,u)}{d(t,v) - d(t,u)}$$

Claim 10.4.1 $\forall i \in [k],\ \exists r \in [0, \frac{1}{2})\ s.t.\ c(\delta(s_i, r)) \leq \alpha\, Vol(s_i, r)$

Let us also consider the following algorithm

---

for i$\in$ [k] **do**
  pick r_i satisfying claim 10.4.1 in $(V, E \backslash \cup_{j<i} F_j)$
  let $F_i = \delta(B(t_i, r))$
end for
return $\cup_i F_i$

Proof: We will prove this by contradiction. Let us suppose $\forall r \in [0, \frac{1}{2}),\ c(\delta(s_i, r)) > \alpha\ \text{Vol}(s_i, r)$ for some i. We will first consider the rate of change of volume of the sphere centered at terminal t.

$$\frac{d}{dr}\text{Vol}_d(t,r) = \sum_{(u,v) \in \delta(t,r)} c_e \frac{d_e}{d(t,v) - d(t,u)} \geq c(\delta(t,r))$$

From this we can conclude the following given our assumption.

$$\frac{d}{dr} Vol(s_i, r) \geq c(\delta(t,r)) > \alpha Vol(s_i, r)$$

$$\implies \int_0^{\frac{1}{2}} \frac{dVol(s_i, r)}{Vol(s_i, r)} > \int_0^{\frac{1}{2}} \alpha dr \tag{10.4.9}$$

$$\implies log(\frac{2V}{\frac{V}{k}}) > log(\frac{Vol(s_i, \frac{1}{2})}{Vol(s_i, 0)}) > \alpha \frac{1}{2} \tag{10.4.10}$$

$$\implies \alpha < 2log(2k) \tag{10.4.11}$$

Hence, we have that as k is just the number of terminal pairs we have, that we can choose α large s.t. $\alpha \geq 2log(2k)$ and our assumption is false. One corollary of this claim is that the provided algorithm returns $\cup_i F_i$ which is an α-approximation ■