**CS880: Approximations Algorithms**

**Scribe:** Dave Andrzejewski **Lecturer:** Shuchi Chawla
**Topic:** Metric multiway cut and multicut, integrality gap **Date:** 3/15/07

The lecture further explores the use of cut metrics, with applications to the multiway cut and multicut problems. Also, the idea of an expander graph is introduced and applied to deriving the integrality gap between an optimal LP solution and the optimal corresponding integral solution.

# 16.1 Metric multiway cut

## 16.1.1 Problem setup and LP

**GIVEN:** a graph $G = (V, E)$ and a set $T \subset V$ of terminals.
**DO:** find the minimum weight cut separating every pair of terminals $t_i, t_j \in T$ from one another.

Just as in the previous lecture, we will formulate this cut problem as a metric LP. We enforce the separation of all pairs of terminals by requiring that our metric assign them distance $\geq 1$.

$$\min \sum_{(u,v)\in E} c_{uv} d(u,v) \qquad \text{obj fcn}$$

$$d \text{ is a metric}$$

$$d(t_i, t_j) \geq 1 \qquad \forall t_i, t_j \in T$$

where $c_{uv}$ is the cost of the edge between vertices $u$ and $v$.

## 16.1.2 Rounding

Once we've found an optimal solution to the metric LP above, we need to transform our metric to a cut metric, which will yield a valid multiway cut. As usual, we use the fact that $LP^* \leq OPT$.

To do this analysis, it will be useful to consider a set of interesting physical analogies for the quantities in our problem.

- edges $\rightarrow$ pipes
- $c_e \rightarrow$ cross-sectional area of pipe $e$
- $d_e \rightarrow =$ length of pipe $e$
- $B(t_i, r) \rightarrow =$ ball of radius $r$, centered at $t_i$
- $f(t_i, r, e) \rightarrow =$ fraction of edge $e$ covered by ball $B(t_i, r)$
- $Vol(B(t_i, r)) \rightarrow =$ total pipe volume enclosed by ball $B(t_i, r)$

- $Area'(B(t_i, r)) \rightarrow =$ the total pipe cross section area on the surface of ball $B(t_i, r)$
- $Area(B(t_i, r)) \rightarrow =$ the total cost ($c_e$) of the edges crossing the ball $B(t_i, r)$

Note that $Area \neq Area'$ in general, because the surface of the ball may not be perpendicular to the pipe (Figure 16.1.1). In fact, $Area' \geq Area$. The expressions for $f$,$Vol$ and $Area$ are as below. For $f$, edge $e = (u, v)$, and $f$ is simply 1 or 0 if both or neither of $(u, v)$ are contained in the ball, with the expression below covering the more interesting case where $u \in B$ and $v \notin B$.

$$f(t_i, r, e) = \frac{r - d(t_i, u)}{d(t_i, v) - d(t_i, u)} \quad (16.1.1)$$

$$Vol(B(t_i, r)) = \sum_{e \in E} f(t_i, r, e) d_e c_e \quad (16.1.2)$$

$$E' = \{(u, v) \text{ s.t. } |B \cap u, v| = 1\} \quad (16.1.3)$$

$$Area(B(t_i, r)) = \sum_{(u,v) \in E'} c_{uv} \quad (16.1.4)$$

$$Area'(B(t_i, r)) = \frac{d}{dr} Vol(B(t_i, r)) = \sum_{(u,v) \in E'} c_{uv} \frac{d(u, v)}{d(t_i, v) - d(t_i, u)} \quad (16.1.5)$$

These quantities will provide an intuitive framework with which to analyze our rounding scheme.

**Algorithm**

- $\forall i$, pick $r_i = \arg\min_{r \in [0,1/2]} Area(B(t_i, r))$
- let $C_i$ be the cut associated with that radius
- pick the $k - 1$ minimum weight cuts from $C_1, C_2, ..., C_k$

Since each pair of terminals $(t_i, t_j)$ must satisfy $d(t_i, t_j) \geq 1$ in the original LP solution, making our cuts with at a radius $r \leq 1/2$ around each terminal will clearly yield a valid set of separating cuts.

We now derive the approximation factor of our resulting scheme.

**Lemma 16.1.1** $Area(t_i, r_i) \leq 2Vol(t_i, 1/2)$ $\forall i$

**Proof:** As the radius of a ball increases, the volume enclosed by the ball will grow proportionally to the surface areas of all pipes currently cut by the surface of the ball.

If all cut pipes were perpendicular to the surface of the expanding ball the volume growth would be exactly equal to the current surface area, but since this is not neccessarily the case (see 16.1.1) $Area$ lower bounds the rate of volume growth ($Area'$).

$$Area(t_i, r) \leq \frac{d}{dr} Vol(t_i, r)$$

Recall that we have chosen $r_i$ to minimize $Area(t_i, r)$ on the interval $r \in [0, 1/2]$. We use this fact to substitute the constant term $Area(t_i, r_i)$ in for $Area(t_i, r)$ above, and then integrate both sides with respect to $r$ with bounds from $r = 0$ to $r = 1/2$. This yields our final result.

$$1/2 Area(t_i, r_i) \leq Vol(t_i, 1/2)$$

■

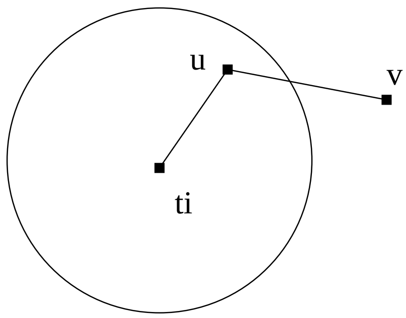


Figure 16.1.1: An edge cut by a ball may not be perpendicular to the ball surface.

**Lemma 16.1.2** $\sum_i Vol(t_i, 1/2) \leq LP^*$

**Proof:** Using the fact that $d(t_i, t_j) \geq 1$ for all terminal pairs, it is clear that all $B(t_i, 1/2)$ will be disjoint. Since the cost of $LP^*$ is equal to the volume of the entire graph, and the set of $B(t_i, 1/2)$ form a disjoint subset of the graph, it must be true that the total ball volume is $\leq LP^*$. ■

The cut produced by our algorithm chooses all edges which are partially cut by the surface of each ball $B(t_i, r_i)$. Thus the total cost is equal to $\sum_i^{k-1} Area(t_i, r_i)$. Recall that we simply omit the most expensive cut, since that terminal is already isolated by the other $k - 1$ cuts. Finally we combine these lemmas to get an approximation factor of $2(1 - 1/k)$.

$$\sum_{i=1}^{k-1} Area(t_i, r_i) \leq 2 \sum_{i=1}^{k-1} Vol(t_i, 1/2) \leq 2(1 - 1/k)LP^* \leq 2(1 - 1/k)OPT$$

## 16.2 Metric multicut

### 16.2.1 Problem setup and LP

Multicut is very similar to multiway cut, except we now have $k$ pairs of terminals $(s_i, t_i)$ and our cut only needs to separate each $s_i$ from its correpsonding $t_i$ for all $i$. The cut does not need to separate different pairs from each other.

The metric LP formulation is identical, except our separation constraint is now $d(s_i, t_i) \geq 1 \ \forall i$.

(As an aside, it is intresting to note that this LP could be reformulated in the 'path-style', where we would require that the path distance $x_p$ between each terminal pair is $\geq 1$, for all paths $p$ between the pair. As in previous lectures, this formulation would have the disadvantage of exponentially many constraints, but could be approached using the ellipsoid method. The ellipsoid method requires an efficient 'separation oracle' to reveal which constraint is violated by any proposed solution. For this problem we could use the results of the polytime all-pairs shortest path algorithm for this purpose.)

### 16.2.2 Rounding

Given an optimal LP solution, how can we round to a valid cut, and how can we analyze this scheme? The method applied to the multiway cut problem is no longer directly applicable, because the fact that we now only separate *pairs* of terminal nodes means that the $B$ centered at each terminal are no longer guaranteed to be disjoint.

The key idea to overcome this is to only charge the area to the sub-ball entirely enclosed by a given cut. Once the edges are charged to that cut, they are then removed from the graph, potentially changing area and volume calculations for subsequent steps. Crucially, these modifications ensure that the volumes remain disjoint.

Our derivation will involve dividing by $Vol(s_i, 0)$ at some point, which would be an undefined divide-by-zero operation. We avoid this by redefining volume slightly. Call $F$ the total volume of the graph. Then redefine volume by assigning volume $F/k$ to each terminal $s_i$.

$$Vol'(s_i, r) = F/k + \sum_e f_e c_e d_e$$

Now the total volume of the graph has doubled, so all previous volume lemmas still hold, with both sides multiplied by 2. The initial volume of a ball is then $F/k$, and its final volume must be $\leq F/k + F$.

**Algorithm**

- for each $i$, pick minimum $r$ such that $Area(s_i, r) \leq \alpha Vol(s_i, r)$
- make the cut $C_i$ at that $r$, remove those edges from the graph
- repeat for next $i$

We pick the $\alpha$ value to be $2\ln(k+1)$.

**Lemma 16.2.1** $r_i \leq 1/2\ \forall i$

**Proof:** As before, we know that $Area(s_i, r) \leq \frac{d}{dr} Vol(s_i, r)$.

Now assume $r_i \geq 1/2$. Then we must have that

$$\frac{d}{dr} Vol(s_i, r) > \alpha Vol(s_i, r)\ \forall r < 1/2$$

because otherwise we would have picked one of those smaller $r$.

We then manipulate the equation, integrating from $r = 0$ to $r = 1/2$.

$$\begin{aligned}
dVol(s_i, r) &> \alpha Vol(s_i, r)\, dr \\
\frac{1}{Vol(s_i, r)} dVol(s_i, r) &> \alpha\, dr \\
\int_0^{1/2} \frac{dVol(s_i, r)}{Vol(s_i, r)} &> \int_0^{1/2} \alpha\, dr \\
\ln\left(\frac{Vol(s_i, 1/2)}{Vol(s_i, 0)}\right) &> \alpha/2
\end{aligned}$$

Recall that the maximum possible value of $Vol$ is now $F + F/k$, and that $Vol(s_i, 0)$ is now defined to be $F/k$. Substitute these values in to get another inequality.

$$\ln(1+k) = \ln\left(\frac{F + F/k}{F/k}\right) > \ln\left(\frac{Vol(s_i, 1/2)}{Vol(s_i, 0)}\right) > \alpha/2$$

This gives us $\alpha < 2\ln(1+k)$. Since we have chosen $\alpha = 2\ln(1+k)$, we have derived a contradiction.

■

**Lemma 16.2.2** $\sum_i Vol(s_i, r_i) \leq 2LP^* = 2F$

**Proof:** This is acheived by construction. Our modified disjoint volumes now sum to no more than $2F$. Since $LP^*$ is equal to the volume of the full graph $F$, this is clearly true. ■

These lemmas will be used in our full derivation of the algoirthm approximation factor.

**Theorem 16.2.3** $\sum_i |C_i| = \sum_i Area(s_i, r_i) \leq 2\alpha LP^* \leq 2\alpha OPT$

This theorem simply follows from the combination of the previous lemmas and definitions. Plugging our chosen $\alpha = 2\ln(1+k)$ in shows that this algorithm achieves an $4\log(1+k)$-approximation [4].

Multicut is known to be APX-hard [1], meaning that it cannot be approximated within every constant factor.

Furthermore, if Unique Games Conjecture is true, it cannot even be approximated within any constant factor. A stronger version the Unique Games Conjecture further implies that it cannot be approximated with a factor $\Omega(\log\log n)$ either [3].

## 16.3 Integrality gap analysis

To derive an integrality gap for a given problem, we find a problem instance in which the optimal integral solution is significantly worse than the optimal fractional solution. The factor by which the integral solution is worse is known as the integrality gap. The existence such a gap bounds the performance of approximation algorithms based on LP rounding, since we cannot approximate $LP^*$ at a factor better than the integrality gap. Our specific approach will be to use a special mathematical stucture known as an expander graph to construct a problem instance with this property.

**Definition 16.3.1** *An* $\alpha$-expander graph $G = (V, E)$ *is a graph with the special property that for any subset* $S \subset V$ *such that* $|S| \leq |V|/2$*, we have* $|E(S, \bar{S})| \geq \alpha|S|$*.*

The study of expander graphs constitutes a very active research area, and expander graphs have been applied to many different problems in multiple fields [2]. For our purposes it is enough to know the basic definition property, and that there exist explicit methods for constructing graphs that have this, and other, properties.

We will now consider a multicut problem instance. Say that we have a degree 3 expander graph with constant $\alpha$. Let $(s_i, t_i)$ be the set of all pairs with distance $\geq \beta \log n$.

A feasible fractional solution is then given by

$$x_e = \frac{1}{\beta \log n} \ \forall e$$

By our problem definition, this is a feasible solution. The total cost is then given the by number of edges, divided by $\beta \log n$, which is $O(\frac{n}{\log n})$.

To get an integral solution, we can break the graph into components, which each must have size $\leq n/2$. The size of the components can be bounded in this way by the specification of the diameter of the expander graph, and our definition of $\beta$ in the problem.

This can be reasoned by observing that no 2 terminals can be in the same component, thus each diameter must be $\leq \beta \log n$. Since every node has degree 3, this bounds the number of nodes in a component by $\leq 3^{\beta \log n} \leq n/2$.

The expander property then shows that we must cut $O(n)$ edges.

$$\text{num edges in cut} \geq \frac{1}{2} \sum_i \alpha |C_i| = \frac{\alpha}{2} n$$

Thus the optimal integral solution is $O(n)$, while we have found a feasible fractional solution of $O(\frac{n}{\log n})$. Therefore we have shown an integrality gap of $\log n$.

## References

[1] The Complexity of Multiterminal Cuts Elias Dahlhaus, David S. Johnson, Christos H. Papadimitriou, P. D. Seymour, Mihalis Yannakakis. SIAM J. Comput. 23 (1994), 864-894.

[2] Expander graphs and their applications Shlomo Hoory; Nathan Linial; Avi Wigderson. Bull. Amer. Math. Soc. 43 (2006), 439-561.

[3] On the Hardness of Approximating Multicut and Sparsest-Cut. Shuchi Chawla, Robert Krauthgamer, Ravi Kumar, Yuval Rabani, D. Sivakumar. IEEE Conference on Computational Complexity 2005: 144-153

[4] Approximate max-flow min-(multi)cut theorems and their applications. Naveen Garg, Vijay V. Vazirani, Mihalis Yannakakis. STOC 1993: 698-707