**CMPUT 675: Approximation Algorithms** **Fall 2014**

Lecture 14 (Oct 3): Multicut (Part 2)

*Lecturer: Zachary Friggstad* *Scribe: Leah Hackman*

## 14.1 Multicut Problem

Recall the Multicut problem, in which we are given a graph $G = (V, E)$ with edge costs $c_e \geq 0$ and pairs of nodes $(s_1, t_1), \ldots, (s_k, t_k)$. We should find the cheapest subset of edges whose removal disconnects $s_i$ from $t_i$ for each $1 \leq i \leq k$. Consider the following LP relaxation.

$$\begin{aligned}
\text{minimize :} \quad & \sum_{e \in E} c_e \cdot x_e & & \\
\text{subject to :} \quad & \sum_{e \in P} x_e \geq 1 & & \text{for each path connecting some pair}(s_i, t_i) \\
& x_e \in \{0, 1\} & & \text{for each edge } e \in E
\end{aligned}$$

Here $x_e = 1$ corresponds to removing $e$. We relax the constraint $x_e \in \{0, 1\}$ to be $x_e \in [0, 1]$ so that we have a linear problem. Next, recall the following notation. Given an LP solution $x^*$, let:

- $d(u, v) =$ the minimum length path from $u$ to $v$ using the values of $x^*$ as edge lengths. Note hat $d(s_i, t_i) \geq 1$ because of the constraints in our LP.

Furthermore, for a subgraph $G'$ of $G$,

- $B_{G'}(v, r) = \{u \text{ in } G' : d(u, v) \leq r\}$ This forms a "ball" around $v$ with radius $r$.

- $V_{G'}(s_i, r) = \frac{OPT_{\text{LP}}}{k} + \sum_{\substack{e=(u,v): \\ u,v \in B_{G'}(s_i,r)}} c_e \cdot x_e + \sum_{\substack{e=(u,v): \\ u \in B_{G'}(s_i,r), \\ v \notin BG_{G'}(s_i,r)}} c_e \cdot (r - d(s_i, u))$.

  We refer to this as the "Volume" of the ball.

- $\delta_{G'}(B_{G'}(s_i, r)) = \{e \text{ an edge of } G' : e = (u, v), u \in B_{G'}(s_i, r), v \notin B_{G'}(s_i, r)\}$. This is the set of edges which cross the boundary of the ball.

Note that the definition of $B_{G'}(v, r)$ uses the original distances $d$ (i.e. the distances in $G$, not $G'$). Algorithm 1 is an approximation for Multicut.

We already have discussed the correctness of this algorithm last lecture, namely that the returned set $F$ disconnects all pairs. Now we seek to prove the approximation bounds for this algorithm.

**Algorithm 1** An approximation algorithm for the MULTICUT Problem using a Linear Program solution

$G' \leftarrow G$
$F \leftarrow \emptyset$
$\mathbf{x}^* \leftarrow$ an optimal LP solution
**while** There is an $s_i - t_i$ path in $G'$ for some $i$ **do**
  i) $r \leftarrow$ some $r \in [0, 1/2)$ such that $c(\delta_{G'}(B_{G'}(s_i, r))) \leq 2 \cdot \ln(k+1) \cdot V_{G'}(s_i, r)$
  ii) $F \leftarrow F \cup \delta_{G'}(B_{G'}(s_i, r))$
  iii) remove $B_{G'}(s_i, r)$ and all incidental edges from $G'$
**end while**
**return** $F$

**Theorem 1** *If step i always succeeds (i.e. we can always find a value* $r \in [0, 1/2)$ *s.t.* $c(\delta_{G'}(B_{G'}(s_i, r))) \leq 2 \cdot \ln(k+1) \cdot V_{G'}(s_i, r)$*) then the cost of the returned set* $F$ *is* $\leq 4 \cdot \ln(k+1) \cdot OPT_{LP}$.

**Proof of Theorem 1.** Consider each iteration of the algorithm and say that $(s_i, t_i)$ is one of the considered pairs. Let $G_i$ denote the graph $G'$ just before the ball around $s_i$ was removed in step iii.

For simplicity, let $B_i$ be the ball $B_{G_i}(s_i, r)$ in this iteration. Let $F_i$ be the edges added to $F$. That is, $F_i = \delta_{G_i}(B_i)$. Finally, let $r_i$ be the radius chosen in this iteration.

Note that $\delta_{G_i}(B_i) \cap \delta_{G_j}(B_j) = \emptyset$ for different $i, j$. This is because if, say, $i$ was considered earlier than $j$ then all edges incident to $B_i$ were removed from $G_i$ and $G_j$ is a subgraph of this graph. More generally, for two balls $B_i, B_j$ for indices $i, j$ considered by the algorithm we have that no edge is incident to both a vertex in $B_i$ and a vertex in $B_j$.

Finally, note that for any $e = (u, w) \in \delta_{G_i}(B_i)$ with $u \in B_i, w \notin B_i$ that $r_i - d(s_i, u) \leq x^*_e$. Otherwise $d(s_i, w) \leq d(s_i, u) + x^*_e < r$ which contradicts $w \notin B_i$.

We bound the cost of $F$ as follows, where all sums in the bounds below are restricted to $i \in \{1, \ldots, k\}$ such that some iteration of the algorithm considered pair $(s_i, t_i)$.

$$\begin{aligned}
cost(F) &= \sum_i cost(F_i) \\
&\leq 2 \cdot \ln(k+1) \cdot \sum_i V(s_i, r) \\
&\leq 2 \cdot \ln(k+1) \cdot \sum_i \left( \frac{OPT_{\text{LP}}}{k} + \sum_{\substack{e \text{ incident to} \\ \text{some } v \in B_i}} \overbrace{c_e \cdot x_e}^{\text{because } r - d(s_i,u) \leq x^*_e} \right) \\
&\leq 2 \cdot \ln(k+1) \cdot \left( \overbrace{OPT_{\text{LP}}}^{\substack{\text{because we consider} \\ \text{at most } k \text{ values of } i}} + \overbrace{OPT_{\text{LP}}}^{\substack{\text{no edge is incident to} \\ \text{more than one } B_i}} \right) \\
&= 4 \cdot \ln(k+1) \cdot OPT_{\text{LP}}
\end{aligned}$$

■

All that is left is to show it is always possible to find a value of $r$ for which $r \in [0, 1/2)$ and $c(\delta_{G_i}(B_{G_i}(s_i, r))) \leq 2 \cdot \ln(k+1) \cdot V(s_i, r)$ (here, $G_i$ also denotes the subgraph $G'$ just before the ball $B_{G'}(s_i, r)$ was removed).

Recall that for a continuous function on an interval $f : [a, b] \to \mathbb{R}$, the average value of $f$ in this interval is:

$$avg(f) = \frac{1}{a - b} \int_a^b f(x)dx.$$

This integral is still defined if $f$ is continuous at all but finitely many values in $[a, b]$ (i.e. the Riemann integral can "handle" finitely many discontinuities). There is some $\bar{x} \in [a, b]$ such that $f(\bar{x}) \leq \frac{1}{a-b} \int_a^b f(x)dx$ (i.e. if $f(\bar{x}) > z$ for all $x \in [a, b]$ then $\int_a^b f(x)dx > (b - a)z$).

Our goal is to show that the average value of $c(\delta_{G'}(B_{G'}(s_i, r)))/V_{G'}(s_i, r)$ is at most $2\ln(k+1)$ over $r \in [0, 1/2]$.

Now, let

- $B(s_i, 1/2) = \{v_1, \ldots, v_m\}$
- $r_j = d(s_i, v_j)$

Also, suppose that $0 = r_1 \leq r_2 \leq \cdots \leq r_m$, where $v_1 = s_i$ and say $r_{m+1} := 1/2$. For ease of notation, let us also define $c(r) := c(\delta_{G_i}(B_{G_i}(s_i, r)))$ and $V(r) := V_{G_i}(s_i, r)$.

Note that the function $V(r)$ is piecewise linear with the only possible "break" points at some $r_j$. Furthermore, for any $r \in (0, 1/2)$ such that $r \neq r_j$ for any $1 \leq j \leq m$ note that $\frac{dV(r)}{dr} = c(r)$.

For the the sake of intuition, first suppose $\frac{dV(r)}{dr}$ was defined at *all* $r \in [0, 1/2]$ and was equal to $c(r)$. If so, then the average value of the function $\frac{V(r)}{c(r)}$ in $[0, 1/2]$ is simply

$$\begin{aligned}
\frac{c(r)}{V(s_i, r)} &= \frac{1}{1/2} \cdot \int_0^{1/2} \frac{c(r)}{V(s_i, r)} dr \\
&= 2 \cdot \int_0^{1/2} \frac{dV(s_i, r)}{V(s_i, r)} \\
&= 2 \cdot (\ln(V(1/2)) - \ln(V(0))) \\
&= 2 \cdot \ln\left(\frac{V(1/2)}{V(0)}\right) \\
&\leq 2 \cdot \ln\left(\frac{OPT_{\text{LP}}/k + OPT_{\text{LP}}}{OPT_{\text{LP}}/k}\right) \\
&= 2 \cdot \ln(k + 1)
\end{aligned}$$

The inequality uses the observation that $V(r) \leq \frac{OPT_{\text{LP}}}{k} + OPT_{\text{LP}}$.

Observe that the inequality would not make sense if we did not include the extra $OPT_{\text{LP}}/k$ term in the definition of volume (otherwise $V(0) = 0$). It must be big enough to ensure that this expression is not too large. On the other hand, it must be small enough so that the value $k \cdot$ (correction term) is not too large in the proof of Theorem 1 above. The value $OPT_{\text{LP}}/k$ strikes the right balance.

This does not complete the proof because $V(r)$ may not be differentiable at points $r_1, \ldots, r_{m+1}$; they may even be points of discontinuity. An example of how such discontinuities can arise is in the Williamson and Shmoys text.

Such discontinuities are not a big problem since $V(r)$ is, at least, an *increasing* function in $[0, 1/2]$. Let us break up our integral into separate integrals over each of the continuous regions between our discontinuities (this is the natural way to handle a Riemann integral over a function with finitely many discontinuities).

For $1 \le j \le m$, let $f_j : [r_j, r_{j+1}] \to \mathbb{R}$ be the linear function with $f_j(r_j) = V(r_j)$ and $f_j(r_{j+1}) = V(r_j) + c(r_j) \cdot (r_{j+1} - r_j)$. Note that $V(r) = f_j(r)$ for all $r \in [r_j, r_{j+1})$ and that $f_j(r_{j+1}) \le V(r_j)$.

Now we can bound the average value of $c(r)/V(r)$ as follows.

$$
\begin{aligned}
\frac{1}{1/2}\int_0^{1/2} \frac{c(r)}{V(r)} dr &= 2\sum_{j=1}^{m} \cdot \int_{r_j}^{r_{j+1}} \frac{c(r)}{V(r)} dr \\
&= 2\sum_{j=1}^{m} \cdot \int_{r_j}^{r_{j+1}} \frac{c(r_j)}{f_j(r)} dr \qquad (c(r) = c(r_j) \text{ and } f_j(r) = V(r) \text{ for all but one } r \in [r_j, r_{j+1}]) \\
&= 2\sum_{j=1}^{m} \cdot \int_{r_j}^{r_{j+1}} \frac{df_j(r)}{f_j(r)} \\
&= 2\sum_{j=1}^{m} \ln(f_j(r_{j+1})) - \ln(f_j(r_j)) \\
&\le 2 \cdot \sum_{j=1}^{m} \ln V(r_{j+1}) - \ln V(r_j) \\
&= 2 \cdot (\ln V(1/2) - \ln V(0)) \qquad \text{(the sum is telescoping)} \\
&\le 2 \cdot \ln(k+1) \qquad \text{(the same arguments as above)}
\end{aligned}
$$

Thus even with the discontinuities, we can still achieve the same bound on our average, and we we are still guaranteed a point in our range which is equal to or less than this average value.

One final note, we require $r \in [0, 1/2)$, not just $r \in [0, 1/2]$. We can prove this for $r \in [0, 1/2)$ with a simple observation. If $c(r)/V(r)$ is constant over $[0, 1/2]$ then choosing any $r \in [0, 1/2)$ suffices. Otherwise, one can show that in fact $f(\bar{x}) < \frac{1}{b-a}\int_a^b f(x)dx$ for some $\bar{x} \in (a, b)$ if $f$ is nonconstant and right-continuous at every point (as in our case). In either case, we may take $r \in [0, 1/2)$.

Finally, we actually need to find such a value $r$ in polynomial time. Note that if $c(r) \le \ln(k+1)V(r)$ where $r_j \le r < r_{j+1}$ then it also holds for $r$ being infinitesimally smaller than $r_{j+1}$. So, we just need to check values that are slightly smaller than each $r_j$.

Even simpler: with a closer inspection of the analysis we can see that it suffices to choose the value $r = r_j$ for some $j = 1, \ldots, m$ with $r_j < 1/2$ that minimizes

$$
\frac{c(r_j)}{\frac{OPT_{\mathrm{LP}}}{k} + \sum\limits_{\substack{e \text{ incident to} \\ \text{some } v \in B_j(r_j)}} c_e x_e^*}.
$$

For example, the cost analysis simply used $r - d(s_i, u) \le x_e^*$ for estimating the volume.

## Discussion

This rounding algorithm is due to Garg, Vazirani, and Yannakakis [GVY96]. To date, it is the best known approximation for the MULTICUT problem in undirected graphs. The only lower bounds known are that it is

NP-hard to approximate better than some constant $c$. Furthermore, under the so-called *unique games conjecture* (which we will discuss later), there is no $c$-approximation for *any* constant $c$ [C+06].

In directed graphs, the situation is much worse. The best approximation is roughly an $O(n^{11/23})$-approximation [AAC07] ("roughly" means some $\log n$ factors are omitted from the expression). In fact, unless NP $\subseteq$ ZPP (i.e. unless SAT can be solved by a randomized algorithm that always returns the correct solution in *expected* polynomial time) there is no $2^{\log^{1-\epsilon}(n)}$-approximation for any constant $\epsilon > 0$ [CK09]. This means there can be no $(\log^c n)$-approximation for any constant $c > 0$. Furthermore, there is strong evidence that in fact we cannot approximate the problem better than $n^\delta$ for some constant $\delta > 0$ ([CK09] again).

Finally, the integrality gap analysis is tight. That is, there are instances of MULTICUT whose optimum solution value is $\Omega(\log k \cdot OPT_{\mathrm{LP}})$. Chapter 20.3 of Vazirani's text describes such an example.

## References

AAC07 A. Agarwal, N. Alon, and M. Charikar, Improved approximation for directed cut problems, In Proceedings of *ACM Symposium on Theory of Computing*, 2007.

C+06 S. Chawla, R. Krauthgamer, R. Kumar, Y. Rabani, and D. Sivakumar, On the hardness of approximating multicut and sparsest cut, *Computational Complexity*, 15:94–114, 2006.

CK09 J. Chuzhoy and S. Khanna, Polynomial flow-cut gaps and the hardness of directed cut problems, *Journal of the ACM*, 52(2): 6, 2009.

GVY96 N. Garg, V. V. Vazirani, and M. Yannakakis, Approximate max-flow min-(multi)cut theorems and their applications, *SIAM Journal on Computing*, 25:235–251, 1996.