Two Logarithmic Approximation Algorithms for Multicut[1]

- In this lecture we consider the multicut problem which generalizes the multiway cut problem. As usual, we are given an undirected graph $G = (V, E)$ with non-negative costs $c(e)$ on edges. We are also given k *pairs* of vertices $\{s_i, t_i\}_{i=1,\ldots,k}$. The objective is to find a subset $F \subseteq E$ of minimum cost such that in $G \setminus F$, s_i is disconnected from t_i. Note that s_i could remain connected to t_j. We describe *two* $O(\log k)$-approximation algorithms for this problem. They are both based on the same distance-based LP relaxation.

$$\mathsf{lp} := \min \sum_{e \in E} c(e) x_e \qquad \text{(Multicut LP)}$$

$$\begin{aligned}
& d_{uv} \le x_e, && \forall e \in E, e = (u,v) && (1)\\
& d_{uw} \le d_{uv} + d_{vw}, && \forall i \in F,\ \forall \{u, v, w\} \subseteq V && (2)\\
& d_{vv} = 0, && \forall v \in V && (3)\\
& d_{s_i t_i} \ge 1, && \forall 1 \le i \le k && (4)
\end{aligned}$$

- ***Randomized Rounding Algorithm.*** The first rounding algorithm we see is a generalization of the multiway cut algorithm. We select a random radius $r \in (0, 0.5)$ uniformly at random. Then, we wish to go over each terminal s_i and "carve out" the region of radius r around S_i. The twist in this algorithm is this: go over the terminals also randomly.

procedure RANDOMIZED MULTICUT($G = (V, E)$, $c(e) \ge 0$ on edges,$\{s_i, t_i\}_{i=1,\ldots,k}$):
  Solve (Multicut LP) to obtain x_e's and d_{uv}'s.
  Randomly sample $r \in (0, 0.5)$ uniformly.
  Randomly sample σ, a permutation of $\{1, \ldots, k\}$.
  Let $S_i := \{v : d_{s_i v} \le r\}$ and let $E[S_i] := \{(u, v) : u, v \in S_i\}$.
  For $1 \le i \le k$: add $\partial S_{\sigma(i)} \setminus \bigcup_{j<i} E[S_{\sigma(j)}]$ to F.
  return F.

- *Analysis.* First let us observe F is a valid multicut.

Claim 1. F separates all s_i, t_i pairs.

Proof. By design, observe that for any i, the subset S_i doesn't contain both s_j and t_j for any j. Now, note that since $\partial S_{\sigma(i)} \setminus \bigcup_{j<i} E[S_{\sigma(j)}]$ is added to F, in $G \setminus F$ the vertex $s_{\sigma(i)}$ is disconnected from all vertices outside $S_{\sigma(i)}$, except maybe those in $S_{\sigma(j)} : j < i$ which contained the vertex $s_{\sigma(i)}$. By the observation above, such $S_{\sigma(j)}$'s don't contain $t_{\sigma(i)}$. Therefore, $s_{\sigma(i)}$ is disconnected from $t_{\sigma(i)}$. ☐

[1] *Lecture notes by Deeparnab Chakrabarty. Last modified : 6th Mar, 2023*
These have not gone through scrutiny and may contain errors. If you find any, or have any other comments, please email me at deeparnab@dartmouth.edu. Highly appreciated!

Theorem 1. The expected cost of the edges F returned by RANDOMIZED MULTICUT is $\leq 2H_k$lp where H_k is the kth Harmonic number.

Proof. Fix an edge (u,v). The proof of the theorem follows if we prove $\mathbf{Pr}[(u,v) \in F] \leq 2H_k \cdot d_{uv}$. Note that the probability is now both over our choice of r *and* the random permutation of the terminals.

Define $\mathcal{E}_i(u,v)$ to be the event that *exactly* one of u or v lies in S_i. That is, $\min(d_{s_iu}, d_{s_iv}) \leq r < \max(d_{s_iu}, d_{s_iv})$. Define $\mathcal{E}'_i(u,v)$ to be the event that *neither* u nor v lie in S_i, that is $r < \min(d_{s_iu}, d_{s_iv})$. Now, note that the edge (u,v) appears in the solution F if and only if there is some i such that $\mathcal{E}_{\sigma(i)}$ occurs **and** for all $j < i$, $\mathcal{E}'_{\sigma(j)}$ occurs. That is,

$$\mathbf{Pr}[(u,v) \in F] = \mathbf{Pr}_{\sigma,r}\left[\exists i:\ \mathcal{E}_{\sigma(i)}(u,v)\ \textbf{ and } \bigwedge_{j<i} \mathcal{E}'_{\sigma(j)}(u,v)\right] \tag{5}$$

Fix an i between 1 and k. Without loss of generality, assume $d_{s_{\sigma(i)}u} \leq d_{s_{\sigma(i)}v}$. Note that $\bigwedge_{j<i} \mathcal{E}'_{\sigma(j)}(u,v)$ occurs only if $r < d_{s_{\sigma(j)}v}$ for all $j < i$. But $\mathcal{E}_{\sigma(i)}(u,v)$ occurs only if $r \geq d(s_{\sigma(i)}, u)$. So, we can upper bound the probability in the RHS above as

$$\mathbf{Pr}_{\sigma,r}\left[\mathcal{E}_{\sigma(i)}(u,v)\ \textbf{ and } \bigwedge_{j:<i} \mathcal{E}'_{\sigma(j)}(u,v)\right] \leq \mathbf{Pr}_{\sigma,r}\left[r \in [d_{s_{\sigma(i)}u}, d_{s_{\sigma(i)}v}]\ \textbf{ and } \bigwedge_{j<i}\left\{d_{s_{\sigma(i)}u} < d_{s_{\sigma(j)}u}\right\}\right]$$

Note that the two events in the RHS above are independent: the first depends only on r, the second depends only on σ, and they were chosen independently. So, by union bound we get that the RHS of (5) is at most

$$\sum_{i=1}^{k} \underbrace{\mathbf{Pr}_{r}\left[r \in [d_{s_{\sigma(i)}u}, d_{s_{\sigma(i)}v}\right]}_{\text{call this } \pi_1(i)} \cdot \underbrace{\mathbf{Pr}_{\sigma}\left[\bigwedge_{j<i}\left\{d_{s_{\sigma(i)}u} < d_{s_{\sigma(j)}u}\right\}\right]}_{\text{call this } \pi_2(i)}$$

We know $\pi_1(i) = Pr_r\left[r \in [d_{s_{\sigma(i)}u}, d_{s_{\sigma(i)}v}\right] \leq 2d_{uv} \leq 2x_e$. This is similar to the mincut argument; r is chosen randomly from an interval of length 0.5 and the length of $[d_{s_{\sigma(i)}u}, d_{s_{\sigma(i)}v}]$, by (2) is at most $d_{uv} \leq x_e$.

To evaluate $\pi_2(i)$, consider the k distances d_{s_iu} from u to each s_i. What π_2 is asking is to figure out the probability that in a random permutation of these k distances, the ith distance is the minimum among the first i. This is precisely $1/i$. Therefore, the probability in the RHS of (5) is at most $\sum_{i=1}^{k} \frac{2x_e}{i} = 2H_k \cdot x_e$. This completes the proof. □

- ***A Region Growing Algorithm.*** We now describe another algorithm for the multicut problem. This algorithm uses a technique called *region growing* which will be useful for the next cut-problem we look at. It also has applications in other related problems.

We start with a couple of definitions. Let's fix a solution to (Multicut LP), and a parameter $r \in [0, 0.5)$. For a subset $U \subseteq V$, define $S_i(r;U) := \{u \in U : d_{s_iu} \leq r\}$. Define $\partial S_i(r;U) := \{(u,v) \in E :$

$u \in S_i(r;U), v \in U \setminus S_i(r)\}$, and define $E[S_i(r;U)] = \{(u,v) \in E : u,v \in S_i(r;U)\}$. These definitions are similar to the ones used above, except we pass on an extra parameter U.

Next, define the "volume" of a ball of radius r around the center s_i.

$$\texttt{Vol}_i(r;U) := \frac{\mathsf{lp}}{k} + \sum_{(u,v)\in E[S_i(r;U)]} c(u,v)d_{uv} + \sum_{(u,v)\in \partial S_i(r;U)} c(u,v)\cdot(r - d_{s_i u}) \quad \text{(LP volume)}$$

It's best to think of this volume as the set $S_i(r;U)$'s contribution to the LP objective. There are three parts above. The first, lp/k is an initialization which is kept for a technical reason that you will make sense soon. The second summation is the contribution to the LP objective due to edges complete present inside $S_i(r;U)$. The third is considering edges in $\partial S_i(r;U)$ and sharing some of the LP contribution on these edges and attributing it to i. Note that for all such edges, $r - d_{s_i u} \leq d_{s_i v} - d_{s_i u} \leq d_{uv}$ where the first inequality follows from the fact that $v \in U \setminus S_i(r)$, and the second is triangle inequality.

The following observation follows from the definition.

Claim 2. Fix any $r \in (0, 0.5)$ and any i and any $U \subseteq V$. The set $S_i(r;U)$ cannot contain s_j and t_j for any $1 \leq j \leq k$.

Proof. For any two vertices $u,v \in S_i(r;U)$, triangle inequality dictates $d_{uv} \leq d_{us_i} + d_{vs_i} \leq 2r < 1$. Since $d_{s_j t_j} \geq 1$, they both can't be in the same $S_i(r;U)$. ☐

This suggests the following algorithm. Figure out certain radii r_i's and peel out the "region of radius r" around the terminal and delete. The boundaries of these "chunks" form a valid multicut.

procedure Region Growing Multicut($G = (V,E)$, $c(e) \geq 0$,$\{s_i,t_i\}_{i=1,\ldots,k}$):
  Solve (Multicut LP) to obtain x_e's and d_{uv}'s.
  $U \leftarrow V$; $\mathcal{B} \leftarrow \emptyset$; $I \leftarrow \emptyset$. ▷ *U is the set of alive vertices; $\mathcal{B}$ is collection of balls.*
  for $1 \leq i \leq k$ **do**:
    If $s_i \in S_j(r_j;U)$ for $j < i$, skip this for loop.
    Otherwise, find $r_i \in [0, 0.5)$ which minimizes $\frac{\sum_{e\in\partial S_i(r_i;U)} c(e)}{\texttt{Vol}_i(r_i;U)}$.
    ▷ *There are at most n different r's such that $S_i(r;U)$ are distinct*
    $U \leftarrow U \setminus S_i(r_i;U)$
    Add $B_i := S_i(r_i;U)$ to $\mathcal{B}$.
  return $F \leftarrow \bigcup_{B\in\mathcal{B}} \partial B$.

- *Analysis.*

Theorem 2. Region Growing Multicut returns a valid multicut F with cost $\sum_{e\in F} c(e) \leq 4\ln(k+1)\mathsf{lp}$.

Observe, by definition, the sets $B \in \mathcal{B}$ are disjoint sets. Furthermore, no $B \in \mathcal{B}$ contains both s_j and t_j for any $1 \le j \le k$; this follows from Claim 2. Therefore, F is a valid multicut. Furthermore, each $B \in \mathcal{B}$ is $S_i(r_i; U_i)$ for some subset $U_i \subseteq V$ which was the alive subset of vertices when this ball was being added. Let $I \subseteq [k]$ be the i's present in this enumeration; these are the s_i's not "gobbled" by other $S_j(r_j; U)$'s.

Claim 3. $\sum_{i \in I} \mathtt{Vol}_i(r_i; U_i) \le 2\mathsf{lp}$.

Proof. Note that the sum of the volumes is at most

$$\mathsf{lp} + \sum_{(u,v) \in \cup_{i \in I} E[S_i(r_i;U_i)]} c(u,v) d_{uv} \; + \; \sum_{i \in I} \sum_{(u,v) \in \partial S_i(r_i;U_i)} c(u,v) d(u,v)$$

Now note that any edge $(u,v) \in E$ appears in at most one $E[S_i(r_i; U_i)]$ or $\partial S_i(r_i; U_i)$: it is the first i for which one of the end points enters $S_i(r_i; U_i)$. Therefore, the last two summations add up to at most $\sum_{(u,v) \in E} c(u,v) d_{uv} \le \sum_{e \in E} c_e x_e = \mathsf{lp}$. □

The heart of the analysis is in the following lemma.

Lemma 1. (Region growing lemma) Fix any subset $U \subseteq V$ and any $s_i \in U$. There exists a $r_i \in [0, 1/2)$ such that

$$\sum_{(u,v) \in \partial S_i(r;U)} c(u,v) \le 2 \ln(k+1) \cdot \mathtt{Vol}_i(r_i; U)$$

Proof. As defined, note that $\mathtt{Vol}_i(r; U)$ is a continuous, piece-wise linear function of r, and the crucial observation is that

$$\frac{d\mathtt{Vol}_i(r;U)}{dr} = \sum_{(u,v) \in \partial S_i(r;U)} c(u,v)$$

This means that if $\sum_{(u,v) \in \partial S_i(r;U)} c(u,v)$ is large, in particular larger than $2\ln(k+1)\mathtt{Vol}_r(r_i; U)$, then the rate of increase of the volume is rather large. On the other hand, even at $r = 0.5$, the volume can be at most the lp. And it began at lp/k (this is the technical reason to have this first term in the definition), and so the rate can't be large throughout, proving the lemma.

A little more formally, for the sake of contradiction, assume that the lemma's assertion is false. Then, we get the partial differential inequality

$$\forall r \in [0, 0.5), \quad \frac{d\mathtt{Vol}_i(r;U)}{dr} > 2\ln(2k) \cdot \mathtt{Vol}_i(r;U) \;\; \Rightarrow \;\; \frac{d\mathtt{Vol}_i(r;U)}{\mathtt{Vol}_i(r;U)} > 2\ln(k+1) \cdot dr$$

Therefore, if we integrate with r going from 0 to 0.5,

$$\int_{\mathtt{Vol}_i(0)}^{\mathtt{Vol}_i(0.5)} \frac{d\mathtt{Vol}_i(r)}{\mathtt{Vol}(r)} > 2\ln(2k) \int_0^{1/2} dr$$

The LHS integrates to $\ln\left(\frac{\mathtt{Vol}_i(0.5;U)}{\mathtt{Vol}_i(0;U)}\right)$. By design, $\mathtt{Vol}_i(0; U) = \mathsf{lp}/k$. And, $\mathtt{Vol}_i(0.5) \le \mathsf{lp}(1 + \frac{1}{k})$. Therefore, the LHS is at most $\ln(k+1)$. The RHS, however, integrates to $\ln(k+1)$, giving the desired contradiction. □

In the algorithm, we pick r_i's which minimize the ration of $c(\partial S_i(r_i; U))/\mathtt{Vol}_i(r_i; U)$, and so this ratio is at most $2\ln(2k)$. Therefore, the cost of the edges deleted is at most

$$c(F) = \sum_{B \in \mathcal{B}} c(\partial B) = \sum_{i \in I} c(\partial S_i(r_i; U_i)) \leq 2\ln(k+1) \cdot \sum_{i \in I} \mathtt{Vol}_i(r_i; U_i) \underbrace{\leq}_{\text{Claim 3}} 4\ln(k+1)\mathsf{lp}$$

completing the proof of Theorem 2.

Notes

The region growing algorithm is from the paper [4] by Garg, Vazirani, and Yannakakis and was the first $O(\log k)$-approximation for the multicut problem. The technique of region growing itself is inspried by the seminal paper [5] by Leighton and Rao on the sparsest cut problem which we will discuss in a subsequent lecture. The randomized rounding algorithm is from the paper [2] by Calinescu, Karloff, and Rabani which followed their paper [1] on the multiway cut problem. On the other hand, it is possible there may not be any constant factor approximations for the multicut problem: the paper [3] by Chawla, Krauthgamer, Kumar, Rabani, and Sivakumar shows that it is UGC-hard to obtain any constant factor approximation.

References

[1] G. Calinescu, H. Karloff, and Y. Rabani. An Improved Approximation Algorithm for Multiway Cut. *J. Comput. Syst. Sci.*, 60(3):564–574, 2000.

[2] G. Calinescu, H. Karloff, and Y. Rabani. Approximation algorithms for the 0-extension problem. *SIAM Journal on Computing*, 34(2):358–372, 2005.

[3] S. Chawla, R. Krauthgamer, R. Kumar, Y. Rabani, and D. Sivakumar. On the hardness of approximating multicut and sparsest-cut. *computational complexity*, 15(2):94–114, 2006.

[4] N. Garg, V. V. Vazirani, and M. Yannakakis. Approximate max-flow min-(multi)cut theorems and their applications. *SIAM J. Comput.*, 25(2):235–251, 1996. Prelim. Version in STOC 1993.

[5] F. T. Leighton and S. Rao. An approximate max-flow min-cut theorem for uniform multicommodity flow problems with application to approximation algorithms. In *Proc., IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 422–431, 1988.