KAIST, School of Computing, Spring 2024 **Scribed By: Eunjung KIM**
Algorithms for NP-hard problems (CS492) **Lecture #24**
Lecture: Eunjung KIM **16 May 2024**

Contents

1 LP-based rounding, Approximation, Max-Flow Min-Cut

1.1 2-Approximation for Weighted Vertex Cover via LP rounding

The integer program for Weighted Vertex Cover is formulated as follows.

$$\begin{aligned} \min \sum_{u \in V(G)} \omega_u \cdot x_u & \\ x_u + x_v \geq 1 & \qquad \forall (u,v) \in E(G) \\ x_u \in \{0,1\} & \qquad \forall u \in V(G) \end{aligned}$$

The linear program relaxation of the above formulation can be obtained simply by replacing the constraint $x_u \in \{0,1\}$ by $x_u \geq 0$. Notice that we dropped the constraint $x_u \leq 1$ as the objective is to minimize $\sum_{u \in V(G)} \omega_u \cdot x_u$ for non-negative coefficients while the constraint will be satisfied once any of the two variables involved in the constraint has value at least 1.

The algorithm is very simple. Solve the LP for Weighted Vertex Cover, and let x^* be a fractional optimal solution. Output the set $S := \{v \in V : x^*_v \geq 0.5\}$.

Analysis of the deterministic rounding algorithm. Let S be the output vertex set.

First, observe that S is indeed a vertex cover of G because if this is not true, there exists an edge $e = uv$ such that $x^*_u + x^*_v < 1$. This is impossible as x^* is a feasible solution to the LP.

Second, we argue that $\omega(S) \leq 2 \cdot \omega(\texttt{opt})$. This is because $2 \cdot x^*_v \geq 1$ for every $v \in S$, and

$$\omega(S) = \sum_{v \in S} \omega_v \cdot 1 \leq \sum_{v \in S} \omega_v \cdot (2x^*_v) \leq 2 \cdot \sum_{v \in V} \omega_v \cdot x^*_v \leq 2 \cdot \omega(\texttt{opt}).$$

1.2 Linear Program and Duality

See Chapter 12 of the textbook "Approximation Algorithms" by Vazirani.

1.3 (s,t)-mincut via LP rounding and Max-Flow Min-Cut Theorem

The material in this subsection is based on Chandra Chekrui's lecture note, Chapter 14. We consider a directed graph $G = (V,E)$ with a fixed pair of vertices, s and t of G.

Let $\mathcal{P}_{s,t}$ be the collection of all (s,t)-path in G. The linear program for Minimum (s,t)-Cut is formulated as follows.

$$\begin{aligned} \min \sum_{e \in E(G)} \omega_e \cdot x_e & \\ \sum_{e \in P} x_e \geq 1 & \qquad \forall P \in \mathcal{P}_{s,t} \\ x_e \geq 0 & \qquad \forall e \in E(G) \end{aligned}$$

The dual LP of the linear program for MINIMUM (s,t)-CUT is:

$$\begin{aligned} \max \sum_{P \in \mathcal{P}_{s,t}} z_P & & \\ \sum_{\substack{P \in \mathcal{P}_{s,t}: \\ P \text{ traverses } e}} z_P \leq \omega_e & & \forall e \in E(G) \\ z_P \geq 0 & & \forall P \in \mathcal{P}_{s,t} \end{aligned}$$

The dual LP program seeks to find a maximum flow; fractional assignment of non-negative value on each (s,t)-path so that the total flow on each edge e does not exceed its capacity c_e. Therefore, by the strong LP duality theorem it holds that

$$\text{max flow value} = \text{weight of min fractional cut} \leq \text{weight of min cut}.$$

In fact, we will show that there is a (integral) minimum (s,t)-cut whose weight matches the weight of a fractional minimum (s,t)-cut. We show this *randomized rounding based an LP solution*, which can be easily derandomized. Once this is established, the above inequality implies that there is a minimum (s,t)-cut whose weight matches the maximum flow value. This is precisely Max-Flow Min-Cut Theorem.

Algorithm 1 Randomized Rounding for MINIMUM (s,t)-CUT

1: **procedure** θ**-mincut**(G, ω, s, t)
2: Solve LP for MINIMUM (s,t)-CUT to obtain an optimal fractional solution x^*.
3: Compute the distance $\texttt{dist}_G(s,v)$ for every $v \in V$ with x^*_e interpreted as the length of edge e.
4: Choose $\theta \in (0,1)$ uniformly at random.
5: Let $S := B(s,\theta) \subseteq V$ be the set of vertices v with $\texttt{dist}_G(s,v) \leq \theta$.
6: **return** $\delta(S)$.

Analysis of the procedure θ**-mincut**(G, ω, s, t)**.** We argue that the output edge set at line 6 is indeed an (s,t)-cut, and the expected cost of an output (s,t)-cut of the procedure is at most $\sum_{e \in E} \omega_e \cdot x^*_e$, namely the optimal objective value of the LP. This implies that there is an (s,t)-cut whose weight equals $\sum_{e \in E} \omega_e \cdot x^*_e$.

Lemma 1. *The output set of the procedure* θ**-mincut**(G, ω, s, t) *is an* (s,t)*-cut.*

Proof: Note that an edge set $\delta(S)$ for a vertex set $S \subseteq V$ is an (s,t)-cut if and only if $|S \cap \{s,t\}| = 1$. As $\texttt{dist}_G(s,s) = 0 \leq \theta$ and $B(s,\theta) \subseteq V$ contains s already, it suffices to argue that $t \notin B(s,\theta) \subseteq V$. Indeed, if t is contained in $B(s,\theta) \subseteq V$, then $\texttt{dist}_G(s,t) \leq \theta < 1$. This means that there is an (s,t)-path P in G such that $\sum_{e \in E(P)} x^*_e < 1$. Then the constraint of the primal LP corresponding to P is violated by x^*, contradicting to the feasibility of x^*. $\square$

Lemma 2. *The expected weight of an output set is at most* $\sum_{e \in E} \omega_e \cdot x^*_e$.

Proof: The expected weight of an output set is

$$\sum_{e = xy \in E} \omega_e \cdot \texttt{prob}[\texttt{dist}_G(s,x) \leq \theta \text{ and } \texttt{dist}_G(s,y) > \theta]$$

and the event of "$\texttt{dist}_G(s,x) \le \theta$ and $\texttt{dist}_G(s,y) > \theta$" will happen when θ is chosen in the interval $[\texttt{dist}_G(s,x), \texttt{dist}_G(s,y))$. The probability of this event is precisely

$$\frac{\texttt{dist}_G(s,y) - \texttt{dist}_G(s,x)}{\text{the length of the interval that } \theta \text{ is chosen from}}.$$

As $\texttt{dist}_G(s,y) \le \texttt{dist}_G(s,x) + x^*_e$, the probability of the said event is at most x^*_e. This yields the desired expectation. □

Lemmas 1 and 4 combined implies that there exists an (s,t)-cut whose weight is at most $\sum_{e\in E} \omega_e \cdot x^*_e$. As an (s,t)-cut, or equivalently an integral feasible solution to the LP, is lower bounded by the optimal fractional objective value which equals $\sum_{e\in E} \omega_e \cdot x^*_e$. Therefore, such an (s,t)-cut is a minimum (s,t)-cut. By the Strong LP duality Theorem, we conclude that the weight of a minimum (s,t)-cut equals the maximum flow value; indeed the dual LP formulates the maximum flow problem.

1.4 $(2-\frac{2}{k})$-approximation for Edge Multiway Cut via LP rounding

Let $T = \{t_1, \ldots, t_k\}$ be the set of terminals and for $1 \le i < j \le k$, let $\mathcal{P}_{i,j}$ be the collection of all (t_i, t_j)-path in G. The linear program for Edge Multiway Cut is formulated as follows.

$$\begin{aligned}
\min \sum_{e\in E(G)} & \omega_e \cdot x_e & & \\
\sum_{e\in P} x_e &\ge 1 & \forall P \in \mathcal{P}_{i,j} \quad & \text{for all } 1 \le i < j \le k \\
x_e &\ge 0 & & \forall e \in E(G)
\end{aligned}$$

Algorithm 2 Randomized Rounding for Edge Multiway Cut

procedure θ-**mwc**(G, ω, T)
- Solve LP for Minimum (s,t)-Cut to obtain an optimal fractional solution x^*.
- For each $i \in [k]$ and $v \in V$, compute $\texttt{dist}_G(t_i, v)$ with x^*_e interpreted as the length of edge e.
- Choose $\theta \in (0, \frac{1}{2})$ uniformly at random.
- For each $i \in [k]$, let $S_i := B(t_i, \theta) \subseteq V$ be the set of vertices v with $\texttt{dist}_G(t_i, v) \le \theta$.
- **return** $\bigcup_{i=1}^k \delta(S_i)$.

Lemma 3. *The output set of the procedure θ-**mwc**(G, ω, T) is a multiway cut.*

Proof: Clearly, $t_i \in S_i$ for each $i \in [k]$. Moreover, S_i does not contain any terminal t_j for $j \ne i$ because $\texttt{dist}_G(t_i, t_j) \ge 1 > \theta$. Therefore $\delta(S_i)$ is an isolating cut for t_i, i.e. an $(t_i, T \setminus t_i)$-cut of G. It follows that $\bigcup_{i=1}^k \delta(S_i)$ pairwise separates all terminals of T. □

Lemma 4. *The expected weight of an output set is at most $2 \cdot \sum_{e\in E} \omega_e \cdot x^*_e$.*

Proof: The expected weight of an output set is at most

$$\sum_{e=xy\in E} \omega_e \cdot \texttt{prob}[\texttt{dist}_G(t,x) \le \theta \text{ and } \texttt{dist}_G(t,y) > \theta \text{ for some } t \in T],$$

where x is an endpoint of e which not farther from t_i than the other endpoint y is, and this choice can be different for a fixed edge e depending on the terminal t_i.

To analyze the probability of the event [$\mathtt{dist}_G(t,x) \le \theta$ and $\mathtt{dist}_G(t,y) > \theta$ for some $t \in T$], we define $B_i^{0.5}$ as the set of vertices v with $\mathtt{dist}_G(t_i, v) < 0.5$.

There are a few cases to consider for the type of an edge $e = xy$. The cases are exclusive, and we assume that none of the preceding cases are applicable for the later case.

1. an endpoint of e, say x, belongs to at least two balls $B_i^{0.5}$ and $B_j^{0.5}$.
2. (1 is not applicable, and) $x \in B_i^{0.5}$ and $y \in B_j^{0.5}$ for $i \neq j$.
3. (1-2 are not applicable, and) there exists a unique ball $B_i^{0.5}$ such that $B_i^{0.5} \cap \{x, y\} \neq \emptyset$.
4. (1-3 are not applicable, and) there is no ball $B_i^{0.5}$ such that $B_i^{0.5} \cap \{x, y\} \neq \emptyset$.

Type 1 edges do not exist because if $x \in B_i^{0.5} \cap B_j^{0.5}$, then $\mathtt{dist}_G(t_i, t_j) < 1$ and x^* is infeasible. Consider a Type 2 edge $e = xy$ and note that $\mathtt{dist}_G(t_i, y) > 0.5$ and $\mathtt{dist}_G(t_j, x) > 0.5$. Therefore,

$$\begin{aligned}
&\mathtt{prob}[\mathtt{dist}_G(t,x) \le \theta \text{ and } \mathtt{dist}_G(t,y) > \theta \text{ for some } t \in T] \\
&\le \max\{\mathtt{prob}[\mathtt{dist}_G(t_i,x) \le \theta], \mathtt{prob}[\mathtt{dist}_G(t_j,y) \le \theta]\} \\
&\le \frac{\max\{0.5 - \mathtt{dist}_G(t_i,x), 0.5 - \mathtt{dist}_G(t_j,y)\}}{\text{the length of the interval that } \theta \text{ is chosen from}}\} \\
&\le 2 \cdot (1 - \mathtt{dist}_G(t_i,x) - \mathtt{dist}_G(t_j,y)) \\
&\le 2 \cdot x_e^*
\end{aligned}$$

where the last inequality comes from $\mathtt{dist}_G(t_i,x) + x_e^* + \mathtt{dist}_G(t_j,y) \ge \mathtt{dist}_G(t_i,t_j) \ge 1$.

For Type 3 edge e, the same analysis for the randomized rounding for MINIMUM (s,t)-CUT easily yields the probability bound at most $2x_e^*$. Notice that both endpoints or exactly one endpoint of e may lie in $B_i^{0.5}$. For Type 4 edge e, the probability is simply zero.

To sum up,

$$\sum_{e=xy \in E} \omega_e \cdot \mathtt{prob}[\mathtt{dist}_G(t,x) \le \theta \text{ and } \mathtt{dist}_G(t,y) > \theta \text{ for some } t \in T] \le 2 \cdot \sum_{e=xy \in E} \omega_e \cdot x_e^*,$$

which completes the proof. $\square$

One can subtract the factor $2/k$ from the approximation ratio by not considering some terminal in both the approximation algorithm and in the analysis.