Part 8 Linear Program of Max Flow

(**Definition 1**) For a **flow network** $G=(V, E)$ with a source vertex s, a sink vertex t, and a capacity c associated with each edge e, we can formulate the **Max Flow Problem** as the following LP:

$$\begin{aligned} \max & \sum_{v\in V} f(s,v) \\ \text{s.t.} & \sum_{u\in V} f(u,v) = \sum_{u\in V} f(v,u), \forall v \in V-\{s,t\} \cdot \\ & f(u,v) \le c(u,v), \forall (u,v)\in E \\ & f(u,v) \ge 0, \forall (u,v) \in E \end{aligned}$$

(**Definition 2**) (LP in **Path Formulation**) We can rewrite the LP of a flow network in **Definition 1** as follow:

$$\begin{aligned} \max & \sum_{p\in P} x_p \\ \text{s.t.} & \sum_{p\in P:(u,v)\in p} x_p \le c(u,v), \forall (u,v)\in E, \\ & x_p \ge 0, \forall p \in P \end{aligned}$$

where we introduce a variable x_p to denote the **flow** on **path** p, for each possible s-t **simple path** (in which there may be repeated vertices). Let P denote the set of all such s-t paths.

Especially, we don't need to consider the **conservation constraint** in the aforementioned LP, because for each path, all the edges have the same flow value, i.e., the flow coming into a vertex equals the flow coming out of the vertex **on this path**. Thus, compared with **Definition 1**, we have a simpler formulation.

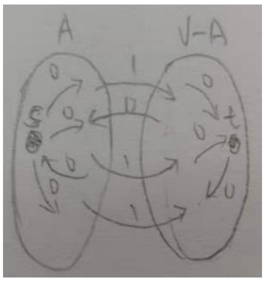
Note that the number of variables $\{x_p\}$ is **exponential**, but we don't care about it, since we don't need to actually solve the LP. We just using it to deepen our understanding about **Max Flow**.

Let the aforementioned LP as the **Primal LP**. We have the following **Dual LP**, where we introduce a variable $y(u, v)$ for each **edge** (u, v):

$$\begin{aligned} \min & \sum_{(u,v)\in E} c(u,v)y(u,v) \\ \text{s.t.} & \sum_{(u,v)\in p} y(u,v) \ge 1, \forall p\in P \cdot \\ & y(u,v) \ge 0, \forall (u,v)\in E \end{aligned}$$

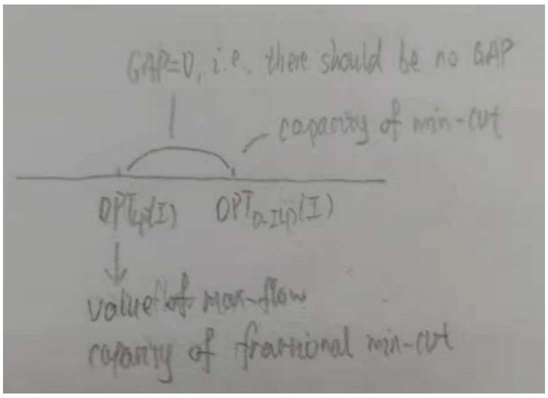
Think of $y(u, v)$ as the '**length**' of **edge** (u, v). The constrain $\sum_{(u,v)\in p} y(u,v) \ge 1$ indicates that the **length** of any s-t **path** is at least 1, i.e., s and t are at distance (length of the shortest path) at least 1 from each other. Our goal is to 'separate' s and t, while minimizing objective $\sum_{(u,v)\in E} c(u,v)y(u,v)$.

Consider an arbitrary s-t **cut** $(A, V\text{-}A)$, where we set $y(u, v)=1$ only for edges (u, v) from A to $V\text{-}A$ and set $y(u, v)=0$ for all the rest edges. Such a setting of $y(u, v)$ is a **feasible solution** to the **Dual LP**. And the objective $\sum_{(u,v)\in E} c(u,v)y(u,v)$ is the **capacity** of the **cut** $(A, V\text{-}A)$, i.e., $c(A)$.



Thus, <u>the aforementioned **Dual LP** is clearly the **LP Relaxation** of the **Min-Cut Problem**</u>.

(**Lemma 1**) For <u>every feasible **cut**</u> (*A*, *V-A*) in the **flow network**, there is a **feasible solution** to the **Dual LP** whose <u>cost is the same as the **capacity** of *A*</u>, i.e., *c*(*A*).



(**Lemma 2**) Given any **feasible solution** *y*(*u*, *v*) for each $(u,v) \in E$, it's possible to find a cut (*A*, *V-A*) such that $c(A) \le \sum_{(u,v)\in E} c(u,v)y(u,v)$.

Proof of **Lemma 2**. (**Second Proof** of the **Max-Flow Min-Cut Theorem**)

Define *d*(*v*) as the **distance** from *s* to *v* according to the **weight** *y*(*u*, *v*) (i.e., *d*(*v*)=min $\{\sum_{(u,v)\in p_s^v} y(u,v)\}$ with p_s^v as a possible path from *s* to *v*), so *d*(*v*) is the **shortest length** over all paths from *s* to *v*. <u>The **constraints** of the **Dual LP** indicates that *d*(*t*)≥1</u> (w.r.t. sink vertex *t*).

Given the values of *y*(*u*, *v*), we can construct a (good) *s*-*t* cut (*A*, *V-A*) as follow. Pick a **threshold** *T* uniformly at random in the interval [0, 1) and let *A* be the set of vertices such that *A*={*v*: *d*(*v*)≤*T*}.

First, we aim to show that the **expected capacity** of the constructed cut (*A*, *V-A*) is <u>less than or equal</u> to the **capacity** of the **fractional cut** (w.r.t. the **Dual LP**), i.e., $\mathrm{E}[c(A)] \le \sum_{(u,v)\in E} c(u,v)y(u,v)$.

For E[*c*(*A*)], we have

$$\mathrm{E}[c(A)] = \mathrm{E}[\sum_{(u,v)\in E} c(u,v)x(u,v)],$$

where the expectation is computed assuming *T* is uniformly chosen at random in [0, 1) and *x*(*u*, *v*) is an auxiliary **random variable** with the following definition:

$$x(u,v) = \begin{cases} 1, u \in A, v \notin A \\ 0, \quad \text{otherwise} \end{cases}.$$

By the **Linearity** of expectation, we have

$$\mathrm{E}[c(A)] = \sum_{(u,v)\in E} \mathrm{E}[x(u,v)] \cdot c(u,v),$$

in which we also have

$$\mathrm{E}[x(u,v)] = P[u \in A, v \notin A] = P[d(u) \le T < d(v)] \cdot$$

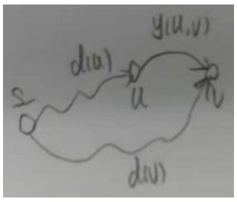
When (1) <u>$d(v)>d(u)$</u>, we have

$$P[d(u) \le T < d(v)] \le (d(v) - d(u))/1 \cdot$$

Moreover, when (2) <u>$d(v)<d(u)$</u>, we have

$$P[d(u) \le T < d(v)] = 0 \cdot$$

Also, for an arbitrary edge (u, v), by the **triangle inequality** (as we consider the length of **shortest path** according to $y(u, v)$), we have



$$d(v) \le d(u) + y(u,v),$$

which holds in case (1) and case (2). It further indicates that

$$d(v) - d(u) \le y(u,v) \cdot$$

Thus, we have

$$\begin{aligned}
\mathrm{E}[c(A)] &= \sum\nolimits_{(u,v)\in E} \mathrm{E}[x(u,v)] \cdot c(u,v) \\
&= \sum\nolimits_{(u,v)\in E} P[d(u) \le T < d(v)] \cdot c(u,v) \\
&\le \sum\nolimits_{(u,v)\in E} (d(v) - d(u)) \cdot c(u,v) \\
&\le \sum\nolimits_{(u,v)\in E} y(u,v) \cdot c(u,v)
\end{aligned}.$$

Since $\mathrm{E}[c(A)] \le \sum\nolimits_{(u,v)\in E} y(u,v) \cdot c(u,v)$, there must be an *s*-*t* cut (A, V-A) constructed via the aforementioned strategy with $c(A) \le \sum\nolimits_{(u,v)\in E} y(u,v) \cdot c(u,v)$. Namely, for some choices of the threshold T, the resulting cut (A, V-A) must satisfy $c(A) \le \sum\nolimits_{(u,v)\in E} y(u,v) \cdot c(u,v) \cdot$