Tutorial 8: Applications of Linear Programming

Problem 1. (Densest Subgraph) Consider a graph $G = (V, E)$. For a non-empty subset of vertices $S \subseteq V$, let $E(S) = \{(u, v) \in E : u, v \in S\}$ be the set of edges within S. Define the density of S as $d(S) = |E(S)|/|S|$. The goal is to find a non-empty subset S of maximum density. To solve this problem, consider the following LP for the problem.

$$
\begin{array}{lll}
\text{LP variables:} & x_u \text{ for every vertex } u \text{ and } y_e \text{ for every edge } e. & \\
\textbf{maximize} & \displaystyle\sum_{e \in E} y_e & \\
\text{s.t.} & & \\
& \displaystyle\sum_{u \in V} x_u \leq 1 & \\
& y_{(u,v)} \leq x_u & \text{for every } u, v \in V \\
& x_u, y_e \geq 0 & \text{for every } u \in V \text{ and } v \in V
\end{array}
$$

Let OPT be the optimal value of the problem and LP be the value of this linear program.

1. Prove that $LP \geq OPT$. To this end, consider an optimal solution S^* of value OPT and define a corresponding LP solution of the same value.
2. Define set $S_t = \{u : x_u \geq t\}$. Let $T = \max_{u \in V} x_u$.
3. Choose t uniformly at random from $[0, T]$. What is the probability that edge (u, v) belongs to $E(S_t)$.
4. What is the expected number of vertices in S_t?
5. What is the expected number of edges in $E(S_t)$?
6. Prove that $\mathbf{E}_t[|E(S_t)| - LP \cdot |S_t|] \geq 0$.
7. Propose an LP-based algorithm for solving the Densest Subgraph Problem.

Solution

1. For each $u \in V$ and $(u, v) \in E$, set

$$
x_u = \begin{cases} 1/|S^*| & \text{if } u \in S^*, \\ 0 & \text{otherwise;} \end{cases} \qquad y_{(u,v)} = \begin{cases} 1/|S^*| & \text{if both } u, v \in S^*, \\ 0 & \text{otherwise.} \end{cases}
$$

We need to show that the variables x_u and $y_{(u,v)}$ make up a feasible solution whose value is S^*. For the latter, notice that

$$
\sum_{e \in E} y_e = \sum_{(u,v) \in E \ \& \ u,v \in S^*} y_{(u,v)} = \frac{|E(S^*)|}{|S^*|} = d(S^*).
$$

Regarding feasibility, notice that we give non-negative values to the variables, and $y_{(u,v)} > 0$ means that $u \in S$, hence $x_u = y_{(u,v)}$. Finally, we have that

$$\sum_{u \in V} x_u = \sum_{u \in S^*} 1/|S^*| = 1.$$

3. The probability that the edge (u, v) belongs to $E(S_t)$ is the probability that both $x_u \geq t$ and $x_v \geq t$, which is equal to $\min\{x_u, x_v\}/T$.

4. To compute the expected number $\mathbf{E}[|S_t|]$ of vertices in S_t, consider the random variables X_u for every $u \in V$, defined as

$$X_u = \begin{cases} 1 & \text{if } x_u \geq t, \\ 0 & \text{otherwise.} \end{cases}$$

We have that $\mathbf{E}[X_u] = P[x_u \geq t] = x_u/T$, and hence

$$\mathbf{E}[|S_t|] = \mathbf{E}\left[\sum_{u \in V} X_u\right] = \sum_{u \in V} \mathbf{E}[X_u] = \frac{1}{T} \sum_{u \in V} x_u.$$

5. Similarly, to compute the expected number $\mathbf{E}[|E(S_t)|]$ of edges in S_t, consider the random variables $Y_{(u,v)}$ for every $(u, v) \in E$, defined as

$$Y_{(u,v)} = \begin{cases} 1 & \text{if } x_u, x_v \geq t, \\ 0 & \text{otherwise.} \end{cases}$$

We have $\mathbf{E}[Y_{(u,v)}] = P[x_u, x_v \geq t] = \min\{x_u, x_v\}/T$, and

$$\begin{aligned} \mathbf{E}\left[|E(S_t)|\right] &= \mathbf{E}\left[\sum_{(u,v) \in E} Y_{(u,v)}\right] = \sum_{(u,v) \in E} \mathbf{E}[Y_{(u,v)}] \\ &= \frac{1}{T} \sum_{(u,v) \in E} \min\{x_u, x_v\} \geq \frac{1}{T} \sum_{(u,v) \in E} y_{(u,v)} = \frac{LP}{T}. \end{aligned}$$

6. We have that

$$LP \cdot \mathbf{E}[|S_t|] = \frac{LP}{T} \sum_{u \in V} x_u \leq \frac{LP}{T} \leq \mathbf{E}[|E(S_t)|],$$

therefore

$$\mathbf{E}[|E(S_t)| - LP \cdot |S_t|] \geq 0.$$

7. Since $\mathbf{E}[|E(S_t)| - LP \cdot |S_t|] \geq 0$, there must be a t^* for which $|E(S_{t^*})| - LP \cdot |S_{t^*}| \geq 0$. For that t^*,

$$d(S_{t^*}) = \frac{|E(S_{t^*})|}{|S_{t^*}|} \geq LP, \tag{1}$$

and therefore $OPT \geq LP$. We already saw that $OPT \leq LP$, thus $OPT = LP$. That is, running our linear program, the value we get is an optimal value for our problem. Moreover, (1) guarantees that such a value will be achieved by some S_{t^*}. How to find such a $t^* \in [0, T]$? Well, there is only a linear number of values we have to try, namely, the values x_u for every $u \in V$. So we can try them all: After our linear program returns with some solution, we can check in linear time whether $d(S_t) = LP$ for some $t = x_u$.