CSCI 270 Lecture 22: Project Selection

Project Selection

We have a set of n projects $P = \{p_1, ..., p_n\}$ with values $v_1, v_2, ..., v_n$ (the values may be negative).

Project i has a set of prerequisites $S_i \subseteq P$.

If we decide to do project i, we get value v_i, but we must also do all projects in S_i (and some of them may have negative values).

Find max-valued subset of projects to do.

$v_1 = 2, S_1 = P_2$
$v_2 = -3, S_2 = P_4$
$v_3 = 5, S_3 = P_2, P_4$
$v_4 = -3$
$v_5 = 2, S_5 = P_4, P_6$
$v_6 = -3$

- What is the total value if we do all projects?
- What is the total value if we do no projects?
- Is there a better subset of projects?
- We need to turn this into a graph. What should the nodes be?
- How should we encode the prerequisites?
- Would Bipartite Matching be a good problem to reduce to?
- Is there any problem with network flow?
- Would Circulations with Lower Bounds solve this problem?
- Are there any problems left?
- How do we make sure that we don't select a project without its prerequisites?
- Which set specifies the projects we do and which we don't do?
- We need to add edges from s to certain projects. If we split one of these edges, what does that say about the project?
- We need to add edges from certain projects to t. If we split one of these edges, what does that say about the project?
- We need to add the appropriate penalties to completing a project with negative value, and appropriate incentives to completing a project with positive value. How should we go about doing this?

Proof of correctness:

For every project with negative value in S, we increase the value of the cut:

$\sum_{i \in S, p_i < 0} -p_i$

We find a min-cut $(S, V - S)$. For every project with positive value in $V - S$, we increase the value of the cut:

$\sum_{i \in V-S, p_i > 0} p_i$

Therefore the value of the cut we find is

$\sum_{i \in V-S, p_i > 0} p_i + \sum_{i \in S, p_i < 0} -p_i$

Let $C = \sum_{i : p_i > 0} p_i$: the sum of capacities of edges outgoing from the source. Then:

$\sum_{i \in V-S, p_i > 0} p_i = C - \sum_{i \in S, p_i > 0} p_i$

Then the total value of our cut is the sum of these two values:

$C - \sum_{i \in S, p_i > 0} p_i + \sum_{i \in S, p_i < 0} -p_i = C - \sum_{i \in S} p_i.$

Since $\sum_{i \in S} p_i$ is exactly our profit, if we minimize $C - \sum_{i \in S} p_i$ then we maximize our profit. Thus, finding min-cut here really works!