1. Read section 7.5 of DPV and prove the equation at the bottom of page 226. The book DPV is available online at http://www.cs.berkeley.edu/ vazirani/algorithms.html

2. Consider the problem of writing an antivirus program that seeks to detect \( n \) different viruses. From an analysis of these viruses, you have found \( m \) code fragments that each appear in one or more viruses. For each \( i \in [m] \), say that fragment \( i \) appears in viruses \( S_i \) for some subset \( S_i \subseteq [n] \). However, since each fragment also may appear in legitimate code (creating false positives), we assign a cost \( c_i \geq 0 \) to each fragment.

Your goal is to choose a minimum-cost valid collection of code fragments \( T \) to search for. The cost of a collection \( T \) is defined to be \( \sum_{i \in T} c_i \). A collection \( T \) is valid if it can identify all \( n \) viruses; i.e. if \( \bigcup_{i \in T} S_i = [n] \). Let \( \text{OPT} \) denote the minimum cost of any valid collection of code fragments.

(a) Consider the following optimization problem:

\[
\min \sum_{i=1}^{m} x_i c_i \quad (1a)
\]

\[
x_1, \ldots, x_m \in \{0, 1\} \quad (1b)
\]

\[
\forall j \in [n], \sum_{i : j \in S_i} x_i \geq 1 \quad (1c)
\]

Prove that the solution to (1) is equal to \( \text{OPT} \).

(b)

\[
\min \sum_{i=1}^{m} x_i c_i \quad (2a)
\]

\[
\forall i \in [m], 0 \leq x_i \leq 1 \quad (2b)
\]

\[
\forall j \in [n], \sum_{i \in S_i \ni j} x_i \geq 1 \quad (2c)
\]

Denote the solution to (2) by \( \text{OPT}_{LP} \). Observe that \( \text{OPT}_{LP} \leq \text{OPT} \). Can any of the constraints in (2) be removed without changing the answer?

(c) Let \( x \in \mathbb{R}^n \) be a solution to (2). Suppose that each element of \([n]\) appears in at most \( f \) subsets. Choose \( T = \{i : x_i \geq 1/f\} \). Prove that \( T \) is a valid collection with cost at most equal to \( f \cdot \text{OPT} \).

(d) Write down the dual of (2).

(e) Again starting with a solution of (2), suppose that we take \( T = \{i : x_i > 0\} \). Prove that the cost of \( T \) is again \( \leq f \cdot \text{OPT} \). Hint: Use complementary slackness.

(f) Extra credit. Consider the following alternate strategy for constructing \( T \). For each \( i \), put \( i \) in \( T \) with probability \( x_i \).

i. What is the expected cost of this strategy?

ii. This strategy will generally not yield a valid collection. Prove that each \( j \in [n] \) is covered with probability \( \geq 1 - 1/e \). Hint: use convexity.
iii. Suppose we repeat this strategy \( \ln(n) \) times and take the union of all of the resulting collections. Prove that with constant probability this yields a valid collection that is \( \leq \OPT \cdot 2 \ln(n) \).

3. Given a directed graph \((V, E)\) with edge capacities \(c\) and vertices \(s, t \in V\), define \(\text{FRAC-MIN-CUT}\) to be the value of the following LP:

\[
\min \sum_{e \in E} c(e)h(e) \quad (3a)
\]

\[
h(v \rightarrow w) \geq 0 \forall (v, w) \in E \quad (3b)
\]

\[
h(v \rightarrow w) \geq g(v) - g(w) \forall (v, w) \in E \quad (3c)
\]

\[
g(s) = 1 \quad (3d)
\]

\[
g(t) = 0 \quad (3e)
\]

(a) Let \(\text{MIN-CUT}\) denote the minimum cost of any \(s-t\) cut. Prove that \(\text{FRAC-MIN-CUT}\) is equal to \(\text{MIN-CUT}\). \textit{Hint: You may use results from lecture such as LP strong duality, and the max-flow/min-cut theorem, without rederiving them.}

(b) Given a solution to (3), choose a random \(\theta \in [0, 1]\) and set \(A = \{v : g(v) \geq \theta\}\) and \(B = \{v : g(v) < \theta\}\). What is the expected (i.e. average) value of \(\|A, B\|\)? \textit{Hint: The only fact about probability that you need to know is linearity of expectation, meaning that the expectation of a sum of random variables is equal to the expectation of the sum.}

(c) Show that any choice of \(\theta \in (0, 1)\) yields a minimum cut.