College of Computer & Information Science
Northeastern University
CS7800: Advanced Algorithms

Fall 2010
Handout 11
29 October 2010

Problem Set 4 (due Monday, November 8)

1. (5 + 2 + 3 = 10 points) Pathalogical example for the basic augmenting path algorithm

In the residual network G_f corresponding to a flow f, define an *augmenting walk* as a directed walk from source s to sink t that visits any arc at most once (it might visit nodes multiple times– in particular, an augmenting walk might visit nodes s and t multiple times). Define the *residual capacity* of an augmenting walk to be the residual capacity (i.e., capacity in G_f) of the minimum-capacity edge in the walk.

(a) Consider the network shown in Figure 1(a) with the edges labeled a, b, c, and d; note that one arc capacity is irrational. Show that this network contains an infinite sequence of augmenting walks whose residual capacities sum to the maximum flow value. (*Hint* : Each augmenting walk of the sequence contains exactly two arcs from node s to node t with finite residual capacities.)

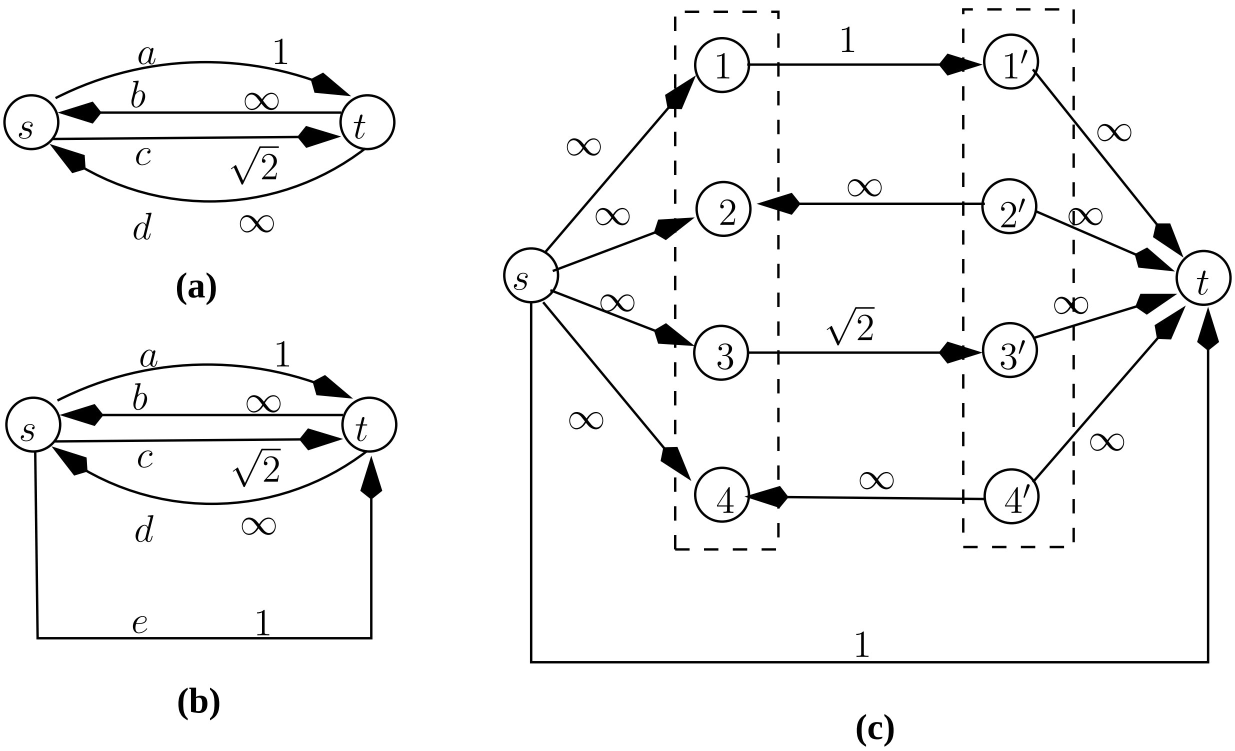

Figure 1: A pathological example for the basic augmenting path algorithm.

(b) Now consider the network shown in Figure 1(b). Show that this network contains an infinite sequence of augmenting walks whose residual capacities sum to a value different than the maximum flow value.

(c) Next consider the network shown in Figure 1(c); in addition to the edges shown, the network contains an infinite capacity edge connecting each node pair in the set $\{1, 2, 3, 4\}$ as well as each node pair in the set $\{1', 2', 3', 4'\}$. Show that each augmenting walk in the solution of part (b) corresponds to an augmenting path in Figure 1(c). Conclude the basic augmenting path algorithm, when applied to a maximum flow problem with irrational capacities, might perform an infinite sequence of augmentations and the terminal flow value might be different than the maximum flow value.

2. (10 points) Covering the 2012 London Summer Olympics

A major sports broadcaster wants to send a group of reporters to cover a set of sporting events in the London Summer Olympics. The sports events will be held in several stadiums and venues in and around the London area. We know the starting time and ending time of each event, and the location where it is held. We are also given the travel time between different locations. Design a polynomial-time algorithm to determine the least number of reporters needed to cover these events. Justify its correctness and analyze its running time.

(*Hint:* Given a positive integer k, determine whether k reporters suffice by solving an appropriate network flow problem.)

3. (10 points) Tight-knit communities in social networks

Social networks are all the craze these days. Assuming that friendship links are symmetric, one can represent a social network such as Facebook as an undirected unweighted graph G. Identifying clusters, communities, and other connectivity properties is a hot topic of research.

Suppose you want to identify a tight-knit community in a social network G. For a given set S of nodes in G, let $n(S)$ denote the number of edges with both endpoints in S. Define the *bonding* of any set S of nodes in G to be $n(S)/|S|$. Larger the $n(S)$, the more tight-knit $n(S)$ is; dividing by $|S|$ normalizes across sets of different sizes.

(a) Give a polynomial time algorithm that takes a rational number α and determines whether there exists a set S with bonding at least α.

(b) Use the above algorithm to find a set S of nodes with maximum bonding.

4. (10 points) Single-variable Linear Program

Exercise 29.5-9 of text.

5. (3 + 7 = 10 points) Markov chains and LP duality

A Markov chain is a set of states (of a system) and the probability, for each pair of states, of transitioning from one state to the other. Markov chains have a wide range of applications including as models for several physical or biological processes, in economics and the social sciences, web search, and statistics.

Formally, a Markov chain is an $n \times n$ matrix P where p_{ij} denotes the probability of transitioning from state i to state j; if the system is in state i at time t, then the probability that it is in state j at time $t + 1$ is p_{ij}. Clearly, P satisfies the property that the sum of the entries along any row is 1.

One nice, and very useful, property of Markov chains is the existence of a stationary distribution π over the state space, where π is an $n \times 1$ vector with π_i being the probability of the system being in state i. We say that π is stationary if the following holds: if at any time t, π gives the probability distribution for the state of the given system, then π is also the probability distribution for the state in time $t + 1$.

For any Markov chain matrix, the existence of a stationary distribution π can be shown easily using LP duality, as we establish in this exercise.

(a) Write an LP (over variables π_i) whose constraints define the stationary distribution. Set $\sum_i \pi$ as the objective function.

(b) Derive the dual for the above LP. Analyze the dual and argue that the primal is always feasible.