College of Computer & Information Science — Fall 2010
Northeastern University — Handout 16
CS7800: Advanced Algorithms — 24 November 2010

Problem Set 5 (due Friday, December 10)

1. ($5 \times 6 = 30$ points) NP-completeness and approximation algorithms

Let G be an undirected graph with k start nodes s_1 through s_k and k end nodes t_1 through t_k.

(a) Give a reduction from 3-SAT to show that it is NP-hard to determine whether there exist k paths, the ith path from s_i to t_i, $1 \leq i \leq k$, such that no two paths share an edge.

We next consider an optimization version of the above problem. Here, we allow paths to share edges, and define the *load* $\ell(e)$ on an edge e to be the number of s_i-t_i paths that use e. The optimization problem then is to determine a set of s_i-t_i paths that minimizes $\max_e \ell(e)$.

(b) Write an integer linear program for the above problem. (*Hint:* You can view each s_i-t_i path as a flow of unit 1 from s_i to t_i.)

(c) Develop a randomized rounding algorithm which proceeds as follows: Relax the integrality constraint, and solve the LP; Decompose each s_i-t_i flow into a set of paths (we have seen this in class); Use randomized rounding to select a path. Fill in the details.

The next step is to show that the above rounding algorithm will achieve a load within an $O(\log n)$-factor of the optimal with high probability, say at least $1 - 1/n$, where n is the number of nodes in the graph.

(d) Show that the expected load of an edge e is equal to the total flow on the edge e in the LP solution.

(e) Using Chernoff bounds (described below) show that with probability at least $1 - 1/n$, the load of the path collection is within an $O(\log n)$ factor of the optimal achievable load.

Chernoff bound: Let $X_1, X_2, \ldots, X_n$ be n independent random variables each taking a value of 0 or 1. Let X denote $\sum_i X_i$. Then, for any $\delta > 0$, we have the following.

$$\Pr\left[X > (1+\delta)E[X]\right] \leq \left(\frac{e^{\delta}}{(1+\delta)^{1+\delta}}\right)^{E[X]}.$$

(*Hint:* For each edge, use the Chernoff bound to place a very small upper bound on the probability that the load of the edge exceeds $O(\log n)$ times the expectation. Then use a union bound to argue that with high probability, the load on every edge is within an $O(\log n)$ factor of the optimal.)

2. (10 points) Random walk in a complex maze

Luke Skywalker and Darth Vader are located on two distinct nodes in a large complex intergalactic maze – represented by an undirected non-bipartite graph G with n nodes and m edges. Each has no other choice but to navigate the maze via a random walk (after all, this is a maze). So, in each step, both move to a neighboring node chosen uniformly at random. Show that they will meet in expected $O(m^2 n)$ time.

3. (20 points) Toeplitz matrices

Problem 30-2.

4. (50 points) Exploratory problem

Do any *one* of the following. You may work in pairs. Please submit a report that includes theoretical analysis, description of programs you have written, and experiments you have developed; also provide a link to any code.

(A) Minimum bottleneck spanning tree of a random graph

Let G be a complete undirected graph on n vertices; that is, there is an edge between every pair of vertices in G. Suppose the weight of each edge is independently and uniformly distributed over the half-open interval $[0, 1)$.

The minimum bottleneck spanning tree of a graph is a spanning tree whose maximum-weight edge has least weight, among all spanning trees of the graph. Derive an estimate for the expected weight of the maximum-weight edge of a minimum bottleneck spanning tree of G. Make your estimate as asymptotically tight as you can. You may provide an answer to the above question in one of two forms (or a combination of both). One form is to provide an estimate and prove it formally. Another form is to perform a set of experiments and present the experimental results, along with some intuition to justify your estimate, derived from the experiments.

(B) Scheduling meetings

We face the following problem when scheduling student meetings with faculty during Open Houses. Similar problems also arise during conferences, and more generally, while managing schedules.

Let S denote a set of students and F denote a set of faculty members. Each student i in S is interested in meeting a set $F_i \subseteq F$ of faculty members. Assume that time is divided into a set T of slots, and each meeting involves exactly one faculty member and one student, and lasts for exactly one slot. Each student or faculty member i has a set A_i of slots where s/he can meet. Multiple meetings may be scheduled at the same time slot as long as no faculty member and no student has more than one meeting scheduled the same time slot.

Given the above – S, F, T, the F_i's, the A_i's – we would like to determine whether it is possible to schedule meetings so that all constraints (determined by the F_i's and A_i's) are met. If it possible, then we would like to compute the schedule. Otherwise, we may want to compute a schedule that maximizes the number of meetings that can be scheduled.

Can you give a polynomial-time algorithm for the above problem? If not, can you show that the problem is NP-hard? If the problem is NP-hard, can you suggest an approximation algorithm?

You may provide an answer to the above question in one of two forms (or a combination of both). One form is to give algorithm(s) and proofs that rigorously establish some claims (e.g., polynomial-time computability, NP-hardness, approximation). Another form is to develop a program for the problem, present sample runs that exercise the program on diverse inputs, and comment on the inputs where the program may fail (if any).