CSE P521: Applied Algorithms (Winter, 2017)

Homework 2 Out: Monday, 16-Jan **Due**: Thursday, 26-Jan (midnight in the Dropbox)

Instructions:

Your proofs and explanations should be clear, well-organized and as concise as possible.

You are allowed to discuss the problems with fellow students taking the class. However, you must write up your solutions completely on your own. Moreover, if you do discuss the problems with someone else, I am asking, on your honor, that you do not take any written material away from the discussion. In addition, for each problem on the homework, I ask that you acknowledge the people you discussed that problem with, if any.

Most of the problems require only one or two key ideas for their solution – spelling out these ideas should give you most of the credit for the problem even if you err in some finer details. So, make sure you clearly write down the main idea(s) behind your solution.

A final piece of advice: Begin work on the problem set early and don't wait until the deadline is only a few days away.

This homework is about balls and bins on steroids. We showed in class that if one throws N balls independently and uniformly into N bins then the maximum loaded bin has $\Theta\left(\frac{\log N}{\log\log N}\right)$ balls with high probability. We also argued that this follows from a multiplicative Chernoff bound.

Here we will see that the same reasoning can be applied in a significantly more complex situation.

Suppose we have a road network represented as a directed graph $D = (V, A)$ and there is a given sequence of city pairs $(s_1, t_1), (s_2, t_2), \dots, (s_k, t_k)$, where $s_1, \dots, s_k, t_1, \dots, t_k \in V$. (Cities could be repeated, and they could be both sources and destinations.)

You want to design an algorithm that decides on a **route** for your trucks that travel from s_i to t_i. A route is simply a directed path through the road network D. Your supply chain gets slowed down to the maximum latency on any edge of the network, and the latency on an edge is proportional to the # of routes that use it.

Formally: For each i, you choose a directed path P_i from s_i to t_i in D. For a given routing $(P_1, \dots, P_k)$, an edge $e \in A$, define the **latency on *e*** by to be the # of routes that use e:

$$\text{latency}_e(P_1, \dots, P_k) = \#\{\, i : e \in P_i \,\}$$

Define the maximum latency of the routing by:

$$\text{maxdelay}(P_1, \dots, P_k) = \max\,\{\, \text{latency}_e(P_1, \dots, P_k) : e \in A \,\}$$

The engineers tell you that finding a routing that minimizes maxdelay is an NP-hard problem. Instead, they are able to find you a "fragmented routing": a weighted collection of routes that could, in principle, feed the supply chains. This means that for every (s_i, t_i) pair, they can give you many paths with positive weights:

$$(Q_{i1}, p_{i1}), (Q_{i2}, p_{i2}), \dots, (Q_{im}, p_{im})$$

such that each Q_{ij} is an s_i-t_i path and $p_{i1} + p_{i2} + \cdots + p_{im} = 1$ (so that if you route a p_{ij}-fraction of the $s_i \to t_i$ trucks down path Q_{ij}, then you would still have the right rate of trucks arriving at the destination). The engineers tell you that the fractional maxdelay is small:

$$\max\left\{\sum_{i=1}^{k} \sum_{Q_{ij}: e \in Q_{ij}} p_{ij} : e \in A\right\} \leq L$$

They are very excited.

Problematically, laws forbid you from using multiple routes, as your routes must be published and approved beforehand by an outdated legal body. Engineers sometimes forget about external considerations. So now your goal is to choose, for each $i = 1, \dots, k$, one of the m routes $Q_{i1}, \dots, Q_{im}$. You are hoping you can do this with maxdelay that is not too much worse than L.

Goal: Show that there is a way to choose routes $P_1, \dots, P_k$ so that the maxdelay is at most $O\left(\frac{\log N}{\log\log N}\right) L$, where N is the total # of edges in the network. In fact, you will give a randomized algorithm that succeeds in finding such routes with high probability.

1. Show how you could implement "balls and bins" in this model.

Design a network D so that balls correspond to (s_i, t_i) pairs and the bins correspond to edges in D. There should be a fragmented routing with maxlatency 1, and the balls-in-bins analysis from class should show that there is a non-fragmented routing with maxdelay $O\left(\frac{\log k}{\log\log k}\right)$.

Extra credit:

Can you find a network D where there is a fragmented routing with fractional maxdelay 1, but **every** non-fragmented routing has large maxdelay? How large can you make the gap?

2. A rounding algorithm

Here is a simple idea: For every $i = 1, \dots, k$ choose a path P_i randomly by choosing path Q_{ij} with probability p_{ij}. Specify such an algorithm, being careful about dependence between the different random choices.

3. Expected latency

Note that the s_i-t_i route $\boldsymbol{P}_i$ chosen by your algorithm is a random variable.

For a fixed edge e, calculate the expected latency on e:

$$\mathbb{E}[\text{latency}_e(\boldsymbol{P}_1, \dots, \boldsymbol{P}_k)]$$

You should be able to derive an exact expression in terms of the fragemented routing and the values p_{ij}. It will be easiest to do this by writing $\text{latency}_e(\boldsymbol{P}_1, \dots, \boldsymbol{P}_k)$ as a sum of $\{0,1\}$ random variables and using linearity of expectation.

Show how to conclude that $\mathbb{E}[\text{latency}_e(\boldsymbol{P}_1, \dots, \boldsymbol{P}_k)] \leq L$ for every edge e.

4. Bounding the maximum delay

Simply knowing the expected latency for every edge is not so great. We want a bound on $\mathrm{maxdelay}(\boldsymbol{P}_1, \dots, \boldsymbol{P}_k)$. You may assume for this part that $L \geq 1$.

Use the multiplicative Chernoff bound from lecture to show that for a fixed edge e,

$$\Pr[\mathrm{latency}_e(\boldsymbol{P}_1, \dots, \boldsymbol{P}_k) \geq \beta L] \leq \frac{\mathrm{e}^{\beta-1}}{\beta^\beta}$$

(Be careful about the random variables you are applying the Chernoff bound to, and why it's OK to do so! Also be aware of one notational hazard: the symbol e on the right-hand side is the base of the natural logarithm $\mathrm{e} = 2.718 \ldots$ The symbol e on the left-hand side is an edge of D.)

Now show how to use a union bound to conclude that for some constant C,

$$\Pr\left[\mathrm{maxdelay}(P_1, \dots, P_k) \geq \frac{C \log N}{\log \log N} L \right] \leq \frac{1}{N}$$