EPFL

Exercise Set VI, Advanced Algorithms 2022

These exercises are for your own benefit. Feel free to collaborate and share your answers with other students. Solve as many problems as you can and ask for help if you get stuck for too long. Problems marked * are more difficult but also more fun :).

These problems are taken from various sources at EPFL and on the Internet, too numerous to cite individually.

1 In class we saw that Karger's min-cut algorithm implies that an undirected graph has at most $\binom{n}{2}$ minimum cuts. Show that this result is tight by giving a graph with n vertices and $\binom{n}{2}$ minimum cuts.

2 Change Karger's algorithm so that it also works for edge-weighted graphs. Also adapt the analysis to prove that it still returns any min cut $(S^*, \overline{S^*})$ with probability at least $1/\binom{n}{2}$. (Hence, edge-weighted graphs also have at most $\binom{n}{2}$ min cuts.)

3 In this exercise, we are going to analyze a beautiful and simple randomized algorithm for the 2-SAT problem (due to Papadimitriou'91). Recall that in the 2-SAT problem we have n variables $x_1, x_2, \ldots, x_n$ and m clauses $C_1, \ldots, C_m$, where each clause is the disjunction of two literals. Examples of possible clauses are $x_i \vee x_j$, $x_k \vee \neg x_\ell$ and so on. The goal is to find a truth assignment to the n variables so that all clauses are satisfied (if one exists).

The algorithm we are going to analyze is ingenious and simple:

1. Select uniformly at random a truth assignment π.

2. Repeat the following for at most $O(n^2)$ steps (or until we find a satisfying assignment):

   Select an unsatisfied clause C_ℓ with variables x_i, x_j.

   Select one of the variables x_i and x_j with equal probability.

   Update π by flipping the truth assignment of the selected variable, i.e., set it to true if it was false and vice versa.

In the following, we assume for simplicity that there is a satisfying truth assignment π^* (where π_i^* denotes the Boolean value of variable x_i). (If there is no such truth assignment we cannot find one so there is nothing to prove.)

For a truth-assignment π, define $\mathrm{dist}(\pi^*, \pi)$ as the number of variables/coordinates where $\pi_i \neq \pi_i^*$ (this is called the *Hamming distance*). Let $\pi = \pi^1, \pi^2, \ldots$ be the truth-assignments generated by the algorithm and let $d_i = \mathrm{dist}(\pi^*, \pi^i)$. Note that as the algorithm only changes one variable in the truth assignment in each iteration, we have $d_{i+1} = d_i + \Delta_i$ where $\Delta_i \in \{-1, 1\}$. Prove that for any i

$$\Pr[\Delta_i = -1] \geq 1/2.$$

Once this is established, we can view the algorithm as a random walk on $\{0, 1, 2, \ldots, n\}$ that always goes to the left (to a smaller number) with probability at least $1/2$. It is known that such a random walk will reach 0 with high probability after $O(n^2)$ many steps. (Show this if you feel that you are up for a challenge.)

4 *(Basic Hashing)* Consider a Hash Family $\mathcal{H}$ where each $h \in \mathcal{H}$ is a function $h : U \to [n]$ that maps the elements U to integers $\{0, 1, \ldots, n - 1\}$. We assume throughout this exercise that $|U| = n$ and so we are in the classic balls-and-bin setting (with n balls and n bins). Assume that $\mathcal{H}$ is a pairwise independent hash family, i.e., it satisfies the following:

1. $\Pr_{h \in \mathcal{H}}[h(x) = y \wedge h(x') = y'] = \frac{1}{n^2}$ for all $x \neq x' \in U$ and $y, y' \in [n]$.

4a Let Y be the number of items that hash to value 1, i.e., $Y = |\{x \in U : h(x) = 1\}|$. Prove that $\mathbb{E}_{h \in \mathcal{H}}[Y] = |U|/n = 1$ and $\text{Var}[Y] \leq 1$.

4b Use the solution to the previous subproblem to prove that

$$\Pr[Y \geq 2\sqrt{n} + 1] \leq \frac{1}{4n}.$$

Note that this implies (using the union bound over the n hash values) that no hash value (i.e., no bin) will receive more than $2\sqrt{n}$ keys with probability at least $3/4$. In the lecture, we obtained the much better bound of $O(\log n / \log \log n)$ when assuming complete independence of the hash family. Here, we only assume pairwise independence which still leads to a good (but worse) bound.

Hint: For the proof use Chebychev's Inequality: Let Y be a random variable with expectation μ and variance σ^2. Then for any real number $k > 0$,

$$\Pr[|Y - \mu| \geq k\sigma] \leq \frac{1}{k^2}.$$