CONCORDIA UNIVERSITY

DEPARTMENT OF COMPUTER SCIENCE AND SOFTWARE ENGINEERING

COMP 6651/4: Algorithm Design Techniques

Fall 2013

Final Exam - Close book exam - 3 hours

Instructor: Professor B. Jaumard
Office: EV 003.189
Email: bjaumard@cse.concordia.ca

First Name	**Last Name**	**ID#**

Question 1. (20 points)

A nucleic acid sequence is a succession of letters that indicate the order of nucleotides within a DNA (using GACT) or RNA (GACU) molecule. *Tandem Repeats* occur in DNA when a pattern of one or more nucleotides is repeated, and the repetitions are directly adjacent to each other. For example, consider the sequence:

ATTCGATTCGATTCG

This contains 9 Tandem Repeats:

ATTCGATTCGATTCG
ATTCGATTCGATTCG
ATTCGATTCGATTCG
ATTCGATTCGATTCG
ATTCGATTCGATTCG
ATTCGATTCGATTCG
ATTCGATTCGATTCG
ATTCGATTCGATTCG
ATTCGATTCGATTCG

(a) Propose an algorithm to answer the following question: Given a nucleotide sequence, how many Tandem Repeats occur in it?

(b) What is the complexity of your algorithm?

Examples:

DNA	# of Tandem Repeats
AGGA	1
AGAG	1
ATTCGATTCGATTCG	9

Question 2. **(20 points)**

Suppose we are given a set of n lines in the plane, where none of the lines passes through the origin $(0, 0)$ and at most two lines intersect at any point. These lines divide the plane into several convex polygonal regions, or cells.

(a) Describe an efficient algorithm to compute the cell containing the origin. The output should be a doubly-linked list of the cells vertices.

(b) Analyze the complexity of your algorithm.

Hint. There are literally dozens of solutions. One solution is to reduce this problem to the convex hull problem. Every other solution looks like a convex hull algorithm.

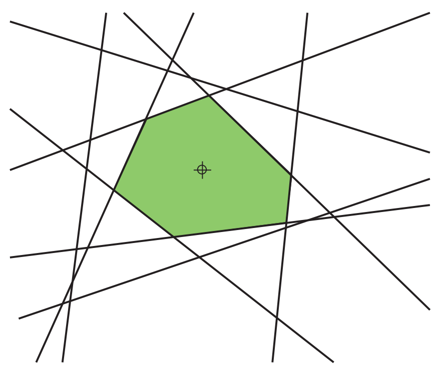


Figure 1: The cell containing the origin in an arrangement of lines.

Question 3. (20 points)

Suppose you are a consultant for the Port Authority of Singapore. As the busiest container port in the world, their revenue is constrained by the rate at which they can unload ships in the port. A ship arrives with n containers of weight $w_1, w_2, \ldots, w_n$. Standing at the dock is a set of trucks, each of which can load K units of weight. Assume that K and w_i are integers. You can stock multiple containers in each truck, subject to the weight restriction K; the goal is to minimize the number of trucks required to carry all containers. This problem is know to be NP-complete.

A greedy algorithm for this problem is the following. Start with an empty truck, and pile containers $1, 2, 3, \ldots$ into it until you get to a container that would overflow the weight limit. Now declare this truck "loaded" and send it off; then continue the process with a fresh truck. This algorithm, by considering trucks one at a time, may not achieve the most efficient way to pack the full set of containers into an available collection of trucks.

(a) Give an example of a set of weights, and a value of K, such that our algorithm does not use the minimum possible number of trucks.

(b) Show that, however, the number of trucks used by our algorithm is within a factor of 2 of the minimum possible number, for any set of weights and any value of K.

Question 4. (20 points)

(a) Recall the definition of Class P, NP, and co-NP.

(b) Recall the definition of an NP-Hard problem

(c) Recall the definition of an NP-Complete problem

(d) Recall the definition of the Dominating Set problem: decision and optimization versions

(e) Prove that the Dominating Set problem is NP-Complete

Question 5. (20 points)

Percolation. Given a composite systems comprised of randomly distributed insulating and metallic materials: what fraction of the materials need to be metallic so that the composite system is an electrical conductor? Given a porous landscape with water on the surface (or oil below), under what conditions will the water be able to drain through to the bottom (or the oil to gush through to the surface)? Scientists have defined an abstract process known as percolation to model such situations.

The model. We model a percolation system using an $N \times N$ grid of sites. Each site is either open or blocked. A full site is an open site that can be connected to an open site in the top row via a chain of neighbouring (left, right, up, down) open sites. We say the system percolates if there is a full site in the bottom row. In other words, a system percolates if we fill all open sites connected to the top row and that process fills some open site on the bottom row. (For the insulating/metallic materials example, the open sites correspond to metallic materials, so that a system that percolates has a metallic path from top to bottom, with full sites conducting. For the porous substance example, the open sites correspond to empty space through which water might flow, so that a system that percolates lets water fill open sites, flowing from top to bottom.)

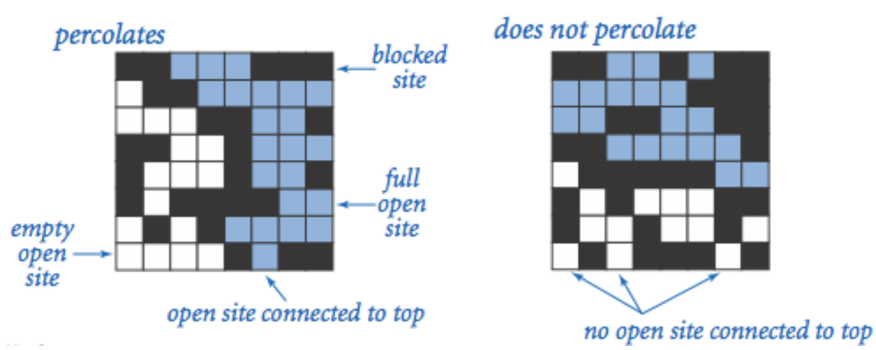
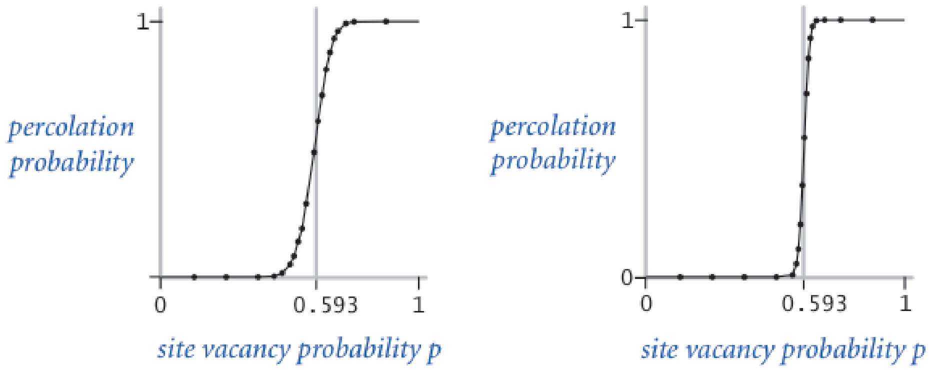


Figure 2: 8×8 grid of sites.

The problem. Researchers are interested in the following question: if sites are independently set to be open with probability p (and therefore blocked with probability $1-p$), what is the probability that the system percolates? When p equals 0, the system does not percolate; when p equals 1,



(a) 20-by-20 random grid (b) 100-by-100 random grid

Figure 3: Site vacancy probability p versus the percolation probability

the system percolates. The plots below show the site vacancy probability p versus the percolation probability for 20-by-20 random grid (3(a)) and 100-by-100 random grid (3(b)).

When N is sufficiently large, there is a threshold value $p^\star$ such that when $p < p^\star$ a random $N \times N$ grid almost never percolates, and when $p > p^\star$, a random $N \times N$ grid almost always percolates.

(a) For given values of N and p value, assume that one $N \times N$ grid G has been computed with probability p, and that the set of open/blocked sites has been determined. Propose an algorithm that determines whether or not the system percolates. Call it $\mathsf{Percol}(G)$ algorithm.

(b) What is the complexity of your $\mathsf{Percol}(G)$ algorithm?

(c) For given N and p, what do you suggest in order to estimate the probability that the system percolates?

(d) (Special Christmas bonus points.) Propose an algorithm in order to estimate $p^\star$.

(e) (Special Christmas bonus points.) What is the complexity of your algorithm?