UNIVERSITY OF TORONTO
Department of Computer Science

Duration – 50 minutes

Examination Aids: One side of one page (A4 or US Letter) handwritten aid sheet.

- Check that your exam booklet has 7 pages (including this cover page).
- This exam has 4 questions.
- Put all answers in this booklet, in the spaces provided. If you run out of space, you can use the back of the page, provided that you ***clearly indicate which question you are answering***.
- For rough work, use pages 6 to 7; ***these will not be marked***.
- Your answers will be marked based on the correctness of your answers ***and*** the clarity of your explanations.
- Good luck!

PLEASE COMPLETE THIS SECTION

First name ______________________________

Family name ______________________________

Problem	Marks Received	Marks Worth
1.		20
2.		10
3.		18
4.		12
TOTAL		60

Question 1. (20 marks) Let $G = (V, E)$ be a connected undirected unweighted graph. Recall that a *cut* of G is given by a partition of the vertices of G into two sets, S and $\bar{S}$. The set of edges that have one endpoint in S and the other in $\bar{S}$ is $E(S, \bar{S})$. For any real number $\alpha \geq 1$, we define an α-approximate minimum cut of G to be a cut $(S, \bar{S})$ such that for every other cut $(T, \bar{T})$, we have

$$|E(S, \bar{S})| \leq \alpha |E(T, \bar{T})|.$$

a. Assume that for every $u \in V$, the cut $(\{u\}, V - \{u\})$ is *not* a 2-approximate minimum cut. Let $(S, \bar{S})$ be a minimum cut of G. Give an upper bound on the probability that an edge e chosen uniformly at random from E is in $E(S, \bar{S})$. Briefly justify your answer.

identify a lower bound on |E|

Answer: Let $k = |E(S, \bar{S})|$, and let d_u be the degree of the vertex u. Because the cut $(\{u\}, V - \{u\})$ is not 2-approximate, and $d_u = |E(\{u\}, V - \{u\})|$, we have $d_u > 2k$ for every u. Summing all degrees, we get $2m = \sum_{u \in V} d_u > 2kn$. The probability that $e \in E(S, \bar{S})$ is $\frac{k}{m} < \frac{1}{n}$.

b. Consider a variant of the Contraction Algorithm, in which the algorithm keeps in memory an additional cut $(T, \bar{T})$. Initially, $T = \{u\}$, where u is the smallest degree vertex in G. After every contraction, the algorithm checks if the smallest degree of a supernode in the current contracted graph is less than $|E(T, \bar{T})|$, and if it is, it replaces T with the set of vertices of G corresponding to the supernode. Finally, the algorithm outputs either the cut that is output by the standard Contraction Algorithm, or $(T, \bar{T})$, whichever has fewer edges. Give a lower bound on the probability that this algorithm outputs a 2-approximate minimum cut of G. Briefly justify your answer.

HINT: If $(T, \bar{T})$ is not a 2-approximate minimum cut of G, then what is a lower bound on the probability that the contraction algorithm outputs a minimum cut of G?

Answer: Let $(S, \bar{S})$ be a minimum cut in G, and let $(R, \bar{R})$ be the (random) cut output by the contraction algorithm. If the algorithm above does *not* output a 2-approximate minimum cut, then $(T, \bar{T})$ is not a 2-approximate minimum cut and $(R, \bar{R})$ is not a 2-approximate minimum cut, and, therefore, is not a minimum cut, and in particular is not equal to $(S, \bar{S})$. So the probability that the algorithm does *not* output a 2-approximate minimum cut is at most the probability that $(R, \bar{R}) \neq (S, \bar{S})$ and $(T, \bar{T})$ is not a 2-approximate minimum cut, which is

$$\mathbb{P}((R, \bar{R}) \neq (S, \bar{S}) \text{ and } (T, \bar{T}) \text{ not 2-approximate}) \leq \mathbb{P}((R, \bar{R}) \neq (S, \bar{S}) \mid (T, \bar{T}) \text{ not 2-approximate})$$

The probability on the right hand side is equal to the probability that at least one of the edges we contract is in $E(S, \bar{S})$, which, using the previous subproblem, is at most

$$1 - \frac{n-1}{n} \cdot \frac{n-2}{n-1} \cdots \frac{2}{3} = 1 - \frac{2}{n}.$$

Therefore the probability that the algorithm *does* output a 2-approximate cut is at least $\frac{2}{n}$.

Question 2. (10 marks) Consider a distance function $d(x, y)$ between two n-bit strings $x, y \in \{0,1\}^n$ defined to equal the sum of the indexes in which x and y differ. I.e. we define

$$d(x, y) = \sum_{i : x_i \neq y_i} i.$$

For example, the 8-bit strings $x = 01000111$ and $y = 00010101$ differ in bits $2, 4, 7$, so $d(x, y) = 2+4+7 = 13$.

Give a random hash function $h : \{0,1\}^n \to \{0,1\}$, so that

$$\mathbb{P}(h(x) = h(y)) = 1 - Cd(x, y),$$

where C is a quantity that depends on n but not on x or y. Give an explicit value of C for your hash function. The function should be computable in time $O(1)$ when x is given as an array of size n. Justify your answer.

Answer: Pick a random index i with probability $\mathbb{P}(i = j) = \frac{2j}{n(n+1)}$, and set $h(x) = x_i$. Then

$$\mathbb{P}(h(x) \neq h(y)) = \mathbb{P}(x_i \neq y_i) = \sum_{j : x_j \neq y_j} \frac{2j}{n(n+1)} = \frac{2}{n(n+1)} d(x, y).$$

Since $\mathbb{P}(h(x) = h(y)) = 1 - \mathbb{P}(h(x) \neq h(y))$, we have our random hash function with $C = \frac{2}{n(n+1)}$.

normalizing factor for a valid probability distribution

Question 3. (18 marks) Let $A[1..n]$ be an array of *distinct* integers. Recall that the *rank* of an integer x in A equals r if and only if there are exactly $r - 1$ integers in A that are strictly smaller than x. E.g. if $A = [8, 3, 19, 6, 5, 7]$, and $x = 6$, then the rank of x in A is 3.

Consider an algorithm that samples k indexes $i_1, \ldots, i_k$, independently and uniformly at random with replacement from $\{1, \ldots, n\}$.

a. Let ℓ be the number of integers in $A[i_1], \ldots, A[i_k]$, counted with repetition, that have rank at least $\frac{2n}{3} + 1$. (Assume n is divisible by 3.) What is the expected value of ℓ? What is the variance of ℓ? You do not need to justify your answer.

Let X_j be indicator random variable which is equal to 1 if $A[i_j]$ has rank at least $\frac{2n}{3} + 1$, and 0 otherwise. Then $\ell = X_1 + \ldots + X_k$. Since exactly $\frac{1}{3}$ of the elements of A have rank at least $\frac{2n}{3} + 1$, we have $\mathbb{E}[X_j] = \mathbb{P}(X_j = 1) = \frac{1}{3}$. Then, by linearity of expectation,

$$\mathbb{E}[\ell] = \mathbb{E}[X_1] + \ldots + \mathbb{E}[X_k] = \frac{k}{3}.$$

Because $i_1, \ldots, i_k$ were picked independently, the random variables $X_1, \ldots, X_k$ are independent, and we have

$$\mathrm{Var}(\ell) = \mathrm{Var}(X_1) + \ldots + \mathrm{Var}(X_k).$$

compute variance of indicator variable, instead of l diretly

For each X_j,

$$\mathrm{Var}(X_j) = \mathbb{E}[X_j^2] - \mathbb{E}[X_j]^2 = \frac{1}{3} - \frac{1}{9} = \frac{2}{9}.$$

Therefore,

$$\mathrm{Var}(\ell) = k\mathrm{Var}(X_j) = \frac{2k}{9}.$$

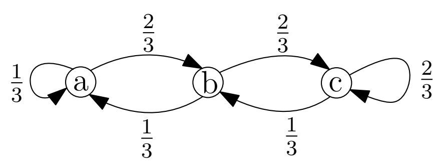
b. Assume that k is odd. Give an upper bound on the probability that the median of $A[i_1], \dots, A[i_k]$ has rank at least $\frac{2n}{3} + 1$ in A. Make your upper bound as tight as possible. You do not need to justify your answer.

Notice that the median of $A[i_1], \dots, A[i_k]$ has rank at least $\frac{2n}{3} + 1$ in A if and only if $\ell \geq \frac{k+1}{2}$. By Chebyshev' we have

$$\mathbb{P}\left(\ell \geq \frac{k+1}{2}\right) \leq \mathbb{P}\left(|\ell - \mathbb{E}[\ell]| \geq \frac{k+3}{6}\right) \leq \frac{\mathrm{Var}(\ell)6^2}{(k+3)^2} = \frac{8k}{(k+3)^2} \leq \frac{8}{k}.$$

i.e. median has rank over 2/3n+1 <=> number of digits over the rank is over half

Question 4. (12 marks) Consider the directed graph G on the vertices a, b, and c drawn below. The numbers next to the edges give the transition probabilities of a random walk: for example when the random walk is at vertex a, the probability of staying at a is $\frac{1}{3}$, and the probability of moving to vertex b is $\frac{2}{3}$.



Draw the transition matrix of this random walk, and give its stationary probability distribution.

The transition matrix is:

$$P = \begin{pmatrix} \frac{1}{3} & \frac{2}{3} & 0 \\ \frac{1}{3} & 0 & \frac{2}{3} \\ 0 & \frac{1}{3} & \frac{2}{3} \end{pmatrix}.$$

Let π be the stationary distribution. We must have $\pi P = \pi$. It is easy to solve this system of equations. The equation for π_a gives

$$\pi_a = \frac{1}{3}\pi_a + \frac{1}{3}\pi_b,$$

so $\pi_b = 2\pi_a$. The equation for π_b gives

$$\pi_b = \frac{2}{3}\pi_a + \frac{1}{3}\pi_c = \frac{1}{3}\pi_b + \frac{1}{3}\pi_c,$$

so $\pi_c = 2\pi_b$. Because the probabilities need to sum to 1, we have $\pi_a = \frac{1}{7}$, $\pi_b = \frac{2}{7}$, and $\pi_c = \frac{4}{7}$.