A

BRAC UNIVERSITY
Department of Computer Science and Engineering

Examination: Semester Final Exam | Semester: Summer 2025
Duration: 2 Hours | Full Marks: 45

CSE 221: Algorithms

Answer the following questions.
Figures in the right margin indicate marks.

Name:	ID:	Section:

1 **a.** [CO1] A directed graph has the following vertices:
1, 2, 3, 4, 5, 6, 7, 8

Following is the adjacency list for the graph:
1: [2, 3]
2: [4]
3: [1, 5]
4: [6]
5: [6]
6: [7]
7: [5, 8]
8: [ ]

i. Using the above information, **draw** the directed graph. **02**
ii. **Identify** and write all the Strongly Connected Components containing at least 2 nodes. **02**
[No need to show any process.]

b. [CO1] Consider the following everyday morning activities:

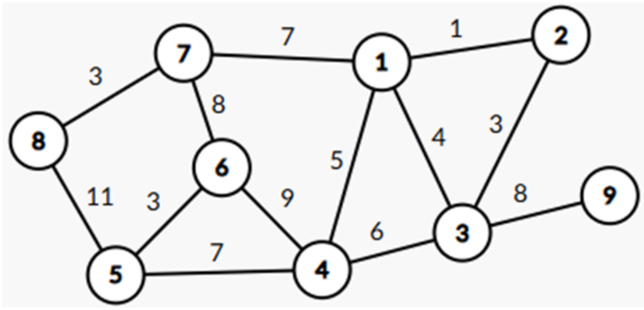
- ➔ Waking Up
- ➔ Brushing
- ➔ Leaving the house
- ➔ Taking a shower
- ➔ Getting dressed
- ➔ Having breakfast

Constraints:
- You must not get dressed before taking a shower
- You are free to have breakfast and then brush, or the opposite
- After waking up: you can either take a shower and then brush, or the opposite
- You must not leave the house without getting dressed
- Waking up is the first thing you do, and leaving the house is the last activity

i. **Draw** a directed graph representing the activities as nodes, and constraints as edges. **03**
ii. From the above activities, **write** two different topologically sorted sequences of activities. **02**
iii. **Explain** in short, how you would use DFS to topologically sort the activities. **01**

2 The city is struck by a massive flood. Several shelters have been established. Rescue teams are coordinating supplies, rescue boats etc. To ensure communication, the government wants to set up temporary bridges between shelters so that all shelters remain connected.

For example, in the graph below, nodes represent shelters. The cost of setting up bridges between two shelters are given as weighted edges.



a. [CO1] **Simulate** a suitable algorithm to find which bridges to build so that all shelters remain connected at the minimum possible cost. Also, report the total cost of this arrangement. **04**

b. [CO3] However, new information arrived that shelter 3 is already equipped with satellite communication devices, which make direct connections to this shelter cheaper than normal. Propose a modification to your algorithm in Q2(a) to directly connect (by building bridges) as many shelters as possible with shelter 3, even if they are not the lowest-weight edges in the overall graph. Just **explain** the modification. **03**

c. [CO2] After connecting all the shelters, you have decided to make shelter 2 as the base of operations. So you want to connect all shelters with shelter 2 at minimum possible cost. Do the same set of bridges you have built for Q2(a) also serve this purpose?
Explain in short with proper reasoning. **03**

3 **a.** [CO1] During the July festival, organizers want to arrange a drone show. There will be N drones carrying lights over an N × N aerial stage. The drones must form a symmetrical light display, but no two drones can align vertically or horizontally.

If the number of drones is 6, **draw** a valid arrangement of drones in a 6x6 grid. **02**

b. [CO1] A delivery robot works in a storage area organized in rows and columns of delivery packages, where each package can be thought of as a cell. There is only one free spot (cell) that allows the robot to slide packages around. Its goal is to arrange all the packages in a specific order so that they are ready for shipping.

While trying different moves, the robot keeps going back to layouts it has already tried, getting stuck in a loop and making no progress.

1. How can the robot avoid repeating the same arrangements again and again? **Explain** within two sentences. **02**
2. **Name** an algorithm that will help to find the shortest sequence of moves to reach the correct order of packages. **01**

4 You are designing a data compression and transmission system for a sensor network that monitors environmental conditions. Each type of sensor reading has a frequency of occurrence and an impact factor indicating its importance for critical decision-making.

Your goals are to compress the data efficiently for transmission using a variable-length encoding system and maximize the total impact factor transmitted within the system's data capacity.

Sensor Data Table:

Sensor Reading	Frequency of occurrence	Impact Factor
Temperature	45	10
Humidity	13	6
Pressure	12	8
Light	16	7
CO2	9	5
Wind Speed	5	4
Rainfall	7	6
UV Index	3	3

a. [CO1] **Construct** a Huffman tree for these sensor readings based on their frequencies of occurrence. Then **assign** binary codes to each reading. **05**

b. [CO1] **Calculate** the total number of bits required for the transmission of all readings. **02**

c. [CO1] Suppose, the system can transmit only 15 units of data. Each sensor reading generates 1 (one) unit of data per occurrence. You can partially transmit the readings but its impact factor drops proportionally if you do so. **03**
Determine which readings to transmit and with what frequency, to maximize the total impact factor within the 15-unit data limit. **Calculate** the maximum total impact factor.

5 A genetics research team is studying DNA sequences to identify whether mutations have occurred during sequencing. When an experimental DNA reading is obtained, it is compared against a known reference gene to evaluate their similarity. However, sequencing errors or natural mutations may lead to insertions, deletions, or substitutions of nucleotides.

To measure this similarity, two approaches are commonly used. The Longest Common Subsequence (LCS) measures similarity by finding the longest subsequence that appears in both sequences without rearranging their order. On the other hand, the Edit Distance measures dissimilarity by calculating the minimum number of edits (insertions, deletions, or substitutions) required to transform one sequence into the other. Both can be solved using Dynamic Programming as there are many overlapping subproblems.

You are provided with the following two DNA sequences:
Experimental reading (Read): ***ACGTA***
Reference gene (Ref): ***AGCTG***

Suppose, *m = length(Read)* and *n = length(Ref)*.

For both LCS, and Edit Distance
- The dimension of the dp table is (m+1)*(n+1)
- The final answer is stored in cell dp[m][n]

a. **[CO1]** Recurrence relation for <u>Longest Common Subsequence</u>:

$$dp[i][j] = \begin{cases} 0 & \text{if } i = 0 \text{ or } j = 0 \\ 1 + dp[i-1][j-1] & \text{if } Read[i-1] = Ref[j-1] \\ \max(dp[i-1][j],\ dp[i][j-1]) & \text{otherwise} \end{cases}$$

Here, each cell dp[i][j] stores the length of the LCS between the first i characters of Read and the first j characters of Ref.

(i) **Construct** the dp table to find the LCS of the provided DNA sequences (Read and Ref) by following this recurrence relation. **04**
(ii) **Determine** the length of LCS from the dp table and **write** a valid LCS. **01**

b. **[CO3]** Recurrence relation for <u>Edit Distance</u>:

$$dp[i][j] = \begin{cases} i & \text{if } j = 0 \\ j & \text{if } i = 0 \\ dp[i-1][j-1] & \text{if } Read[i-1] = Ref[j-1] \\ 1 + \min(dp[i-1][j],\ dp[i][j-1],\ dp[i-1][j-1]) & \text{otherwise} \end{cases}$$

Here, each cell dp[i][j] stores the minimum number of edits required to transform the first i characters of Read into the first j characters of Ref.

(i) **Construct** the dp table to find the Edit Distance between the provided DNA sequences (Read and Ref) by following this recurrence relation. **03**
(ii) From the dp table, **determine** the minimum number of edits required to transform one sequence into the other. **01**

c. **[CO2]** Prove the following using the answers from Q5(a-ii) and Q5(b-ii). **01**
EditDistance(Read, Ref) ≤ m + n − 2 × LCS(Read, Ref)

BRAC UNIVERSITY
Department of Computer Science and Engineering

Examination: Semester Final Exam | Semester: Summer 2025
Duration: 2 Hours | Full Marks: 45

CSE 221: Algorithms

Answer the following questions.
Figures in the right margin indicate marks.

Name:	ID:	Section:

1 **a.** [CO1] A directed graph has the following vertices:
1, 2, 3, 4, 5, 6, 7, 8

Following is the adjacency list for the graph:
1: [2]
2: [3]
3: [1, 4]
4: [5]
5: [6]
6: [4]
7: [6, 8]
8: [7]

i. Using the above information, **draw** the directed graph. **02**
ii. **Identify** and write all the Strongly Connected Components containing at least 2 nodes. **02**
[No need to show any process.]

b. [CO1] Consider the following nighttime activities:

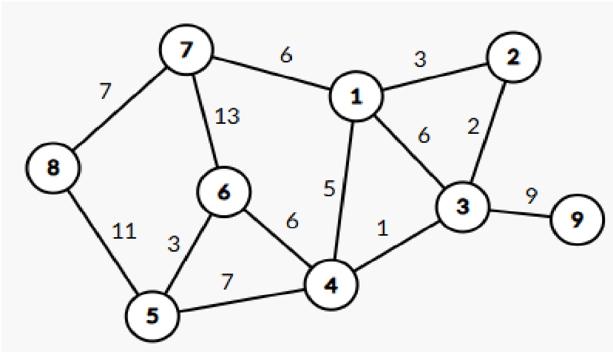
- ➔ Having dinner
- ➔ Brushing teeth
- ➔ Going to bed
- ➔ Turning off the lights
- ➔ Changing into pajamas
- ➔ Fending off the ghosts

Constraints:
- You must not go to bed before brushing your teeth
- You must change into pajamas before going to bed
- You are free to have dinner before or after fending off the ghosts
- You must turn off the lights before going to bed
- Fending off the ghosts must be done before turning off the lights
- Going to bed is the last activity

i. **Draw** a directed graph representing the activities as nodes, and constraints as edges. **03**
ii. From the above activities, **write** two different topologically sorted sequences of activities. **02**
iii. **Explain** in short, how you would use DFS to topologically sort the activities. **01**

2 The city is struck by a massive flood. Several shelters have been established. Rescue teams are coordinating supplies, rescue boats etc. To ensure communication, the government wants to set up temporary bridges between shelters so that all shelters remain connected.

For example, in the graph below, nodes represent shelters. The cost of setting up bridges between two shelters are given as weighted edges.



a. [CO1] **Simulate** a suitable algorithm to find which bridges to build so that all shelters remain connected at the minimum possible cost. Also, report the total cost of this arrangement. **04**

b. [CO3] However, new information arrived that shelter 2 is already equipped with satellite communication devices, which make direct connections to this shelter cheaper than normal. Propose a modification to your algorithm in Q2(a) to directly connect (by building bridges) as many shelters as possible with shelter 2, even if they are not the lowest-weight edges in the overall graph. Just **explain** the modification. **03**

c. [CO2] After connecting all the shelters, you have decided to make shelter 4 as the base of operations. So you want to connect all shelters with shelter 4 at minimum possible cost. Do the same set of bridges you have built for Q2(a) also serve this purpose? **03**
Explain in short with proper reasoning.

3 **a.** [CO1] During the July festival, organizers want to arrange a drone show. There will be N drones carrying lights over an N × N aerial stage. The drones must form a symmetrical light display, but no two drones can align vertically or horizontally.

If the number of drones is 6, **draw** a valid arrangement of drones in a 6x6 grid. **02**

b. [CO1] A delivery robot works in a storage area organized in rows and columns of delivery packages, where each package can be thought of as a cell. There is only one free spot (cell) that allows the robot to slide packages around. Its goal is to arrange all the packages in a specific order so that they are ready for shipping.

While trying different moves, the robot keeps going back to layouts it has already tried, getting stuck in a loop and making no progress.

1. How can the robot avoid repeating the same arrangements again and again? **Explain** within two sentences. **02**
2. **Name** an algorithm that will help to find the shortest sequence of moves to reach the correct order of packages. **01**

4 You are designing a data compression and transmission system for a sensor network that monitors environmental conditions. Each type of sensor reading has a frequency of occurrence and an impact factor indicating its importance for critical decision-making.

Your goals are to compress the data efficiently for transmission using a variable-length encoding system and maximize the total impact factor transmitted within the system's data capacity.

Sensor Data Table:

Sensor Reading	**Frequency of occurrence**	**Impact Factor**
Temperature	50	12
Humidity	10	7
Pressure	15	9
Light	20	8
CO2	8	6
Wind Speed	6	5
Rainfall	12	7
UV Index	4	3

a. [CO1] **Construct** a Huffman tree for these sensor readings based on their frequencies of occurrence. Then **assign** binary codes to each reading. **05**

b. [CO1] **Calculate** the total number of bits required for the transmission of all readings. **02**

c. [CO1] Suppose, the system can transmit only 20 units of data. Each sensor reading generates 1 (one) unit of data per occurrence. You can partially transmit the readings but its impact factor drops proportionally if you do so. **03**
Determine which readings to transmit and with what frequency, to maximize the total impact factor within the 20-unit data limit. **Calculate** the maximum total impact factor.

5 A genetics research team is studying DNA sequences to identify whether mutations have occurred during sequencing. When an experimental DNA reading is obtained, it is compared against a known reference gene to evaluate their similarity. However, sequencing errors or natural mutations may lead to insertions, deletions, or substitutions of nucleotides.

To measure this similarity, two approaches are commonly used. The Longest Common Subsequence (LCS) measures similarity by finding the longest subsequence that appears in both sequences without rearranging their order. On the other hand, the Edit Distance measures dissimilarity by calculating the minimum number of edits (insertions, deletions, or substitutions) required to transform one sequence into the other. Both can be solved using Dynamic Programming as there are many overlapping subproblems.

You are provided with the following two DNA sequences:
Experimental reading (Read): ***ACGTA***
Reference gene (Ref): ***GACGC***

Suppose, *m = length(Read)* and *n = length(Ref).*

For both LCS, and Edit Distance
- The dimension of the dp table is (m+1)*(n+1)
- The final answer is stored in cell dp[m][n]

a. **[CO1]** Recurrence relation for Longest Common Subsequence:

$$dp[i][j] = \begin{cases} 0 & \text{if } i = 0 \text{ or } j = 0 \\ 1 + dp[i-1][j-1] & \text{if } Read[i-1] = Ref[j-1] \\ \max(dp[i-1][j],\ dp[i][j-1]) & \text{otherwise} \end{cases}$$

Here, each cell dp[i][j] stores the length of the LCS between the first i characters of Read and the first j characters of Ref.

(i) **Construct** the dp table to find the LCS of the provided DNA sequences (Read and Ref) by following this recurrence relation. **04**
(ii) **Determine** the length of LCS from the dp table and **write** a valid LCS. **01**

b. **[CO3]** Recurrence relation for Edit Distance:

$$dp[i][j] = \begin{cases} i & \text{if } j = 0 \\ j & \text{if } i = 0 \\ dp[i-1][j-1] & \text{if } Read[i-1] = Ref[j-1] \\ 1 + \min(dp[i-1][j],\ dp[i][j-1],\ dp[i-1][j-1]) & \text{otherwise} \end{cases}$$

Here, each cell dp[i][j] stores the minimum number of edits required to transform the first i characters of Read into the first j characters of Ref.

(i) **Construct** the dp table to find the Edit Distance between the provided DNA sequences (Read and Ref) by following this recurrence relation. **03**
(ii) From the dp table, **determine** the minimum number of edits required to transform one sequence into the other. **01**

c. **[CO2]** Prove the following using the answers from Q5(a-ii) and Q5(b-ii). **01**
EditDistance(Read, Ref) ≤ m + n − 2 × LCS(Read, Ref)