Digest Recommendation

Problem ID: digest

Quora's mission is to share and grow the world's knowledge. A vast amount of the knowledge that would be valuable to many people is currently only available to a few - either locked in people's heads, or only accessible to select groups. We want to connect the people who have knowledge to the people who need it, to bring together people with different perspectives so they can understand each other better, and to empower everyone to share their knowledge for the benefit of the rest of the world.

There's a lot of content on Quora and we need a way of recommending the top stories for a user. Recommender systems drive many features at Quora, but in this problem we'll focus on Digest recommendation. Quora Digests are emails which we send to users containing top recommended stories based on their interests.

Note: This is a fictitious representation of our Digest recommender system.

There are n stories $(S_1, \ldots, S_n)$ and m users $(U_1, \ldots, U_m)$. Each story is created by a single user. A user can follow both other users and other stories.

The Digest recommendation score for story S_k to user U_i is defined as follows:

- If U_i created or follows S_k, the score is -1.
- Otherwise, the score is

$$\sum_{j=1}^{m} a(U_i, U_j) \times b(U_j, S_k)$$

where

$$a(U_i, U_j) = \begin{cases} 0 \text{ if } i = j \\ \text{otherwise, } 3 \text{ if } U_i \text{ follows } U_j \\ \text{otherwise, } 2 \text{ if } U_i \text{ follows stories created by } U_j \\ \text{otherwise, } 1 \text{ if } U_i \text{ follows stories followed by } U_j \\ \text{otherwise, } 0 \end{cases}$$

$$b(U_j, S_k) = \begin{cases} 2 \text{ if } U_j \text{ created } S_k \\ \text{otherwise, } 1 \text{ if } U_j \text{ follows } S_k \\ \text{otherwise, } 0 \end{cases}$$

Given the stories, users, and their associations, find the top three recommended Digest stories for each user.

Input

Your program will receive input from standard input.

The first line contains two space-separated positive integers n and m, representing the number of stories and the number of users.

The following n lines each contain a single integer. The k-th line contains an integer j indicating that U_j created S_k.

After that, there will be two integers p and q, representing the number of follows between users and the number of story follows.

The next p lines each contain two integers i and j representing that user U_i follows user U_j.

The final q lines each contain two integers i and k representing that user U_i follows story S_k.

Output

Your program should write to standard output.

Print m lines. The i-th line should contain three integers representing the recommended stories for U_i. To select the stories to output, sort all the stories in non-increasing order by recommendation score, and in increasing order by story index among stories with the same score. Then, print the indexes of the first 3 stories in that order.

Constraints

- $5 \leq n, m \leq 200$
- $0 \leq p \leq n^2 - n$
- $0 \leq q \leq nm$
- It is guaranteed that there will be at least 3 stories to recommend for each user.
- There will be no duplication on follow relations.
- When U_i created S_k, it is guaranteed that U_i does not follow S_k.

Scoring

There are 25 test cases, each worth 4 points. Your submission score will be the sum of the points you get from each test case you pass.

Sample Input 1

```
7 5
1
1
2
3
4
5
5
2 4
1 2
4 5
1 7
5 2
3 3
4 1
```

Sample Output 1

```
3 6 4
1 2 4
1 2 5
7 2 6
1 3 4
```

Escape

Problem ID: escape

Victor is trapped in a $n \times m$ maze monitored by k guards and we want to help him escape without being seen by any of the guards. The i-th guard is initially located at (r_i, c_i) and covers distance d_i, which means guard i monitors any cell (r, c) in the maze as long as (r, c) is reachable from (r_i, c_i) in d_i moves. The maze has walls which neither Victor nor the guards can occupy or pass through. The guards and Victor may move to any of the 4 adjacent cells in a single move as long as it is empty.

Given Victor's initial location, find the shortest path to the exit without touching cells monitored by guards.

Input

Your program will receive input from standard input.

The first line contains three space-separated positive integers n, m, and k, representing the number of rows of the maze, the number of columns of the maze, and the number of guards, respectively. The following n lines each contain m characters representing a map of the maze. Each character is one of the following:

- `.`: represents an empty cell
- `#`: represents a wall
- `S`: represents Victor's initial location
- `E`: represents the exit

There is only one `S` and `E`. The border of the maze is always a wall `#`.

In the next k lines, the i-th line contains three space-separated integers r_i, c_i, and d_i, representing the initial location and movement range of the i-th guard. The top-left corner is represented as $(1, 1)$, and the bottom-right corner is represented as (n, m).

Output

Your program should write to standard output.

Print exactly one line. The line should contain the length of the shortest path that Victor can take to move to the exit without touching cells monitored by guards if it is possible, otherwise print `IMPOSSIBLE` instead. Note that Victor's initial location or the exit may already be monitored by a guard, in which case you should print `IMPOSSIBLE`.

Constraints

- $5 \leq n, m \leq 10^3$
- $0 \leq k \leq 10^4$
- $1 < r_i < n; 1 < c_i < m; 0 \leq d_i \leq nm$
- Guards cannot be placed on a wall
- The border of the maze is always a wall
- No two guards are placed at the same location

Subtasks

You will get points for each subtask when you pass all of the testcases of the subtask.

1. $n, m \leq 100, k \leq 10^3$ (43 points)
2. No additional constraints (57 points)

Sample Input 1

```
10 10 2
##########
#........#
#........#
#...#...E#
#........#
#........#
#.S......#
#.......##
#........#
##########
6 7 2
4 4 1
```

Sample Output 1

```
15
```

Sample Input 2

```
9 9 2
#########
#.......#
#.#.###.#
#.......#
#.###.E.#
#.....#.#
#...S##.#
#.......#
#########
4 4 1
5 6 0
```

Sample Output 2

```
8
```

Students
Problem ID: students

It's time for lunch break at a certain school, and the n^2 students are out and about in the playground. This playground is an $n \times n$ grid, and there is one student standing on each cell of the grid. Since the students have been playing and chatting with each other for awhile, some m pairs of adjacent (cells sharing an edge) students have become friends. The teacher wants to take some of the students on a field trip, but does not want it to be too rowdy, so no student on the field trip can be friends with any other student on the trip.

Write a program to find the maximum number of students the teacher can take on the field trip.

Input

Your program will receive input from standard input.

The first line of the input contains two space-separated integers, n and m.

m lines follow. The i-th line contains 4 space separated integers $x_{i,1}, y_{i,1}, x_{i,2}, y_{i,2}$ indicating that the student located in row $x_{i,1}$, column $y_{i,1}$ and the student located in row $x_{i,2}$, column $y_{i,2}$ are friends.

Output

Your program should write to standard output.

Print exactly one line containing the maximum number of students the teacher can take on the field trip.

Constraints

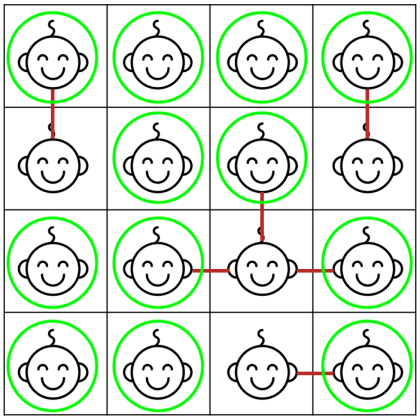
- $2 \leq n \leq 3 \cdot 10^3$
- $0 \leq m \leq \min(2n(n-1), 2 \cdot 10^5)$
- $1 \leq x_{i,1}, y_{i,1}, x_{i,2}, y_{i,2} \leq n$
- $|x_{i,1} - x_{i,2}| + |y_{i,1} - y_{i,2}| = 1$

Subtasks

You will get points for each subtask when you pass all of the testcases of the subtask.

1. $n \leq 5$ (21 points)
2. $n \leq 100$ (26 points)
3. No additional constraints (53 points)

Sample Explanation



This image illustrates the sample Input/Output. Friendships are indicated by the red lines. The students circled in green illustrate one possible optimal group the teacher could choose to take on the field trip.

Sample Input 1	Sample Output 1
4 6 3 2 3 3 3 3 3 4 4 3 4 4 1 1 2 1 1 4 2 4 2 3 3 3	12

Walls
Problem ID: walls

Due to COVID-19, Quora decided to split the office space into a grid of $3 \times n$ cells and set up the walls between each cell to isolate employees as much as possible. Unfortunately, the walls ended up affecting the productivity significantly in a negative way. Once our employees are all vaccinated we want to break down the walls and connect all our employees again.

The cost to break each wall is different and we would like to minimize the cost to connect people. So, given the locations of m employees and cost of individual walls, find the minimum cost to connect employees again.

Input

Your program will receive input from standard input.

The first line contains two space-separated positive integers n and m, representing the number of columns in the grid and the number of employees, respectively.

3 lines follow. The i-th line of this set contains $n-1$ integers $a_1, \ldots a_{n-1}$. a_j is the cost to break the vertical wall between the cells in columns j and $j+1$ in the i-th row.

Then, 2 lines follow. The i-th line of this set contains n integers $b_1, \ldots b_n$. b_j is the cost to break the horizontal wall between rows i and $i+1$ in the j-th column.

Finally, m lines follow. The i-th line of this set contains two integers r_i and c_i, representing the row and column of the i-th employee's cell.

Output

Your program should write to standard output.

Print exactly one line containing the minimum cost to connect all employees.

Constraints

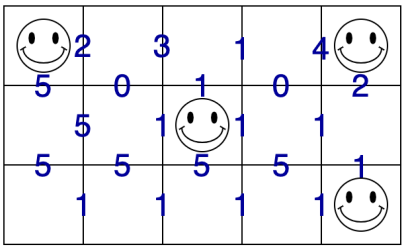
- $n \leq 10^4$
- $1 \leq m \leq 3n$
- $0 \leq a_j \leq 10^6$ for all j.
- $0 \leq b_j \leq 10^6$ for all j.
- $1 \leq r_i \leq 3$ for all i.
- $1 \leq c_i \leq n$ for all i.

Subtasks

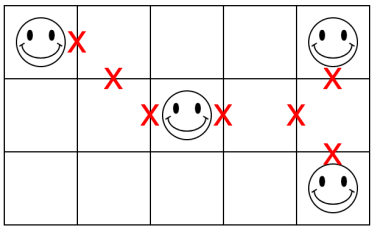
You will get points for each subtask when you pass all of the testcases of the subtask.

1. $n \leq 5$ (14 points)
2. $n \leq 20$ (13 points)
3. $n \leq 100$ (22 points)
4. No additional constraints (51 points)

Sample Explanation



This image illustrates the grid in the sample Input/Output.



This image illustrates the optimal walls to break for the sample Input/Output.

Sample Input 1	Sample Output 1
5 4 2 3 1 4 5 1 1 1 1 1 1 1 5 0 1 0 2 5 5 5 5 1 1 1 1 5 2 3 3 5	8

Tourism
Problem ID: tourism

You are a tourist and want to consider some possible trip scenarios.

There are n cities and $n-1$ bidirectional roads such that each city is reachable from any other city (in other words, the graph of cities is a tree). Some roads require tolls, and some cities have friends who can give you money.

You have m independent queries to answer. Each query is a trip you are considering: you would start in some city a, and want to reach some city b along the shortest path (in other words, you must travel along the unique simple path from city a to city b). Whenever you pass a road with a toll, you must pay that amount of money to pass. Whenever you pass a city with friends, they will give you money. In some scenarios, you would need to have some initial amount of money with you before you start your trip, or else you would run out of money before reaching your destination city.

Note: If there are friends in the starting city, you do receive money from them.

For each independent query, determine the minimum amount of money you would have to start with in order to travel from starting city a to ending city b. You are considering several trips, so you will be given several queries, each consisting of a single pair of starting and ending cities. Only consider one query trip at a time; the queries are not cumulative.

Input

Your program will receive input from standard input.

The first line of the input contains two space-separated integers, n, the number of cities (numbered from 1 to n) and m, the number of independent queries to answer.

The second line contains n integers $x_1, \ldots, x_n$ representing your friends. Friends in city i are willing to give you x_i dollars.

Next, there are $n-1$ lines describing the roads between cities. The i-th contains three integers v_i, w_i, t_i indicating that there is a road from city v_i to city w_i, which charges toll t_i, to pass in either direction. It is guaranteed that the resulting graph will form a tree.

Finally, there are m more lines. The i-th of these lines represents the i-th query and contains two integers a_i and b_i. a_i is the start city and b_i is the destination city.

Output

Your program should write to standard output.

Write m lines. The i-th line should be the minimum amount of money you must have to successfully complete the i-th trip.

Constraints

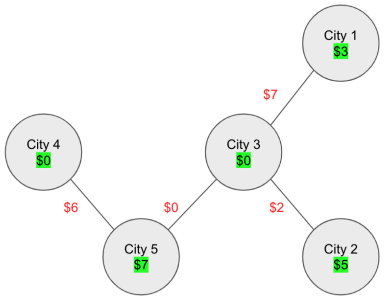
- $2 \leq n \leq 2 \cdot 10^5$
- $2 \leq m \leq 2 \cdot 10^5$
- $0 \leq x_i, t_i \leq 10^9$
- $1 \leq v_i, w_i \leq n; v_i \neq w_i$
- $1 \leq a_i, b_i \leq n; a_i \neq b_i$

Subtasks

You will get points for each subtask when you pass all of the testcases of the subtask.

1. $nm \leq 5 \cdot 10^5$ (13 points)
2. The given graph is a line (18 points)
3. No additional constraints (69 points)

Sample Explanation



This image illustrates the graph in the sample Input/Output.

In the first query, you need 4 to start with in order to travel from city 1 to city 3 (you get $\$3$ from your friends in the starting city, city 1, and you need $\$4$ more in order to pay the $\$7$ toll).

In the second query, you need $\$0$ to start with in order to travel from city 2 to city 4.

In the third query, you need $\$6$ to start with in order to travel from city 4 to city 2.

Sample Input 1	Sample Output 1
5 3 3 5 0 0 7 3 1 7 2 3 2 5 4 6 3 5 0 1 3 2 4 4 2	4 0 6

[ML] Flipped Data

Problem ID: flipped

You were carrying your data to your machine learning model, but along the way you tripped and fell! You accidentally flipped b rows of your training data when this happened but aren't sure which ones. Your data consists of an array of n rows and f columns. Thankfully, you know that your data initially was nicely distributed. Originally, the data in the j-th column was normally distributed with mean u_j, and standard deviation d_j.

Determine which rows were flipped so you can train your model.

Input

Your program will receive input from standard input.

The first line will contain three space-separated integers, n, f, and b, representing the number of rows and columns in each row, and the number of flipped rows. The following n lines will each contain f space separated floats.

Output

Your program should write to standard output.

Print b lines, each containing a single integer o_i indicating that row o_i was flipped. Rows numbers start at 1 and go up to n.

Constraints

- $10^2 \leq n \leq 10^5$
- $10 \leq b \leq 0.4 \times n$
- $10 \leq f \leq 30$
- $0 \leq u_i \leq 1$
- $0 \leq d_i \leq 0.5$

Scoring

The score awarded for this problem is the percentage of flipped rows which were correctly identified. Scores are rounded to the nearest integer.

Sample Input 1

```
10 6 3
0.48 0.97 0.24 0.58 0.69 0.29
0.16 0.61 0.74 0.41 0.96 0.41
0.26 0.61 0.63 0.32 0.81 0.47
0.50 0.90 0.45 0.68 0.64 0.20
0.42 0.95 0.46 0.64 0.60 0.24
0.06 0.51 0.66 0.31 0.90 0.44
0.11 0.54 0.74 0.40 0.86 0.40
0.17 0.74 0.62 0.37 0.96 0.44
0.19 0.51 0.67 0.40 0.97 0.39
0.18 0.57 0.50 0.27 0.87 0.43
```

Sample Output 1

```
1
4
5
```

[ML] Identifying Spammers

Problem ID: spam

One issue that social platforms like Quora need to deal with is spam. A potential approach we could leverage to identify users producing spam (spammers) is to analyze the visit patterns. Normal users will tend to act differently from spammers, so this can help us identify and ban users producing this unwanted content. In this problem, your task is to determine which users are producing spam content given the list of `page_type:time` pairs from the users' visits to Quora. You will be given a training set of previous normal and spam user visit lists and labels of whether the user was a spammer. You need to predict the labels of users in a test set.

Notes

- The datasets for this problem are synthetically generated and **not** real user data. Therefore, they may not match intuitions on how spammers behave. In particular, we have modeled users as state machines with each `page_type` as a separate state. Normal users and spam users will have differing transition probability matrices, and the time each type of user spends on a page will be pulled from different distributions.
- To help test before submitting, a sample input is provided as an attachment to this problem.

Input

Your program will receive input from standard input.

The first row will be two space-separated integers n and m, representing the number of training and test samples.

Then, n lines will follow, each with up to v_{max} space-separated pairs of integers representing n users' visits to Quora. Let $p_{i,j} : t_{i,j}$ be the j-th pair of integers on the i-th of these n lines. This represents that user i's j-th visit to Quora was on page $p_{i,j}$ at time $t_{i,j}$.

Following that, there will be n lines, each with a single integer s_i representing whether the i-th training line was a spammer, where 1 represents that this user is a spammer, while 0 represents a normal user.

Finally, m lines of test data are provided, each with up to v_{max} space-separated pairs of integers, with the same format as the training data.

Output

Your program should write to standard output.

Print m lines representing the labels for each of the test users, where 1 represents that this user is a spammer, while 0 represents a normal user.

Constraints

- Train and test data are generated from the same distribution for each test case.
- $5 \cdot 10^3 \leq n \leq 3 \cdot 10^4$
- $10^3 \leq m \leq 5 \cdot 10^3$
- $0 \leq p_{i,j} \leq 12$
- $0 \leq t_{i,j} \leq 2^{31}$
- If $j_1 < j_2$, then $t_{i,j_1} < t_{i,j_2}$
- $v_{\text{max}} = 200$
- $0.05 \leq \frac{\sum_i s_i}{n} \leq 0.3$

Scoring

Score will be based on the F_1 score of predictions on the test set compared to the F_1 score of simply outputting all 1s.

$$F_1 = \frac{2}{\text{recall}^{-1} + \text{precision}^{-1}}$$

A 70% improvement over this baseline F_1 score will net full points, and scores in between will award points proportionally.

For example, if the dataset contains 30% positive labels, an F_1 score 0.4615 or below will award 0 points, while a score of 0.7846 or greater will score full points.

Sample Explanation

In this example, there are 11 training and 5 test users. There are two spam entries in the training set which are spammers, the 4th and 5th users. In the test set, a single user is a spammer, the 5th user. In this simple example, the spammers are easily identified as they tend land on a different page initially (page type 1).

Sample Input 1

```
11 5
2:0 2:14669 3:29642 1:41182 1:48298 3:53637
3:0
2:0 3:14874
1:0 0:0 1:5747
1:0 3:2049 2:12642 2:30991 2:49382 2:67816 0:86316 3:92663 0:102103 3:109224
2:0
2:0 3:14640
2:0 0:14974 1:17997 2:28943 0:43528 2:53765 1:68575 3:78493 1:89310 3:100206 3:110391
2:0 0:14995
2:0 1:14685 3:21289 1:32281 3:39379 2:52504 2:67318 1:82063 2:88576 1:103469
2:0 2:14872 2:29748 0:44401
0
0
0
1
1
0
0
0
0
0
0
2:0 1:18415 3:27095 2:37543 0:55935
3:0 2:7977 2:22721 0:37463 2:45641 2:60371 0:75187 2:82992 0:97862 1:102783 3:110944
3:0 2:11921 0:26812 2:29149
2:0
1:0 3:2794 1:13316 3:23960 0:33973 0:41125
```

Sample Output 1

```
0
0
0
0
1
```