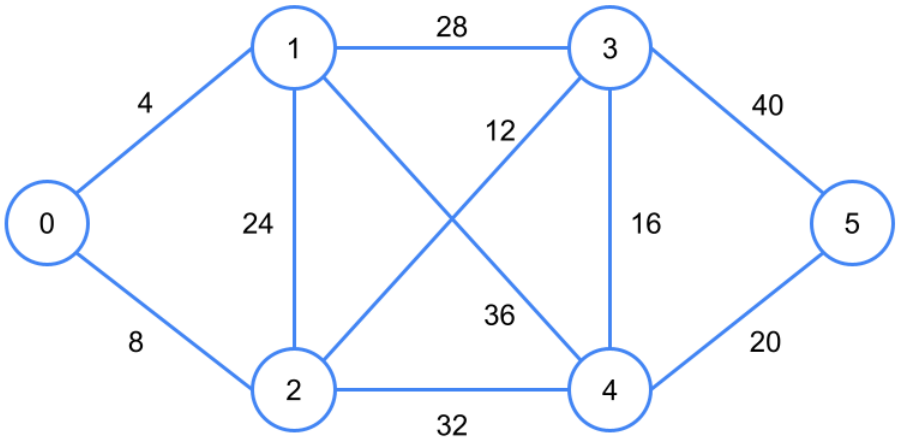
資料結構 Data Structures
Spring 2021
10920EECS204001
Professor Yi-Shin Chen

Final Exam (D)

1. [8 points] Given the following graph, use **Dijkstra's Algorithm** to find the shortest path and the shortest distance from vertex 0 to vertex 5. Write down the process step by step; otherwise, no points will be awarded.



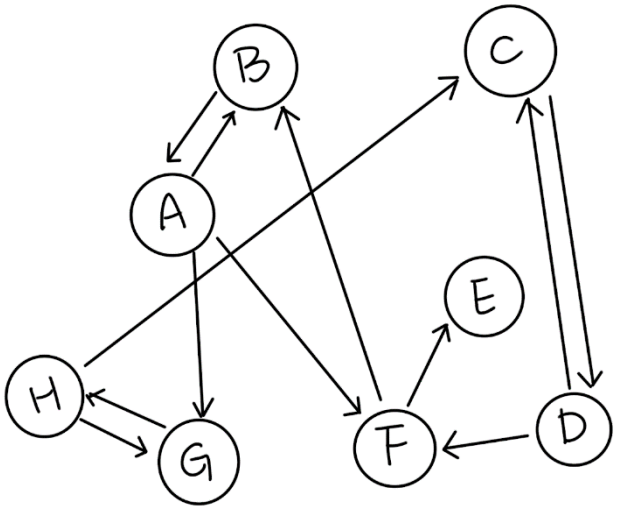
2. [10 points] Given the following sequence, sort the sequence with **Quick Sort** into ascending order. Write down the process and highlight the pivot step by step; otherwise, no points will be awarded.

91, 74, 59, 46, 35, 26, 19, 14, 11

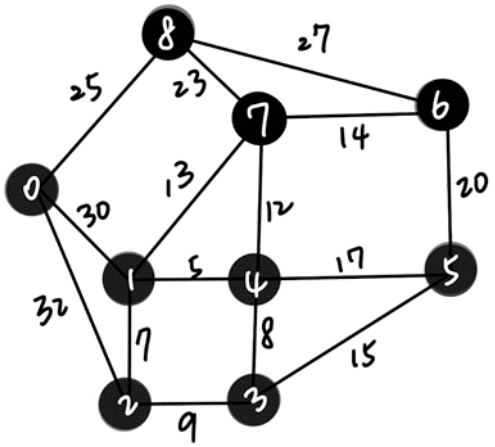
3. [7 points] Construct an **AVL Tree** given the following insertion sequence. Write down the process step by step; otherwise, no points will be awarded.

40, 15, 19, 6, 18, 3, 5, 17, 60, 47

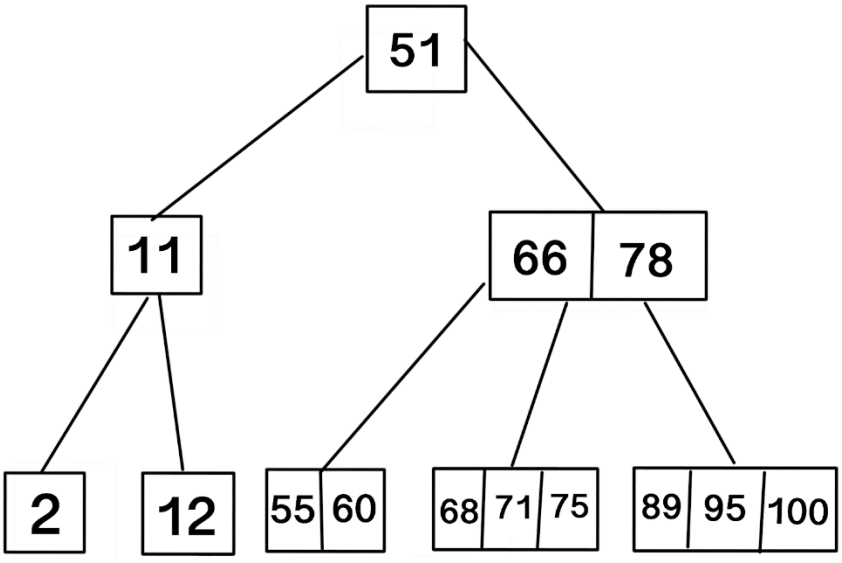
4. [5 points] Given the following graph, starting from node A, write down the corresponding traversal order of Depth-first search (DFS) and Breadth-first search (BFS). If there are multiple unvisited vertices, visit them in lexicographic order.



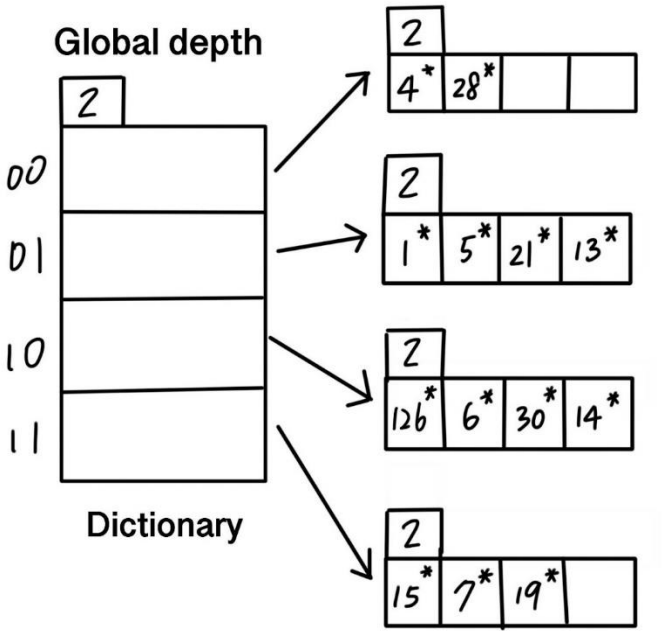
5. Karen wants to construct a minimum-cost spanning tree (MST). Karen recalls that she has learned **Kruskal's algorithm**, **Prim's Algorithm** and **Sollin's Algorithm** in classes. Answer the following questions.
   a. [5 points] These 3 different algorithms are based on the greedy approach. Explain the concept of what is a greedy algorithms.
   b. [10 points] Construct the MST using **Prim's algorithm** and **Sollin's algorithm** (starting from vertex 0 if necessary). Write down the process step by step; otherwise, no points will be awarded.



6. [5 points] Consider the B-tree of order 51 below. Delete the node with value 12 and write down the deletion process step by step; otherwise, no points will be awarded.



7. [5 points] Consider the following dynamic hashing (extendible hashing) example with hash function $\mathrm{h}(\mathrm{key}) = \mathrm{key} \,\%\, 2^{\mathrm{d}}$, where d is the global depth, and $\%$ is the modulo operation. Draw the outcome of the hash table after inserting 29 and 62.



8. Consider the data about public companies (some examples are listed below):

Ticker Symbol	Company Name	Market Capital (Billion)	Revenue Growth Rate (YoY)	Price-to-Earnings-Ratio (P/E ratio)	Selected In S&P 500 Index
NASDAQ: AAPL	Apple Inc.	2166	54	28.05	Yes
NASDAQ: MSFT	Microsoft Corp.	1872	19	33.88	Yes
NASDAQ: SBUX	Starbucks Corp.	133.3	-10.6	133.33	Yes
NYSE: T	AT&T Inc.	209.6	-3.5	N/A	Yes
NYSE: PLTR	Palantir tech Inc.	43.14	49	N/A	No
NYSE: NET	Cloudflare Corp.	26.06	110	N/A	No

Assume now you received the data about all public companies from an agent. You need to construct your own investment portfolio (the choices of companies to invest in). **Justify** your answers to the following questions.

a. [4 points] Assume that you would like to choose the companies ranking top 5 based on market capital. Also, these companies should be selected in S&P 500 Index. Therefore, the computer is capable enough to store all the relevant data in the RAM. Design an algorithm and write down the pseudo code to find such companies. Note that the table values are not sorted, and you just need to use a few statements to express your ideas.
b. [8 points] Assume that you would like to choose the companies ranking in the top 25% based on P/E ratio. Also, these companies should be selected in the S&P 500 Index. Therefore, the computer is capable enough to store all the relevant data in the RAM. Design an algorithm and write down the pseudo code to find such companies. Note that you might need to apply concepts about sorting in your algorithm, and you just need to use a few statements to express your ideas.
c. [8 points] Assume that you would like to choose the companies ranking in the top 1% based on revenue growth rate. In this time, you should consider ALL the data in the table. Therefore, the computer is NOT capable enough to store all the relevant data in the RAM. Design an algorithm and write down the pseudo code to find such companies. Note that you might need to apply concepts about sorting in your algorithm, and you just need to use a few statements to express your ideas.

9. The Emperor of the Galactic Empire announced that each planet needs to pay tribute every year. The tributes will be collected by Trantor, the planet where the Emperor lives. On each tribute, the codename of the planet is printed. The following table shows the codename of the planets (Galaxy_ID, Star_Cluster_ID, Planetary_System_ID, Planet_ID) on the tributes. We assume that the following table only shows a small part of the planets in the Galactic Empire.

Planet_Name	Galaxy_ID	Star_Cluster_ID	Planetary_System_ID	Planet_ID
Alpha-303	0	6	7	1
Earth	10	9	0	3
Beta-404	4	9	7	1
Beta-957	3	1	4	5
Siwenna	6	1	5	10
Terminus	10	9	3	3
Gamma-007	4	7	1	5
Mars	10	9	0	10
Ogma	3	3	4	3
Aurora	2	4	6	1

All the tributes need to be checked and sorted by the military before being sent to the palace. The military utilized codenames to sort the tributes, where the Planet_ID is the least significant and the Galaxy_ID is the most significant. After collecting all the tributes, the sorting process is shown as follow:

1) Sort the tributes in ascending order based on the Planet_IDs.
2) Sort the tributes in ascending order based on the Planetary_System_IDs.
3) Sort the tributes in ascending order based on the Star_Cluster_IDs.
4) Sort the tributes in ascending order based on the Galaxy_IDs.

For example, given the codenames of three planets p_1, p_2, and p_3:
p_1: (Galaxy_ID, Star_Cluster_ID, Planetary_System_ID, Planet_ID) = (1,3,9,9)
p_2: (Galaxy_ID, Star_Cluster_ID, Planetary_System_ID, Planet_ID) = (9,3,0,1)
p_3: (Galaxy_ID, Star_Cluster_ID, Planetary_System_ID, Planet_ID) = (1,4,9,1)

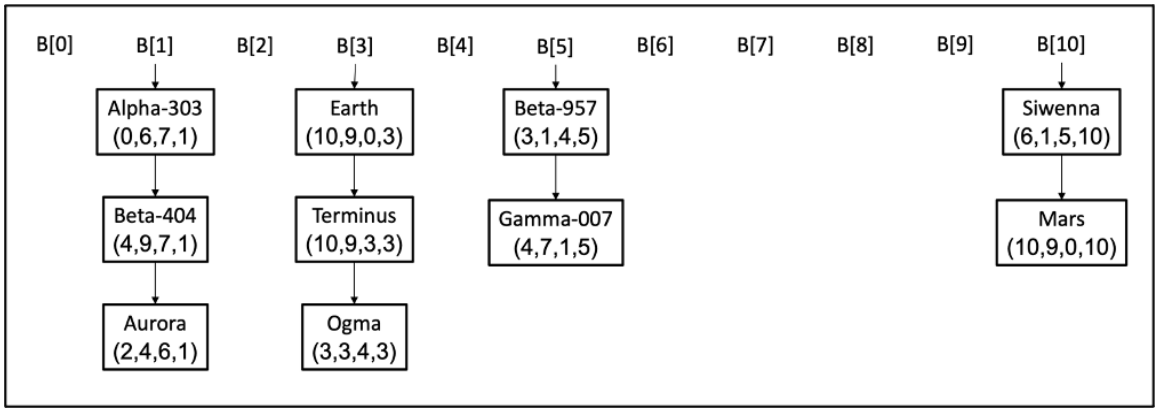
The sorting result will be: p_1(1,3,9,9), p_3(1,4,9,1), and p_2(9,3,0,1)

a. [10 points] Now, Hober Marlow, the sergeant who is responsible for managing the tributes, wants to choose sorting algorithms that can be utilized in the sorting scenario mentioned above. The candidate algorithms are listed below. Help Hober to decide which algorithm(s) is/are applicable. Also, elaborate the reasons why other algorithms are not suitable for the scenario mentioned above.

(A) Insertion sort
(B) Bubble sort
(C) Quick sort
(D) Heap sort
(E) Merge sort

b. [10 points] Hober Marlow has just learned the radix sort, and he tried to practice the algorithm by sorting the codenames shown in the table above. However, he got stuck after finishing the first round of sorting (Planet_ID). Please follow the schema below to help him finish the sorting. Note that you need to show the round-by-round process and the final results; otherwise, no points will be awarded.

The status of the buckets after first round of sorting (Planet_ID):



Result of the first round of sorting (Planet_ID):

Alpha-303 (0,6,7,1)	Beta-404 (4,9,7,1)	Aurora (2,4,6,1)	Earth (10,9,0,3)	Terminus (10,9,3,3)	Ogma (3,3,4,3)	Beta-957 (3,1,4,5)	Gamma-007 (4,7,1,5)	Siwenna (6,1,5,10)	Mars (10,9,0,10)

10. Instagram is a social media where the users can follow other users and see their posts. We can use a graph to model the social structure on Instagram. In the graph, the nodes represent the users, while the directed edges represent the “following” relationship. For example, a directed edge from User 0 to User 1 means that User 0 is following User 1 on Instagram. Now, given the table that records the following activities of users, answer the questions below.

Following_Record_ID	User_ID (Follower)	User_ID (Followee)	Timestamp
0	0	2	2019-02-12 04:12:34
1	3	0	2019-02-12 04:12:50
2	3	2	2019-02-12 04:13:01
3	2	0	2019-02-12 04:20:20
4	1	3	2019-02-12 04:24:30
5	0	2	2019-02-12 04:25:08
6	2	0	2019-02-12 04:30:28
7	1	3	2019-02-12 04:31:59
8	0	3	2019-02-12 04:48:12
9	2	3	2019-02-12 04:49:20
10	3	4	2019-02-12 04:58:33
11	4	5	2019-02-12 05:15:40
12	5	2	2019-02-12 05:18:22
13	2	5	2019-02-12 05:25:35
14	5	5	2019-02-12 05:29:39

a. [5 points] In the table, the Timestamp column keeps the time of the following activities. Each record means that User_ID(Follower) follows User_ID(Followee) at the specific timestamp on Instagram. According to the activities shown in the table, we can construct a graph to model the social structure. However, there are some records that violate the restrictions of a graph we mentioned in class. Write down the Following_Record_IDs of the invalid records first; then, draw the valid graph of the remaining valid records. (The valid graph is termed SW_Graph from now on.)

b. [5 points] Draw all the complete subgraphs in SW_Graph.
c. [5 points] Draw all the strongly connected components in SW_Graph.

11.

a. [5 points] What is/are the benefit(s) of the balanced binary search tree as compared to normal binary search tree? Illustrate with examples.
b. [10 points] Consider the E-Commerce Shipping Data listed below to answer the following question.

Above is the distribution of some attributes in the data set. Build an index for each of the attributes listed above with the most appropriate indexing technique and **justify** your answers. The possible candidates are: **static hashing**, **dynamic hashing**, and **B-Trees**. (You can bundle attributes with similar characteristics and justify the most appropriate indexing technique.)