CS 5800-Spring 20: Final Exam 04/20/20 5:30 PM PST - 9:30 PM PST

1. Complexity Theory (5*2=10 pts.)

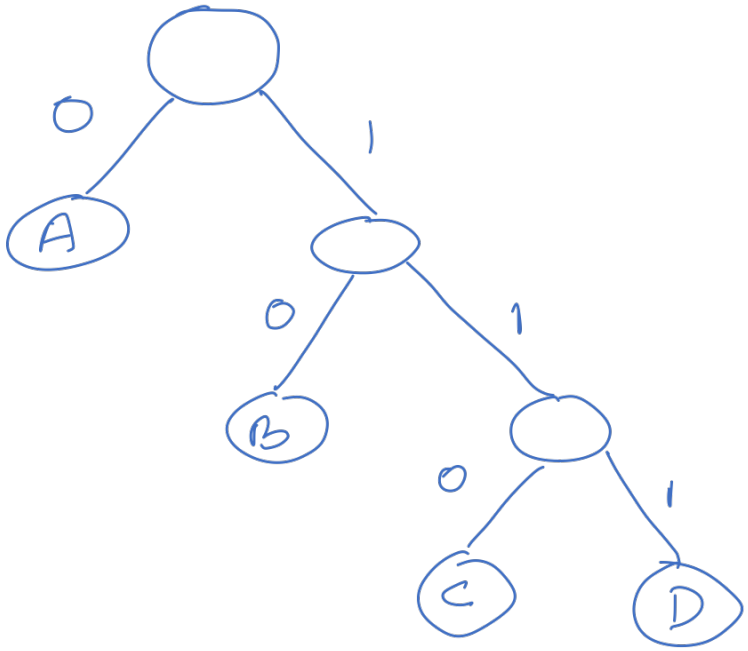
For each of the statements before, describe if the statement is TRUE, FALSE or OPEN PROBLEM. For full credit, explain your answer in 1-2 sentences.

1. Every problem in NP is either in P or is NP-Complete
2. For every optimization problem, there is an equivalent decision problem
3. NP stands for "Not Polynomial" time algorithm
4. P = NP
5. The shortest path problem is in NP

2. Tax Calculator (17+3=20 pts.)

You started a new job at Internal Revenue Services (IRS). As part of your job as IRS algorithm administrator, you are given a giant list of tax bills (two columns: SSN, Tax_Bill-Y/N). You are also given a list of payable checks made against these bills (two columns: SSN, Check_Paid-Y/N). Your task is to find out which bills do not have a checks paid against them. Describe an efficient algorithm to perform this task. Also provide the run time of the algorithm. You can assume you are memory constrained so you CANNOT load all the data in memory at once but only in smaller K chunks. Note: you either need to describe the algorithm in short sentences or just provide a high level pseudo code.

3. Optimal Coding (15 pts.)



You are given a variable length encoding of four alphabets A, B, C, D in the encoding shown above. The relative frequencies of each of the alphabet is as follows:

Symbol	Frequency
A	60%
B	25%
C	10%
D	5%

What is the % average savings in the total number of bits used using the variable length encoding as compared to a fixed 2-bit encoding for each symbol?

4. The coolest messaging App (30 pts.)

Nikke is a computer science researcher in Ibadan, Nigeria. She wants to build a new messaging app for people in her country. She knows that building a fast messaging app requires strong cellular network. However, building a strong cellular network is also very expensive. Nike realizes that there are a large number of people who use cellphones in Nigeria and neighboring countries. She wonders - what if I build a strong network of cellphones itself (instead of constructing cellphone towers)? What if her new App sends messages hopping from one phone to another nearby phone until it reaches its final destination? Now to send the messages in the fastest possible way, all Nike has to do is to figure out a shortest route from the sender to the receiver of the message.

Write an algorithm / pseudocode to print out the shortest route for each message sender and recipient pair. For this problem, let us assume you don't need to worry about privacy since messages are heavily encrypted. Let us also assume that cost/ length of each hop of the message is constant.

Example:
There are 3 inputs to the algorithm: network, sender, recipient. Let us suppose the network of users which is input to the algorithm is below:

Input:

```
network = {
   'Min'       : ['William', 'Jayden', 'Omar'],
   'William' : ['Min', 'Noam'],
   'Jayden'   : ['Min', 'Amelia', 'Ren', 'Noam'],
   'Ren'       :  ['Jayden', 'Omar'],
   'Amelia'   : ['Jayden', 'Adam', 'Miguel'],
   'Adam'     : ['Amelia', 'Miguel', 'Sofia', 'Lucas'],
   'Miguel'   : ['Amelia', 'Adam', 'Liam', 'Nathan'],
   'Noam'     : ['Nathan', 'Jayden', 'William'],
   'Omar'     : ['Ren', 'Min', 'Scott'],
   ...
}

sender = 'Jayden'
```

recipient = 'Adam'

Output:

Shortest route possible between sender and recipient is: [‘Jayden', 'Amelia', 'Adam']

5. Web Search Engine (25 pts.)

You need to design a large scale web search engine. Few use cases you need to take care of are:

(A) Handling phrase queries & multi-word queries

(B) Apply Boolean logic (AND / OR) to compute relevance.

Discuss what specific choices you will make in designing the index for the search engine. Describe storage trade-off for your design and discuss potential improvements for managing storage constraint. Also discuss how different choices of Boolean relevance can impact the run time performance of your query.