

## Algorithms and Data structures Final assignment

Khan Inan

Due: Dec 19<sup>th</sup>, 11:59pm

1. What does it mean when an algorithm is P or NP complete? (10pts)

Algorithms that are labeled as P are ones that are associated with problems that can be solved by a deterministic approach. NP is the opposite where, they are non-deterministic. In layman's terms, P is a set of easy problems, while NP is difficult ones. The term  $P=NP$  is implying that hard problems have easy solutions

2. What is the difference between Big O, Big Omega, and Big Theta functions? (10pts)

In terms of asymptotes, Big O is the upper asymptotic bound and Big Omega is the lower asymptotic bound. Big Theta is in between these two. Big O and Big Omega each represent the opposite extremes of how long it can possibly take to complete a particular algorithm, while Big Theta can be described as the "average" case for how long it takes

3. Explain the difference between an adjacency list and an adjacency matrix. How do they store networks and in what scenarios is one preferred over the other? (20pts)

An adjacency list is an array that has the address of all the other lists that are contained within. The first node are the vertices. An adjacency matrix is a 2d array that has the dimensions of  $V/V$ . For when you would use one or the other, you would use an adjacency list when the data is sparse and there are gaps in between points, and you would use a matrix if the data set is very dense.

4. Give a summary of BFS and DFS. Make sure to explain the difference between them. What is the time complexity of both? Explain why this is the case. (20pts)

A BFS is a breadth first search which is used to find the shortest path in a graph. It follows the rule that the first piece of data that enters is the first one that comes out, and as a result it is slower than DFS. A DFS or Depth first search has its "children" visited before the siblings if you visualize it as a family tree. A BFS would not be that useful for decision trees that are present in games and puzzles, compared to a DFS which explores all the paths and when it finds a confirmation it stops.

5. In Lecture 7 we discussed different algorithms for string matching. Describe one and give an example of how it can be used in genomics research today. (20pts)

One example would be EPMA which is Efficient Pattern Matching algorithm. This method uses 2-bit binary code that has a fixed length. It uses multiple searcher agents at the same time and the results that are obtained are confirmed by comparing with experimental results. The way that this can be used in Genomics research today is in DNA sequence pattern matching which can benefit from this very efficient method that simplifies the extremely long length that DNA patterns usually have

6. In Lectures 10 and 11 we discussed different algorithms for sorting. Describe one and give an example of how it can be used in genomics research today. (20pts)

One example of a sorting algorithm is a quick sort which is also coincidentally one of the most popular sorting algorithms. This sorting algorithm is useful because it doesn't make any unnecessary predictions about the data that you are working with. In terms of genomic research, a quick sort can be used to find a cluster or a snippet of a sequence that several others also have in common.