CS 212 Homework 2

Due 11:59PM on Tuesday, October 11, 2022.

Each question carries 5 points. Recall that every question here requires a proof. You can discuss the questions in groups of 3 (please list other students you discussed the problems with). Please remember that the assignment solutions need to be submitted individually and not as a group (and written/ typset with no collaboration). Solutions are only accepted in PDF format. Also please be as clear and legible as possible; any step that is ambiguous because of unclear writing will be interpreted as a mistake. See the Canvas page for instructions on how to submit.

Each question carries 5 points. Recall that every question here requires a proof. You can discuss the questions in groups of 3 (please list other students you discussed the problems with). Please remember that the assignment solutions need to be submitted individually and not as a group (and written/ typeset with no collaboration). Solutions are only accepted in PDF format. Also please be as clear and legible as possible; any ambiguous step because of illegible writing will be interpreted as a mistake.

Problem 1

Consider a set $S = \{1, 2, ..., n\}$. Prove that there is a bijection between the set

$$\{(S_1, S_2, S_3) | S_1 \subseteq S_2 \subseteq S_3 \subseteq S\}$$

and DNA sequences of length n. Here a DNA sequence can be expressed as a string of length n on the alphabet a, c, g, t (e.g., the string "acaagccagg" is a valid DNA sequence of length 10).

Note: For this problem, you need to specify the bijective map, and give a short explanation for why it is indeed a bijective function.

Problem 2

There are n websites $W = \{w_1, w_2, w_3, \ldots, w_n\}$, and k different content distribution servers (CDS) hosting subsets of the websites. In particular, the set $C_i \subseteq W$ denotes the websites hosted by the ith CDS. Note that the sets $C_1, \ldots, C_k$ could overlap. Let the quantity $m(C_1, C_2, \ldots, C_k)$ represent how many websites are hosted in total by these k different CDS servers (this notation indicates that it is a function of the sets $C_1, C_2, \ldots, C_k$). The question is to understand $m(C_1, \ldots, C_k)$ by proving upper and lower bounds.

(a) Express the number of websites hosted by the k CDS servers $m(C_1, C_2, \ldots, C_k)$ in terms of the cardinality of an appropriate set.

(b) Prove that $m(C_1, \ldots, C_k) \leq |C_1| + |C_2| + \ldots + |C_k|$.

(c) $m(C_1, \ldots, C_k) \geq |C_1| + \ldots + |C_k| - \sum_{1 \leq i < j \leq k} |C_i \cap C_j|$.

Hint: You can try to use induction to prove both parts. You may want to use part (b) to prove part (c).

Problem 3

Below is a list of 5 functions in terms of n. Sort these 5 functions into a list $f_1(n), f_2(n), \ldots, f_5(n)$ of increasing growth rate such that, for each $i \in [5]$, either $f_i(n) = o(f_{i+1}(n))$ or $f_i(n) = \Theta(f_{i+1}(n))$. For each i, explicitly state and prove the relationship between $f_i(n)$ and $f_{i+1}(n)$. Note that in this problem, $\log n = \log_2 n$ and $\ln n = \log_e n$.

$$n^{\log 10},\ (\log 10)^n,\ n!,\ \binom{n}{2},\ n^{2n},$$

Also $\binom{n}{r}$ is called a binomial coefficient and is given by $\binom{n}{r} = \frac{n*(n-1)*\ldots*(n-r+1)}{r*(r-1)*\ldots*1}$

Problem 4

Below is a list of 5 functions in terms of n. Sort these 5 functions into a list $f_1(n), f_2(n), \ldots, f_5(n)$ of increasing growth rate such that, for each $i \in [5]$, either $f_i(n) = o(f_{i+1}(n))$ or $f_i(n) = \Theta(f_{i+1}(n))$. For each i, explicitly state and prove the relationship between $f_i(n)$ and $f_{i+1}(n)$.

$$2^{\log n},\ 2^{3\log n},\ 2^{n \log n},\ 2^{\sqrt{\log n}},\ 2^n$$

Note for the last two questions: Recall the following definitions:
Little-oh: $f(n) = o(g(n))$ if $\lim_{n\to\infty} \frac{f(n)}{g(n)} = 0$.
Theta: $f(n) = \Theta(g(n))$ if $f(n) = O(g(n))$ and $g(n) = O(f(n))$.