

## Homework 2

1. Given: a graph  $G$   
 question: is  $G$  connected?

This problem corresponds to the language that puts all the string encodings of the positive instances of the problem into:

$\{ \langle G \rangle : G \text{ is a connected graph} \}$ , where  $\langle G \rangle$  is the string encoding of  $G$ . Please indicate the languages corresponding to the following four problems:

\*

Given: an NFA  $A$ ,

Question: Is there a word  $w$  such that  $A$  accepts  $w$ ?

$\{ \langle A \rangle : \text{there exists a word } w \text{ such that NFA } A \text{ accepts } w \}$   
 ↳ language includes the encodings of NFAs that accept at least one word

\* NFA Acceptance of Any word \*

Given: a number  $n$  and two primes  $p$  and  $q$ ,  
 question: is it the case that  $n = p \cdot q$ ?

language  $\{ \langle n, p, q \rangle : n = p \cdot q, \text{ where } p \text{ and } q \text{ are primes} \}$  ↳ where a number  $n$  is the product of two primes  $p$  and  $q$

• Number Factorization into two Primes

Given: a number  $n$ ,

question: is it the case that  $n = p \cdot q$  for some primes  $p$  and  $q$ ?

→  $\{ \langle n \rangle : n = p \cdot q, \text{ for some primes } p \text{ and } q \}$  ↳ includes the encoding of numbers that can be expressed as the product of two prime numbers

• Prime Factorization Check

\* Given: an NFA  $A$  and a word  $w$ , Question: Does  $A$  accept  $w$ ?  
 →  $\{ \langle A, w \rangle : \text{NFA } A \text{ accepts word } w \}$

• NFA Acceptance of a Specific word

2. prove why the following statements are true:

(1) Function  $2n^3 - 18n$  is  $O(n^3)$  and also it is  $O(n^4)$  but is it not  $O(n^2 \log n)$ .

(2) Function  $3n^2 2^{2n}$  is  $2^{O(n)}$ .

(1)

Function  $2n^3 - 18n$  is  $O(n^3)$  and  $O(n^4)$ , but not  $O(n^2 \log n)$

• Proving  $2n^3 - 18n$  is  $O(n^3)$ :

$f(n)$  is  $O(g(n))$  if there exist constants  $c$  and  $n_0$  such that  $0 \leq f(n) \leq c \cdot g(n)$  for all  $n \geq n_0$ .

Let  $f(n) = 2n^3 - 18n$  and  $g(n) = n^3$

If  $c=3$  and  $n_0=1$ , then  $2n^3 - 18n \leq 3n^3$  for all  $n \geq 1$

which means,  $2n^3 - 18n$  is  $O(n^3)$

• Proving  $2n^3 - 18n$  is  $O(n^4)$

$g(n) = n^4$

$c$  and  $n_0$  such that  $2n^3 - 18n \leq c \cdot n^4$  for all  $n \geq n_0$

$c=1$  and  $n_0=1$ , then  $2n^3 - 18n \leq n^4$  for all  $n \geq 1$

$n \geq 1 \rightarrow$  which means  $2n^3 - 18n$  is  $O(n^4)$

• Proving  $2n^3 - 18n$  is not  $O(n^2 \log n)$ :

$f(n)$  being  $2^{O(n)}$  means there's a  $g(n)$  that is  $O(n)$  such that  $f(n) = 2^{g(n)}$

$\hookrightarrow f(n) = 3n^2 2^{2n} \rightarrow 3n^2 2^{2n} = 2^{\log_2(3n^2) + 2n}$

$\hookrightarrow \log_2(3n^2)$  is  $O(n)$ , and  $2n$  is also  $O(n)$ , their sum is  $O(n)$ :

$\downarrow$

This means  $3n^2 2^{2n}$  is  $2^{O(n)}$ ,  $g(n) = \log_2(3n^2) + 2n$  is  $O(n)$

This would mean  $3n^2 2^{2n}$  is  $2^{O(n)}$

3.

### A. Importance of Algorithm M in Drug Development

→ Algorithm M, designed to compute a similarity metric between two protein molecules, could revolutionize medical drug development. Proteins are fundamental to biological processes, and their malfunction or abnormal behavior is often linked to diseases. By comparing protein structures, scientists can identify proteins with similar functions or binding sites. This type of information can be critical in drug discovery for several reasons which includes:

- Target identification: Understanding protein structure helps identify potential drug targets. Similar proteins might be implicated in similar diseases, suggesting new avenues for treatment.
  - Drug Repurposing: Drugs developed for one protein might be effective against another with a similar structure, speeding up drug discovery and reducing costs.
  - Personalized medicine: Similarity metrics will end up helping in designing drugs tailored to individual genetic profiles, especially in cases where slight protein variations influence drug effectiveness.
  - Understanding disease mechanisms: one way is comparing healthy and diseased protein structures, researchers can gain insights into disease mechanisms at a molecular level.
- Thus, Algorithm M has the potential to accelerate the discovery of new drugs and contribute to personalized and effective medical treatments.

## B. Data Structure for Representing protein molecules

→ Representing a protein molecule in computer memory is Challenging due to its complex structure. one efficient way is to use a Graph data structure. Each atom in the protein can be represented as a node, and the bonds between atoms as edges. This structure effectively captures the 3D nature of proteins, this would include the spatial relationships between atoms.

pros: Reflects molecular structure, analysis of protein folding interactions

cons: potentially large and complex, requiring a lot of computational power.

## C. Similarity metrics and Algorithms.

→ Structure similar metric, this would compare the structures of proteins. It could involve measuring the distance between atoms in protein graphs.

- pros, more accurate in identifying similarities

- cons, its computationally heavy

Sequence similarity metric: it would compare the amino acid sequences, it could compare Needleman or Smith algorithms.

The different approaches have trade offs. But while a structural approach is more thorough, its also very comp. intensive. The sequence approach is more thorough and 'straightforward' but could miss details. Both are critical to making discoveries in drug development