A11: PROVING ALGORITHM CORRECTNESS. GRAPHS AND TREES

You can graph human evolution, which is mostly a straight line, but we do get better and change over time, and you can graph technological evolution, which is a line that's going straight up. They are going to intersect each other at some point, and that's happening now. —Daniel H. Wilson

Course: CS 5002
Fall 2018
Due: Dec 7, 2018, Midnight

OBJECTIVES

After you complete this assignment, you will be comfortable with:

- Several proof techniques, including:
  - Proof by counterexample
  - Proof by induction
  - Proof by loop invariance
- Some basic tree terminology
- Some special trees
- Tree traversals
- Some basic graph terminology
- Graph representations

RELEVANT READING

Rosen:

- Chapter 5.1. Mathematical induction
- Chapter 5.2 Strong Induction and Well-Ordering
- Chapter 5.3 Recursive Definitions and Structural Induction
- Chapter 5.5. Program Correctness
- Chapter 11.1 Introduction to Trees
- Chapter 11.3 Tree Traversals
- Chapter 10.1 Graphs and Graph Models
- Chapter 10.2 Graph Terminology and Special Types of Graphs

NEXT WEEK'S READING

Rosen,

- Chapter 7: Discrete Probability

EXERCISES

Question 1

Let $P(n)$ be the statement that $1^2 + 2^2 + \cdots + n^2 = n(n+1)(2n+1)/6$ for a positive integer n.

(a) What is the statement $P(1)$?

(b) Show that $P(1)$ is true, completing the basis step of the proof.

(c) What is the inductive hypothesis?

(d) What do you need to prove in the inductive step?

(e) Complete the inductive step, identifying where you use the inductive hypothesis.

Solution $P(1)$ represents the basis case for the given statement, and it is expressed as follows:

$$P(1) := 1^2 = \frac{1(1+)(2 \cdot 1 + 1)}{6}$$

Truthfulness of statement $P(1)$ can easily be shown. We simply need to evaluate the right-hand side of the expression:

$$P(1) := 1^2 = \frac{1(1+)(2 \cdot 1 + 1)}{6} \rightarrow 1 = 1$$

The inductive hypothesis requires that for every positive integer n, expression:

$$1^2 + 2^2 + \cdots + n^2 = n(n+1)(2n+1)/6$$

holds true.

As a part of the inductive step, we need to prove that the given expression also holds true for $n + 1$.

To somplete the inductive step of the proof ot correctness for statement: $P(n+1) := \sum_{i=1}^{k+1} i^2 = \frac{(k+1)(k+2)(2k+3)}{6}$ we will focus on its left-hand side, and we will show that it is equal to the right hand side.

$$\begin{aligned}
\sum_{i=1}^{k+1} i^2 &= \underbrace{\sum_{i=1}^{k} i^2}_{\frac{k(k+1)(2k+1)}{6}} + (k+1)^2 \\
&= \frac{k(k+1)(2k+1)}{6} + (k+1)^2 \\
&= \frac{k(k+1)(2k+1) + 6(k+1)^2}{6} \\
&= \frac{(k+1)(2k^2 + k + 6k + 6)}{6} \\
&= \frac{(k+1)(k+2)(2k+3)}{6}
\end{aligned} \tag{1}$$

Question 2

Briefly explain what is wrong with the following "proof":

Theorem: For every positive integer n, if x and y are positive integers with max(x, y) = n, then $x = y$.

Basis step: Suppose that $n = 1$. If max(x, y) = 1, and x and y are positive integers, then we have $x = 1$ and $y = 1$.

Inductive step: Let k be a positive integer. Assume that whenever max$(x, y) = k$ and x and y are positive integers, then $x = y$. Now let max$(x, y) = k + 1$, where x and y are positive integers. Then max$(x - 1, y - 1) = k$, so by inductive hypothesis, $x - 1 = y - 1$. It follows that $x = y$, completing the inductive step.

The first wrong thing with this problem is the theorem itself . It simply does not hold true that if max(x, y) = n, then $x = y$ for all possible positive integers. The basis step is correct, but the theorem and the proof are still not. The problem with the proof happens in the inductive step, which involves a **circular argument**. Inductive hypothesis was made for some arbitrary $n = k$. Inductive step is taken for $n = k + 1$, so we cannot just jump back to $x - 1$ and $y - 1$.

Question 3

Prove that for every positive integer n:

$$1 \cdot 2 + 2 \cdot 3 + \cdot + n(n+1) = n(n+1)(n+2)/3$$

Basis step: For the basis step, let's assume that $n = 1$. Then it follows:

$$1 \cdot (1+1) = \frac{1(1+1)(1+2)}{3} \rightarrow 2 = 2$$

The basis step is satisfied, and we proceed to the inductive hypothesis.

Inductive hypothesis: The given expression:

$$1 \cdot 2 + 2 \cdot 3 + \cdot + n(n+1) = n(n+1)(n+2)/3$$

holds true for some positive integer k.

Inductive step: As a part of the inductive step, let's show that the given expression also holds for step $k+1$. We can write:

$$1 \cdot 2 + 2 \cdot 3 + \cdot + n(n+1) + (n+1)(n+2) = \frac{(n+1)(n+2)(n+3)}{3} \tag{2}$$

To show that the equation (2) holds, we focus on the left-hand side again, and manipulate it, to show that it is equal to the right-hand side:

$$\begin{aligned}
\sum_{i=1}^{k+1} i(i+1) &= \underbrace{\sum_{i=1}^{k} i(i+1)}_{k(k+1)(k+2)/3} + (k+1)(k+2) \\
&= \frac{k(k+1)(k+2)}{3} + (k+1)(k+2) \\
&= \frac{k(k+1)(k+2) + 3(k+1)(k+2)}{3} \\
&= \frac{(k+1)(k+2)(k+3)}{3} \qquad (3)
\end{aligned}$$

Question 4

Give a recursive algorithm for computing nx whenever n is a positive integer, and x is an integer, using just addition. Please provide pseudocode fro your algorithm.

To compute $n \cdot x$, we can just add n to sum initialized to 0, but such that we add it x times. A pseudocode that does that is given below.

```
multiplyByAddition(int n, int x)
  if (x = 0) || (n = 0)
    return 0;
  if (x = 1)
    return n;
  if (n = 1)
    return x;
  else
    return multiplyByAddition (x - 1, n) + n;
```

Question 5

Use merge sort to sort $b, d, a, f, g, h, z, p, o, k$ into increasing order. Show all steps used by the algorithm.

A graphical solution to this problem is given below, in Figure 1.

Question 6

Use quick sort to sort 3, 5, 7, 8, 1, 9, 2, 4, 6 into increasing order. Show all steps used by the algorithm.

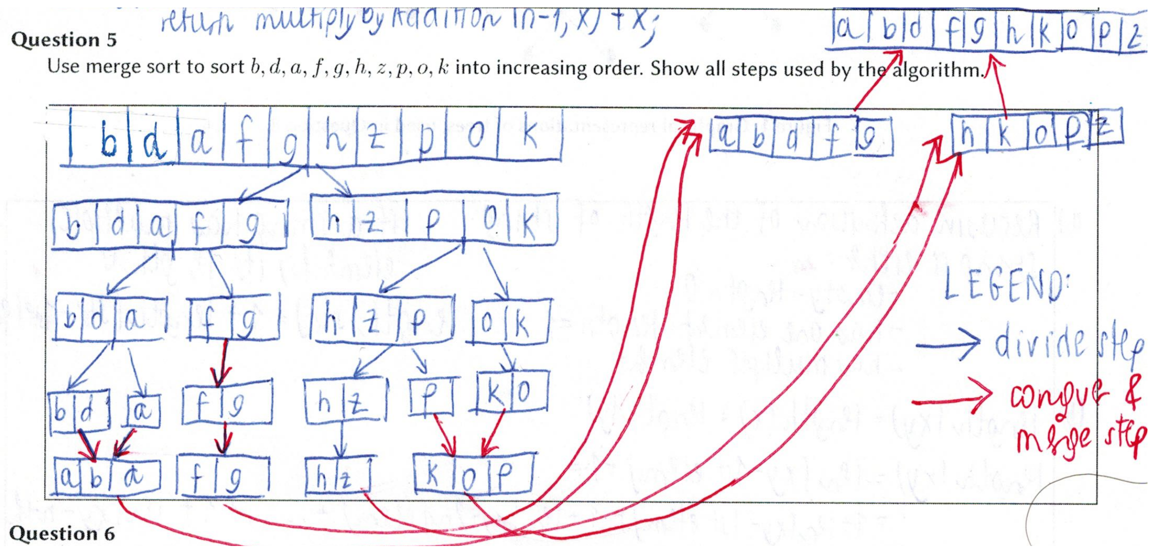

Figure 1: Graphical solution to Problem 5.

A graphical solution to this problem is given below, in Figure 2.

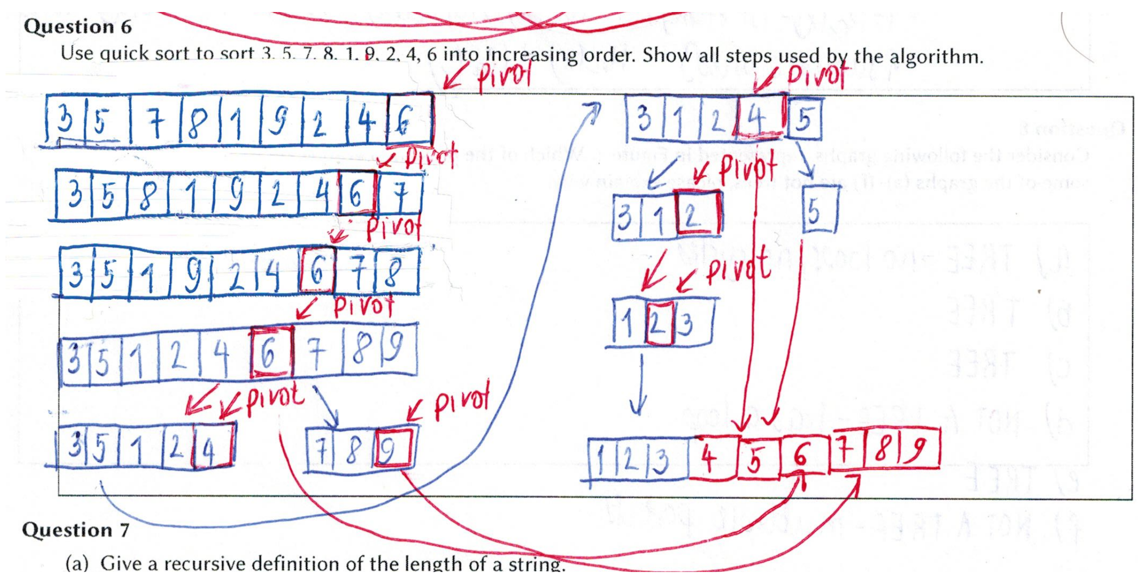

Figure 2: Graphical solution to Problem 6.

Question 7

(a) Give a recursive definition of the length of a string.

(b) Use the recursive definition from part (a) to prove that, given two strings x and y, it holds that $l(xy) = l(x) + l(y)$, where $l(\cdot)$ denotes the length of a string.

Let's consider some arbitrary string. With respect to its lenght, that string could be:

- Empty - its length is equal to 0
- Contain one element - its length is equal to 1

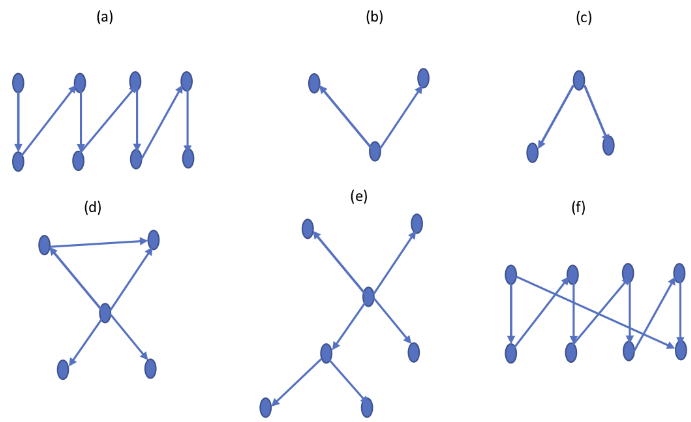

Figure 3: Graphical representations of trees, used in Question 8.

- Contain more than one element - it's length is equal to 1 + length of a substring that we get when we remove the first element from the original string.

So, taking these observations into account, we can define the length of a string as follows:

$$\text{length}(string) = \begin{cases} 0, \text{if empty string} \\ 1 + \text{length}(\text{substring}(string, 1)) \end{cases}$$

where $(\text{subset}(string, 1))$ represents a substring that we get by removing the first element from our original string.

Using the given recursive definition, we can now prove that, given two strings x and y, it holds that $l(xy) = l(x) + l(y)$, where $l(\cdot)$ denotes the length of a string as follows.

Let's assume that we have two arbitrary strings x and y, with some arbitrary lengths length(x) and length(y). Let's define their concatenation as a new string xy. The length of a string xy can now be found as:

$$\text{length}(xy) = \begin{cases} 0, \text{if empty string} \\ 1 + \text{length}(\text{substring}(xy, 1)) \end{cases}$$

Under the assumption that neither x nor y are empty, unrolling our recursion we get:

$$\begin{aligned} \text{length}(xy) &= \underbrace{\text{length}(\text{substring}(xy, 1)) + \text{length}(\text{substring}(\text{substring}(xy, 1), 1))}_{\text{length(x)}} \\ &+ \underbrace{\cdots + \text{length}(\text{substring}(\ldots(\text{substring}(\text{substring}(xy, 1), \ldots), 1))}_{\text{length(y)}} = \text{length}(x) + \text{length}(y) \end{aligned}$$

Question 8

Consider the following graphs, represented in Figure 3. Which of the presented graphs are trees? If you think that some of the graphs (a)–(f) are not trees, please explain why.

Graphs (a), (b), (c) and (e) represent **trees**, because it have a single parent, nad not loops or cycles. Graph (d) is not a tree because it has a loop, and graph (f) is not a tree because it has multiple parents.

Question 9

Draw graphs that have the following adjacency matrices:

(a) $\begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 2 \\ 1 & 2 & 0 \end{bmatrix}$

(b) $\begin{bmatrix} 0 & 2 & 0 \\ 2 & 1 & 2 \\ 0 & 0 & 1 \end{bmatrix}$

(c) Find adjacency lists that correspond to the adjacency matrices from parts (a) and (b).

Corresponding graphs and their adjacency lists are given below, in Figure 4.

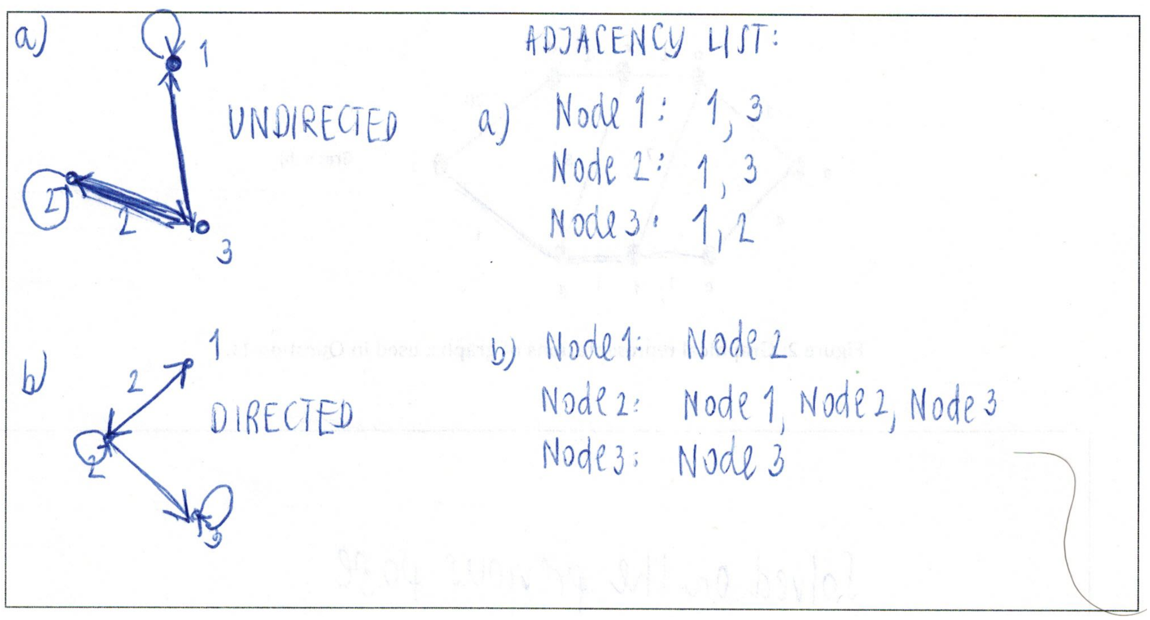

Figure 4: Graphical solution to Problem 9.

Question 10

(a) Represent the following expressions using binary trees, such that operands are leaves and operations are internal nodes:

- $(x + xy) + (x/y)$
- $x + ((xy + x)/y)$

Now, traverse the constructed trees using the following traversals:

(b) Pre-order traversal.

(c) In-order traversal.

(d) Post-order traversal.

Graphs representing the given arithmetic expressions are depicted in Figures ?? and ??.

The traversals are given as follows:

Graph 1:

- **Pre-order:** + – + – X – * – X – Y – / – X – Y

- **In-order:** X – + – X – * – Y – + – X – / – Y
- **Post-order:** X – X – Y – * – + – X – Y – / – +

Graph 2:

- **Pre-order:** + – X – / – + – * – X – Y – Y – Y
- **In-order:** X – + – X – * – Y – + – Y – / – Y
- **Post-order:** X – X – Y – * – Y – + – Y – / – +

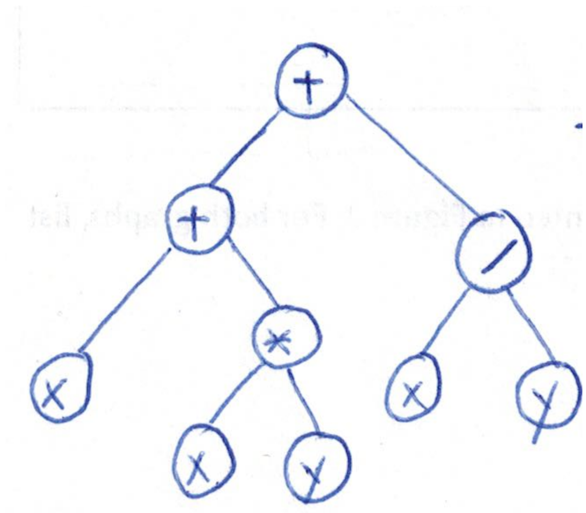

Figure 5: Graphical solution to Problem 10 (a).

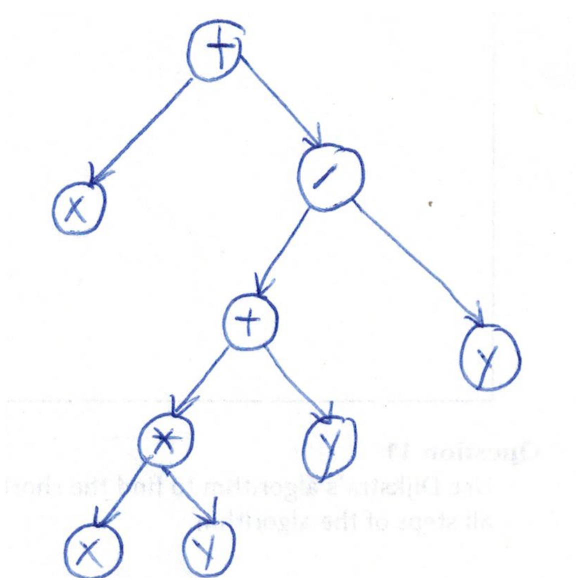

Figure 6: Graphical solution to Problem 10 (b).

Question 11

Use Dijkstra's algorithm to find the shortest path from a to z for graphs represented in Figure 7. For both graphs, list all steps of the algorithm.

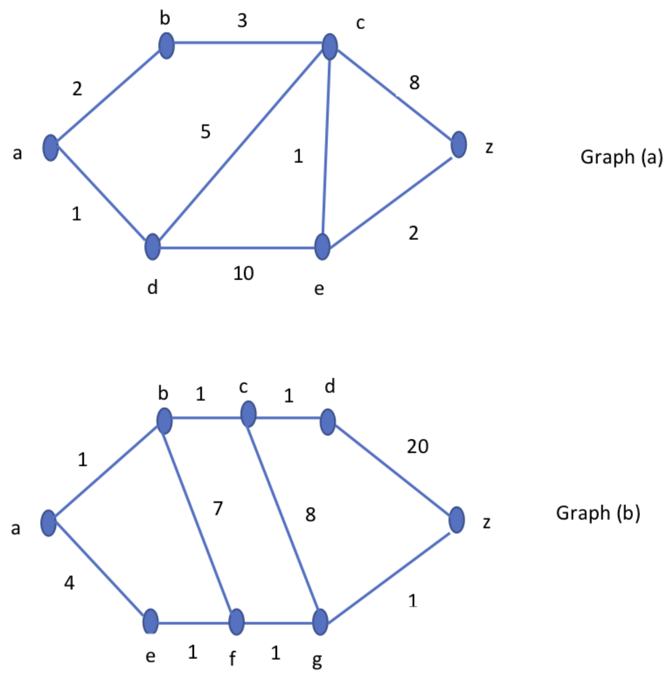

Figure 7: Graphical representations of graphs, used in Question 11.

The steps of the Dijkstra's algorithm for the first graph are presented below:

Step 1:

$Source$	Via	$Destintion$	$Cost$
A	A	A	0
A	A	B	2
A	A	C	∞
A	A	D	1
A	A	E	∞
A	A	Z	∞

(4)

Step 2:

$Source$	Via	$Destintion$	$Cost$
A	A	A	0
A	A	B	2
A	D	C	5
A	A	D	1
A	D	E	10
A	A	Z	∞

(5)

Step 3:

$Source$	Via	$Destintion$	$Cost$
A	A	A	0
A	A	B	2
A	B	C	3
A	A	D	1
A	D	E	10
A	A	Z	∞

(6)

Step 4:

$Source$	Via	$Destintion$	$Cost$
A	A	A	0
A	A	B	2
A	B	C	3
A	A	D	1
A	C	E	1
A	E	Z	8

(7)

Step 5:

$Source$	Via	$Destintion$	$Cost$
A	A	A	0
A	A	B	2
A	B	C	3
A	A	D	1
A	C	E	1
A	E	Z	2

(8)

The algorithm terminates after step 5, because no other reductions in cost are possible. From the table represented in step 5, we find the optimal path from node A to node Z to be a path $A--B--C--E--Z$, with the associated cost equal to 8.

The steps of the Dijkstra's algorithm for the second graph are presented below:

Step 1:

$Source$	Via	$Destintion$	$Cost$
A	A	A	0
A	A	B	1
A	A	C	∞
A	A	D	∞
A	A	E	4
A	A	F	∞
A	A	G	∞
A	A	Z	∞

(9)

Step 2:

$Source$	Via	$Destintion$	$Cost$
A	A	A	0
A	A	B	1
A	A	C	∞
A	A	D	∞
A	A	E	4
A	E	F	1
A	A	G	∞
A	A	Z	∞

(10)

Step 3:

$Source$	Via	$Destintion$	$Cost$
A	A	A	0
A	A	B	1
A	B	C	1
A	A	D	∞
A	A	E	4
A	E	F	1
A	A	G	∞
A	A	Z	∞

(11)

Step 4:

$Source$	Via	$Destintion$	$Cost$
A	A	A	0
A	A	B	1
A	B	C	1
A	C	D	1
A	A	E	4
A	E	F	1
A	C	G	8
A	A	Z	∞

(12)

Step 5:

$Source$	Via	$Destintion$	$Cost$
A	A	A	0
A	A	B	1
A	B	C	1
A	C	D	1
A	A	E	4
A	E	F	1
A	C	G	8
A	D	Z	20

(13)

Step 6:

$Source$	Via	$Destintion$	$Cost$
A	A	A	0
A	A	B	1
A	B	C	1
A	C	D	1
A	A	E	4
A	E	F	1
A	C	G	8
A	G	Z	1

(14)

The algorithm terminates after step 6, because no other reductions in cost are possible. From the table represented in step 5, we find the optimal path from node A to node Z to be a path $A--E--F--G--Z$, with the associated cost equal to 7.

PROBLEMS

Question 12

Using mathematical induction, prove that 2 divides $n^2 + n$ whenever n is a positive integer.

Basis step: To prove the basis step, we set $n = 1$, and get:

$$1^2 + 1 = 2 \bmod 2 \equiv 0$$

Therefore, the given statement holds true for the basis step.

Inductive hypothesis: In the inductive hypothesis, we assume that for every positive integer n it holds that:

$$n^2 + n \bmod 2 \equiv 0 \tag{15}$$

Inductive step: In the inductive step, we chose some $n + 1$, and we want to prove that:

$$(n+1)^2 + n + 1 \bmod 2 \equiv 0 \tag{16}$$

Equation (16) can now be rewritten as:

$$(k+1)^2 + k + 1 = k^2 + 2k + k + 2 = \underbrace{k^2 + k}_{\text{inductive hypothesis}} + 2(k+1) \tag{17}$$

Using the properties of modular arithmetic (please check earlier lecture notes on number theory again, where we talked about linear combinations under modular arithmetic), we can now examine individual summands as follows:

$$\begin{aligned} &k^2 \bmod 2 \rightarrow \text{divisible my inductive hypothesis} \\ &2(k+1) \bmod \rightarrow \text{divisible by definition} \end{aligned} \tag{18}$$

Therefore, it follows that the whole expression is divisible by 2, and therefore the statement holds for step $n+1$. This completes the proof.

Question 13

Prove that $f_1 + f_3 + f_3 + \cdots + f_{2n-1} = f_{2n}$.

Let's start proving the given expression by recalling the initial conditions for the Fibonacci sequence: $f_1 = f_2 = 1$.

With this assumption in mind, we can prove the given statement using **mathematical induction**:

Basis step: Let's set $n = 2$. We know that:

$$\begin{aligned} f_1 &= f_2 = 1 \\ f_3 &= f_1 + f_2 = 1 + 1 = 2 \\ f_4 &= f_2 + f_3 = 1 + 2 = 3 \end{aligned} \tag{19}$$

We can now write:

$$\begin{aligned} f_1 + f_{2n-1} &= f_{2n} \\ f_1 + f_3 &= f_4 \\ 1 + 2 &= 3 = f_4 \end{aligned} \tag{20}$$

Thus, the expression holds for the basis step.

Inductive hypothesis: Let's assume that the given statement holds for every n:

$$f_1 + f_3 + f_3 + \cdots + f_{2n-1} = f_{2n} \tag{21}$$

Inductive step: Let's prove that the given statement also holds for $n = k+1$. We can write:

$$f_1 + f_3 + f_3 + \cdots + f_{2n-1} + f_{2n+1} = f_{2n+2} \tag{22}$$

Using inductive hypothesis, equation (22) can be rewritten as:

$$\begin{aligned} \underbrace{f_1 + f_3 + f_3 + \cdots + f_{2n-1}}_{f_{2n}} + f_{2n+1} &= f_{2n+2} \\ f_{2n} + f_{2n+1} = f_{2n+2} \end{aligned} \tag{23}$$

Equation (23) holds, however, by the defintion of the Fibonacci sequence. This equation also completes the proof.

Question 14

A chain letter starts when a person sends a letter to five others. Each person who receives the letter either sends it to five other people who have never received it, or does not send it to anyone. Suppose that 10000 people send out the letter before the chain ends, and that no one receives more than one letter. How many people receive the letter, and how many do not send it out?

To solve this problem, let's list out the simplifying assumptions given in the problem:

- The only person who sends the letter, but does not receives it first is the originator of this chain letter.
- A person who receives the letter either sends it to five other people who have never received it, or does not send it to anyone.
- No one receives more than one letter.

We can now start thinking about this problem as a tree:

- The originator of the chain letter is the root,
- The people who forward the letter are the internal nodes,
- The people who forward the letter, forward it always to their five children,
- Those nodes that receive the letter, but do not forward it, are the leaves.

So, the question of how many people do not forward the letter boils down to how many leaves are there in a 5-ary tree with 10000 internal nodes?

Using online resource G-Fact 11 (available here: https://www.geeksforgeeks.org/g-fact-42/), we know that in any n-ary tree, in which every node has either 0 or n children, the number of leaves equals to:

$$L = (n - 1) * i + 1 \tag{24}$$

where i represents the number of internal nodes.

Using equation (24), we now know that the number of people who have received the letter, but did not forward it equals:

$$L = (n - 1) * i + 1 = 40001 \tag{25}$$

Combining all of the internal nodes, with all of the leave nodes, we know that the total number of people who have received the letter equals to 50001.

Question 15

Prove that if some adjacency matrix A is an $m \times m$ symmetric matrix, then A^2 is also symmetric.

Note: This is a simplified proof, shown on a 2×2 matrix. The proof requires a bit more work to prove that the given statement holds for a matrix of an arbitrary size.

Let's consider some arbitrary 2×2 matrix A, given as:

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \tag{26}$$

Since the matrix is said to be symmetric, we know that $a_{12} = a_{21}$. Therefore, we can rewrite it as follows:

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{bmatrix} \tag{27}$$

When we now take the given matrix, and multipy it by itself, we can write:

$$A \cdot A = A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = \begin{bmatrix} a_{11}a_{11} + a_{12}a_{12} & a_{11}a_{12} + a_{12}a_{22} \\ a_{11}a_{12} + a_{12}a_{22} & a_{12}a_{12} + a_{22}a_{22} \end{bmatrix} \tag{28}$$

Clearly, the given resulting matrix is still symmetric. Therefore, the given statement holds for 2×2 matrices.

Question 16

Show that a simple graph is a tree if and only if it is connected, but the deletion of any of its edges produces a graph that is not connected.

Please note that this is an **if and only if theorem**, so we need to prove two direction:

- $\rightarrow$ If simple graph that is a connected, but deletion of any edge produces a graph that is not connected, is a tree.
- $\leftarrow$ A tree that is connected, but a delection of any edge procudes a graph that is not connected is a simple graph.

We can prove both statements by contradiction. Let's show that now.

- $\rightarrow$ Let's assume there exists a simple graph that is connected, but deletion of any edge produces a graph that is not connected, that **is not a tree**. If such a graph is not a tree, then it either isn't acyclic, and/or some node has mutltiple parents. Let's examine those conditions now:
  - If there exists a node that has multiple parents, then erasing some of edges incident to it would not create a disconnected graph. But that is contradiction, because we assumed otherwise.
  - If the graph is not acyclic, there there exists at least one node that can be reached in more than one way. Therefore, there exists at least one node where erasing edges incident to it does not render the graph disconnected. We reached a contradiction again.

  Therefore, it follows that every simple graph that is connected, but deletion of any edge produces a graph that is not connected is a tree.
- $\rightarrow$ Let's assume there exists a tree that is connected, but a deletion of any edge produces a graph that is not connected, but that such a tree is not a simple graph. If such a tree is not a simple graph, then it allows loops and multiple edges between some two nodes. However, if multiple edges are allowed, then our assumption about being able to render such a graph disconnected with a deletion of an edge does not hold. Therefore, a tree that is connected, but a deletion of any edge produces a graph that is not connected, but that such a tree is a simple graph.

Question 17

Suppose there exists an integer k such that every man on a desert island is willing to marry exactly k of the women on the island, and every woman on the island is willing to marry exactly k of the men. Let's also assume that a man is willing to marry a woman if and only if she is willing to marry him. Show that it is possible to match men and women on the island so that everyone is matched with someone that they are willing to marry.

The given problem is a simplified version of the **stable matching problem**. A stable matching problem is the problem of finding a stable matching between two equally sized sets of element, given a preference for every element, where matching is defined as mapping of the elements from one set of the other.

You can find more information about the stable matching problem here: https://www.cs.princeton.edu/ wayne/kleinberg-tardos/pdf/01StableMatching.pdf.

Typically, the stable marriage problem can be stated as: given n men and n women, where each person has ranked all members of the opposite sex in order of preference, marry the men and women together such that there are no two people of opposite sex who would both rather have each other than their current partners. When there are no such pairs of people, the set of marriages is deemed stable.

An intuitive method that guarantees to find a stable matching is known as the **Gale-Shapley algorithm**, and its pseudocode is given below (pseudocode source: https://www.cs.princeton.edu/ wayne/kleinberg-tardos/pdf/01StableMatching.pdf.).

```
INITIALIZE M to empty matching.
WHILE (some man m is unmatched and has not proposed to every woman)
  s <-- first woman on m's list to whommh has not yet proposed.
  IF (s is unmatched)
    Add m-s to matching M.
  ELSE IF (s prefers m to current partner m')
    Replace m'-s with m-s in matching M.
  ELSE
    s rejects m.
RETURN stable matching M.
```

Question	Points	Score
Simple Proof By Induction	6	
Lame Proof	4	
Another Proof by Induction	4	
Simple Recursive Algorithm	3	
Merge Sort	2	
Quick Sort	2	
Strings and Recursion	5	
Trees	3	
Adjacency Matrix	4	
Binary Trees	10	
Dijkstra's Algorithm	6	
Mathematical Induction	6	
Mathematical Induction	4	
Trees and Proofs	10	
Adjacency Matrix and Proofs	10	
Connected Graphs and Treees	10	
Marriage Problem	11	
Total:	100	

SUBMISSION DETAILS

Things to submit:

- Submit the following on Blackboard for Assignment 11:
  - The written parts of this assignment as a .pdf named "CS5002_[lastname]_A11.pdf". For example, Ben Bitdiddle's file would be named "CS5002_Bitdiddle_A11.pdf". (There should be no brackets around your name).
  - Make sure your name is in the document as well (e.g., written on the top of the first page).