

### **TREES**

Dr. Isaac Griffith

**IDAHO STATE UNIVERSITY** 

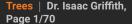
## Outline



#### The lecture is structured as follows:

- Tree Basics
- Applications of Trees
  - Prefix Trees and Huffman Coding
  - Decision Trees
  - BSTs
  - Game Trees
- Tree Traversals
  - Pre-, In-, Post-order
- Tree Induction
- Spanning Trees
  - BFS and DFS
- Minimum Spanning Trees









**CS 1187** 

### What's Trees?



• A tree is a connected undirected graph with no simple circuits







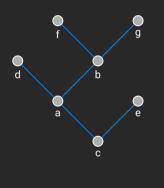


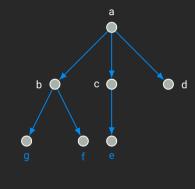
• Theorem: An undirected graph is a tree iff there is a unique simple path between any two of its vertices

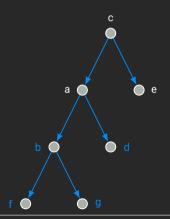
### **Rooted Trees**



• A **rooted tree** is a tree in which one vertex has been designated as the **root** and every edge is directed away from the root.

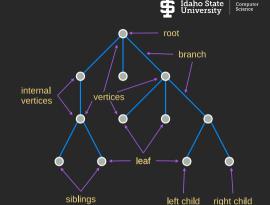






### Terminologies of Rooted Trees

- If v is a vertex in T other than the root, the parent of v is the unique vertex u such that there is a directed edge from u to v
- If u is the parent of v, v is called a child of u
- Vertices with the same parent are called siblings





# Terminologies of Rooted Trees



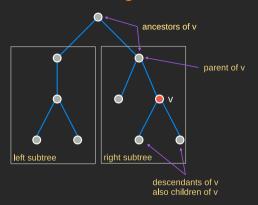
- The ancestors of a vertex other than the root are the vertices in the path from the root to this vertex, excluding the vertex itself and including the root
- The descendants of a vertex v are those vertices that have v as an ancestor
- A vertex of a tree is called a leaf if it has no children
- Vertices that have children are called internal vertices
- If a is a vertex is a tree, the subtree with a as its root is the subgraph of the tree consisting of a
  and its descendants and all edges incident to these descendants.

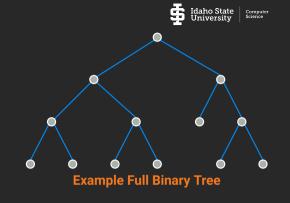
### m-Ary Trees



- A rooted tree is called an m-ary tree if every internal vertex has no more than m children. The
  tree is called a full m-ary tree if every internal vertex has exactly m children. An m-ary tree with
  m = 2 is called a binary tree
- An ordered rooted tree is a rooted tree where the children of each internal vertex are ordered.
   Ordered rooted trees are drawn so that the children of each internal vertex are shown in order from left to right.
- In an ordered binary tree (usually just called a binary tree), if an internal vertex has two children,
  the first child is called the left child and the second child is called the right child. The tree rooted
  at the left child (or right child, respectively) of a vertex is called the left subtree (or right subtree,
  respectively) of this vertex.

# Tree Terminologies





## **Properties of Trees**

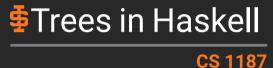


- Theorem: A tree with n vertices has n-1 edges
- Theorem: A full m-ary tree with *i* internal vertices contains n = mi + 1 vertices

## **Properties of Trees**



- Theorem: A full m-ary tree with
  - 1. n vertices has i = (n-1)/m internal vertices and I = [(m-1)n + 1]/m leaves
  - 2. *i* internal vertices has n = mi + 1 vertices and l = (m 1)i + 1 leaves
  - 3. I leaves has n = (ml 1)/(m 1) vertices and \$i = (I 1) / (m 1) internal vertices
- Theorem: There are at most  $m^h$  leaves in any m-ary tree of height h



# Representing Trees in Haskell



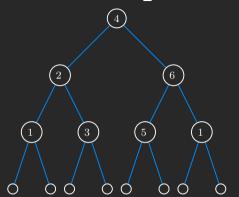
• The representation of a *binary tree* in Haskell is as follows (limited to integer data)

- · That is a tree is either
  - A leaf without a value, or
  - A node with a value and a right and a left subtree

# Haskell Example

#### Idaho State | comp University | Scient

#### **Example tree**





**CS 1187** 

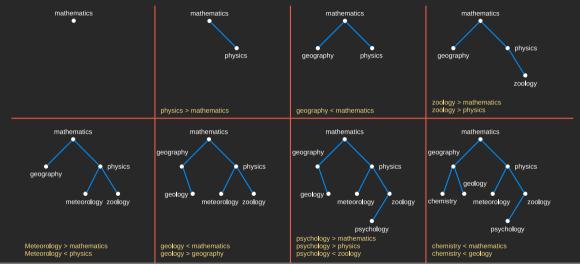
## **Binary Search Trees**



- Binary Search Tree: a binary tree in which each child of a vertex is designated as a right or left child, no vertex has more than one right or left child, and each vertex is labeled with a key, which is one of the items.
- Vertices are assigned keys so that the key of a vertex is both larger than keys of all vertices in its left subtree and smaller than the keys of all vertices in its right subtree.

## **Binary Search Trees**





Trees | Dr. Isaac Griffith, Page 16/70



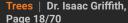
# Searching a BST



• We can implement a BST in Haskell as follows:

```
bstSearch :: Ord a => a -> BinTree (a, b) -> Maybe b
bstSearch key BinLeaf = Nothing
bstSearch key (BinNode (x, y) t1 t2) =
  if key == x
   then Just y
  else if key < x
    then bstSearch key t1
   else bstSearch key t2</pre>
```

```
procedure INSERTION(T: binary search tree, x item)
   v := \text{root of } T
   {a vertex is not present in T has the value null}
   while v \neq null and label(v) \neq x do
       if x < label(v) then
           if left child of v \neq null then
               v := left child of v
               add new vertex as a left child of v and set v := null
       else
           if right child of v \neq null then
               v := \text{right child of } v
           else
               add new vertex as a right child of v and set v := null
   if root of T = null then
       add a vertex v to the tree and label it with x
   else if v is null or label(v) \neq x then
       label new vertex with x and let v be this new vertex
   return v \{ v = \text{location of } x \}
```





### **BST Insertion**



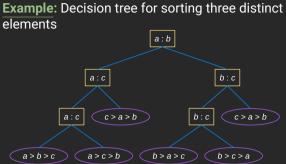
• The prior algorithm for insertion can be implemented as follows in Haskell:

```
insert :: Ord a => (a, b) -> BinTree (a, b) -> BinTree (a, b)
insert (key, d) BinLeaf = BinNode (key, d) BinLeaf BinLeaf
insert (key, d) (BinNode (x, y) t1 t2) =
  if key == x
    then BinNode (key, d) t1 t2
  else if key < x
    then BinNode (x, y) (insert (key, d) t1) t2
    else BinNode (x, y) t1 (insert (key, d) t2)</pre>
```

#### **Decision Trees**

Idaho State Computer Science

- Decision tree: a rooted tree in which each internal vertex corresponds to a decision with a subtree at these vertices for each possible outcome fo the decision.
- Such trees can be used to model problems in which a series of decisions leads to a solution.





#### **Prefix Codes**



- Prefix Codes: codes wherein a bit string represents a letter, and no bit strings corresponds to more than one sequence of letters.
- Huffman coding: a special case of prefix codes
  - Uses frequencies of symbols in a string and produces as output a prefix code that encodes the string using the fewest possible bits, among all possible binary prefix codes for these symbols.

# **Huffman Coding**



**procedure** HUFFMAN(C: symbols  $a_i$  with frequencies  $w_i$ , i = 1, ..., n)

F := forest of n rooted trees, each consisting of a single vertex  $a_i$  and assigned weight  $w_i$  while F is not a tree **do** 

Replace the rooted trees T and T' of least weights from F with  $w(T) \ge w(T')$  with a tree having a new root that has T as it left subtree and T' as its right subtree. Label the new edge to T with 0 and the new edge to T' with 1.

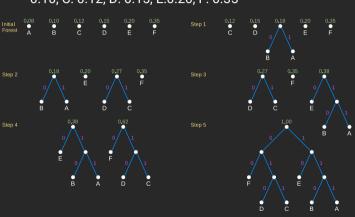
Assign w(T) + w(T') as the weight of the new tree

{the Huffman coding for the symbol  $a_i$  is the concatenation of the labels of the edges in the unique path from the root to the vertex  $a_i$ }

# **Huffman Coding Example**



• Example: Use Huffman coding to encode the following symbols with the frequencies: A: 0.08, B: 0.10, C: 0.12, D: 0.15, E:0.20, F: 0.35



 Solution: This encodes A by 111, B by 110, C by 011, D by 010, E by 10, and F by 00.

The average number of bits used to encode a symbol using this encoding is:

$$3 \cdot 0.08 + 3 \cdot 0.10 + 3 \cdot 0.12 +  $3 \cdot 0.15 + 2 \cdot 0.020 + 2 \cdot 0.35 = 2.45$$$



### **Game Trees**



- Trees can be used to analyze certain types of games
  - Tic-tac-toe
  - Checkers
  - Chess
  - Nim
- These types of games are called perfect-information games
  - both players know the moves of the other players
  - the state of the game is known by both players
  - No chance involved

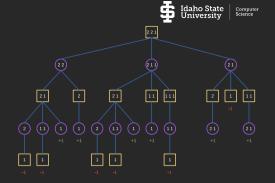
#### Game Trees



- Game Trees: where the vertices represent the positions that a game can be in as it progresses, and the edges represent legal moves.
  - Tend to be very large but can be simplified by combining all symmetric positions into the same node
  - Root node represents the starting position.
  - Even levels are represented by boxes and are the first player's move
  - · Odd levels are represented by circles and are the second player's moves
  - We assign values to leaves to represent the payoff to the first player if the game terminates (terminal vertices are represented by boxes)

### The Game of Nim

- At the start of the game there are a number of piles of stones.
- Players take turns making moves
  - A legal move consists of removing one or more stones from one of the piles, without removing all stones left.
  - A player without a legal move loses
- A game tree for this game is shown

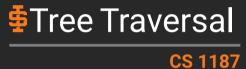




### **Minmax**



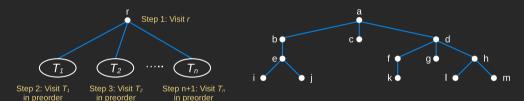
- Definition: The value of a vertex in a game tree is defined recursively as:
  - i. the value of a leaf is the payoff to the first player when the game terminates in the position represented by this leaf.
    - ii. the value of an internal vertex at an even level is the maximum of the values of its children, and the value of an internal vertex at an odd level is the minimum of the values of its children.
- Theorem: The value of a vertex of a game tree tells us the payoff to the first player if both
  players follow the minmax strategy and play starts from the position represented by this vertex.
- Minmax Strategy: where the first player moves to a position represented by a child with maximum value and the send player moves to a position of a child with minimum value.
- Various approaches can be used to devise good strategies
  - alpha-beta pruning eliminates computation by pruning portions of game trees based on ancestor vertex values
  - evaluation functions to estimate the value of internal vertices
- More information can be learned by studying combinatorial game theory



#### **Preorder Traversal**

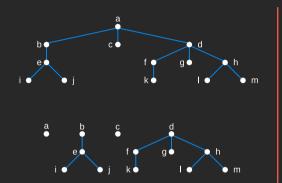


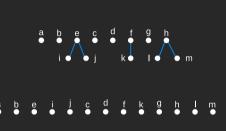
Definition: Let T be an ordered rooted tree with root r. If T consists only of r, then r is the
preorder traversal of T. Otherwise, suppose that T<sub>1</sub>, T<sub>2</sub>,..., T<sub>n</sub> are the subtrees at r from left to
right in T. The preorder traversal begins by visiting r. It continues by traversing T<sub>1</sub> in preorder,
then T<sub>2</sub> in preorder, and so on, until T<sub>n</sub> is traversed in preorder



# **Examples of Preorder Traversal**







### Pseudocode of Preorder Traversal



#### Algorithm:

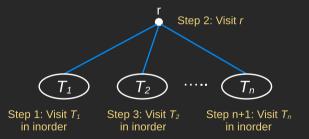
```
procedure PREORDER(T: ordered rooted tree)
r := \text{root of } T
\textit{listr}
\textit{for each child } c \text{ of } r \text{ from left to right } \textit{do}
T(c) := \text{subtree with } c \text{ as its root}
\texttt{PREORDER}(T(c))
```

```
preorder :: BinTree a -> [a]
preorder BinLeaf = []
preorder (BinNode x t1 t2) =
  [x] ++ preorder t1 ++ preorder t2
```

### **Inorder Traversal**

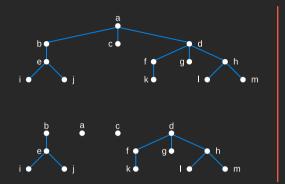


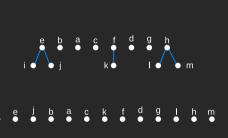
Defintion: Let T be an ordered rooted tree with root r. If T consists only of r, then r is the inorder traversal of T. Otherwise, suppose that T<sub>1</sub>, T<sub>2</sub>,..., T<sub>n</sub> are the subtrees at r from left to right. The inorder traversal begins by traversing T<sub>1</sub> in inorder, then visiting r. It continues by traversing T<sub>2</sub> in inorder, then T<sub>3</sub> in inorder, ..., and finally T<sub>n</sub> in inorder.



# **Examples of Inorder Traversal**







## Pseudocode of Inorder Traversal



#### Algorithm

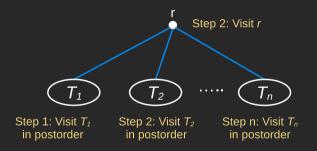
```
procedure INORDER(T: ordered rooted tree)
   r := \text{root of } T
   if r is a leaf then
       listr
   else
       l := first child of r from left to right
        T(I) := subtree with I as its root
       INORDER(T(I))
       list r
       for each child c of r except for I from
left to right do
           T(c) := subtree with c as its root
           INORDER(T(c))
```

#### Haskell Implementation:

```
inorder :: BinTree a -> [a]
inorder BinLeaf = []
inorder (BinNode x t1 t2) =
  inorder t1 ++ [x] ++ inorder t2
```

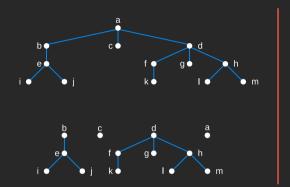
#### Postorder Traversal

• **Definition:** Let T be an ordered rooted tree with root r. If T consists only of r, then r is the **postorder traversal** of T. Otherwise, suppose that  $T_1, T_2, \ldots, T_n$  are the subtrees at r from left to right in T. The **postorder traversal** begins by traversing  $T_1$  in postorder, then  $T_2$  in postorder, ..., then  $T_n$  in postorder, and end by visiting r.



## **Examples of Postorder Traversal**







## Pseudocode of Postorder Traversal



#### Algorithm

```
procedure POSTORDER(T: ordered rooted tree)
r := \text{root of } T
for each child c of r from left to right do
T(c) := \text{subtree with } c as its root
POSTORDER(T(c))
list r
```

#### Haskell Implementation

```
postorder :: BinTree a -> [a]
postorder BinLeaf = []
postorder (BinNode x t1 t2) =
   postorder t1 ++ postorder t2 ++ [x]
```

#### **Procession Tree Structure**



- STDM provides several functions to process a tree, including measuring tree size or the ability to affect its shape
  - reflect takes a binary tree and returns its mirror image

```
reflect :: BinTree a -> BinTree a
reflect BinLeaf = BinLeaf
reflect (BinNode n 1 r) = BinNode n (reflect r) (reflect 1)
```

height - measures the height of a tree between its root and its deepest leaf (an empty tree has height zero)

```
height :: BinTree a -> Integer
height BinLeaf = 0
height (BinNode x t1 t2) = 1 + max (height t1) (height t2)
```

#### **Procession Tree Structure**



- STDM provides several functions to process a tree, including measuring tree size or the ability to affect its shape
  - size calculates the size of a tree, as the number of nodes a tree has

```
size :: BinTree a -> Integer
size BinLeaf = 0
size (Node x t1 t2) = 1 + size t1 + size t2
```

balanced - determines if the binary tree is balanced or not

```
balanced :: BinTree a -> Bool
balanced BinLeaf = True
balanced (BinNode x t1 t2) =
  balanced t1 && balanced t2 && (height t1 == height t2)
```

• Theorem: Let h = height t. If balanced t, then size  $t = 2^h - 1$ 

## **Evaluating Expression Trees**

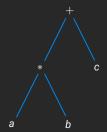


- Often when working with text we can use trees to represent documents in the language.
- Trees represent the structure of the text while omitting unimportant details
- Examples of this includes programs that manipulate language, as well as compilers and interpreters for programming languages.

## Infix, Prefix, and Postfix Notation



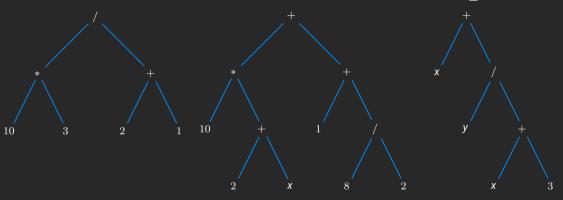
- Examples: infix, prefix, and postfix notations of  $a \times b + c$ 
  - Infix: a \* b + c: uses inorder traversal of an expression tree
  - Prefix: +\*abc (also called Polish notation) uses preorder traversal of an expression tree
  - Postfix: ab \* c+ uses postorder traversal of an expression tree
- Represented by ordered rooted trees



## Examples of Expression Tree Representation



Idaho State University Computer Science



## **Evaluation Expression Trees**



• Let's consider the following simple expression language

```
data Exp
= Const Integer
| Add Exp Exp
```

• We could then implement a simple programming language interpreter as a tree traversal, The interpreter function eval takes an expression tree and returns the value it represents

```
eval :: Exp -> Integer
eval (Const n) = n
eval (Add e1 e2) = eval e1 + eval e2
eval (Mult e1 e2) = eval e1 * eval e2
```

| Mult Exp Exp

# **\$**Tree Induction CS 1187

#### Tree Induction



- Theorem: Principle of induction on binary trees. Let BinTree a be a binary tree type as prior
  defined, and let P(t) be a proposition on trees. Suppose the following two requirements hold:
  - Base Case: P(BinLeaf)
  - Induction Case: For all t1 and t2 of type BinTree a, and all x :: a, suppose that the proposition holds for a tree consisting of a node, the value a, and the subtrees t1 and t2, provided that the proposition holds for t1 and t2.
    - This can be written as:  $P(t_1) \wedge P(t_2) o P(\mathtt{BinNode} \ \mathtt{x} \ \mathtt{t1} \ \mathtt{t2})$
  - Then  $\forall t :: BinTree a. P(t)$ , thus the proposition holds for all trees of finite size

## Tree Induction Example



Example: Let t :: BinTree a be any finite binary tree. Then length (inorder t) = size t

### **Proof:**

```
Base Case:
```

= length []

= 0

Trees | Dr. Isaac Griffith. Page 46/70

length (inorder BinLeaf)

= size BinLeaf

Assume the induction hypothesis:

length (inorder t1) = size t1 length (inorder t2) = size t2

Then

length (inorder (BinNode x t1 t2))

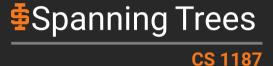
= length (inorder t1 ++ [x] ++ inorder t2

= length (inorder t1) + length [x] + length

= size t1 + 1 + size t2

= size (BinNode x t1 t2)

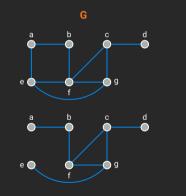
Therefore the theorem holds by tree induction

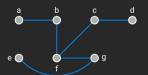


## What is a Spanning Tree



• **Definition:** Let *G* be a simple graph. A **spanning tree** of *G* is a subgraph of *G* that is a tree containing every vertex of *G*.





• Theorem: A simple graph is connected iff it has a spanning tree

## How to Construct Spanning Trees?



Computer Science

Two approaches

- Depth-first search (DFS)
- Breadth-first search (BFS)

## Algorithm: Depth-First Search



- Idea form a rooted tree by arbitrarily selecting a root
  - Form a path by successively adding vertices and edges incident with the last vertex selected
    - Only select those edges/vertices not already in the graph
  - If we can no longer go forward, we backtrack and try again.
  - This process is repeated until all nodes are visited.
- Also called **backtracking** as the algorithm returns to vertices previously visited to add new paths. **procedure** DFS(G: connected graph with vertices  $v_1, v_2, \ldots, v_n$ )

```
procedure VISIT(v: vertex of G)

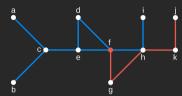
for each vertex w adjacent to v and not yet in T do

add vertex w and edge \{v, w\} to T

VISIT(w)
```

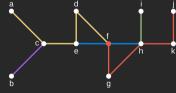
## Example of DFS













## **Backtracking Applications**



- Backtracking as used in DFS can also be used as a technique for the exhaustive search of all
  possible solutions, application of this technique include
  - Graph colorings
  - The *n*-Queens problem
  - · Sums of Subsets
- Additionally, we can apply DFS (and of course BFS) in graphs and digraphs. When applied to digraphs some useful applications occur
  - Webcrawlers and Internet Search Engines
- It provides the ability to backtrack and try another path when it is known that the current path will provide no viable solutions.
  - Additionally, another *algorithmic strategy* can be applied called **branch-and-bound**
  - In branch-and-bound when we see a similar subtree which has already been shown to been shown that it will not render the optimal solution, rather than continuing down that path, we simply bound over it to the next subtree.



#### **Breadth-First Search**



- Rather than processing all nodes and backtracking when we reach an end, here we instead process each all successive vertices of vertex.
- The idea is as follows: select an arbitrary root
  - Add all edges incident to the selected vertex
  - We then visit each of these nodes in the same way before processing their successive nodes.

## Algorithm: Breadth-First Search

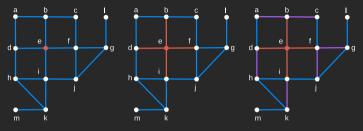


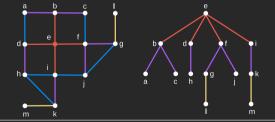
```
procedure BFS(G: connected graph with vertices v_1, v_2, \ldots, v_n)
T := tree consisting only of the vertex v_1
L := empty list
put v_1 in the list L of unprocessed vertices
while L is not empty do
remove the first vertex, v, from L
for each neighbor w of v and not yet in T do
if w is not in L and not in T then
add w to the end of the list L
add w and edge \{v, w\} to T
```



## **Example of Breadth-First Search**







#### DFS vs. BFS



 Although we can specify the DFS and BFS algorithms as already stated, we can create a general iterative algorithm for both:

```
procedure GENERALSEARCH(G: connected graph with vertices v_1, v_2, \ldots, v_n)
T := tree consisting only of the vertex v_1
L := empty list
ENQUE(L, v_1)
while L is not empty do
V := DEQUE(L)
for each neighbor W of V and not yet in V do
    if W is not in U and not in U then
ENQUE(L, W)
add U and edge U and U to U
```

#### DFS vs. BFS



- For DFS, L is a Stack (LIFO)
  - enque is the stack push method (adding an item to the front of the list)
  - deque is the stack pop method (removing the first item from the list)
- For BFS, L is a Queue (FIFO)
  - enque is the queue offer method (adding an item to the end of the list)
  - deque is the queue poll method (removing the first item from the list)



**CS 1187** 

• If T is a spanning tree in a weighted graph G(V, E, w), the weight of T, denoted by w(T), is the sum of weights of edges in T.

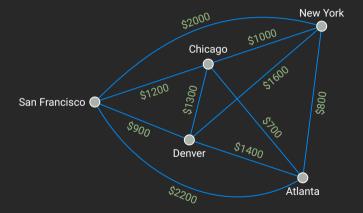
$$w(t) = \sum_{e \in T} w(e)$$

• Given a weighted graph G(V, E, w), the minimum spanning tree problem is to find a spanning tree in G that has the smallest weight

## Cost of a Computer Network



• What is the smallest total cost to maintain a connected network between those five cities?



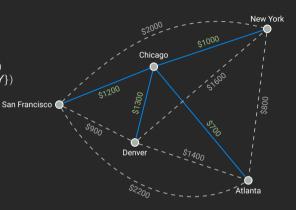


## Some Spanning Trees

```
Idaho State Computer Science
```

```
 \bullet \  \  \, \textit{T}_1 = \left\{ \begin{array}{l} \{\textit{Chicago}, \textit{SF}\}, \, \{\textit{Chicago}, \textit{Denver}\}, \\ \{\textit{Chicago}, \textit{Atlanta}\}, \, \{\textit{Chicago}, \textit{NY}\} \end{array} \right\}
```

```
\begin{array}{lll} \textit{w}(\textit{T}_1) & = & \textit{w}(\{\textit{Chicago}, \textit{SF}\}) + \textit{w}(\{\textit{Chicago}, \textit{Denver}\}) \\ & + \textit{w}(\{\textit{Chicago}, \textit{Atlanta}\}) + \textit{w}(\{\textit{Chicago}, \textit{NY}\}) \\ & = & \$1200 + \$1300 + \$700 + \$1000 = \$4200 \\ \end{array}
```

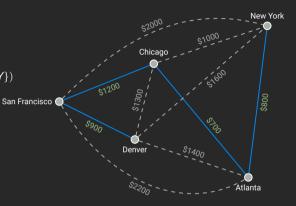


## Some Spanning Trees

```
Idaho State Computer Science
```

```
 \bullet \quad \textit{T}_2 = \left\{ \begin{array}{l} \{\textit{Chicago}, \textit{SF}\}, \ \{\textit{SF}, \textit{Denver}\}, \\ \{\textit{Chicago}, \textit{Atlanta}\}, \ \{\textit{Atlanta}, \textit{NY}\} \end{array} \right\}
```

```
\begin{array}{lcl} \textit{w}(\textit{T}_1) & = & \textit{w}(\{\textit{Chicago}, \textit{SF}\}) + \textit{w}(\{\textit{SF}, \textit{Denver}\}) \\ & + \textit{w}(\{\textit{Chicago}, \textit{Atlanta}\}) + \textit{w}(\{\textit{Chicago}, \textit{NY}\}) \\ & = & \$1200 + \$900 + \$700 + \$800 = \$3600 \\ & & \$\text{Sar} \end{array}
```



## Some Spanning Trees

```
 \bullet \  \  \, \textit{T}_{3} = \left\{ \begin{array}{l} \{\textit{Chicago}, \textit{Denver}\}, \ \{\textit{Denver}, \textit{SF}\}, \\ \{\textit{Denver}, \textit{Atlanta}\}, \ \{\textit{Atlanta}, \textit{NY}\} \end{array} \right\} 
                                                                                                                                                                                                                       New York
        w(T_1)
                         = w(\{Chicago, Denver\}) + w(\{Denver, SF\})
                                  +w(\{Denver, Atlanta\}) + w(\{Atlanta, NY\})
                                  \$1300 + \$900 + \$1400 + \$800 = \$4400
                                                                                                                   San Francisco
                                                                                                                                                                   Denver
```

• Problem: Which one is the one with the smallest weight among all possible spanning trees?

## Prim's Algorithm

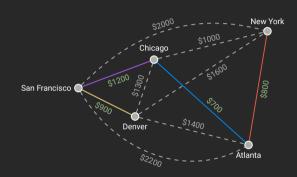


```
procedure PRIM(G: weighted connected undirected graph with n vertices)
T := a minimum-weighted edge
for i := 1 to n-2 do
e := an edge of minimum weight incident to a vertex in T and not forming a simple circuit in T if added to T
T := T with e added
T is a minimum spanning tree of T
```

## An Example of Prim's Algorithm



Choice	Edge	Cost
	{Atlanta, Chicago}	\$700
	{Atlanta, NY}	\$800
	{Chicago, SF}	\$900
4	{Denver, SF}	\$1200
	Total	\$3600





## Kruskal's Algorithm



**procedure** KRUSKAL(G: weighted connected undirected graph with n vertices)

e := an edge in G with smallest weight that does not form a simple circuit when added to T

```
T := empty graph
for i := 1 to n - 1 do
```

T := T with e added

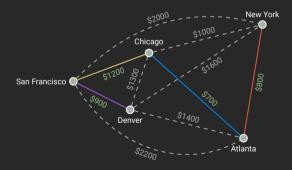
{*T* is a minimum spanning tree of *G*}

## An Example of Kruskal's Algorithm



- · First, sort all edges based on their weight in ascending order.
  - {Atlanta, Chicago}, {Atlanta, NY}, {Denver, SF}, {Chicago, NY}, {Chicago, SF}, {Chicago, Denver}, {Atlanta, Denver}, {Denver, NY}, {NY, SF}, {Atlanta, SF}
- Examine each edge one by one until a spanning tree is constructed

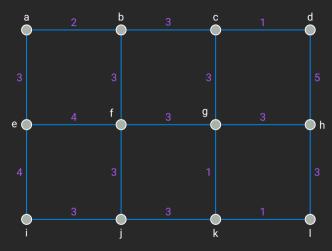
Choice	Edge	Cost
1	{Atlanta, Chicago}	\$700
	{Atlanta, NY}	\$800
	{Denver, SF}	\$900
4	{Chicago, SF}	\$1200
	Total	\$3600





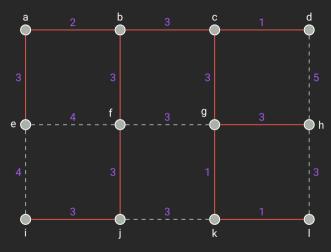
## Finding a Spanning Tree with Minimum Weight





## Finding a Spanning Tree with Minimum Weight









## Are there any questions?