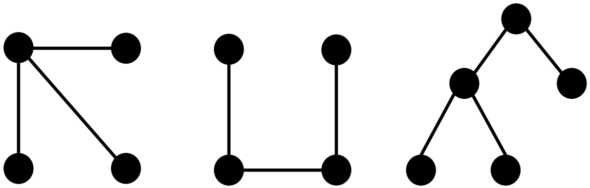
ET 1201
Matematika Diskrit
Prodi Teknik Telekomunikasi, Institut Teknologi Bandung
2024

Outline

- Tree
- Spanning Tree
- Huffman Code

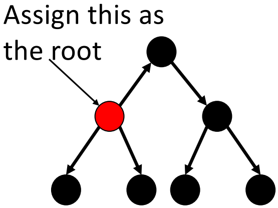
What is a Tree ?

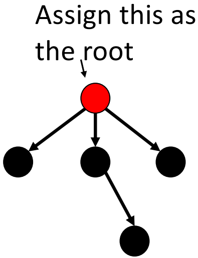
- An un-directed graph is a tree if and only if there is a unique simple path between any two of its vertices.
  - Only a **unique** simple path between two vertices.
  - No loops, no multiple edges.
  - A connected graph with no simple circuits.
- Examples of the tree.



Rooted Trees

- A rooted tree is:
  - One vertex has been designated as the root.
  - Every edge is directed away from the root.
- We usually put the root at the top, and point each edge downwards.

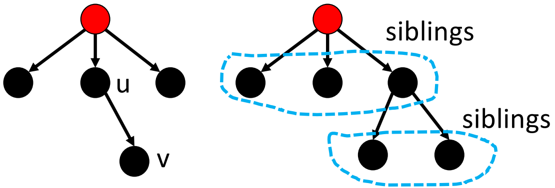




-

Rooted Trees

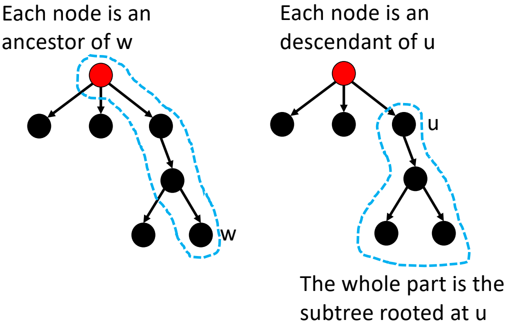
- Each edge is from a parent to a child
  - The parent of a vertex is **unique**
- Vertices with the same parent are siblings



u is the parent of v
v is a child of u

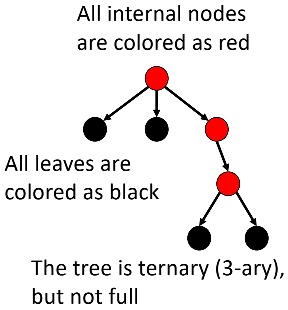
Rooted Trees

- The ancestors of a vertex **w** include all the nodes in the path from the root to **w**
- The descendants of a vertex **u** include all the nodes that have **u** as its ancestor.
- The subtree rootd at **u** includes all the descendants of **u**, and all edges that connect between them.



Rooted Trees

- Vertices with no children are called leaves.
- Otherwise, they are called internal nodes (vertices).
- A m-ary tree is every internal node has no more than m children.
- A full m-ary tree is every internal vertex has exactly m children.
  - A m-ary tree with $\mathrm{m} = 2$ is called **binary tree**.

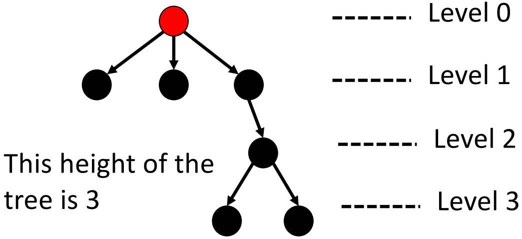


Properties of Trees

- **Theorem:** A tree with **n** nodes has **n - 1** edges.
- **Proof:**
  - Pick a vertex **u** in the tree and make **u** the root.
  - Each edge links a parent and a child.
  - The root has no parent, but every node has exactly one parent.
  - Therefore, the number of edges = **n - 1**.

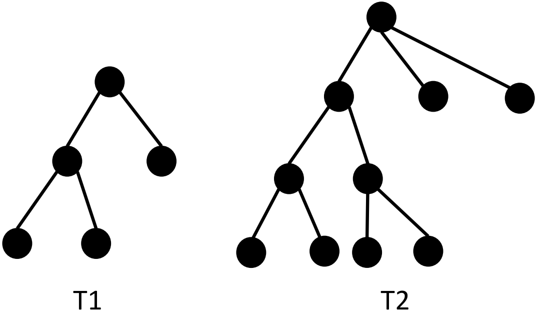
Properties of Trees

- The level of vertex **v** in a rooted tree is the length of the path from root to **v**
- The height of the tree is the maximum level of all the vertices.



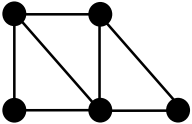
Properties of Trees

- A rooted tree of height **h** is balanced if all the leaves are at level **h** or **h - 1**
- Which of the following trees are balanced ?
  - T1 is balanced, because all its leaves are at level 2.
  - T2 is not balanced, because it has leaves at level 1 and level 3.

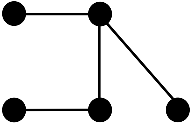


What is a Spanning Tree ?

- A spanning tree of G is a subgraph of G that is a tree containing every vertex of G.



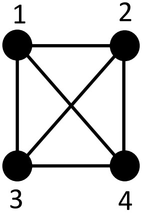
G



A spanning tree of G

Counting Spanning Trees

- Let $G = (V, E)$, and n is the number of vertices and m is the number of edges of G. K_n denotes a complete graph with n vertices.
- How many spanning trees are there in the complete graph K_n ?

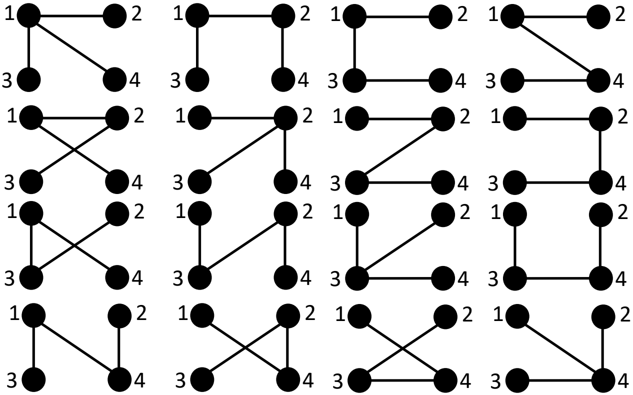


A four-vertex complete graph K_4

-

Counting Spanning Trees

- **16** spanning trees are in K_4.



-

Text Encoding

- Each English character is represented in the same number of bits (8 bits) in ASCII.

  - ASCII uses fixed-length encoding

  - A text contains n characters, which take 8n bits in total to store the text in ASCII.

  - Is it possible to find a coding scheme of these letters such that fewer bits are used ?

Text Encoding

- In real-life English texts, characters do not appear with the same frequency.
- Using bit strings of different lengths to encode letters – variable-length encoding.
  - Frequent characters are encoded in fewer bits.
  - In-frequent characters are encoded in more bits.
- Then, we can reduce the total storage.

Text Encoding

- Supposed a file contains 100K chars composed of A, B, C, D, E letters only.
  - A occurs 45K times, others 11K times each.
- Using fixed-length
  - Each character is encoded in 3 bits, total takes **300Kb**
- Using variable-length:
  - A $->$ 0, B $->$ 100, C $->$ 101, D $->$ 110, E $->$ 111
  - $45K \times 1 + 44K = 177Kb$ (41% savings.)

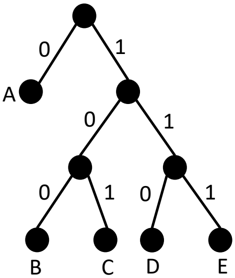
Prefix Code

- If the encoding code book becomes:
  - $A->0, B->1, C->00, D->01, E->010.$
  - Suppose the encoded text is: 0101
  - The original texts can have several possibilities such as
  - ABAB or ABD or DAB or DD or EB
- The problem comes from:
  - One codeword is a prefix of another.

Prefix Code

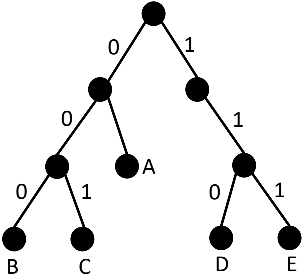
- Prefix code encoding scheme is used to resolve that each codeword is a prefix of another.
- For a text encoded by a prefix code, we can easily decode it in the following way:
  - Scan from left to right to extract the first code
  - Recursively decode the remaining part.

Prefix Code Tree

- A prefix code tree is a rooted tree such that:
  - each edge is labeled by a bit
  - each leaf denoted by a character.
  - The codeword for the character is based on the labels on root-to-leaf path.
  - $A- > 0, B- > 100, C- > 101, D- > 110, E- > 111$
- 

Optimal Prefix Code

- **Problem:** Given the frequencies of each character, design the optimal prefix code whose encoded text requires the least storage.
- **Property 1:** In an optimal prefix code tree, each internal node must have two children.



- This is not an optimal prefix code tree

Optimal Prefix Code

- **Property 2:**

  - The leaves corresponding to the two least frequent characters are siblings.

  - The leaves are farthest from the root.

- **Proof:** Consider an optimal prefix code tree.

  - Let y and z be the least frequent characters.

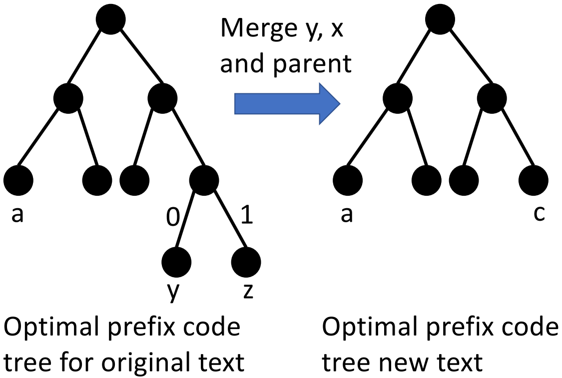
  - Let x be a character whose leaf is the farthest from the root. Its sibling must be a leaf for some character x'.

Optimal Prefix Code

- Let **y** and **z** be the two least frequent characters.
- Let **T** be an optimal tree such that y and z are sibling leaves and fartest from the root.
- When a new text shows: Replace each y and z by a common character c in the original text
- **Property 3:** We get an optimal prefix code tree for the new text if we merge y, z and their parent into a leaf in T, and correspond this leaf to c.

Optimal Prefix Code

- Graphically, the property 3 says:

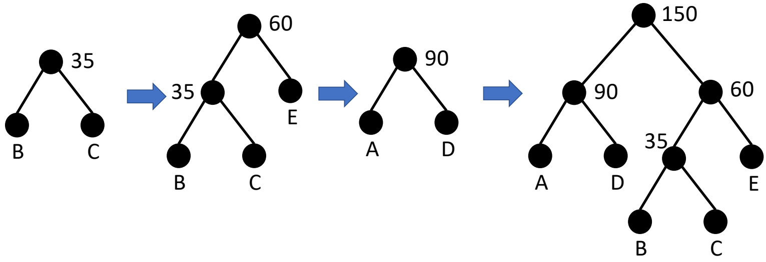


Optimal Prefix Code

- Steps to obtain an optimal prefix code (David Huffman in 1952)
  - Find the least frequent characters x and y.
  - For two leaves for x and y, and join them with a common parent p.
  - Replace x and y by a common character c.
  - Recursively find the optimal prefix code tree for the new text (and replace the leaf for c with p, x, y).

Example

- Suppose the relative frequencies are as follows:
  - A: 40, B: 20, C: 15, D: 50, E:25



Referensi

- Lecture slides on DCP 1244 Discrete Mathematics, Tsung Tai Yeh, 2021

Available:
https://people.cs.nycu.edu.tw/~ttyeh/course/2021_Spring/DCP1244/outline.html