Computer Science & Engineering 423/823
Design and Analysis of Algorithms

Lecture 04 — Elementary Graph Algorithms (Chapter 22)

Stephen Scott
(Adapted from Vinodchandran N. Variyam)

sscott@cse.unl.edu

Introduction

- Graphs are abstract data types that are applicable to numerous problems
  - Can capture *entities*, *relationships* between them, the *degree* of the relationship, etc.
- This chapter covers basics in graph theory, including representation, and algorithms for basic graph-theoretic problems (some content was covered in review lecture)
- We'll build on these later this semester

Breadth-First Search (BFS)

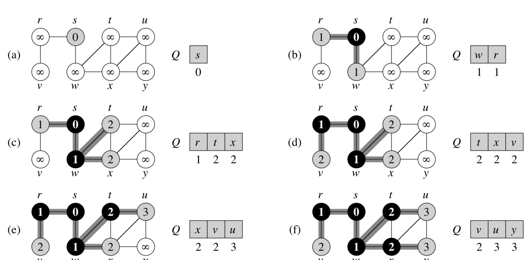
- Given a graph $G = (V, E)$ (directed or undirected) and a *source* node $s \in V$, BFS systematically visits every vertex that is reachable from s
- Uses a queue data structure to search in a breadth-first manner
- Creates a structure called a **BFS tree** such that for each vertex $v \in V$, the distance (number of edges) from s to v in tree is a shortest path in G
- Initialize each node's **color** to WHITE
- As a node is visited, color it to GRAY ($\Rightarrow$ in queue), then BLACK ($\Rightarrow$ finished)

BFS(G, s)

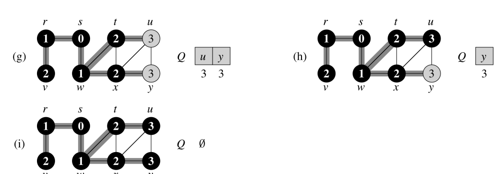
```
for each vertex u ∈ V \ {s} do
    color[u] = WHITE
    d[u] = ∞
    π[u] = NIL
end
color[s] = GRAY
d[s] = 0
π[s] = NIL
Q = ∅
ENQUEUE(Q, s)
while Q ≠ ∅ do
    u = DEQUEUE(Q)
    for each v ∈ Adj[u] do
        if color[v] == WHITE then
            color[v] = GRAY
            d[v] = d[u] + 1
            π[v] = u
            ENQUEUE(Q, v)

    end
    color[u] = BLACK
end
```

BFS Example



BFS Example (2)



BFS Properties

- What is the running time?
  - Hint: How many times will a node be enqueued?
- After the end of the algorithm, $d[v]$ = shortest distance from s to v
  - $\Rightarrow$ Solves unweighted shortest paths
  - Can print the path from s to v by recursively following $\pi[v]$, $\pi[\pi[v]]$, etc.
- If $d[v] == \infty$, then v not reachable from s
  - $\Rightarrow$ Solves reachability

Depth-First Search (DFS)

- Another graph traversal algorithm
- Unlike BFS, this one follows a path as deep as possible before backtracking
- Where BFS is "queue-like," DFS is "stack-like"
- Tracks both "discovery time" and "finishing time" of each node, which will come in handy later

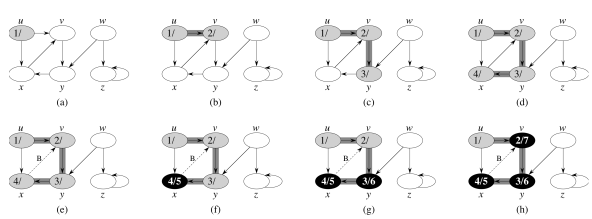
DFS(G)

for *each vertex* $u \in V$ **do**
    $color[u] = \textsc{white}$
    $\pi[u] = \textsc{nil}$
end
$time = 0$
for *each vertex* $u \in V$ **do**
    if $color[u] == \textsc{white}$ **then**
        $\textsc{DFS-Visit}(u)$
end

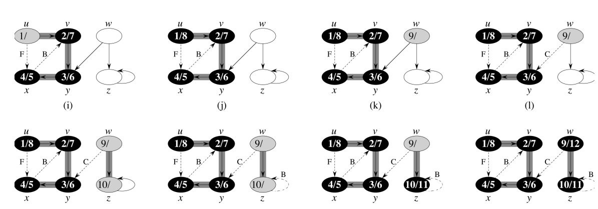
DFS-Visit(u)

$color[u] = \textsc{gray}$
$time = time + 1$
$d[u] = time$
for *each* $v \in Adj[u]$ **do**
    if $color[v] == \textsc{white}$ **then**
        $\pi[v] = u$
        $\textsc{DFS-Visit}(v)$
end
$color[u] = \textsc{black}$
$f[u] = time = time + 1$

DFS Example



DFS Example (2)



DFS Properties

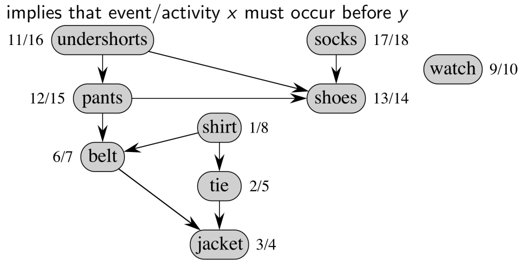
- Time complexity same as BFS: $\Theta(|V| + |E|)$
- Vertex u is a proper descendant of vertex v in the DF tree iff $d[v] < d[u] < f[u] < f[v]$
  - $\Rightarrow$ **Parenthesis structure:** If one prints "(u" when discovering u and "u)" when finishing u, then printed text will be a well-formed parenthesized sentence

DFS Properties (2)

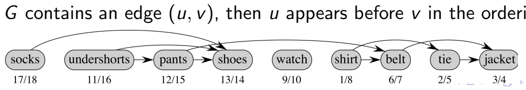
- Classification of edges into groups
  - A **tree edge** is one in the depth-first forest
  - A **back edge** (u, v) connects a vertex u to its ancestor v in the DF tree (includes self-loops)
  - A **forward edge** is a nontree edge connecting a node to one of its DF tree descendants
  - A **cross edge** goes between non-ancestral edges within a DF tree or between DF trees
  - See labels in DFS example
- Example use of this property: A graph has a cycle iff DFS discovers a back edge (application: deadlock detection)
- When DFS first explores an edge (u, v), look at v's color:
  - $color[v] == \text{WHITE}$ implies tree edge
  - $color[v] == \text{GRAY}$ implies back edge
  - $color[v] == \text{BLACK}$ implies forward or cross edge

Application: Topological Sort

A directed acyclic graph (dag) can represent precedences: an edge (x, y) implies that event/activity x must occur before y



A **topological sort** of a dag G is an linear ordering of its vertices such that if G contains an edge (u, v), then u appears before v in the ordering



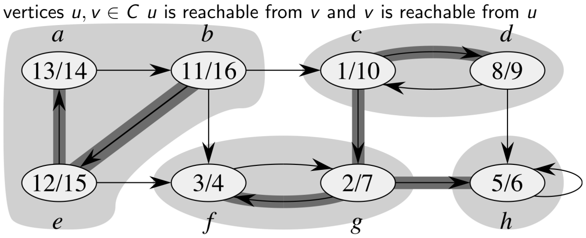
Topological Sort Algorithm

1. Call DFS algorithm on dag G
2. As each vertex is finished, insert it to the front of a linked list
3. Return the linked list of vertices

- Thus topological sort is a descending sort of vertices based on DFS finishing times
- What is the time complexity?
- Why does it work?
  - When a node is finished, it has no unexplored outgoing edges; i.e., all its descendant nodes are already finished and inserted at later spot in final sort

Application: Strongly Connected Components

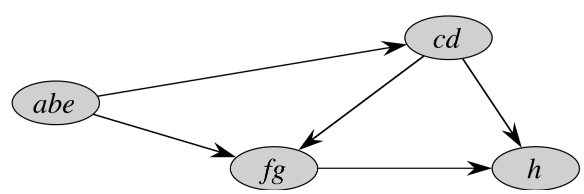
Given a directed graph $G = (V, E)$, a **strongly connected component** (SCC) of G is a maximal set of vertices $C \subseteq V$ such that for every pair of vertices $u, v \in C$ u is reachable from v and v is reachable from u



What are the SCCs of the above graph?

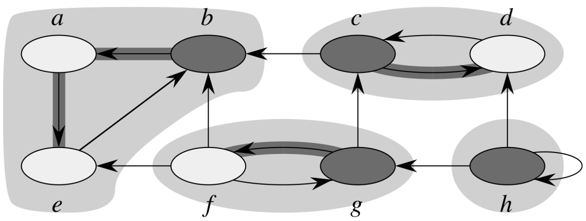
Component Graph

Collapsing edges within each component yields acyclic **component graph**



Transpose Graph

- Algorithm for finding SCCs of G depends on the **transpose** of G, denoted $G^{\top}$
- $G^{\top}$ is simply G with edges reversed
- Fact: $G^{\top}$ and G have same SCCs. Why?

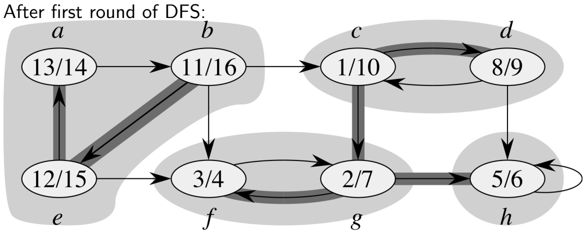


SCC Algorithm

1. Call DFS algorithm on G
2. Compute $G^{\top}$
3. Call DFS algorithm on $G^{\top}$, looping through vertices in order of decreasing finishing times from first DFS call
4. Each DFS tree in second DFS run is an SCC in G

SCC Algorithm Example

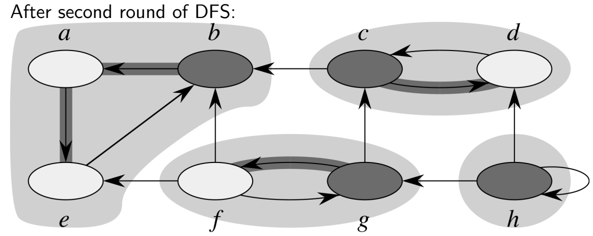
After first round of DFS:



Which node is first one to be visited in second DFS?

SCC Algorithm Example (2)

After second round of DFS:



SCC Algorithm Analysis

- What is its time complexity?
- How does it work?
  1. Let x be node with highest finishing time in first DFS
  2. In $G^{\top}$, x's component C has no edges to any other component (Lemma 22.14), so the second DFS's tree edges define exactly x's component
  3. Now let x' be the next node explored in a new component C'
  4. The only edges from C' to another component are to nodes in C, so the DFS tree edges define exactly the component for x'
  5. And so on...
- In other words, DFS on $G^{\top}$ visits components in order of a topological sort of G's component graph
  - $\Rightarrow$ First component node of $G^{\top}$ visited has no outgoing edges (since in G it has only incoming edges), second only has edges into the first, etc.

Intuition

- For algorithm to work, need to start second DFS in component *abe*
- How do we know that some node in *abe* will have largest finish time?
  - If first DFS in G starts in *abe*, then it visits all other reachable components and finishes in *abe* $\Rightarrow$ one of $\{a, b, e\}$ will have largest finish time
  - If first DFS in G starts in component "downstream" of *abe*, then that DFS round will not reach *abe* $\Rightarrow$ to finish in *abe*, you have to start there at some point $\Rightarrow$ you will finish there last (see above)