CS 374: Algorithms & Models of Computation

Chandra Chekuri   Manoj Prabhakaran

University of Illinois, Urbana-Champaign

Fall 2015

Today

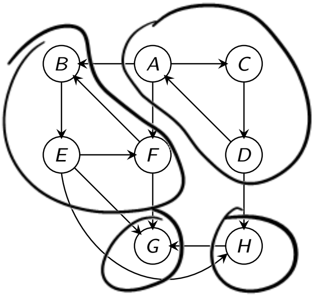
Two topics:

- Structure of directed graphs
- **DFS** and its properties
- One application of **DFS** to obtain fast algorithms

Strong Connected Components (SCCs)

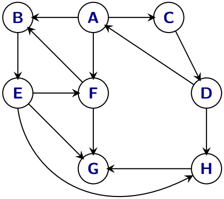
Algorithmic Problem

Find all SCCs of a given directed graph.

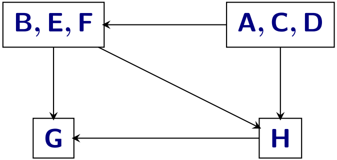


Previous lecture:
Saw an $\mathbf{O(n \cdot (n + m))}$ time algorithm.
This lecture: sketch of a $\mathbf{O(n + m)}$ time algorithm.

Graph of SCCs



Graph G



Graph of SCCs G^{SCC}

Meta-graph of SCCs

Let $S_1, S_2, \ldots S_k$ be the strong connected components (i.e., SCCs) of G. The graph of SCCs is G^{SCC}

1. Vertices are $S_1, S_2, \ldots S_k$
2. There is an edge (S_i, S_j) if there is some $u \in S_i$ and $v \in S_j$ such that (u, v) is an edge in G.

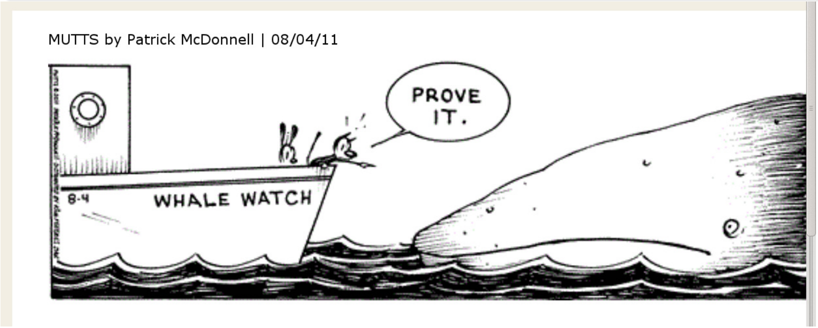
Reversal and SCCs

Proposition

For any graph G, the graph of SCCs of $\mathbf{G}^{\mathbf{rev}}$ is the same as the reversal of G^{SCC}.

Proof.

Exercise. □



SCCs and DAGs

Proposition

For any graph G, the graph G^{SCC} has no directed cycle.

Proof.

If G^{SCC} has a cycle $\mathbf{S_1}, \mathbf{S_2}, \ldots, \mathbf{S_k}$ then $\mathbf{S_1} \cup \mathbf{S_2} \cup \cdots \cup \mathbf{S_k}$ should be in the same SCC in G. Formal details: exercise. □

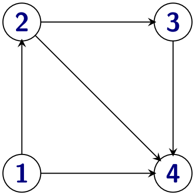
Part I

Directed Acyclic Graphs

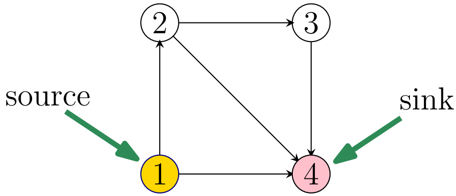
Directed Acyclic Graphs

Definition

A directed graph G is a **directed acyclic graph** (DAG) if there is no directed cycle in G.



Sources and Sinks



Definition

1. A vertex **u** is a **source** if it has no in-coming edges.
2. A vertex **u** is a **sink** if it has no out-going edges.

Simple DAG Properties

Proposition

Every DAG G *has at least one source and at least one sink.*

Simple DAG Properties

Proposition

Every DAG G has at least one source and at least one sink.

Proof.

Let $\mathbf{P} = \mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_k}$ be a longest path in $\mathbf{G}$. Claim that $\mathbf{v_1}$ is a source and $\mathbf{v_k}$ is a sink. Suppose not. Then $\mathbf{v_1}$ has an incoming edge which either creates a cycle or a longer path both of which are contradictions. Similarly if $\mathbf{v_k}$ has an outgoing edge. □

Simple DAG Properties

Proposition

Every DAG G has at least one source and at least one sink.

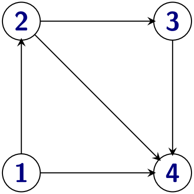
Proof.

Let $\mathbf{P} = \mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_k}$ be a longest path in $\mathbf{G}$. Claim that $\mathbf{v_1}$ is a source and $\mathbf{v_k}$ is a sink. Suppose not. Then $\mathbf{v_1}$ has an incoming edge which either creates a cycle or a longer path both of which are contradictions. Similarly if $\mathbf{v_k}$ has an outgoing edge. □

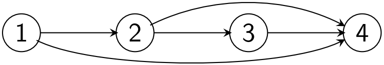
1. G is a DAG if and only if G^{rev} is a DAG.
2. G is a DAG if and only each node is in its own strong connected component.

Formal proofs: exercise.

Topological Ordering/Sorting



Graph G



Topological Ordering of G

Definition

A **topological ordering**/**topological sorting** of $G = (V, E)$ is an ordering $\prec$ on V such that if $(u, v) \in E$ then $u \prec v$.

Informal equivalent definition:

One can order the vertices of the graph along a line (say the x-axis) such that all edges are from left to right.

DAGs and Topological Sort

Lemma

A directed graph G can be topologically ordered iff it is a DAG.

Need to show both directions.

DAGs and Topological Sort

Lemma

A directed graph G can be topologically ordered if it is a DAG.

Proof.

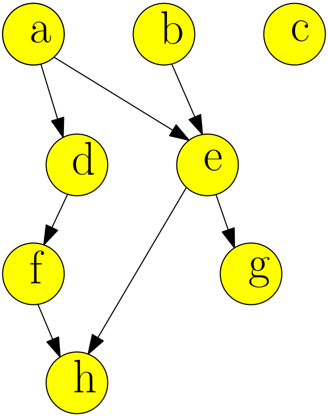
Consider the following algorithm:

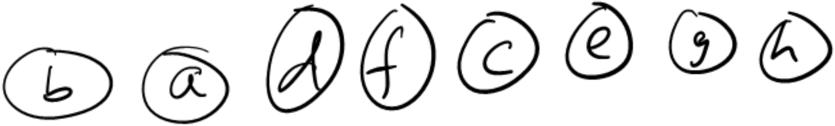
1. Pick a source $\mathbf{u}$, output it.
2. Remove $\mathbf{u}$ and all edges out of $\mathbf{u}$.
3. Repeat until graph is empty.

Exercise: prove this gives toplogical sort. □

Exercise: show algorithm can be implemented in $\mathbf{O(m+n)}$ time.

Topological Sort: Example





DAGs and Topological Sort

Lemma

A directed graph G can be topologically ordered only if it is a DAG.

Proof.

Suppose G is not a DAG and has a topological ordering $\prec$. G has a cycle $\mathbf{C} = \mathbf{u_1}, \mathbf{u_2}, \ldots, \mathbf{u_k}, \mathbf{u_1}$.

Then $\mathbf{u_1} \prec \mathbf{u_2} \prec \ldots \prec \mathbf{u_k} \prec \mathbf{u_1}$!

That is... $\mathbf{u_1} \prec \mathbf{u_1}$.

A contradiction (to $\prec$ being an order).

Not possible to topologically order the vertices. □

DAGs and Topological Sort

Note: A DAG G may have many different topological sorts.

Question: What is a DAG with the most number of distinct topological sorts for a given number n of vertices?

Question: What is a DAG with the least number of distinct topological sorts for a given number n of vertices?

Cycles in graphs

Question: Given an *undirected* graph how do we check whether it has a cycle and output one if it has one?

Question: Given an *directed* graph how do we check whether it has a cycle and output one if it has one?

To Remember: Structure of Graphs

Undirected graph: connected components of $\mathbf{G = (V, E)}$ partition $\mathbf{V}$ and can be computed in $\mathbf{O(m + n)}$ time.

Directed graph: the meta-graph G^{SCC} of $\mathbf{G}$ can be computed in $\mathbf{O(m + n)}$ time. G^{SCC} gives information on the partition of $\mathbf{V}$ into strong connected components and how they form a DAG structure.

Above structural decomposition will be useful in several algorithms

Part II

Depth First Search (DFS)

Depth First Search

DFS is a special case of Basic Search but is a versatile graph exploration strategy. John Hopcroft and Bob Tarjan (Turing Award winners) demonstrated the power of **DFS** to understand graph structure. **DFS** can be used to obtain linear time ($\mathbf{O(m+n)}$) algorithms for

1. Finding cut-edges and cut-vertices of undirected graphs
2. Finding strong connected components of directed graphs
3. Linear time algorithm for testing whether a graph is planar

Many other applications as well.

DFS in Undirected Graphs

Recursive version. Easier to understand some properties.

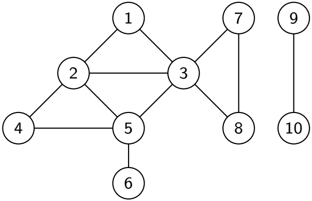
```
DFS(G)
    for all u ∈ V(G) do
        Mark u as unvisited
        Set pred(u) to null
    T is set to ∅
    while ∃ unvisited u do
        DFS(u)
    Output T
```

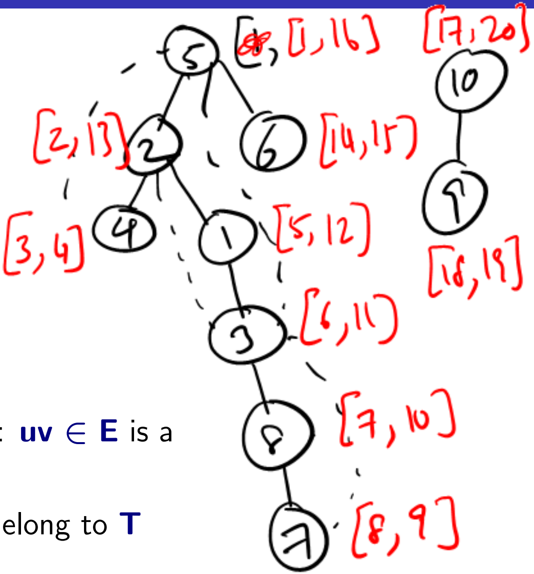
```
DFS(u)
    Mark u as visited
    for each uv in Out(u) do
        if v is not visited then
            add edge uv to T
            set pred(v) to u
            DFS(v)
```

Implemented using a global array **Visited** for all recursive calls.

T is the search tree/forest.

Example





Edges classified into two types: **uv** ∈ **E** is a

1. tree edge: belongs to **T**
2. non-tree edge: does not belong to **T**

Properties of DFS tree

Proposition

1. $\mathbf{T}$ *is a forest*
2. *connected components of* $\mathbf{T}$ *are same as those of* $\mathbf{G}$.
3. *If* $\mathbf{uv} \in \mathbf{E}$ *is a non-tree edge then, in* $\mathbf{T}$, *either:*
    1. $\mathbf{u}$ *is an ancestor of* $\mathbf{v}$, *or*
    2. $\mathbf{v}$ *is an ancestor of* $\mathbf{u}$.

Question: Why are there no *cross-edges*?

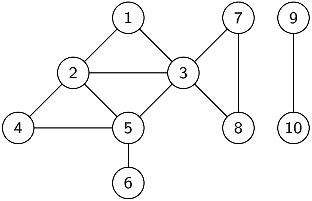
DFS with Visit Times

Keep track of when nodes are visited.

```
DFS(G)
    for all u ∈ V(G) do
        Mark u as unvisited
    T is set to ∅
    time = 0
    while ∃unvisited u do
        DFS(u)
    Output T
```

```
DFS(u)
    Mark u as visited
    pre(u) = ++time
    for each uv in Out(u) do
        if v is not marked then
            add edge uv to T
            DFS(v)
    post(u) = ++time
```

Example



pre and post numbers

Node $\mathbf{u}$ is **active** in time interval $[\mathrm{pre}(\mathbf{u}), \mathrm{post}(\mathbf{u})]$

Proposition

For any two nodes $\mathbf{u}$ *and* $\mathbf{v}$*, the two intervals* $[\mathrm{pre}(\mathbf{u}), \mathrm{post}(\mathbf{u})]$ *and* $[\mathrm{pre}(\mathbf{v}), \mathrm{post}(\mathbf{v})]$ *are disjoint or one is contained in the other.*

pre and post numbers

Node $\mathbf{u}$ is **active** in time interval $[\mathrm{pre}(\mathbf{u}), \mathrm{post}(\mathbf{u})]$

Proposition

For any two nodes $\mathbf{u}$ *and* $\mathbf{v}$*, the two intervals* $[\mathrm{pre}(\mathbf{u}), \mathrm{post}(\mathbf{u})]$ *and* $[\mathrm{pre}(\mathbf{v}), \mathrm{post}(\mathbf{v})]$ *are disjoint or one is contained in the other.*

Proof.

pre and post numbers

Node **u** is **active** in time interval $[\mathrm{pre}(\mathbf{u}), \mathrm{post}(\mathbf{u})]$

Proposition

For any two nodes **u** *and* **v**, *the two intervals* $[\mathrm{pre}(\mathbf{u}), \mathrm{post}(\mathbf{u})]$ *and* $[\mathrm{pre}(\mathbf{v}), \mathrm{post}(\mathbf{v})]$ *are disjoint or one is contained in the other.*

Proof.

- Assume without loss of generality that $\mathrm{pre}(\mathbf{u}) < \mathrm{pre}(\mathbf{v})$. Then **v** visited after **u**.

pre and post numbers

Node **u** is **active** in time interval $[\mathrm{pre}(\mathbf{u}), \mathrm{post}(\mathbf{u})]$

Proposition

For any two nodes **u** *and* **v**, *the two intervals* $[\mathrm{pre}(\mathbf{u}), \mathrm{post}(\mathbf{u})]$ *and* $[\mathrm{pre}(\mathbf{v}), \mathrm{post}(\mathbf{v})]$ *are disjoint or one is contained in the other.*

Proof.

- Assume without loss of generality that $\mathrm{pre}(\mathbf{u}) < \mathrm{pre}(\mathbf{v})$. Then **v** visited after **u**.
- If **DFS**$(\mathbf{v})$ invoked before **DFS**$(\mathbf{u})$ finished, $\mathrm{post}(\mathbf{v}) < \mathrm{post}(\mathbf{u})$.

pre and post numbers

Node **u** is **active** in time interval $[\mathrm{pre}(\mathbf{u}), \mathrm{post}(\mathbf{u})]$

Proposition

For any two nodes **u** *and* **v***, the two intervals* $[\mathrm{pre}(\mathbf{u}), \mathrm{post}(\mathbf{u})]$ *and* $[\mathrm{pre}(\mathbf{v}), \mathrm{post}(\mathbf{v})]$ *are disjoint or one is contained in the other.*

Proof.

- Assume without loss of generality that $\mathrm{pre}(\mathbf{u}) < \mathrm{pre}(\mathbf{v})$. Then **v** visited after **u**.
- If **DFS**$(\mathbf{v})$ invoked before **DFS**$(\mathbf{u})$ finished, $\mathrm{post}(\mathbf{v}) < \mathrm{post}(\mathbf{u})$.
- If **DFS**$(\mathbf{v})$ invoked after **DFS**$(\mathbf{u})$ finished, $\mathrm{pre}(\mathbf{v}) > \mathrm{post}(\mathbf{u})$. □

pre and post numbers

Node $\mathbf{u}$ is **active** in time interval $[\mathrm{pre}(\mathbf{u}), \mathrm{post}(\mathbf{u})]$

Proposition

For any two nodes $\mathbf{u}$ *and* $\mathbf{v}$*, the two intervals* $[\mathrm{pre}(\mathbf{u}), \mathrm{post}(\mathbf{u})]$ *and* $[\mathrm{pre}(\mathbf{v}), \mathrm{post}(\mathbf{v})]$ *are disjoint or one is contained in the other.*

Proof.

- Assume without loss of generality that $\mathrm{pre}(\mathbf{u}) < \mathrm{pre}(\mathbf{v})$. Then $\mathbf{v}$ visited after $\mathbf{u}$.
- If **DFS**$(\mathbf{v})$ invoked before **DFS**$(\mathbf{u})$ finished, $\mathrm{post}(\mathbf{v}) < \mathrm{post}(\mathbf{u})$.
- If **DFS**$(\mathbf{v})$ invoked after **DFS**$(\mathbf{u})$ finished, $\mathrm{pre}(\mathbf{v}) > \mathrm{post}(\mathbf{u})$. □

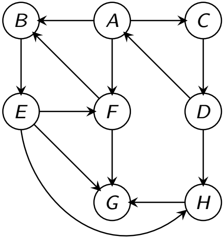
pre and post numbers useful in several applications of **DFS**

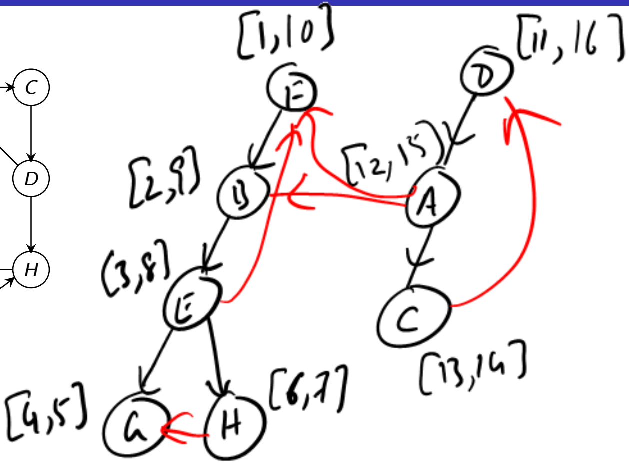
DFS in Directed Graphs

```
DFS(G)
    Mark all nodes u as unvisited
    T is set to ∅
    time = 0
    while there is an unvisited node u do
        DFS(u)
    Output T
```

```
DFS(u)
    Mark u as visited
    pre(u) = ++time
    for each edge (u, v) in Out(u) do
        if v is not visited
            add edge (u, v) to T
            DFS(v)
    post(u) = ++time
```

Example





DFS Properties

Generalizing ideas from undirected graphs:

1. **DFS**(G) takes $O(m + n)$ time.

DFS Properties

Generalizing ideas from undirected graphs:

1. **DFS(G)** takes $\mathbf{O(m+n)}$ time.
2. Edges added form a *branching*: a forest of out-trees. Output of **DFS(G)** depends on the order in which vertices are considered.

DFS Properties

Generalizing ideas from undirected graphs:

1. **DFS(G)** takes $\mathbf{O(m + n)}$ time.
2. Edges added form a *branching*: a forest of out-trees. Output of **DFS(G)** depends on the order in which vertices are considered.
3. If $\mathbf{u}$ is the first vertex considered by **DFS(G)** then **DFS(u)** outputs a directed out-tree $\mathbf{T}$ rooted at $\mathbf{u}$ and a vertex $\mathbf{v}$ is in $\mathbf{T}$ if and only if $\mathbf{v} \in \mathrm{rch}(\mathbf{u})$

DFS Properties

Generalizing ideas from undirected graphs:

1. **DFS(G)** takes $O(m + n)$ time.
2. Edges added form a *branching*: a forest of out-trees. Output of **DFS(G)** depends on the order in which vertices are considered.
3. If u is the first vertex considered by **DFS(G)** then **DFS(u)** outputs a directed out-tree T rooted at u and a vertex v is in T if and only if $v \in \text{rch}(u)$
4. For any two vertices x, y the intervals $[\text{pre}(x), \text{post}(x)]$ and $[\text{pre}(y), \text{post}(y)]$ are either disjoint or one is contained in the other.

DFS Properties

Generalizing ideas from undirected graphs:

1. $\mathbf{DFS(G)}$ takes $\mathbf{O(m+n)}$ time.
2. Edges added form a *branching*: a forest of out-trees. Output of $\mathbf{DFS(G)}$ depends on the order in which vertices are considered.
3. If $\mathbf{u}$ is the first vertex considered by $\mathbf{DFS(G)}$ then $\mathbf{DFS(u)}$ outputs a directed out-tree $\mathbf{T}$ rooted at $\mathbf{u}$ and a vertex $\mathbf{v}$ is in $\mathbf{T}$ if and only if $\mathbf{v} \in \mathrm{rch}(\mathbf{u})$
4. For any two vertices $\mathbf{x}, \mathbf{y}$ the intervals $[\mathrm{pre}(\mathbf{x}), \mathrm{post}(\mathbf{x})]$ and $[\mathrm{pre}(\mathbf{y}), \mathrm{post}(\mathbf{y})]$ are either disjoint or one is contained in the other.

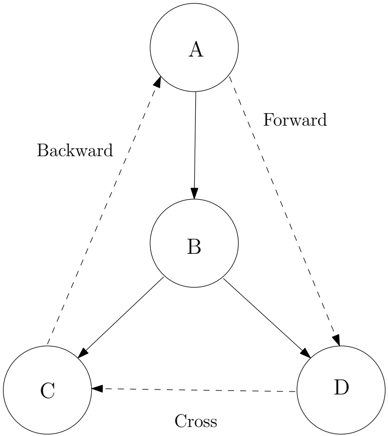
Note: Not obvious whether $\mathbf{DFS(G)}$ is useful in dir graphs but it is.

DFS Tree

Edges of $\mathbf{G}$ can be classified with respect to the **DFS** tree $\mathbf{T}$ as:

1. **Tree edges** that belong to $\mathbf{T}$
2. A **forward edge** is a non-tree edges $(\mathbf{x}, \mathbf{y})$ such that $\mathrm{pre}(\mathbf{x}) < \mathrm{pre}(\mathbf{y}) < \mathrm{post}(\mathbf{y}) < \mathrm{post}(\mathbf{x})$.
3. A **backward edge** is a non-tree edge $(\mathbf{y}, \mathbf{x})$ such that $\mathrm{pre}(\mathbf{x}) < \mathrm{pre}(\mathbf{y}) < \mathrm{post}(\mathbf{y}) < \mathrm{post}(\mathbf{x})$.
4. A **cross edge** is a non-tree edges $(\mathbf{x}, \mathbf{y})$ such that the intervals $[\mathrm{pre}(\mathbf{x}), \mathrm{post}(\mathbf{x})]$ and $[\mathrm{pre}(\mathbf{y}), \mathrm{post}(\mathbf{y})]$ are disjoint.

Types of Edges



Cycles in graphs

Question: Given an *undirected* graph how do we check whether it has a cycle and output one if it has one?

Question: Given an *directed* graph how do we check whether it has a cycle and output one if it has one?

Using DFS...

... to check for Acylicity and compute Topological Ordering

Question

Given G, is it a DAG? If it is, generate a topological sort. Else output a cycle $\mathbf{C}$.

Using DFS...

... to check for Acylicity and compute Topological Ordering

Question

Given G, is it a DAG? If it is, generate a topological sort. Else output a cycle $\mathbf{C}$.

DFS based algorithm:

1. Compute $\mathbf{DFS(G)}$
2. If there is a back edge $\mathbf{e = (v, u)}$ then G is not a DAG. Output cyclce $\mathbf{C}$ formed by path from $\mathbf{u}$ to $\mathbf{v}$ in $\mathbf{T}$ plus edge $\mathbf{(v, u)}$.
3. Otherwise output nodes in decreasing post-visit order. Note: no need to sort, $\mathbf{DFS(G)}$ can output nodes in this order.

Algorithm runs in $\mathbf{O(n + m)}$ time.

Using DFS...

... to check for Acylicity and compute Topological Ordering

Question

Given G, is it a DAG? If it is, generate a topological sort. Else output a cycle $\mathbf{C}$.

DFS based algorithm:

1. Compute $\mathbf{DFS(G)}$
2. If there is a back edge $\mathbf{e = (v, u)}$ then G is not a DAG. Output cyclce $\mathbf{C}$ formed by path from $\mathbf{u}$ to $\mathbf{v}$ in $\mathbf{T}$ plus edge $\mathbf{(v, u)}$.
3. Otherwise output nodes in decreasing post-visit order. Note: no need to sort, $\mathbf{DFS(G)}$ can output nodes in this order.

Algorithm runs in $\mathbf{O(n + m)}$ time.
Correctness is not so obvious. See next two propositions.

Back edge and Cycles

Proposition

G has a cycle iff there is a back-edge in **DFS(G)**.

Proof.

If: (u, v) is a back edge implies there is a cycle C consisting of the path from v to u in **DFS** search tree and the edge (u, v).

Only if: Suppose there is a cycle $C = v_1 \to v_2 \to \ldots \to v_k \to v_1$.
Let v_i be first node in C visited in **DFS**.
All other nodes in C are descendants of v_i since they are reachable from v_i.
Therefore, (v_{i-1}, v_i) (or (v_k, v_1) if $i = 1$) is a back edge. □

Proof

Proposition

If G is a DAG and $\mathrm{post}(\mathbf{v}) > \mathrm{post}(\mathbf{u})$, then $(\mathbf{u}, \mathbf{v})$ is not in G.

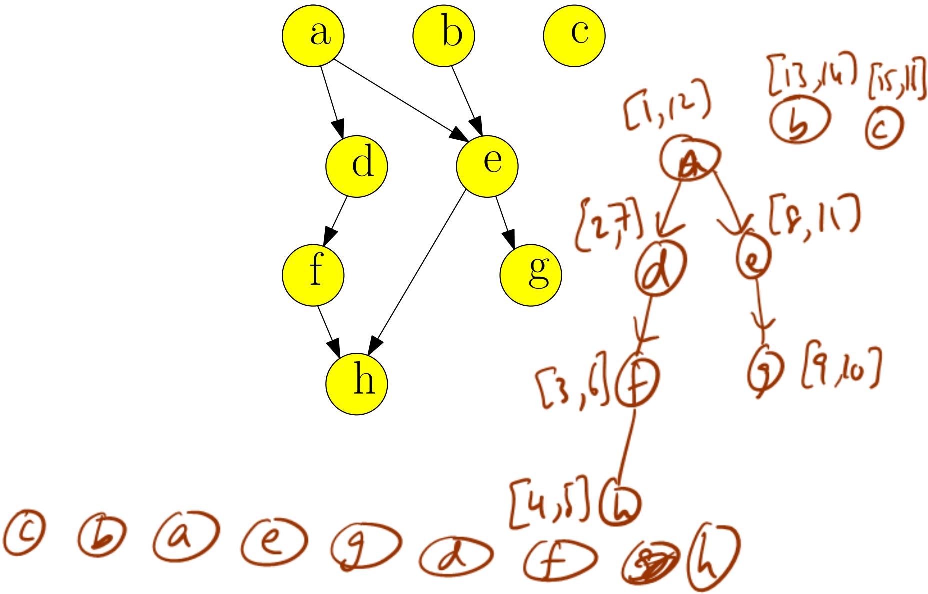
Proof.

Assume $\mathrm{post}(\mathbf{v}) > \mathrm{post}(\mathbf{u})$ *and* $(\mathbf{u}, \mathbf{v})$ is an edge in $\mathbf{G}$. We derive a contradiction. One of two cases holds from DFS property.

- Case 1: $[\mathbf{pre}(\mathbf{u}), \mathbf{post}(\mathbf{u})]$ is contained in $[\mathbf{pre}(\mathbf{v}), \mathbf{post}(\mathbf{v})]$. Implies that $\mathbf{u}$ is explored during **DFS**$(\mathbf{v})$ and hence is a descendent of $\mathbf{v}$. Edge $(\mathbf{u}, \mathbf{v})$ implies a cycle in G but G is assumed to be DAG!
- Case 2: $[\mathbf{pre}(\mathbf{u}), \mathbf{post}(\mathbf{u})]$ is disjoint from $[\mathbf{pre}(\mathbf{v}), \mathbf{post}(\mathbf{v})]$. This cannot happen since $\mathbf{v}$ would be explored from $\mathbf{u}$.

□

Example



Part III

Linear time algorithm for finding all strong connected components of a directed graph

Finding all SCCs of a Directed Graph

Problem

Given a directed graph $\mathbf{G} = (\mathbf{V}, \mathbf{E})$, output *all* its strong connected components.

Finding all SCCs of a Directed Graph

Problem

Given a directed graph $G = (V, E)$, output *all* its strong connected components.

Straightforward algorithm:

```
Mark all vertices in V as not visited.
for each vertex u ∈ V not visited yet do
    find SCC(G, u) the strong component of u:
        Compute rch(G, u) using DFS(G, u)
        Compute rch(G^rev, u) using DFS(G^rev, u)
        SCC(G, u) ⇐ rch(G, u) ∩ rch(G^rev, u)
        ∀u ∈ SCC(G, u):  Mark u as visited.
```

Running time: $O(n(n + m))$

Finding all SCCs of a Directed Graph

Problem

Given a directed graph $\mathbf{G} = (\mathbf{V}, \mathbf{E})$, output *all* its strong connected components.

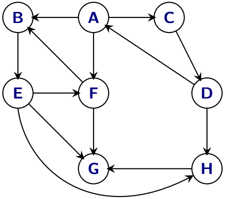
Straightforward algorithm:

```
Mark all vertices in V as not visited.
for each vertex u ∈ V not visited yet do
    find SCC(G, u) the strong component of u:
        Compute rch(G, u) using DFS(G, u)
        Compute rch(G^rev, u) using DFS(G^rev, u)
        SCC(G, u) ⇐ rch(G, u) ∩ rch(G^rev, u)
        ∀u ∈ SCC(G, u):  Mark u as visited.
```

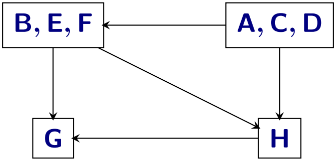
Running time: $\mathbf{O(n(n+m))}$

Is there an $\mathbf{O(n+m)}$ time algorithm?

Structure of a Directed Graph



Graph G



Graph of SCCs G^{SCC}

Reminder

G^{SCC} is created by collapsing every strong connected component to a single vertex.

Proposition

For a directed graph G, its meta-graph G^{SCC} is a DAG.

Linear-time Algorithm for SCCs: Ideas

Exploit structure of meta-graph...

Wishful Thinking Algorithm

1. Let **u** be a vertex in a *sink* SCC of G^{SCC}
2. Do **DFS**$(\mathbf{u})$ to compute $\mathrm{SCC}(\mathbf{u})$
3. Remove $\mathrm{SCC}(\mathbf{u})$ and repeat

Linear-time Algorithm for SCCs: Ideas

Exploit structure of meta-graph...

Wishful Thinking Algorithm

1. Let **u** be a vertex in a *sink* SCC of G^{SCC}
2. Do **DFS(u)** to compute $\mathrm{SCC}(\mathbf{u})$
3. Remove $\mathrm{SCC}(\mathbf{u})$ and repeat

Justification

1. **DFS(u)** only visits vertices (and edges) in $\mathrm{SCC}(\mathbf{u})$

Linear-time Algorithm for SCCs: Ideas

Exploit structure of meta-graph...

Wishful Thinking Algorithm

1. Let $\mathbf{u}$ be a vertex in a *sink* SCC of G^{SCC}
2. Do **DFS**$(\mathbf{u})$ to compute $\mathrm{SCC}(\mathbf{u})$
3. Remove $\mathrm{SCC}(\mathbf{u})$ and repeat

Justification

1. **DFS**$(\mathbf{u})$ only visits vertices (and edges) in $\mathrm{SCC}(\mathbf{u})$
2. ... since there are no edges coming out a sink!
3.
4.

Linear-time Algorithm for SCCs: Ideas

Exploit structure of meta-graph...

Wishful Thinking Algorithm

1. Let **u** be a vertex in a *sink* SCC of G^{SCC}
2. Do **DFS(u)** to compute $SCC(\mathbf{u})$
3. Remove $SCC(\mathbf{u})$ and repeat

Justification

1. **DFS(u)** only visits vertices (and edges) in $SCC(\mathbf{u})$
2. ... since there are no edges coming out a sink!
3. **DFS(u)** takes time proportional to size of $SCC(\mathbf{u})$
4.

Linear-time Algorithm for SCCs: Ideas

Exploit structure of meta-graph...

Wishful Thinking Algorithm

1. Let $\mathbf{u}$ be a vertex in a *sink* SCC of G^{SCC}
2. Do **DFS**$(\mathbf{u})$ to compute $\mathrm{SCC}(\mathbf{u})$
3. Remove $\mathrm{SCC}(\mathbf{u})$ and repeat

Justification

1. **DFS**$(\mathbf{u})$ only visits vertices (and edges) in $\mathrm{SCC}(\mathbf{u})$
2. ... since there are no edges coming out a sink!
3. **DFS**$(\mathbf{u})$ takes time proportional to size of $\mathrm{SCC}(\mathbf{u})$
4. Therefore, total time $\mathbf{O(n + m)}$!

Big Challenge(s)

How do we find a vertex in a sink SCC of G^{SCC}?

Big Challenge(s)

How do we find a vertex in a sink SCC of G^{SCC}?

Can we obtain an *implicit* topological sort of G^{SCC} without computing G^{SCC}?

Big Challenge(s)

How do we find a vertex in a sink SCC of G^{SCC}?

Can we obtain an *implicit* topological sort of G^{SCC} without computing G^{SCC}?

Answer: **DFS(G)** gives some information!

Linear Time Algorithm

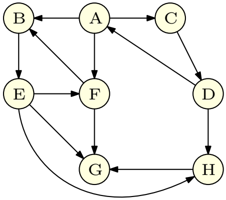
...for computing the strong connected components in $\mathbf{G}$

```
do DFS(G^rev) and output vertices in decreasing post order.
Mark all nodes as unvisited
for each u in the computed order do
    if u is not visited then
        DFS(u)
        Let S_u be the nodes reached by u
        Output S_u as a strong connected component
        Remove S_u from G
```

Theorem

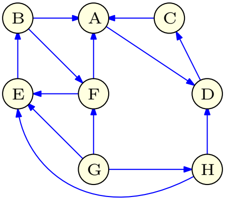
Algorithm runs in time $\mathbf{O(m + n)}$ and correctly outputs all the SCCs of $\mathbf{G}$.

Graph G:



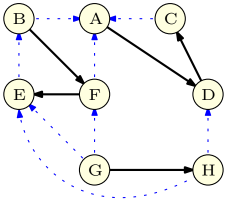
$\Longrightarrow$

Reverse graph $\mathbf{G}^{\mathbf{rev}}$:



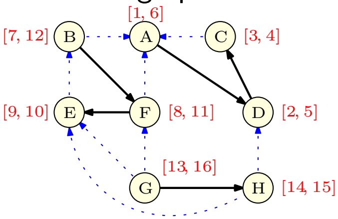
$\Longrightarrow$

DFS of reverse graph:



$\Longrightarrow$

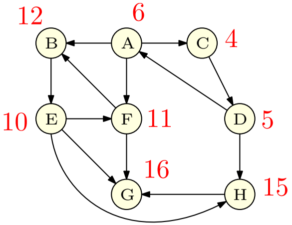
Pre/Post **DFS** numbering of reverse graph:



Linear Time Algorithm: An Example

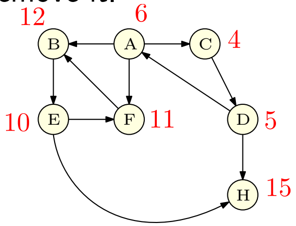
Removing connected components: 1

Original graph G with rev post numbers:



$\Longrightarrow$

Do **DFS** from vertex G remove it.

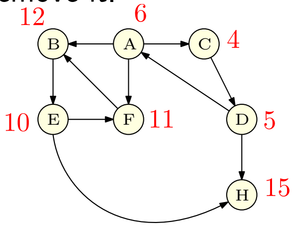


SCC computed:
{G}

Linear Time Algorithm: An Example

Removing connected components: 2

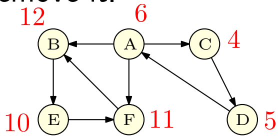
Do **DFS** from vertex G remove it.



SCC computed:
{G}

⟹

Do **DFS** from vertex **H**, remove it.

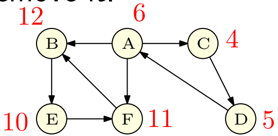


SCC computed:
{G}, {H}

Linear Time Algorithm: An Example

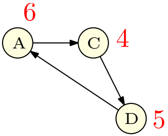
Removing connected components: 3

Do **DFS** from vertex **H**, remove it.



⟹

Do **DFS** from vertex **B**
Remove visited vertices: **{F, B, E}**.



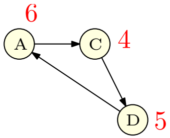
SCC computed:
{G}, {H}

SCC computed:
{G}, {H}, {F, B, E}

Linear Time Algorithm: An Example

Removing connected components: 4

Do **DFS** from vertex **F**
Remove visited vertices:
{F, B, E}.



SCC computed:
{G}, {H}, {F, B, E}

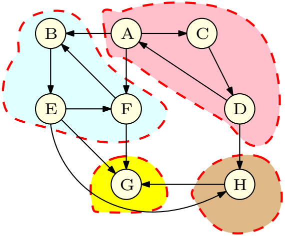
⟹

Do **DFS** from vertex **A**
Remove visited vertices:
{A, C, D}.

SCC computed:
{G}, {H}, {F, B, E}, {A, C, D}

Linear Time Algorithm: An Example

Final result



SCC computed:

{G}, {H}, {F, B, E}, {A, C, D}

Which is the correct answer!

Obtaining the meta-graph...

Once the strong connected components are computed.

Exercise:

Given all the strong connected components of a directed graph $\mathbf{G = (V, E)}$ show that the meta-graph G^{SCC} can be obtained in $\mathbf{O(m + n)}$ time.

Solving Problems on Directed Graphs

A template for a class of problems on directed graphs:

- Is the problem solvable when **G** is strongly connected?
- Is the problem solvable when **G** is a DAG?
- If the above two are feasible then is the problem solvable in a general directed graph **G** by considering the meta graph G^{SCC}?

Part IV

An Application to `make`

Make/Makefile

(A) I know what make/makefile is.
(B) I do NOT know what make/makefile is.

make Utility [Feldman]

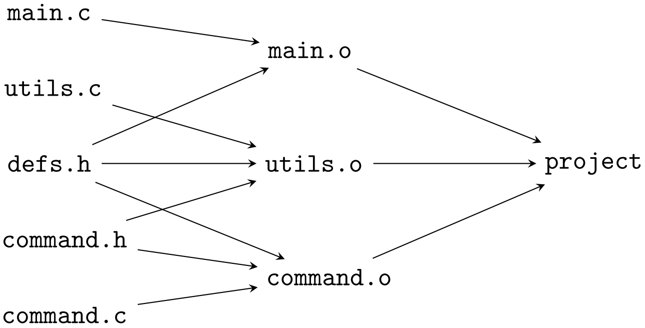
1. Unix utility for automatically building large software applications
2. A makefile specifies
    1. Object files to be created,
    2. Source/object files to be used in creation, and
    3. How to create them

An Example `makefile`

```
project:  main.o utils.o command.o
    cc -o project main.o utils.o command.o

main.o:  main.c defs.h
    cc -c main.c
utils.o:  utils.c defs.h command.h
    cc -c utils.c
command.o:  command.c defs.h command.h
    cc -c command.c
```

makefile as a Digraph



Computational Problems for `make`

1. Is the `makefile` reasonable?
2. If it is reasonable, in what order should the object files be created?
3. If it is not reasonable, provide helpful debugging information.
4. If some file is modified, find the fewest compilations needed to make application consistent.

Algorithms for `make`

1. Is the `makefile` reasonable? Is G a DAG?
2. If it is reasonable, in what order should the object files be created? Find a topological sort of a DAG.
3. If it is not reasonable, provide helpful debugging information. Output a cycle. More generally, output all strong connected components.
4. If some file is modified, find the fewest compilations needed to make application consistent.
    1. Find all vertices reachable (using **DFS**/**BFS**) from modified files in directed graph, and recompile them in proper order. Verify that one can find the files to recompile and the ordering in linear time.

Take away Points

1. Given a directed graph G, its SCCs and the associated acyclic meta-graph G^{SCC} give a structural decomposition of G that should be kept in mind.
2. There is a **DFS** based linear time algorithm to compute all the SCCs and the meta-graph. Properties of **DFS** crucial for the algorithm.
3. DAGs arise in many application and topological sort is a key property in algorithm design. Linear time algorithms to compute a topological sort (there can be many possible orderings so not unique).