ساختمان داده و الگوريتم ها (CE203)

جلسه شانزدهم: جستجوی عمق اول (DFS)

> سجاد شیرعلی شهرضا پاییز 1401 شنبه، 19 آذر 1401

اطلاع رساني

• بخش مرتبط كتاب براى اين جلسه: 22

جستجوى عمق اول (DFS)

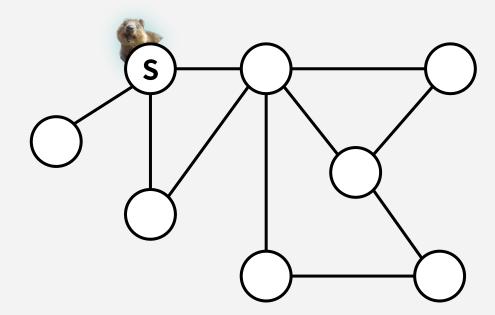
یک روش دیگر برای پیمایش گراف

BFS vs. DFS

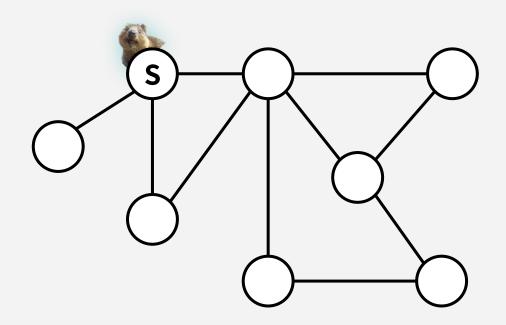
Literally just BREADTH vs DEPTH:

While BFS first explores the nodes closest to the "source" and then moves outwards in layers, DFS goes as far down a path as it can before it comes back to explore other options.

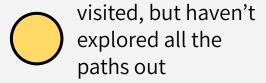
An analogy:

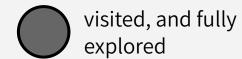


An analogy:

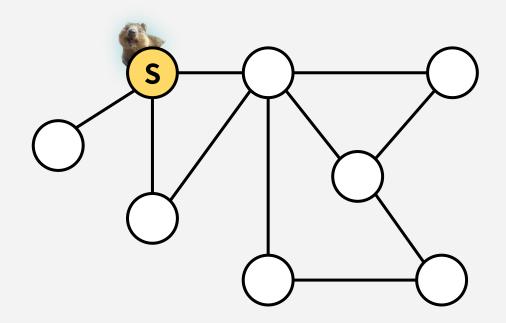




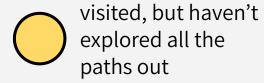




An analogy:

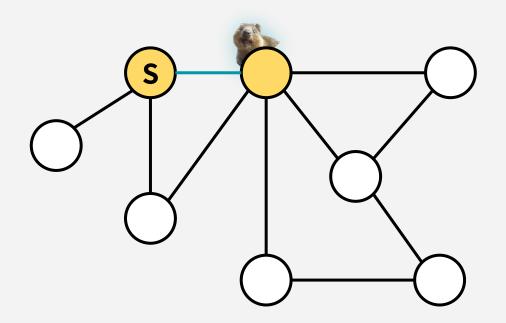




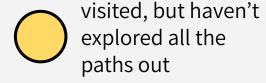




An analogy:

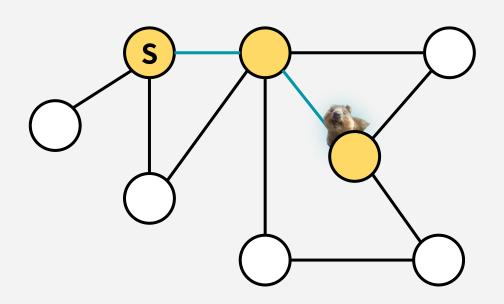








An analogy:

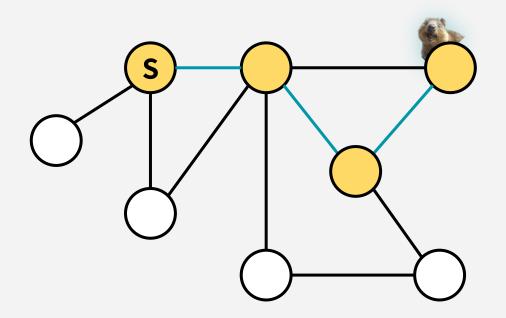




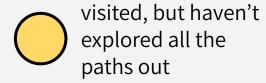


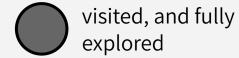


An analogy:

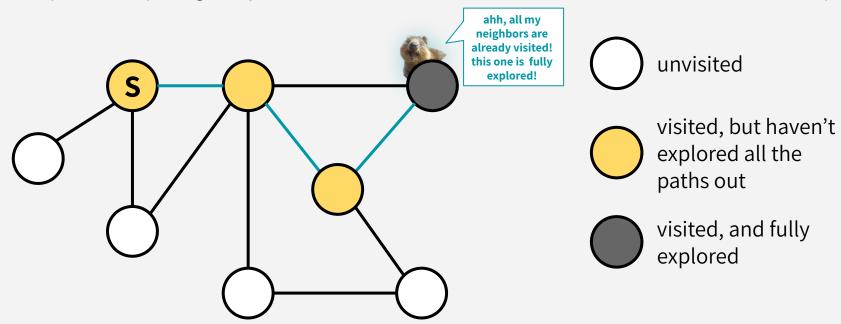




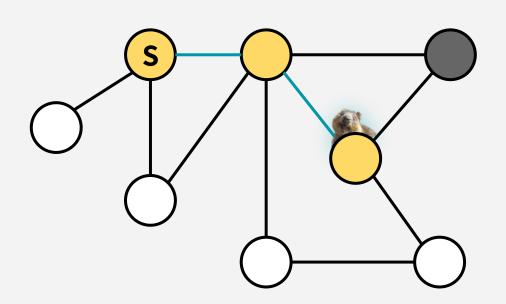




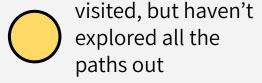
An analogy:



An analogy:

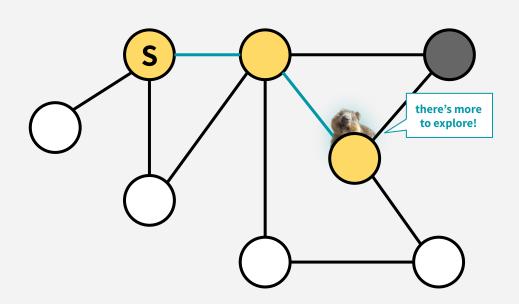




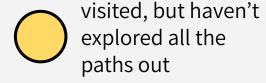


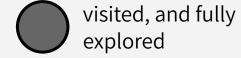


An analogy:

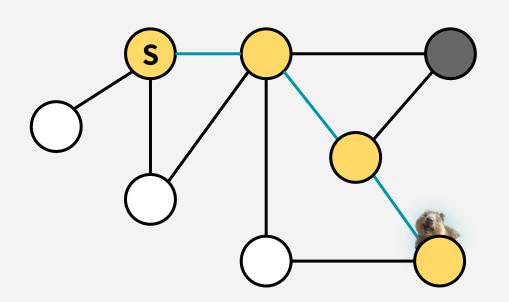




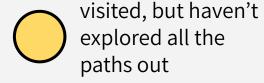




An analogy:

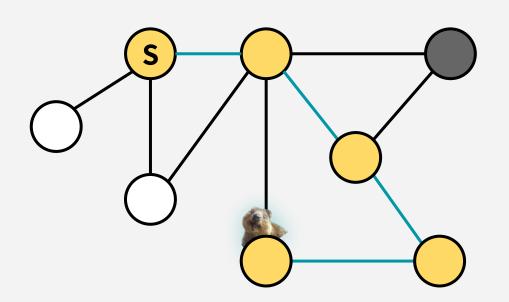




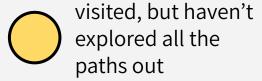


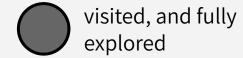


An analogy:

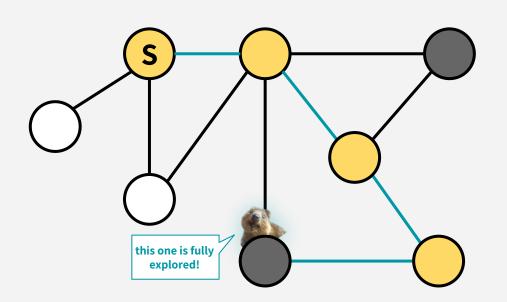




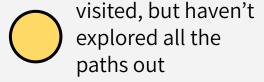




An analogy:

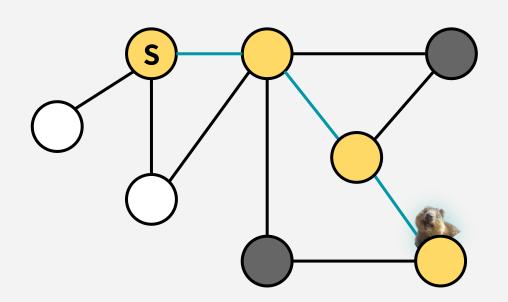




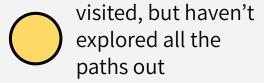




An analogy:

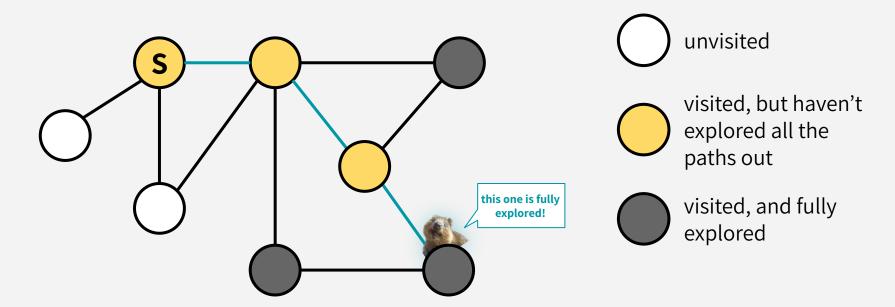




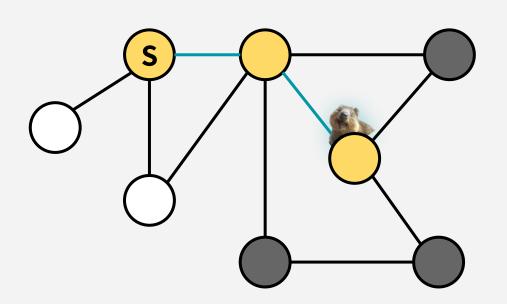




An analogy:



An analogy:

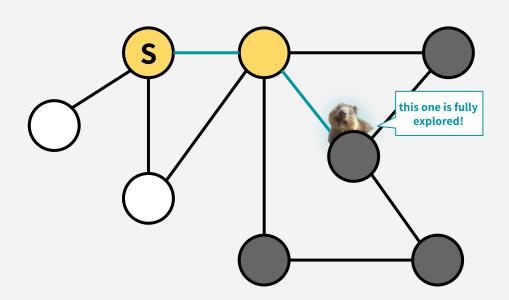








An analogy:

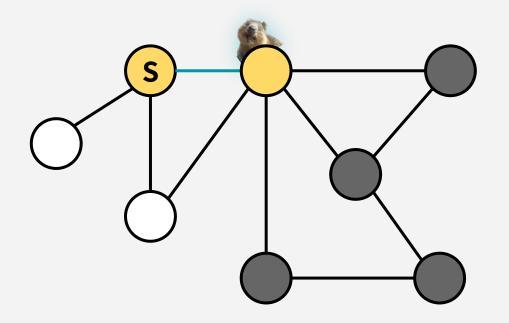




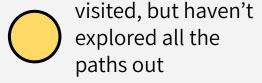




An analogy:

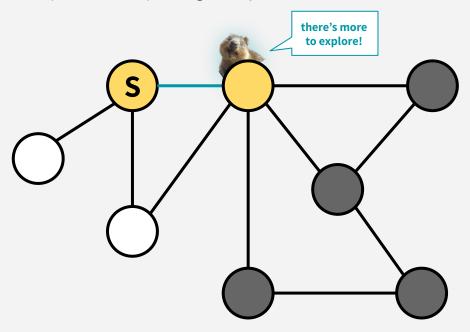








An analogy:

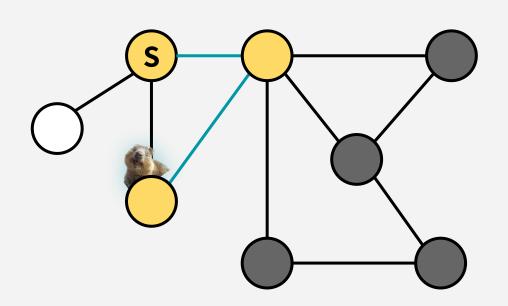








An analogy:

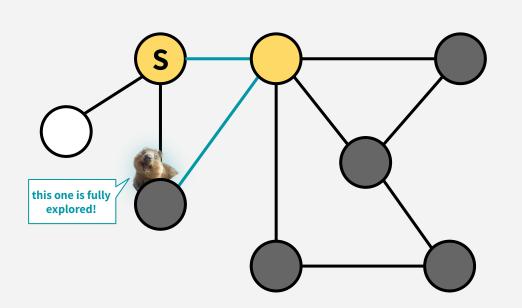








An analogy:

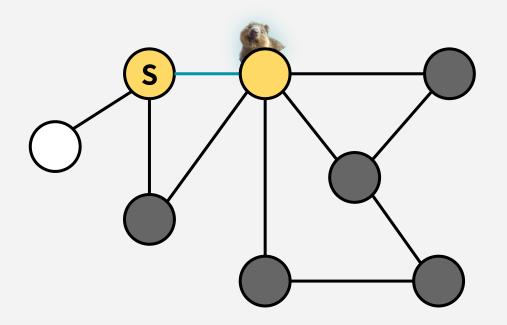




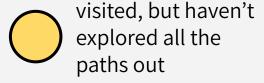




An analogy:

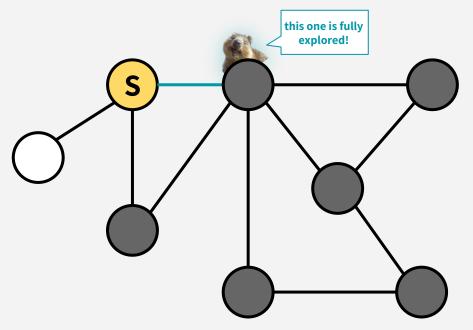








An analogy:

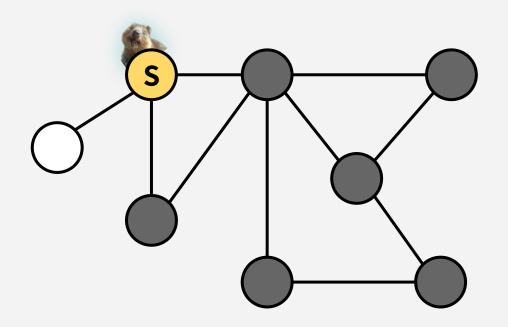




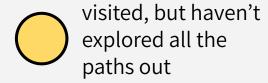




An analogy:

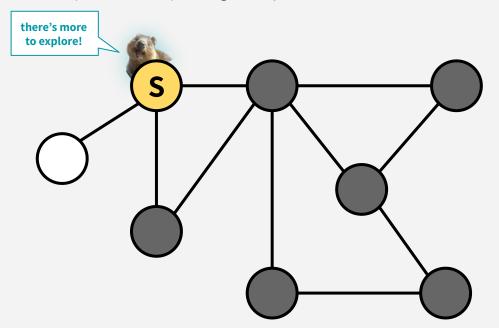




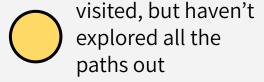




An analogy:

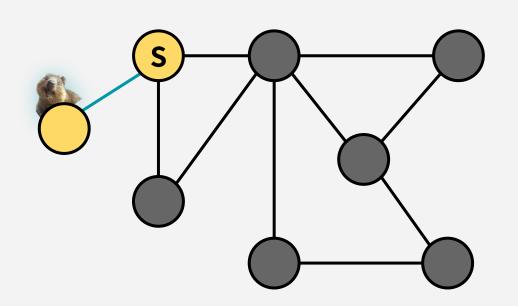




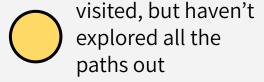




An analogy:

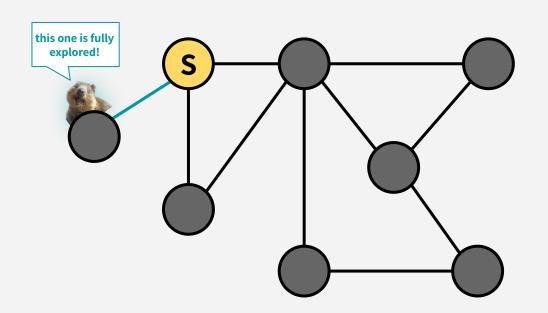




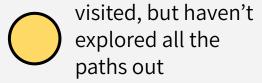




An analogy:

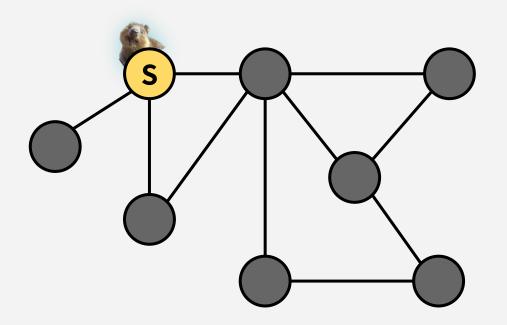




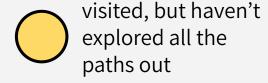


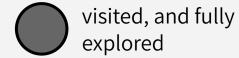


An analogy:

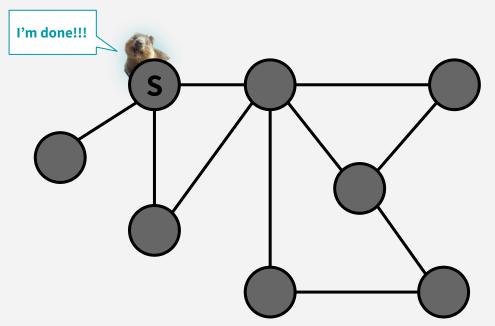




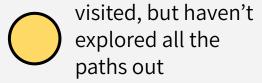




An analogy:









An analogy:

A smart quokka is exploring a labyrinth with chalk (to mark visited destinations) & thread (to retrace steps)

I'm done!!!

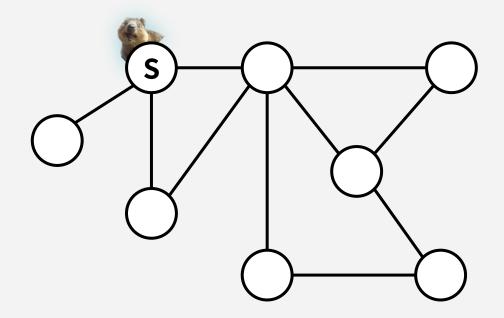
In addition to keeping track of the visited status of nodes, we're going to keep track of:

ut haven't all the

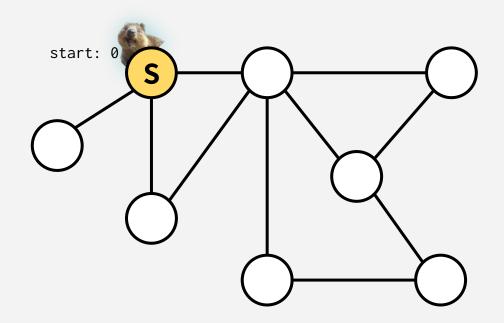
nd fully

You've probably seen other ways to implement DFS, all this extra bookkeeping will be useful for us later!

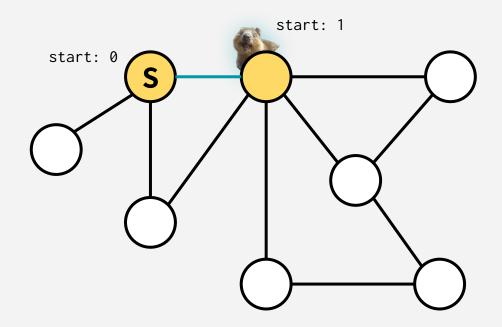
An analogy:



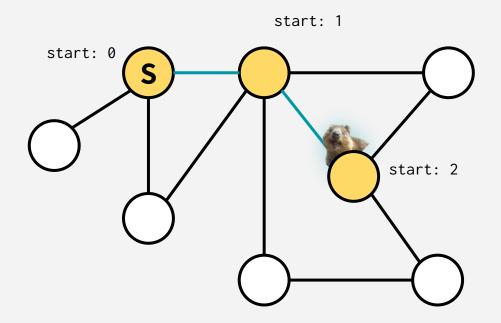
An analogy:



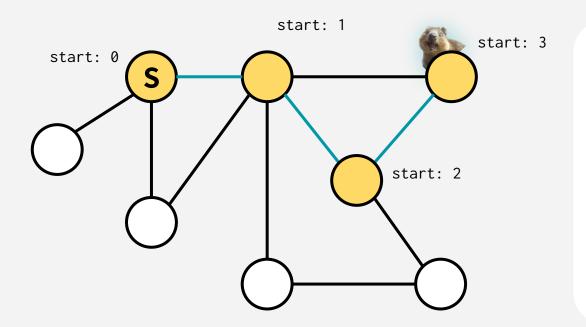
An analogy:



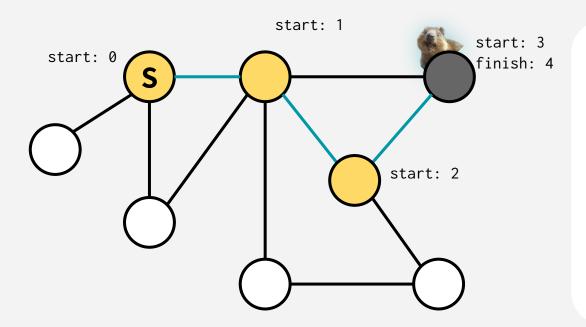
An analogy:



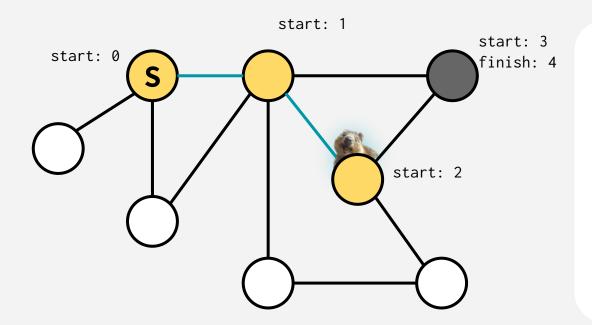
An analogy:



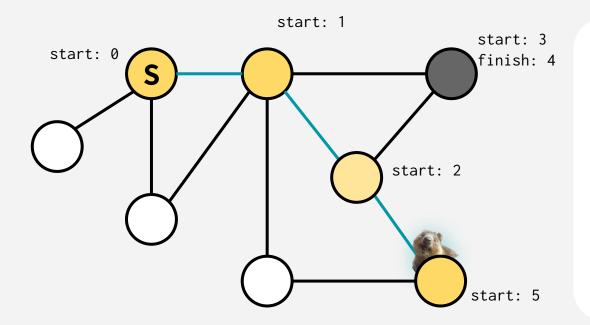
An analogy:



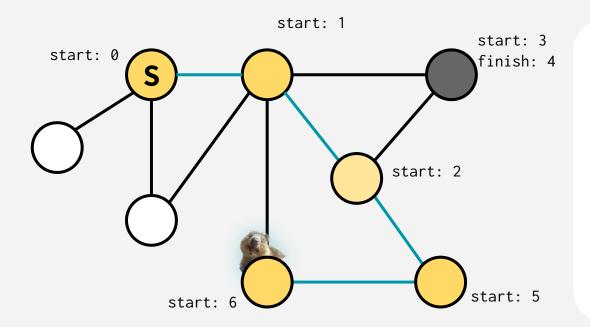
An analogy:



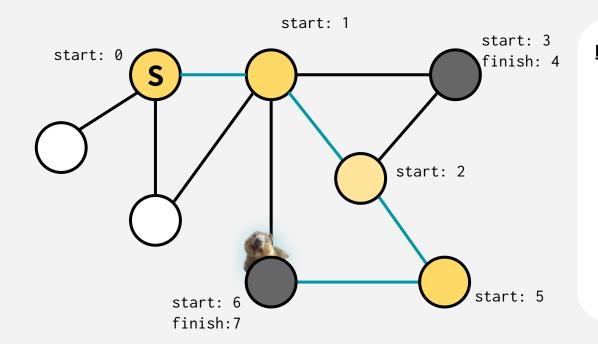
An analogy:



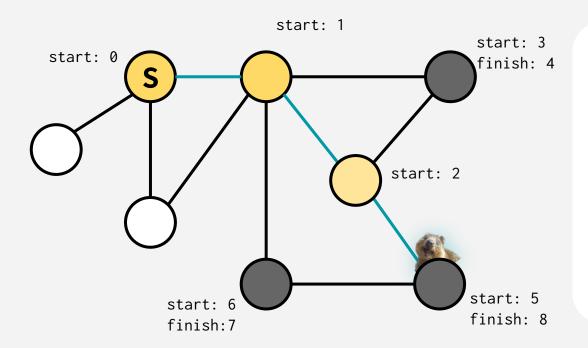
An analogy:



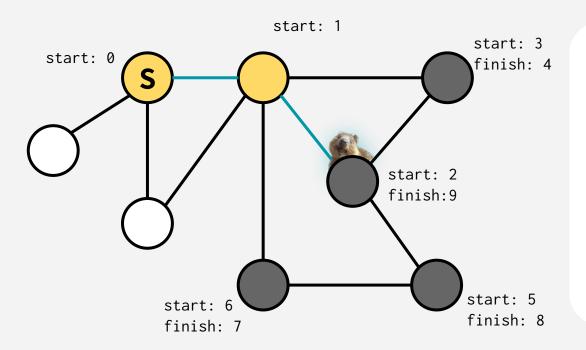
An analogy:



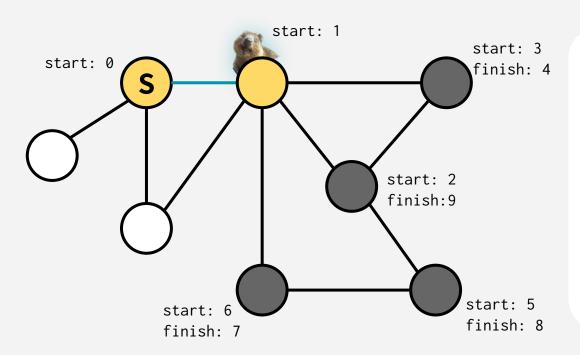
An analogy:



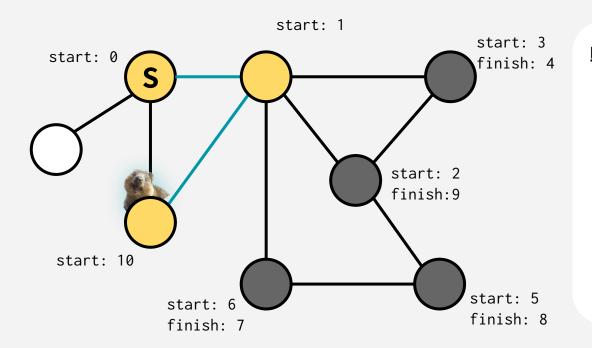
An analogy:



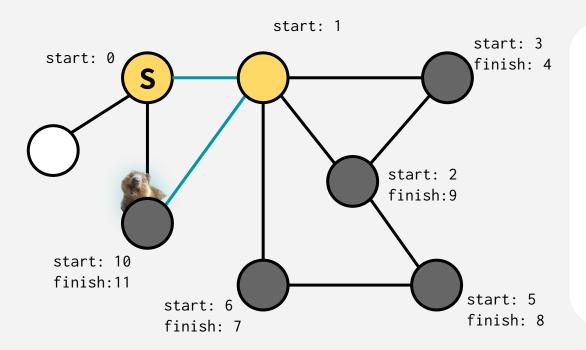
An analogy:



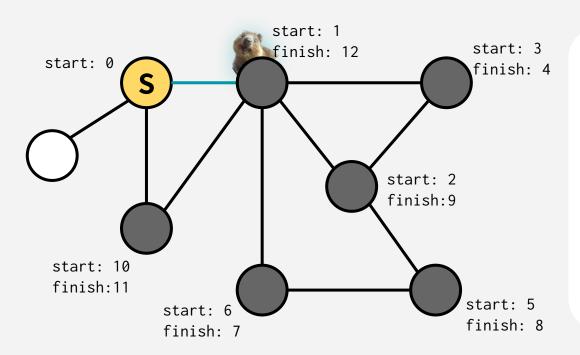
An analogy:



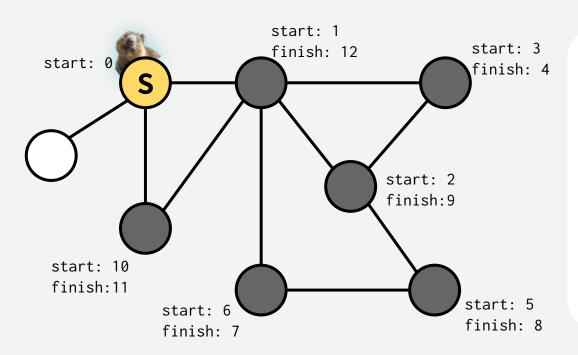
An analogy:



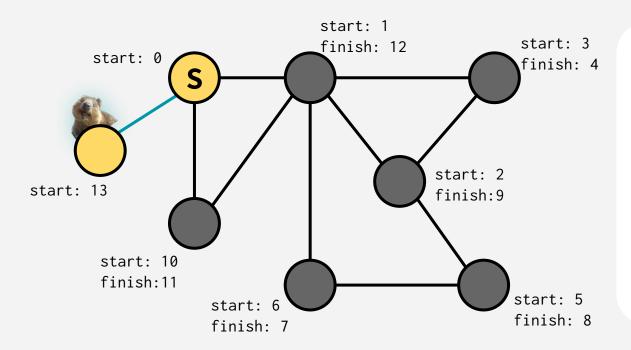
An analogy:



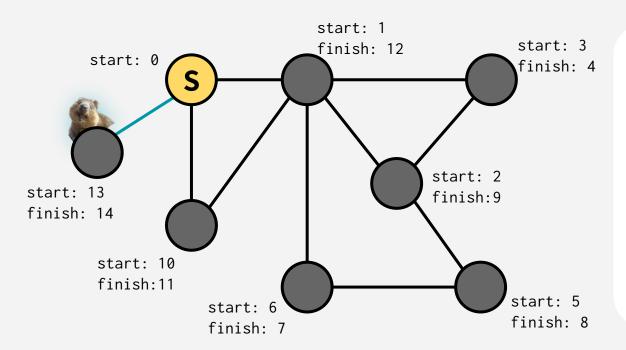
An analogy:



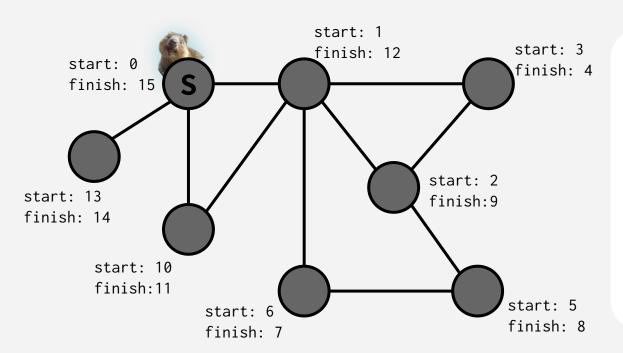
An analogy:



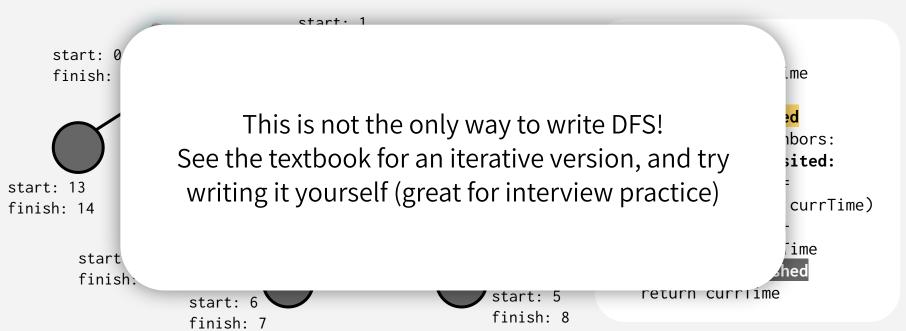
An analogy:



An analogy:

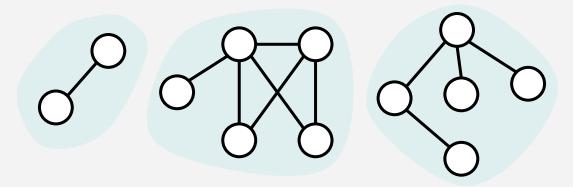


An analogy:



Like BFS, DFS finds all the nodes reachable from the starting point!

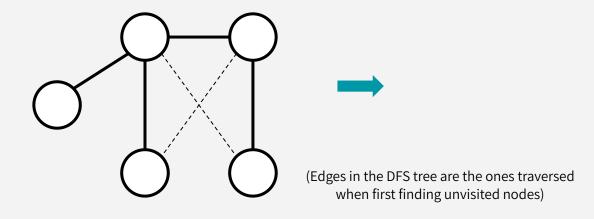
In undirected graphs, this is equivalent to finding a **connected component.**



Why is it called depth-first?

We are implicitly building a **tree**!

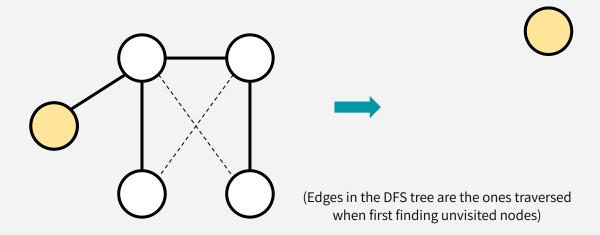
(It's a tree because we never revisit a node)



Why is it called depth-first?

We are implicitly building a **tree**!

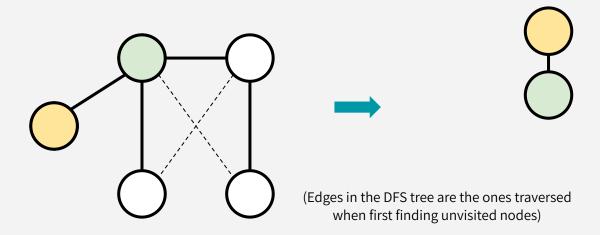
(It's a tree because we never revisit a node)



Why is it called depth-first?

We are implicitly building a **tree**!

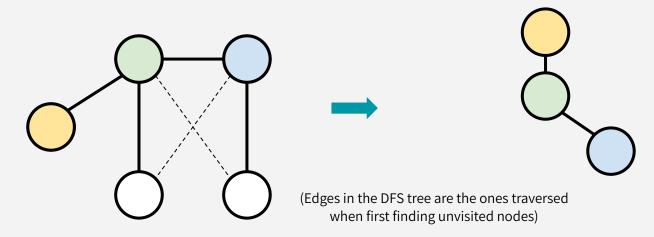
(It's a tree because we never revisit a node)



Why is it called depth-first?

We are implicitly building a **tree**!

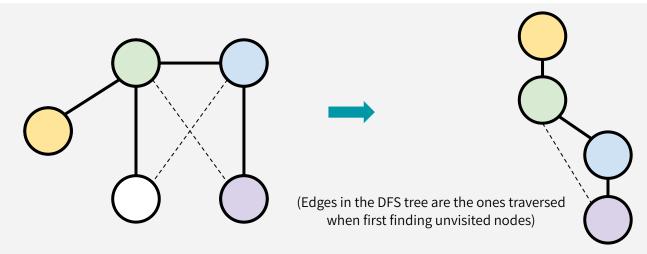
(It's a tree because we never revisit a node)



Why is it called depth-first?

We are implicitly building a **tree**!

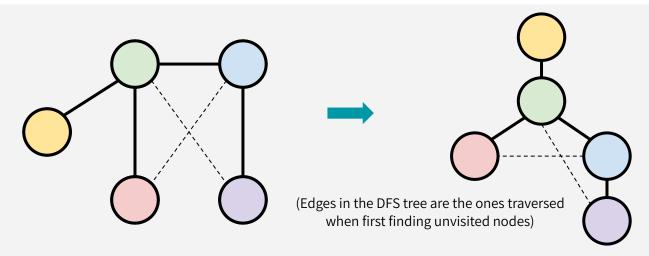
(It's a tree because we never revisit a node)



Why is it called depth-first?

We are implicitly building a **tree**!

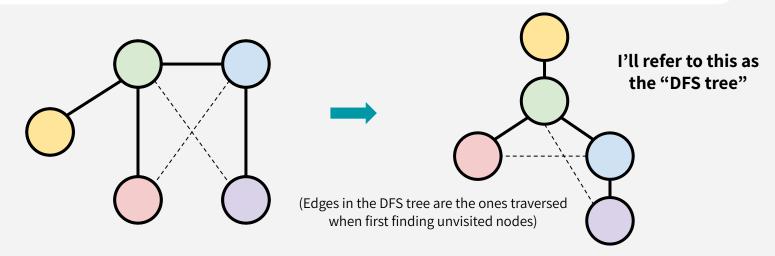
(It's a tree because we never revisit a node)



Why is it called depth-first?

We are implicitly building a **tree**!

(It's a tree because we never revisit a node)



DEPTH-FIRST SEARCH: RUNTIME

To explore a graph's **i**th **connected component** (n_i nodes, m_i edges):

We visit each vertex in the CC exactly once ("visit" = "call DFS on"). At each vertex v, we:

- Do some bookkeeping: O(1)
- Loop over v's neighbors & check if they are visited (& then potentially make a recursive call): O(1) per neighbor → O(deg(v)) total.

Total:
$$\sum_{v} O(deg(v)) + \sum_{v} O(1) = O(m_i + n_i)$$

DEPTH-FIRST SEARCH: RUNTIME

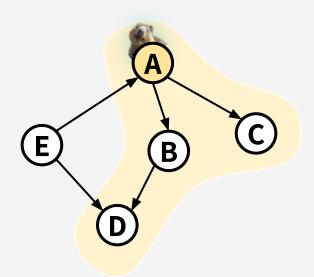
To explore **the entire graph** (n nodes, m edges):

A graph might have multiple connected components! To **explore the whole graph**, we would call our DFS routine once for each connected component (note that each vertex and each edge participates in exactly one connected component). The combined running time would be:

$$O(\sum_{i} m_{i} + \sum_{i} n_{i}) = O(m + n)$$

DFS works fine on directed graphs too!

From a start node x, DFS would find all nodes *reachable* from x. (In directed graphs, "connected component" isn't as well defined... more on that later!)



Verify this on your own:

running DFS from A would still find all nodes reachable from A (E isn't reachable from A in this directed graph).



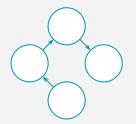
مرتب سازی توپولوژیکی

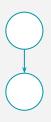
یک کاربرد از جستجوی عمق اول برای مسائل دارای پیش نیاز

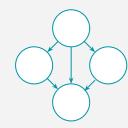
ASIDE: DIRECTED ACYCLIC GRAPHS

A **Directed Acyclic Graph (DAG)** is a directed graph with *no directed cycles*.

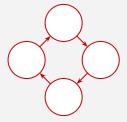
These are DAGs:

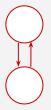


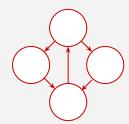




These are not DAGs:







Given a DAG, find an ordering of vertices so that all of the dependency requirements are met

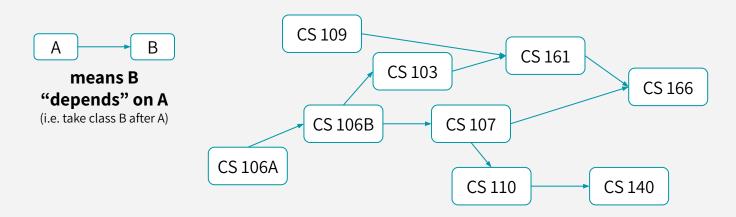
Example applications:

Given a package dependency graph, in what order should packages be installed? Given a course prerequisites graph, in what order should we take classes?

Given a DAG, find an ordering of vertices so that all of the dependency requirements are met

Example applications:

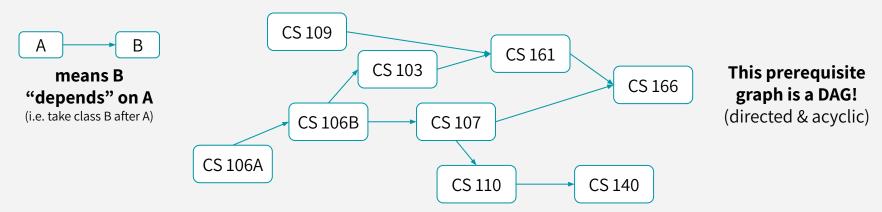
Given a package dependency graph, in what order should packages be installed? Given a course prerequisites graph, in what order should we take classes?



Given a DAG, find an ordering of vertices so that all of the dependency requirements are met

Example applications:

Given a package dependency graph, in what order should packages be installed? Given a course prerequisites graph, in what order should we take classes?



Given a DAG, find an ordering of vertices so that all of the dependency requirements are met

What does "meeting the dependency requirements" mean?

We want to produce an ordering such that:

for every edge (v, w) in E, v must appear before w in the ordering (e.g. CS103 must come before CS161)

(i.e. take class B atter A)

CS 106B

CS 107

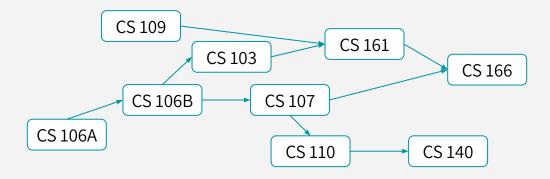
CS 106B

CS 107

CS 140

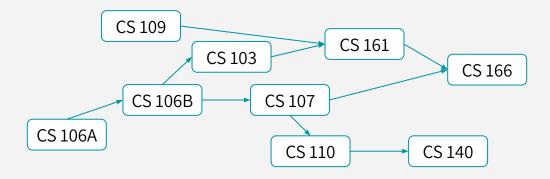
te

It's helpful to think of this as "linearizing" the graph, where all edges point to the right

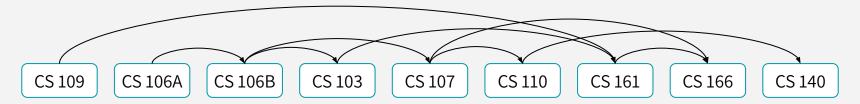


A correct "toposort" of this DAG:

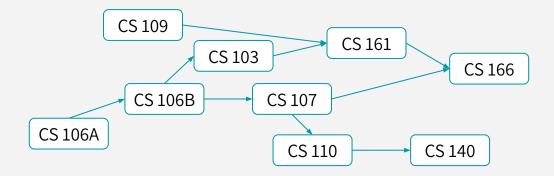
It's helpful to think of this as "linearizing" the graph, where all edges point to the right



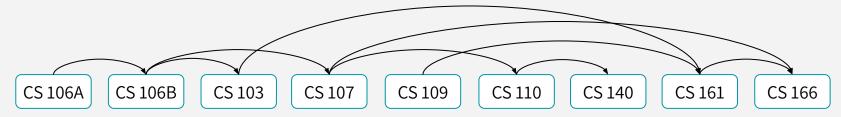
A correct "toposort" of this DAG:



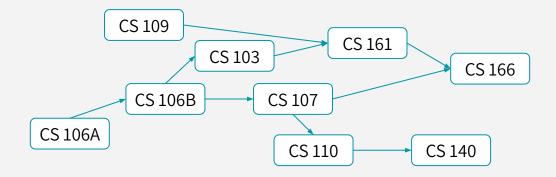
It's helpful to think of this as "linearizing" the graph, where all edges point to the right



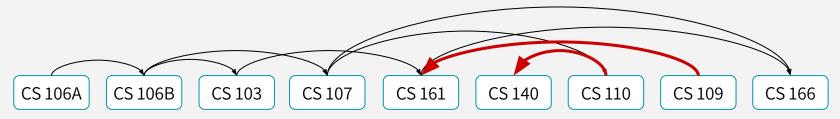
Also a correct toposort of this DAG:



It's helpful to think of this as "linearizing" the graph, where all edges point to the right



Not a correct toposort of this DAG:



TOPOSORT ON NON-DAGS?

We assume these "dependency" graphs are all DAGs!

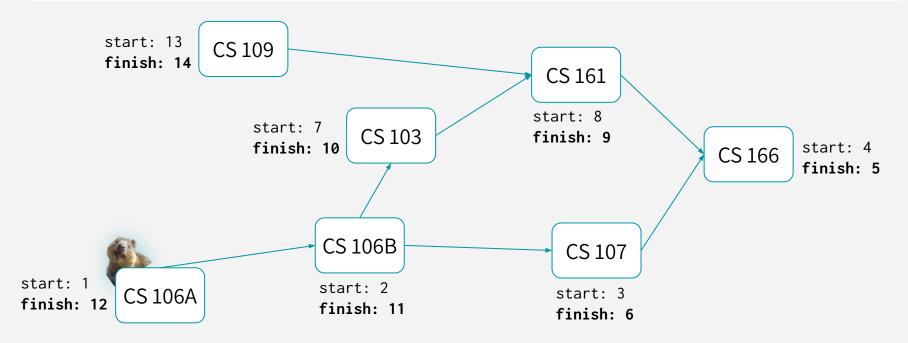
What about other graphs? Undirected graphs? Directed graphs with cycles?

Toposort gives us a priority ordering of nodes (e.g. more intro classes are "higher priority" than more advanced classes). Edges in DAGs clearly illustrate priority: edge from **x** to **y** means **x** has priority over **y**.

In an undirected graph, if there's an **x-y** edge, which node has "priority"?

In a graph with cycles, if **x** and **y** are part of a cycle, then **x** can reach **y** and **y** can also reach **x**... so which node has "priority"?

Let's run DFS. What do you notice about the finish times? What does it have to do with toposort?



Let's run DFS. What do you notice about the finish times? What does it have to do with toposort?

CLAIM: In general, if there's an edge from **v** → **w**, **v**'s finish time will be *larger* than **w**'s finish time

Let's consider two cases: (1) DFS visits **v** first, or (2) DFS visits **w** first.

start: 1

Cinish: 12 CS 106

start: 2

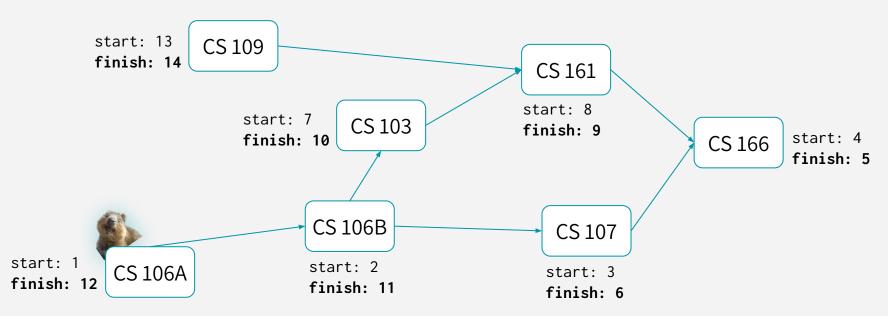
finish: 11

start: 3

finish: 6

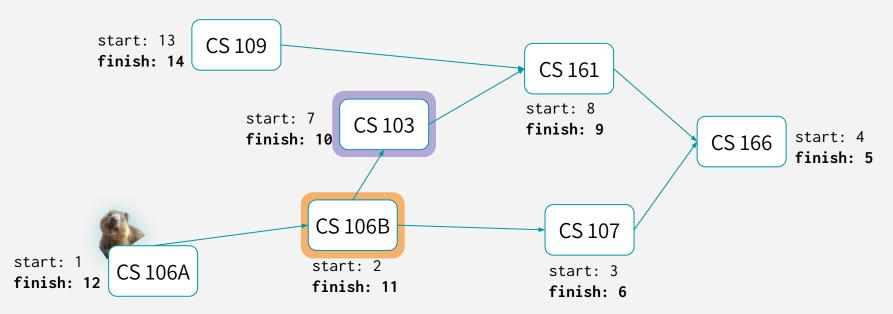
CLAIM: In general, if there's an edge from $\mathbf{v} \to \mathbf{w}$, \mathbf{v} 's finish time will be *larger* than \mathbf{w} 's finish time

<u>CASE 1</u>: $\mathbf{v} \rightarrow \mathbf{w}$, and \mathbf{v} is discovered first by DFS



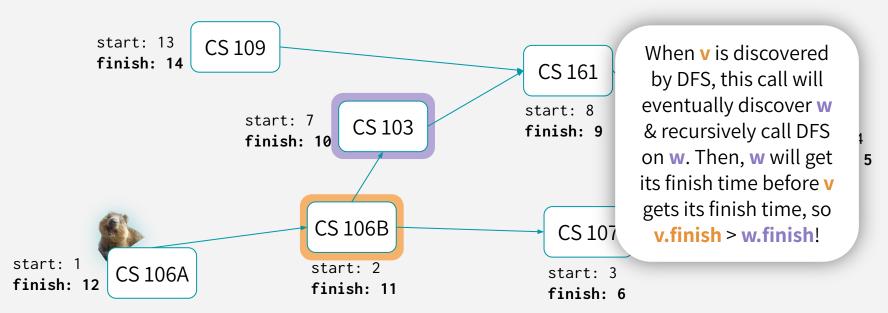
CLAIM: In general, if there's an edge from $\mathbf{v} \to \mathbf{w}$, \mathbf{v} 's finish time will be *larger* than \mathbf{w} 's finish time

<u>CASE 1</u>: $\mathbf{v} \rightarrow \mathbf{w}$, and \mathbf{v} is discovered first by DFS



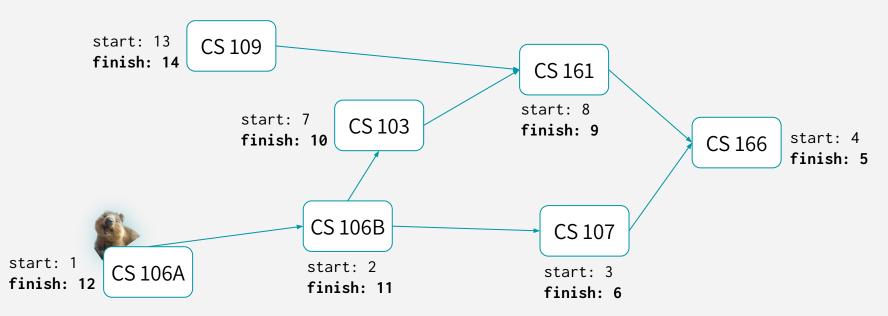
CLAIM: In general, if there's an edge from $\mathbf{v} \to \mathbf{w}$, \mathbf{v} 's finish time will be *larger* than \mathbf{w} 's finish time

<u>CASE 1</u>: $\mathbf{v} \rightarrow \mathbf{w}$, and \mathbf{v} is discovered first by DFS



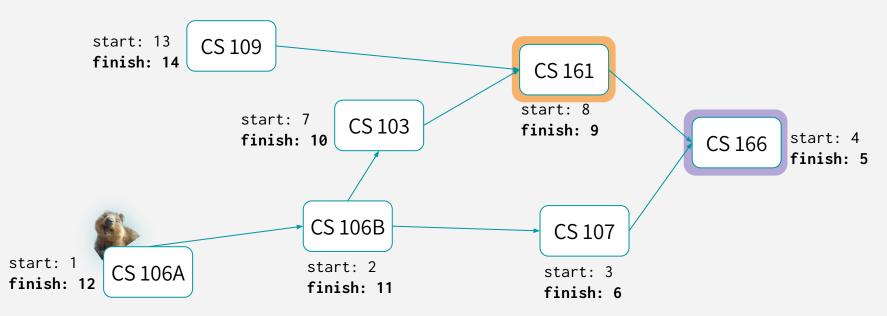
CLAIM: In general, if there's an edge from $\mathbf{v} \to \mathbf{w}$, \mathbf{v} 's finish time will be *larger* than \mathbf{w} 's finish time

<u>CASE 2</u>: $\mathbf{v} \rightarrow \mathbf{w}$, and \mathbf{w} is discovered first by DFS



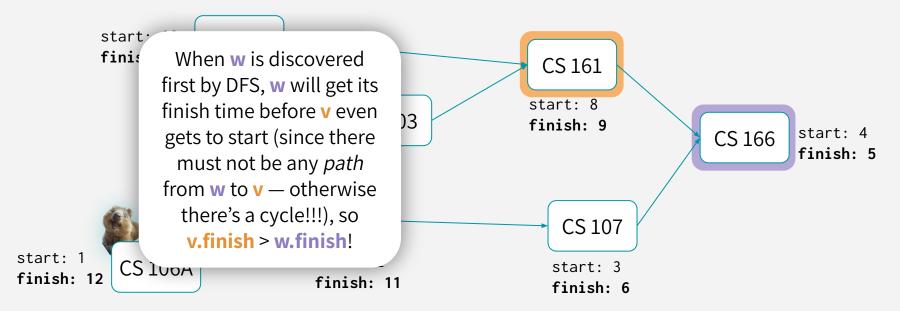
CLAIM: In general, if there's an edge from $\mathbf{v} \to \mathbf{w}$, \mathbf{v} 's finish time will be *larger* than \mathbf{w} 's finish time

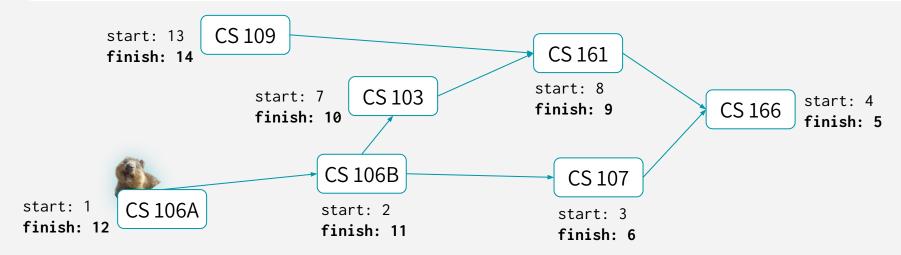
<u>CASE 2</u>: $\mathbf{v} \rightarrow \mathbf{w}$, and \mathbf{w} is discovered first by DFS



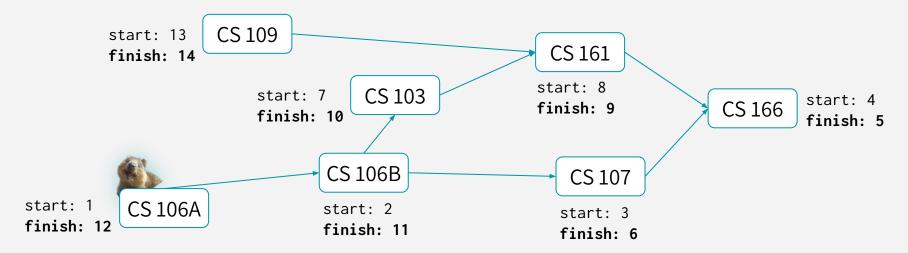
CLAIM: In general, if there's an edge from $\mathbf{v} \to \mathbf{w}$, \mathbf{v} 's finish time will be *larger* than \mathbf{w} 's finish time

<u>CASE 2</u>: $\mathbf{v} \rightarrow \mathbf{w}$, and \mathbf{w} is discovered first by DFS

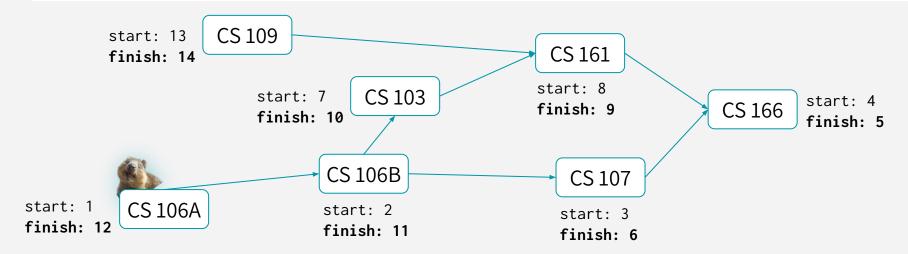




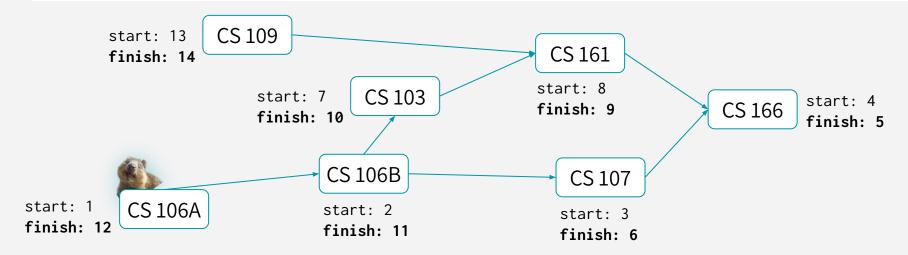
TOPOSORT: Perform DFS. When a vertex gets its finish time, insert it at the start of the list.



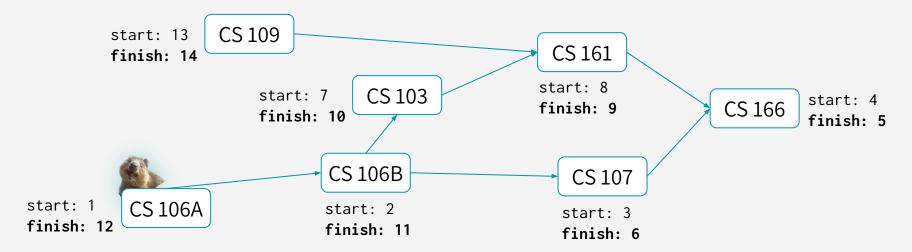
CS 166 **f: 5**

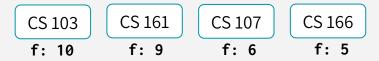


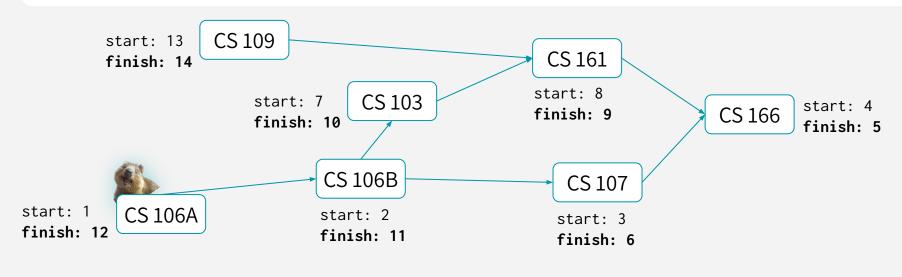


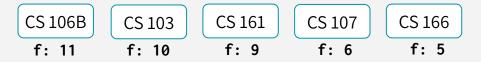


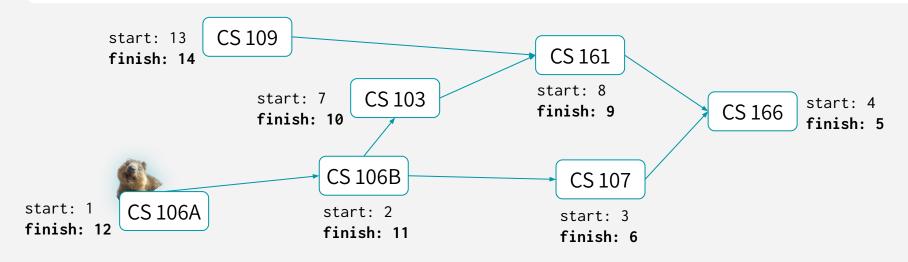




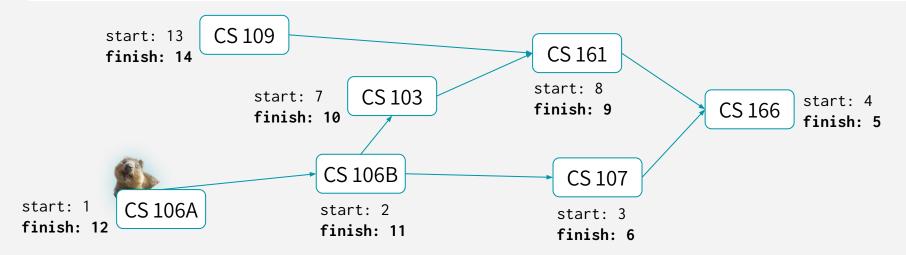


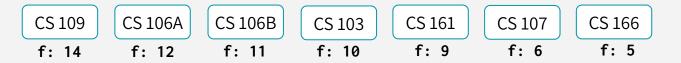


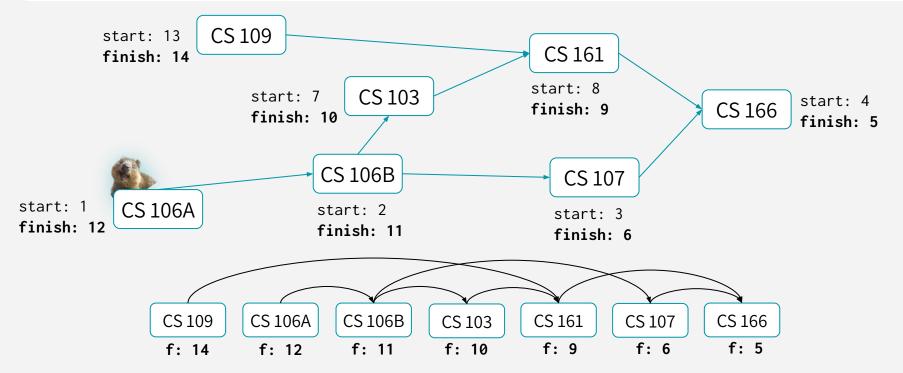


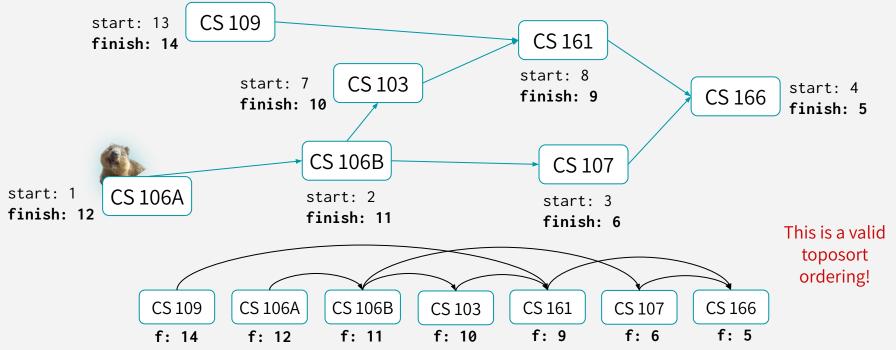








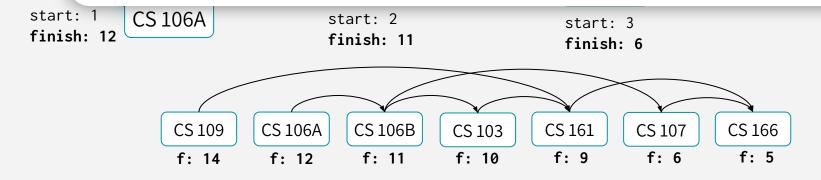




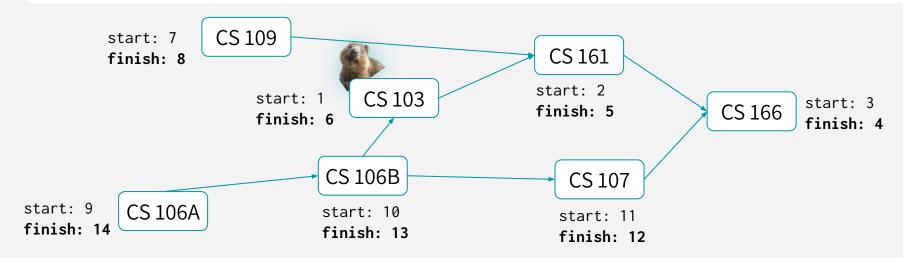
TOPOSORT: Perform DFS. When a vertex gets its finish time, insert it at the start of the list.



Regardless of which vertex your DFS starts, it'll get you a correct Toposort ordering of your DAG

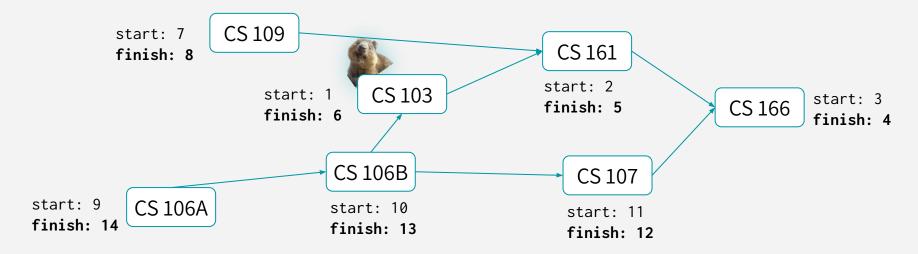






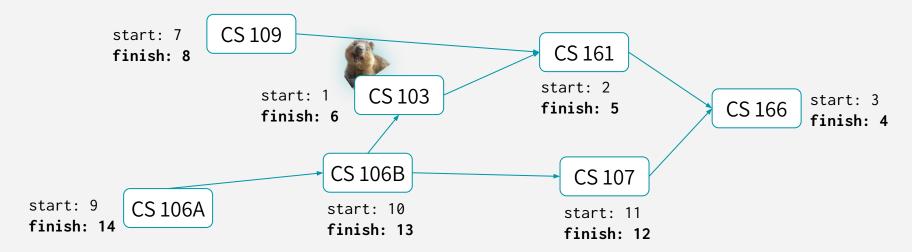


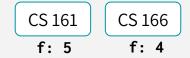
TOPOSORT: Perform DFS. When a vertex gets its finish time, insert it at the start of the list.



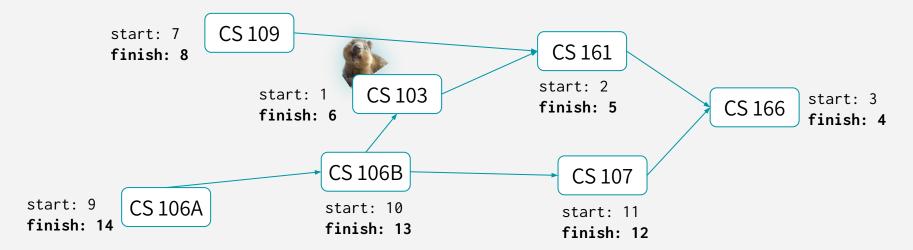
CS 166 **f: 4**

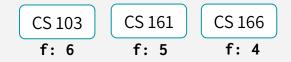




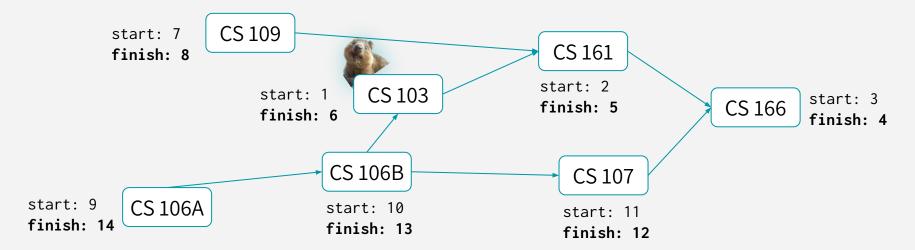






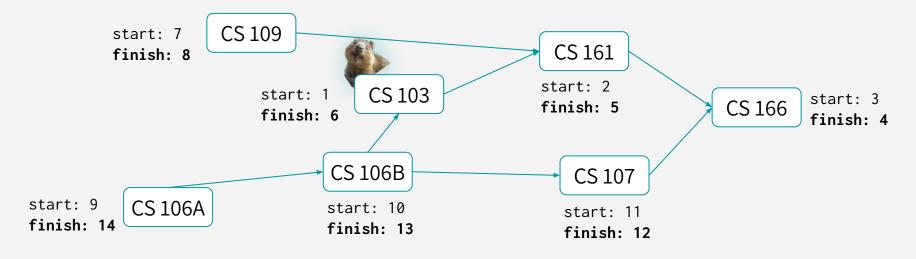


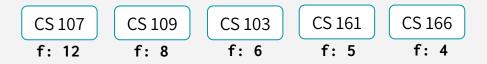




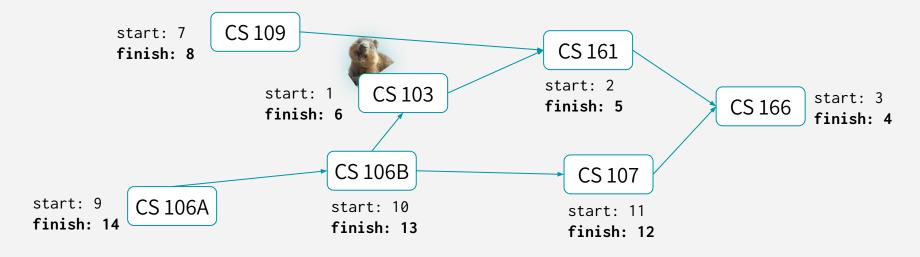


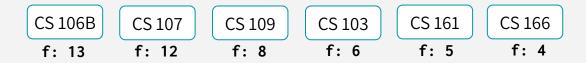




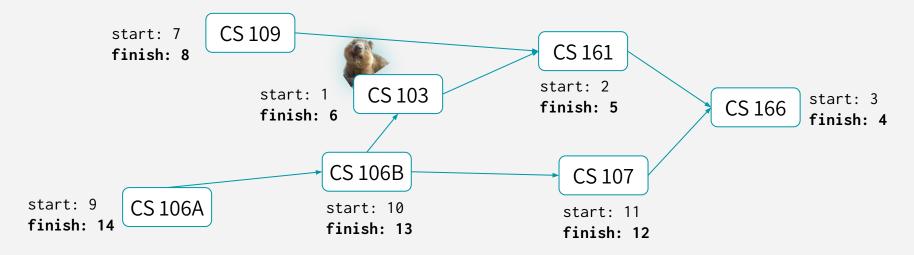


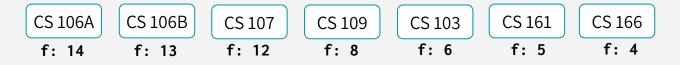




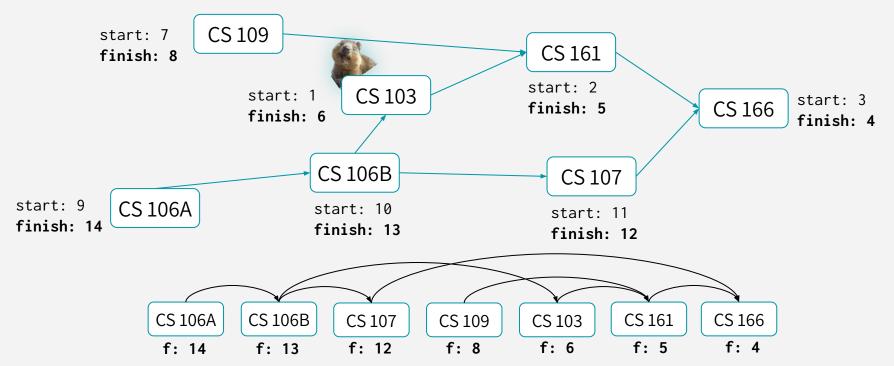




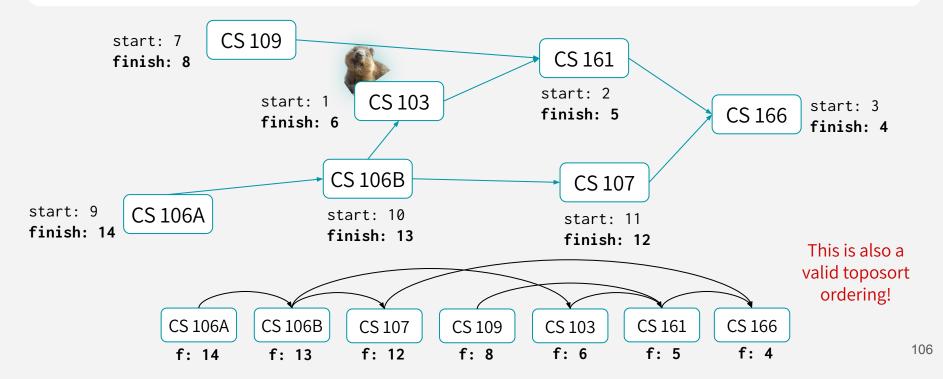












DFS & TOPOSORT RECAP

DFS can help you solve the Topological Sorting Problem.

That's just the fancy name for the problem of finding an ordering of the vertices which respect all the dependencies.

