CS 170 Homework 3

Due **Saturday 2/15/2025, at 10:00 pm (grace period until 11:59pm)**

1 Study Group

List the names and SIDs of the members in your study group. If you have no collaborators, explicitly write "none".

2 Depth First Search

Depth first search is a useful and often efficient way to organize computations on a graph.

Let G be an undirected connected tree, and let $wt : E \to \mathbb{R}^+$ be positive weights on its edges. We show a template for DFS-based computations below.

```
1: Input: Undirected connected tree G = (V, E) and positive weights wt(u, v) for each
   edge (u, v) ∈ E
2:
3: Initialization:
4: visited[v] ← False for all vertices v.
5: A[v] ← 0 and B[v] ← 0 for all vertices v.
6: t ← 1.
7:
8: function Explore(Vertex u)
9:     visited[u] ← True
10:    for each edge (u, v) in E do
11:        if NOT visited[v] then
12:            PreVisit(u, v)
13:            Explore(v)
14:            PostVisit(u, v)
```

DFS can be used for different purposes by defining the procedures PreVisit and PostVisit appropriately.

(a) In each of the following cases, PreVisit and PostVisit have been defined for you. After execution, the array $A[v]$ will hold a value for each vertex v. Describe in words what $A[v]$ represents.

(i)
```
1: procedure PreVisit(u, v)
2:     return
3:
4: procedure PostVisit(u, v)
5:     A[u] ← max(A[u], A[v] + 1)
```

Solution:

$A[v]$ represents the length of the longest path from v to a leaf in its subtree.

(ii)
```
1: procedure PreVisit(u, v)
2:     B[u] ← B[u] + 1
3:
4: procedure PostVisit(u, v)
5:     A[u] ← max(A[u], B[v] + 1)
```

Solution: $A[v]$ represents the maximum degree among all the children of v.

(b) In each of the following cases, write down pseudocode for PreVisit and PostVisit routines to perform the computation needed.

(i) For each vertex v, compute the maximum weight of an edge along the path from root r to vertex v and store it in array $A[v]$.

Solution:
```
1: procedure PreVisit(u, v)
2:     A[v] ← max(A[u], wt[u, v])
3:
4: procedure PostVisit(u, v)
5:     return
```

(ii) For each vertex v, compute the maximum weight of any edge in the subtree rooted at vertex v and store it in array $A[v]$.

Solution:
```
1: procedure PreVisit(u, v)
2:     return
3:
4: procedure PostVisit(u, v)
5:     A[u] ← max(A[u], A[v], wt[u, v])
```

(iii) For each vertex v, compute the maximum pre-order number of any of its children and store it in array $A[v]$. If v has no children, then $A[v]$ should be 0.

Solution:
```
1: procedure PreVisit(u, v)
2:     t ← t + 1
3:     A[u] ← t
4:
5: procedure PostVisit(u, v)
6:     t ← t + 1
```

3 Biconnected Components

Consider any undirected connected graph $G = (V, E)$. We say that an edge $(u, v) \in E$ is *critical* if removing it disconnects the graph. In other words, the graph $(V, E \setminus (u, v))$ is no longer connected.
Similarly, we call a vertex $v \in V$ critical if removing v (and all its incident edges) leaves the graph disconnected.

(a) Suppose that $|V| \geq 2$. Can you always find a vertex $v \in V$ that is **not** critical? What about an edge that is not critical?

(b) Give a linear time algorithm to find all the critical edges of G.

(c) Modify your algorithm above to find all the critical vertices of G.

Solution:

(a) Consider running DFS from any vertex on the graph, and any leaf in the resulting DFS tree. The leaf can be removed without disconnecting the graph, since the remaining vertices are connected using the DFS tree edges. But we can't always find such an edge. For example, every edge in a tree is critical.

(b) Perform DFS on G while keeping track of $pre[v]$ for each vertex. We also maintain a low value for each vertex, where $low[v]$ denotes the smallest $pre[u]$ such that there is a back edge to u from the subtree of v. The only potential critical edges are the tree edges in the DFS. An edge between v and its parent p is critical if and only if $low[v] > pre[p]$ (i.e. there does not exist a back edge from v to p or any of its ancestors).

(c) The root of the DFS tree is critical iff it has more than one child. A non-root vertex v is critical iff for at least one of its children c, $low[c] > pre[v]$.

4 Topological Sort Proofs

(a) A directed acyclic graph G is *semiconnected* if for any two vertices A and B, there is a path from A to B or a path from B to A. Show that G is semiconnected if and only if there is a directed path that visits all of the vertices of G. Make sure to prove both sides of the "if and only if" condition.

Hint: Is there a specific arrangement of the vertices that can help us solve this problem?

Solution: First, we show that the existence of a directed path p that visits all vertices implies that G is semiconnected. For any two vertices A and B, consider the subpath of p between A and B. If A appears before B in p, then this subpath will go from A to B. Otherwise, it will go from B to A. In either case, A and B are semiconnected for all pairs of vertices (A, B) in G.

Now we show that if G is semiconnected, then there is a directed path that visits all of the vertices. Consider a topological ordering $v_1, v_2, \ldots, v_n$ of the vertices in G. For any pair of consecutive vertices v_i, v_{i+1}, we know that there is a path from v_i to v_{i+1} or from v_{i+1} to v_i by semiconnectedness. But topological orderings do not have any edges from later vertices to earlier vertices. Therefore, there is a path from v_i to v_{i+1} in G. This path cannot visit any other vertices in G because the path cannot travel from later vertices to earlier vertices in the topological ordering. Therefore, the path from v_i to v_{i+1} must be a single edge from v_i to v_{i+1}. This edge exists for any consecutive pair of vertices in the topological ordering, so there is a path from v_1 to v_n that visits all vertices of G.

(b) Show that a DAG has a unique topological ordering if and only if it has a directed path that visits all of its vertices.

Remark: This means that a semiconnected DAG always has a unique topological ordering.

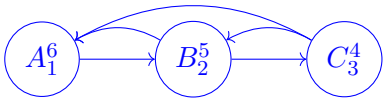
Solution: If a DAG has a directed path that visits all of its vertices, then arranging the vertices in the order they appear in the path will yield a topological ordering, as no backward edges can exist in this ordering since the graph has no cycles. There also clearly cannot be any other ordering as it would conflict with an edge of the directed path.

To prove the other direction, we'll proceed by contraposition. If a DAG does not have a directed path that visits all of its vertices, then by the previous part there exist two vertices A and B with no path between them. Then A and B can be interchanged and the resulting ordering will still be a valid topological ordering. Therefore, there exist at least two topological orderings so the topological ordering is not unique.

(c) This subpart is unrelated to the notion of semiconnectedness. Consider what would happen if we ran the topological sorting algorithm from class on a directed graph that had cycles.

Prove or disprove the following: The algorithm would output an ordering with the least number of edges pointing backwards.

Solution: The statement is false. Consider the possible DFS traversal on this graph yielding the following pre- and post-numbers:



The SCC-finding algorithm would output the ordering specified in the graph, which has 3 edges pointing backwards. However, ordering the vertices in the reverse order would yield an ordering with only 2 edges pointing backwards.

5 Distant Descendants

You are given a tree $T = (V, E)$ with a designated root node r and a positive integer K. For each vertex v, let $d[v]$ be the number of descendants of v that are a distance of at least K from v. In this problem, we will find an $O(|V|)$ algorithm to output $d[v]$ for every v.

(a) Write an $O(|V|)$ algorithm that computes the total size of the subtree (number of descendants plus 1 for the vertex itself) of each vertex v in an array $s[v]$. Give a brief justification that your algorithm is correct and runs in $O(|V|)$ time. Do not just cite an algorithm from class; reproduce anything you use in your solution.

Solution:

```
s[v] ← 1 for all vertices v.
visited[v] ← False for all vertices v.

function EXPLORE(Vertex u)
    visited[u] ← True
    for each edge (u, v) in E do
        if NOT visited[v] then
            EXPLORE(v)
            s[u] = s[u] + s[v]

EXPLORE(r)
```

To compute the size of the subtree of a vertex v, the DFS algorithm adds the sizes of the subtrees of each of its children plus the initial 1 for the vertex itself.

(b) Write an $O(|V|)$ algorithm that computes the K-th level ancestor of each vertex v (null if the depth of v is less than K) in an array $a[v]$. Give a brief justification that your algorithm is correct and runs in $O(|V|)$ time. Make sure your algorithm runs in $O(|V|)$ time and not $O(K|V|)$ time.

Solution:

```
a[v] ← 0 for all vertices v.
visited[v] ← False for all vertices v.
ancestors = []

function EXPLORE(Vertex u)
    visited[u] ← True
    if len(ancestors) > K then
        a[u] = ancestors[−(K + 1)]
    else
        a[u] = null

    for each edge (u, v) in E do
        if NOT visited[v] then
            ancestors.append(v)
```



```
        EXPLORE(v)
        ancestors.poplast()

EXPLORE(r)
```

When a vertex u is explored in the DFS, the list *ancestors* contains all the ancestors of u, ending in u itself. This allows us to find the K-th ancestor at $ancestors[-(K+1)]$ and assign it to $a[u]$.

(c) Write an $O(|V|)$ algorithm to compute $d[v]$ for each vertex v using $s[v]$ and $a[v]$. Give a brief justification that your algorithm is correct and runs in $O(|V|)$ time.

Solution:

```
d[v] ← 0 for all vertices v.
for each edge v in V do
    if a[v] then
        d[a[v]] = d[a[v]] + s[v]
```

For each vertex, we add the size of the subtree of each of the children at the K-th descendant level. Combined, this counts all the vertices that are at least depth K. Since there is a single for loop, this algorithm takes time $O(|V|)$.

6 [Coding] DFS & Edge Classification

For this week's homework, you'll implement implement DFS and use DFS to classify edges in a graph as forward/tree, backward, or cross edges. There are two ways that you can access the notebook and complete the problems:

1. **On Datahub**: click here and navigate to the `hw03` folder.
2. **On Local Machine**: `git clone` (or if you already cloned it, `git pull`) from the coding homework repo,

   https://github.com/Berkeley-CS170/cs170-sp25-coding

   and navigate to the `hw03` folder. Refer to the `README.md` for local setup instructions.

Notes:

- *Submission Instructions:* Please download your completed submission `.zip` file and submit it to the Gradescope assignment titled "Homework 3 Coding Portion".
- *Getting Help:* Conceptual questions are always welcome on Edstem and office hours; *note that support for debugging help during OH will be limited.* If you need debugging help first try asking on the public Edstem threads. To ensure others can help you, make sure to:

  1. Describe the steps you've taken to debug the issue prior to posting on Ed.
  2. Describe the specific error you're running into.
  3. Include a few small but nontrivial test cases, alongside both the output you expected to receive and your function's actual output.

  If staff tells you to make a private Ed post, make sure to include *all of the above items* plus your full function implementation. If you don't provide them, we will ask you to provide them.
- *Academic Honesty Guideline:* We realize that code for some of the algorithms we ask you to implement may be readily available online, but we strongly encourage you to not directly copy code from these sources. Instead, try to refer to the resources mentioned in the notebook and come up with code yourself. That being said, we **do acknowledge** that there may not be many different ways to code up particular algorithms and that your solution may be similar to other solutions available online.

Depth First Search and Edge Classification

If you're using Datahub:

- Run the cell below **and restart the kernel if needed**

If you're running locally:

You'll need to perform some extra setup.

First-time setup

- Install Anaconda following the instructions here: https://www.anaconda.com/products/distribution (https://www.anaconda.com/products/distribution)
- Create a conda environment: `conda create -n cs170 python=3.10`
- Activate the environment: `conda activate cs170`
  - See for more details on creating conda environments https://conda.io/projects/conda/en/latest/user-guide/tasks/manage-environments.html (https://conda.io/projects/conda/en/latest/user-guide/tasks/manage-environments.html)
- Install pip: `conda install pip`
- Install jupyter: `conda install jupyter`

Every time you want to work

- Make sure you've activated the conda environment: `conda activate cs170`
- Launch jupyter: `jupyter notebook` or `jupyter lab`
- Run the cell below **and restart the kernel if needed**

In [45]:
```
# Install dependencies
!pip install -r requirements.txt --quiet
```

In [46]:
```
import otter
assert (
    otter.__version__ >= "5.5.0"
), "Please reinstall the requirements and restart your kernel."
import networkx as nx
import typing
import numpy as np
import tqdm
import pickle
grader = otter.Notebook("dfs-edge-classification.ipynb")

rng_seed = 42
```

In [47]:
```
# Load test cases
file_path = "generated_testcases.pkl"

# Load the variables from the pickle file
with open(file_path, "rb") as file:
    loaded_data = pickle.load(file)
file.close()
inputs, outputs = loaded_data
```

Representing graphs in code

There are multiple ways to represent graphs in code. In class we covered adjacency matrices (https://people.eecs.berkeley.edu/~vazirani/algorithms/chap3.pdf#page=2) and adjacency lists (https://people.eecs.berkeley.edu/~vazirani/algorithms/chap3.pdf#page=3). There is also the edge list representation, in which you store the edges in a single 1 dimensional list. In general for CS170 and in most cases, we choose to use the adjacency list representation since it allows us to efficiently search over a node's neighbors.

In many programming problems, verticies are typically labelled 0 through $n-1$ for convenience (recall that arrays and lists in most languages begin at index 0). This allows us to represent an adjacency list using a list of lists that store ints. Given an edge list, the following code will create an adjacency list for an **unweighted directed graph**.

In [48]:
```
def generate_adj_list(n, edge_list):
    """ Generates an adjacency list given a set of edges.

    Args:
        n (int): Number of nodes in the graph. The nodes are labelled with intege
rs 0 through n-1
        edge_list (List[Tuple(int,int)]): Edge list where each tuple (u,v) repres
ents the directed edge (u,v) in the graph.

    Returns:
        List[List[int]]: The adjacency list.
    """
    adj_list = [[] for i in range(n)]
    for u, v in edge_list:
        adj_list[u].append(v)
    for nodes in adj_list:
        nodes.sort()
    return adj_list

def draw_graph(adj_list):
    """ Utility method for visualizing graphs

    Args:
        adj_list (List[List[int]]): Adjacency list of the graph given by generate
_adj_list.

    Returns:
        None
    """
    G = nx.DiGraph()
    for u in range(len(adj_list)):
        for v in adj_list[u]:
            G.add_edge(u, v)
    nx.draw(G, with_labels=True)
```

Q1: Reconstructing the DFS Path

In class we showed how to use DFS to check if there exists a path between two nodes, topologically sort nodes, and find SCCs. In those algorithms, pre and post numbers were used.

Here you'll implement a variation of DFS to print out the path between two nodes. In many problems, we want to be able to find the actual path between two nodes, not just determine if it exists.

> **Task 1:** Compute a path from s to t using DFS and return the path as a list of nodes on that path.

For example, the path $s \to a \to b \to c \to t$ corresponds to the list `[s, a, b, c, t]`. If no path exists, return the empty list `[]`.

You do not need to implement calculating pre and post numbers for this exercise.

Hint:

1. If you want to start with the recursive DFS implementation from DPV, you can use mutable types or the `nonlocal` keyword (https://cs61a.org/study-guide/mutation/#local-state) to preserve state across recursive calls.
2. It may be helpful to maintain an extra data structure which tracks the previous node we visited each time we visit a new node.

In [49]:
```
def dfs_path(adj_list, s, t):
    """ Finds a path from s to t using DFS or returns an empty list if no path ex
ists.

    Args:
        adj_list (List[List]): An adjacency list.
        s (int): An int representing the starting node.
        t (int): An int representing the destination node.

    Returns:
        List[int]: A list of nodes starting with s and ending with t representing
an s to t path if it exists.
                   Returns an empty list otherwise.
    """
    def explore(adj_list, curr):
        """
        Implements the explore subroutine from DPV, which is used in DFS. feel fr
ee to delete this
        function and use an alternative implementation if you prefer.

        Args:
            adj_list (List[List]): An adjacency list.
            curr (int): The node currently being traversed.

        Returns:
            None
        """
        # BEGIN SOLUTION
        nonlocal visited, prev # share the same visited and prev arrays across al
l calls to explore()

        visited[curr] = True
        for v in adj_list[curr]:
            if not visited[v]:
                prev[v] = curr
                explore(adj_list, v)
        # END SOLUTION

    # implement the dfs and path reconstruction here
    # BEGIN SOLUTION
    # initialize
    n = len(adj_list)
    visited = [False]*n # an array of booleans representing if a vertex has been
visited
    prev = [-1]*n       # an array of ints representing the previous node on a pa
th from start to the current node

    # unlike DPV algorithm, only need to start the dfs from s
    explore(adj_list, s)

    # if t was not visited, then there is no path from s to t
    if not visited[t]:
        return []

    # if path exists, backtrack through the prev array to find the s-t path
    path = []
    curr = t
    while curr != s:
        path.append(curr)
        curr = prev[curr]
    path.append(curr)
```

```
    path.reverse()
    return path
    # END SOLUTION
```

Debugging

You can create sample tests in the following cells to help debug your solution. We provide a few small tests as an example, but they might not be comprehensive.

To add a new graph to the test, append a new edge list to `edge_lists` as shown in the next cell.
Remember that these edges are directed, so do not add both directions of an edge to the edge list.

In [50]:
```
edge_lists = []
edge_lists.append([(0,1), (0,2), (1,2), (2,3), (3,4), (3,5), (4,5)])   # edge lis
t of first graph
edge_lists.append([(0,1), (0,2), (1,2), (3,4), (3,5), (4,5)])          # edge lis
t of second graph
# add any additional tests here
```

For each test case you also need to add a starting node s, a destination node t, and n the number of nodes in the graph, add them to the following lists.

In [51]:
```
s_list = []
s_list.append(0)  # s for first graph
s_list.append(1)  # s for second graph
# add any additional tests here

t_list = []
t_list.append(3)  # t for first graph
t_list.append(4)  # t for second graph
# add any additional tests here

n_list = []
n_list.append(6)  # n = 6 for first graph
n_list.append(6)  # n = 6 for second graph
# add any additional tests here
```

The following is a simplified version of the autograder, you may want to add more print statements or other debugging statements to check your function.

In [52]:
```
import matplotlib.pyplot as plt
index = 1
for s, t, n, edge_list in zip(s_list, t_list, n_list, edge_lists):
    print("Testing graph:", index)
    index += 1

    adj_list_graph = generate_adj_list(n, edge_list) # function defined earlier

    path = dfs_path(adj_list_graph, s, t)

    nx_graph = nx.DiGraph(edge_list)

    # uncomment the following to plot each graph
    '''
    nx.draw(nx_graph, with_labels=True)
    plt.title(f"Graph with {n} vertices and start node {s} and destination {t}")
    plt.show()
    '''


    if not nx.has_path(nx_graph,s,t):
        assert len(path) == 0, f"your dfs_path found an s-t path when there isn't
one."
    else:
        # checks that the path returned is a real path in the graph and that it s
tarts and ends
        # at the right vertices
        assert nx.is_simple_path(nx_graph, path), f"your dfs_path did not return
a valid simple path"
        assert path[0] == s, f"your dfs_path returned a valid simple path, but it
does not start at node s"
        assert path[-1] == t, f"your dfs_path returned a valid simple path, but i
t does not end at node t"

print("Success")
```

```
Testing graph: 1
Testing graph: 2
Success
```

In [ ]:
```
grader.check("q1")
```

Q2: Pre and Post Numbers

In order to topologically sort or find strongly connected components, we need to be able to calculate pre and post numbers for each node.

In this part, you will rework your implementation of DFS to allow it to generate pre and post order numbers for each node. It might be a good idea to copy/paste your solution from the previous part and modify it here.

> **Task 2:** Implement a function that computes DFS pre and post numbers for each node in the graph.

To pass the autograder, your smallest preorder number should be 1. Your largest postorder number should be $2 \times (\text{number of vertices})$. Return two lists of tuples, a `pre` list should containing tuples `(node, pre-number)`, and a `post` list containing tuples `(node, post-number)`.

Both lists should be ordered according to the pre/post number in the tuple. **You should not use any sorting functions to accomplish this!**

> **Reflect:** Why might returning pre/post numbers in this way be helpful for finding strongly connected components?

Feel free to delete the starter code and implement your own solution.

For this part, you can no longer assume that the entire graph is guaranteed to be reachable from some certain start node. How will this change your implementation?

Finally, break ties by choosing the node with the smallest number value. The autograder may fail implementations which are otherwise correct but break ties in a different way.

In [54]:
```
def get_pre_post(adj_list):
    """ Computes pre and post numbers for each node in the graph.

    Args:
        adj_list (List[List[int]]): The adjacency list that represents our input
graph.

    Returns:
        List[Tuple(int, int)], List[Tuple(int, int)]: The pre and post order valu
es respectively.
            Each tuple should have a vertex as its first entry, and the pre/post
order value as its second entry.
    """
    time = 1
    pre = []
    post = []

    # YOUR CODE HERE
    # BEGIN SOLUTION
    n = len(adj_list)
    visited = [False]*n

    def explore(u):
        nonlocal time
        nonlocal visited
        visited[u] = True
        pre.append((u, time))
        time += 1
        for v in adj_list[u]:
            if not visited[v]:
                explore(v)
        post.append((u, time))
        time += 1
    for i in range(n):
        if not visited[i]:
            explore(i)
    # END SOLUTION

    return pre, post
```

In [ ]:
```
grader.check("q2")
```

Q3: Identifying Tree, Forward, Back, Cross Edges

As we perform DFS traversals and create DFS trees and DFS forests within our graph, we would like to classify our edges according to how they appear in the resulting DFS forest. These classifications can provide us with insights about our graph. For example, the presence of a back edge (u, v) tells us that we have a cycle within this graph that includes all the tree edges on the path from v to u and the back edge (u, v).

> **Task 3:** Given the adjacency list of a graph, add each edge present in the edge set to the correct classification according the DFS traversal you implemented in part 1.

Don't modify the initialization of the edges_lookup dictionary.

In [56]:
```
def categorize_edges(adj_list):
    """ Categorizes all edges of the graph.

    Args:
        adj_list (List[List[int]]): The adjacency list that represents our input
graph.

    Returns:
        Dictionary({
            'tree': set(),
            'forward': set(),
            'cross': set(),
            'back': set()
        }) where each set() contains the edges that belong to the corresponding e
dge type
    """
    edges_lookup = {
        'tree': set(),
        'forward': set(),
        'cross': set(),
        'back': set()
    }
    # BEGIN SOLUTION
    # Generate prev array. This time, we do need to loop through all nodes
    visited = [False]*len(adj_list)
    prev = [-1]*len(adj_list)
    # explore subroutine from earlier
    def explore(adj_list, curr):
        nonlocal visited, prev # share the same visited and prev arrays across al
l calls to explore()

        visited[curr] = True
        for v in adj_list[curr]:
            if not visited[v]:
                prev[v] = curr
                explore(adj_list, v)
    for v in range(len(adj_list)):
        if not visited[v]:
            explore(adj_list, v)

    preorder, postorder = get_pre_post(adj_list)

    pre, post = {}, {}
    for u, time in preorder:
        pre[u] = time
    for u, time in postorder:
        post[u] = time

    for u in range(len(adj_list)):
        for v in adj_list[u]:
            edge = (u, v)
            if pre[u] < pre[v] < post[v] < post[u]:
                if prev[v] == u:
                    edges_lookup['tree'].add(edge)
                else:
                    edges_lookup['forward'].add(edge)
            elif pre[v] < pre[u] < post[u] < post[v]:
                edges_lookup['back'].add(edge)
            else:
                edges_lookup['cross'].add(edge)
```

```
    # END SOLUTION
    return edges_lookup
```

In [ ]:
```
grader.check("q3")
```

Submission

Make sure you have run all cells in your notebook in order before running the cell below, so that all images/graphs appear in the output. The cell below will generate a zip file for you to submit.

In [ ]:
```
grader.export(pdf=False, force_save=True, run_tests=True)
```