CS41 Homework 5

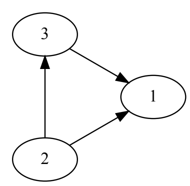
This homework is due at 11:59PM on Sunday, October 11. Write your solution using LaTeX. Submit this homework using **github** as a file called **hw4.tex**. This is a **partnered homework**. You should primarily be discussing problems with your homework partner.

It's ok to discuss approaches at a high level with others; however, you should **primarily discuss approaches with your homework partner.**

If you do discuss problems with others, you should not reveal specific details of a solution, nor should you show your written solution to anyone else. The only exception to this rule is work you've done with a lab partner *while in lab*. In this case, note (in your **homework submission poll**) who you've worked with and what parts were solved during lab.

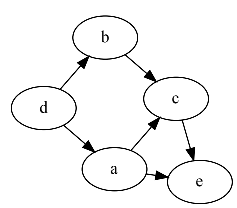
1. **Topological Sorting.** For each of the following directed graphs, determine whether or not the graph is *acyclic*. If the graph is cyclic, identify a cycle. If the graph is acyclic, give a topological ordering of vertices. To get you started, we gave the solution to part(a).

   (a) Directed graph G_1.

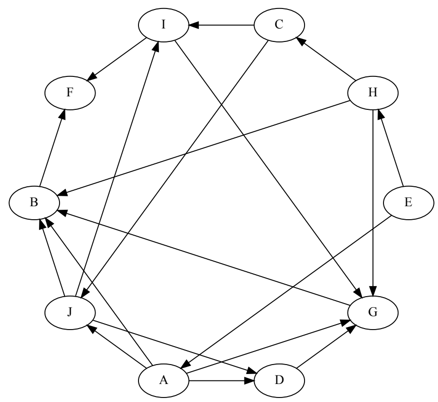


   Solution. G_1 is a directed acyclic graph. Here is a topological ordering: [2,3,1]

   (b) Directed graph G_2.



(c) Directed graph G_3.



2. **Butterfly Classification(K& T 3.4)** Some of your friends are *lepidopterists* — they study butterflies. Part of their recent work involves collecting butterfly specimens and identifying what species they belong to. Unfortunately, determining distinct species can be difficult because many species look very similar to one another.

   During their last field expedition, your friends collected n butterfly specimens and believe the specimens come from one of two butterfly species (call them species A and B.) They'd like to divide the n specimens into two groups—those that belong to A and those that belong to B. However, it is very hard for them to directly label any one specimen. Instead, they adopt the following approach:

   For each pair of specimens i and j, they study them carefully side by side. If they're confident enough in their judgement, they will label the pair as *same* (meaning they are confident that both specimens belong to the same species) of *different* (meaning they believe that the specimens belong to different species). If they are not confident, they do not label the specimens. Call this labeling (either (i, j) are the same or (i, j) are different) a *judgement.*

   A set of judgements is **consistent** if it is possible to label each specimen either A or B in such a away that for each pair (i, j) labeled "same", it is the case that i and j have the same label, and for each pair (i, j) labeled "different", it is the case that i and j have different labels.

   Design and analyze an algorithm which takes n butterfly specimens and m judgements, and outputs whether or not the judgements are consistent. Your algorithm should run in $O(n+m)$ time.

3. **Ethnographers.** (Kleinberg and Tardos, 3.12) You're helping a group of ethnographers analyze some oral history data they've collected by interviewing members of a village to learn about the lives of people who have lived there over the past two hundred years.

   From these interviews, they've learned about a set of n people (all now deceased), whom we'll denote $P_1, P_2, \ldots, P_n$. They've also collected facts about when these people lived relative to one another. Each fact has one of the following two forms:

- for some i and j, person P_i died before person P_j was born; or
- for some i and j, the lifespans of P_i and P_j overlapped at least partially.

Naturally, the ethnographers are not sure that all these facts are correct; memories are not very good, and a lot of this was passed down by word of mouth. So what they'd like you to determine is whether the data they've collected is at least *internally consistent*, in the sense that there could have existed a set of people for which all the facts they've learned simultaneously hold.

Give an efficient algorithm to do this: either it should propose dates of birth and death for each of the n people so that all the facts hold true, or it should report (correctly) that no such dates can exist—that is, the facts collected by the ethnographers are not internally consistent.

4. **(Extra Challenge).** Let $n \geq 2$ and let G be a directed acyclic graph on n vertices. Show that between any two nodes s, t in G there are at most 2^{n-2} simple $s \rightsquigarrow t$ paths. Give a graph where this is tight. i.e. identify an n-vertex graph $H = (V, E)$ and two vertices $s, t \in V$ such that there are exactly 2^{n-2} $s \rightsquigarrow t$ paths in H.

5. **(Extra Challenge).** Design and analze an $O(n + m)$ algorithm `SCC` for identifying all strongly conneted components in a directed graph. `SCC` should take a directed graph as input and return an integer array `COMP`, where `COMP[i] = k` if i is in the k-th strongly connected component.