CS41 Homework 4

This homework is due at 11:59pm, October 4, 2017. Write your solution using \LaTeX. Submit this homework using github as a file called hw4.tex. Make sure it compiles before you submit. Try to make your answers as clear and concise as possible.

For this homework, you will work with an assigned partner. It’s ok to discuss approaches with others at a high level, but most of your discussions should just be with your homework partner. The only exception to this rule is work you’ve done while in lab. In this case, note who you’ve worked with and what parts were solved during lab. If there are questions about academic integrity, please visit the section on Academic Integrity on the course website, or contact us.

0. Before final submission, make sure to fill out the README file.

1. Analysis. Let \( f(n) = 12n^{4/5} \) and \( g(n) = n^{3/5}(\log n)^6 \). Prove that \( g(n) = O(f(n)) \). You may use techniques and facts from class and the textbook; your proof should be formal and complete.

2. Social Distance In analyzing social networks, sociologists are often interested in measuring the social distance between individuals or groups. The social distance between two people \( d(v_1, v_2) \) can be defined as the number of edges along the shortest path between them. However, there are several ways to measure the social distance between groups. For each of the following measures of group social distance, design and analyze a polynomial time algorithm that takes an undirected social network graph \( G = (V, E) \), along with two (possibly overlapping) groups \( V_1 \subset V \) and \( V_2 \subset V \), and returns the group distance \( D(V_1, V_2) \).

(a) Min distance: \( D(V_1, V_2) = \min_{v_1 \in V_1, v_2 \in V_2} d(v_1, v_2) \)

(b) Max distance: \( D(V_1, V_2) = \max_{v_1 \in V_1, v_2 \in V_2} d(v_1, v_2) \)

(c) Average distance: \( D(V_1, V_2) = \frac{1}{|V_1||V_2|} \sum_{v_1 \in V_1} \sum_{v_2 \in V_2} d(v_1, v_2) \)

3. Bus Transfers. In the midst of a long backpacking trip, you’ve arrived in a new city that you’re visiting for only one day. Starting from the train station at the southeast corner of the city, you need to reach your hostel at the city’s northern boundary by the end of the day. Along the way, you want to visit as many tourist attractions as you can, but on the cheap. Luckily, the city has lots of good bus routes, and lets you transfer to another bus for free as long as it’s going in the same direction.

We can model this problem using a directed acyclic graph, where the nodes represent tourist attractions, and the edges correspond to northbound bus routes. Design and analyze a polynomial time algorithm to determine how many sites you can visit. Your algorithm should take a DAG \( G \), a start vertex \( s \), and a destination vertex \( t \), and should return the maximum number of nodes that can be visited along a path \( s \rightarrow t \).

4. 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.

5. **Extra credit.** In class/lab we saw an algorithm for testing bipartiteness which used BFS to color the vertices. It should be possible to use DFS to test bipartiteness to color the vertices. Give an algorithm (in pseudocode) which uses DFS to test bipartiteness. Rigorously prove that your algorithm works.