Assignment 5
Minimum Spanning Trees, NP-Complete Problems, and Approximation Algorithms

Paul Jones and Matthew Klein
Professor Kostas Bekris
Design and Analysis of Computer Algorithms (01.198.344)

May 3, 2014

Problem 1

You are a programmer in an electronic commerce company, where you are asked to design a new recommendation system. The idea is to divide all users into k groups based on their purchase history. After the grouping, your system can provide recommendations to users based on the purchases of other users in the same group.

To perform the grouping, we define a function $s(u_i, u_j)$ to reflect the similarity of two users u_i and u_j. The similarity function is defined as:

$$s(u_i, u_j) = \begin{cases} 1 + \text{the number of common purchases } u_i \text{ and } u_j \text{ have made} & (i \neq j) \\ 0 & (i = j) \end{cases}$$

To optimize the grouping, we need to minimize the similarity of the most similar pair of users that have been assigned to different groups.

(a) Provide an efficient algorithm that groups all users into k groups and satisfies the optimality requirement mentioned above (i.e., minimizes the similarity of users in different groups).

(b) What data structures do you assume for the implementation of your algorithm? What is the running time of your approach given the data structures you employ?

(c) Prove the correctness of your algorithm.

Problem 2 (20 points):

In wireless sensor networks, it is an important task to periodically collect data from an area of interest for time-sensitive applications. The sensed data must be gathered and transmitted from the sensors to a base station through the most efficient way in terms of energy and speed.

The sensor nodes are deployed randomly in the target field. Then, the network establishment begins with the formation of communication links between the sensor nodes. Consider that $E(a, b)$ is the energy needed to send message from node a to node b, which typically depends on the distance between the two nodes.

An idea in the sensor network literature is to use a minimum spanning tree (MST) as the underlying communication structure in order to shorten the total transmission distance, while guaranteeing that there is a way for a message to reach every node in the network. This means that we need to construct a MST of the nodes in the network.

It is possible, however, that a node in the resulting MST will be connected with a lot of neighboring nodes, e.g., consider the case the MST looks like a star-like structure where a central node is connected to all other nodes. Such nodes with many edges may need to fuse more data collected from its neighbors than other nodes and to consume more energy. This may cause nodes with a lot of neighbors to die earlier because they exhaust their energy. In order to avoid this situation, we limit the number of connections for each node to be at most d in the communication tree, for a given positive integer d.

A. For $d = 2$, describe an algorithm for finding such a communication tree and show that this problem is NP-complete problem. (Hint: consider the relation with finding a RUDRATA-PATH on a graph.)

B. Describe a polynomial time algorithm (in detail) that solves this problem, if you have an algorithm that solves TSP in a polynomial time.

Problem 3 (30 points):

Lord Tywin Lannister is organizing the feast for the upcoming wedding of his grandson, King Tommen Baratheon, with Lady Margaery Tyrell. The Queen of Thornes, Olenna Redwyne, and grandmother of Margaery, has purchased two exquisite tables from Bravos for the event for the guests to sit and attend the feast. These tables are big enough to accommodate all n guests but she bought two of them because it is better if some of the guests are separated. Weddings can lead to quite unfortunate events in the land of Westeros...

In particular, the list of n attendees includes members of the Tyrell family of Highgarden, members of the Martell family of Dorne as well as members of the Lannister family of Casterly Rock. Unfortunately, not all attendees have good relationships one with another. For instance, Oberyn Martell accuses Gregor Clegane from the Lannisters for the death of his sister and her children. So, it would not be wise to allow both Oberyn and Gregor to sit on the same table, if possible.

Lord Tywin and the Queen of Thornes have drafted a list of all m pairs of attendees who do not have good relationships and should better be placed on different tables. Their objective is to find an assignment of attendees to the two tables such that the total number of conflicts between any two guests from different tables is maximized.

Based on the above description:

1. **Formulate the problem that Lord Tywin and the Queen of Thornes are facing as a graph problem and define the cost function clearly.**
2. **Propose a local search algorithm that is a 2-approximation polynomial time solution. Show that the algorithm provides a 2-approximation solution.**
3. **Propose a greedy algorithm that is a 2-approximation polynomial time solution. Show that the algorithm provides a 2-approximation solution.**
4. **Propose a randomized algorithm that is a 2-approximation polynomial time solution. Show that the algorithm provides a 2-approximation solution.**

[Note: Though the problem is similar to the problem 4 in homework 3, the assignment here does not require that there is no conflict between guests placed on the same table. You may have a situation where it is impossible to assign the guests so that no conflict arises on a table. For instance, both Oberyn Martell and Willas Tyrell hate Gregor Clegane, but they are also not very fond of one another... Instead, you are required to find an assignment that maximizes the number of conflicts between guests sitting in different tables.]

Problem 4 (40 points):

Daenerys Targaryen is moving her army of elite warriors, the Unsullied, in the land of Essos trying to liberate as many cities as possible from the slave-owners that rule them. The major cities of Essos, such as Yunkai, Meereen and Astapor, are trying to coordinate their defenses.

The strategy suggested by Daenerys' advisor Ser Barristan Selmy is to try to isolate the major cities one from another so that they cannot communicate and coordinate against Daenerys. Ser Jorah Mormont generated for this

purpose a map of Essos that provides the road network connecting the biggest n cities on the continent.

Daenerys agreed with Ser Barristan Selmy and decided to place teams of 100 Unsullied along different road segments so that it would not be possible for a messenger from one of the major cities to reach another through the road network without being detected and... facing the consequences... Daenerys is worried, however, of spending too many of her Unsullied for this purpose and reducing her main military force. So, she wants to identify the minimum number of road segments that her forces will have to guard so as to achieve this objective.

Without knowing how the road network with the n cities and the roads connecting them looks like:

a. **Provide a polynomial time algorithm that solves Daenerys' problem exactly, when the objective is to separate only Yunkai and Astapor.**

b. **Provide a solution that achieves an approximation ratio of at most 2 when Daenerys wants to separate all three of Yunkai, Meereen and Astapor.**

c. **Propose a local search approach for the general case where Daenerys wants to make sure that at least m cities remain isolated one from another.**

d. **The cities of Essos are working on a counter-strategy after they heard the news regarding Daenerys' strategy from spies.**

They want to place soldiers along the minimum set of roads that need to be protected so that all the cities can still communicate one with another. When a soldier detects the 100 Unsullied along a road segment, they will send a large enough force to take them out. To be able to monitor a road, it is necessary to place a soldier along every mile of the corresponding road. The cities need to minimize the number of soldiers used for this purpose so that they are sufficiently defended in case Daenerys attacks them directly. Consequently, the cost of protecting a longer road is directly related to its length [Hint: Thus, the cost of selecting a road segment to monitor is a metric function and satisfies the triangular inequality.].

Assume that the road network between the n cities of Essos is a complete graph. Consider the problem of identifying the minimum set of roads that need to be monitored so that at least the m major cities will remain connected given the above strategy. Provide a polynomial time approximation algorithm for this problem that has an approximation ratio of 2.