DA 507 – modeling and optimization
Fall 2015

Question 1. A supermarket chain is planning to open its first retail outlet stores in a small town. The town is districted into 8 regions as A, B, C, D, E, F, G and H; a central location in each region is considered as a candidate facility location for the stores. While locating a competitive retail outlet store, accessibility within a critical distance is crucial; it is assumed that a district is directly serviced if it is within a distance of ***D*** from the region where the store is located.

(a) Write down the linear integer programming formulation to find the optimal location of ***P*** facilities where the objective function maximizes the total population directly serviced by the store(s). Suppose that d_{ij} denotes the distance from region *i* to region *j* and w_i denotes the population of region.

(b) The following table shows the distances between regions and the population of each region.

d_{ij}	A	B	C	D	E	F	G	H	w_i
A	0	28	55	42	57	37	31	34	10900
B	28	0	67	26	29	48	57	22	12000
C	55	67	0	61	96	19	60	52	18000
D	42	26	61	0	35	43	70	12	9500
E	57	29	96	35	0	77	85	44	6000
F	37	48	19	43	77	0	52	34	12700
G	31	57	60	70	85	52	0	59	13000
H	34	22	52	12	44	34	59	0	4000

The managers are concerned about the trade-off between the accessible population and the critical distance; a reasonable critical distance ranges from 30 to 60. They want to also investigate the options of opening either only one facility or two facilities. How would you help them understand this trade-off for both options? How do the facility locations change when the distance is modified between 30 and 60? How do the facility locations change when the number of facilities is increased from 1 to 2?

(c) An analysis has shown that locating an additional facility is worth the corresponding investment only if an additional population of 20,000 is directly serviced? Is there a range of reasonable critical distance where locating the second facility is worth the investment?

Question 2. A group of plant biologists has identified 12 new species in a restricted geographical area. Based on a DNA sequence study, they have calculated a dissimilarity index (out of 100) for each pair. The resulting dissimilarity matrix is given below. They would like to understand how these species are related with each other. Are they all closely related to each other? Or, are they related to each other in group(s)? In order to answer these questions, they aim to cluster/organize them in groups for which

- dissimilarity within the group is minimal, and

- dissimilarity between the groups is maximal.

To achieve this goal, they use the following procedure:

Step 1. Given a threshold value (D) for dissimilarity index, construct a dissimilarity network where there is an arc between any two species with a dissimilarity index ***less than or equal to D***. Identify each connected component and declare them as groups.

Step 2. Within each group construct a minimum spanning tree of dissimilarity so that the group's similarity score can be calculated as the ratio of number of species in the group to the total dissimilarity index of the corresponding spanning tree.

As a result of this study, they will also identify a viable threshold value D.

Implement the above classification algorithm using a generic programming language (such as Python) for different values of D, identify each class and calculate the group similarity index for at least three different threshold values where the elements of the clusters and/or clusters' similarity scores change.

	1	2	3	4	5	6	7	8	9	10	11	12
1		50	21	41	44	18	91	12	90	77	89	31
2	50		81	88	36	70	19	49	67	100	35	74
3	21	81		40	98	79	65	84	60	15	37	60
4	41	88	40		18	49	31	96	78	93	41	38
5	44	36	98	18		36	77	94	70	52	49	31
6	18	70	79	49	36		87	40	15	90	0	46
7	91	19	65	31	77	87		40	31	34	22	60
8	12	49	84	96	94	40	40		15	25	3	87
9	90	67	60	78	70	15	31	15		60	17	100
10	77	100	15	93	52	90	34	25	60		21	83
11	89	35	37	41	49	0	22	3	17	21		15
12	31	74	60	38	31	46	60	87	100	83	15	