Set Covering Problem

Problem

The Ministry of Health wants to build some orthopedic-specialized hospitals, able to serve the following cities within the range of 200 kms: Latina, Lecce, Matera, Napoli, Potenza, Salerno e Roma. Here follows, for every city, the list other cities far less then 200 kms:

- Latina: Latina, Napoli, Roma;
- Lecce: Lecce, Matera;
- Matera: Lecce, Matera, Potenza;
- Napoli: Latina, Napoli, Potenza, Salerno;
- Potenza: Matera, Napoli, Potenza, Salerno;
- Salerno: Napoli, Potenza, Salerno;
- Roma: Latina, Roma.

For example, if an hospital is built in Napoli, it would be able to serve the cities of Latina, Potenza and Salerno too, that are less than 200 kms far by Napoli.
We want to decide in which of the seven cities build the hospitals, in order that every city is served at least by one hospital, far less than 200 kms, and considering that two hospitals cannot be built in the same city.

Analysis

This is a classical problem of Operations Research but not only, also of Combinatorics and Computer Science.
Given a certain number of regions (or places), we must choose where to build or install a set of elements (or buildings, like in this case). The information about the installation cost can be known (and the problem will be named *weighted*). Let $M = \{1, \dots, m\}$ be the set of regions, while $B = \{1, \dots, b\}$ is the set of potential buildings. Let $S_j \subseteq M$ be the set of regions served by the building j, $j \in B$, and c_j its cost of installation.

The problem is to determine the collection of subsets T such that:

$$\min_{T \subseteq B} \left\{ \sum_{j \in T} c_j : \cup_{j \in T} S_j = M \right\}.$$

The solution to the Set Covering Problem is the smallest sub-collection of S, whose union equals the universe M.

In this case, our universe M is composed of all the cities, whereas the information about the set N corresponds to the lists of cities covered by the hospitals built. Since in this example the number of cities (cardinality of M) is equal to the number of potential buildings, we can rearrange the information about cities and buildings using a a matrix N x N, where N is the number of cities or places and where we assign an index to each city (that is simply its code):

City	**LT**	**LE**	**MT**	**NA**	**PT**	**SA**	**RM**
LT	1	0	0	1	0	0	1
LE	0	1	1	0	0	0	0
MT	0	1	1	0	1	0	0
NA	1	0	0	1	1	1	0
PT	0	0	1	1	1	1	0
SA	0	0	0	1	1	1	0
RM	1	0	0	0	0	0	1

Formulation

Sets: we can define M as the set of cities and B the set of potential buildings.

Parameters: we know data about which cities will be covered if a building is built; this information can be represented, as we did above, with an incidence matrix A made just by 1 and 0 values (in this instance, the matrix is squared because it is possible to build an hospital in every city).

Variables: we introduce a boolean variable for every building, whose value will be 1 if the hospital i is built, 0 otherwise: x_i, for every i in B.

Constraints:

- There must be at most one hospital in every city: $x_i \leq 1$, for every i in B; since our variables are boolean (0 or 1), this constraint is automatically satisfied.
- Every city must be covered by at least one hospital: $Ax \geq 1$, where A is the incidence matrix of cities and buildings and x is the column-vector composed of the buildings in B.

  For example, from the first line of the incidence matrix we know that

  $$x_{LT} + x_{NA} + x_{RM} \geq 1.$$

Objective function: since we do not have information about the cost of building a hospital in a specific city, we just want to minimize the number of hospitals to build: $\min \sum_{i \in B} x_i$.

The complete formulation is the following:

$$\begin{array}{lrcll} \min \sum\limits_{i \in B} & x_i & & & \\ \text{s.t.} & Ax & \geq & 1 & \\ & x_i & \in & \{0,1\} & \forall i \in B \end{array}$$

Weighted Set Covering Problem

If every building had its related cost of installation, a more general version of the problem could be modeled, where the purpose was not to minimize the number of buildings to build but the total installation costs:

$$\min \sum_{i \in B} c_i x_i$$

where c_i is the cost of building i, for every i in B. The rest of the formulation does not need to be changed, because the cost only impacts the objective function. The first model can be seen as a particular version of the Weighted Set Covering Problem where the cost coefficient of every building is 1.