IEOR 151 – Lecture 15
Set Covering Problem

1 Problem Setup

The set covering problem is a specific type of a discrete location model. In this model, a facility can serve all demand nodes that are within a given coverage distance D^c from the facility. The problem is the place the minimum number of facilities so as to ensure that all demand nodes can be served. In this model, there are no capacity constraints at the facilities.

1.1 Mathematical Model

We represent the problem formulation using an undirected graph $G = (I, J, E)$, where the demand nodes are represented by a set of vertices $i \in I$, the possible locations of the facilities are given by another set of vertices $j \in J$, and edges $e_{i,j} \in E$ only exist between vertices from $i \in I$ to those in $j \in J$. Furthermore, we assign positive weights to the edges $d_{i,j} \geq 0$, which represents a distance between vertices i and j. In contrast to the p-median problem, we do *not* assign demand weights to each demand node. In contrast to the vertex p-center problem, the number of facilities to place is not fixed. However, we do have a new value: The coverage distance D^c is the maximum distance between a facility and a demand node that it can serve. Furthermore, in this formulation we do not need to explicitly use the distances. Instead, we make use of an indicator variable

$$a_{i,j} = \begin{cases} 1, \text{if } d_{i,j} \leq D^c \\ 0, \text{otherwise} \end{cases}$$

that describes if demand node at $i \in I$ can be covered by a facility at $j \in J$ if the coverage distance is D^c.

1.2 ILP Formulation

To formulate an integer linear program (ILP) to solve this problem, we define a decision variable

$$X_j = \begin{cases} 1, \text{if facility located at } j \in J \\ 0, \text{otherwise} \end{cases}$$

that describes the locations at which a facility is placed. Given these decision variables, we can now formulate the set covering problem as the following ILP

$$
\begin{aligned}
\min \sum_{j \in J} & X_j \\
\text{s.t.} \sum_{j \in J} & a_{i,j} X_j \geq 1, \forall i \in I \\
& X_j \in \{0,1\}, \forall j \in J.
\end{aligned}
$$

Each of these terms has associated intuition. The objective $\sum_{j \in J} X_j$ is stating that we wish to minimize the number of placed facilities. The first constraint $\sum_{j \in J} a_{i,j} X_j \geq 1$ implies that every demand node i needs to be served by at least facility. In this model, a demand node can be served by more than one facility because there is no assignment of a demand node to a facility (c.f., the variable $Y_{i,j}$ in the p-median and vertex p-center problem). The constraints $X_j \in \{0,1\}$ force the decision variables to be binary.

1.3 Inefficient Algorithm

The ILP can be difficult to solve, and so we discuss an algorithm to solve this problem. The idea of this algorithm is to use an oracle (i.e., a black box) that solves the vertex p-center problem. Let #I be the number of demand nodes, and assume that there is a feasible solution. Then the algorithm is given by

1. For $p = 1, \ldots, \#I$
   - (a) Solve the vertex p-center problem;
   - (b) If the optimal Q^* is such that $Q^* \leq D^c$, then break the loop;

The solution to the problem is given by the solution to the final p-center problem that was solved.