1 Basic problem

Given a matrix $A \in \mathbb{R}^{n\times n}_{\geq 0}$, representing a *target* molecule and a *database* $B^{(i)} \in \mathbb{R}^{n_i \times n_i}$, $i \in \{1,\ldots,k\}$. The n_i are usually small, around 7.

A *solution* is a permutation $\pi \in S_n$, a multiset $S \subseteq \{1,\ldots,k\}$ and a mapping $p : S \to \{1,\ldots,n\}$. The *objective value* of the solution is $\|A_\pi - A_{S,p}\|_2^2$, where A_π is the matrix stemming from A where the rows and columns were permuted by π and A_{S_p} is the block-diagonal matrix that stems from the elements in S that are put with upper leftmost component on the corresponding position described by p.

A feasibility constraint is that S and p induce a partition of the set $\{1,\ldots,n\}$.

Overarching goal for Chemistry: Suggest new ways to synthesize a target molecule.

- Bigger database of pieces should make sense. Efficiency is a problem can be overcome. Also the pieces need a score (objective value) on the cost to synthesize them. This can be modeled in the objective function.
- Diversity of solutions: One way would be to list pareto optimal solutions using distance, sums of costs of pieces, ... Another way would be to model diversity
- Objective function should also model the number of smaller pieces. The more pieces in the solution, the more synthesis steps.

2 Optimal placement of molecules

The first goal is to find a way to optimally place a small molecule represented by the matrix $M \in \mathbb{R}^{m\times m}$ inside the target represented by the matrix $T \in \mathbb{R}^{n\times n}$. In other words, we want to find the mapping $\phi : \{1,\ldots,m\} \to \{1,\ldots,n\}$. This is an index assignment problem where the indices $x_{ij} \in \{0,1\}$ are assigned:

$$x_{ij} = \begin{cases} 1 & \text{if } \phi(i) = j \\ 0 & \text{otherwise} \end{cases}$$

under the constraints:

$$\forall i \in \{1,...,m\} : \sum_{j=1}^{n} x_{ij} = 1$$
$$\forall j \in \{1,...,n\} : \sum_{i=1}^{m} x_{ij} \leq 1$$

This consists of minimising the objective function:

$$\min \sum (m_{ij} - T_{kl})^2 x_{ik} * x_{jl}$$