QUBO Bus routes

Febrero 2025

1 Introduction

1.1 Problem description

Find the best routing strategy for several lines of urban transport vehicles, considering the following statements:

- Distances are not symmetric (in general, $d(A, B) \neq d(B, A)$).
- Bus stops are set based on social and demographic factors.
- The number of bus stops per line, as well as the number of lines are hyperparameters of the problem.

We want to set L bus lines in N stops. As the distances between stops are not symmetric, we can define the binary variables as

$$x_{ij}^{l} \in \{0, 1\} \tag{1}$$

Where the value is 1 only if the line l goes at some point from stop i to stop j. The global function that we want to minimize is

$$f(x_{00}^{0}, x_{01}^{0}, ..., x_{0N-1}^{0}, x_{10}^{0}, ..., x_{N-1N-1}^{0}, x_{00}^{1}, ..., x_{N-1N-1}^{L-1}) = \sum_{l}^{L-1} \sum_{i}^{N-1} \sum_{j}^{N-1} D_{ij} x_{ij}^{l} \tag{2}$$

where the matrix of distances D_{ij} is not symmetric and of order $N \times N$.

1.2 QUBO formulation

A Quadratic unconstrained binary optimization (QUBO) is a optimization problem which is defined for a set of binary variables $z = (z_1, z_2, ..., z_n) \in \{0, 1\}^n$ which minimizes the quantity

$$f(z) = z^T Q z \tag{3}$$

for Q a symmetric (or triangular superior) matrix of size $(n \times n)$. We need to reformulate our problem to fit into this definition. Some simplifications may be necessary in order to do so.

2 First Approach: Just one line in symmetric distances.

To start with a simple approach we can consider that the distances are in fact symmetric. To obtain them we can calculate the following new cost matrix, which have 0 in its diagonal and in all terms under the diagonal.

$$D_{ij}^{'} = D_{ij} + D_{ji} \quad \text{for } i < j, \quad 0 \quad \text{otherwise} \tag{4}$$

The variables in this simplification will x_{ij} where they represents that the bus line pass from i to j or viceversa. Therefore we can get rid of all the terms x_{ij} with $j \leq i$.

QUBO cost function In order to have a vector of binary variables we will define the following new variable y with $R = N(N-1)/2$ components as

$$y = \begin{pmatrix} x_{01} & x_{02} & ... & x_{0N-1} & x_{1,2} & ... & ... & x_{N-2N-1} \end{pmatrix} \tag{5}$$

The translation between subindices is

$$\{i,j\} \rightarrow k = S_i + (j-i-1) = \sum_{m=0}^{i-1}(N-1-m) + (j-i-1) \tag{6}$$

$$k \rightarrow \{i,j\} \quad \text{such that} \quad S_i \leq k < S_{i+1} \quad \text{and} \quad j = i+1+k-S_i \tag{7}$$

With a similar indexing, the distances matrix should be placed into de diagonal of a bigger matrix, Q, of shape $N(N-1)/2 \times N(N-1)/2$. Then, the cost function can be written as

$$f_{\text{cost}}(y) = y^T Q y = \sum_{k=0}^{R-1}\sum_{l=k+1}^{R-1} Q_{kl} y_k y_l = \sum_{k=0}^{R-1} Q_{kk} y_k^2 = \sum_{k=0}^{R-1} Q_{kk} y_k \tag{8}$$

Constaint 1: Length of routes Once we have the cost function, we must impose some constraints to find a suitable solution. The first one will be to impose that the route should have a certain lenght $p \in \{1, ..., N-1\}$ (note that the lenght measures the edges, not the number of stops). To achieve that, we should add a quadratic penalty term of the shape $\lambda_1 H_1$ to the cost function. The restriction is

$$\sum_{k=0}^{R-1} y_k = p \tag{9}$$

Therefore, the term in the Hamiltonian reads

$$\lambda_1 H_1 = \lambda_1 \Big(\sum_{k=0}^{R-1} y_k - p \Big)^2 = \lambda_1 \Big(\sum_{k=0}^{R-1} y_k^2 + 2\sum_{k \neq l} y_k y_l - 2p\sum_{k=0}^{R-1} y_k + p^2 \Big) \tag{10}$$

WARNING: In the coding of this condition, in order to mantain the simmetry of the matrix but obtain better solutions, we divide the term $2\sum_{k\neq l} y_k y_l$ by 2.

Constraint 2: If a node is visited, it has two edges. If not, it should have no edges The second constraint is more difficult to translate to the new indexing, so we will work first in the previous one. The condition reads

$$d_i = \sum_{j<i} y_{\{j,i\}} + \sum_{k=i+1}^{N-1} y_{\{i,k\}} \in \{0,2\} \quad \forall i \in \{1, ..., N-2\} \tag{11}$$

Where $y_{\{i,j\}}$ symbolizes the two indices i,j which lead to k. We have no consider the first and last nodes. We have to construct a quadratic term which has a penalty when the sum is not 0 or 2. We have

$$\lambda_2 H_2 = \lambda_2 \sum_{i=1}^{N-2}(d_i(d_i-2)) = \lambda_2 \sum_{i=1}^{N-2} \Big(\sum_{j<i} y_{\{j,i\}} + \sum_{k=i+1}^{N-1} y_{\{i,k\}} \Big)\Big(\sum_{l<i} y_{\{l,i\}} + \sum_{r=i+1}^{N-1} y_{\{i,r\}} - 2 \Big) \tag{12}$$

Constraint 3: The bus line should start in node 0 This condition serves as a starting point of the route. Therefore, the condition is

$$\sum_{j=1}^{N-1} y_{\{0,j\}} = 1 \tag{13}$$

The penalti term is then

$$\begin{aligned}\lambda_3 H_3 &= \lambda_3(\sum_{j=1}^{N-1} y_{\{0,j\}} - 1)^2 = \lambda_3\Big(\sum_{j=1}^{N-1} y^2_{\{0,j\}} + 2\sum_{j\neq k} y_{\{0,j\}}y_{\{0,l\}} - 2\sum_{j=1}^{N-1} y_{\{0,j\}} + 1\Big) = \\ &= \lambda_3\Big(2\sum_{j\neq k} y_{\{0,j\}}y_{\{0,l\}} - \sum_{j=1}^{N-1} y_{\{0,j\}} + 1\Big)\end{aligned} \tag{14}$$

Constraint 4: The bus line should end in node N-1 Dual to the previous one. The condition reads

$$\sum_{i=0}^{N} y_{\{i,N-2\}} = 1 \tag{15}$$

Then, the Hamiltonian is

$$\begin{aligned}\lambda_4 H_4 &= \lambda_4(\sum_{i=0}^{N-2} y_{\{i,N-1\}} - 1)^2 = \lambda_4\Big(\sum_{i=0}^{N-2} y^2_{\{i,N-1\}} + 2\sum_{i\neq l} y_{\{i,N-1\}}y_{\{l,N-1\}} - 2\sum_{i}^{N-2} y_{\{i,N-1\}} + 1\Big) = \\ &= \lambda_4\Big(2\sum_{i\neq l} y_{\{i,N-1\}}y_{\{l,N-1\}} - \sum_{i}^{N-2} y_{\{i,N-1\}} + 1\Big)\end{aligned} \tag{16}$$

2.1 Conclusions

With those constraints, adjusting the parameters λ_i and p we should be able to obtain bus routes which go from stop 0 to stop $N-1$ with lenght p. This is so because we have two open paths (one starting from stop 0 and another ending in stop $N-1$) and every node should have two edges except from those starting and ending points. The only way of fulfill all the constraints is to join the two paths, creating the desired route. However, as the tuning of the parameters is being made manually, it is common that one of the constraints is not fulfilled.

Another issue is how a new bus route can be added to the stops graph. If the election of parameters λ_i, p is too sensible, the only way is to relabel the stops and selects two new stops $(0, N-1)$.