1 Problem Setting

In this project, we want to investigate how to develop marco network to connect a nation. More specifically, given some OD flow data, we want to solve for the optimal network which satisfies all demand while minimize construction cost. Next, we will discuss several problem formulations, from the easiest to the hardest, along with algorithms to solve them and the results.

1.1 Static Formulation

First, let us tackle the simplest question: if the OD demand are invariant across time, how can we design optimal network?

Assume we are given OD pairs $i \in \{1, \cdots, N\}$, along with distance $l_i \geqslant 0$ and demand $d_i \geqslant 0$ (notice that pairs (x, y) and (y, x) are encoded in different indices). We want to find a network construction plan $y_i \in \{0, 1\}$ with capacity $c_i \geqslant 0$ that can satisfy all demand d_j with allocation plan s_{ij}. The full formulation is shown below:

$$
\begin{aligned}
\min \sum_i c_i l_i + C \sum_i y_i l_i \\
s.t. \sum_j s_{ij} \leqslant c_i,\ \forall i \\
\sum_i s_{ij} \geqslant d_j,\ \forall j \\
M y_i \geqslant s_{ij},\ \forall i, j \\
s_{ij} \geqslant 0, c_i \geqslant 0, y_i \in \{0, 1\}
\end{aligned}
\tag{1}
$$

where C is the fixed cost for a new construction, and M is the big-M coefficient to ensure that the problem is feasible.

The problem of this formulation is that, when the number of possible OD pairs grows, the size of the problem increase significantly and become hard to solve at once. In our dataset, there are $\sim 30,000$ OD with non-trivial demand, so the above formulation has $\sim 10^9$ variables. Thus, we need to develop a more practical formulation.

1.2 Simple Dynamic Formulation

Next, we can extend the above framework into a more dynamic setting. Again, assume we are given OD pairs $i \in \{1, \cdots, N\}$, along with distance $l_i \geqslant 0$; but now we have time-varying demand $d_i^t \geqslant 0$. We want to find a network construction plan $y_i \in \{0, 1\}$ with capacity $c_i \geqslant 0$ that can satisfy all demand d_j^t with allocation plan s_{ij}^t. The full formulation is shown below:

$$
\begin{aligned}
\min \sum_i c_i l_i + C \sum_i y_i l_i \\
s.t. \sum_j s_{ij}^t \leqslant c_i,\ \forall t, i \\
\sum_i s_{ij}^t \geqslant d_j^t,\ \forall t, j \\
M y_i \geqslant s_{ij}^t,\ \forall i, j, t \\
s_{ij}^t \geqslant 0, c_i \geqslant 0, y_i \in \{0, 1\}
\end{aligned}
\tag{2}
$$

where C is the fixed cost for a new construction, and M is the big-M coefficient to ensure that the problem is feasible.

Again, this direct formulation is huge in size; by applying similar approximation approach above, we have