Appendix A: Implementation details of Benders decomposition

In this Appendix, we introduce the process of Benders decomposition for CMApHMPC only. CMAHLPC can be solved with similar process with minor modification. In the CMApHMPC, there is a constraint on the node capacity for each hub, i.e., the total flow through each hub cannot exceed its capacity. The formulation for CMApHMPC with $O(n^3)$ variables is shown in Section 2.4. In addition to this formulation, Ebery et al. (2000) also proposed a formulation with $O(n^4)$ variables as follows:

$$\min \sum_{i\in V}\sum_{j\in V}\sum_{k\in V}\sum_{m\in V}\left(\delta_1 c_{ik}+\alpha c_{km}+\delta_2 c_{mj}\right) w_{ij} X_{ijkm} \tag{43}$$

$$\text{subject to } \sum_{k\in V} Y_k = p \tag{44}$$

$$\sum_{k\in V}\sum_{m\in V} X_{ijkm} = 1, \forall i,j\in V \tag{45}$$

$$\sum_{m\in V} X_{ijkm} + \sum_{m\in V, m\neq k} X_{ijmk} \le Y_k, \forall i,j,k\in V \tag{46}$$

$$\sum_{i\in V}\sum_{j\in V} w_{ij} \sum_{m\in V} X_{ijkm} \le \lambda_k Y_k, \forall k\in V \tag{47}$$

$$Y_k \in \{0,1\}, \forall k\in V \tag{48}$$

$$X_{ijkm}\in[0,1], \forall i,j,k,m\in V \tag{49}$$

Here $X_{ijkm}\in[0,1]$ denotes the ratio of travel demands from node i to node j via hubs k and m. Set variables Y_k to fixed values $\hat{Y}_k$. The sub-problem (SP) is generated:

$$\min \sum_{i\in V}\sum_{j\in V}\sum_{k\in V}\sum_{m\in V}\left(\delta_1 c_{ik}+\alpha c_{km}+\delta_2 c_{mj}\right) w_{ij} X_{ijkm} \tag{50}$$

$$\text{subject to } \sum_{k\in V}\sum_{m\in V} X_{ijkm} = 1, \forall i,j\in V \tag{51}$$

$$\sum_{m\in V} X_{ijkm} + \sum_{m\in V, m\neq k} X_{ijmk} \le \hat{Y}_k, \forall i,j,k\in V \tag{52}$$

$$\sum_{i\in V}\sum_{j\in V} w_{ij} \sum_{m\in V} X_{ijkm} \le \lambda_k \hat{Y}_k, \forall k\in V \tag{53}$$

$$X_{ijkm}\in[0,1], \forall i,j,k,m\in V \tag{54}$$

Let $\sigma_{ij}\in R$, $\pi_{ijk}\ge 0$ and $\eta_k\ge 0$ be the dual variables for Equations (51–53). The dual sub-problem (DSP) is formulated below:

$$\max \sum_{i\in V}\sum_{j\in V}\left(\sigma_{ij}-\sum_{k\in V}\hat{Y}_k\pi_{ijk}\right)-\sum_{k\in V}\lambda_k\hat{Y}_k\eta_k \tag{55}$$

$$\text{subject to } \sigma_{ij}-\pi_{ijk}-\pi_{ijm}-w_{ij}\eta_k \le (\delta_1 c_{ik}+\alpha c_{km}+\delta_2 c_{mj})w_{ij}, \forall i,j,k,m\neq k\in V \tag{56}$$

$$\sigma_{ij}-\pi_{ijk}-w_{ij}\eta_k \le (\delta_1 c_{ik}+\delta_2 c_{kj})w_{ij}, \forall i,j,k\in V \tag{57}$$

$$\sigma_{ij}\in R, \forall i,j\in V \tag{58}$$

$$\pi_{ijk}\ge 0, \forall i,j,k\in V \tag{59}$$

$$\eta_k\ge 0, \forall k\in V \tag{60}$$

After solving the DSP, the obtained solution is used to construct the Benders cut as follows:

$$\theta \geq \hat{\sigma}_{ij} - \sum_{k\in V} \hat{\pi}_{ijk} Y_k - \sum_{k\in V} \lambda_k \hat{\eta}_k Y_k \tag{61}$$

In each iteration, the newly generated Benders cut is added to the following master problem (MP):

$$\min \ \theta \tag{62}$$

$$\text{subject to } \theta \geq 0 \tag{63}$$

$$\sum_{k\in V} Y_k = p \tag{64}$$

$$\sum_{k\in V} Y_k \lambda_k \geq \sum_{i\in V} O_i \tag{65}$$

$$Y_k \in \{0,1\}, \forall k \in V \tag{66}$$

Here, an additional equation (65) is added to guarantee the feasibility of the problem, i.e., the sum of the capacity of all hubs should not be smaller than the total travel demands. The complete process of Benders decomposition for solving CMApHMPC is summarized below, inspired by de Camargo et al. (2008):

1. Set $UB = +\infty$ and $LB = 0$.
2. If $LB = UB$, terminate the algorithm. The optimal solution of the original problem is obtained.
3. Solve the MP and obtain the optimal values of the objective function $\hat{z}_{MP}$ and variables $\hat{Y}_k$.
4. Let $LB = \max(LB, \hat{z}_{MP})$. Update the values of $\hat{Y}_k$ in the new DSP.
5. Solve the new DSP. Obtain the optimal values of the objective function $\hat{z}_{DSP}$ and variables $\beta_{ij}, \eta_m, \sigma_{im}, \gamma_k$.
6. Add the new Benders cut (61) to the MP. Let $UB = \min(UB, \hat{z}_{DSP})$
7. Go back to Step 2.

Based on the above process of Benders decomposition for solving the CMApHMPC, it is straightforward to implement Benders decomposition for the CMAHLPC. We only need to add the setup-cost for hubs, i.e., $\sum_{k\in V} f_k Y_k$ to the objective function of the MP and the value of the upper bound (UB). In addition, the p-hub constraint needs to be removed. Benders decomposition for the USApHMPI and USAHLPI can also be implemented based on (de Camargo et al. 2017) with some modifications. The process of Benders decomposition for all other types of hub location problems can be found in the literature (see Table 3 in the main text for details).

Appendix B: Implementation details of RG

We introduce the process of RG for USApHMPC. The row generation on Euclidean data is proposed based on the quadratic formulations with $O(n^2)$ variables (O'Kelly 1987):

$$\min \sum_{i\in V}\sum_{k\in V} A_{ik}(\delta_1 c_{ik} O_i + \delta_2 c_{ki} D_i) + \sum_{i\in V}\sum_{j\in V}\sum_{k\in V}\sum_{m\in V} \alpha c_{km} w_{ij} A_{ik} A_{jm} \tag{67}$$

$$\text{subject to } \sum_{k\in V} A_{ik} = 1, \forall i \in V \tag{68}$$

$$\sum_{k\in V} A_{kk} = p \tag{69}$$

$$A_{ik} \le A_{kk}, \forall i,k \in V \tag{70}$$

$$A_{ik} \in \{0,1\}, \forall i,k \in V \tag{71}$$

The second term (the costs for transporting travel demands through the hub network) in the objective function can be replaced by $\sum_{i\in V}\sum_{j\in V} w_{ij} T_{ij}$ and the following constraints:

$$T_{ij} \ge \sum_{k\in V}\sum_{m\in V} \alpha c_{km} w_{ij} A_{ik} A_{jm}, \forall i,j \in V \tag{72}$$

$$T_{ij} \ge 0, \forall i,j \in V \tag{73}$$

The key concept of the algorithm is to replace Equation (72) by Equation (74) as follows:

$$T_{ij} \ge \sum_{k\in V} \lambda_k^{hl} A_{ik} - \sum_{m\in V} \lambda_m^{hl} A_{jm}, \forall i,j,h,l \in V, h \ne l \tag{74}$$

where $\lambda_k^{hl} = \frac{\alpha c_{hl}}{2} + \frac{\alpha(c_{kl}^2 - c_{kh}^2)}{2c_{hl}}$. It has been proven that the above replacement is equivalent (Meier and Clausen 2018). Since there are $O(n^4)$ constraints in Equation (74), they cannot be added to the model formulation directly when solving large-scale instances. Therefore, a row generation procedure is applied. Let P, R-P and P($\hat{H}$) be the problem without Equation (74), the linear relaxation of P and the restriction of P (where hubs can only be selected from $\hat{H} \in V$), respectively. Then the row generation procedure is as follows (Meier and Clausen 2018):

1. Solve R-P with (74)-type constraints for which $i = h$ and $j = l$. A solution $\hat{A}, \hat{T}$ is obtained.
2. Let $\hat{H}$ be the set of k with $\hat{Y}_{kk} > 0$. Solve P($\hat{H}$) with all (74)-type constraints that have been added to R-P already. The obtained variable Y is recorded as $\bar{Y}_{ik}$.
3. For each $i,j \in V$, find h,l with $\bar{Y}_{ih} = \bar{Y}_{jl} = 1$. If $h \ne l$, check whether Equation (74) is satisfied. If not, add it to R-P.
4. If there are newly added (74)-type constraints, resolve R-P and go back to Step 2. Otherwise, terminate the procedure.

The above process can be adapted to CSApHMPC, USAHLPC, and CSAHLPC directly.

Appendix C: Algorithm for computing similarity index of two hub sets

Algorithm 1 Computation of similarity index for two hub sets.

```
Input: The distance (cost) $c_{ij}$ between each pair of node $(i,j)$, two hub sets $H_1$ and $H_2$.
Output: The similarity index $S_{H_1,H_2}$ of $H_1$ and $H_2$.
Let set $L = \{(i,j) | i \in H_1, j \in H_2\}$. Sort the node pairs in $L$ in an ascending order of $c_{ij}$.
Let $SeenHubs = \emptyset$, $SumDistance = 0$, $MinP = \min\{|H_1|, |H_2|\}$, and $MaxP = \max\{|H_1|, |H_2|\}$. Let $MedianC$ =the median value of all $c_{ij}$ in the dataset.
for $(i,j) \in L$ do
    if $i \notin SeenHubs$ and $j \notin SeenHubs$ then
        Let $SumDistance = SumDistance + c_{ij}$.
        Let $SeenHubs = SeenHubs \cup \{i,j\}$
        if $|SeenHubs| \geq 2 * MinP$ then
            break
        end if
    end if
end for
Let $S_{H_1,H_2} = 1 - (SumDistance + (MaxP - MinP) * MedianC)/(MaxP * MedianC)$.
```

References

de Camargo, Ricardo S., Gilberto de Miranda, Morton E. O'Kelly, James F. Campbell. 2017. Formulations and decomposition methods for the incomplete hub location network design problem with and without hop-constraints. *Applied Mathematical Modelling* **51** 274 – 301.

de Camargo, Ricardo Saraiva, G d Miranda, HP Luna. 2008. Benders decomposition for the uncapacitated multiple allocation hub location problem. *Computers & Operations Research* **35**(4) 1047–1064.

Ebery, Jamie, Mohan Krishnamoorthy, Andreas Ernst, Natashia Boland. 2000. The capacitated multiple allocation hub location problem: Formulations and algorithms. *European Journal of Operational Research* **120**(3) 614–631.

Meier, J Fabian, Uwe Clausen. 2018. Solving single allocation hub location problems on euclidean data. *Transportation Science* **52**(5) 1141–1155.

O'Kelly, Morton E. 1987. A quadratic integer program for the location of interacting hub facilities. *European Journal of Operational Research* **32**(3) 393–404.