Resilient Relay Logistics Network Design: A k-Shortest Path Approach

Onkar Kulkarni[1], Mathieu Dahan[1], Benoit Montreuil[1,2]
[1]H. Milton Stewart School for Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA 30363
[2]Coca-Cola Chair in Material Handling and Distribution
onkar.kulkarni@gatech.edu, mathieu.dahan@isye.gatech.edu, benoit.montreuil@isye.gatech.edu

Problem definition: We study the problem of designing large-scale resilient relay logistics hub networks. We propose a model of *k-Shortest Path Network Design*, which aims to improve a network's efficiency and resilience through its topological configuration, by locating relay logistics hubs to connect each origin-destination pair with k paths of minimum lengths, weighted by their forecasted demand share. ***Methodology:*** We formulate this problem as a path-based mixed-integer program with an exponential number of variables and constraints, and leverage its structure to design two scalable solution approaches based on tailored implementations of Benders decomposition and branch-and-price, respectively. In the first approach, we analytically characterize the optimal dual solutions of the exponential-sized Benders subproblem and generate feedback cuts in pseudo-polynomial time using single- and multiple-shortest path subroutines. In the second approach, we show using complementary slackness that the pricing subproblem employed within branch-and-price can also be solved in pseudo-polynomial time by computing multiple shortest paths in a carefully defined graph. ***Results:*** We apply our methodology to design large-scale resilient relay networks for a China-based parcel delivery partner. Our computational experiments demonstrate that our developed approaches can obtain optimal solutions for practically relevant instances. The resulting logistics networks showcase a significant improvement in capabilities to sustain hub disruptions with marginal compromise on nominal efficiency, in comparison with relay networks designed to only optimize nominal efficiency. ***Implications:*** This work offers valuable insights for logistics service providers aiming to design resilient relay networks with limited information regarding future demand and disruption risks. Our analysis provides decision-makers with recommendations on adjusting the topology of network designs to achieve a desired tradeoff between nominal efficiency and performance under disruptions.

Key words: Relay Network Design; Resilience; k Shortest Paths; Benders Decomposition; Branch-and-Price

1. Introduction

1.1. Motivation

The market size, measured by revenue, of the Global Parcel Delivery Services industry was \$491.5 billion in 2021 (IBISWorld 2023). In terms of parcel volume, the number crossed 159 billion across the globe (Placek 2022). In order to handle such a large volume of parcels coupled with customers' expectations of quicker parcel delivery, it is paramount for the parcel delivery industry to operate both efficiently and reliably. This requires meticulous planning, starting when designing these logistics networks (Cordeau et al. 2006, Peng et al. 2011). Such networks majorly comprise hub facilities spread out across a geographical region of interest where parcels are sorted and then consolidated to ship towards their respective destinations. Benefits of such consolidation opportunities include economies of scale, better service, and lesser emissions (Grove and O'Kelly 1986, Hall 1989).

However, traditional parcel delivery operations take a toll on the mental and physical health of delivery drivers, and present unsustainable working conditions (Sieber 2015, Nosowitz 2017). This is primarily due to the long-haul delivery trips, requiring drivers to cover distances exceeding 2,000 miles in a single journey and keeping them away from their homes and families for extended periods of time. As a direct consequence, drivers are more reluctant to work for such parcel delivery companies, leading to an increase in driver shortage, reaching 2.6 million drivers globally in 2022 (IRU 2022). In order to remedy this situation, one solution that has been identified is to build relay facilities to permit the fulfillment of demand via short-haul transportation segments (Hunt 1998, Montreuil 2011, Hirsch 2020). These relay facilities or hubs act as sortation facilities where commodities are transferred between delivery vehicles, and serve as pit stops for drivers. So, in relay logistics, the delivery drivers can advance commodities for half of their daily driving limit from one relay to the next, and then return to the original relay—ideally with other commodities—before reaching their home by the end of the day (Ali et al. 2002, Hu et al. 2019).

Logistics networks regularly face disruptions of varied nature, ranging from frequent and low-impact events such as congestion delays on roads or temporary power failures at hub facilities, to low-probability high-impact events such as reduced labor due to pandemics or localized interruption of logistics operations due to hurricanes or wildfires. Such disruptions cause delivery delays, increased logistics costs, and a dip in customer satisfaction (Kouvelis et al. 2011, Morris 2022). While the optimization literature on relay network design does not consider disruption risks (Yıldız et al. 2018, Kewcharoenwong et al. 2023), the literature on logistics network resilience primarily designs small-scale non-relay logistics networks that can mitigate supply-demand disruptions (Contreras et al. 2011, Shahabi and Unnikrishnan 2014, Ang et al. 2017). Since network structure has been shown to impact network resilience (Craighead et al. 2007), we then aim to address the following research question: *How to design efficient and resilient logistics hub network configurations for relay transportation?*

1.2. Contributions

To address the research question, we introduce the problem of *k-Shortest Path Network Design* (k-SPND), which focuses on improving a network's efficiency and resilience through its topology. We formulate the problem as an exponential-sized mixed-integer program (MIP) that locates relay logistics hubs to connect each origin-destination (O-D) pair with k paths of minimum lengths, weighted by the O-D pair's forecasted demand share.

We develop two approaches for solving the large-scale MIP optimally: In the first approach, based on a tailored implementation of Benders decomposition (Algorithm 3), we provide an analytical characterization of the optimal dual solutions of the exponential-sized Benders subproblem to generate the feedback cuts (Proposition 1). This leads to a pseudo-polynomial time approach to generate these cuts based on Yen's algorithm (for computing k shortest paths; Yen (1971)), which we accelerate using breadth-first-search and shortest-path subroutines (Algorithms 1 and 2).

In the second solution approach, we tailor an implementation of branch-and-price (Algorithm 7): At each node of the branch-and-bound tree, we solve the master problem—a linear program with an exponential number of variables and constraints—using column generation (Algorithm 5). Using complementary slackness we show in Proposition 2 that at each iteration of column generation, the pricing subproblem can also be solved in pseudo-polynomial time by adjusting Yen's algorithm in an auxiliary graph with edge lengths depending on the optimal dual variables of the restricted master problem (Algorithm 4). To further accelerate our approach, we also develop a greedy algorithm (Algorithm 6) that improves the branching process.

We conduct an extensive case study pertaining to the design of large-scale resilient relay networks for a China-based parcel delivery service provider that partnered with our research team. We show that our solution approaches can obtain optimal solutions for practically relevant instances and that our tailored implementation of Benders decomposition (respectively, branch-and-price) converges faster for smaller (respectively, larger) instances. We assess the performance of the generated networks by determining minimum-cost consolidation plans to satisfy average parcel demand in the presence versus absence of disruptions, and compare it with that of relay networks designed to only optimize efficiency in nominal situations. Our designed networks showcase significantly higher capabilities to sustain disruptions, as they can fulfill more demand by short-haul transportation and at a lower cost, with limited compromise on nominal efficiency. Finally, we conduct a sensitivity analysis to quantify the impact of the number of paths and hubs to open on the performance of the designed networks, offering valuable insights to logistics service providers on designing networks that achieve a desired tradeoff between nominal efficiency and resilience.

The remainder of this article is organized as follows: Section 2 reviews the existing literature on relay network design and network resilience. We then formulate the k-SPND problem in Section 3. In Sections 4 and 5, we develop our two solution approaches. Section 6 presents results from a case study to validate our model and solution approaches when designing large-scale resilient relay networks for a China-based parcel delivery partner. Section 7 lays out the concluding remarks and provides avenues for future research. Finally, the electronic companion (EC) contains mathematical proofs and additional computational results.

2. Literature Review

2.1. Relay Network Design

For the past 35 years, researchers have studied the Hub Location Problem (HLP) extensively (see the reviews by Campbell and O'Kelly (2013), Alumur et al. (2021)). In HLP, the aim is to select locations from a discrete set of candidates to build hub facilities and serve demand between a given set of O-D pairs (Hall 1989). Such network design models seek to find the optimal tradeoff between hub construction costs and transportation costs. Due to typically large hub construction costs and substantial transportation cost savings through consolidation, the resulting network designs contain very few hubs and rely heavily on long-haul delivery trips, thus negatively impacting the drivers' quality of life.

In order to address this issue, more recent work has proposed to instead build relay facilities with lower construction costs and operate them to support commodity delivery for a given set of O-D pairs while respecting the drivers' driving limit (Jacquillat et al. 2022, Kewcharoenwong et al. 2023). This *relay network design problem* aims to select relay hub locations that minimize total relay hub construction and transportation costs to satisfy commodity demand while respecting the driver tour length constraints (Hunt 1998, Ali et al. 2002, Cabral et al. 2007, Üster and Maheshwari 2007, Ballot et al. 2012). Due to a high number of potential paths for delivering commodities, even with innovative solution methodologies, investigations in the literature are only able to solve small instances and design relay-hub networks for narrower geographical regions (Kulturel-Konak and Konak 2008, Üster and Kewcharoenwong 2011, Konak 2012, Kewcharoenwong and Üster 2014, Yıldız et al. 2018, Leitner et al. 2019, 2020, Ziaeifar and Üster 2023, Kewcharoenwong et al. 2023).

Furthermore, an important characteristic of relay networks designed in the literature is their hierarchical nature: The non-hub nodes such as commodity origins and destinations, commonly referred to as spoke nodes, are allocated to specific hubs, forcing commodities to travel only along the associated legs. However, such restriction constrains commodity flows and leads to an increase in traveled distances (Tu and Montreuil 2019). This restriction coupled with the parcel cut-off times usually implemented by parcel delivery companies create severe parcel congestion at the hub facilities, which ultimately cause parcel delivery delays (Montreuil et al. 2018). To overcome the aforementioned issues, the concept of Physical Internet recently emerged to design hyperconnected networks, which are multi-tier hub networks that interconnect open-access hub facilities at multiple planes (Montreuil 2011, Montreuil et al. 2018). This hyperconnectivity provides better degrees of freedom for parcel movement while preserving cost savings achieved through consolidation opportunities (Kulkarni et al. 2023). In our work, we leverage this framework to design densely connected relay-hub networks that are spread out across a wide geographical region.

2.2. Network Resilience

Logistics hub networks face numerous disruption events during their lifetime and it is paramount to mitigate their effects. Owing to this fact, network resilience has been gaining substantial attention from academia and industry across the globe. Concerning logistics networks, disruptions could occur at logistics hub facilities, transportation legs, or spoke nodes. Majorly, they are classified as either total disruptions where network components (hubs, legs, spokes) are totally dysfunctional, or as partial disruptions where components are functional but at reduced capacity (Mohammadi et al. 2016).

Disruptions that usually occur at spoke nodes are due to commodity supply-demand variability. To hedge against this uncertainty, the networks are generally designed for various demand scenarios, and in each scenario, the commodity routing is adaptive to the uncertain demand realization. Such network design models either use stochastic optimization or robust optimization toolkits (Contreras et al. 2011, Shahabi and Unnikrishnan 2014, Ang et al. 2017, Zetina et al. 2017, Taherkhani et al. 2020, Basciftci et al. 2021). However, network designs obtained through robust optimization may lead to conservative networks. On the other hand, stochastic optimization models require a significant number of demand scenarios, in which case the models become too large to solve for realistic size instances that are of interest to the industry. Furthermore, because of the dramatic and uncertain commodity demand fluctuation, it can be challenging to properly select future demand scenarios when designing desired logistics networks.

In order to tackle disruptions at hubs or legs, researchers have proposed several reactive strategies from either an operational or tactical perspective (An et al. 2015). These strategies include potential premeditated action plans that are to be employed in case of any disruptions. They include but are not limited to re-routing commodities through alternate paths in case the primary path is unavailable (Kim 2008), using pre-assigned backup hubs when the original hub is totally disrupted (Azizi 2017), fortification of a small number of hubs as an additional protection layer to avoid those hubs being disrupted (Ghaffarinasab and Atayi 2018), and re-allocation of non-hub nodes to functional hub facilities during the disruptions (Shen et al. 2011, Chen et al. 2022). Although these strategies help mitigate the impact of disruptions in a proper manner, they are usually difficult and expensive to implement (Mohammadi et al. 2019). Therefore, to achieve both economic advantage and network reliability, the network design itself should also consider alternative methods to induce resilience in case of such disruptions.

One set of approaches to design resilient logistics hub networks in a scalable manner is by focusing on their structure. Topological properties of a logistics network's structure have been employed as surrogates to measure network resilience (Craighead et al. 2007, Ip and Wang 2009, 2011, Osei-Asamoah and Lownes 2014, Kulkarni et al. 2021). However, very few investigations have incorporated such topological measures to induce resilience at the network design stage, especially for relay logistics. To the best of our knowledge, only Kulkarni et al. (2021, 2022) proposed to design resilient hyperconnected logistics networks by minimizing the total length of k shortest paths and k edge-disjoint shortest paths between each O-D pair. While the authors optimally solve the edge-disjoint formulation (which leads to more conservative network designs), they simply designed a matheuristic to solve the non-edge-disjoint formulation. Furthermore, they did not include any commodity demand information. As a result, their computational study simply evaluated the impact of edge disruptions on the length of the shortest paths between each O-D pair, which has limitations in terms of approximating logistics operations. In contrast, we develop scalable exact solution approaches to minimize the demand-share-weighted total length of the k shortest paths between each O-D pair. Furthermore, we conduct an extensive computational study to better estimate the resilience of the designed networks by computing minimum-cost consolidation plans.

3. *k*-Shortest Path Network Design Modeling

In this section, we first present the k-Shortest Path Network Design Problem. We then derive a mixed-integer programming formulation of the problem using path-based decisions.

3.1. Problem Description

We consider a logistics service provider or a consortium of such providers interested in designing a large-scale logistics hub network for efficient and resilient relay transportation. This problem is motivated by existing unsustainable long-haul delivery trip schedules that affect drivers' mental and physical health (Sieber 2015, Nosowitz 2017). The objective is then to open relay hubs that will permit the efficient fulfillment of long-haul demand using daily driver trips that will be resilient against logistics disruptions.

We consider the initial planning phase of the design process and assume that the service provider has limited information regarding future demand and disruption risks. As a result, we introduce the problem of *k-Shortest Path Network Design* (k-SPND), which consists of locating logistics hubs to connect each origin-destination (O-D) pair with at least $k \geq 1$ routes of minimum total lengths. The premise is that by connecting O-D pairs with multiple short routes, it will then be possible to cost-effectively transport commodities with appropriate consolidation given the realized demand. In addition, if a multi-day disruption occurs at a hub or a transportation leg (e.g., due to flooding or a wildfire), then the service provider will be capable of transporting commodities via a different route, with a marginal impact on delivery cost and time.

Formally, let $\mathcal{S}$ (respectively, $\mathcal{T}$) represent the discrete set of origin (respectively, destination) locations of future commodities. For each O-D pair $p \in \mathcal{P} \subseteq \mathcal{S} \times \mathcal{T}$, we denote by $d_p \in [0,1]$ the associated demand share that the logistics service provider plans to serve; the demand shares satisfy $\sum_{p\in\mathcal{P}} d_p = 1$. For instance, d_p can represent the fraction of the total volume estimated to be fulfilled for O-D pair p over a long planning horizon. To serve the forecasted demand share, the service provider intends to open N relay logistics hubs from a pre-selected set of discrete candidate locations $\mathcal{H}$ aimed to facilitate commodity deliveries from their origins to their corresponding destinations through short-haul transportation. These logistics hubs act as sortation facilities where commodities are transferred between delivery vehicles depending on the respective destinations, and serve as pit stops for drivers. Hubs will be assumed to have sufficient capacity to handle large commodity volumes and sustain future logistics operations.

Building upon the principles of the Physical Internet, we permit each origin, destination, and potential hub location to be connected to multiple other locations to allow the design of a hyper-connected network. We represent as $\mathcal{A} \subseteq (\mathcal{S}\cup\mathcal{T}\cup\mathcal{H})^2$ the set of potential (directed) transportation legs, which satisfy the traveled distance or driving time regulations to ensure a daily return for all drivers to their respective homes. For every leg $(i,j)\in\mathcal{A}$, we let $\tau_{i,j}$ denote the travel time from i to j, which can also incorporate processing times at locations i and j. We define by $\mathcal{G} \coloneqq (\mathcal{S}\cup\mathcal{T}\cup\mathcal{H},\mathcal{A})$ the directed graph of potential network designs.

To increase the resilience of service operations to disruptions, we consider that the service provider aims to guarantee the existence of $k \in \mathbb{Z}_{>0}$ routes (or paths) in the relay network between each O-D pair. For every O-D pair $p=(s,t)\in\mathcal{P}$, we define an $s-t$ (simple) path in the graph $\mathcal{G}$ as a sequence of adjacent and distinct nodes $\lambda = \{s, h_1, \ldots, h_n, t\}$ starting at origin $s\in\mathcal{S}$, visiting hub locations via edges in $\mathcal{A}$, and ending at destination $t\in\mathcal{T}$. For every O-D pair $p=(s,t)\in\mathcal{P}$, we denote the set of $s-t$ paths in $\mathcal{G}$ as Λ_p. We also denote as $\Lambda \coloneqq \cup_{p\in\mathcal{P}}\Lambda_p$ the set of all paths connecting all O-D pairs. Finally, for every path $\lambda=\{s,h_1,\ldots,h_n,t\}\in\Lambda_p$, we denote its total travel time (or length) by $\tau_\lambda \coloneqq \tau_{s,h_1} + \sum_{i=1}^{n-1}\tau_{h_i,h_{i+1}} + \tau_{h_n,t}$.

The goal of the k-SPND problem is then to select a subset of hub locations $\mathcal{H}_o \subseteq \mathcal{H}$ of size at most N so as to minimize the demand-share-weighted total length of the k shortest paths between each O-D pair in the subgraph of $\mathcal{G}$ induced by the set of nodes $\mathcal{S}\cup\mathcal{T}\cup\mathcal{H}_o$, which we denote as $\mathcal{G}_{\mathcal{H}_o}$. The k-SPND problem can be modeled as follows:

$$\min_{\mathcal{H}_o,\,(\lambda_1^p,\ldots,\lambda_k^p)_{p\in\mathcal{P}}} \quad \sum_{p\in\mathcal{P}}\sum_{i=1}^{k} d_p\cdot\tau_{\lambda_i^p} \tag{1a}$$

$$\text{s.t.} \quad |\mathcal{H}_o| \le N, \tag{1b}$$

$$\lambda_i^p \ne \lambda_j^p, \qquad \forall\, p\in\mathcal{P},\ \forall\, i,j\in[k] \mid i\ne j, \tag{1c}$$

$$\lambda_i^p \setminus \{s,t\} \subseteq \mathcal{H}_o, \quad \forall\, p\in\mathcal{P},\ \forall\, i\in[k], \tag{1d}$$

$$\mathcal{H}_o \subseteq \mathcal{H}, \tag{1e}$$

$$\lambda_1^p,\ldots,\lambda_k^p \in \Lambda_p, \quad \forall\, p\in\mathcal{P}, \tag{1f}$$

where $[k] \coloneqq \{1, \ldots, k\}$. Constraints (1d) ensure that the selected paths only traverse opened hubs.

This model is useful in situations where the service provider has limited information regarding the exact volume of commodities to transport—as it can vary significantly throughout the hubs' lifetime—or the disruption risks faced by the designed network (e.g., hub or route disruptions caused by extreme events such as natural disasters or pandemics). Despite the uncertainty faced by the service provider over the long planning horizon at the design stage, this model focuses on improving efficiency and resilience through the structure of the network configuration.

Next, we illustrate the k-SPND problem with the following example:

EXAMPLE 1. Consider the graph $\mathcal{G}$ of possible network designs illustrated in Figure 1. In this example, the sets of origins, destinations, and candidate hub locations are respectively given by $\mathcal{S} = \{s_1, s_2\}$, $\mathcal{T} = \{t_1, t_2\}$, and $\mathcal{H} = \{h_1, h_2, h_3, h_4\}$. The set of transportation legs $\mathcal{A}$ is represented by the dashed arrows. In this example, we consider a set of two O-D pairs $\mathcal{P} = \{(s_1, t_1), (s_2, t_2)\}$ with equal demand shares $d_p = 0.5$, and that the logistics company wishes to open up to $N = 3$ relay hubs. In Figure 2, we depict the optimal solutions of the k-SPND problem when $k = 1$ and $k = 2$.

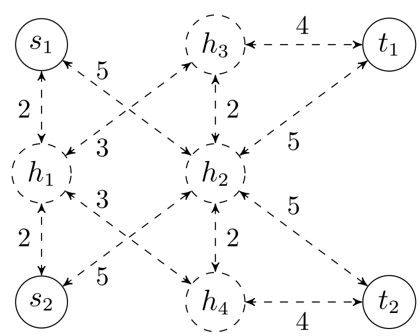

Figure 1 Graph $\mathcal{G}$ of potential network designs to fulfill the demand between two O-D pairs (s_1, t_1) and (s_2, t_2). Dashed nodes represent candidate hub locations and edge labels represent travel times (in hours).

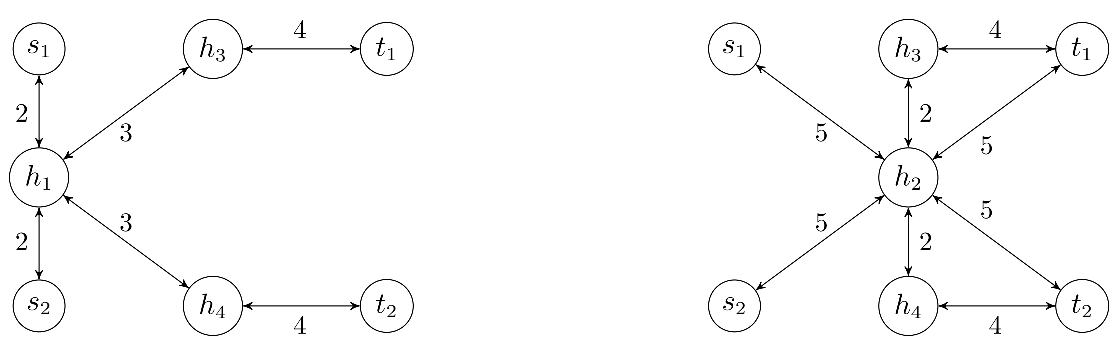

Figure 2 Optimal solutions of k-SPND for $k = 1$ (left) and for $k = 2$ (right), with $N = 3$ opened hubs.

When $k = 1$, the optimal set of hubs to open is given by $\mathcal{H}_o = \{h_1, h_3, h_4\}$, and the length of the shortest path between each O-D pair is 9 hours. In contrast, when $k = 2$, the optimal set of hubs to open is $\mathcal{H}_o = \{h_2, h_3, h_4\}$ and each O-D pair is connected by 2 paths of respective lengths 10 and 11 hours. While the former network permits the commodities to travel over shorter distances, we observe that any hub or edge disruption in the network will disconnect at least one O-D pair and prevent the service provider from fulfilling the demand using short-haul transportation. On the other hand, the second network can better withstand random disruptions and has a higher chance of maintaining connectivity between each O-D pair via a short path. △

3.2. Path-Based Mixed-Integer Programming Formulation

The combinatorial model (1) cannot be solved directly. Instead, we reformulate it as a mixed-integer program (MIP) using path-based decisions. We consider for each hub $i \in \mathcal{H}$ a *binary* variable y_i that takes a value of 1 if hub i is opened, and 0 otherwise. Additionally, for every O-D pair $p = (s,t) \in \mathcal{P}$ and every $s-t$ path $\lambda \in \Lambda_p$, we define a *continuous* variable z_λ that will be equal to 1 if λ is selected as one of the k shortest $s-t$ paths in the subgraph induced by the opened hubs. We then derive the following MIP:

$$\mathcal{I}_{path}: \quad \min_{y,z} \quad \sum_{p \in \mathcal{P}} \sum_{\lambda \in \Lambda_p} d_p \cdot \tau_\lambda \cdot z_\lambda \tag{2a}$$

$$\text{s.t.} \quad \sum_{i \in \mathcal{H}} y_i \leq N, \tag{2b}$$

$$\sum_{\lambda \in \Lambda_p} z_\lambda = k, \qquad \forall\, p \in \mathcal{P}, \tag{2c}$$

$$\sum_{\{\lambda \in \Lambda_p \,|\, i \in \lambda\}} z_\lambda \leq k \cdot y_i, \quad \forall\, p \in \mathcal{P},\ \forall\, i \in \mathcal{H}, \tag{2d}$$

$$0 \leq z_\lambda \leq 1, \qquad \forall\, \lambda \in \Lambda, \tag{2e}$$

$$y_i \in \{0,1\}, \qquad \forall\, i \in \mathcal{H}. \tag{2f}$$

Objective (2a) minimizes the total travel time on the selected paths, weighted by the O-D pair demand shares. Constraint (2b) ensures at most N hubs are opened while Constraints (2c) guarantee the selection of k $s-t$ paths for each O-D pair $p = (s,t) \in \mathcal{P}$. Finally, from Constraints (2d), if an $s-t$ path traverses a hub $i \in \mathcal{H}$, then i must be opened. In the next lemma, we show that for fixed hub selection variables, the feasible region of the path selection variables for each O-D pair is an integral polyhedron, thus guaranteeing the validity of the formulation (2).

Lemma 1. *$\mathcal{I}_{path}$ is a MIP formulation of the k-SPND problem.*

The MIP formulation (2) is a challenging optimization problem to solve for practical real-world problem instances, as it contains an exponential number of variables ($|\mathcal{H}| + |\Lambda|$) and an exponential number of constraints $((|\mathcal{H}|+1)\cdot(|\mathcal{P}|+1)+|\Lambda|)$. Kulkarni et al. (2022) considered a similar formulation with only binary variables, for which they developed a matheuristic to obtain good-quality solutions. Our formulation differs in the path-hub linking constraints (2d), which allows the path selection variables to be continuous. This in turn will enable us to tailor Benders decomposition in Section 4 and the branch-and-price algorithm in Section 5 to optimally solve the k-SPND problem for a wide range of instances in Section 6.

4. Solution Methodology 1: Benders Decomposition

In this section, we leverage the network and block-angular structure of the k-SPND problem to develop a scalable solution approach based on Benders decomposition.

4.1. Two-Stage Decomposition

The MIP formulation (2) exhibits a special structure: If we first fix the hub selection variables, then the path selection variables for each O-D pair $(s,t)\in\mathcal{P}$ can be separately determined either by solving a linear program with an exponential number of variables and constraints, or by computing the k shortest $s-t$ paths in the subgraph of $\mathcal{G}$ induced by the opened hubs using Yen's algorithm (Yen 1971). We leverage this property to design and accelerate our first solution approach based on Benders decomposition. Specifically, we decompose $\mathcal{I}_{path}$ as follows:

$$
\begin{aligned}
\min_{y} \quad & \sum_{p\in\mathcal{P}} d_p\cdot\theta_p^*(y) \\
\text{s.t.} \quad & \sum_{i\in\mathcal{H}} y_i \le N, \\
& y_i\in\{0,1\}, \quad \forall\, i\in\mathcal{H},
\end{aligned}
$$

where for every O-D pair $p\in\mathcal{P}$, the subproblem and its dual are given by:

$$
\begin{aligned}
\theta_p^*(y) = \min_{z} \quad & \sum_{\lambda\in\Lambda_p}\tau_\lambda\cdot z_\lambda && = \max_{\alpha,\beta,\gamma}\; k\cdot\alpha - \sum_{i\in\mathcal{H}} k\cdot y_i\cdot\beta_i - \sum_{\lambda\in\Lambda_p}\gamma_\lambda \\
\text{s.t.} \quad & \sum_{\lambda\in\Lambda_p} z_\lambda = k, && \quad \text{s.t.}\;\; \alpha - \sum_{i\in\mathcal{H}\cap\lambda}\beta_i - \gamma_\lambda \le \tau_\lambda, \qquad \forall\,\lambda\in\Lambda_p, \\
& \sum_{\{\lambda\in\Lambda_p\,|\,i\in\lambda\}} z_\lambda \le k\cdot y_i,\ \forall\, i\in\mathcal{H}, && \qquad\;\; \beta_i\ge 0, \qquad \forall\, i\in\mathcal{H}, \\
& 0\le z_\lambda\le 1, \quad \forall\,\lambda\in\Lambda_p, && \qquad\;\; \gamma_\lambda\ge 0, \qquad \forall\,\lambda\in\Lambda_p.
\end{aligned}
\tag{3}
$$

To circumvent feasibility issues in the primal subproblem (3), we add dummy paths of very large lengths between each O-D pair that can be traversed even when no hub is opened. Then, $\mathcal{I}_{path}$ is equivalent to the following master problem:

$$\mathcal{M}_{path}(\mathcal{C}): \quad \min_{y,\theta} \quad \sum_{p\in\mathcal{P}} d_p \cdot \theta_p \tag{4a}$$

$$\text{s.t.} \quad \theta_p \geq k\cdot\alpha - \sum_{i\in\mathcal{H}} k\cdot y_i\cdot\beta_i - \sum_{\lambda\in\Lambda_p}\gamma_\lambda, \quad \forall\, p\in\mathcal{P},\ \forall\,(\alpha,\beta,\gamma)\in\mathcal{C}_p, \tag{4b}$$

$$\sum_{i\in\mathcal{H}} y_i \leq N, \tag{4c}$$

$$y_i\in\{0,1\}, \qquad \forall\, i\in\mathcal{H}, \tag{4d}$$

where for every O-D pair $p\in\mathcal{P}$, the set $\mathcal{C}_p$ contains all feasible solutions to the dual subproblem (3) (or simply its basic feasible solutions).

4.2. Constraint Generation

Due to the very large number of constraints (4b), we solve the master problem $\mathcal{M}_{path}(\mathcal{C})$ using constraint generation: We initially solve a relaxed version of the master problem $\mathcal{M}_{path}(\mathcal{C}')$ where $\mathcal{C}'_p = \emptyset$ for every $p\in\mathcal{P}$; let (y^*,θ^*) be its optimal solution. We then determine if this solution violates any of the constraints (4b) in the original master problem. To this end, we compute for each O-D pair $p\in\mathcal{P}$ an optimal dual solution $(\alpha^*,\beta^*,\gamma^*)$ of the subproblem (3). If the corresponding constraint (4b) is violated by (y^*,θ^*), then $(\alpha^*,\beta^*,\gamma^*)$ is added to $\mathcal{C}'_p$ and the relaxed master problem is solved again. Benders decomposition terminates when the optimal solution to the relaxed master problem satisfies every constraint (4b).

However, at each iteration of the algorithm, we must solve the dual subproblem (3), which has an exponential number of variables and constraints. By leveraging the subproblem's structure, we provide an analytical characterization of its optimal dual solutions:

Proposition 1. *Consider a vector of hub selection variables $y^*\in\{0,1\}^{|\mathcal{H}|}$ and an O-D pair $p=(s,t)\in\mathcal{P}$. Let $\mathcal{H}_o=\operatorname{supp}(y^*)$ be the set of opened hubs, $\tau^\dagger_p$ be the shortest $s-t$ path length in the original graph $\mathcal{G}$, and $(\lambda^p_1,\ldots,\lambda^p_k)$ be the k shortest $s-t$ paths in $\mathcal{G}_{\mathcal{H}_o}$ ordered by their lengths. Then, an optimal solution to the dual subproblem (3) is given by*

$$\alpha^* = \tau_{\lambda^p_k}, \qquad \beta^*_i = \begin{cases} w_i\cdot(\tau_{\lambda^p_k}-\tau^\dagger_p) & \text{if } i\in\mathcal{H}\setminus\mathcal{H}_o,\\ 0 & \text{otherwise,}\end{cases} \qquad \gamma^*_\lambda = \begin{cases}\tau_{\lambda^p_k}-\tau_\lambda & \text{if } \lambda\in\{\lambda^p_1,\ldots,\lambda^p_k\},\\ 0 & \text{otherwise,}\end{cases}$$

for any $w\in\mathbb{R}^{|\mathcal{H}\setminus\mathcal{H}_o|}_{\geq 0}$ that satisfies $\sum_{i\in\lambda\cap\mathcal{H}\setminus\mathcal{H}_o} w_i \geq \mathbb{1}_{\{\lambda\cap\mathcal{H}\setminus\mathcal{H}_o\neq\emptyset\}}$ for every $\lambda\in\Lambda_p$.

The corresponding Benders cut is given by:

$$\theta_p \geq \sum_{\ell=1}^{k}\tau_{\lambda^p_\ell} - \sum_{i\in\mathcal{H}\setminus\mathcal{H}_o} k\cdot w_i\cdot(\tau_{\lambda^p_k}-\tau^\dagger_p)\cdot y_i.$$

From Proposition 1, we find that although the dual subproblem (3) is of exponential size, we can still optimally solve it and construct a cut in pseudo-polynomial time: For each O-D pair $p = (s,t) \in \mathcal{P}$, we compute the k shortest $s-t$ paths $(\lambda_1^p, \ldots, \lambda_k^p)$ in the subgraph of $\mathcal{G}$ induced by the opened hubs $\mathcal{H}_o = \text{supp}(y^*)$ (given by the support of the vector of variables y^* at optimality of the relaxed master problem $\mathcal{M}_{path}(\mathcal{C}'))$ using Yen's algorithm. Then, the optimal dual variables α^*, $(\gamma_\lambda^*)_{\lambda\in\Lambda_p}$, and $(\beta_i^*)_{i\in\mathcal{H}_o}$ can directly be determined from $(\lambda_1^p, \ldots, \lambda_k^p)$. The remaining variables $(\beta_i^*)_{i\in\mathcal{H}\setminus\mathcal{H}_o}$ must be selected to ensure feasibility of the dual constraint associated with each $s-t$ path $\lambda \in \Lambda_p$ that traverses an unopened hub location. In fact, there is an infinite number of such solutions. One option is to simply assign a very large number to every β_i^* for $i \in \mathcal{H}\setminus\mathcal{H}_o$. However, such an approach would result in a weak cut. Instead, we first compute the shortest $s-t$ path length $\tau_p^\dagger$ in the original graph $\mathcal{G}$, and show that by selecting β_i^* for $i \in \mathcal{H}\setminus\mathcal{H}_o$ to be $w_i \cdot (\tau_{\lambda_k^p} - \tau_p^\dagger)$ with $\sum_{i\in\lambda\cap\mathcal{H}\setminus\mathcal{H}_o} w_i \geq \mathbb{1}_{\{\lambda\cap\mathcal{H}\setminus\mathcal{H}_o\neq\emptyset\}}$ for every $\lambda \in \Lambda_p$, we ensure that *all* dual constraints are satisfied.

We propose two alternative algorithms for selecting the weights $(w_i)_{i\in\mathcal{H}\setminus\mathcal{H}_o}$, thus generating two different cuts. In the first approach described in Algorithm 1 below, we implement an adjusted breadth-first search that assigns a weight $w_i = 1$ to any hub $i \in \mathcal{H}\setminus\mathcal{H}_o$ that is the first unopened hub encountered by at least one $s-t$ path.

Algorithm 1: Adjusted Breadth-First Search (ABFS$(p, \mathcal{H}_o)$)

Input : O-D pair $p = (s,t) \in \mathcal{P}$, set of opened hubs $\mathcal{H}_o$

Output: Vector $w \in \mathbb{R}_{\geq 0}^{|\mathcal{H}\setminus\mathcal{H}_o|}$ satisfying $\sum_{i\in\lambda\cap\mathcal{H}\setminus\mathcal{H}_o} w_i \geq \mathbb{1}_{\{\lambda\cap\mathcal{H}\setminus\mathcal{H}_o\neq\emptyset\}}$, $\forall\, \lambda \in \Lambda_p$

$\forall\, i \in \mathcal{H}\setminus\mathcal{H}_o$, $w_i \leftarrow 0$;
$Q \leftarrow (s)$;
while $Q \neq \emptyset$ **do**
    $i \leftarrow$ Dequeue Q;
    Mark i as explored;
    for every $j \in \mathcal{H} \mid (i,j) \in \mathcal{A}$ **do**
        if j is not explored **then**
            if $j \in \mathcal{H}\setminus\mathcal{H}_o$ **then**
                $w_j \leftarrow 1$;
            else
                Enqueue j to Q;
return w

In the second approach, we instead spread the assignment of these weights w_i depending on the number of unopened hubs traversed by each $s-t$ path. Specifically, for each unopened hub $i \in \mathcal{H} \setminus \mathcal{H}_o$, we assign the following weight:

$$w_i = \frac{1}{\min\{|\lambda \cap \mathcal{H} \setminus \mathcal{H}_o|,\ \forall\, \lambda \in \Lambda_p \text{ s.t. } i \in \lambda\}},$$

which can be implemented in polynomial time by determining the $s-t$ path $\lambda \in \Lambda_p$ traversing i that traverses the minimum number of unopened hub locations. This in turn can be implemented by setting the length of every edge $(i,j) \in \mathcal{A}$ to $\mathbb{1}_{\{j \in \mathcal{H} \setminus \mathcal{H}_o\}}$ and computing the shortest path that traverses i (see Algorithm 2 below).

Algorithm 2: Spread Assignment (SA$(p, \mathcal{H}_o)$)

Input : O-D pair $p = (s,t) \in \mathcal{P}$, set of opened hubs $\mathcal{H}_o$

Output: Vector $w \in \mathbb{R}_{\geq 0}^{|\mathcal{H} \setminus \mathcal{H}_o|}$ satisfying $\sum_{i \in \lambda \cap \mathcal{H} \setminus \mathcal{H}_o} w_i \geq \mathbb{1}_{\{\lambda \cap \mathcal{H} \setminus \mathcal{H}_o \neq \emptyset\}}$, $\forall\, \lambda \in \Lambda_p$

$\mathcal{G}' \leftarrow$ Copy of $\mathcal{G}$;
for every $(i,j) \in \mathcal{A}$ **do**
  if $j \in \mathcal{H} \setminus \mathcal{H}_o$ **then**
    Set the length of (i,j) in $\mathcal{G}'$ to 1;
  else
    Set the length of (i,j) in $\mathcal{G}'$ to 0;
for every $i \in \mathcal{H} \setminus \mathcal{H}_o$ **do**
  $\lambda' \leftarrow$ shortest $s-t$ path in $\mathcal{G}'$ traversing i, computed using Dijkstra's algorithm;
  $w_i \leftarrow \dfrac{1}{|\lambda' \cap \mathcal{H} \setminus \mathcal{H}_o|}$;
return w

Both approaches guarantee that $\sum_{i \in \lambda \cap \mathcal{H} \setminus \mathcal{H}_o} w_i \geq \mathbb{1}_{\{\lambda \cap \mathcal{H} \setminus \mathcal{H}_o \neq \emptyset\}}$ for every $s-t$ path $\lambda \in \Lambda_p$ and generate strong cuts. Empirically, we find that simultaneously adding both cuts to the relaxed master problem when they are violated by the current optimal solution of $\mathcal{M}_{path}(\mathcal{C}')$ provides the best results (see Table EC.1 for a numerical comparison). Once the cuts have been added via the solver callback function, the relaxed master problem $\mathcal{M}_{path}(\mathcal{C}')$ is solved again, and this process is repeated until an optimal solution to $\mathcal{I}_{path}$ is obtained.

4.3. Overall Algorithm

Our tailored implementation of Benders decomposition for solving $\mathcal{I}_{path}$ is detailed in Algorithm 3.

Algorithm 3: Benders Decomposition for Solving $\mathcal{I}_{path}$

Input : Graph $\mathcal{G}=(\mathcal{S}\cup\mathcal{H}\cup\mathcal{T},\mathcal{A})$, vector of demand shares $d\in[0,1]^{|\mathcal{P}|}$, number of hubs to open $N\in\mathbb{Z}_{>0}$, integer $k\in\mathbb{Z}_{>0}$, optimality gap $\epsilon\geq 0$

Output: Optimal subset of hubs to open $\mathcal{H}_o$ and k shortest paths between each O-D pair $(\lambda_\ell^p)_{(p,\ell)\in\mathcal{P}\times[k]}$

$L\leftarrow-\infty,\quad U\leftarrow+\infty,\quad \mathcal{C}'_p\leftarrow\emptyset\ \ \forall p\in\mathcal{P}$;

$\forall p=(s,t)\in\mathcal{P},\ \tau_p^\dagger\leftarrow$ shortest $s-t$ path length in $\mathcal{G}$ computed using Dijkstra's algorithm;

while $U-L>\epsilon$ **do**

  Solve $\mathcal{M}_{path}(\mathcal{C}')$: $(y^*,\theta^*)\leftarrow$ optimal solution, $\quad L\leftarrow$ optimal value;

  $\mathcal{H}_o\leftarrow\operatorname{supp}(y^*)$;

  for every $p=(s,t)\in\mathcal{P}$ **do**

    $(\lambda_1^p,\ldots,\lambda_k^p)\leftarrow k$ shortest $s-t$ paths in $\mathcal{G}_{\mathcal{H}_o}$ computed using Yen's algorithm and ordered by their lengths;

    if $\theta_p^*<\sum_{\ell=1}^k\tau_{\lambda_\ell^p}$ **then**

      $w^1\leftarrow\text{ABFS}(p,\mathcal{H}_o)$;

      $\mathcal{C}'_p\leftarrow\mathcal{C}'_p\cup\{(\tau_{\lambda_k^p},(w_i^1\cdot(\tau_{\lambda_k^p}-\tau_p^\dagger)\cdot\mathbb{1}_{\{i\in\mathcal{H}\setminus\mathcal{H}_o\}})_{i\in\mathcal{H}},((\tau_{\lambda_k^p}-\tau_\lambda)\cdot\mathbb{1}_{\{\lambda\in\{\lambda_1^p,\ldots,\lambda_k^p\}\}})_{\lambda\in\Lambda_p})\}$;

      $w^2\leftarrow\text{SA}(p,\mathcal{H}_o)$;

      $\mathcal{C}'_p\leftarrow\mathcal{C}'_p\cup\{(\tau_{\lambda_k^p},(w_i^2\cdot(\tau_{\lambda_k^p}-\tau_p^\dagger)\cdot\mathbb{1}_{\{i\in\mathcal{H}\setminus\mathcal{H}_o\}})_{i\in\mathcal{H}},((\tau_{\lambda_k^p}-\tau_\lambda)\cdot\mathbb{1}_{\{\lambda\in\{\lambda_1^p,\ldots,\lambda_k^p\}\}})_{\lambda\in\Lambda_p})\}$;

  $U\leftarrow\min\{U,\sum_{p\in\mathcal{P}}\sum_{\ell=1}^k d_p\cdot\tau_{\lambda_\ell^p}\}$;

return $(\mathcal{H}_o,(\lambda_\ell^p)_{(p,\ell)\in\mathcal{P}\times[k]})$

5. Solution Methodology 2: Branch-and-Price Algorithm

Although the MIP $\mathcal{I}_{path}$ is of exponential size, we remark that the large number of constraints comes from the upper bounds (2e) on the path selection variables. In this section, we propose a solution approach that builds upon the branch-and-price algorithm, which is typically used to solve MIPs with a small number of constraints.

5.1. Column Generation

In the branch-and-price algorithm, the linear programming (LP) relaxation of $\mathcal{I}_{path}$, called the master problem, is solved using column generation. Specifically, the column generation algorithm solves a restricted version of the master problem where only a subset of variables are considered; let (y^*,z^*) be the resulting solution. Then, the pricing subproblem determines the variable in the original master problem with the lowest reduced cost. If the lowest reduced cost is negative, the corresponding variable is then added to the restricted master problem, which is solved again. Column generation terminates when the lowest reduced cost is nonnegative, at which point (y^*,z^*)

optimally solves the master problem. If (y^*, z^*) has fractional components, then branching occurs and two new LPs are solved using column generation. The state space search is governed by a branch-and-bound tree. Furthermore, to accelerate the computations at each tree node, column generation is initialized by considering the variables generated when solving its parent node.

We now describe our approach for solving the LP relaxation at each node of the branch-and-bound tree. We consider a tree node, which is associated with a set of hubs $\mathcal{Y}^0 \coloneqq \{i \in \mathcal{H} \mid y_i = 0\}$ that must not be opened and a set of hubs $\mathcal{Y}^1 \coloneqq \{i \in \mathcal{H} \mid y_i = 1\}$ that must be opened (due to branching), as well as a set of generated paths $\tilde{\Lambda} = \cup_{p\in\mathcal{P}} \tilde{\Lambda}_p \subseteq \Lambda$ between all O-D pairs (obtained at termination of the column generation algorithm when solving the parent node). The (unrestricted) master problem to solve at the tree node is then given by the following LP:

$$
\begin{aligned}
\mathcal{L}_{path}(\mathcal{Y}^0, \mathcal{Y}^1, \Lambda): \quad & \min_{y,z} \quad \sum_{p\in\mathcal{P}} \sum_{\lambda\in\Lambda_p} d_p \cdot \tau_\lambda \cdot z_\lambda && \text{(5a)}\\
& \text{s.t.} \quad \sum_{i\in\mathcal{H}} y_i \le N, && \text{(5b)}\\
& \sum_{\lambda\in\Lambda_p} z_\lambda = k, \qquad \forall\, p \in \mathcal{P}, && \text{(5c)}\\
& \sum_{\{\lambda\in\Lambda_p \mid i\in\lambda\}} z_\lambda \le k \cdot y_i, \quad \forall\, p\in\mathcal{P},\ \forall\, i \in \mathcal{H}, && \text{(5d)}\\
& 0 \le z_\lambda \le 1, \qquad \forall\, \lambda \in \Lambda, && \text{(5e)}\\
& 0 \le y_i \le 1, \qquad \forall\, i \in \mathcal{H}, && \text{(5f)}\\
& y_i = 0, \qquad \forall\, i \in \mathcal{Y}^0, && \text{(5g)}\\
& y_i = 1, \qquad \forall\, i \in \mathcal{Y}^1. && \text{(5h)}
\end{aligned}
$$

We split our approach for solving $\mathcal{L}_{path}(\mathcal{Y}^0, \mathcal{Y}^1, \Lambda)$ into two cases. If $|\mathcal{Y}^1| = N$ (i.e., N specific hubs are selected to be opened), then $\mathcal{L}_{path}(\mathcal{Y}^0, \mathcal{Y}^1, \Lambda)$ boils down to finding the k shortest paths between each O-D pair in the subgraph induced by the opened hubs $\mathcal{Y}^1$, which we conduct using Yen's algorithm. Otherwise, we solve $\mathcal{L}_{path}(\mathcal{Y}^0, \mathcal{Y}^1, \Lambda)$ using column generation: Specifically, $\mathcal{L}_{path}(\mathcal{Y}^0, \mathcal{Y}^1, \tilde{\Lambda})$ (with the path selection variables restricted to $\tilde{\Lambda}$) is solved using an optimization solver; let $(\alpha'_p)_{p\in\mathcal{P}}, (\beta'_{pi})_{(p,i)\in\mathcal{P}\times\mathcal{H}}, (\gamma'_\lambda)_{\lambda\in\tilde{\Lambda}}$ denote the optimal dual variables of $\mathcal{L}_{path}(\mathcal{Y}^0, \mathcal{Y}^1, \tilde{\Lambda})$ associated with constraints (5c), (5d), (5e), respectively. Then, for any O-D pair $p \in \mathcal{P}$, the reduced cost associated with each path selection variable z_λ (for $\lambda \in \Lambda_p$) is given by $d_p \cdot \tau_\lambda - \alpha'_p - \sum_{i\in\mathcal{H}\cap\lambda} \beta'_{pi} - \gamma'_\lambda \cdot \mathbb{1}_{\{\lambda\in\tilde{\Lambda}_p\}}$. Note that we used the fact that constraint $z_\lambda \le 1$ is not binding for $\lambda \in \Lambda_p \setminus \tilde{\Lambda}_p$ and its corresponding dual variable is 0 by complementary slackness. Then, the pricing subproblem is given as follows:

$$
\min_{p\in\mathcal{P}} \min_{\lambda\in\Lambda_p} \left\{ d_p \cdot \tau_\lambda - \alpha'_p - \sum_{i\in\mathcal{H}\cap\lambda} \beta'_{pi} - \gamma'_\lambda \cdot \mathbb{1}_{\{\lambda\in\tilde{\Lambda}_p\}} \right\}. \tag{6}
$$

The challenge in solving this combinatorial problem comes from the dual variables $(\gamma'_\lambda)_{\lambda\in\tilde{\Lambda}}$. However for each O-D pair $p \in \mathcal{P}$, we observe that by complementary slackness and Constraint (5c), at most k dual variables in $(\gamma'_\lambda)_{\lambda\in\tilde{\Lambda}_p}$ are nonzero, and they are associated with upper bounds on variables z_λ included in solving the restricted master problem $\mathcal{L}_{path}(\mathcal{Y}^0, \mathcal{Y}^1, \tilde{\Lambda})$, for which the reduced costs are nonnegative. We then derive the following result:

Proposition 2. *For every O-D pair $p = (s,t) \in \mathcal{P}$, the lowest reduced cost associated with an $s-t$ path selection variable is given by:*

$$\min\left\{0, \min_{\lambda\in\Lambda_p\setminus\mathrm{supp}(\gamma')}\left\{d_p\cdot\tau_\lambda - \alpha'_p - \sum_{i\in\mathcal{H}\cap\lambda}\beta'_{pi}\right\}\right\}.$$

For each O-D pair $p = (s,t) \in \mathcal{P}$, we utilize Proposition 2 to derive an approach for computing the $s-t$ path with the lowest reduced cost. We create a graph $\mathcal{G}^p$ that is identical to $\mathcal{G}$ except that the length of each edge (i,j) in $\mathcal{G}^p$ is $d_p\cdot\tau_{i,j} - \beta'_{pj}\cdot\mathbb{1}_{\{j\in\mathcal{H}\}} - \alpha'_p\cdot\mathbb{1}_{\{j=t\}}$. In this new graph, the length of each $s-t$ path $\lambda\in\Lambda_p$ is given by $d_p\cdot\tau_\lambda - \alpha'_p - \sum_{i\in\mathcal{H}\cap\lambda}\beta'_{pi}$. Given $\kappa_p := |\,\mathrm{supp}((\gamma'_\lambda)_{\lambda\in\tilde{\Lambda}_p})| \le k$, we then adapt Yen's algorithm to compute up to (κ_p+1) shortest $s-t$ paths in $\mathcal{G}^p$ until we obtain the shortest $s-t$ path, denoted λ^p, that is not in the support of γ'. If its reduced cost $\bar{c}_p$ is nonpositive, then this is the lowest reduced cost among $s-t$ paths. On the other hand, it may occur that $\bar{c}_p$ is positive, in which case Proposition 2 guarantees that the lowest reduced cost among $s-t$ paths is 0. We repeat this procedure for each O-D pair and obtain Algorithm 4 that solves the pricing subproblem (6) in pseudo-polynomial time.

To accelerate Algorithm 4, we remove the hub locations in $\mathcal{Y}^0$ and connecting arcs when constructing the graph $\mathcal{G}^p$. We also note that α'_p is included in the reduced cost calculation after Yen's algorithm is run on $\mathcal{G}^p$, as the algorithm requires the length of each edge in $\mathcal{G}^p$ to be nonnegative.

If the reduced costs of every λ^p (for $p\in\mathcal{P}$) are nonnegative, then column generation terminates with an optimal solution to the master problem $\mathcal{L}_{path}(\mathcal{Y}^0, \mathcal{Y}^1, \Lambda)$. Otherwise, we add each λ^p (for $p\in\mathcal{P}$) with a negative reduced cost to $\tilde{\Lambda}$ and solve the new restricted master problem $\mathcal{L}_{path}(\mathcal{Y}^0, \mathcal{Y}^1, \tilde{\Lambda})$. In our numerical experiments, we find that adding at each iteration of column generation every λ^p with negative reduced cost leads to a better convergence, as opposed to only adding a single path selection variable with the lowest reduced cost. Our detailed implementation of column generation is described in Algorithm 5.

To ensure that $\mathcal{L}_{path}(\mathcal{Y}^0, \mathcal{Y}^1, \Lambda)$ is feasible at each tree node, we always open k additional dummy hubs to connect each O-D pair with k feasible paths (yet with very high lengths).

5.2. Branching Scheme

After solving $\mathcal{L}_{path}(\mathcal{Y}^0, \mathcal{Y}^1, \Lambda)$ using column generation and obtaining (continuous) hub selection variables $y\in[0,1]^{|\mathcal{H}|}$ at a tree node, the branch-and-price algorithm branches on one of possibly

Algorithm 4: Solving the Pricing Subproblem (6) (PSP($\mathcal{Y}^0, \alpha', \beta', \gamma'$))

Input : Set of closed hubs $\mathcal{Y}^0$, dual variable values $(\alpha', \beta', \gamma')$

Output : For every O-D pair $p=(s,t) \in \mathcal{P}$, an $s-t$ path λ^p and associated reduced cost $\bar{c}_p$

for every $p=(s,t) \in \mathcal{P}$ **do**
  $\mathcal{G}^p \leftarrow (\mathcal{S} \cup \mathcal{H} \cup \mathcal{T} \setminus \mathcal{Y}^0, \mathcal{A} \setminus \{(i,j) \in \mathcal{A} \mid i \text{ or } j \in \mathcal{Y}^0\})$;
  for every edge (i,j) in $\mathcal{G}^p$ **do**
    Set length of (i,j) in $\mathcal{G}^p$ to $d_p \cdot \tau_{i,j} - \beta'_{pj} \cdot \mathbb{1}_{\{j \in \mathcal{H}\}}$;
  $\ell \leftarrow 0$;
  do
    $\ell \leftarrow \ell + 1$;
    $\lambda^p_\ell \leftarrow \ell^{th}$ shortest $s-t$ path in $\mathcal{G}^p$ using Yen's algorithm;
  while $\lambda^p_\ell \notin \operatorname{supp}(\gamma')$;
  $\lambda^p \leftarrow \lambda^p_\ell$, $\bar{c}_p \leftarrow d_p \cdot \tau_{\lambda^p_\ell} - \alpha'_p - \sum_{i \in \mathcal{H} \cap \lambda^p_\ell} \beta'_{pi}$;
return $((\lambda^p)_{p \in \mathcal{P}}, (\bar{c}_p)_{p \in \mathcal{P}})$

Algorithm 5: Column Generation for Solving $\mathcal{L}_{path}(\mathcal{Y}^0, \mathcal{Y}^1, \Lambda)$ (CG($\mathcal{Y}^0, \mathcal{Y}^1, \tilde{\Lambda}$))

Input : Set of closed hubs $\mathcal{Y}^0$, set of opened hubs $\mathcal{Y}^1$, warm-starting path set $\tilde{\Lambda}$

Output : Optimal solution (y', z'), generated path set $\tilde{\Lambda}'$, and optimal value ξ'

$\tilde{\Lambda}' \leftarrow \tilde{\Lambda}$;
do
  Solve $\mathcal{L}_{path}(\mathcal{Y}^0, \mathcal{Y}^1, \tilde{\Lambda}')$: $(y', z') \leftarrow$ optimal primal solution, $(\alpha', \beta', \gamma') \leftarrow$ optimal dual variables associated with Constraints (5c), (5d), (5e), $\xi' \leftarrow$ optimal value;
  Solve the pricing subproblem (6): $((\lambda^p)_{p \in \mathcal{P}}, (\bar{c}_p)_{p \in \mathcal{P}}) \leftarrow$ PSP($\mathcal{Y}^0, \alpha', \beta', \gamma'$);
  for every $p \in \mathcal{P}$ **do**
    if $\bar{c}_p < 0$ **then**
      $\tilde{\Lambda}' \leftarrow \tilde{\Lambda}' \cup \{\lambda^p\}$;
while $\min_{p \in \mathcal{P}} \bar{c}_p < 0$;
return $(y', z', \tilde{\Lambda}', \xi')$

many fractional variables. To further fine-tune the algorithm, we derive a scheme that branches on a fractional variable that minimizes the demand-share-weighted total length of the k shortest paths between each O-D pair in the subgraph induced by $\{i \in \mathcal{H} \mid y_i = 1\} \cup \{i'\}$ (also including the above-mentioned dummy hubs for feasibility). This greedy procedure, described in Algorithm 6, reduces the overall running time of the branch-and-price algorithm compared to other branching rules (see Table EC.2 for a numerical comparison).

Algorithm 6: Greedy Hub Selection (GHS(y'))

Input : Hub selection variable values y'

Output : Hub $i^\dagger$ for branching

$\mathcal{K} \leftarrow \{i \in \mathcal{H} \mid y'_i = 1\}, \quad M \leftarrow +\infty;$

for every $i \in \mathcal{H}$ such that $0 < y'_i < 1$ **do**

    $D \leftarrow 0$;

    for every $p = (s,t) \in \mathcal{P}$ **do**

        $\delta_p \leftarrow$ total k-shortest $s-t$ path length in $\mathcal{G}_{\mathcal{K}\cup\{i\}}$ using Yen's Algorithm;

        $D \leftarrow D + d_p \cdot \delta_p$;

    if $D < M$ **then**

        $i^\dagger \leftarrow i, \qquad M \leftarrow D;$

return $i^\dagger$

5.3. Overall Algorithm

Our tailored implementation of branch-and-price for solving $\mathcal{I}_{path}$ is detailed in Algorithm 7.

To compute the optimality gap of the incumbent solution to $\mathcal{I}_{path}$, we associate each node of the branch-and-bound tree with a lower bound ξ on the optimal value of the corresponding master problem. This lower bound is computed from the optimal value of the master problem at the parent tree node.

We also note that line 14 in Algorithm 7 leverages the fact that at optimality of $\mathcal{L}_{path}(\mathcal{Y}^0, \mathcal{Y}^1, \Lambda)$, if the hub selection variables are integral, so are the path selection variables (from Lemma 1). Furthermore, when branching occurs and creates a child node with a new hub $i^\dagger$ that must be closed, line 18 removes the generated paths that go through $i^\dagger$, as they will not be selected in the master problem at that child node.

6. Case Study

This section presents the results from the computational experiments conducted using data instances from a large China-based parcel delivery partner to (i) validate the k-SPND model, (ii) test the scalability of the solution approaches developed in Sections 4 and 5, and (iii) analyze the efficiency and resilience of the designed networks when facing hub disruptions. All algorithms were implemented in Python v.3.8 and all optimization problems were solved using Gurobi v.9.1.1 on an Intel i7-8565 processor with a 4.60 GHz (1 core) CPU on a single thread with 12 GB of RAM.

Algorithm 7: Branch-and-Price for Solving $\mathcal{I}_{path}$

Input : Graph $\mathcal{G} = (\mathcal{S} \cup \mathcal{H} \cup \mathcal{T}, \mathcal{A})$, vector of demand shares $d \in [0,1]^{|\mathcal{P}|}$, number of hubs to open $N \in \mathbb{Z}_{>0}$, integer $k \in \mathbb{Z}_{>0}$, optimality gap $\epsilon \geq 0$

Output: Optimal subset of hubs to open $\mathcal{H}_o$ and k shortest paths between each O-D pair $(\lambda_\ell^p)_{(p,\ell)\in\mathcal{P}\times[k]}$

$L \leftarrow -\infty, \quad U \leftarrow +\infty, \quad Q \leftarrow ((\emptyset, \emptyset, \emptyset, -\infty))$;
while $U - L > \epsilon$ or $Q \neq \emptyset$ **do**
  $(\mathcal{Y}^0, \mathcal{Y}^1, \tilde{\Lambda}, \xi) \leftarrow$ Dequeue Q;
  if $|\mathcal{Y}^1| = N$ **then**
    for every $p = (s,t) \in \mathcal{P}$ **do**
      $(\lambda_1^{p'}, \ldots, \lambda_k^{p'}) \leftarrow k$ shortest $s-t$ paths in $\mathcal{G}_{\mathcal{Y}^1}$ computed using Yen's algorithm;
    if $\sum_{p\in\mathcal{P}} \sum_{\ell=1}^{k} d_p \cdot \tau_{\lambda_\ell^{p'}} < U$ **then**
      $\mathcal{H}_o \leftarrow \mathcal{Y}^1, \quad \lambda_\ell^p \leftarrow \lambda_\ell^{p'} \; \forall (p,\ell) \in \mathcal{P}\times[k], \quad U \leftarrow \sum_{p\in\mathcal{P}} \sum_{\ell=1}^{k} d_p \cdot \tau_{\lambda_\ell^{p'}}$;
  else
    Solve $\mathcal{L}_{path}(\mathcal{Y}^0, \mathcal{Y}^1, \Lambda)$ using column generation, warm-started with $\tilde{\Lambda}$;
      $(y', z', \tilde{\Lambda}', \xi') \leftarrow \mathrm{CG}(\mathcal{Y}^0, \mathcal{Y}^1, \tilde{\Lambda})$;
    if $\xi' < U$ **then**
      if $y' \in \{0,1\}^{|\mathcal{H}|}$ **then**
        $\mathcal{H}_o \leftarrow \mathrm{supp}(y'), \quad (\lambda_\ell^p)_{\ell\in[k]} \leftarrow \mathrm{supp}((z'_\lambda)_{\lambda\in\tilde{\Lambda}'\cap\Lambda_p}) \; \forall p \in \mathcal{P}, \quad U \leftarrow \xi'$;
      else
        $i^\dagger \leftarrow \mathrm{GHS}(y')$;
        Enqueue $(\mathcal{Y}^0, \mathcal{Y}^1 \cup \{i^\dagger\}, \tilde{\Lambda}', \xi')$ to Q;
        Enqueue $(\mathcal{Y}^0 \cup \{i^\dagger\}, \mathcal{Y}^1, \tilde{\Lambda}' \setminus \{\lambda \in \tilde{\Lambda}' \mid i^\dagger \in \lambda\}, \xi')$ to Q;
        $L \leftarrow \min\{\xi, \; \forall (\mathcal{Y}^0, \mathcal{Y}^1, \tilde{\Lambda}, \xi) \in Q\}$;
return $(\mathcal{H}_o, (\lambda_\ell^p)_{(p,\ell)\in\mathcal{P}\times[k]})$

6.1. Network Instances

Using the national level data of one of the largest parcel delivery companies in China that partnered with our research team, we created 6 representative problem instances of increasing size and complexity. These instances differ in the number of O-D pairs, hub candidate locations, and transportation legs with each instance representing middle-mile logistics operations spanning over a broad geographical region of Central China. In every instance, the parcel demand originates at one of the existing outbound logistics facilities of a city owned by the company ($\mathcal{S}$) and is destined for one of the company's existing last-mile delivery centers ($\mathcal{T}$). Based on the annual parcel demand data of the company, we computed the demand share d_p for each O-D pair $p \in \mathcal{P}$. As the company intends to implement relay transportation for its middle-mile operations, it identified a set $\mathcal{H}$ of candidate locations to open relay hubs, given by the company's existing intercity logistics hubs or major highway intersections. These relay hubs will serve as facilities where parcels are unloaded from incoming trucks, sorted, and then loaded into the outgoing trucks. We note that these facilities will not serve as fulfillment centers or locations where parcels are stored for a longer duration (e.g., more than a few hours).

To determine the set $\mathcal{A}$ of feasible transportation legs between facilities, we first computed the travel time τ_{ij} between every pair of locations $(i,j) \in (\mathcal{S} \cup \mathcal{T} \cup \mathcal{H})^2$ using the Haversine formula for the estimated traveled distance and using an average driving speed of 60 km/hr. Since the Chinese government imposes an 11-hour daily driving limit for truck drivers, we then only retained the transportation legs for which the drive time does not exceed 5.5 hours. This ensures that parcels travel towards their respective destinations while drivers return home daily.

Table 1 **Data instance characteristics**

Data instance	# Origins $\lvert\mathcal{S}\rvert$	# Destinations $\lvert\mathcal{T}\rvert$	# O-D pairs $\lvert\mathcal{P}\rvert$	# Candidates $\lvert\mathcal{H}\rvert$	# Edges $\lvert\mathcal{A}\rvert$
1	17	19	50	18	189
2	18	21	72	31	491
3	24	26	93	53	926
4	36	33	183	82	1284
5	36	33	576	87	1298
6	36	33	1091	90	1342

Table 1 describes the characteristics of all data instances used in this case study. We note that among these data instances, Instance 4 most closely represents the demand data of the parcel delivery company and requires the handling of 9,685,798 parcels annually. The networks designed for this instance would be able to serve 91.88% of the Chinese population, and would cover 93.81% Chinese inhabitable land with 92.17% of total Chinese GDP (Li et al. 2018). Figure 3 shows the locations of hub candidates, origins, and destinations for Instance 4.

We remark that designing relay networks for such instances is significantly complex: For the smallest instance, the number of potential paths between a single O-D pair reaches 10^7 and off-the-shelf optimization solvers cannot be used to directly solve $\mathcal{I}_{path}$. Next, we test our developed solution methodologies and showcase their computational performance in solving the large-scale k-SPND problem for these representative data instances.

6.2. Solution Methodology Performance

We run the solution approaches developed in Sections 4 and 5 to solve $\mathcal{I}_{path}$ for each data instance and with parameters ranging from 10 to 60 for the number of hubs N to open, and from 1 to

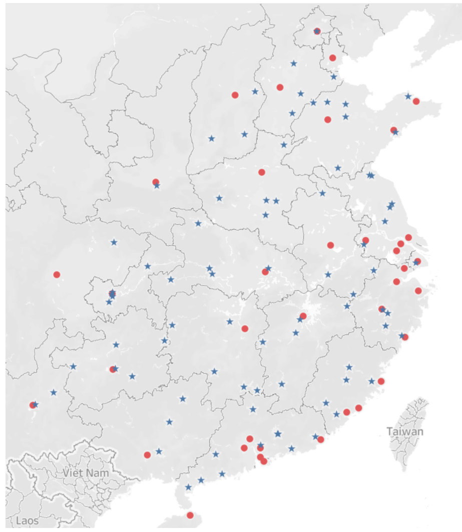

Figure 3 **Instance 4 with hub candidate locations (blue asterisks), and origins and/or destinations (red circles)**

8 for the number of shortest paths k. We find that for the smallest instances (#1 to #3), both approaches optimally solve $\mathcal{I}_{path}$. Figure 4 depicts the corresponding solution times.

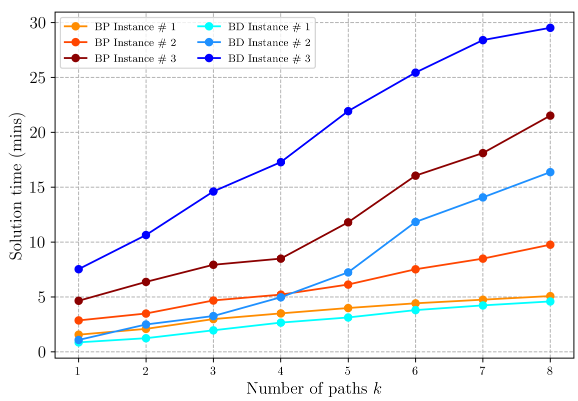

Figure 4 **Solution times for Benders decomposition (BD) and branch-and-price (BP) applied to Instance 1 with $N = 10$, Instance 2 with $N = 20$, and Instance 3 with $N = 30$.**

As expected, the solution times increase with k, which is in great part attributed to the solution approaches relying on repeatedly computing k shortest paths between each O-D pair. We also observe that for smaller instances and lower values of k, Benders decomposition converges faster than branch-and-price.

Next, we compare in Table 2 the optimality gaps of the solutions to $\mathcal{I}_{path}$ provided by both approaches after 12 hours of execution on the data instances with varied parameters. To determine the optimality gap of each solution, we computed the percentage difference between its objective value and the best lower bound found by *either* of the solution approaches.

Table 2 Optimality gaps from the solution approaches with a 12-hour time limit. Italic numbers represent instances for which optimality gaps are computed from the lower bound provided by the other solution approach.

Solution approach	Data instance	# Hubs N	Optimality gap (%) $k=1$	$k=2$	$k=3$	$k=4$	$k=5$	$k=6$	$k=7$	$k=8$
Benders decomposition (Algorithm 3)	1	10	0	0	0	0	0	0	0	0
	2	20	0	0	0	0	0	0	0	0
	3	30	0	0	0	0	0	0	0	0
	4	40	0	0	0	0	*1.23*	*1.58*	*2.76*	*3.92*
	5	50	*4.38*	*4.71*	*4.96*	*5.08*	*5.22*	*7.48*	*8.62*	*12.76*
	6	60	*13.64*	*14.07*	*15.09*	*15.86*	*16.57*	*18.56*	*20.47*	*23.77*
Branch-and-price (Algorithm 7)	1	10	0	0	0	0	0	0	0	0
	2	20	0	0	0	0	0	0	0	0
	3	30	0	0	0	0	0	0	0	0
	4	40	0	0	0	0	0	0	0	0
	5	50	0	0	0	0	0	1.59	1.74	2.08
	6	60	2.52	3.68	4.26	5.91	7.05	7.88	9.67	11.75

We observe that despite the large-scale nature of the k-SPND problem, our solution approaches are capable of solving $\mathcal{I}_{path}$ to optimality for most of the instances. In particular, our tailored branch-and-price algorithm reaches optimality up to Instance 5 with $N = 50$ and $k = 5$. We recall that the number of O-D pairs in Instance 5 is 3 times larger than in our partner's demand data, represented by Instance 4. A closer inspection of Table 2 reveals that Benders decomposition is more sensitive to the number of hubs to open N, but is still able to reach optimality for Instance 4 with $N = 40$ and $k = 4$. As a result, we find that Benders converges faster for smaller instances and branch-and-price obtains better solutions for larger instances.

Overall, Figure 4 and Table 2 show the effectiveness of our tailored solution methods in solving the k-SPND problem and in designing relay networks for large-scale logistics operations. Next, we analyze the designed networks and quantify their capabilities to sustain disruptions.

6.3. Network Efficiency Assessment

To evaluate the efficiency of any given relay network with a set of opened hubs and transportation legs, we determine a minimum-cost consolidation plan to transport parcels from their respective outbound facilities to their corresponding last-mile delivery centers over a planning horizon of one week. We adapt a classical service network design model (Bakir et al. 2021) to select the number of delivery trips (termed as services) along each transportation leg over the planning horizon to satisfy the average weekly parcel demand. The objective is to minimize costs given by parcel handling, truck scheduling, and fuel consumption. See Section EC.2 for the detailed optimization model and parameter values. We note that the consolidation plan will leverage the backhauling opportunities offered by the back-and-forth trips in the relay network to efficiently transport parcels that may travel in opposite directions.

To validate the proposed k shortest paths relay logistics networks, we compare their performance against relay logistics networks constructed with only cost considerations to support parcel delivery.

Specifically, given the graph $\mathcal{G}$ of potential network designs and a number N of hubs to open, we construct an *efficiency-optimized network*, obtained by selecting up to N hubs to open to minimize the cost of the consolidation model described above (see Section EC.3 for the detailed optimization model). This provides a baseline against which we can assess the capabilities of our proposed networks.

First, we compare in Table 3 the efficiency of the k shortest paths networks with that of the efficiency-optimized (E-O) networks under nominal conditions, i.e., in the absence of disruptions.

Table 3 **Networks' average delivery costs ($/parcel) in the absence of disruptions**

Data instance	# Hubs N	E-O network	k shortest paths networks				
			$k=1$	$k=2$	$k=3$	$k=4$	$k=5$
1	10	2.638 (0.00%)	2.798 (6.04%)	2.881 (9.19%)	2.980 (12.94%)	2.983 (13.05%)	2.983 (13.05%)
2	20	2.551 (0.00%)	2.643 (3.59%)	2.671 (4.69%)	2.737 (7.29%)	2.738 (7.31%)	2.738 (7.31%)
3	30	2.288 (0.00%)	2.362 (3.23%)	2.386 (4.28%)	2.426 (6.03%)	2.426 (6.04%)	2.426 (6.04%)
4	40	2.270 (0.00%)	2.316 (2.03%)	2.330 (2.64%)	2.359 (3.92%)	2.360 (3.96%)	2.360 (3.96%)
5	50	2.057 (0.00%)	2.083 (1.26%)	2.087 (1.46%)	2.091 (1.65%)	2.092 (1.70%)	2.092 (1.70%)
6	60	1.890 (0.00%)	1.912 (1.16%)	1.913 (1.21%)	1.915 (1.32%)	1.915 (1.32%)	1.915 (1.32%)

While the efficiency-optimized networks expectedly achieve the lowest average parcel delivery costs, we observe that the k shortest paths networks with $k=1$ lead to similar costs, more particularly as the problem size increases. Indeed for the largest two instances, the k shortest paths networks with $k=1$ achieve delivery costs within $\sim 1\%$ of those of efficiency-optimized networks: As the number of O-D pairs and the total parcel volume increase, the minimum-cost consolidation plan transports parcels along the shortest paths while still exploiting the backhauling opportunities from the back and forth trips. This validates the use of shortest path lengths as a surrogate for the operational costs in the objective function of the network design problem.

From Table 3, we also observe that as k increases, the average parcel delivery costs in the k shortest paths networks increase as well. Enforcing each O-D pair to be connected with k paths further constrains the network design, which in turn leads to a decrease in nominal efficiency. Nonetheless, as the instance size increases, we observe that the impact of considering higher values of k on average delivery costs becomes marginal (e.g., they increase on average by \$0.003 between $k=1$ and $k=5$ for Instance 6). This can be explained by the k-SPND model smartly leveraging the higher density of the network to connect each O-D pair with a higher number of paths while maintaining short path lengths. Next, we study the impact of disruptions on the logistics networks.

6.4. Network Resilience Assessment

We conduct a set of experiments, where we subject the networks to random hub disruptions. In each disruption scenario, occurring uniformly at random, we suppose a relay hub becomes dysfunctional and no parcel can be routed through it during the planning horizon. We consider week-long disruptions that can be observed by the service provider before determining a consolidation plan (e.g., after an extreme event such as a hurricane or wildfire that severely impacted a hub). For each network and disruption scenario, we determine a minimum-cost consolidation plan (using (EC.1)). We assume that if an O-D pair becomes disconnected in the relay network as a result of a disruption, the demand for that O-D pair cannot be fulfilled during the planning horizon.

Table 4 compares the networks' performances using two quantifiable metrics: The first metric computes the expected percentage of demand that cannot be fulfilled through the relay network under a disruption scenario. For the demand that can be fulfilled through the relay network, the second metric computes the expected delivery cost, averaged over the parcel demand of the O-D pairs that remain connected under a disruption scenario.

Table 4 Network resilience comparison, averaged across uniform hub disruptions

Degradation metric	Data instance	# Hubs N	E-O network	k shortest paths networks				
				$k=1$	$k=2$	$k=3$	$k=4$	$k=5$
Unfulfilled demand (%)	1	10	13.86	13.24	11.63	10.25	9.43	9.43
	2	20	10.78	10.90	10.24	9.28	8.58	8.58
	3	30	7.39	7.00	6.93	6.11	5.99	5.99
	4	40	6.08	6.13	5.84	5.46	5.12	5.12
	5	50	5.17	5.20	5.01	4.77	4.39	4.39
	6	60	4.92	4.12	4.02	3.98	3.00	3.00
Average delivery costs ($/parcel)	1	10	3.835	3.726	3.716	3.596	3.521	3.521
	2	20	3.389	3.375	3.372	3.370	3.317	3.317
	3	30	2.920	2.891	2.885	2.877	2.868	2.868
	4	40	2.762	2.732	2.722	2.711	2.691	2.691
	5	50	2.488	2.360	2.355	2.355	2.353	2.353
	6	60	2.193	2.128	2.126	2.125	2.122	2.122

The results show that for every data instance, our proposed k shortest paths networks with $k \geq 2$ outperform the efficiency-optimized networks when facing hub disruptions with respect to both performance metrics: When $k = 2$ (respectively, $k = 4$), the expected amount of unfulfilled demand between efficiency-optimized networks and k shortest paths networks decreases by 8.77% (respectively, 23.54%) on average across all data instances. In addition to guaranteeing the delivery of a higher proportion of parcel demand through short-haul transportation, our networks also achieve lower average delivery costs per parcel as compared to efficiency-optimized networks under disruptions. Figures 5 and 6 further demonstrate that besides improving the performance metrics in expectation, our proposed networks also reduce their variances and their maximum values, as k increases.

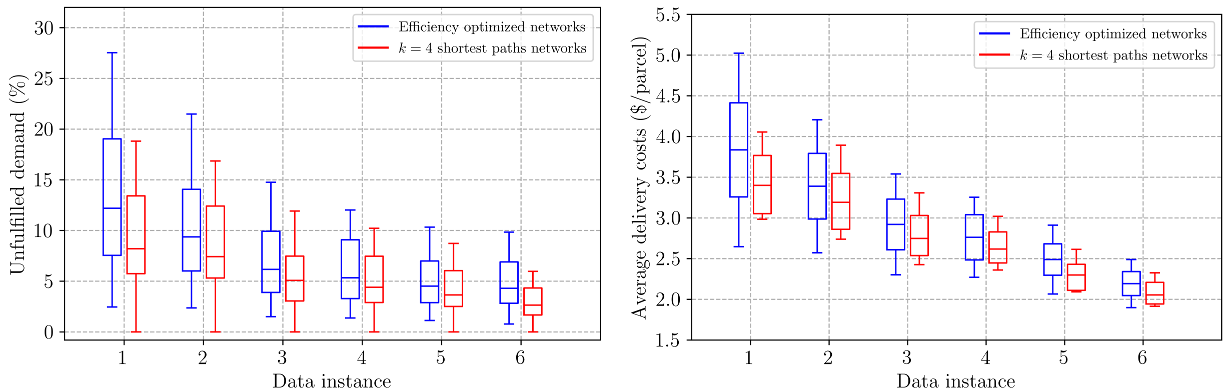

Figure 5 **Performance metrics comparison under uniform hub disruptions.**

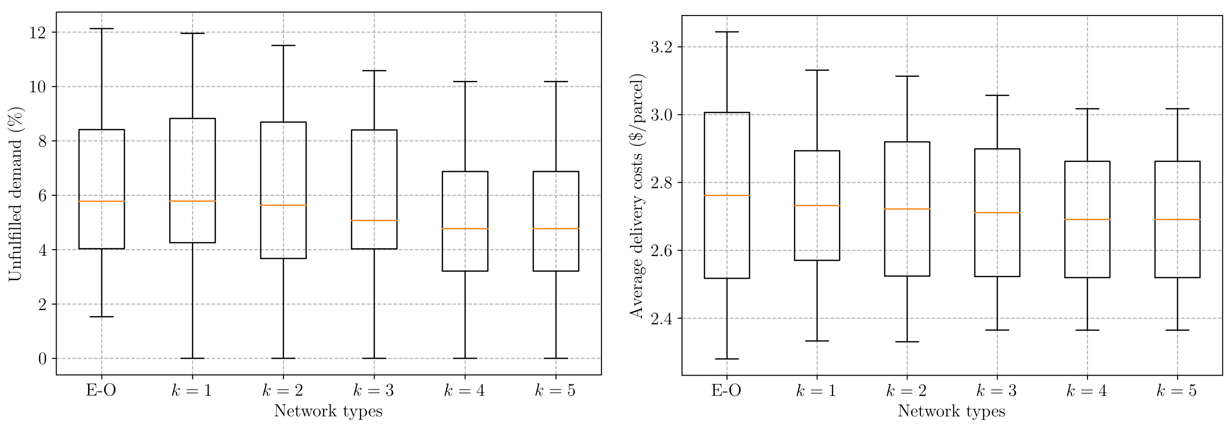

Figure 6 **Performance metrics comparison for Instance #4 under uniform hub disruptions.**

This numerical analysis shows that by ensuring the existence of an increased number of paths k between each O-D pair, demand can more likely be fulfilled through the relay network when facing disruptions. Furthermore, by minimizing the demand-weighted total length of the k shortest paths, the designed networks are more likely to contain more paths of short lengths during a disruption, which provides better consolidation opportunities and results in a limited increase in parcel delivery costs in comparison with nominal situations.

For the sake of completeness, we also conduct in Section EC.5 a resilience study under uniform transportation leg disruptions to capture instances of geographical disruptions due to severe natural events such as hurricanes or wildfires. The results demonstrate similar insights to the hub disruption study, with lower magnitude of performance degradation. Finally, we conduct a sensitivity analysis on the resilience capabilities of our proposed networks for Instance 4 under uniform hub disruptions as both k and N vary, and illustrate the results in Figure 7 below.

We observe that as the number of hubs N to open increases, the capability of the networks to sustain disruptions improves, due to their increased density. Interestingly, Figure 7 shows that the

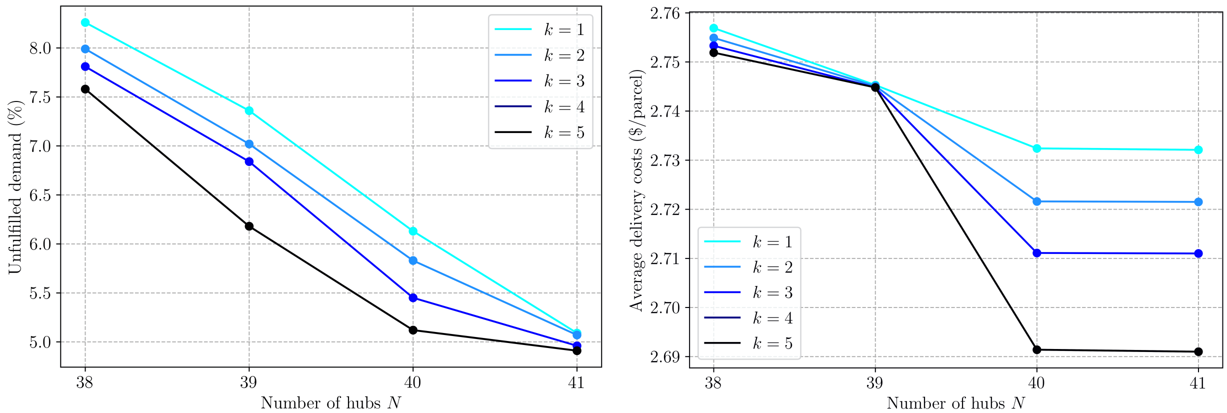

Figure 7 **Resilience comparison of proposed networks for Instance 4, averaged across uniform hub disruptions**

improvement in average delivery costs for the O-D pairs that remain connected during disruptions as k increases is more significant when the number of opened hubs N is higher. Indeed, designing a k shortest paths network with $k = 5$ in comparison with $k = 1$ leads to a reduction in average delivery costs for connected O-D pairs of \$0.005 per parcel when $N = 38$, and \$0.041 per parcel when $N = 40$. This shows how the k-SPND model exploits additional hubs to create more dense and efficient relay networks, providing an increased number of short paths to improve consolidation during disruptions.

These numerical results validate the use of the k-SPND model to design relay logistics networks that provide significant improvement in resilience to disruptions with marginal impact on nominal efficiency. In particular, our sensitivity analysis can be utilized by service providers to design relay logistics networks that achieve a desired tradeoff between nominal efficiency and performance under disruptions.

7. Conclusion

In this article, we studied the problem of designing large-scale resilient relay logistics hub networks with limited information regarding future demand and disruption risks. We introduced a model that focuses on improving efficiency and resilience through the topology of the network configuration. This model, for *k-Shortest Path Network Design*, consists of locating relay logistics hubs to connect each origin-destination pair with k paths of minimum lengths, weighted by their forecasted demand share. We formulated this problem as a mixed-integer program with exponential path selection variables and constraints, and leveraged its structure to design two scalable solution methodologies based on tailored implementations of Benders decomposition and branch-and-price, respectively. In the first solution methodology, we provided an analytical characterization of the optimal solutions of the exponential-sized dual subproblem used to generate Benders feedback cuts. We then derived a pseudo-polynomial time approach to generate these cuts, based on Yen's algorithm and shortest-path subroutines. In the second solution methodology, we used complementary slackness to prove that the pricing subproblem in the column generation algorithm employed as part of the branch-and-price algorithm can be also solved in pseudo-polynomial time by computing multiple shortest paths in a carefully adjusted graph. We also developed a specific branching scheme to further accelerate the branch-and-price algorithm.

We then conducted extensive computational experiments to design large-scale resilient relay logistics networks using data from a large China-based parcel delivery partner. We found that our developed algorithms can obtain an optimal or near-optimal solution in a reasonable time, with our tailored implementation of Benders decomposition converging faster for smaller instances and our tailored branch-and-price scaling better as the number of hubs N increases. To validate our designed networks, we computed their performance by determining minimum-cost consolidation plans to satisfy average weekly parcel demand in nominal situations and when facing random hub disruptions. For comparison purposes, we also created baseline networks, designed by solely optimizing the cost of delivering the parcel demand in nominal situations. Overall, we observed that our designed networks significantly outperform the efficiency-optimized networks when facing disruptions, as more demand can be fulfilled by short-haul transportation and at a lower cost. Furthermore, our sensitivity analysis quantified the impact on resilience capabilities of our k shortest paths networks as the numbers of paths k and hubs to open N jointly vary. Finally, we observed that our designed networks achieved these resilience capabilities with only limited compromise in efficiency in the absence of disruptions in comparison with the efficiency-optimized networks. This analysis validates the use of the k-Shortest Path Network Design problem by service providers to design efficient and resilient relay logistics hub networks.

While this work has focused on designing networks that can sustain random and independent disruptions, a natural extension is to design networks that can also sustain correlated (or even worst-case) disruptions caused by extreme events (e.g., pandemics). This will require evaluating the impact of such disruptions on various network configurations in order to guide the design of appropriate logistics networks. Another extension is to incorporate factors such as hub facility sizes, demand variability and uncertainty, consolidation opportunities, and service level requirements at the network design stage. This will result in complex multi-stage optimization problems, which will require new heuristics and approximation algorithms to provide practically relevant solutions. Finally, it would be worthwhile to study how the models and solution techniques proposed in this work can be extended to design multi-layered hyperconnected logistics networks for fast (e.g., same-day), efficient, and resilient delivery service.

Acknowledgments

We are grateful to Alan Erera and Alejandro Toriello for their valuable feedback.

References

Ali TH, Radhakrishnan S, Pulat S, Gaddipati NC (2002) Relay network design in freight transportation systems. *Transportation Research Part E: Logistics and Transportation Review* 38(6):405–422.

Alumur SA, Campbell JF, Contreras I, Kara BY, Marianov V, O'Kelly ME (2021) Perspectives on modeling hub location problems. *European Journal of Operational Research* 291(1):1–17.

An Y, Zhang Y, Zeng B (2015) The reliable hub-and-spoke design problem: Models and algorithms. *Transportation Research Part B: Methodological* 77:103–122.

Ang E, Iancu DA, Swinney R (2017) Disruption risk and optimal sourcing in multitier supply networks. *Management Science* 63(8):2397–2419.

Azizi N (2017) Managing facility disruption in hub-and-spoke networks: formulations and efficient solution methods. *Annals of Operations Research* 272:159–185.

Bakir I, Erera A, Savelsbergh M (2021) Motor carrier service network design. Crainic TG, Gendreau M, Gendron B, eds., *Network Design with Applications to Transportation and Logistics*, 427–467 (Springer International Publishing).

Ballot E, Gobet O, Montreuil B (2012) *Physical Internet Enabled Open Hub Network Design for Distributed Networked Operations*, 279–292 (Berlin, Heidelberg: Springer Berlin Heidelberg).

Basciftci B, Ahmed S, Shen S (2021) Distributionally robust facility location problem under decision-dependent stochastic demand. *European Journal of Operational Research* 292(2):548–561.

Cabral EA, Erkut E, Laporte G, Patterson RA (2007) The network design problem with relays. *European Journal of Operational Research* 180(2):834–844.

Campbell J, O'Kelly M (2013) Twenty-five years of hub location research. *Transportation Science* 46:153–169.

Chen D, Sun D, Yin Y, Dhamotharan L, Kumar A, Guo Y (2022) The resilience of logistics network against node failures. *International Journal of Production Economics* 244:108373.

Contreras I, Cordeau JF, Laporte G (2011) Stochastic uncapacitated hub location. *European Journal of Operational Research* 212(3):518–528.

Cordeau JF, Pasin F, Solomon M (2006) An integrated model for logistics network design. *Annals OR* 144:59–82.

Craighead CW, Blackhurst J, Rungtusanatham MJ, Handfield RB (2007) The severity of supply chain disruptions: Design characteristics and mitigation capabilities. *Decision Sciences* 38(1):131–156.

Ghaffarinasab N, Atayi R (2018) An implicit enumeration algorithm for the hub interdiction median problem with fortification. *European Journal of Operational Research* 267(1):23–39.

Greening LM, Dahan M, Erera AL (2023) Lead-time-constrained middle-mile consolidation network design with fixed origins and destinations. *Transportation Research Part B: Methodological* 174:102782.

Grove PG, O'Kelly ME (1986) Hub networks and simulated schedule delay. *Papers in Regional Science* 59(1):103–119.

Hall RW (1989) Configuration of an overnight package air network. *Transportation Research Part A: General* 23(2):139–149.

Hirsch J (2020) Industry trends, driver preference pave way for hub-to-hub routes. Transport Topics, https://www.ttnews.com/articles/industry-trends-driver-preference-pave-way-hub-hub-routes.

Hu Z, Askin RG, Hu G (2019) Hub relay network design for daily driver routes. *International Journal of Production Research* 57(19):6130–6145.

Hunt GW (1998) *Transportation Relay Network Design*. Ph.D. thesis, Georgia Institute of Technology.

IBISWorld (2023) Global courier & delivery services - market size 2005–2029. Technical report, IBISWorld, https://www.ibisworld.com/global/market-size/global-courier-delivery-services/.

Ip WH, Wang D (2009) Resilience evaluation approach of transportation networks. *Proceedings of the 2009 International Joint Conference on Computational Sciences and Optimization, CSO 2009* 2:618–622.

Ip WH, Wang D (2011) Resilience and friability of transportation networks: Evaluation, analysis and optimization. *IEEE Systems Journal* 5:189–198.

IRU (2022) Driver shortage global report 2022. Technical report, IRU, https://www.iru.org/resources/iru-library/driver-shortage-global-report-2022.

Jacquillat A, Schmid A, Wang K (2022) Relay logistics: A multi-variable generation approach. *SSRN Electronic Journal* URL http://dx.doi.org/10.2139/ssrn.4241031.

Kewcharoenwong P, Li Q, Üster H (2023) Lagrangean relaxation algorithms for fixed-charge capacitated relay network design. *Omega* 121:102926.

Kewcharoenwong P, Üster H (2014) Benders decomposition algorithms for the fixed-charge relay network design in telecommunications. *Telecommunication Systems* 56:441–453.

Kim H (2008) *Reliable p-hub location problems and protection models for hub network design*. Ph.D. thesis, The Ohio State University.

Konak A (2012) Network design problem with relays: A genetic algorithm with a path-based crossover and a set covering formulation. *European Journal of Operational Research* 218(3):829–837.

Kouvelis P, Dong L, Boyabatli O, Li R (2011) *Handbook of integrated risk management in global supply chains* (John Wiley & Sons).

Kulkarni O, Cohen YM, Dahan M, Montreuil B (2021) Resilient hyperconnected logistics hub network design. *8th International Physical Internet Conference*.

Kulkarni O, Dahan M, Montreuil B (2022) Resilient hyperconnected parcel delivery network design under disruption risks. *International Journal of Production Economics* 251:108499.

Kulkarni O, Dahan M, Montreuil B (2023) Resilience assessment of hyperconnected parcel logistic networks under worst-case disruptions. *9th International Physical Internet Conference*.

Kulturel-Konak S, Konak A (2008) A local search hybrid genetic algorithm approach to the network design problem with relay stations. Raghavan S, Golden B, Wasil E, eds., *Telecommunications Modeling, Policy, and Technology*, 311–324 (Boston, MA: Springer US).

Leitner M, Ljubić I, Riedler M, Ruthmair M (2019) Exact approaches for network design problems with relays. *INFORMS Journal on Computing* 31(1):171–192.

Leitner M, Ljubić I, Riedler M, Ruthmair M (2020) Exact approaches for the directed network design problem with relays. *Omega* 91:102005.

Li M, He B, Guo R, Li Y, Chen Y, Fan Y (2018) Study on population distribution pattern at the county level of china. *Sustainability (Switzerland)* 10.

Mohammadi M, Jula P, Tavakkoli-Moghaddam R (2019) Reliable single-allocation hub location problem with disruptions. *Transportation Research Part E: Logistics and Transportation Review* 123:90–120.

Mohammadi M, Tavakkoli-Moghaddam R, Siadat A, Rahimi Y (2016) A game-based meta-heuristic for a fuzzy bi-objective reliable hub location problem. *Engineering Applications of Artificial Intelligence* 50:1–19.

Montreuil B (2011) Toward a physical internet : meeting the global logistics sustainability grand challenge. *Logistics Research* 3:71–87.

Montreuil B, Buckley SM, Faugère L, Khir R, Derhami S (2018) Urban parcel logistics hub and network design: The impact of modularity and hyperconnectivity. *Progress in Material Handling Research*.

Morris C (2022) Warning shipping delay problems to continue this year. https://www.bbc.com/news/business-60131947.

Nosowitz D (2017) The long white line: The mental and physical effects of long-haul trucking. Pacific Standard, http://tinyurl.com/ypsp3vvf.

Osei-Asamoah A, Lownes NE (2014) Complex network method of evaluating resilience in surface transportation networks. *Transportation Research Record* 2467:120–128.

Peng P, Snyder LV, Lim A, Liu Z (2011) Reliable logistics networks design with facility disruptions. *Transportation Research Part B: Methodological* 45(8):1190–1211, supply chain disruption and risk management.

Placek M (2022) Global parcel shipping volume between 2013 and 2027. Technical report, Statista, https://www.statista.com/statistics/1139910/parcel-shipping-volume-worldwide/.

Shahabi M, Unnikrishnan A (2014) Robust hub network design problem. *Transportation Research Part E: Logistics and Transportation Review* 70:356–373.

Shen ZJM, Zhan RL, Zhang J (2011) The reliable facility location problem: Formulations, heuristics, and approximation algorithms. *INFORMS Journal on Computing* 23(3):470–482.

Sieber K (2015) Long-haul truck driver health survey results. NIOSH Science Blog, CDC, http://tinyurl.com/ypsp3vvf.

Taherkhani G, Alumur SA, Hosseini M (2020) Benders decomposition for the profit maximizing capacitated hub location problem with multiple demand classes. *Transportation Science* 54(6):1446–1470.

Tu D, Montreuil B (2019) Hyper-connected megacity logistics: Multi-tier territory clustering and multi-plane meshed hub network design. *6th International Physical Internet Conference*, 159–167.

US Bureau of Labor Statistics (2023) Occupational outlook handbook, heavy and tractor-trailer truck drivers. Technical report, U.S. Bureau of Labor Statistics, https://www.bls.gov/ooh/transportation-and-material-moving/heavy-and-tractor-trailer-truck-drivers.

htm.

Üster H, Kewcharoenwong P (2011) Strategic design and analysis of a relay network in truckload transportation. *Transportation Science* 45(4):505–523.

Yen JY (1971) Finding the k shortest loopless paths in a network. *Management Science* 17:712–716.

Yıldız B, Karaşan OE, Yaman H (2018) Branch-and-price approaches for the network design problem with relays. *Computers & Operations Research* 92:155–169.

Zetina CA, Contreras I, Cordeau JF, Nikbakhsh E (2017) Robust uncapacitated hub location. *Transportation Research Part B: Methodological* 106:393–410.

Ziaeifar A, Üster H (2023) Relay network design with direct shipment and multi-relay assignment. *Annals of Operations Research* .

Üster H, Maheshwari N (2007) Strategic network design for multi-zone truckload shipments. *IIE Transactions* 39(2):177–189.

Proofs and Additional Details

EC.1. Proofs of Statements

Proof of Lemma 1. Given $y \in \{0,1\}^{|\mathcal{H}|}$, the feasible region of the path selection variables for each O-D pair $p \in \mathcal{P}$ is given by

$$\begin{cases} \sum_{\lambda \in \Lambda_p} z_\lambda = k, \\ 0 \le z_\lambda \le 1, \quad \forall \lambda \in \Lambda_p, \\ z_\lambda = 0, \qquad \forall \lambda \in \Lambda_p \mid \lambda \cap \{i \in \mathcal{H} \mid y_i = 0\} \ne \emptyset, \end{cases}$$

which is an integral polyhedron. Therefore, there exists an optimal solution (y^*, z^*) of $\mathcal{I}_{path}$ such that $z^* \in \{0,1\}^{|\Lambda|}$, and $\mathcal{I}_{path}$ is indeed a MIP formulation of the k-SPND problem (1). □

Proof of Proposition 1. Consider a vector of hub selection variables $y^* \in \{0,1\}^{|\mathcal{H}|}$ and an O-D pair $p = (s,t) \in \mathcal{P}$. Let $\mathcal{H}_o = \text{supp}(y^*)$ be the set of opened hubs, $\tau_p^\dagger$ be the shortest $s-t$ path length in the original graph $\mathcal{G}$, and $\lambda_1^p, \dots, \lambda_k^p \in \Lambda_p$ be the k shortest $s-t$ paths in $\mathcal{G}_{\mathcal{H}_o}$ ordered so that $\tau_{\lambda_1^p} \le \dots \le \tau_{\lambda_k^p}$. Then, the optimal value of the subproblem (3) is given by $\sum_{\ell=1}^k \tau_{\lambda_\ell^p}$.

Let $w \in \mathbb{R}_{\ge 0}^{|\mathcal{H} \setminus \mathcal{H}_o|}$ be a vector satisfying $\sum_{i \in \lambda \cap \mathcal{H} \setminus \mathcal{H}_o} w_i \ge \mathbb{1}_{\{\lambda \cap \mathcal{H} \setminus \mathcal{H}_o \ne \emptyset\}}$ for every $s-t$ path $\lambda \in \Lambda_p$. Then, we consider the following solution to the dual subproblem (3):

$$\alpha^* = \tau_{\lambda_k^p}, \qquad \beta_i^* = \begin{cases} w_i \cdot (\tau_{\lambda_k^p} - \tau_p^\dagger) & \text{if } i \in \mathcal{H} \setminus \mathcal{H}_o, \\ 0 & \text{otherwise,} \end{cases} \qquad \gamma_\lambda^* = \begin{cases} \tau_{\lambda_k^p} - \tau_\lambda & \text{if } \lambda \in \{\lambda_1^p, \dots, \lambda_k^p\}, \\ 0 & \text{otherwise.} \end{cases}$$

First, we verify that $(\alpha^*, \beta^*, \gamma^*)$ is a dual feasible solution: Indeed, $\beta_i^* \ge 0$ for every $i \in \mathcal{H}$ since $w_i \ge 0$ and $\tau_p^\dagger \le \tau_\lambda$ for every $\lambda \in \Lambda_p$. Similarly, $\gamma_\lambda^* \ge 0$ for every $\lambda \in \Lambda_p$ since for every $\ell \in [k]$, $\tau_{\lambda_\ell^p} \le \tau_{\lambda_k^p}$. Furthermore,

$$\forall \ell \in [k], \ \alpha^* - \sum_{i \in \mathcal{H} \cap \lambda_\ell^p} \beta_i^* - \gamma_{\lambda_\ell^p}^* = \tau_{\lambda_k^p} + \tau_{\lambda_\ell^p} - \tau_{\lambda_k^p} = \tau_{\lambda_\ell^p},$$

where we used the fact that for every $\ell \in [k]$, λ_ℓ^p only traverses opened hubs $i \in \mathcal{H}_o$ (for which $\beta_i^* = 0$).

Next, we consider an $s-t$ path $\lambda \in \Lambda_p \setminus \{\lambda_1^p, \dots, \lambda_k^p\}$. If λ only traverses opened hubs, then necessarily $\tau_\lambda \ge \tau_{\lambda_k^p}$ by optimality of the k shortest $s-t$ paths $\lambda_1^p, \dots, \lambda_k^p$. In this case,

$$\alpha^* - \sum_{i \in \mathcal{H} \cap \lambda} \beta_i^* - \gamma_\lambda^* = \tau_{\lambda_k^p} \le \tau_\lambda.$$

On the other hand, if λ traverses at least one unopened hub, then:

$$\alpha^* - \sum_{i \in \mathcal{H} \cap \lambda} \beta_i^* - \gamma_\lambda^* = \tau_{\lambda_k^p} - (\tau_{\lambda_k^p} - \tau_p^\dagger) \cdot \sum_{i \in \lambda \cap \mathcal{H} \setminus \mathcal{H}_o} w_i \le \tau_p^\dagger \le \tau_\lambda.$$

Therefore, $(\alpha^*, \beta^*, \gamma^*)$ is a dual feasible solution. The corresponding objective value is given by:

$$k \cdot \alpha^* - \sum_{i \in \mathcal{H}_o} k \cdot \beta_i^* - \sum_{\lambda \in \Lambda_p} \gamma_\lambda^* = k \cdot \tau_{\lambda_k^p} - \sum_{\ell=1}^{k} (\tau_{\lambda_k^p} - \tau_{\lambda_\ell^p}) = \sum_{\ell=1}^{k} \tau_{\lambda_\ell^p}.$$

By weak duality, we conclude that $(\alpha^*, \beta^*, \gamma^*)$ is an optimal dual solution of the subproblem (3). Therefore, a valid cut for the master problem is given as follows:

$$\theta_p \geq \sum_{\ell=1}^{k} \tau_{\lambda_\ell^p} - \sum_{i \in \mathcal{H} \setminus \mathcal{H}_o} k \cdot w_i \cdot (\tau_{\lambda_k^p} - \tau_p^\dagger) \cdot y_i,$$

with $w \in \mathbb{R}_{\geq 0}^{|\mathcal{H} \setminus \mathcal{H}_o|}$ satisfying $\sum_{i \in \lambda \cap \mathcal{H} \setminus \mathcal{H}_o} w_i \geq \mathbb{1}_{\{\lambda \cap \mathcal{H} \setminus \mathcal{H}_o \neq \emptyset\}}$ for every $s-t$ path $\lambda \in \Lambda_p$. □

Proof of Proposition 2. Consider an O-D pair $p = (s,t) \in \mathcal{P}$, and let $\lambda \in \Lambda_p$. If $\lambda \in \text{supp}(\gamma')$, then by complementary slackness, $z_\lambda = 1 > 0$ at optimality of $\mathcal{L}_{path}(\mathcal{Y}^0, \mathcal{Y}^1, \tilde{\Lambda})$. As a consequence, its reduced cost is 0. If on the other hand $\lambda \notin \text{supp}(\gamma')$, then the reduced cost of z_λ is given by $d_p \cdot \tau_\lambda - \alpha'_p - \sum_{i \in \mathcal{H} \cap \lambda} \beta'_{pi}$. Thus, the lowest reduced cost associated with the $s-t$ paths (for O-D pair p) is given by:

$$\min \left\{ 0, \min_{\lambda \in \Lambda_p \setminus \text{supp}(\gamma')} \left\{ d_p \cdot \tau_\lambda - \alpha'_p - \sum_{i \in \mathcal{H} \cap \lambda} \beta'_{pi} \right\} \right\}.$$

□

EC.2. Network Consolidation Model

To evaluate the performance of any given relay network in our case study, we develop an optimization model that generates a minimum-cost consolidation plan to transport parcels from their respective origins to their corresponding destinations. This model generates a tactical plan for the trips along each transportation leg to fulfill the parcel demand over a pre-specified time horizon, in our case, a week.

More formally, we consider a relay network $\mathcal{G}' = (\mathcal{S} \cup \mathcal{H}' \cup \mathcal{T}, \mathcal{A}')$, where $\mathcal{H}'$ and $\mathcal{A}'$ respectively represent the functional relay hubs and transportation legs that can be used to transport parcels. We suppose that for each O-D pair $p = (s,t) \in \mathcal{P}$, the service provider aims to transport an average parcel weekly demand volume v_p from its origin outbound facility s to its destination last-mile delivery center t through the relay network $\mathcal{G}'$. In order to transport parcels on each leg $(i,j) \in \mathcal{A}'$, we account for a fixed scheduling cost c_S for each required truck, and a variable fuel-related cost c_F per parcel and per unit of distance traveled. Such fixed plus linear cost structure realistically represents several real-world freight costs (Greening et al. 2023). Then at the commodity origin location and at each relay hub, we suppose that parcels are sorted depending on their next transportation leg. So, the commodities face a unit handling cost c_H at every visited relay hub and their origin locations. After sorting, these parcels are consolidated in trucks. We denote by q the average cubic volume of a parcel, and by Q the total cubic volume capacity of each truck. The distance of each leg $(i,j)\in\mathcal{A}'$ is denoted by $\ell_{i,j}$.

Consequently, to devise a minimum-cost parcel consolidation plan, we adapt the arc-based flow planning model for consolidation trucking (Bakir et al. 2021) for the relay-network case. In this model, we define continuous decision variables $f^p_{i,j}\in\mathbb{R}_{\geq 0}$ that denote the amount of parcels transported on leg $(i,j)\in\mathcal{A}'$ for O-D pair $p\in\mathcal{P}$, and discrete variables $x_{i,j}\in\mathbb{Z}_{\geq 0}$ that represent the number of delivery trucks used to transport parcels on leg $(i,j)\in\mathcal{A}'$. The optimization model is given as follows:

$$\min_{x,f} \quad \sum_{p\in\mathcal{P}}\sum_{(i,j)\in\mathcal{A}'}(c_F\cdot\ell_{i,j}+c_H)\cdot f^p_{i,j}+\sum_{\{(i,j)\in\mathcal{A}'\,|\,i<j\}} c_S\cdot x_{i,j} \tag{EC.1a}$$

$$\text{s.t.} \quad \sum_{\substack{\{j\in\mathcal{H}'\cup\{t\}\\ |\,(i,j)\in\mathcal{A}'\}}} f^p_{i,j}-\sum_{\substack{\{j\in\mathcal{H}'\cup\{s\}\\ |\,(j,i)\in\mathcal{A}'\}}} f^p_{j,i}=\begin{cases} v_p & \text{if } i=s\\ 0 & \text{if } i\in\mathcal{H}'\\ -v_p & \text{if } i=t\end{cases}, \quad \forall\, p=(s,t)\in\mathcal{P},\ \forall\, i\in\mathcal{H}'\cup\{s,t\}, \tag{EC.1b}$$

$$\sum_{p\in\mathcal{P}} q\cdot f^p_{i,j}\leq Q\cdot x_{i,j}, \quad \forall\,(i,j)\in\mathcal{A}', \tag{EC.1c}$$

$$x_{i,j}=x_{j,i}, \quad \forall\,(i,j)\in\mathcal{A}'\mid i<j, \tag{EC.1d}$$

$$f^p_{i,j}\geq 0, \quad \forall\,(i,j)\in\mathcal{A}',\ \forall\, p\in\mathcal{P}, \tag{EC.1e}$$

$$x_{i,j}\in\mathbb{Z}_{\geq 0}, \quad \forall\,(i,j)\in\mathcal{A}'. \tag{EC.1f}$$

The objective (EC.1a) minimizes the total cost of parcel handling, truck scheduling, and truck fuel consumption for parcel delivery. Constraints (EC.1b) are flow conservation constraints to route parcels from their respective origins to their destinations. Constraints (EC.1c) ensure that enough delivery vehicles are scheduled on each leg to feasibly transport the parcels planned to travel along that leg. Finally, Constraints (EC.1d) ensure that scheduled trucks on each transportation leg return to their origin locations, which in turn facilitates the single-day driver trips.

It may occur that parcels may not feasibly travel between certain O-D pairs throughout the logistics network (e.g., due to a disruption). Hence, before solving the consolidation model (EC.1) for a given graph $\mathcal{G}'$, we create dummy direct long-haul transportation legs in $\mathcal{A}'$ between each O-D pair to ensure feasibility of the problem. However, we ensure that utilizing these long-haul legs is a very expensive parcel delivery option and will only be used when there is no available alternative to fulfill the corresponding demand.

In the computational experiments, we utilize a handling cost c_H of \$0.02576 per parcel and per location, a truck scheduling cost c_S of \$750, and a fuel cost c_F of \$0.0002318 per parcel for each kilometer traveled. We obtained these values from data related to average truck mileage, fuel price, truck cubic volume capacity, average parcel cubic volume, and labor wage statistics (U.S. Bureau of Labor Statistics 2023, Greening et al. 2023).

EC.3. Efficiency-Optimized Relay-Hub Network Design

To benchmark our proposed k shortest paths networks, we design relay logistics networks that are optimized only based on efficiency considerations. In particular, the objective is to select relay hubs to open that minimize the total cost to fulfill parcel demand through consolidation-based transportation under nominal operating conditions—i.e., in the absence of disruptions—while satisfying single-day delivery trips. Formally, we consider the same graph $\mathcal{G} = (\mathcal{S} \cup \mathcal{T} \cup \mathcal{H}, \mathcal{A})$ of potential network designs in which each O-D pair $p \in \mathcal{P}$ has an average parcel weekly demand volume of v_p units. The aim is to select a subset of hub locations $\mathcal{H}_o \subseteq \mathcal{H}$ of size at most N that minimizes the cost of the consolidation plan—computed from (EC.1)—in the induced subgraph $\mathcal{G}_{\mathcal{H}_o}$.

We define a binary relay-hub opening variable y_i for each hub candidate $i \in \mathcal{H}$ that takes a value of 1 if a relay hub is opened at location i, and 0 otherwise. Similarly to the network consolidation model (EC.1), we define a parcel flow variable $f^p_{i,j} \in \mathbb{R}_{\geq 0}$ for each O-D pair $p \in \mathcal{P}$ and for each transportation leg $(i,j) \in \mathcal{A}$, and a truck variable $x_{i,j} \in \mathbb{Z}_{\geq 0}$ for each $(i,j) \in \mathcal{A}$. The problem of designing *efficiency-optimized relay-hub networks* can be formulated as follows:

$$\min_{y,x,f} \quad \sum_{p \in \mathcal{P}} \sum_{(i,j) \in \mathcal{A}} (c_F \cdot \ell_{i,j} + c_H) \cdot f^p_{i,j} + \sum_{\{(i,j) \in \mathcal{A} \,|\, i<j\}} c_S \cdot x_{i,j} \tag{EC.2a}$$

$$\text{s.t.} \quad \sum_{i \in \mathcal{H}} y_i \leq N, \tag{EC.2b}$$

$$\sum_{\substack{\{j \in \mathcal{H} \cup \{t\} \\ |\, (i,j) \in \mathcal{A}\}}} f^p_{i,j} - \sum_{\substack{\{j \in \mathcal{H} \cup \{s\} \\ |\, (j,i) \in \mathcal{A}\}}} f^p_{j,i} = \begin{cases} v_p & \text{if } i = s \\ 0 & \text{if } i \in \mathcal{H} \\ -v_p & \text{if } i = t \end{cases}, \quad \forall\, p = (s,t) \in \mathcal{P},\ \forall\, i \in \mathcal{H} \cup \{s,t\}, \tag{EC.2c}$$

$$\sum_{p \in \mathcal{P}} q \cdot f^p_{i,j} \leq Q \cdot x_{i,j}, \quad \forall\, (i,j) \in \mathcal{A}, \tag{EC.2d}$$

$$f^p_{i,j} \leq v_p \cdot y_i, \quad \forall\, (i,j) \in \mathcal{A} \mid i \in \mathcal{H},\ \forall\, p \in \mathcal{P}, \tag{EC.2e}$$

$$f^p_{i,j} \leq v_p \cdot y_j, \quad \forall\, (i,j) \in \mathcal{A} \mid j \in \mathcal{H},\ \forall\, p \in \mathcal{P}, \tag{EC.2f}$$

$$x_{i,j} = x_{j,i}, \quad \forall\, (i,j) \in \mathcal{A} \mid i < j, \tag{EC.2g}$$

$$f^p_{i,j} \geq 0, \quad \forall\, (i,j) \in \mathcal{A},\ \forall\, p \in \mathcal{P}, \tag{EC.2h}$$

$$x_{i,j} \in \mathbb{Z}_{\geq 0}, \quad \forall\, (i,j) \in \mathcal{A}, \tag{EC.2i}$$

$$y_i \in \{0,1\}, \quad \forall\, i \in \mathcal{H}. \tag{EC.2j}$$

Constraints (EC.2e)-(EC.2f) ensure that parcels travel only between opened facilities.

EC.4. Fine-Tuning of the Solution Algorithms

To fine-tune our implementation of Benders decomposition for solving $\mathcal{I}_{path}$, we compared the running time of three variants of the algorithm when solving small instances. The first variant only generates cuts through the adjusted breadth-first search (Algorithm 1), the second variant only generates cuts through the spread assignment (Algorithm 2), and the third variant generates both cut types. The numerical results are summarized in Table EC.1.

Table EC.1 Solution times of Benders decomposition for solving $\mathcal{I}_{path}$ with different cut generation techniques

Cut generation	Data instance	# Hubs N	Solution times (mins) $k=2$	$k=3$	$k=4$
ABFS	1	10	2.38	2.89	3.51
	2	20	5.27	8.25	11.37
	3	30	19.36	30.04	52.71
	4	40	165.27	235.61	492.47
SA	1	10	4.29	4.73	5.16
	2	20	6.11	10.58	18.72
	3	30	32.01	49.66	61.13
	4	40	248.39	480.19	591.04
ABFS + SA	1	10	1.24	1.96	2.66
	2	20	2.49	3.26	4.97
	3	30	10.64	14.61	17.28
	4	40	90.28	167.19	384.67

We observe a significant time reduction by adding both cut types at each iteration of Benders decomposition, and we selected this variant for the computational experiments (Algorithm 3).

Similarly, in fine-tuning our implementation of the branch-and-price algorithm, we compared three rules for branching on continuous hub selection variables $y \in [0,1]^{|\mathcal{H}|}$ at a tree node. The first one branches on the hub $i' \in \arg\max\{y_i \mid y_i < 1,\ \forall\, i \in \mathcal{H}\}$, that is, with the largest fractional variable value. The second rule branches on the hub $i' \in \arg\min\{|y_i - 0.5|,\ \forall\, i \in \mathcal{H}\}$, that is, with the most fractional variable value. The third one branches according to the greedy hub selection algorithm (Algorithm 6). The numerical results are summarized in Table EC.2.

Table EC.2 Solution times of branch-and-price for solving $\mathcal{I}_{path}$ with different branching rules

Branching technique	Data instance	# Hubs N	Solution times (mins) $k=2$	$k=3$	$k=4$
Largest fractional value	1	10	4.91	5.16	5.47
	2	20	6.14	7.23	10.58
	3	30	8.75	11.68	19.06
	4	40	37.26	65.34	88.92
	5	50	175.86	304.54	457.66
Most fractional value	1	10	3.67	3.99	4.27
	2	20	5.02	5.44	6.56
	3	30	7.78	10.88	16.95
	4	40	26.15	43.16	64.74
	5	50	137.29	278.35	394.07
GHS	1	10	2.09	2.98	3.5
	2	20	3.49	4.68	5.21
	3	30	6.37	7.93	8.49
	4	40	16.42	20.61	33.76
	5	50	94.06	145.22	310.65

We observe that branching using the greedy hub selection leads to the best running times, and we selected this branching rule for the computational experiments (Algorithm 7).

EC.5. Network Resilience Under Transportation Leg Disruptions

We conduct another set of disruption experiments where we subject the networks to random transportation leg disruptions. Leg disruptions occur uniformly at random and prevent parcels from being routed through the associated dysfunctional leg. Table EC.3 compares the networks' performances utilizing the same degradation metrics, namely the expected percentage of unfulfilled demand and the expected delivery costs averaged over the parcel demand of the O-D pairs that remain connected.

Table EC.3 Network resilience comparison under uniform transportation leg disruptions

Degradation metric	Data instance	# Hubs N	E-O network	k shortest paths networks				
				$k=1$	$k=2$	$k=3$	$k=4$	$k=5$
Unfulfilled demand (%)	1	10	6.64	5.67	3.55	0.11	0.04	0.04
	2	20	3.67	4.03	2.50	0.10	0.03	0.03
	3	30	2.74	2.82	1.68	0.07	0.02	0.02
	4	40	1.91	1.74	1.32	0.04	0.02	0.02
	5	50	0.39	0.33	0.25	0.01	0.00	0.00
	6	60	0.17	0.14	0.13	0.00	0.00	0.00
Average delivery costs ($/parcel)	1	10	3.227	3.255	3.234	3.187	3.170	3.170
	2	20	3.008	2.984	2.954	2.938	2.913	2.913
	3	30	2.618	2.585	2.584	2.574	2.554	2.554
	4	40	2.533	2.520	2.487	2.476	2.466	2.466
	5	50	2.251	2.210	2.196	2.175	2.167	2.167
	6	60	2.061	2.014	1.975	1.972	1.965	1.965

Similarly to hub disruption experiments, the results show that the k shortest paths networks designed with $k \geq 2$ outperform the efficiency-optimized networks in the ability to sustain edge disruptions. Notably, this effect is quite pronounced for networks designed with $k \geq 3$ as they are able to fulfill almost all the parcel demand, and at a lower cost. This occurs because of the presence of a larger number of paths of comparable length that connect each O-D pair.

Tables 4 and EC.3 show that hub disruptions have more impact on the consolidation plans in comparison with transportation leg disruptions. As the designed networks are dense in nature, each hub has a high degree of connectivity to other hubs. Hence, when a hub is disrupted, the adjacent transportation legs are also dysfunctional and no parcel can be routed through them. In contrast, an edge disruption renders only a single leg dysfunctional, which in turn degrades less the consolidation plans.