http://pubsonline.informs.org/journal/ijoc

INFORMS JOURNAL ON COMPUTING
Vol. 00, No. 0, X 0000, pp. 00–00
ISSN 2694-4022, EISSN 2694-4030

Submitted to *INFORMS Journal on Computing*

A Quantum-Inspired Bi-level Optimization Algorithm for the First Responder Network Design Problem

Anthony Karahalios, Sridhar Tayur, Ananth Tenneti
Tepper School of Business, Carnegie Mellon University, Pittsburgh, PA 15213, USA

Amirreza Pashapour, F. Sibel Salman, Barış Yıldız
Department of Industrial Engineering, Koç University, Istanbul, 34450, Türkiye

Abstract. In the aftermath of a sudden catastrophe, First Responders (FRs) strive to reach and rescue immobile victims. Simultaneously, civilians use the same roads to evacuate, access medical facilities and shelters, or reunite with their relatives, via private vehicles. The escalated traffic congestion can significantly hinder critical FR operations. A proposal from the Türkiye Ministry of Transportation and Infrastructure is to allocate a lane on specific road segments exclusively for FR use, mark them clearly, and pre-communicate them publicly. For a successful implementation of this proposal, an FR path should exist from designated entry points to each FR-demand point in the network. The reserved FR lanes along these paths will be inaccessible to evacuees, potentially increasing evacuation times. Hence, in this study, we aim to determine a subset of links along which an FR lane should be reserved and analyze the resulting evacuation flow under evacuees' selfish routing behavior. We introduce this problem as the First Responder Network Design Problem (FRNDP) and formulate it as a mixed-integer non-linear program. To efficiently solve FRNDP, we introduce a novel bi-level nested heuristic, the Graver Augmented Multi-Seed Algorithm (GAMA) within GAMA, called GAGA. We test GAGA on synthetic graph instances of various sizes as well as scenarios related to a potential Istanbul earthquake. Our comparisons with a state-of-the-art exact algorithm for network design problems demonstrate that GAGA offers a promising alternative approach and highlights the need for further exploration of quantum-inspired computing to tackle complex real-world problems.

Key words: Disaster Preparedness, First Responder Network Design Problem, Quantum-inspired algorithm, Quadratic Unconstrained Binary Optimization (QUBO), Graver Augmented Multi-seed Algorithm (GAMA)

1. Introduction

Disaster Management (DM) is a critical and growing area of study in the Operations Research community (Kamyabniya et al. 2024). DM research is typically classified into four phases: mitigation, preparedness, response, and recovery (McLoughlin 1985, Galindo and Batta 2013). The present work contributes to the preparedness and response phases by addressing pre-disaster planning, aiming to improve traffic management for a swift and effective post-disaster first response and evacuation. To this end, we introduce and study the First Responder Network Design Problem (FRNDP).

After a sudden-onset disaster, such as a massive earthquake, mobile individuals naturally attempt to evacuate the affected region, access medical facilities or shelters, and reunite with their relatives (Goltz and Mileti 2011). Simultaneously, FRs - comprising search-and-rescue teams, firefighters, police officers, ambulance crews, debris clearance, and road restoration teams, relief aid distribution teams, and volunteers - rely on the same road infrastructure to promptly reach immobile victims in urgent need (Alexander and Klein 2009). The inevitable traffic congestion caused by mobile citizens has the potential to significantly hinder critical first-response operations.

The complex and time-sensitive nature of first-response operations requires meticulous pre-disaster planning. This involves employing analytical methodologies to swiftly implement pre-identified and pre-communicated policies when a disaster strikes. The two consecutive earthquakes of magnitudes 7.8 and 7.5 that hit Türkiye on February 6, 2023, and affected over 15 million people in 11 provinces in southeastern Türkiye and northwestern Syria exemplify the complexity of the required first-response operations. The earthquakes caused over 50,000 deaths, more than 100,000 injuries, and over 2.7 million displaced people in Türkiye alone due to 36,932 collapsed and more than 311,000 damaged buildings (World Bank and Global Facility for Disaster Reduction and Recovery 2023). After the earthquakes, 271,060 personnel were deployed to the region, including 35,250 search and rescue personnel, public employees, personnel of NGOs, international search and rescue personnel, and volunteers (The European Union Delegation in Türkiye and the EU Joint Research Centre 2023). Although rescue and response teams tried to reach the disaster-stricken region immediately, not all cities and villages could be provided with sufficient emergency services within the critical first 24 hours due to accessibility difficulties (METU 2023).

A proposal to ensure the accessibility of FRs is to reserve a lane on specific road segments exclusively for FR use during the crucial initial hours following a disaster referred to as “golden hours” (Lerner and Moscati 2001). However, this course of action requires careful planning, as this allocation diminishes the already congested road capacity for evacuees in private vehicles (Xie and Turnquist 2009). Moreover, road reservations must be decided and announced before the occurrence of a potential disaster. Communicating such a strategy amidst disaster chaos, compounded by potential communication breakdowns, would be an extremely challenging, if not impossible, task. Another advantage of identifying FR lanes to be reserved a priori is that the corresponding roads and the buildings along them can be strengthened via retrofitting projects in the pre-disaster stage to withstand the disaster (Peeta et al. 2010, Yücel et al. 2018, Zhang et al. 2023). The Transportation and Infrastructure Ministry of Türkiye announced, on August 26th, 2023, which highways would be reserved for FRs as part of preparation to better respond to a potential Istanbul earthquake and stated that critical structures on these roads would be strengthened. Accordingly, the Istanbul Municipality has been actively working on identifying emergency routes on urban roads.

In FRNDP, we seek to answer the questions, “On which links should we reserve a lane to be exclusively used by FRs?” and “What will be the resulting evacuation flow under evacuees’ selfish routing mechanism?”. Essentially, the problem is a complex mixed-integer non-linear program (MINLP) where the objective function depends on the outcome of another optimization problem, i.e., evacuees’ selfish routing decisions. Thus, the effort in FRNDP is to acquire the best combination of FR paths so that FRs can access all FR-demand regions via reserved lanes while minimizing the aggregate public evacuation time, constrained by the chosen FR paths and the evacuees’ behaviors. We assume that a lane must be reserved in both directions along the FR path so that immobile individuals can be quickly taken to exit points and medical centers as well.

Recognizing the important practical application of FRNDP, the goal of this study is to develop an algorithmic framework capable of leveraging advancements in quantum computing to tackle this challenging optimization problem. To this end, we note that our goal is not merely to develop a new heuristic for solving FRNDP that outperforms existing methods. With this perspective, the primary contributions of this study are twofold. First, we propose FRNDP for the first time in the DM literature. Although lane reservation problems for transportation networks have been addressed in DM literature (Xie and Turnquist 2009, Kimms and Maassen 2012), the existing models focus on other mechanisms for managing transportation networks such as lane reversals and intersection crossing elimination, or differ in their requirements such as needing to reserve an entire road for an FR instead of one lane. FRNDP exploits single-lane reservations in both directions along FR paths while maintaining public evacuation road capacities. Second, from a methodological perspective, our study introduces a novel quantum-inspired approach using test sets to solve FRNDP instances. We build on Graver Augmented Multi-Seed Algorithm (GAMA) (Alghassi et al. 2019b) to develop an effective heuristic, GAMA within GAMA, which we call GAGA. Our work introduces a different class of algorithms than the studies that have utilized quantum computing in hybrid approaches for bi-level problems, such as the quantum-inspired evolutionary algorithm coupled with a Frank-Wolfe algorithm (Yan et al. 2008) or with genetic algorithms (Fan et al. 2021). To the best of our knowledge, a bi-level optimization problem has not been solved using graver test sets. Our computational experiments benchmark GAGA with an exact state-of-the-art algorithm and give promising results with encouraging implications for the emerging area of quantum-inspired computing in addressing complex real-world problems.

The remainder of this paper is structured as follows. Section 2 reviews the related literature. Section 3 formally describes FRNDP and presents a MINLP model. In Section 4, we propose our novel quantum-inspired bi-level algorithm. Section 5 implements a Branch-and-Bound (B&B) algorithm for comparison purposes. Section 6 conducts computational experiments and discusses managerial insights. Section 7 concludes our paper.

2. Related works

FRNDP is a new variant of the Discrete Network Design Problem (DNDP), a well-studied bi-level optimization problem in transportation. In DNDP, the leader (outer) problem identifies an optimal set of candidate links to be adjusted in the network to minimize a travel-oriented metric (e.g., total travel time) subject to a budget constraint. The follower (inner) problem models the users' reactions, typically governed by a static traffic assignment problem (TAP) under user equilibrium (Dafermos 1980, Bard 2013).

Three primary classes of exact methodologies have been proposed to solve the DNDP or its linear approximation (derived by piece-wise linear link travel time approximation or linear outer approximation (Wang et al. 2015)): (i) a B&B algorithm (LeBlanc 1975, Farvaresh and Sepehri 2013, Bagloee et al. 2017, Yin et al. 2022), (ii) a generalized Benders' decomposition approach (Gao et al. 2005, Fontaine and Minner 2014, Bagloee et al. 2017, Fontaine and Minner 2018, Afkham et al. 2022), and (iii) a single-level reformulation (Farvaresh and Sepehri 2011, Fontaine and Minner 2014). LeBlanc (1975) proposed a B&B as the first exact method to solve DNDP. The authors used the single-level system-optimum DNDP (SO-DNDP) relaxation of the TAP, which ignores the follower objective function, to generate lower bounds and solve DNDP iteratively. Variational inequalities (Luathep et al. 2011), as well as interdiction cuts (Wang et al. 2013), have also been suggested to complement the SO-DNDP formulation. Bagloee et al. (2017) developed a B&B algorithm in which a generalized Benders decomposition approach is utilized to solve the SO-relaxation of the TAP at each node of the B&B tree. Bi-level optimization problems with a convex follower problem can be reformulated into a single-level model via the Karush-Kuhn-Tucker (KKT) conditions accompanied by binary variables to model complementary slackness (Bard 2013). This also applies to the DNDP since the follower TAP problem is a convex nonlinear program (NLP) (Beckmann et al. 1956, Rey 2020). A survey by Rey (2020) highlights that the largest DNDP instances solved to optimality are still small in scale.

Apart from exact methods, a range of heuristic approaches–evolutionary and greedy algorithms, followed by a classical paradigm– have been developed to solve DNDP variants. Chang and Chang (1993) introduced a heuristic algorithm based on expected generalized travel costs in an iterative assignment procedure to solve a DNDP variant of upgrading investment decisions under a budget limitation. Janson and Husaini (1987) presented two heuristics, one based on ranking and deleting alternative network modifications and the other on a constrained random sampling approach. Wu et al. (2009) proposed a greedy heuristic for a lane reservation problem for the Olympic games to minimize the total travel time for ordinary traffic while ensuring strict time windows for athletes and equipment among venues. Shangin and Pardalos (2016) focused on the network design problem with connectivity requirements, developing algorithms based on a greedy strategy and variable neighborhood search metaheuristics. Cheng et al. (2022) studied the problem of selecting some lanes on the road network to reserve for vehicles with special tasks (such as the FRs in our problem) that need to arrive at their destination within a target travel time. Their objective is to minimize the impact of lane reservations on regular vehicles while ensuring the target travel time for special vehicles. The authors developed an artificial bee colony (ABC) algorithm for the upper-level reservation decisions, and the Frank-Wolfe algorithm to solve the lower-level user equilibrium TAP.

Only a few studies provided quantum mechanics-inspired ideas (e.g., superposition, quantum gates, and probabilistic states) to improve classical methods for the bi-level network design problems. Yan et al. (2008) formulated a bi-level model to optimize transportation network capacity improvement under budget constraints, with the upper level minimizing total travel time and the lower level addressing user equilibrium TAP. The authors utilized a Quantum-Inspired Evolutionary Algorithm (QEA) coupled with a Frank-Wolfe algorithm. Recently, Fan et al. (2021) examined a QEA combined with genetic algorithms to schedule regional integrated energy systems. The QEA uses quantum bits to represent possible solutions, and quantum gates to update these quantum bits in each generation, enabling a more efficient search of the solution space compared to classical algorithms. Additionally, a repair operation is included to ensure that solutions respect the budget constraints. Different from these QEAs, GAMA is based on a concept from algebraic geometry and integer programming called a Graver basis. A Graver basis is a set of directions or vectors that guide the optimization process to iteratively improve a feasible solution. GAMA has found successful applications in various fields, including portfolio management (Venturelli and Kondratyev 2019, Alghassi et al. 2019c), cancer genomics (Alghassi et al. 2019a), and the quadratic assignment problem (Alghassi et al. 2019b).

Focusing on DM applications, a few DNDP variants have been proposed to establish emergency routes while considering the evacuation problem at the inner level (Afkham et al. 2022, Nikoo et al. 2018, Shariat Mohaymany and Nikoo 2020, Xie and Turnquist 2009, Kimms and Maassen 2012). Afkham et al. (2022) developed a bi-level optimization model for mass self-evacuation planning for non-compliant evacuees and employed an accelerated Benders decomposition. Nikoo et al. (2018) identified the emergency transportation routes to be closed to public traffic using a Branch-and-Cut (B&C) algorithm. Shariat Mohaymany and Nikoo (2020) developed a multi-objective stochastic program to obtain disaster response routes considering earthquake scenarios with arc failure probabilities, solved by a B&C algorithm. Xie and Turnquist (2009) studied a problem that included decisions for lane reversals, intersection crossing elimination and emergency vehicle lane reservations, and proposed a hybrid Lagrangian relaxation and tabu search algorithm for a triggered nuclear alarm case study as a slow-onset disaster context. Their model differs from FRNDP as it focuses on other mechanisms for network management, and their FR lane reservations only affect traffic in one direction - from FRs to the demand nodes - and not in reverse. Kimms and Maassen (2012) proposed a three-staged heuristic for evacuation planning with the assignment of rescue teams. Their problem differs from FRNDP as FR lane reservations require an entire road to be blocked off instead of allowing single lane reservations. The main similarity between FRNDP and DNDP variants is that they share the same structure of a bi-level formulation with a discrete outer problem of minimizing the total travel time and a continuous inner problem that solves a user equilibrium TAP. A complicating difference of FRNDP is that the outer problem requires reserving sets of lanes for FRs that connect each FR-demand node to an FR entry point, in both directions. A second difference between these problems is that FRNDP lacks budget constraints since lane reservations for FRs are cost-negligible.

Our study pioneers the development of GAGA, a novel quantum-inspired bi-level optimization algorithm to solve FRNDP instances, where both levels employ GAMA. GAMA augments a given feasible solution by using Graver basis elements as a set of candidate improving directions, to progress toward the optimal solution. The algorithm uses a multi-seed approach, meaning it starts from multiple feasible points (seeds) to explore the solution space. Picking the best solution (among the local optima) helps us come closer to the global optimum. For solving the outer level problem in FRNDP, GAGA is similar to the aforementioned heuristics that use search methods like ABC and neighborhood search. The key difference is the underlying GAMA heuristic within the GAGA, where the search method relies on Graver basis elements (generated by quantum or quantum-inspired methods) to obtain candidate feasible solutions for improving the objective function. Typically, the nested heuristics solve the inner problem exactly using methods such as Frank-Wolfe. A further difference in the GAGA approach is to also use GAMA for approximately solving the inner problem. At the end of the inner GAMA search, we solve the inner problem exactly to ensure a feasible final solution. The fundamentals of GAMA and the GAGA structure will be detailed in Section 4.

3. First responder network design problem

Section 3.1 provides an outline of FRNDP, and Section 3.2 formulates a MINLP model for FRNDP with a glossary of notation provided in Appendix A of the Online Supplement. To clarify the concepts of Sections 3 and 4, an illustrative example is presented in Appendix B of the Online Supplement.

3.1. Problem overview

Consider a network $G = (N, A)$ with node set N and directed link set A. Let F be the set of nodes that must be visited by FRs, and S denote the set of nodes with disaster-affected individuals (agents) who aim to evacuate the network. Let E be the set of entry/exit nodes from which FRs can enter and agents can evacuate the network. Without loss of generality, we assume that these entrance and exit nodes are the same and originate from highways within the network or medical or shelter points, even though they can theoretically be distinct. Let d_i denote the number of agents, in terms of the number of vehicles, residing at $i \in S$. For each $(i,j) \in A$, let u_{ij} be the maximum vehicle capacity, T_{ij} be the free-flow travel time of the link for a single vehicle, $h_{ij} \in \{1, 2, \cdots\}$ be the number of lanes, and x_{ij} be the number of agents traveling on it. Let $t_{ij}(x_{ij}, u_{ij}, T_{ij})$ denote the travel time function from i to j along link $(i,j) \in A$, which depends on traffic flow, link capacity, and free-flow travel time. It is modeled as the classic Bureau of Public Roads (BPR), a strictly convex function of the form

$$t_{ij}(x_{ij}, u_{ij}, T_{ij}) = T_{ij}\left(1 + \alpha\left(\frac{x_{ij}}{u_{ij}}\right)^{\beta}\right), \tag{1}$$

where we set $\alpha = 0.15$ and $\beta = 4$ following the standard by US Bureau of Public Roads. Office of Planning. Urban Planning Division (1964). We assume that agents decide to travel to an exit node in E under optimal selfish routing behavior, which can be altered depending on the behavior of other agents. Finally, it is assumed that the reserved FR lanes will not be congested.

The key tradeoff in FRNDP involves determining the allocation of limited link capacity between FRs and agents. Therefore, the primary decision in FRNDP is to select a subset of links along which one lane in both directions will be reserved for FRs. Reserving one lane in both directions is necessary to allow FRs to visit nodes and return to their origin points to transport immobile individuals to exit points or medical centers. Additionally, since the traffic volume for FRs is relatively low, reserving a single lane on selected links is deemed sufficient and consistent with the literature on lane reservation problems (Fang et al. 2011, Wu et al. 2018, Li et al. 2022). FRNDP ensures that each node in F is connected to a node in E via a path of reserved lanes. Reserving a lane on link (i,j) reduces its capacity by $u_{ij} \leftarrow \frac{h_{ij}-1}{h_{ij}} u_{ij}$. We note that if more than one lane exists on a link, agents can use its unreserved lanes. Furthermore, an entrance node in E can be connected to multiple nodes in F via several common reserved lanes, opting not to overfill the capacity. The objective is to minimize the total evacuation time for agents located in S under user equilibrium traffic conditions. An alternative objective could be to minimize the maximum travel time over the agents.

3.2. A MINLP model

Based on the described problem characteristics in Section 3.1, in this section, we formulate FRNDP as a bi-level MINLP. The necessity of developing a bi-level model is discussed in Appendix C of the Online Supplement. The outer problem $P_1(\mathbf{x})$ corresponds to choosing which links to reserve a lane on for a given inner problem solution $\mathbf{x}$. Similarly, the inner problem $P_2(\mathbf{y})$ solves the user equilibrium traffic problem for a given outer problem solution $\mathbf{y}$. A glossary of notations, including sets, parameters, and decision variables, has been provided in Appendix A of the Online Supplement.

$$P_1(\mathbf{x}) \qquad \min_{y} \sum_{(i,j)\in A} x_{ij} \cdot t_{ij}(x_{ij}, u_{ij}(y_{ij}), T_{ij}) \tag{2a}$$

$$\text{s.t.} \sum_{j\in N:(i,j)\in A} w_{ijk} - \sum_{j\in N:(j,i)\in A} w_{jik} = \delta_{ik} \qquad \forall i \in N\setminus E, \forall k \in F \tag{2b}$$

$$w_{ijk} \le y_{ij} \qquad \forall (i,j) \in A, \forall k \in F \tag{2c}$$

$$\sum_{j\in N:(i,j)\in A} w_{ijk} \le 1 \qquad \forall i \in N\setminus E, k \in F \tag{2d}$$

$$y_{ij} = y_{ji} \qquad \forall (i,j) \in A \tag{2e}$$

$$y_{ij} \in \{0,1\} \qquad \forall (i,j) \in A \tag{2f}$$

$$w_{ijk} \in \{0,1\} \qquad \forall (i,j) \in A, \forall k \in F \tag{2g}$$

$$\mathbf{x} \in P_2^* \tag{2h}$$

$$P_2(\mathbf{y}) \qquad \min_x \sum_{(i,j)\in A} \int_0^{x_{ij}} t_{ij}(v, u_{ij}(y_{ij}), T_{ij}) dv \tag{3a}$$

$$\text{s.t.} \sum_{j\in N:(i,j)\in A} x_{ij} - \sum_{j\in N:(j,i)\in A} x_{ji} = d_i \qquad \forall i \in N \backslash E \tag{3b}$$

$$x_{ij} \geq 0 \qquad \forall (i,j) \in A \tag{3c}$$

$$\mathbf{y} \in P_1^* \tag{3d}$$

The outer problem $P_1(\mathbf{x})$ comprises a binary variable y_{ij} $\forall (i,j) \in A$ that represents the choice of link (i,j) for FR lane reservation. Also, decision variables w_{ijk} $\forall (i,j) \in A, k \in F$ represent the choice of using link (i,j) to reach node k by FRs. The objective function (2a) minimizes the total travel times of all agents from their settled nodes to an exit node. Note that the capacity u_{ij} and thus the travel times t_{ij} depend on variables $\mathbf{y}$ as discussed in Section 3.1. Constraints (2b) ensure that reserved lanes connect each FR-demand node to an FR entry node. Constraints (2c) guarantee that one or more FR-demand nodes can be connected to an FR entry node through a path comprising link (i,j) if a lane on link (i,j) is reserved. Although cyclic paths are admissible in $P_1(\mathbf{x})$, constraints (2d) avoid such paths by requiring that there can be a maximum of one outward edge at each node $i \in N \backslash E$. Constraints (2e) impose the assumption that lane reservation is made bidirectional. Constraints (2f) and (2g) define decision variable domains. Constraint (2h) ensures that $P_1(\mathbf{x})$ uses $\mathbf{x}$ as the optimal solution to $P_2(\mathbf{y})$, where $\mathbf{x}$ corresponds to inner problem decision variables and $P_2^*(\mathbf{y})$ is the set of optimal solutions to $P_2(\mathbf{y})$ for a given outer solution $\mathbf{y}$.

The $P_2(\mathbf{y})$ model takes as input a given feasible solution $\mathbf{y}$ from the outer problem $P_1(\mathbf{x})$. We define a continuous variable x_{ij} $\forall (i,j) \in A$ that represents the flow of agents evacuating the network using link (i,j). The inner objective (3a) is the standard objective function used in selfish routing problems (Beckmann et al. 1956, Ata et al. 2017), albeit with a modified travel time function that captures the dynamic link capacities with respect to lane reservations. We note that the flow on any edge can exceed its capacity. However, the travel time in such cases increases drastically, as given by Equation (1). Constraints (3b) ensure that all agents at nodes in S reach an exit node E. Constraints (3c) enforce the non-negativity of the $\mathbf{x}$ variables. Continuous flow is a standard assumption in classic TAP models, upon which the stability of results and user equilibrium have been demonstrated in the literature (Lu and Nie 2010, Jin 2015, Rey 2020). Constraint (3d) ensures that $P_2(\mathbf{y})$ uses $\mathbf{y}$ as the optimal solution to $P_1(\mathbf{x})$, where $\mathbf{y}$ corresponds to outer problem decision variables and $P_1^*(\mathbf{x})$ is the set of optimal solutions to $P_1(\mathbf{x})$ for a given inner solution $\mathbf{x}$. We note that there is no need to balance flow conservation for entry/exit nodes E, as indicated by constraints (2b) and (3b) because the agent flows begin/terminate at these points. To tackle the complex FRNDP models $P_1(\mathbf{x})$ and $P_2(\mathbf{y})$, we propose a novel quantum-inspired algorithm named *GAGA* in Section 4.

4. Quantum-inspired algorithm

We begin by introducing the fundamentals of GAMA in Section 4.1. Section 4.2 overviews the computation of a partial Graver basis in GAMA. Section 4.3 describes a customized GAMA for FRNDP, following the studies of Alghassi et al. (2019b) and Alghassi et al. (2019c). Finally, we introduce GAGA, a novel bi-level optimization algorithm for FRNDP, in Section 4.4. The concepts discussed in Section 4 are illustrated in Appendix B of the Online Supplement.

4.1. GAMA

GAMA is a heuristic algorithm for solving general non-linear integer optimization problems of the form presented in model (4),

$$\begin{cases} \min \ f(x) \\ Ax = b \\ l \leq x \leq u, \ x \in \mathbb{Z}^n \end{cases} \tag{4}$$

where $A \in \mathbb{Z}^{m \times n}$, $b, l, u \in \mathbb{Z}^n$, and $f : \mathbb{R}^n \to \mathbb{R}$ is a non-linear real-valued objective function with linear constraints. One approach to solving such models is to use an augmentation procedure. It begins with an initial feasible solution to model (4) and iteratively seeks improvement by exploring directions from a set of pre-determined candidate improving directions, known as a *test set* or *optimality certificate*. The algorithm terminates once no candidate direction results in an improved objective function value. The test set is defined in Definition 1, taken verbatim from Alghassi et al. (2019c).

Definition 1 *A set $\Omega \subseteq \mathbb{Z}^n$ is a test set or an optimality certificate if for every feasible, non-optimal solution of model (4), say x_0, there exists $t \in \Omega$ and $\lambda \in Z_+$ such that $f(x_0 + \lambda t) < f(x_0)$. The vector $t \in \Omega$ is called an augmenting direction.*

In principle, GAMA uses a Graver basis (Graver 1975) as its test set. Definitions 2 and 3 introduce the Graver basis, taken verbatim from Alghassi et al. (2019c).

Definition 2 *Given vectors $x, y \in \mathbb{R}^n$, we define x to be conformal to y, written as $x \sqsubseteq y$, if $x_i \times y_i \geq 0$ (x and y lie in the same orthant), and $|x_i| \leq |y_i|$, $\forall\ i \in \{1, \cdots, n\}$. Similarly, a sum $u = \sum_i v_i$ is called conformal if $v_i \sqsubseteq u$, $\forall i$.*

Definition 3 *For a matrix $A \in \mathbb{Z}^{m \times n}$, define the lattice $L^*(A) = \{x | Ax = 0, x \in \mathbb{Z}^n, x \neq 0\}$. The Graver basis, $\mathcal{G}(A) \subset \mathbb{Z}^n$, of an integer matrix A is defined as the finite set of $\sqsubseteq$-minimal (indecomposable) elements in $L^*(A)$. A set of elements $\mathcal{G}$ is $\sqsubseteq$-minimal if $\forall g_i, g_j \in \mathcal{G}$, $g_i \not\sqsubseteq g_j$, when $i \neq j$.*

The Graver basis of an integer matrix $A \in \mathbb{Z}^{m\times n}$ is known to be a test set for integer linear programs (Graver 1975). Graver basis also serves as a test set for specific non-linear objective functions, such as Separable convex minimization (Murota et al. 2004), Convex integer maximization (De Loera et al. 2009), Norm p minimization (Hemmecke et al. 2011), Quadratic (Lee et al. 2010, Murota et al. 2004) and Polynomial minimization (Lee et al. 2010). These works show that the required number of augmentation steps for their studied problems is polynomial. The Graver basis can be computed for relatively small problems using classical methods, such as those developed by Pottier (1996) and Sturmfels and Thomas (1997). However, obtaining a complete test set for larger problems is computationally expensive. Thus, we employ the *partial Graver basis*, as described in Section 4.2.

4.2. Partial Graver basis computation

Acquiring a complete test set for the complex FRNDP is computationally prohibitive. We utilize *partial Graver basis*, which is a subset of the Graver basis. Using a partial Graver basis, GAMA provides a high-quality, feasible, but not provably optimal solution. Alghassi et al. (2019b) and Alghassi et al. (2019c) studied such a method where the heuristic is improved by starting GAMA from multiple initial feasible solutions as starting seeds instead of just one. A partial Graver basis is systematically obtained as follows. We begin by collecting elements in $L^*(A)$ by taking the differences of the feasible solutions of model (4). Subsequently, a classical post-processing step by $\sqsubseteq$-minimal filtering (Alghassi et al. 2019b) yields a partial Graver basis. We provide the details of this post-processing procedure in Appendix D of the Online Supplement. To obtain a sample of feasible solutions, we express the constraints of model (4) as $(Ax-b)^\top(Ax-b)=0$, which can be rewritten as $x^\top Q_I x - 2b^\top Ax + b^\top b = 0$, where $Q_I = A^\top A$. This is equivalent to solving a quadratic unconstrained integer optimization (QUIO) given by model (5),

$$\min \quad x^\top Q_I x - 2b^\top Ax, \quad x \in \mathbb{Z}^n, \tag{5}$$

where the objective function value cannot be less than $-b^\top b$, and in case of equality, the x variable is guaranteed to be feasible. For an integer x, an integer-to-binary transformation from x to X is required, which can be performed using a binary encoding or a unary encoding, expressed as $x = L + EX$, with L being the lower bound vector and E as the encoding matrix.

In general, for any $i \in \{1, ..., n\}$, if the lower and upper bounds are l_i and u_i respectively, then $x_i = l_i + e_i^\top X_i$. For binary encoding, $e_i^\top = [2^0, 2^1, ..., 2^{k_i-1}]$ where $k_i = \lceil \log_2(u_i - l_i) \rceil$ and for unary encoding, $e_i^\top = [1, 1, ..., 1]$ with $|e_i| = u_i - l_i$. Usually, the transformation that results in fewer additional variables leads to better performance. However, in FRNDP, we only deal with binary variables, so no encoding is required. From the transformation described above, we attain a Quadratic Unconstrained Binary Optimization (QUBO) with binary variables X as in model (6),

$$\min \quad X^\top Q_B X, \quad X \in \{0,1\}^{nk}, \tag{6}$$

where $Q_B = E^\top Q_I E + 2diag[(L^\top Q_I - b^\top A)E]$ and $k = \sum_{i=0}^{n-1} k_i$. The above QUBO can be solved on behalf of model (5) via an annealer or by means of simulated annealing (SA) (Kirkpatrick et al. 1983) to obtain a sample of feasible solutions. An annealer refers to a specially constructed quantum or semi-classical hardware (Lucas 2014, McMahon et al. 2016, Glover et al. 2018, King et al. 2018, Harris et al. 2018, Wang and Roychowdhury 2019, Chou et al. 2019, Mohseni et al. 2022) used for solving QUBO problems mapped to an Ising model.

4.3. GAMA for FRNDP

In this section, we customize GAMA to FRNDP by only searching on FR paths. GAMA solves the outer problem $P_1(\mathbf{x})$ by searching over a feasible set of FR paths and solves the inner problem $P_2(\mathbf{y})$ for each feasible set of FR paths by using a gradient descent algorithm from LeBlanc et al. (1975), which gives a provably optimal solution to the inner problem. To search over feasible sets of FR paths, we start by obtaining a sample of n_{paths} feasible paths for each node $k \in F$, each connected to some FR entry node in E. These feasible paths are attained by solving a QUBO over constraints (2b) using a quantum annealing device (i.e., D-Wave), classical SA, or a path-finding algorithm such as (Yen 1971)'s algorithm for k-shortest paths. Then, we combine feasible FR paths for each node in F to create M feasible solutions (seeds) for the entire network. We generate each feasible solution by randomly choosing a path for each $k \in F$ from the corresponding set of n_{paths}. In our experiments, we checked that the M feasible solutions correspond to different initial FR paths in our experiments. Next, we obtain a set of lattice elements (recall Definition 3) by taking all pairwise differences of the M feasible solutions. Finally, we obtain a partial Graver basis from the lattice elements by using the procedure in Alghassi et al. (2019b,c), ensuring that all elements are conformal minimal.

Given a partial Graver basis and M initial seeds, GAMA for FRNDP performs as follows. For each initial seed, it considers the corresponding initial feasible solution. Then, it applies each direction in the partial Graver basis to update variables $\mathbf{w}$ and $\mathbf{y}$, with a step size $\lambda = 1$. Next, it runs the gradient descent algorithm of LeBlanc et al. (1975) to calculate the values of the inner $\mathbf{x}$ variables. If a direction exhibits improvement, the solution is updated along the step. The algorithm terminates when no direction improves the objective function value compared to the recent one.

GAMA is particularly suitable for solving FRNDP because it has significant flexibility in handling complex objective functions. In crisis situations, it is critical to minimize the time-to-destination for every agent (i.e., evacuees and FRs). Consequently, most studies in the literature use total travel time as the objective. Considering its practical relevance and to maintain consistency with the literature, we adhere to the same objective function in our study. Still, other objective functions can be incorporated based on different measures of efficiency and fairness, such as the Gini index representing the differences between travel time across the locations. GAMA has a significant advantage over classical methods in this respect as it can handle arbitrary objective functions, provided the solution evaluation is fast enough to perform improvement checks in all directions of the Graver basis. GAGA, which we introduce next as an extension of GAMA, also inherits this advantage.

4.4. GAGA: Bi-level GAMA within GAMA

As discussed in Section 4.3, we use the GAMA heuristic to solve the outer problem, $P_1(\mathbf{x})$. A methodological enhancement in this work is to also solve the inner problem $P_2(\mathbf{y})$ using GAMA, motivated by the fact that approximately solving $P_2(\mathbf{y})$ at each step can significantly accelerate the solution procedure. While the GAMA solution to $P_2(\mathbf{y})$ is an approximation, we enhance the final inner GAMA solution with an additional step by solving $P_2(\mathbf{y})$ using the gradient descent algorithm proposed by LeBlanc et al. (1975). To the best of our knowledge, this is the first time a heuristic that uses GAMA (for $P_2(\mathbf{y})$) within GAMA (for $P_1(\mathbf{x})$) has been explored. We refer to this nested algorithm as GAGA.

As before, to generate a partial Graver basis with respect to the constraint matrix in $P_2(\mathbf{y})$, we find lattice elements that are differences of paths between nodes with evacuees in S and exit nodes in E, and then apply the same procedure to end up with conformally minimal solutions. To create many of these paths, for each node, $k \in S$, we generate a sample of n_{paths} feasible paths to exit nodes in E. As before, this can be done with an annealer, SA, or a path-finding algorithm. We construct each initial feasible solution to $P_2(\mathbf{y})$ in GAGA as follows. For each node $k \in S$, we randomly sample d_k paths, with replacement, from the set of n_{paths} paths and send one evacuee along each path. Similar to generating initial FR paths, we generate M initial seeds for FR paths, where it is verified that the M seeds correspond to different full evacuee paths.

The inner GAMA solves $P_2(\mathbf{y})$ as follows. We start from each of the M different initial solutions. At each step, GAMA checks each possible improving direction in the partial Graver basis, using a step size of $\lambda = 1$, and updates the current solution when an objective improving direction is attained. The inner GAMA terminates when there is no improving direction up to a tolerance. Without a tolerance threshold, the Graver augmentation runs until there is no improving direction amongst all Graver basis elements. We experiment with/without a tolerance threshold of 10^{-3} for the inner problem.

We note that the Graver augmentation for $P_2(\mathbf{y})$ is carried out using a partial Graver basis. Moreover, the resulting flow amounts will have integer values since the step size, $\lambda = 1$, despite the underlying model allowing for continuous flow values. Hence, it is possible and can be expected that the user equilibrium flow obtained by the GAMA heuristic is a close approximation to, but not the exact user equilibrium flow solution. Therefore, as a final step, we run the gradient descent algorithm from LeBlanc et al. (1975) on each of the local optimal FR solutions at the end of the GAGA and report the optimal total evacuation time under this user flow equilibrium as the best result. In the rest of the paper, we denote the final solution obtained at the end of the GAGA procedure as GAGA-*only* and the solution after running the additional step of the gradient descent as GAGA-*LeBlanc*. Unless explicitly stated, the final solutions presented refer to the GAGA-*LeBlanc* values.

5. A Branch-and-Bound algorithm

A branch-and-bound algorithm is considered among the most efficient exact solution methodologies for DNDP (Rey 2020). We implement a B&B algorithm in this section to solve FRNDP and compare its performance with GAGA. In theory, the B&B algorithm is an exact method, but in practice, it would require far too long to enumerate all feasible solutions. The method mimics the algorithm that was first developed by LeBlanc (1975) and later extended in Farvaresh and Sepehri (2013) by solving the SO-DNDP instead of the TAP. Our executed B&B is adapted to consider the additional constraints of FRNDP, such as ensuring reserved links for FRs form paths from entry to FR-demand nodes. We acknowledge that other heuristics and exact methods would also be interesting to investigate. Next, we will describe the details of the B&B algorithm, including the primal heuristic, branching rule, dual-bound heuristic, and node selection heuristic.

We use the following primal heuristic to attain a feasible solution to $P_1(\mathbf{x})$. First, we compute the shortest path from each node in E to each node in F using Dijkstra's algorithm. Then, for each node in F, we reserve the links along the shortest of these paths, which corresponds to an assignment of the variables $\mathbf{y}$. Next, we solve $P_2(\mathbf{y})$ using a gradient descent algorithm (LeBlanc et al. 1975), which solves each subproblem by calculating several shortest paths for each node in S. The algorithm applies a halving method on the step size λ until a relative tolerance of $\epsilon = 0.001$ as the gradient descent stopping criterion. We use a greedy heuristic at the root node to find an initial feasible solution to begin the gradient descent algorithm. At a non-root node, we initialize a feasible solution with the optimal solution of the parent node to accelerate the gradient descent algorithm.

Next, we describe the branching rule. We develop a branching procedure that creates two child nodes in the B&B tree for a given link: one reserves this link while the other does not. We use a random order of the FR nodes F. If node $i \in F$ were considered in the previous branch, then the next branch would use the next node $j \in F$ from the given order. Also, we order the links connecting j to an exit node $k \in E$ in the solution $\mathbf{y}$, starting from the exit node. We select the first link that has not yet been fixed for branching. We note that this branching rule creates one child node with the identical optimal solution to its parent.

A dual bound can be obtained by solving for the total evacuation time of a *system optimal* (i.e., authoritarian) solution using the same gradient descent algorithm. Aligned with the literature, we noticed that the lower bounds are not effective in fathoming search nodes in the B&B tree. Thus, to improve the B&B efficiency, we forgo calculating lower bounds in our experiments. As a consequence of not calculating lower bounds, we implement a breadth-first node selection heuristic instead of another commonly used heuristic that chooses the next node to solve based on lower bounds.

6. Computational experiments

In this section, we evaluate the performance of GAGA on two classes of instances. Section 6.1 examines synthetic instances. In Section 6.2, we design a case pertinent to a region in Istanbul, the most populated city in Türkiye, which is under serious earthquake risk. As pointed out in recent news (Duvar English: Turkey's own independent gazette 2023) and scientific studies (Parsons et al. 2000), Istanbul is expected to be hit by an earthquake with a magnitude above 7.0 Richter in the upcoming decade at a probability of 64%±15%. In our experiments, we compare the performance of GAGA with the B&B method described in Section 5. Appendix C of the Online Supplement gives an analysis of the necessity of a bi-level formulation, and Section 6.3 gives an analysis of our experimental results. Throughout the experiments, we varied the following GAGA parameters.

- $n_{samples} \in \{1000, 10000\}$: The number of reads or samples in the SA step or D-Wave annealer.
- $n_{paths} \in \{20, 25, 30, 35, 40\}$: The number of paths from/to each node $i \in N \backslash E$.

All the relevant source codes and instance data are available on Github (Karahalios et al. 2024). GAGA is implemented on a single core of Intel Xeon Gold 6252 @ 2.10 GHz (Processor-A) with 192GB memory and 48 cores. The B&B has been carried out on a single core of Intel(R) Xeon(R) Gold 6248R CPU @ 3.00GHz (Processor-B) with 64 GB memory and 8 cores. The performance of these processors has been compared in Appendix E of the Online Supplement. We note that only a single core is used for the computational experiments in both the GAGA and B&B runs. There is no parallel processing involved, although there appears to be a high possibility of developing such algorithms. So, the different number of cores in the two processors is not relevant for the runtime comparison.

6.1. Synthetic graphs

A set of graph instances is constructed using synthetic data to facilitate a practical assessment of GAGA performance and its comparison with B&B. In these graphs, we randomly generate n nodes on a two-dimensional grid, with coordinates within a uniform [0,1] distribution. Next, we produce $\lfloor \frac{n}{10} \rfloor$ extra nodes as entry/exit nodes E on the boundary of the unit square. The remaining nodes constitute the set F. An edge connects each pair of nodes with a probability that is the product of p and an additional factor that takes into account the distance between the two nodes. Capacities of edges follow a Normal(50,20) distribution. All edges possess two lanes in both directions. The evacuee population at each node follows a Normal(100,10) distribution. The free-flow travel time, T_{ij} in Equation (1) is chosen to be 1 unit for all edges equally. All non-positive samples in these random variables have been replaced with positive samples. The instance numbers are denoted as i. As a remark, if a generated instance is found to be infeasible, we replace it with another new instance. In total, we design 10 instances $i \in \{0, \cdots, 9\}$ for each $n \in \{10, 20, 30\}$ and $p \in \{0.5, 0.75\}$ combination.

Table 1 compares the objective function values for instances of $n \in \{20, 30\}$ nodes (The $n = 10$ results are analyzed in Appendix F of the Online Supplement) obtained from GAGA using feasible solutions from SA, with B&B. Alternative to SA, we generate a set of feasible solutions using Yen (1971) k-shortest loopless path algorithm. The B&B algorithm is given a runtime of $14,400$ seconds, which is longer than the maximum time taken for the GAGA run on any of these instances. The SA runs are carried out with $n_{samples} \in \{1000, 10000\}$ (referred to as GAGA-SA-1000 and GAGA-SA-10000) to assess the impact of the $n_{samples}$ parameter on the solution quality. The results using feasible solutions from Yen's algorithm are called GAGA-Yen. Yen's algorithm is used to obtain the first k-shortest paths without cycles, and accordingly, we set $k = 25$ to be consistent with $n_{paths} = 25$. The bold numbers in Table 1 highlight the best objective function values (i.e., total travel times) achieved among the four methods discussed. We observe in Table 1 that the solutions are comparable among both GAGA-SAs and GAGA-Yen. Compared with B&B, GAGA yields better solution quality. We conduct a thorough computational analysis of these graphs in Appendix G of the Online Supplement.

6.2. Case study: Istanbul instances

To assess GAGA on a real-life urban road network and demonstrate the impact of FR lane reservation on evacuees' travel times, we generate three sample instances based on the city of Istanbul, Türkiye in Section 6.2.1. The performance comparison of GAGA and B&B is detailed in Section 6.2.2.

6.2.1. Instance generation

In this section, we designate a realistic case study on a specific region in Istanbul, including the Avcılar district and a neighborhood around the Kucukcekmece Lake. Istanbul is subject to major earthquake risk due to the fault line passing under the sea to the city's south (Erdik and Durukal 2008). The region of study is home to nearly 500,000 residents. We leveraged Istanbul's OpenStreetMap (OSM) dataset to generate the region's road network, focusing on the road infrastructure information layer, analyzed in ArcGIS 10.5. The links are clustered into three groups based on their road-type features. The first class encompasses *primary*, *secondary*, *motorway*, and *trunk* road types. The second group includes *tertiary*, *residential*, *service*, *track*, and *living street* lanes. While the second group cannot accommodate FR lane reservations due to capacity and lane limitations, they remain accessible for evacuees to navigate the network. Given the substantial number of links in the second class, the selected arcs comprise only roads featuring two or more lanes on both sides. Finally, vehicle-impassable link types such as *cycleway*, *pedestrian*, *footway*, *steps*, and *path* have been excluded. The resulting network, depicted in Figure 1, consists of 179 vertices, each representing a residential area or a potential origin node for FR demands, and 234 arcs connecting the vertices in

Table 1 Objective function value comparison of GAGA implemented with (i) SA-1000, (ii) SA-10000, and (iii) Yen (1971)'s k-shortest loopless path algorithm, with B&B

Instance			GAGA-SA-1000		GAGA-SA-10000		GAGA-Yen		B&B	
n	p	i	obj	time (s)	obj	time (s)	obj	time (s)	obj	time (s)
20	0.5	0	2.83 e+06	2,931.32	**2.64 e+06**	4,368.71	2.69 e+06	3,697.60	4.03 e+06	14,400
20	0.5	1	1.45 e+06	2,387.13	1.47 e+06	3,463.57	**1.42 e+06**	2,390.27	2.15 e+06	14,400
20	0.5	2	**0.71 e+06**	1,931.70	0.73 e+06	2,503.33	0.75 e+06	1,813.50	9.50 e+05	14,400
20	0.5	3	2.18 e+05	1,625.86	**2.09 e+05**	2,383.20	2.15 e+05	1,406.08	3.12 e+05	14,400
20	0.5	4	**7.39 e+05**	2,788.98	7.71 e+05	2,917.41	8.37 e+05	2,186.68	8.35 e+05	14,400
20	0.5	5	0.96 e+06	2,011.54	0.97 e+06	2,678.38	**0.90 e+06**	1,676.48	1.07 e+06	14,400
20	0.5	6	3.24 e+06	2,536.07	2.89 e+06	3,327.69	**2.65 e+06**	2,685.08	3.95 e+06	14,400
20	0.5	7	2.08 e+05	2,199.30	**2.01 e+05**	3,456.32	2.03 e+05	2,229.30	2.56 e+05	14,400
20	0.5	8	2.43 e+05	1,641.64	2.42 e+05	2,354.32	**2.41 e+05**	1,514.79	3.37 e+05	14,400
20	0.5	9	6.20 e+05	2,529.35	6.24 e+05	3,374.94	5.80 e+05	2,948.75	**5.51 e+05**	14,400
20	0.75	0	**6.90 e+04**	1,783.51	7.30 e+04	3,277.64	7.30 e+04	1,188.81	7.86 e+04	14,400
20	0.75	1	1.12 e+05	1,900.06	1.12 e+05	3,649.95	**1.12 e+05**	1,753.16	1.18 e+05	14,400
20	0.75	2	**0.83 e+05**	1,691.69	0.87 e+05	2,724.63	0.88 e+05	1,159.22	1.01 e+05	14,400
20	0.75	3	**7.00 e+04**	2,235.44	7.50 e+04	2,782.49	7.90 e+04	1,148.20	9.35 e+04	14,400
20	0.75	4	**0.99 e+05**	1,735.64	1.00 e+05	2,965.26	1.02 e+05	1,159.36	1.06 e+05	14,400
20	0.75	5	**0.97 e+05**	1,963.95	0.97 e+05	3,402.85	0.97 e+05	2,020.67	1.01 e+05	14,400
20	0.75	6	**0.91 e+05**	1,508.97	0.93 e+05	3,555.80	0.92 e+05	1,476.82	9.74 e+04	14,400
20	0.75	7	**0.89 e+05**	1,710.47	0.89 e+05	3,122.01	0.90 e+05	1,367.58	1.01 e+05	14,400
20	0.75	8	**0.98 e+05**	1,776.35	0.99 e+05	3,183.02	1.01 e+05	2,031.30	1.11 e+05	14,400
20	0.75	9	**0.99 e+05**	1,779.24	1.02 e+05	4,046.86	1.02 e+05	1,470.36	1.05 e+05	14,400
30	0.5	0	1.97 e+05	8,131.35	1.97 e+05	9,303.11	**1.84 e+05**	6,661.51	1.98 e+05	14,400
30	0.5	1	2.09 e+05	6,741.38	**2.05 e+05**	8,337.86	2.06 e+05	6,223.99	2.36 e+05	14,400
30	0.5	2	1.62 e+05	6,310.00	1.63 e+05	5,466.08	**1.61 e+05**	2,403.93	1.80 e+05	14,400
30	0.5	3	**1.76 e+05**	7,258.44	1.77 e+05	8,490.38	1.83 e+05	5,364.12	2.38 e+05	14,400
30	0.5	4	3.04 e+05	8,005.14	3.22 e+05	8,665.91	**2.89 e+05**	6,750.19	3.49 e+05	14,400
30	0.5	5	**1.88 e+05**	7,435.17	1.88 e+05	7,963.70	1.91 e+05	4,415.26	2.26 e+05	14,400
30	0.5	6	1.88 e+05	7,244.06	1.95 e+05	8,679.78	**1.88 e+05**	5,036.13	2.33 e+05	14,400
30	0.5	7	**3.82 e+05**	8,138.68	4.05 e+05	9,202.13	4.20 e+05	4,607.10	4.39 e+05	14,400
30	0.5	8	**1.51 e+05**	5,786.28	1.52 e+05	7,515.79	1.57 e+05	4,728.17	1.86 e+05	14,400
30	0.5	9	2.78 e+05	7,180.90	2.73 e+05	7,172.85	**2.59 e+05**	4,259.48	3.30 e+05	14,400
30	0.75	0	1.04 e+05	5,948.39	**1.02 e+05**	11,937.47	1.05 e+05	3,976.56	1.15 e+05	14,400
30	0.75	1	1.23 e+05	5,546.67	**1.22 e+05**	10,765.75	1.28 e+05	3,371.02	1.37 e+05	14,400
30	0.75	2	0.98 e+05	7,392.68	**0.96 e+05**	12,556.06	1.01 e+05	3,642.89	1.14 e+05	14,400
30	0.75	3	1.03 e+05	6,083.00	**0.99 e+05**	12,773.28	1.04 e+05	2,800.32	1.08 e+05	14,400
30	0.75	4	0.98 e+05	6,049.69	**0.95 e+05**	12,036.30	1.02 e+05	3,160.86	1.15 e+05	14,400
30	0.75	5	1.04 e+05	6,026.44	**1.02 e+05**	13,553.03	1.06 e+05	3,695.71	1.13 e+05	14,400
30	0.75	6	0.99 e+05	6,660.11	**0.98 e+05**	11,316.65	0.99 e+05	3,509.40	1.11 e+05	14,400
30	0.75	7	1.07 e+05	6,519.37	**1.05 e+05**	11,515.46	1.09 e+05	3,668.31	1.27 e+05	14,400
30	0.75	8	**0.93 e+05**	6,524.25	0.94 e+05	11,144.66	0.96 e+05	3,562.64	1.06 e+05	14,400
30	0.75	9	0.89 e+05	7,337.74	**0.88 e+05**	12,763.55	0.93 e+05	4,114.14	1.04 e+05	14,400

both directions. The thick pink and the thin grey links correspond to the first and second categories of edges, respectively. The yellow squares (highlighted with 'E') represent entry/exit nodes.

Yabe et al. (2019) studied a cross-comparative analysis of individuals' evacuation behavior in post-earthquake. Their results indicate that after an earthquake with a 7 seismic intensity, nearly 60% of individuals strive to evacuate the region immediately. This result, however, neglects the earthquake hitting time, which can affect the magnitude of the evacuation. Therefore, jointly considering such an evacuation ratio with uncertainties in evacuees' travel volume, we generated three instances on the same network while varying the evacuation level. Considering the population size pop_i provided in the Istanbul Metropolitan Municipality (2023) report and distributed over nodes i, 4-person vehicle capacities indicating an agent entity and the evacuation ratio reported by Yabe et al. (2019), we assume that among the three instance sets 1, 2, and 3, the number of evacuating vehicles d_i emanating from each node i is measured by $\lfloor 0.25p \times U[0.1, 0.3] \rfloor$, $\lfloor 0.25p \times U[0.3, 0.5] \rfloor$, and $\lfloor 0.25p \times U[0.5, 0.7] \rfloor$, respectively. The uniform distribution U accounts for evacuation randomness and captures different evacuation ratios depending on the occurrence time of an earthquake with a seismic intensity of 7. In the designated network, 154 nodes exist with varying evacuation levels, and 69 nodes require an FR visit, taking into account the estimated severely damaged areas (Istanbul Metropolitan Municipality 2023) among the nodes that are reachable via thick FR lane hosting road segments.

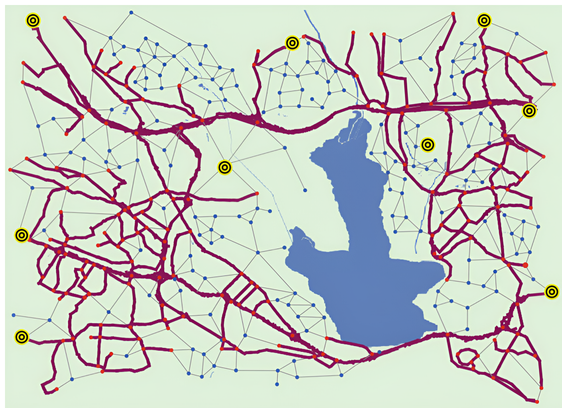


Figure 1 The network associated with case study instances. FR lanes can be reserved along the thick edges. The nine circles represent entry/exit nodes.

6.2.2. Comparing GAGA and B&B

Considering the traffic equilibrium, we generate feasible paths from each FR-demand node to the exit nodes using SA. The graph density in these instances is sufficiently low so that any cyclic paths obtained in the feasible solutions can be simply eliminated in a post-processing step using Dijkstra's algorithm. So, for these instances, we suppress the constraints (2d) in the QUBO model. Subsequently, we construct the partial Graver basis for both FR and evacuee paths. We compare the 'GAGA-SA-10000' with the B&B results. Four different GAGA settings are discussed below.

1. *Unnormalized*: The evacuation level remains the same. The inner problem optimization is terminated when no further augmentation is possible at that solution.

2. *Unnormalized, Tol = 1e-3*: The evacuation level remains the same. The Graver walk in the inner problem is terminated when the fractional difference in optima before and after one full pass over all the Graver elements is less than 10^{-3}.

3. *Normalized*: The evacuation level is normalized such that the maximum value is 100. The inner problem optimization is terminated when no further augmentation is possible at that solution.

4. *Normalized, Tol=1e-3*: The evacuation level is normalized such that the maximum value is 100. The tolerance before and after one full pass over all the Graver elements is less than 10^{-3}.

Considering the instance sizes, B&B and GAGA are given a time limit of 72 hours. We note that a pre-Graver-walk time is incorporated for each of the four settings when computing the initial feasible solutions and partial Graver bases while the B&B runtime starts from 0. Except for the *Unnormalized* setting, the total runtime of GAGA for the other three settings is less than B&B. For the *Unnormalized* setting, the total runtime for the first instance is less than B&B. In the other two instances, however, the Graver augmentation time increases due to increased evacuation levels. Hence, the Graver walk is not fully completed within the time limit for all initial seeds. The results comparing each of the four settings and comparing GAGA-SA-10000 with B&B are provided in Tables 2 and 3, where the objective function (total travel times) is in units of minutes. The performance of the algorithms over time for the three instances is plotted in Figure 2. The initial seeds are ordered both "randomly" or "improving" among the SA solutions. In the latter, the seed with the most inferior optimal solution comes first, and the optimal solutions improve as we proceed to the next seeds.

Table 2 ***Unnormalized* v.s. *Unnormalized,Tol=1e-3*: GAGA-SA-10000 and B&B results on case study instances**

Instance			Unnormalized				Unnormalized, Tol = 1e-3			
			GAGA-SA-10000		B&B		GAGA-SA-10000		B&B	
n	m	i	obj (minutes)	time (s)	obj (minutes)	time (s)	obj (minutes)	time (s)	obj (minutes)	time (s)
179	234	1	1.54 e+06 (M=100)	236439	1.55 e+06	259200	1.43 e+06 (M=100)	97056	1.55 e+06	259200
179	234	2	2.56 e+07 (M=85)	303312	2.56 e+07	259200	2.46 e+07 (M=85)	99077	2.56 e+07	259200
179	234	3	1.16 e+08 (M=43)	301813	1.23 e+08	259200	1.01 e+08 (M=43)	140761	1.23 e+08	259200

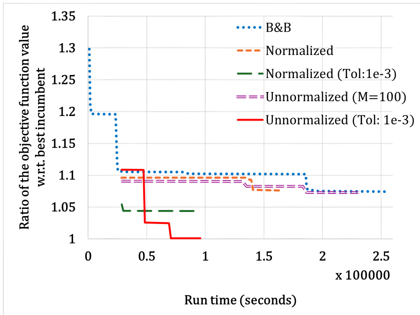
Table 3 ***Normalized* v.s. *Normalized,Tol=1e-3*: GAGA-SA-10000 and B&B results on case study instances**

Instance			Normalized				Normalized, Tol = 1e-3			
			GAGA-SA-10000		B&B		GAGA-SA-10000		B&B	
			obj		obj			obj		
n	m	i	(minutes)	time (s)	(minutes)	time (s)	(minutes)	time (s)	(minutes)	time (s)
179	234	1	1.55 e+06 (M=100)	166785	1.55 e+06	259200	1.50 e+06 (M=100)	95280	1.55 e+06	259200
179	234	2	2.59 e+07 (M=85)	195092	2.56 e+07	259200	2.46 e+07 (M=85)	98936	2.56 e+07	259200
179	234	3	1.13 e+08 (M=43)	244233	1.23 e+08	259200	1.03 e+08 (M=43)	118397	1.23 e+08	259200

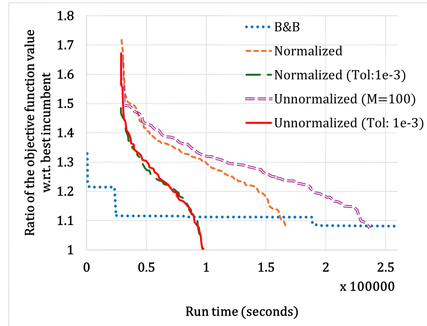
Figure 2 visualizes the performance results for the case study instances. We observe that the optimal solution from B&B is similar to the GAGA result for the *Unnormalized* and *Normalized* settings, but nearly $\sim 10\%$ larger when a tolerance threshold is included. Comparing among the GAGA runs, the time to solution is much longer for the *Unnormalized* and *Normalized* settings without tolerance. The best GAGA results are obtained with the *Unnormalized* and *Normalized* settings with a tolerance threshold and outperform B&B. Thus, we can conclude that running GAGA in an *Unnormalized* or *Normalized* setting and including a tolerance criterion in the inner user equilibrium problem tends to produce high-quality solutions in a shorter time compared to B&B.

6.2.3. Effect of FR-demand level on evacuation times

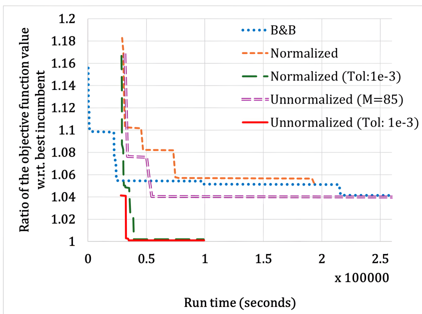
To examine the impact of increasing the number of FR-demand nodes (set F), which signifies the severity of a disaster, we generated 10 samples for each combination of $n \in \{20, 30\}$ and $p \in \{0.5, 0.75\}$, resulting in a total of 40 samples. These samples are based on the synthetic instance with ID 0, with the cardinality of set F gradually increasing. The results are illustrated in Figure 3, where the x-axis indicates the percentage of nodes with FR-demand, and the y-axis represents the total evacuation



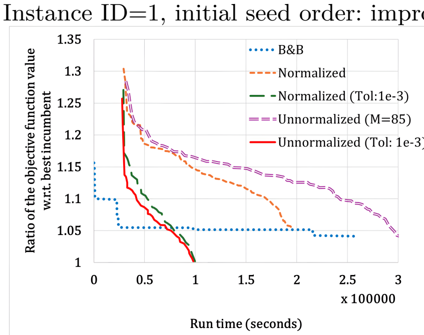
(a) Instance ID=1, initial seed order: random



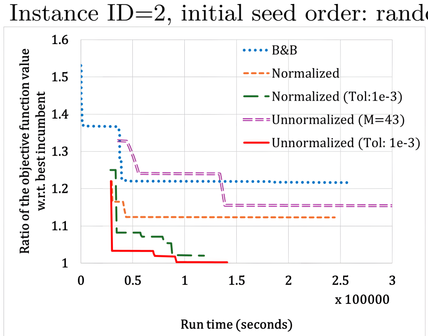
(b) Instance ID=1, initial seed order: improving



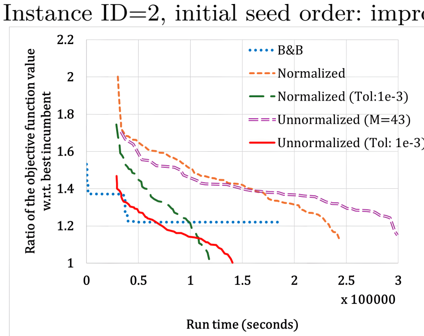
(c) Instance ID=2, initial seed order: random



(d) Instance ID=2, initial seed order: improving



(e) Instance ID=3, initial seed order: random



(f) Instance ID=3, initial seed order: improving

Figure 2 Comparison of GAGA (under four different Graver walks) with B&B for case instances

time. As shown in Figure 3, our objective function, i.e., the total evacuation time, rises in almost an S-shaped curve as the number of FR-demand nodes increases where the objective function value increased significantly after 50% FR-demand level amongst the nodes on the map.

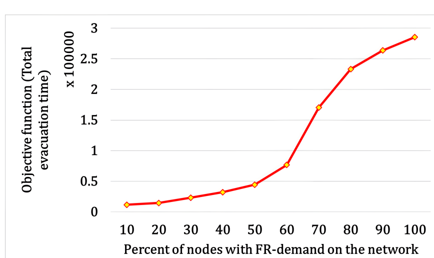


Figure 3 Effect of varying FR-demand as a result of disaster damage level on the total evacuation time (seconds) on the synthetic graph instance 0 averaged over $n \in \{20, 30\}$ and $p \in \{0.5, 0.75\}$ settings

6.3. Insights

We point out several critical insights from the multi-faceted analysis in Section 6 and Appendices F and G. First, the lane reservation decisions, as depicted in green in Figure 4, are remarkably consistent across the three Avcilar instances, regardless of variations in evacuation level. This consistency suggests that FRNDP strategies, driven by a single reservation plan, remain robust under different disaster scenarios and magnitudes. Second, among the solutions of the three Avcilar instances, FR lanes are reserved on approximately 60% of all potential FR road segments. This high percentage underscores the critical role of FRNDP in effective DM preparedness and response planning. Expectedly, all highways and primary road segments in the study region were selected for FR lane reservation. This outcome validates the proposed model and emphasizes the importance of prioritizing these segments for lane reservations. Third, we observed that the total evacuation time follows an S-shaped curve as the number of FR-demand nodes increases. Notably, the objective function experienced a sharp rise after the share of nodes having FR-demand across the map exceeded 50%. Hence, besides optimization of FR-lane reservation, proactive reinforcement of vulnerable areas in the network to avoid increased total evacuation times is imperative. Lastly, ignoring individuals' selfish routing behavior during the network design phase to simplify computations can significantly increase total evacuation times by up to 5.5% in certain scenarios, with evacuation times for some regions increasing by up to 10%. This implies the importance of accounting for evacuees' selfish routing behavior in a bi-level model for choosing lane reservations.

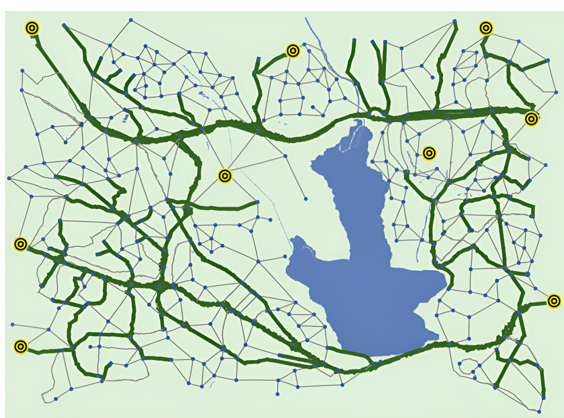


Figure 4 **FR lane reservation plan for case study instances**

7. Conclusion

Following a disaster, the immediate priority lies in locating and rescuing individuals who may be trapped or injured while ensuring others can travel to safe locations independently. An effective mechanism is to pre-determine and publicly pre-communicate a set of lanes restricted to FRs. However, identifying such roads requires a careful assessment of the tradeoff between effective FR responses and reducing the travel time for disaster-affected people. Our studied FRNDP is formulated as a bi-level optimization problem, where the arcs reserved for the FRs are selected at the first level to minimize total travel time, which depends on the second-level problem where the public travels between their origin-destination pairs over the eligible roads, adopting a selfish routing principle. Unlike the existing methodologies for similar network design problems, we develop a a novel quantum-inspired solution algorithm: Nested GAMA within the GAMA algorithm, called GAGA. We test several variations of GAGA on synthetic graph instances and three realistic scenarios for a predicted Istanbul earthquake. In most instances, we obtain superior solution quality and run-time performance of GAGA compared to a state-of-the-art exact algorithm for a traditional formulation.

We hope that our promising results encourage applying quantum-inspired methods to other disaster preparedness problems, and explore further development of this novel methodology in other application domains. In this context, the results of our study are important to show that the proposed approach - primarily implemented on classical computers due to limitations in current quantum computing environments - can provide high-quality solutions. We envision that, with future hardware advancements that enable quantum advantages, this alternative approach could also demonstrate significant computational benefits to generate better solutions much faster.

Acknowledgements. Sridhar Tayur and Ananth Tenneti acknowledge Raytheon BBN for their support through a CMU-BBN contract as part of a DARPA project on Quantum-Inspired Classical Computing (QuICC). Anthony Karahalios is supported by the National Science Foundation Graduate Research Fellowship Program under Grant No. DGE1745016, DGE2140739. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

References

Afkham M, Ramezanian R, Shahparvari S (2022) Balancing traffic flow in the congested mass self-evacuation dynamic network under tight preparation budget: An Australian bushfire practice. *Omega* 111:102658.

Alexander DA, Klein S (2009) First responders after disasters: a review of stress reactions, at-risk, vulnerability, and resilience factors. *Prehospital and Disaster Medicine* 24(2):87–94.

Alghassi H, Dridi R, Robertson AG, Tayur S (2019a) Quantum and quantum-inspired methods for de novo discovery of altered cancer pathways. *bioRxiv* 845719.

Alghassi H, Dridi R, Tayur S (2019b) GAMA: a novel algorithm for non-convex integer programs. *arXiv preprint arXiv:1907.10930* .

Alghassi H, Dridi R, Tayur S (2019c) Graver bases via quantum annealing with application to non-linear integer programs. *arXiv preprint arXiv:1902.04215* .

Ata B, Skaro A, Tayur S (2017) Organjet: Overcoming geographical disparities in access to deceased donor kidneys in the United States. *Management Science* 63(9):2776–2794.

Bagloee SA, Sarvi M, Patriksson M (2017) A hybrid branch-and-bound and Benders decomposition algorithm for the network design problem. *Computer-Aided Civil and Infrastructure Engineering* 32(4):319–343.

Bard JF (2013) *Practical bilevel optimization: algorithms and applications*, volume 30 (Springer Science & Business Media).

Beckmann M, McGuire CB, Winsten CB (1956) Studies in the economics of transportation. Technical report, Yale University Press.

Chang CJ, Chang SH (1993) A heuristic algorithm for solving the discrete network design problem. *Transportation Planning and Technology* 17(1):39–50.

Cheng Q, Chen Y, Liu Z (2022) A bi-level programming model for the optimal lane reservation problem. *Expert Systems with Applications* 189:116147.

Chou JB, Bramhavar S, Ghosh S, Herzog W (2019) Analog coupled oscillator based weighted Ising machine. *Scientific Reports* 9.

Dafermos S (1980) Traffic equilibrium and variational inequalities. *Transportation Science* 14(1):42–54.

De Loera JA, Hemmecke R, Onn S, Rothblum U, Weismantel R (2009) Convex integer maximization via Graver bases. *Journal of Pure and Applied Algebra* 213(8):1569–1577.

Duvar English: Turkey's own independent gazette (2023) Istanbul to be hit by major quake in next 7 years with 64% chance, Kandilli Observatory professor warns. last accessed 06.14.2024.

Erdik M, Durukal E (2008) Earthquake risk and its mitigation in Istanbul. *Natural Hazards* 44:181–197.

Fan W, Liu Q, Wang M (2021) Bi-level multi-objective optimization scheduling for regional integrated energy systems based on quantum evolutionary algorithm. *Energies* 14(16):4740.

Fang Y, Chu F, Mammar S, Zhou M (2011) Optimal lane reservation in transportation network. *IEEE Transactions on Intelligent Transportation Systems* 13(2):482–491.

Farvaresh H, Sepehri MM (2011) A single-level mixed integer linear formulation for a bi-level discrete network design problem. *Transportation Research Part E: Logistics and Transportation Review* 47(5):623–640.

Farvaresh H, Sepehri MM (2013) A branch and bound algorithm for bi-level discrete network design problem. *Networks and Spatial Economics* 13:67–106.

Fontaine P, Minner S (2014) Benders decomposition for discrete–continuous linear bilevel problems with application to traffic network design. *Transportation Research Part B: Methodological* 70:163–172.

Fontaine P, Minner S (2018) Benders decomposition for the hazmat transport network design problem. *European Journal of Operational Research* 267(3):996–1002.

Galindo G, Batta R (2013) Review of recent developments in OR/MS research in disaster operations management. *European Journal of Operational Research* 230(2):201–211.

Gao Z, Wu J, Sun H (2005) Solution algorithm for the bi-level discrete network design problem. *Transportation Research Part B: Methodological* 39(6):479–495.

Glover FW, Kochenberger GA, Du Y (2018) Quantum Bridge Analytics I: a tutorial on formulating and using QUBO models. *4OR* 17:335 – 371.

Goltz JD, Mileti DS (2011) Public response to a catastrophic Southern California earthquake: a sociological perspective. *Earthquake Spectra* 27(2):487–504.

Graver JE (1975) On the foundations of linear and integer linear programming I. *Mathematical Programming* 9:207–226.

Harris R, Sato Y, Berkley AJ, Reis M, Altomare F, Amin M, Boothby K, Bunyk P, Deng C, Enderud C, et al. (2018) Phase transitions in a programmable quantum spin glass simulator. *Science* 361(6398):162–165.

Hemmecke R, Onn S, Weismantel R (2011) A polynomial oracle-time algorithm for convex integer minimization. *Mathematical Programming* 126:97–117.

Istanbul Metropolitan Municipality (2023) AVCILAR - Deprem ve Zemin Inceleme Müdürlüğü. last accessed 06.14.2024.

Janson BN, Husaini A (1987) Heuristic ranking and selection procedures for network design problems. *Journal of Advanced Transportation* 21(1):17–46.

Jin WL (2015) Advances in Dynamic Traffic Assignment: TAC: A New Relationship Between Wardrop's User Equilibrium and Nash Equilibrium. *Networks and Spatial Economics* 15:617–634.

Kamyabniya A, Sauré A, Salman FS, Bénichou N, Patrick J (2024) Optimization models for disaster response operations: a literature review. *OR Spectrum* 1–47.

Karahalios A, Tayur S, Tenneti A, Pashapour A, Salman FS, Yildiz B (2024) A Quantum Inspired Bi-level Optimization Algorithm for the First Responder Network Design Problem. URL http://dx.doi.org/10.1287/ijoc.2024.0574.cd, available for download at https://github.com/INFORMSJoC/2024.0574.

Kimms A, Maassen KC (2012) Cell-transmission-based evacuation planning with rescue teams. *Journal of Heuristics* 18:435–471.

King AD, Carrasquilla J, Raymond J, Ozfidan I, Andriyash E, Berkley A, Reis M, Lanting T, Harris R, Altomare F, et al. (2018) Observation of topological phenomena in a programmable lattice of 1,800 qubits. *Nature* 560(7719):456–460.

Kirkpatrick S, Gelatt Jr CD, Vecchi MP (1983) Optimization by simulated annealing. *Science* 220(4598):671–680.

LeBlanc LJ (1975) An algorithm for the discrete network design problem. *Transportation Science* 9(3):183–199.

LeBlanc LJ, Morlok EK, Pierskalla WP (1975) An efficient approach to solving the road network equilibrium traffic assignment problem. *Transportation Research* 9(5):309–318.

Lee J, Onn S, Romanchuk L, Weismantel R (2010) The quadratic Graver cone, quadratic integer minimization, and extensions. *Mathematical Programming* 136:301–323.

Lerner EB, Moscati RM (2001) The golden hour: scientific fact or medical "urban legend"? *Academic Emergency Medicine* 8(7):758–760.

Li T, Wang N, Jiang B, Zhang M (2022) A bi-objective lane reservation problem considering dynamic traffic flow. *IEEE Transactions on Intelligent Transportation Systems* 24(1):367–381.

Lu S, Nie YM (2010) Stability of user-equilibrium route flow solutions for the traffic assignment problem. *Transportation Research Part B: Methodological* 44(4):609–617.

Luathep P, Sumalee A, Lam WH, Li ZC, Lo HK (2011) Global optimization method for mixed transportation network design problem: a mixed-integer linear programming approach. *Transportation Research Part B: Methodological* 45(5):808–827.

Lucas A (2014) Ising formulations of many NP problems. *Frontiers in Physics* 2:5.

McLoughlin D (1985) A framework for integrated emergency management. *Public Administration Review* 45:165–172.

McMahon PL, Marandi A, Haribara Y, Hamerly R, Langrock C, Tamate S, Inagaki T, Takesue H, Utsunomiya S, Aihara K, et al. (2016) A fully programmable 100-spin coherent Ising machine with all-to-all connections. *Science* 354(6312):614–617.

METU (2023) *Preliminary Reconnaissance Report on February 6, 2023, Pazarcık Mw=7.7 and Elbistan Mw=7.6, Kahramanmaraş-Türkiye Earthquakes* (METU - Earthquake Engineering Research Center).

Mohseni N, McMahon PL, Byrnes T (2022) Ising machines as hardware solvers of combinatorial optimization problems. *Nature Reviews Physics* 4(6):363–379.

Murota K, Saito H, Weismantel R (2004) Optimality criterion for a class of nonlinear integer programs. *Operations Research Letters* 32(5):468–472.

Nikoo N, Babaei M, Mohaymany AS (2018) Emergency transportation network design problem: Identification and evaluation of disaster response routes. *International Journal of Disaster Risk Reduction* 27:7–20.

Parsons T, Toda S, Stein RS, Barka A, Dieterich JH (2000) Heightened odds of large earthquakes near Istanbul: an interaction-based probability calculation. *Science* 288(5466):661–665.

Peeta S, Salman FS, Gunnec D, Viswanath K (2010) Pre-disaster investment decisions for strengthening a highway network. *Computers & Operations Research* 37(10):1708–1719.

Pottier L (1996) The Euclidean algorithm in dimension n. *Proceedings of the 1996 International Symposium on Symbolic and Algebraic Computation*, 40–42.

Rey D (2020) Computational benchmarking of exact methods for the bilevel discrete network design problem. *Transportation Research Procedia* 47:11–18.

Shangin RE, Pardalos P (2016) Heuristics for the network design problem with connectivity requirements. *Journal of Combinatorial Optimization* 31(4):1461–1478.

Shariat Mohaymany A, Nikoo N (2020) Designing large-scale disaster response routes network in mitigating earthquake risk using a multi-objective stochastic approach. *KSCE Journal of Civil Engineering* 24(10):3050–3063.

Sturmfels B, Thomas RR (1997) Variation of cost functions in integer programming. *Mathematical Programming* 77:357–387.

The European Union Delegation in Türkiye and the EU Joint Research Centre (2023) *Türkiye Earthquakes Recovery and Reconstruction Assessment* (Government of Türkiye).

US Bureau of Public Roads Office of Planning Urban Planning Division (1964) *Traffic assignment manual for application with a large, high speed computer* (US Department of Commerce).

Venturelli D, Kondratyev A (2019) Reverse quantum annealing approach to portfolio optimization problems. *Quantum Machine Intelligence* 1(1-2):17–30.

Wang DZ, Liu H, Szeto WY (2015) A novel discrete network design problem formulation and its global optimization solution algorithm. *Transportation Research Part E: Logistics and Transportation Review* 79:213–230.

Wang S, Meng Q, Yang H (2013) Global optimization methods for the discrete network design problem. *Transportation Research Part B: Methodological* 50:42–60.

Wang T, Roychowdhury J (2019) OIM: Oscillator-based Ising machines for solving combinatorial optimisation problems. *Unconventional Computation and Natural Computation: 18th International Conference, UCNC 2019, Tokyo, Japan, June 3–7, 2019, Proceedings 18*, 232–256 (Springer).

World Bank and Global Facility for Disaster Reduction and Recovery (2023) *Global Rapid Post-Disaster Damage Estimation (GRADE) Report: February 6, 2023 Kahramanmaraş Earthquakes - Türkiye Report* (World Bank).

Wu P, Chu F, Che A, Zhao Y (2018) Dual-objective optimization for lane reservation with residual capacity and budget constraints. *IEEE Transactions on Systems, Man, and Cybernetics: Systems* 50(6):2187–2197.

Wu Y, Chu C, Chu F, Wu N (2009) Heuristic for lane reservation problem in time constrained transportation. *2009 IEEE International Conference on Automation Science and Engineering*, 543–548 (IEEE).

Xie C, Turnquist MA (2009) Integrated evacuation network optimization and emergency vehicle assignment. *Transportation research record* 2091(1):79–90.

Yabe T, Sekimoto Y, Tsubouchi K, Ikemoto S (2019) Cross-comparative analysis of evacuation behavior after earthquakes using mobile phone data. *PLOS One* 14(2):e0211375.

Yan X, Lv N, Liu Z, Xu K (2008) Quantum-inspired evolutionary algorithm for transportation network design optimization. *2008 Second International Conference on Genetic and Evolutionary Computing*, 189–192 (IEEE).

Yen JY (1971) Finding the k shortest loopless paths in a network. *Management Science* 17(11):712–716.

Yin R, Xing J, Mo P, Zheng N, Liu Z (2022) BO-B&B: A hybrid algorithm based on Bayesian optimization and branch-and-bound for discrete network design problems. *Electronic Research Archive* 30(11):3993–4014.

Yücel E, Salman FS, Arsik I (2018) Improving post-disaster road network accessibility by strengthening links against failures. *European Journal of Operational Research* 269(2):406–422.

Zhang T, Niu C, Nair DJ, Robson EN, Dixit V (2023) Transportation resilience optimization from an economic perspective at the pre-event stage. *Sustainability Analytics and Modeling* 3:100027.