Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
TLQP: Early-stage transportation lock-down and quarantine problem

Yida Ding\textsuperscript{a}, Sebastian Wandelt\textsuperscript{b}, Xiaoqian Sun\textsuperscript{a,b,*}

\textsuperscript{a} School of General Engineering, Beihang University, 100191 Beijing, China
\textsuperscript{b} School of Electronic and Information Engineering, Beihang University, 100191 Beijing, China

\textbf{ARTICLE INFO}

\textbf{Keywords:}
Pandemic
Transportation
Lock-down
Prevention
COVID-19

\textbf{ABSTRACT}

The advent of COVID-19 is a sensible reminder of the vulnerability of our society to pandemics. We need to be better prepared for finding ways to stem such outbreaks. Except from social distancing and wearing face masks, restricting the movement of people is one important measure necessary to control the spread. Such decisions on the lock-down/reduction of movement should be made in an informed way and, accordingly, modeled as an optimization problem. We propose the Early-stage Transportation Lock-down and Quarantine Problem (TLQP), which can help to decide which parts of the transportation infrastructure of a country should be restricted in early stages. On top of the network-based Susceptible-Exposed-Infectious-Recovered (SEIR) model, we establish a decision recommendation framework, which considers the lock-down of cross-border traffic, internal traffic, and movement inside individual populations. The combinatorial optimization problem aims to find the best set of actions which minimize the social cost of a lock-down. Given the inherent intractability of this problem, we develop a highly-efficient heuristic based on the Effective Distance (ED) path and the Cost-Effective Lazy Forward (CELF) algorithm. We perform and report experiments on the global spread of COVID-19 and show how individual countries may protect their population by taking appropriate measures against the threatening pandemic. We believe that our study contributes to the orchestration of measures for dealing with current and future epidemic outbreaks.

1. Introduction

The increased mobility of our society, together with enhanced connectivity and efficiency of transportation systems has become a double-edged sword. While it is easier than ever to travel long distances within short time, we can reach the remotest destinations within one day of traveling at affordable prices (Janic, 2000; Lee et al., 2009; Diaconu, 2012; Kotegawa et al., 2014), global mobility has also significantly contributed to the risk of spreading diseases worldwide. Research on the evolution of previous epidemics has revealed that air transportation plays a crucial role in the initial spread of a disease (Brockmann and Helbing, 2013; Sun et al., 2020); for instance, three high-impact outbreaks in the last 20 years cover SARS in 2003 (Likhacheva, 2006), MERS in 2012 (Zaki et al., 2012), and Ebola in 2014 (Bogoch et al., 2015). These diseases have been harmful to the public, but they remained rather at a local scale, mostly due to epidemiological properties of the disease and timely reactions of local authorities. COVID-19, however, has turned into a fully blown pandemic. The disease was first observed in December 2019 around Wuhan and is caused by Severe Acute Respiratory...
Syndrome Coronavirus 2 (SARS-CoV-2) (Andersen et al., 2020). By March 11th, 2020, the World Health Organization declared COVID-19 as a pandemic (World Health Organization, 2020c). As of April 9th, 2021, there are 132.7 million confirmed cases and more than 2.8 million fatalities related to COVID-19. The long-term effects of COVID-19 are still unforeseeable, but the impact in recent months is remarkable. The pandemic is estimated to have caused the largest global recession since the severe worldwide economic downturn in the 1930s (Great Depression), with millions of people falling into extreme poverty (Sumner et al., 2020).

The reactions to COVID-19 in different regions are various and (seemingly) uncoordinated; mainly due to the lack of methods/tools to efficiently recommend appropriate measures and incomplete information about the actual virus. Despite knowing the potential role of air transportation in disease spreading, countries were reluctant to close their borders in a coordinated manner. Avoiding to import cases is the best way to prevent a local outbreak (Organization et al., 2020). Naturally, air travel restrictions are most effective at the early stage of an epidemic; researchers have found that the restrictions of travel from Wuhan came too late (Kraemer et al., 2020; Chinazzi et al., 2020). Once the disease had spread into other countries and regions, restricting air travel was just not helpful to prevent infections taking place locally, e.g., via public transportation (Sipetas et al., 2020; Ceder and Jiang, 2020; Sun et al., 2021). Again, it is important to stress that it is about the right time to impose travel bans and other stricter measures; such as wearing face masks, extensive hand washing, social distancing, increased monitoring and potentially self-quarantine.

An extensive number of studies have explored metapopulation models to investigate the relationship between mobility and epidemic spreading (Chinazzi et al., 2020; Brockmann and Helbing, 2013; Badr et al., 2020; Gardner et al., 2012b; Lai et al., 2020). Different mobility data at a wide range of scales, including air and railway transportation data (Merler and Ajelli, 2010), census data on commuting patterns (Balcan et al., 2009) and aggregated mobile phone data (Badr et al., 2020) have been collected. Upon aggregation, these mobility data can help refine interventions by anticipating changes of human movement patterns. Tian et al. (2020) found that the Wuhan lock-down led to the delayed arrival of COVID-19 in other cities by 2.91 days based on human movement data and case reports. Furthermore, Badr et al. (2020) devised a social distancing metric by aggregated cell phone data and it was found that mobility patterns are strongly correlated with decreased COVID-19 case growth. Dalziel et al. (2013) showed that variation of mobility patterns in city level can lead to different infectious disease dynamics among cities. However, a gap remains in the literature considering real-time predictive models that directly contribute to the decision making of policy makers among a set of potential mitigation strategies. The aforementioned studies that investigate the relationship of mobility and epidemics focus on the consequence (in terms of epidemiology) of an existing mitigation policy. Currently, few studies provide a decision recommendation framework for allocating policies to at-risk regions in the context of an outbreak. Chen et al. (2016) proposed a metapopulation epidemic model to evaluate and rank the control strategies based on their effectiveness in reducing the spread of outbreaks given limited budget. Based on various existing airport-rank metrics, their study identifies the optimal set of airports for deploying border control, i.e., passenger screening upon arrival at airports, subject to budget constraint. Our study, in a similar manner, aims to decide which parts of the transportation infrastructure of a country should be restricted to contain the epidemic, and find the best set of actions which minimizes the social cost of a lock-down.

In our study, we first propose a decision recommendation framework by illustrating the traditional Susceptible-Exposed-Infectious-Recovered (SEIR) model assigned to a network of vertices and units, extending the model to a formal decision problem of administrative units, and introducing possible mitigation strategies with different assigned costs. A highly-efficient heuristic is proposed to solve this combinatorial optimization problem. Based on this theoretical framework, we extend our model to the global level inspired by the Global Epidemic and Mobility Model (Balcan et al., 2009), with the adaptation of the definitions of vertices, administrative units, external and internal transportation edges into subpopulations, countries, international/domestic flight and commuting. We show the advantage of our approach by several detailed scenario analysis at the global scale. The major contribution of our study is summarized as follows:

1. We propose a novel and practical decision recommendation framework, Early-stage Transportation Lock-down and Quarantine Problem (TLQP), to help decide which parts of the transportation infrastructure of a country should be restricted in face of an outbreak.
2. We show the efficiency and efficacy of exploiting the Effective Distance (ED) path to accelerate the computation of standard SEIR model given a targeted administrative unit or country.
3. We show the computational complexity of TLQP, and develop an efficient and accurate heuristic based on the Cost-Effective Lazy Forward (CELF) algorithm.
4. We conduct real-world scenario analysis on Germany in face of COVID-19 and find that Germany should focus on the lock-down of international flights with country of origin at an early response stage, while intensify the lock-down of its domestic transportation system and inner-subpopulation mobility in face of the second wave of pandemic.

The remainder of this paper is organized as follows: Section 2 conducts a detailed literature review on epidemic modelling and related problems. Section 3 formally introduces TLQP and discusses its properties. Section 4 introduces two solution techniques for solving TLQP. Section 5 reports the experimental results of the solution techniques and applies them to the decision making of Germany in face of the first wave and second wave of global pandemic. Section 6 concludes this study and gives recommendations for future work.
2. Related work

2.1. Epidemic modelling

The increasing computational resources and data availability throughout recent years have spurred the development of epidemic models, especially in the forecasting and risk assessment of infectious disease outbreaks (Vespignani et al., 2020). These epidemic models can generally be categorized into two groups: agent-based models (Longini et al., 2005; Germann et al., 2006; Merler and Ajelli, 2010; Degli Atti et al., 2008; Ferguson et al., 2006; Stefanoff et al., 2010; Mamelund, 2011) and metapopulation models (Chinazzi et al., 2020; Hufnagel et al., 2004; Balcan et al., 2009; Rvachev and Longini Jr, 1985; Rader et al., 2020; Kraemer et al., 2020; Zlojutro et al., 2019; Gardner et al., 2016). We discuss the advantages of each group and their limitations below.

The agent-based models feature highly-detailed input data and high computational cost, with the goal to keep tracking the infection of single individuals considered in the model (Longini et al., 2005; Germann et al., 2006; Merler and Ajelli, 2010; Stefanoff et al., 2010; Mamelund, 2011; Takahashi et al., 2020; Lucas et al., 2020). Merler and Ajelli (2010) found that household groups and the fraction of single individuals considered in the model will greatly influence $R_0$ and cumulative attack rates in the course of pandemic influenza based on air and railway transportation data. Mamelund (2006) utilized individual-level data and household-level data, showing that apartment size and social status of place of residence influence mortality during Spanish influenza. However, the computational complexity of agent-based models restricted their use to country-level scenarios (Degli Atti et al., 2008), not exceeding continent level (Merler and Ajelli, 2010). The metapopulation models consider a network where connections among regions represent the individual fluxes given transportation and mobility infrastructures (Chinazzi et al., 2020; Hufnagel et al., 2004; Rader et al., 2020; Kraemer et al., 2020; Zlojutro et al., 2019; Gardner et al., 2016; Skog et al., 2014; Skog, 2014). Through thousands of stochastic realizations, metapopulation models can identify epidemic patterns at the worldwide scale. The Global Epidemic and Mobility Model (GLEaM) integrated high resolution census data worldwide, human mobility pattern and disease dynamics to predict anticipation of the spatio-temporal patterns of global epidemic spreading (Balcan et al., 2009). (Hufnagel et al., 2004) introduced a probabilistic model which integrates infection dynamics among individuals with aviation traffic in a worldwide network.

To model the worldwide spread of COVID-19 outbreak and introduce the Early-stage Transportation Lock-down and Quarantine Problem (TLQP), we use a metapopulation-based model to describe the spatial and temporal patterns of epidemic spreading. Although the agent-based models are capable of providing targeted interventions, their high demand on detailed input data and computing resources makes it difficult to model the disease spreading at the global level. The use of a metapopulation-based model has enabled us to simulate the mobility of infectious individuals on the global scale, and estimate the disease arrival timelines for cities or countries without assumptions on case importation (Balcan et al., 2009). In addition, we also analyze and discuss various epidemic scenarios where parts of the transportation infrastructure of a country are restricted, including the lock-down of cross-border traffic, internal traffic, and movement inside individual populations.

2.2. Related problems and methodology

To evaluate ongoing epidemic and implement a real-time control of infectious disease outbreaks, researchers mainly adopted approaches that represent the spreading process as a tree. Rey et al. (2016) sought the most likely infection tree that spans a set of known infected nodes. Their solution method utilized the network topology, estimated disease parameters and available infection reports to evaluate a region that has been exposed to infection. Gardner et al. (2012a) and Fajardo and Gardner (2013) modeled the contagion process by a social-contact network, and infer the most likely paths of infection through a contact network based on the assumption of partially available infection data. Vazquez (2004) represented an epidemic outbreak by its causal tree of infection transmission, where nodes represent infected agents and arcs represent the disease transmission between agents, to obtain the expected outbreak size.

The study on infection process has roots in various fields of research, especially medical sciences and sociology. In medical sciences, it was applied to study epidemic spreading and prevention strategies (Bóta et al., 2013; Bóta et al., 2014; Saad-Roy et al., 2020; Morris et al., 2015; Nie et al., 2020; Tang et al., 2020). In sociology, it was used to study information diffusion, which has been successfully applied to viral marketing (Domingos and Richardson, 2001), network monitoring (Leskovec et al., 2007) and rumor control (Borodin et al., 2010). The core algorithmic problem behind information diffusion is the well-known Influence Maximization (IM), which aims to find a set of $k$ users in an online social network to maximize influence spread (Kempe et al., 2003). Our proposed TLQP shares certain similarity with the IM problem, with the former aimed to minimize social-economic cost in face of an outbreak and the latter aimed to maximize the influence spread given limited influencers. Another problem called Targeted Immunization (TI) is also relevant, which aims to identify and immunize at-risk individuals in order to contain the disease (Cheng et al., 2020). The models studied in IM, e.g., Independent Cascade (IC) model and Linear Threshold (LT) model are also similar with network-based SEIR and ED-SEIR model in TLQP. The IC model considers a stochastic activation process based on the influence probability of each edge, and the LT model considers a user activation if a “sufficient” number of its incoming neighbors are active (Li et al., 2018). Similarly, the network-based SEIR and ED-SEIR model receive a mitigation strategy (i.e., a set of actions attached with rates) as the input, and conduct compartmental transitions and transportation connections to simulate the spatial and temporal spreading of epidemic. The Epidemic Spreading Index (ESI), which is defined to be the weighted sum of cumulative cases and growth rate of cumulative cases in specific administrative unit at the time horizon, is utilized to evaluate the effectiveness of the mitigation strategy. Thus, the aforementioned models all simulate a diffusion process and output a specific metric to evaluate the performance of the input set.

Kempe et al. (2003) proposed an approximate solution to IM based on greedy algorithm, which iteratively selects the node that
provides maximum marginal gain to the influence function. They further proved that the greedy algorithm guarantees a \((1 - \frac{1}{2} - \epsilon)\) bound to the optimal solution for diffusion models with monotonic and submodular objective functions. However, the number of Monte Carlo simulations in greedy algorithm is relatively large, leading to inefficiency. Based on the consideration to reduce number of simulations, the Cost-Effective Lazy Forward (CELF) algorithm (Leskovec et al., 2007) further exploits submodularity to find near-optimal node selections. It effectively estimates an upper bound of influence spread to prune the nodes with insignificant influence.

Leskovec et al. (2007) proved that the solutions obtained by CELF achieve at least a fraction of \(\frac{1}{4}(1 - \frac{1}{2})\) of the optimal solution and up to 700 times improvement in performance compared with the greedy algorithm.

In our study, we use Effective Distance (ED) path (Brockmann and Helbing, 2013) to simplify SEIR simulations and focus on the epidemic status of vertices within a specific administrative unit. The concept of effective distance reflects the phenomenon that large traffic flow is equivalent to a small distance, and vice versa. We have found that the disease arrival time of each vertex in the administrative unit obtained by ED-SEIR model has 0.20% mean deviation \((s = 0.0098)\) in terms of that obtained by the standard SEIR model, based on the experiments of six artificial examples. Since the cumulative cases of a vertex grow exponentially soon after the disease arrives (if no effective mitigation strategy is implemented), the disease arrival time indicated by the shortest path is the major influence of cumulative cases. Experimental results show that our ED-SEIR model maintains a high accuracy of Epidemic Spreading Index \((ESI)\) (less than 4% error) in terms of standard SEIR model, while significantly reduces the number of simulated vertices and edges, leading to a high-efficient simulation.

### 3. Problem formulation

This section formally introduces the Early-stage Transportation Lock-down and Quarantine Problem (TLQP), by illustrating the SEIR model on a network of vertices and edges (Section 3.1), extending the model to a formal decision problem of administrative units (Section 3.2), and introducing possible mitigation strategies (Section 3.3). An overview on the key parameters in our model is provided in Table 1.

Table 1

| Symbol | Description |
|--------|-------------|
| \(\tau, \theta, \mu\) | Transmission rate of disease, inverse of mean latency period and inverse of mean infectious period in SEIR model |
| \(F_i\) | Passenger flow travelling from vertex \(i\) to vertex \(j\) per day |
| \(F_m^{ij}(t)\) | Passenger flow of a specific compartment \(m\) travelling from vertex \(i\) to vertex \(j\) at time \(t\). \(m \in \{S,E,I,R\}\) |
| \(N_i(t), N_i^m(t)\) | Population of vertex \(i\) at time \(t\), population of vertex \(i\) of compartment \(m\) at time \(t\) |
| \(v_h\) | The disease origin vertex |
| \(u, u'\) | The administrative unit; the administrative unit of interest |
| \(V, V_u\) | The set of vertices; the set of vertices in administrative unit \(u\) |
| \(E, E_u\) | The set of edges; the set of edges in administrative unit \(u\) |
| \(V_u', E_u'\) | The set of vertices/edges passed along dominant spreading paths from disease origin to vertices in the administrative unit of interest \(u'\) |
| \(A\) | The set of actions for the administrative unit of interest \(u'\) |
| \(A_q\) | The set of actions that quarantine the vertices in \(u'\) |
| \(A_{EF}, A_{UT}\) | The set of actions that reduce the flow of important/unimportant transportation edges that lie within/beyond the dominant spreading paths from disease origin to vertices in \(V_u\) |
| \(x\) | The rate attached to an action |
| \(X\) | The set of rate alternatives |
| \((a, x)\) | An action-rate tuple, i.e., an action attached with a rate |
| \(\Omega(a, x)\) | The cost of an action-rate tuple \((a, x)\) |
| \(Y\) | The universe set of action-rate tuples under consideration in TLQP-H model, i.e., \(Y = \{(a, x)|a \in A_q \cup A_{EF}, x \in X \setminus \{0\}\}\) |
| \(M\) | Mitigation strategy that is composed of a set of action-rate tuples taken by \(u'\) |
| \(M^*, \tilde{M}\) | Optimal mitigation strategy obtained by TLQP-E model; best mitigation strategy obtained by TLQP-H model |
| \(S(t), E(t), I(t), R(t)\) | The number of Susceptible, Exposed, Infectious and Recovered/Removed individuals in vertex \(i\) at time \(t\) under the influence of mitigation strategy \(M\) |
| \(C_0^M(t)\) | The expected number of cumulative cases/the growth rate of cumulative cases of vertex \(i\) at time \(t\) under the influence of mitigation strategy \(M\) |
| \(t_0\) | A finite time horizon, i.e., the simulation end time of SEIR or ED-SEIR model |
| \(ESI(M, t_0)\) | The Epidemic Spreading Index at time \(t_0\) under mitigation strategy \(M\) |
| \(U_{ESI}\) | Upper limit of Epidemic Spreading Index |
3.1. Classic SEIR model

We begin with a classic compartmental model, the Susceptible-Exposed-Infectious-Recovered (SEIR) for human-to-human transmission (Kermack and McKendrick, 1927). The SEIR model divides the population of a region into four compartments: Susceptible compartment ($S$), Exposed compartment ($E$), Infectious compartment ($I$), and Recovered/Removed compartment ($R$). The epidemiological status of all individuals is initially set to susceptible, where it can contract the virus through contacts with individuals in the Infectious compartment and proceed to the Exposed compartment at the transmission rate ($\tau$). The exposed individuals proceed to the infectious stage with a rate inversely proportional to the mean latency period ($\theta$), and the infectious individuals proceed to the Recovered/Removed compartment with a rate inversely proportional to the mean infectious period ($\mu$) (Berger et al., 2020).

\[
\frac{dS(t)}{dt} = - \tau S(t)I(t) / N(t)
\]

\[
\frac{dE(t)}{dt} = - \theta E(t) I(t) / N(t) + \frac{\tau S(t)I(t)}{N(t)}
\]

Fig. 1. The schematic representation of SEIR compartmental transitions. Individuals in the Susceptible compartment can contract the virus through contacts with individuals in the Infectious compartment and proceed to the Exposed compartment at the transmission rate ($\tau$). The exposed individuals proceed to the infectious stage with a rate inversely proportional to the mean latency period ($\theta$), and the infectious individuals proceed to the Recovered/Removed compartment with a rate inversely proportional to the mean infectious period ($\mu$) (Berger et al., 2020).

Fig. 2. The spatial structure of an epidemic network (center) and its compartmental transitions for Example 3.1. The width of the transportation edge is proportionally correlated to the passenger flow. The disease is assumed to originate from vertex $\alpha_0$ with 0.5% of the population being initially infected at $t = 0$, i.e., being assigned to $I$ compartment of $\alpha_0$ (Grais et al., 2004). The disease arrival time and peak period of infections in each vertex are related to the distance from disease origin $\alpha_0$.

3.1. Classic SEIR model

We begin with a classic compartmental model, the Susceptible-Exposed-Infectious-Recovered (SEIR) for human-to-human transmission (Kermack and McKendrick, 1927). The SEIR model divides the population of a region into four compartments: Susceptible compartment ($S$), Exposed compartment ($E$), Infectious compartment ($I$), and Recovered/Removed compartment ($R$). The epidemiological status of all individuals is initially set to susceptible, where it can contract the virus through contacts with individuals in the Infectious compartment and proceed to the Exposed compartment at the transmission rate ($\tau$). The exposed individuals, who are infected but not yet infectious, proceed to the infectious stage with a rate inversely proportional to the mean latency period ($\theta$), and the infectious individuals proceed to the Recovered/Removed compartment with a rate inversely proportional to the mean infectious period ($\mu$). In particular, the model assumes no recurrence of infection, so that once the individuals change to the Recovered/Removed compartment, they can no longer contract the disease or infect other individuals. The situation of birth and death of individuals is not considered in our model. A schematic representation of the SEIR compartmental transition is visualized in Fig. 1. The dynamics of the four epidemiological states over time is represented by:

\[
\frac{dS(t)}{dt} = - \tau S(t)I(t) / N(t)
\]

\[
\frac{dE(t)}{dt} = - \theta E(t) I(t) / N(t) + \frac{\tau S(t)I(t)}{N(t)}
\]
\[ \frac{dI_i(t)}{dt} = \mu I_i(t) + \theta E_i(t) \]  
\[ \frac{dR_i(t)}{dt} = \mu I_i(t) \]  

Given a populated vertex \( i \), the SEIR model has the following features: (a) The time step \( dt \) is assumed to be a single day. (b) Population of vertex \( i \) at time \( t \) is given by the sum of individuals in four compartments, i.e., \( N_i(t) = S_i(t) + E_i(t) + I_i(t) + R_i(t) \). (c) Transmission rate \( \tau \) which is independent of time reflects the speed of individuals in \( S_i \) proceeding to \( E_i \), while \( I_i(t)/N_i(t) \) represents the density of infected individuals in the population. (d) The expected cumulative number of confirmed cases (abbreviated as cumulative cases) of a vertex \( i \) at time \( t \) is the sum of infectious and recovered, i.e., \( C_i(t) = I_i(t) + R_i(t) \), which indicates the total number of cases that have been diagnosed up to a certain point in time regardless of whether they have recovered.

Based on the SEIR model, which is tailored to one populated region, we consider a generic epidemic network, composed of a set of populated vertices \( V \) and a set of transportation edges \( E \) connecting the vertices. Note that \( E \) represents the movement of population with the time unit of a single day, covering all means of transportation, including flights and ground mobility. The weight of an edge represents the passenger flow per day. For the sake of simplicity, the edges in our model are undirected and, accordingly, the flow symmetric. The flow of a specific compartment \( m \) travelling from vertex \( i \) to vertex \( j \) at time \( t \) is given by:

\[ F_{ij}^{[m]}(t) = \frac{N_i^{[m]}(t)}{N_i(t)} F_{ij} \]  

where \( N_i^{[m]}(t) \) is the population of compartment \( m \) in vertex \( i \) at time \( t. N_i(t) > 0 \), and the time-independent flow \( F_{ij} \) is the flow travelling from vertex \( i \) to vertex \( j \) per day, which satisfies \( F_{ij} = \sum_{m} F_{ij}^{[m]}(t) \), with \( m \in \{ S, E, I, R \} \).

**Example 3.1. (Epidemic spreading)** In order to illustrate the proposed model, we use a numerical example where a vertex set \( V = \{a_0, a_1, \ldots, a_7\} \) is considered. The spatial structure of the example is visualized in the center of Fig. 2, where the width of the transportation edge is proportionally correlated to the passenger flow. For this example, we assume the disease originates from vertex \( a_0 \) with 0.5% of the population being initially infected at \( t = 0 \), i.e., being assigned to \( I \) compartment of \( a_0 \) (Grais et al., 2004).
3.2. Administrative units

In our model, the cumulative cases in each vertex grows exponentially shortly after the disease arrival. This is natural in the absence of any mitigation actions. In reality, however, the goal is to contain the disease spreading early, without hitting the exponential part of the curve (commonly known as Flatten the curve during COVID-19). Therefore, in the following we present our formal problem definition, which considers the implementation of mitigation actions. Specifically, we assume that a subset of the vertices is under the control of an administrative unit which aims to avoid an aggressive spread. An administrative unit has a set of actions, which are (a) reducing the flow of internal transportation, (b) reducing the flow of external transportation and (c) enforcing vertex-specific lockdowns. For the formal definitions of these actions, we assume that the set of vertices are assigned to a group of administrative units \( U \), with \( \psi(v) \) denoting the unique administrative unit of vertex \( v \in V \). The expression \( V_u \) denotes the set of vertices assigned to administrative unit \( u \in U \), i.e., \( V_u = \{ v | \psi(v) = u \} \).

**Definition 3.1. (Actions of an administrative unit)** Given a set of vertices \( V \), a set of edges \( E \), a partitioning function \( \psi \), and an administrative unit of interest \( u^* \in U \), the following actions are defined:

1. For each \( e \in E \), if \( e \) is an internal transportation edge between two vertices in \( u^* \), i.e., \( e = (v_i, v_j) (v_i, v_j \in V_{u^*}) \), then action \( \text{INT}(e) \) represents reducing the flow on \( e \).
2. For each \( e \in E \), if \( e \) is an external transportation edge between a vertex of \( u^* \) and another vertex of \( u^* \), i.e., \( e = (v_i, v_j) (v_i \in V_{u^*}, v_j \in V_{u^*}, u^* \neq u^*) \), then action \( \text{EXT}(e) \) represents reducing the flow on \( e \).
3. For each \( v \in V_{u^*} \), action \( \text{QUA}(v) \) represents reducing the transmission rate of \( v \); this action is at its essence equivalent to the quarantine of \( v \).

Let \( x \) denotes the rate attached to an action (\( x \in [0, 1] \)) and \( X \) denotes the set of rate alternatives. If \( x \neq 0 \), the corresponding action is applied, and vice versa. An action attached with a rate is denoted by a tuple \(( a, x)\).

**Example 3.2. (Actions of an administrative unit)** We extend Example 3.1 with a partitioning function \( \psi \) introducing administrative units. In particular, the vertices in Example 3.1 are divided into three administrative units, \( \alpha, \beta \), and \( \gamma \), with \( \psi = \{ u_i \rightarrow u | u \in \{ \alpha, \beta, \gamma \} \land i \in \{0, 1, 2\} \} \) (\( u_i \) represents vertices). This example with administrative units is shown in Fig. 3. Below, the administrative unit of interest is \( \beta \), thus we have the following actions for \( \beta \):

- **Actions on internal transportation edges:** \( \{ \text{INT}(\beta_0, \beta_1), \text{INT}(\beta_0, \beta_2), \text{INT}(\beta_1, \beta_2) \} \)
- **Actions on external transportation edges:** \( \{ \text{EXT}(\beta_1, \alpha_0), \text{EXT}(\beta_1, \alpha_1), \text{EXT}(\beta_2, \gamma_0) \} \)
- **Actions on intra-vertex movement:** \( \{ \text{QUA}(\beta_0), \text{QUA}(\beta_1), \text{QUA}(\beta_2) \} \)

Accordingly, the administrative unit \( \beta \) in our example has nine actions. Strong actions with large \( x \) should influence the flow of internal and external transportation (or degree of quarantine) to a large extent. In this study we take the view that a full lock-down, i.e., reducing the flow to zero is not possible in most countries; translating into some leakage of flow in the real world. The action rate will be \( 0 \) if the action is not applied. Thus, we assume the action rate to satisfy \( x \in \{0, 0.3, 0.6, 0.9\} \). For this example, the search space of action-rate tuples has a size of \( 4^9 \).

In order to avoid the spread within \( \beta \) growing exponentially, we identify the actions with high priority; which mainly relates to the flow of external transportation with \( \alpha_0 \) where the disease originates, and quarantine of \( \beta_1 \) which is directly linked to the disease origin; leading to (\( \text{EXT}(\beta_1, \alpha_0), 0.9 \)) and (\( \text{QUA}(\beta_1), 0.9 \)). The compartmental transitions of each vertex after the implementation of both actions are visualized in Fig. 3. Compared with Fig. 2, the exponential growth phase of cumulative cases in \( \beta_1 \) disappeared, as a consequence of intensive quarantine action. Meanwhile, the disease arrival time and peak period of infections of the other vertices in \( \beta \) and \( \gamma \) are all postponed for over 30 days, compared to Example 3.1.

3.3. Mitigation strategies

**Definition 3.2. (Mitigation strategy and cost)** The combination of action-rate tuples taken by the administrative unit \( u^* \) is labeled as a mitigation strategy \( M = \{(a_i, x_i)| i = 1, 2, ..., |A|\} \), where \(|A|\) denotes the number of actions taken by \( u^* \). The cost of a specific mitigation strategy \( M \) is the sum of each of its action-rate tuple’s cost; the cost of an individual action-rate tuple is positively correlated to the number of people affected. In the following, we define the cost for different types of actions:

\[
\Omega(a, x) = \begin{cases} 
F_{mn} \cdot x_i & \text{if } a_i = \text{EXT}(m, n) \\
\langle F_{mn} \cdot x_i \rangle \log(F_{mn} \cdot x_i) & \text{if } a_i = \text{INT}(m, n) \\
\langle N_{mn} \cdot x_i \rangle^2 & \text{if } a_i = \text{QUA}(m)
\end{cases}
\]  

\((m \in V_{u^*}, n \in V_{u^*}, u^* \neq u^*, F_{mn} \cdot x_i > 0)\)

The definition of cost comes with the following intuition: (a) restrictions on internal transportation have higher cost than restriction of external transportation for the same number of affected people; (b) quarantine of a vertex is the costliest action; (c) since the cost expression of quarantine action is quadratic to the number of affected people, quarantining a few people has far less cost than
Table 2
Three distinct mitigation strategies from the search space of Example 3.2. The cost of strategy \( M_1 \) is much higher than \( M_2 \) and \( M_3 \) since the cost expression of \( QUA(\beta_1, 0.9) \) is quadratic to the number of affected people. As shown in Example 3.2, \( QUA(\beta_1, 0.9) \) and \( EXT(\beta_1, \alpha_0, 0.9) \) are quite effective in preventing the spread of disease, thus leading to a relatively small \( ESI \) for strategy \( M_1 \).

| Strategy | Action-rate tuples | Cost | ESI |
|----------|--------------------|------|-----|
| \( M_1 \) | \{(EXT(\beta_1, \alpha_0), 0.9), (EXT(\beta_1, \alpha_1), 0.9), (QUA(\beta_1), 0.9)\} | 2.37 \times 10^2 | 1.49 \times 10^3 |
| \( M_2 \) | \{(EXT(\beta_1, \alpha_0), 0.6), (INT(\beta_1, \beta_2), 0.6), (INT(\beta_2, \beta_2), 0.9)\} | 1.03 \times 10^4 | 5.15 \times 10^5 |
| \( M_3 \) | \{(EXT(\beta_2, \alpha_0), 0.3), (EXT(\beta_1, \alpha_1), 0.6), (INT(\beta_2, \beta_2), 0.6)\} | 3.13 \times 10^3 | 6.94 \times 10^5 |

quarantining a large fraction of population. These intuitions are derived from the real world: Restricting domestic (internal) transportation, including commuting of people for purpose of business, leisure or migration (Chopra and Meindl, 2007) and medical quarantining a large fraction of population. These intuitions are derived from the real world: Restricting domestic (internal) transport, including commuting of people for purpose of business, leisure or migration (Chopra and Meindl, 2007) and medical emergency (Skinner, 1962), has a greater social-economic impact than restricting international (external) transportation which includes mostly international tourism and trade. The quarantine action, on the other hand, directly restricts the freedom and mobility of individuals. Thus, the quarantine action has a much larger cost. In addition, putting a few people into quarantine has a smaller social-economic impact than putting a large fraction of people into isolation. While the former poses threat on the freedom of individuals and has negative psychological consequences, the latter will lead to significant economic paralysis of manufacturing and service industries since large proportional of employees are absent from work (Hsieh et al., 2005). Based on this consideration, we model the cost for action \( QUA(v) \) to be quadratic of the number of affected people. The specific choice of a cost function could be adapted in future studies. Here, given the preliminary character of this study, we propose the above simplified cost function for developing a proof of concept.

In the following, we develop a general index to evaluate the performance of different mitigation strategies in controlling epidemic, i.e., Epidemic Spreading Index \( ESI \). The cumulative cases can reflect the past and current epidemic situation well. However, the prediction of future epidemic situation should also be considered in this general index. The same cumulative cases but at different stages (e.g., one at exponential growing stage while another at slow growing stage) may lead to different situations. Accordingly, we take into account the growth rate of cumulative cases, which determines the future epidemic trend to some extent. Besides, for a standard SEIR model that does not consider birth and death of individuals, the number of cumulative cases of a vertex is non-decreasing with respect to time, leading to a non-negative growth rate of cumulative cases.

**Definition 3.3. (Epidemic Spreading Index)** Given a mitigation strategy \( M \), a time horizon \( t_0 \), the Epidemic Spreading Index \( ESI \) is defined as

\[
ESI(M, t_0) = p \sum_{i \in V^e} C^M_i(t_0) + q \sum_{i \in V^e} \frac{dc^M_i(t_0)}{dt} \left( t_0 \in Z^+, p \geq 0, q \geq 0 \right)
\]  

(7)

where \( C^M_i(t_0) \) and \( \frac{dc^M_i(t_0)}{dt} \) denote the cumulative cases and the growth rate of cumulative cases in vertex \( i \) at time horizon \( t_0 \) under the influence of mitigation strategy \( M \), respectively. \( p \) and \( q \) are non-negative proportionality constants.

Here, we assume that the cumulative cases and the growth rate of cumulative cases are equally weighted, i.e. \( p = 0.5 \) and \( q = 0.5 \). The specific choice of these weighting factors could be determined by decision makers depending on their application.

**Example 3.3. (Mitigation strategy, cost and ESI)** Here, we select three distinct mitigation strategies from the search space of Example 3.2 and report their combinations of action-rate tuples, costs and \( ESI \) in Table 2. Since the cost expression of quarantine action is quadratic to the number of affected people, the cost of strategy \( M_1 \) is much higher than \( M_2 \) and \( M_3 \). However, as shown in Example 3.2, \( QUA(\beta_1, 0.9) \) and \( EXT(\beta_1, \alpha_0, 0.9) \) are quite effective in preventing the spread of disease, thus leading to a relatively small \( ESI \) for strategy \( M_1 \). In contrast, although strategy \( M_2 \) and \( M_3 \) have a relatively small cost, their action-rate tuples are not as effective as that of \( M_1 \), thus leading to a relatively large \( ESI \).

**Definition 3.4. (TLQP)** Given a collection of possible mitigation strategies \( \{M_k\} \) and a fixed time horizon \( t_0 \), the goal of TLQP is to identify the minimum-cost mitigation strategy \( M^* \) such that the Epidemic Spreading Index \( ESI \) remains below a fixed upper limit \( U_{ESI} \), i.e.,

\[
\min \sum_{(a, x) \in M} \Omega(a, x) \\
\text{s.t.} \quad ESI(M, t_0) \leq U_{ESI}
\]

(8)

4. Solution techniques

In this section, we introduce two techniques for solving TLQP. First, TLQP-E is an exact solution technique based on full enumeration (Section 4.1). The second technique, TLQP-H, is a heuristic which exploits fast simulation and highly-efficient searching
4.1. TLQP-E: Exact solution based on full enumeration

In the following, we present an exact solution algorithm for TLQP, labeled as TLQP-E, and also explore how its complexity increases with the size of the network. The idea behind TLQP-E is straightforward: We perform a full enumeration of the search space of mitigation strategies and select the strategy with minimum cost that satisfies the given upper limit of $\text{ESI}$. 

Example 4.1. (TLQP-E) Continuing with Example 3.2, we set the upper limit of $\text{ESI}$ as $2 \times 10^3$, and obtain the following optimal mitigation strategy $M^*$:

$$M^* = \{(\text{EXT} (\beta_1, \alpha_0), 0.9), (\text{EXT} (\beta_1, \alpha_1), 0.9), (\text{QUA} (\beta_1), 0.9)\}$$

In addition to the optimal mitigation strategy, we further explore the structure of the search space. The changes of feasible mitigation strategies with respect to the changes of $U_{\text{ESI}}$ are visualized in Fig. 4. Three distinct $U_{\text{ESI}}$ are considered to generate three clusters of mitigation strategies, i.e., $2 \times 10^5$ for Cluster 1 (A,B), $1 \times 10^4$ for Cluster 2 (C,D) and $2 \times 10^3$ for Cluster 3 (E,F). For the mitigation strategies in each cluster, the mean value of $\text{ESI}$ with respect to time series is visualized on the left. The categorical bubble plots on the right show the rate distribution of actions that constitute each mitigation strategy. For each action, the relative size of bubbles represents the probability of mitigation strategies that apply specific rate (i.e., $x \in \{0, 0.3, 0.6, 0.9\}$). In Cluster 2 (D) and Cluster 3 (F), $\text{EXT} (\beta_1, \alpha_0), \text{EXT} (\beta_1, \alpha_1)$ and $\text{QUA}(\beta_1)$ cause extreme distributions of rates, which indicates their dominant effects on the magnitude of $\text{ESI}$.

(Section 4.2).

4.1. TLQP-E: Exact solution based on full enumeration

In the following, we present an exact solution algorithm for TLQP, labeled as TLQP-E, and also explore how its complexity increases with the size of the network. The idea behind TLQP-E is straightforward: We perform a full enumeration of the search space of mitigation strategies and select the strategy with minimum cost that satisfies the given upper limit of $\text{ESI}$.

Example 4.1. (TLQP-E) Continuing with Example 3.2, we set the upper limit of $\text{ESI}$ as $2 \times 10^3$, and obtain the following optimal mitigation strategy $M^*$:

$$M^* = \{(\text{EXT} (\beta_1, \alpha_0), 0.9), (\text{EXT} (\beta_1, \alpha_1), 0.9), (\text{QUA} (\beta_1), 0.9)\}$$

In addition to the optimal mitigation strategy, we further explore the structure of the search space. The changes of feasible mitigation strategies with respect to the changes of $U_{\text{ESI}}$ are visualized in Fig. 4. Three distinct $U_{\text{ESI}}$ are considered to generate three clusters of mitigation strategies, i.e., $2 \times 10^5$ for Cluster 1 (A,B), $1 \times 10^4$ for Cluster 2 (C,D) and $2 \times 10^3$ for Cluster 3 (E,F). For the mitigation strategies in each cluster, the mean value of $\text{ESI}$ with respect to time series is visualized on the left. The categorical bubble plots on the right show the rate distribution of actions that constitute each mitigation strategy. For each action, the relative size of bubbles represents the probability of mitigation strategies that apply specific rate (i.e., $x \in \{0, 0.3, 0.6, 0.9\}$). In Cluster 2 (D) and Cluster 3 (F), $\text{EXT} (\beta_1, \alpha_0), \text{EXT} (\beta_1, \alpha_1)$ and $\text{QUA}(\beta_1)$ cause extreme distributions of rates, which indicates their dominant effects on the magnitude of $\text{ESI}$. Therefore, we may infer that $\text{QUA}(\beta_1)$ can effectively constrain the growth
rate of ESI, while $\text{EXT}(\beta_1, \alpha_0), \text{EXT}(\beta_2, \alpha_1)$ are able to postpone the disease arrival time to $\beta$.

Clearly, such an enumeration algorithm faces combinatorial challenges, with an increasing number of actions. In order to highlight the computation difficulties, we discuss an extended set of artificial examples next. To generate these artificial examples with increasing size, we randomly partition the vertex set $V$ with administrative units in $U$ by K-means algorithm (with x and y coordinates as the attributes for partition), where $|V| \in \{10, 15, 20, \ldots, 35\}, U = \{\alpha, \beta, \gamma, \delta, \epsilon\}$. Random internal or external transportation edges are generated among vertices, represented by solid and dashed lines, respectively. The width of internal and external transportation edges is proportionally correlated with passenger flow.

Fig. 5. The spatial structure of six artificial examples with increasing size and complexity. The vertex set $V$ is randomly partitioned with administrative units in $U$ by K-means algorithm (with x and y coordinates as the attributes for partition), where $|V| \in \{10, 15, 20, \ldots, 35\}, U = \{\alpha, \beta, \gamma, \delta, \epsilon\}$. Random internal or external transportation edges are generated among vertices, represented by solid and dashed lines, respectively. The width of internal and external transportation edges is proportionally correlated with passenger flow.

Table 3
The network parameters of eight artificial example datasets, including six binary rate examples and two non-binary (NB) examples. The disease origin vertex (denoted as $v_s$) is randomly selected, while the administrative unit of interest remains $\beta$ for all eight artificial examples. The time horizon $t_h$ is 60 days, and the initial rate of infection is 0.5%. The exponential growth of the number of mitigation strategies make the execution of TLQP-E intractable for large networks.

| Dataset | $u^*$ | $v_s$ | $|V|$ | $|E|$ | $t_h$ | Actions | Rates | Strategies | TLQP-E time (s) |
|---------|--------|--------|--------|--------|------|---------|-------|------------|----------------|
| AE I    | $\beta$ | $\delta_2$ | 10     | 9      | 60   | 5       | 2     | 32         | 9.8            |
| AE II   | $\beta$ | $\delta_0$ | 15     | 19     | 60   | 8       | 2     | 256        | 35.5           |
| AE III  | $\beta$ | $\delta_0$ | 20     | 33     | 60   | 13      | 2     | 8,192      | 1,129.9        |
| AE IV   | $\beta$ | $\delta_0$ | 25     | 45     | 60   | 15      | 2     | 32,768     | 5,403.3        |
| AE V    | $\beta$ | $\delta_0$ | 30     | 56     | 60   | 18      | 2     | 262,144    | 49,766.5       |
| AE VI   | $\beta$ | $\delta_0$ | 35     | 69     | 60   | 20      | 2     | 1,048,576  | 228,714.3      |
| AE I (NB)| $\beta$ | $\delta_0$ | 10     | 9      | 60   | 5       | 4     | 1,024      | 96.1           |
| AE II (NB)| $\beta$ | $\delta_0$ | 15     | 19     | 60   | 8       | 4     | 65,536     | 7,554.9        |
as the attributes for partition), where \(|V| \in \{10, 15, 20, \ldots, 35\} \), \(U = \{a, \beta, \gamma, \delta, \varepsilon\} \). Random internal or external transportation edges are generated among vertices. The spatial structures of these six artificial examples are visualized in Fig. 5. The width of internal and external transportation edges is proportionally correlated with passenger flow. The network parameters of growing artificial examples are reported in Table 3. While the disease origin vertex is randomly selected, we always focus on the decision making of administrative unit \(\beta\) for all artificial examples. The time horizon \(t_h\) is 60 days and the initial rate of infection remains the same with Example 3.1.

The mechanism of TLQP-E model in solving this problem indicates two features: (a) it simulates every transportation edge (both internal and external), as well as the compartmental transitions of every vertex in \(V\) throughout the time horizon \(t_h\). (b) the optimal mitigation strategy \(M^*\) is identified through full enumeration over the search space, whose size is exponential in the number of actions of the administrative unit. Both features lead to a relative high runtime complexity. If we assume the runtime of compartmental simulator and a search method based on the Cost-Effective Lazy Forward (CELF) algorithm. The ED-SEIR simulator first identifies the mitigation strategy can be obtained efficiently. Furthermore, to avoid the full enumeration of the whole search space of mitigation strategies, we use CELF to reduce the number of ED-SEIR simulation by exploiting the submodularity of ESI function. The details of TLQP-H are described below.

### 4.2. TLQP-H: Heuristic solution based on fast simulation and highly-efficient searching

Given that the full enumeration of all action combinations does not scale up well with larger datasets, we develop an efficient solution heuristic, labeled as TLQP-H. There are two core ingredients of our heuristic: the network-based ED-SEIR model as the simulator and a search method based on the Cost-Effective Lazy Forward (CELF) algorithm. The ED-SEIR simulator first identifies the dominant epidemic spreading paths to the administrative unit \(u^*\), predicts the disease arrival time of each vertex along these paths, and executes a fast SEIR simulation for the vertices along these paths. With the help of the ED-SEIR simulator, the ESI under the influence of a mitigation strategy can be obtained efficiently. Furthermore, to avoid the full enumeration of the whole search space of mitigation strategies, we use CELF to reduce the number of ED-SEIR simulation by exploiting the submodularity of ESI function. The details of TLQP-H are described below.

#### 4.2.1. ED-SEIR simulator

The concept of effective distance (ED) of an edge, proposed by Brockmann and Helbing (2013), reflects the phenomenon that large traffic flow is equivalent to a small distance, and vice versa. The key idea behind the effective distance is that, despite the structural internal and external), as well as the compartmental transitions of every vertex in \(V\) throughout the time horizon \(t_h\), the effective distance \(D_{mn}\) is given by

\[
d_{mn} = (1 - \log P_{mn}) \geq 1
\]

The directed length \(d(\Gamma)\) of an ordered path \(\Gamma = \{v_1, v_2, \ldots, v_7\}\) is the sum of effective lengths along the legs of the path. Thus, the effective distance \(D_{mn}\) from an arbitrary reference vertex \(m\) to another vertex \(n\) is defined by the length of the shortest path from \(m\) to \(n\), i.e.,

\[
D_{mn} = \min d(\Gamma)
\]

From the perspective of the disease origin \(v_0\), the set of shortest paths to all the vertices in the epidemic network constitutes an effective distance tree, illustrating the most probable path from the disease origin to the other vertices. Based on the effective distance tree rooted at the disease origin, we can obtain the dominant epidemic spreading paths to the vertices in administrative unit \(u^*\). The vertices that are passed along these dominant paths form a subset of \(V\), denoted as \(V'_{u^*}\), and the edges included in these dominant paths form a subset of \(E\), denoted as \(E'_{u^*}\). Note that \(V'_{u^*} \subseteq V\) and \(E'_{u^*} \subseteq E\). In general, this subnetwork constituted by vertex subset \(V'_{u^*}\) and edge subset \(E'_{u^*}\) is a good candidate for SEIR simulation, based on the following two reasons:

1. **Accuracy.** Since the cumulative number of cases in a vertex tends to grow exponentially once the disease arrives if no effective mitigation strategy is involved (we have discussed this fact in Section 3.1), the determining factor for the present/future epidemic situation of a vertex is the disease arrival time. The disease arrival time for each vertex in \(u^*\) can be estimated from the shortest path of ED, which are all included in edge subset \(E'_{u^*}\). We have found that the disease arrival time of each vertex in the administrative unit obtained by ED-SEIR model has 0.20% mean deviation \((s = 0.0098)\) in terms of that obtained by the standard SEIR model, based on the experiments of artificial examples, thus verifying that ED-SEIR can estimate disease arrival time accurately. After the disease arrival time, there may still exist some newly imported infections by other parallel spreading paths, but since the vertex has already entered a fast-growing stage, these newly imported infections will not make a significant difference on the present or future epidemic situation of the vertex, according to our experiments. We conduct experiments and multiple sensitivity tests on ED-SEIR...
The administrative unit of interest is \( \beta \). The annotations over the white arrows are the disease arrival times obtained by ED-SEIR simulator, while the annotations below the black arrows are the disease arrival times obtained by standard SEIR simulator. These disease arrival times obtained by different approaches are highly coincided.

The spatial structures of ED tree for Artificial Example II and III. The width of the edges is proportionally correlated to the magnitude of flow, while the length of the edges is the effective distance obtained by Eqs. (9) and (10). The unit length of ED is shown in the scale bar. In both examples, the administrative unit of interest is \( \beta \). The annotations over the white arrows are the disease arrival times obtained by ED-SEIR simulator, while the annotations below the black arrows are the disease arrival times obtained by standard SEIR simulator. These disease arrival times obtained by different approaches are highly coincided.

The spatial structures of ED tree for Artificial Example II and III. The width of the edges is proportionally correlated to the magnitude of flow, while the length of the edges is the effective distance obtained by Eqs. (9) and (10). The unit length of ED is shown in the scale bar. In both examples, the administrative unit of interest is \( \beta \). The annotations over the white arrows are the disease arrival times obtained by ED-SEIR simulator, while the annotations below the black arrows are the disease arrival times obtained by standard SEIR simulator. These disease arrival times obtained by different approaches are highly coincided.

2. Efficiency. The exclusion of the other vertices and edges that lay outside the dominant epidemic spreading paths reduces the runtime complexity. The number of simulated vertices and edges (i.e., vertices that perform compartmental transitions and edges that transport passenger flow) of ED-SEIR simulator are largely reduced comparing with those of standard SEIR simulator, i.e., \(|V_\text{u}^*| < |V|\) and \(|E_\text{u}^*| < |E|\).

Example 4.2. (ED-SEIR simulator) Specifically, we illustrate the ED-SEIR simulator with the Artificial Example II and Artificial Example III. The spatial structures of ED tree for both artificial examples are shown in Fig. 6. The width of the edges is proportionally correlated to the magnitude of flow, while the length of the edges is the effective distance obtained by Eqs. (9) and (10). The unit length of ED is shown in the scale bar. In both examples, the administrative unit of interest is \( \beta \). For instance, in Artificial Example II, the shortest path to \( \beta_0 \) is \( \gamma_0 = \{ \beta_1, \beta_2, \beta_3, \beta_0 \} \), then the effective distance from disease origin \( \delta_0 \) to \( \beta_0 \) is the sum of effective length along the legs of \( \gamma \). For Artificial Example II, the number of simulated vertices and edges are largely reduced, i.e., \(|V_\text{u}^*|/|V| = 0.27\) and \(|E_\text{u}^*|/|E| = 0.16\). Note that the vertices in \( u^* \) are not necessarily clustered together in a branch of the tree, for instance in Artificial Example III, \( \beta_0 \) is not connected to the other vertices in \( \beta \), thus we need to take into account the spreading path \( \Gamma = \{ v_0, \gamma_3, \gamma_2, \beta_0 \} \) when executing SEIR simulation. The annotations over the white arrows are the disease arrival times obtained by ED-SEIR simulator, while the annotations below the black filled arrows are the disease arrival times obtained by standard SEIR simulator. These disease arrival times obtained by different approaches are highly coincided.

4.2.2. Searching method based on CELF

In order to avoid full enumeration over the search space, we develop a highly-efficient searching method based on the Cost-Effective Lazy Forward (CELF) algorithm. In the following, we screen out influential actions from action space based on ED tree, establish a greedy framework which iteratively selects the cost-effective action-rate tuple, and finally apply CELF to reduce the number of ED-SEIR simulation by exploiting the submodularity of ESI function.

Definition 4.1. (Action classes) We reinforce the idea of ED tree to classify actions into the following three action classes:

- Action class \( A_0 \): actions of quarantining vertices in administrative unit \( u^* \), i.e., \( \text{QUA}(v) (v \in V_\text{u}^*)\)
- Action class \( A_1 \): actions of reducing flow of important transportation edges that lies within the dominant epidemic spreading paths, i.e., \( \text{EXT}(e) \) and \( \text{INT}(e) \), with \( e \in E_\text{u}^* \).
- Action class \( A_2 \): actions of reducing flow of unimportant transportation edges that lies beyond the dominant epidemic spreading paths, i.e., \( \text{EXT}(e) \) and \( \text{INT}(e) \), with \( e \not\in E_\text{u}^* \).

Example 4.3. (Action classes) Here, we illustrate the action classification of Artificial Example II as follows:

- \( A_0 = \{ \text{QUA}(\beta_2), \text{QUA}(\beta_1), \text{QUA}(\beta_0) \} \)
- \( A_1 = \{ \text{EXT}(\beta_3, \delta_0), \text{INT}(\beta_1, \beta_2), \text{INT}(\beta_0, \beta_1) \} \)
- \( A_2 = \{ \text{EXT}(\beta_1, \epsilon_0), \text{EXT}(\beta_1, \alpha_1) \} \)
Since the dominant spreading paths largely determine the epidemic situation of vertices and parallel spreading paths have relatively small influence on vertices, actions of $A_Q$ have negligible effect on reducing $ESI$, thus we only consider the actions of $A_Q$ and $A_{IT}$ in the following. To find the general properties of $ESI$ function, we conduct two experiments on actions in $A_Q$ and $A_{IT}$: The first experiment tests the effect of individual action-rate tuple on $ESI$ qualitatively. Let $\mathcal{N}$ denotes the event that the application of an action-rate tuple decreases or maintains $ESI$ given an empty mitigation strategy set, i.e., $ESI(\{(a, x)\}) - ESI(\emptyset) \leq 0 \forall a \in A_Q \cup A_{IT}, x \in \{0.3, 0.6, 0.9\}$. Experimental results on six artificial examples show that $P(\mathcal{N}) = 88.11\%(x = 0.18)$, i.e., the majority of action-rate tuples will lower $ESI$. The second experiment tests how the effect of individual action-rate tuple on $ESI$ changes with respect to the growth of mitigation strategy. Let $(a_0, x_0)$ denotes the action-rate tuple that is to be tested, $\mathcal{N}_h = \{(a_j, x_j) | 1 \leq j \leq h, j \in N\}$ denotes the mitigation strategy set of size $h \geq 0$ on each binary-rate artificial example. Fig. 7, we show the negative relationship between marginal loss of $ESI$ and $\mathcal{N}_h$.

**Definition 4.2. (ESI properties)** Based on the results of $P(\mathcal{N})$ and $P(\exists)$, we formally make two assumptions on $ESI(\cdot)$ function as follows:

- Monotonicity: $ESI(\cdot)$ is a monotonically decreasing function with respect to mitigation strategy set $M$, i.e., $ESI(M') - ESI(M) \leq 0, \forall M \subseteq M' \subseteq Y$.
- Submodularity: $ESI(\cdot)$ satisfies diminishing return property, i.e., $ESI(M) - ESI(M \cup (a,x)) \geq ESI(M') - ESI(M' \cup (a,x)), \forall M \subseteq M' \subseteq Y$ and $(a,x) \notin M$.

where $Y$ denotes the universe set of action-rate tuples, i.e., $Y = \{(a,x) | a \in A_Q \cup A_{IT}, x \in X \setminus \{0\}\}$

**Algorithm 1. CELF-based Algorithm**

**Input**: set of action-rate tuples $Y$, upper limit $U_{ESI}$, function $ESI(\cdot)$ and $\Omega(\cdot)$

**Output**: best mitigation strategy $M$

1: initialize $M = \emptyset$, max-heap $\mathcal{H} = \emptyset$
2: for each $(a, x) \in Y$ do
3: \hspace{1cm} $Q \leftarrow \Delta((a,x)|\emptyset)/\Omega(a,x)$
4: \hspace{1cm} add node tuple $((a,x), Q, 1)$ to $\mathcal{H}$ in decreasing order of $Q$
5: end for
6: while $ESI(M) > U_{ESI}$ do
7: \hspace{1cm} extract node tuple $((a,x), Q, r)$ from the heap root
8: \hspace{1cm} if $r = |M| + 1$ then
9: \hspace{1cm} $M \leftarrow M \cup \{(a,x)\}$
10: \hspace{1cm} delete node tuple $((a,x), Q, r)$ from the heap $\mathcal{H}$
11: \hspace{1cm} else
12: \hspace{1cm} delete node tuple $((a,x), Q, r)$ from the heap $\mathcal{H}$
13: \hspace{1cm} if $\Delta((a,x)|M) > ESI(M) - U_{ESI}$ then
14: \hspace{1cm} $Q \leftarrow (ESI(M) - U_{ESI})/\Omega(a,x)$
15: \hspace{1cm} end if
16: \hspace{1cm} if $\Delta((a,x)|M) \leq ESI(M) - U_{ESI}$ then
17: \hspace{1cm} $r \leftarrow |M| + 1$
18: \hspace{1cm} insert node tuple $((a,x), Q, r)$ to the heap $\mathcal{H}$
19: end if
20: end if
21: end while
22: Return $M$

The reason why we apply CELF to the search space of mitigation strategy is that despite the fact that many action alternatives may be applied, the control and prevention of epidemic given an upper limit of $ESI$ is dominated by a set of cost-effective actions. We first consider the greedy framework that CELF is based on here. Let $M'$ denote the set of selected action-rate tuples after the $i$-th iteration, and $\Delta((a,x)|M') = ESI(M') - ESI(M' \cup \{(a,x)\}) \geq 0$ denote the marginal loss of $ESI$ of $(a,x)$ w.r.t $M'$. The cost-effectiveness (or quality) of $(a,x)$ in the $i$-th iteration is measured by the ratio of marginal loss of $ESI$ and the corresponding cost, i.e.,

\[
Q(a,x) = \frac{\Delta((a,x)|M')}{\Omega(a,x)} \quad \left((a,x) \in Y, M' \subseteq Y\right) \quad (11)
\]

\[
if \Delta((a,x)|M') \leq ESI(M') - U_{ESI} \quad (12)
\]

The $(a,x)$ that can induce more marginal loss of $ESI$ or possess less cost will be considered as priority. In the $i$-th iteration, the $(a,x)$ with
the highest \( Q' \) will be selected and added to the mitigation strategy set \( M' \). Note that each selected action \( a \) can only take on one rate \( x (x \in X \setminus \{0\}) \) in \( M' \). The algorithm terminates when current \( ESI \) is lower than the upper limit \( U_{ESI} \), and returns the mitigation strategy set as the best solution of TLQP-H, denoted with \( \hat{M} \). Besides, we weaken the role of \( \Delta((a,x)\mid M') \) by constraining it with an upper bound; if \( \Delta((a,x)\mid M') \) is larger than the expected marginal loss of \( ESI \), then replace \( \Delta((a,x)\mid M') \) with the expected marginal loss of \( ESI \), i.e.,

\[
Q'(a,x) = \frac{ESI(M') - U_{ESI}}{\Omega(a,x)} \bigg( (a,x) \in Y, M' \subseteq Y \bigg)
\]

if \( \Delta((a,x)\mid M') > ESI(M') - U_{ESI} \)

(13)

(14)

This modification is especially effective when \( ESI(M') \) draws near to \( U_{ESI} \), and at this time the \( (a,x) \) that induces relatively small marginal loss but just passes the \( U_{ESI} \) may leads to a relatively high \( Q'(a,x) \) compared to another one that induces larger marginal loss but with higher cost, since they share the same numerator but different denominators.

Based on the greedy framework, CELF can effectively reduce the number of ED-SEIR simulations by exploiting the submodularity of \( ESI \) function. The intuition behind CELF is that most actions in the search space have negligible influence on reducing \( ESI \) and thus can be easily pruned at subsequent iterations (Leskovec et al., 2007; Li et al., 2018). According to the submodularity of \( ESI \) function, \( \Delta((a,x)\mid M') \) is an upper bound for any \( \Delta((a,x)\mid M') \) s.t. \( M' \subseteq M \subseteq Y \). Based on this property, CELF first computes \( \Delta((a,x)\mid \emptyset) \) for each \( (a,x) \in Y \) and selects the most cost-effective one to form \( M' \). Then, \( \Delta((a,x)\mid \emptyset) \) can be utilized as an upper bound as follows. At each iteration \( j = 2, \ldots, k \) until the terminate condition \( ESI(M^k) \leq U_{ESI} \) is satisfied, CELF visits \( (a,x) \in Y \setminus M^{j-1} \) in a descending order of their upper bounds of \( \Delta((a,x)\mid M^{j-1}) \), and computes \( \Delta((a,x)\mid M^{j-1}) \) using ED-SEIR simulation. Instead of visiting all \( (a,x) \in Y \setminus M^{j-1} \), CELF implements an early termination whenever the maximum upper bound of unvisited action-rate tuple is already smaller than the maximum \( \Delta((a,x)\mid M^{j-1}) \) of visited action-rate tuples. Then, CELF updates the upper bound of each visited action-rate tuple as \( \Delta((a,x)\mid M^{j-1}) \) and proceeds to the next iteration \( j + 1 \). The pseudocode of CELF based on a max-heap data structure is shown in Algorithm 1.

5. Experimental results

In this section, we verify the efficiency and efficacy of TLQP-H. In Section 5.1, we report the experimental results of TLQP-H for solving artificial example problems compared to TLQP-E. In Section 5.2, we extend the definition of TLQP to the real-world scenario by referring to GLEaM (Balcan et al., 2010), and show the experimental results of TLQP-H in solving Planet Top-k problems compared to TLQP-E. In Section 5.3 and Section 5.4, we conduct two scenario analysis which focus on the decision making of a specific country, Germany, in face of the first wave and second wave of global pandemic.
5.1. Experimental results on artificial example datasets

We apply both the TLQP-H and TLQP-E to the artificial example problems presented in Section 4. To measure the accuracy of simulation results of TLQP-H with respect to TLQP-E, the gap of TLQP-H in terms of TLQP-E is computed as:

\[ \text{Gap} = \frac{|\text{Simulation result of TLQP-H} - \text{Simulation result of TLQP-E}|}{\text{Simulation result of TLQP-E}} \times 100\% \]

Fig. 8. The gap of TLQP-H in terms of TLQP-E with respect to the full range of \( U_{ISI} \) on artificial examples. The experiments on six binary rate examples \( x \in \{0, 0.9\} \) and two non-binary (NB) rate examples \( x \in \{0, 0.3, 0.6, 0.9\} \) are considered. The lower/upper bound of \( U_{ISI} \) are set to be the minimum/maximum value of the \( ISI \) obtained by TLQP-E. Gaps at relatively high upper limits are larger than those at relatively low upper limits. However, all the gaps remain below 0.2%.

Fig. 9. The comparison of efficiency of TLQP-H and TLQP-E in solving artificial example problems. (A) The number of simulated vertices and edges (i.e. vertices that perform compartmental transitions and edges that transport passenger flow) of ED-SEIR simulator and standard SEIR simulator on Artificial Example I to VI. ED-SEIR simulator largely reduces the number of simulated vertices and edges. (B) The runtime of TLQP-E and TLQP-H in solving six binary rate examples. The presented runtime for TLQP-H is the mean runtime of simulations over the full range of \( U_{ISI} \). Note that the y-axis is in log scale. The combinatorial challenge of TLQP-E clearly leads to large magnitude of runtime with growing sizes of datasets. TLQP-H remains low runtime (within 1s).

5.1. Experimental results on artificial example datasets

We apply both the TLQP-H and TLQP-E to the artificial example problems presented in Section 4. To measure the accuracy of simulation results of TLQP-H with respect to TLQP-E, the gap of TLQP-H in terms of TLQP-E is computed as:
The disease origin

Table 4
infected (World Health Organization, 2020a). The country of interest

where TLQP: experiments, using planet-scale data to solve the TLQP, we refer to GLEaM (Balcan et al., 2010) and adapt the following definitions of

5.2. Experimental results on planet dataset
increase much with growing sizes of datasets.
magnitude of runtime with growing sizes of datasets. On the other hand, TLQP-H remains low runtime (within 1

Y. Ding et al.
5.2. Experimental results on planet dataset

ED-SEIR simulator has largely reduced the number
vertices and edges (i.e. vertices that perform compartmental transitions and edges that transport passenger flow) of ED-SEIR simulator
gaps remain below 0
Experimental results show that gaps at relatively high upper limits are larger than those at relatively low upper limits. However, all the
gaps remain below 0.2%, thus indicating the efficacy of TLQP-H in solving artificial example problems.

We further compare the efficiency of TLQP-H and TLQP-E, visualized in Fig. 9. In Fig. 9(A), we compare the number of simulated
vertices and edges, and both numbers do not increase explicitly with growing sizes of datasets. In Fig. 9(B), we compare
the runtime of TLQP-E and TLQP-H in solving six binary rate examples. The presented runtime for TLQP-H is the mean runtime of
simulations over the full range of ESIs, which does not
per day is given by

Gap = \frac{\left| \Omega(\hat{\mathcal{M}}) - \Omega(\mathcal{M}^*) \right|}{\Omega(\mathcal{M}^*)} \times 100\% \tag{15}

where \(\Omega(\mathcal{M}^*)\) represents the cost of optimal mitigation strategy obtained by TLQP-E, and \(\Omega(\hat{\mathcal{M}})\) represents the cost of best mitigation strategy obtained by TLQP-H, i.e.,

Since TLQP-E will clearly lead to combinatorial challenges with growing sizes of datasets, we consider the experiments on six
binary rate examples (\(x \in \{0.0.9\}\)) and two non-binary (NB) rate examples (\(x \in \{0.0.3.0.6.0.9\}\)). In order to verify that our heuristic
performs well under any specific upper limit of ESIs, we set the lower/upper bound of \(U_{\text{ESI}}\) to be the minimum/maximum value of the
ESI obtained by TLQP-E, respectively. The gap of TLQP-H in terms of TLQP-E with respect to the full range of \(U_{\text{ESI}}\) is visualized in Fig. 8.
Experimental results show that gaps at relatively high upper limits are larger than those at relatively low upper limits. However, all the
gaps remain below 0.2%, thus indicating the efficacy of TLQP-H in solving artificial example problems.

We further compare the efficiency of TLQP-H and TLQP-E, visualized in Fig. 9. In Fig. 9(A), we compare the number of simulated
vertices and edges (i.e. vertices that perform compartmental transitions and edges that transport passenger flow) of ED-SEIR simulator
and standard SEIR simulator on Artificial Example I to VI. The bar chart shows that ED-SEIR simulator has largely reduced the number of
simulated vertices and edges, and both numbers do not increase explicitly with growing sizes of datasets. In Fig. 9(B), we compare
the runtime of TLQP-E and TLQP-H in solving six binary rate examples. The presented runtime for TLQP-H is the mean runtime of
simulations over the full range of \(U_{\text{ESI}}\). Note that the y-axis is in log scale. The combinatorial challenge of TLQP-E clearly leads to large
magnitude of runtime with growing sizes of datasets. On the other hand, TLQP-H remains low runtime (within 1s), which does not
increase much with growing sizes of datasets.

5.2. Experimental results on planet dataset
To further verify the efficacy of TLQP-H, we perform experiments on real-world scenarios. In preparation for the real-world ex-
periments, using planet-scale data to solve the TLQP, we refer to GLEaM (Balcan et al., 2010) and adapt the following definitions of
TLQP:

- The vertices (\(V\)) are extended to the subpopulations that are centered at major airports or transportation hubs with aggregated
population.
- The administrative units (\(U\)) are countries in the real world.
- The external transportation edges are international flights with other countries. The action \(\text{EXT}(v_1, v_2)\) (\(v_1 \in U^*, v_2 \notin U^*\)) denotes
reducing the flow of international flights of \(U^*\).
- The internal transportation edges cover both domestic flights and commuting links within a country. The action \(\text{INT}(v_1, v_2)\) (\(v_1 \in U^*, v_2 \in U^*\)) denotes reducing the flow of domestic flights or commuting links within \(U^*\).

The flow of international flights and domestic flights per day is obtained from OpenFlights database. The commuting flow data is
calculated through gravity law, calibrated by the real-world data of 29 countries in five continents (Balcan et al., 2010). The
commuting flow between subpopulation \(i\) and \(j\) per day is given by

\[ F_{ij} = \frac{C_{\alpha \beta} N_i^\alpha N_j^\beta}{d_{ij}^\beta} \]

where \(d_{ij}\) is the distance (in km) between the two major airports of subpopulation \(i\) and \(j\). The values of exponents \(\alpha\) and \(\beta\), the inverse
characteristic distance \(\beta\) and the proportionality constant \(C\) are reported in Appendix.

To evaluate TLQP-H in the real-world cases, we generate a series of Planet Top-k datasets, which consists of \(k\) subpopulations with
the largest number of inhabitants, and the transportation edges connecting them, here \(k \in \{50, 60, 70, 80, 90, 100\}\). The network

| Dataset | \(U\) | \(V\) | \(|V|\) | \(|E|\) | \(t_0\) | --Actions-- | --Rates-- | --Strategies-- | TLQP-E time (s) |
|---------|------|------|------|------|------|------------|-----------|--------------|----------------|
| PT 50   | EGY  | WUH  | 50   | 167  | 60   | 6          | 2         | 64           | 34.5           |
| PT 60   | PHL  | WUH  | 60   | 217  | 60   | 7          | 2         | 128          | 67.4           |
| PT 70   | KOR  | WUH  | 70   | 265  | 60   | 12         | 2         | 4,096        | 2,077.5        |
| PT 80   | JPN  | WUH  | 80   | 290  | 60   | 12         | 2         | 4,096        | 2,392.1        |
| PT 90   | ETH  | WUH  | 90   | 357  | 60   | 17         | 2         | 131,072      | 84,645.3       |
| PT 100  | IDN  | WUH  | 100  | 448  | 60   | 19         | 2         | 524,288      | 377,355.1      |
| PT 50 (NB) | EGY  | WUH  | 50   | 167  | 60   | 6          | 4         | 4,096        | 1,545.7        |
| PT 60 (NB) | PHL  | WUH  | 60   | 217  | 60   | 7          | 4         | 16,384       | 7,276.3        |
parameters of eight Planet Top-k datasets are reported in Table 4. The disease origin is assumed to be in WUH (the subpopulation centered at Wuhan Tianhe International Airport) with 44 individuals being initially infected (World Health Organization, 2020a). The country of interest is selected from the Asia continent and its neighborhood.

The gap of TLQP-H with respect to TLQP-E in solving eight Planet Top-k problems, under the full range of upper limits of \( U_{\text{ESI}} \) are shown in Fig. 10. Comparable to the artificial example problems, the majority of gaps remain trivial. For experiments on non-binary datasets (PT50 (NB) and PT60 (NB)), all gaps remain below 0.5%. Meanwhile, in Fig. 11, ED-SEIR simulator significantly reduces the number of simulated vertices and edges, and the runtime of TLQP-H is four orders of magnitude shorter than TLQP-E. Accordingly, the combinatorial challenge of TLQP-E clearly leads to large magnitude of runtime with growing sizes of datasets, but TLQP-H remains low runtime (within 10s).

Fig. 10. The gap of TLQP-H in terms of TLQP-E with respect to a wide range of \( U_{\text{ESI}} \) on Planet Top-k datasets. The experiments on six binary rate datasets \((x \in \{0.0, 0.9\})\) and two non-binary (NB) rate datasets \((x \in \{0.3, 0.6, 0.9\})\) are considered. The lower/upper bound of \( U_{\text{ESI}} \) are set to be the minimum/maximum value of the ESI obtained by TLQP-E. Compared to the gaps in solving artificial example problems, the gaps in solving Planet Top-k problems remain below 0.5%.

Fig. 11. The comparison of efficiency of TLQP-H and TLQP-E in solving Planet Top-k problems. (A) The number of simulated vertices and edges (i.e. vertices that perform compartmental transitions and edges that transport passenger flow) of ED-SEIR simulator and standard SEIR simulator on six Planet Top-k datasets. ED-SEIR simulator largely reduces the number of simulated vertices and edges. (B) The runtime of TLQP-E and TLQP-H on six binary rate datasets. The presented runtime for TLQP-H is the mean runtime of simulations over the full range of \( U_{\text{ESI}} \). Note that the y-axis is in log scale. The combinatorial challenge of TLQP-E clearly leads to large magnitude of runtime with growing sizes of datasets, but TLQP-H remains low runtime (within 10s).
TLQP-H is both efficient and accurate with respect to TLQP-E in solving Planet Top-k problems.

5.3. Scenario analysis I (avoiding the first wave)

By using TLQP-H, we can solve TLQP on much larger instances of the planet dataset, compared to TLQP-E. In the following, we consider a real-world scenario which focuses on the decision making of a specific country, Germany, in face of the first wave of COVID-19, with the origin of the disease being in Wuhan City, Hubei Province of China. In order to compare our simulation results with the real-world epidemic data, we modify TLQP model by adjusting the proportionality constants of $ESI$ function, i.e., $p = 1$ and $q = 0$; indicating that $ESI$ is equivalent to cumulative cases in the country. We concentrate on a realistic decision problem for the administration: What should Germany do in the first three months to avoid the epidemic spreading inside the country? In other words: How could Germany avoid the first wave of global pandemic?

First, we report the relevant subpopulations centered around Germany airports, external transportation edges (international flights) and internal transportation edges (domestic flights and commuting links) of Germany are reported. Binary rates are considered in this scenario analysis, i.e., $x \in (0, 0.9)$. The total number of possible actions is huge, which leads to an intractable number of mitigation strategies.

![Fig. 12.](image-url) The situation of unconstrained subpopulations, internal transportation edges and external transportation edges before the initial outbreak of COVID-19 (January 3rd, 2020) in the real world. In (A), the relative sizes of the subpopulations represent their population; In (B) and (C), the relative sizes of the subpopulations represent node degree. The top ten largest subpopulations are annotated by the IATA codes of the major airports. These subpopulations also feature high node degrees.

### Table 5
The network parameters of Germany based on TLQP-H model. The number of relevant subpopulations centered around Germany airports, external transportation edges (international flights) and internal transportation edges (domestic flights and commuting links) of Germany are reported. Binary rates are considered in this scenario analysis, i.e., $x \in (0, 0.9)$. The total number of possible actions is huge, which leads to an intractable number of mitigation strategies.

| Parameter | Description | Value |
|-----------|-------------|-------|
| $|V|$ | Number of subpopulations centered around major airports | 31 |
| $|EXT|$ | Number of external edges (international flights) | 3141 |
| $|INT|$ | Number of internal edges (domestic flights and commuting links) | 288 |
| $|A|$ | Number of possible actions | 1734 |
| $|M|$ | Number of possible mitigation strategies | $2^{1734}$ |

### Table 6
The real-world COVID-19 statistics of Germany and the model settings for TLQP-H in face of the first-wave and second wave global pandemic. In first wave scenario, the cumulative cases in Germany has reached 120,479 on April 12th, 2020 ([World Health Organization, 2020b](#)). We set the upper limit of cumulative cases for Germany to be 10,000, which is approximately ten times smaller than the number of (confirmed) real-world cases. In second wave scenario, the global situation of COVID-19 on November 5th, 2020 is selected as the initial condition, and the time horizon is set to be 14 days. We set the upper limit of cumulative cases for Germany to be 700,000, accordingly.

| Scenario | Initial date | Initial cases | Final date | Final cases | Duration (days) | Upper limit for TLQP-H |
|----------|--------------|---------------|------------|-------------|----------------|-----------------------|
| First wave | 2020/01/03 | 0 | 2020/04/12 | 120,479 | 100 | 10,000 |
| Second wave | 2020/11/05 | 597,583 | 2020/11/19 | 855,916 | 14 | 700,000 |

TLQP-H is both efficient and accurate with respect to TLQP-E in solving Planet Top-k problems.

5.3. Scenario analysis I (avoiding the first wave)

By using TLQP-H, we can solve TLQP on much larger instances of the planet dataset, compared to TLQP-E. In the following, we consider a real-world scenario which focuses on the decision making of a specific country, Germany, in face of the first wave of COVID-19, with the origin of the disease being in Wuhan City, Hubei Province of China. In order to compare our simulation results with the real-world epidemic data, we modify TLQP model by adjusting the proportionality constants of $ESI$ function, i.e., $p = 1$ and $q = 0$; indicating that $ESI$ is equivalent to cumulative cases in the country. We concentrate on a realistic decision problem for the administration: What should Germany do in the first three months to avoid the epidemic spreading inside the country? In other words: How could Germany avoid the first wave of global pandemic?

First, we report the relevant subpopulations centered around Germany airports, external transportation edges (international flights) and internal transportation edges (domestic flights and commuting links) of Germany in Table 5. Since Germany has hundreds of internal edges and external edges, the total number of possible actions is huge, which leads to an intractable number of mitigation strategies for TLQP-E. We consider binary rates in this scenario analysis, i.e., $x \in (0, 0.9)$, so the application of an action indicates a rate
generally regarded as the first wave of global pandemic of COVID-19 (World Health Organization, 2020b). The time horizon is set from January 3rd, 2020 to April 12th, 2020, with a duration of 100 days, which is the upper limit of cumulative cases for Germany to be 10,000, which is approximately ten times smaller than the number of confirmed real-world cases.

![Fig. 13. The situation of constrained/unconstrained subpopulations, unconstrained internal transportation edges and external transportation edges during the first wave of global pandemic, generated by TLQP-H. In (A), the relative sizes of the subpopulations represent their population; In (B) and (C), the relative sizes of the subpopulations represent node degree. The cross marker of subpopulations indicates the application of QUA($v$|$v \in DE$) action, and the removal of external or internal edges indicates the application of EXT($v_1$, $v_2$)($v_1 \in DE$, $v_2 \notin DE$) or INT($v_1$, $v_2$)($v_1 \in DE$, $v_2 \in DE$) action. The QUA action is applied to FRA and MUC, and both of their degree in external/internal transportation network decrease to a large extent.]

The information of best mitigation strategy ($\hat{M}$) under the first wave scenario and second wave scenario for Germany, generated by TLQP-H. In the first wave scenario, the probability of applying action EXT($v_1$, $v_2$)($v_1 \in DE$, $v_2 \in CN$) is up to 90%, but the probability of locking down external edges which are not from China, internal edges and subpopulations of quarantine are generally low. In the second wave scenario, nearly 80% of the external edges, internal edges and subpopulations are restricted or quarantined.

| Action type | First wave | Second wave | Action example |
|-------------|------------|-------------|----------------|
| EXT($v_1$, $v_2$)($v_1 \in DE$, $v_2 \in CN$) | 90.91% | 100.00% | EXT(FRA, PVG) |
| EXT($v_1$, $v_2$)($v_1 \in DE$, $v_2 \notin CN$) | 29.40% | 80.33% | EXT(FRA, LHR) |
| QUA($v$|$v \in DE$) | 6.45% | 80.65% | QUA(FRA) |
| INT($v_1$, $v_2$)($v_1 \in DE$, $v_2 \in DE$) | 28.17% | 78.87% | INT(MUC, TXL) |

of 0.9, and vice versa. Thus, an efficient heuristic such as TLQP-H is required for solving this optimization problem. In our scenario analysis, an initial infection of 44 cases is set to the region WUH (subpopulation centered at Wuhan Tianhe International Airport); according to the WHO case report that a total of 44 case-patients with SARS-CoV-2 were detected on January 3rd, 2020 (World Health Organization, 2020a). The time horizon is set from January 3rd, 2020 to April 12th, 2020, with a duration of 100 days, which is generally regarded as the first wave of global pandemic of COVID-19 (World Health Organization, 2020b).

As shown in Table 6, the cumulative number of cases in Germany has reached 120,479 on April 12th, 2020 (World Health Organization, 2020b). We are thus interested in whether a more reasonable decision making (compared to the real world) could have effectively controlled the epidemic spreading, while maintaining the social and economic cost not reaching too far. Accordingly, we set the upper limit of cumulative cases for Germany to be 10,000, which is approximately ten times smaller than the number of confirmed real-world cases.

The restriction of international flight edges and quarantine of subpopulations during the first wave are shown in Fig. 13. In subfigure A, the relative sizes of the subpopulations represent their population; in subfigure B and subfigure C, the relative sizes of the subpopulations represent node degree. The cross marker of subpopulations indicates the application of QUA($v$|$v \in DE$) action, and the removal of external or internal edges indicates the application of EXT($v_1$, $v_2$)($v_1 \in DE$, $v_2 \notin DE$) or INT($v_1$, $v_2$)($v_1 \in DE$, $v_2 \in DE$) action. The QUA action is applied to FRA and MUC, and both of their degree in external/internal transportation network decrease to a large extent.

Compared to the situation before initial outbreak (visualized in Fig. 12), the QUA action is applied to FRA (the subpopulation centered at Frankfurt Airport) and MUC (the subpopulation centered at Munich Airport), and both of their degree in external transportation network and internal transportation network decrease significantly, which indicates the EXT and INT actions have been applied to the majority of external or internal transportation of FRA and MUC. This implies that those subpopulations that frequently receive imported cases should be treated with special care, because they tend to further spread out the disease than the other regions.

The information of the best mitigation strategy $\hat{M}$ under the first-wave scenario is reported in Table 7. The probability of applying action EXT($v_1$, $v_2$)($v_1 \in DE$, $v_2 \in CN$) is 90.91%, which indicates that at an early response stage, it is critical to cut down the international flight edges with the country where disease originates from. However, we can observe that the probability of locking down external edges which are not from China, internal edges and subpopulations of quarantine are generally low, which indicates that at an early response stage, Germany may not pose too much restriction on other means of transportation or activity since they would lead to severe social-economic consequences.
5.4. Scenario analysis II (avoiding the second wave)

In the next scenario analysis, we focus on the decision making of Germany in face of the second wave of global pandemic. In reality, a relaxation of lock-downs and the public’s loosening of precautionary behaviours has seen recorded cases and risen fatalities across Europe (Looi, 2020). Contrary to the first wave, there are multiple high-risk countries and subpopulations with large number of infections. As reported in Table 6, the global situation of COVID-19 on November 5th, 2020 is selected as the initial condition of this second-wave simulation, when the cumulative number of cases reached 47,959,190 in the real world (World Health Organization, 2020b). For the sake of simplicity, we assume the current infection of a country is evenly distributed to each of its subpopulation. The time horizon is set to be 14 days, from 5th November 2020 to 19th November 2020. As of 5th November 2020, the number of cumulative cases in Germany has reached 597,583 (World Health Organization, 2020b). Under this circumstance, what is the best possible mitigation strategy that Germany can take so as to control the cumulative cases within 700,000 after two weeks?

In Fig. 14, the best mitigation strategy of the second wave scenario indicates that the majority of external/internal edges and subpopulations are restricted or quarantined, which forms a sharp contrast with the situation before the initial outbreak (shown in Fig. 12) and the situation during the first wave (shown in Fig. 13).

6. Conclusions

In this study, we proposed the Early-stage Transportation Lock-down and Quarantine Problem (TLQP), which addresses the problem of deciding how an administrative unit (e.g., country) should face a pandemic threat by locking down its transportation infrastructure and possibly implementing quarantine. Given the intrinsic complexity of the model, we designed a solution heuristic which is based on Effective Distance paths and the Cost-Effective Lazy Forward algorithm. We solved several artificial examples to show the efficiency and efficacy of our heuristics compared to full enumeration. In addition, we solved several real-world use cases to highlight the practicality of model and solution heuristics in the presence of COVID-19. Future work could further aim at improving the heuristic in terms of runtime and solution quality, e.g., by using Deep Learning (Wang et al., 2019). Besides, future studies can extend the epidemic spreading network to a directed one, since restricting the in-flow tends to be more severe than restricting the out-flow, from the perspective of an administrative unit. Moreover, future studies could use our model to perform scenario analysis on specific regions of interest and under specific (disease/cost) parameters, including action costs and action rates. The ongoing research on vaccines has the potential to gradually shift the need for optimizing for quarantine/lock-downs towards medical distribution planning (Escribano Macias et al., 2020); however, it has to be shown how effective these vaccines are and whether they can prevent reinfections in the long run. There is also a need for studies which could aim at better understanding how unmanned mobility could facilitate the life in face of pandemics (Gonzalez-R et al., 2020). Finally, it would be very interesting, yet challenging to treat the action rates as a decision variable, which could possibly be achieved by using Bayesian optimization.
CRediT authorship contribution statement

Yida Ding: Conceptualization, Methodology, Software, Validation, Writing - original draft. Sebastian Wandelt: Conceptualization, Formal analysis, Software, Resources, Writing - original draft, Supervision. Xiaoqian Sun: Conceptualization, Writing - original draft, Supervision, Funding acquisition.

Acknowledgements

This study is supported by the Research Fund from National Natural Science Foundation of China (Grants No. 61861136005, No. 61851110763, No. 71731001).

Appendix A

A.1. Epidemiological parameters

The epidemiological parameters of TLQP-E and TLQP-H in the context of COVID-19 are reported in Table 8. As is explained in Section 3.1, the epidemiological status of all individuals is initially set to susceptible, where it can contract the virus through contacts with individuals in the infectious compartment and proceed to the exposed compartment at the transmission rate $\tau = 0.5$ (Zou et al., 2020). Exposed individuals proceed to the infectious stage with a rate inversely proportional to the mean latency period $\theta = 0.2$ (Chinazzi et al., 2020), and the infectious individuals proceed to the recovered/removed compartment with a rate inversely proportional to the mean infectious period $\mu = 0.2$ (Chinazzi et al., 2020). The adaptation of SEIR model into GLEaM model requires two more parameters, $P_a$ and $P_t$. The probability of an infected individual to become asymptomatic is $P_a = 0.4$ (Zou et al., 2020), and its probability to travel and further spread the disease is $P_t = 0.2$ (Balcan et al., 2009).

A.2. Sensitivity analysis on ED-SEIR model

To verify that a parallel path of similar ED as the shortest path from the disease origin to a specific vertex $v$ has negligible influence on the ESI of this vertex (denoted as $ESI_v$), we conduct experiments and sensitivity analysis on Artificial Example V as follows. We consider the fully unconstrained case in these experiments, i.e. actions and mitigation strategies are not involved here. The length of the shortest path from disease origin to vertex $v$ (denoted as $D(\Gamma_0)$) is obtained by Dijkstra’s Algorithm, while the length of second shortest path from disease origin to vertex $v$ (denoted as $D(\Gamma_1)$) is obtained by Yen’s Algorithm. The deviation of ED is calculated by

$$\text{Dev of } ED = \frac{|D(\Gamma_1) - D(\Gamma_0)|}{D(\Gamma_0)} \times 100\% \left( D(\Gamma_0) > 0 \right)$$ (16)

Table 9 shows the deviation of $ESI_v (v \in V_\delta)$ with respect to the deviation of ED. Smaller deviation of ED indicates higher similarity of two parallel paths. $ESI_v$ is computed by both ED-SEIR simulator and standard SEIR simulator. The latter is utilized as the baseline regarding the deviation of $ESI_v$. The results show that there is no direct relationship between deviation of ED and deviation of $ESI_v$, and even when deviation of ED is sufficiently small (e.g.g, for $\beta_0$ is 4.44%, indicating that two parallel paths to $\beta_0$ are very similar), the deviation of $ESI_{\beta_0}$ remains trivial *** (see Table 10).

We further conduct a sensitivity analysis on the relationship between deviation of ED and deviation of $ESI_v (v \in V_\delta)$. Based on the values of default parameters, we consider multiple variation of parameters one at a time, including transmission rate $\tau$, mean latency period $\mu^{-1}$, mean infectious period $\theta^{-1}$, time horizon $\delta$, and two $ESI$ proportionality constants $p, q$. The network’s configuration of Artificial Example V remains unchanged, in order to fix the deviation of ED for vertices in administrative unit $\delta$. The scatter plot of deviation of $ESI_v (v \in V_\delta)$ with respect to deviation of ED under the variation of different parameters is visualized in Fig. 15. The results show that the variation of some parameters (e.g., $\tau, \mu, \theta$ and $\delta$) influence the deviation of $ESI_v$, but to a small extent (always less than 5%). Meanwhile, there remains no negative correlation between deviation of ED and deviation of $ESI_v$, in other words, similar parallel paths have negligible influence on $ESI_v$.

| Symbol | Description | Value |
|--------|-------------|-------|
| $\theta$ | Inverse of mean latency period | 0.2 |
| $\mu$ | Inverse of mean infectious period | 0.2 |
| $\tau$ | Transmission rate | 0.5 |
| $P_a$ | Probability of asymptomatic infections | 0.4 |
| $P_t$ | Probability of travelling infections | 0.2 |
In addition, we conduct one-at-a-time parameter sensitivity analysis on the performance of ED-SEIR with each artificial example, visualized in Fig. 16. Based on the values of default parameters, we consider multiple variation of parameters, including transmission rate $\tau$, mean latency period $\mu$, mean infectious period $\theta$, time horizon $t_h$, two ESI proportionality constants $p$ and $q$, the proportion of initial infection in disease origin vertex $P_I$, average rate constant $x$ and average flow rate of vertices $P_F$ (i.e. $P_F = \frac{1}{|V|} \sum_{i \in V} F_i$). The deviation of ESI is measured by the ESI obtained by ED-SEIR in terms of that obtained by standard SEIR simulator. The results show that variation of some disease parameters (e.g., $\tau$, $\mu$, $\theta$ and $t_h$) will increase deviation of ESI, for instance $\tau$ and $t_h$ at 175% of default values lead to 10% to 20% deviation of ESI in Artificial Example II and III.

### Table 9
The relationship between the deviation of ED and deviation of $ESI_v$ ($v \in V_\beta$) on Artificial Example V. $D(\Gamma_0)$ represents the length of the shortest path from disease origin to vertex $v$, while $D(\Gamma_1)$ represents the length of second shortest path from disease origin to vertex $v$. Deviation of ED is derived by Eq. 16. $ESI_v$ is computed by both ED-SEIR simulator and standard SEIR simulator. The latter is utilized as the baseline regarding the deviation of $ESI_v$.

| $v$  | $D(\Gamma_0)$ | $D(\Gamma_1)$ | Dev of ED | $ESI_v$ by SEIR | $ESI_v$ by ED-SEIR | Dev of ESI |
|------|---------------|---------------|-----------|----------------|------------------|------------|
| $v_0$ | 7.11          | 7.43          | 4.44%     | 10,141.50      | 10,121.50        | 0.20%     |
| $v_3$ | 7.24          | 9.39          | 29.65%    | 3,705.00       | 3,668.50         | 0.99%     |
| $v_1$ | 4.94          | 6.84          | 38.41%    | 337,692.50     | 336,594.50       | 0.33%     |
| $v_4$ | 4.90          | 7.22          | 47.43%    | 337,426.00     | 336,775.00       | 0.19%     |

### Table 10
Default Parameters of the sensitivity analysis in Section A.2.

| Symbol | Description                                      | Value       |
|--------|--------------------------------------------------|-------------|
| $\theta$ | Inverse of mean latency period                   | 0.2         |
| $\mu$  | Inverse of mean infectious period                | 0.2         |
| $\tau$ | Transmission rate                                | 1.0         |
| $t_h$  | Time horizon                                     | 60          |
| $p, q$ | ESI proportionality constants                    | 0.5, 0.5    |
| $P_I$  | Proportion of initial infection in disease origin vertex | 0.005     |
| $x$    | Average rate constant                            | 0.5         |
| $P_F$  | Average flow rate of vertices                    | 0.002       |

![Fig. 15](image-url)  
Fig. 15. One-at-a-time sensitivity analysis on the relationship between deviation of ED and deviation of $ESI_v$ ($v \in V_\beta$) regarding Artificial Example V. The network’s configuration of Artificial Example V remains unchanged, in order to fix the deviation of ED for $v \in V_\beta$. Based on the values of default parameters, multiple variation of parameters are considered, including transmission rate $\tau$, mean latency period $\mu$, mean infectious period $\theta$, time horizon $t_h$ and two ESI proportionality constants $p$ and $q$. The variation of some parameters (e.g., $\tau, \mu, \theta$ and $t_h$) influence the deviation of $ESI_v$, but to a small extent (always less than 5%). No clear negative correlation between deviation of ED and deviation of $ESI_v$ is observed.

In addition, we conduct one-at-a-time parameter sensitivity analysis on the performance of ED-SEIR with each artificial example, visualized in Fig. 16. Based on the values of default parameters, we consider multiple variation of parameters, including transmission rate $\tau$, mean latency period $\mu$, mean infectious period $\theta$, time horizon $t_h$, two ESI proportionality constants $p$ and $q$, the proportion of initial infection in disease origin vertex $P_I$, average rate constant $x$ and average flow rate of vertices $P_F$ (i.e. $P_F = \frac{1}{|V|} \sum_{i \in V} F_i$). The deviation of ESI is measured by the ESI obtained by ED-SEIR in terms of that obtained by standard SEIR simulator. The results show that variation of some disease parameters (e.g., $\tau, \mu, \theta$ and $t_h$) will increase deviation of ESI, for instance $\tau$ and $t_h$ at 175% of default values lead to 10% to 20% deviation of ESI in Artificial Example II and III.
The gravity law of Eq. 5.2 has 4 free parameters: the exponents $\alpha$ and $\gamma$, the inverse characteristic distance $\beta$ and the proportionality constant $C$. A multivariate regression analysis is applied to obtain the values of the parameters that better fit the data as well as an estimation of their statistical significance. The values estimated for these parameters are reported in Table 11 along with their p-values and the regression coefficients (Balcan et al., 2010).

Figs. 17 and 18.

### A.3. Gravity law and commuting data statistical analysis

The gravity law of Eq. 5.2 has 4 free parameters: the exponents $\alpha$ and $\gamma$, the inverse characteristic distance $\beta$ and the proportionality constant $C$. A multivariate regression analysis is applied to obtain the values of the parameters that better fit the data as well as an estimation of their statistical significance. The values estimated for these parameters are reported in Table 11 along with their p-values and the regression coefficients (Balcan et al., 2010).

Figs. 17 and 18.

### Table 11

Exponents of gravity law as obtained by applying a multivariate analysis to global commuting data.

| $d$ (km) | Parameter | Estimate | Standard Error | $p$ – value | $R^2$ |
|----------|-----------|----------|----------------|-------------|-------|
| $\leq 300$ | $\alpha$ | 0.46 | 0.01 | $< 2E-16$ | 0.7972 |
| $\gamma$ | 0.64 | 0.01 | $< 2E-16$ | 0.5369 |
| $\beta$ | 0.0122 | 0.0002 | $< 2E-16$ | 0.5369 |
| $> 300$ | $\alpha$ | 0.35 | 0.06 | 6.91E-09 | 0.5369 |
| $\gamma$ | 0.37 | 0.06 | 2.12E-09 | 0.5369 |
Fig. 17. The runtime of TLQP-H in solving increasing sizes of Planet Top-k problems given five upper limits of ESI. The Planet Top-k datasets are subsets of planet dataset consisting of k subpopulations with the largest number of inhabitants and the transportation edges connecting them, here k ranges from 500 to 3267. In each dataset, five upper limits of ESI are imposed, which are 10%, 30%, 50%, 70% and 90% of the baseline ESI (i.e. the maximum ESI when no action or mitigation strategy is involved). The runtime remains below 500 s.

Fig. 18. Sensitivity analysis on the relationship between the running time of TLQP-H and $U_{ESI}$. The first six artificial examples consider binary rates, i.e. $x \in \{0, 0.9\}$ and the last two artificial examples consider non-binary (denoted as NB) rates, i.e. $x \in \{0, 0.3, 0.6, 0.9\}$. The set of upper limits range from the minimum ESI to the maximum ESI obtained by the full enumeration of TLQP-E. The running time of TLQP-H in solving all eight examples remain below 5 s. As the upper limit reduces, the running time of TLQP-H increases at some specific points and then remain steady, thus it does not highly depend on upper limits.

References
Andersen, K.G., Rambaut, A., Lipkin, W.I., Holmes, E.C., Garry, R.F., 2020. The proximal origin of SARS-CoV-2. Nat. Med. 26, 450–452.
Badr, H.S., Du, H., Marshall, M., Dong, E., Squire, M.M., Gardner, L.M., 2020. Association between mobility patterns and covid-19 transmission in the usa: a mathematical modelling study. Lancet Infect. Dis. 20, 1247–1254.
Balcan, D., Colizza, V., Gonçalves, B., Hu, H., Ramasco, J.J., Vespignani, A., 2009. Multiscale mobility networks and the spatial spreading of infectious diseases. Proc. Nat. Acad. Sci. 106, 21484–21489.
Saad-Roy, C.M., Wagner, C.E., Baker, R.E., Morris, S.E., Farrar, J., Graham, A.L., Levin, S.A., Mina, M.J., Metcalf, C.J.E., Grenfell, B.T., 2020. Immune life history, vaccination, and the dynamics of sars-cov-2 over the next 5 years. Science 370, 811–818.

Sipetas, C., Kehlilkoglou, A., Gonzales, E.J., 2020. Estimation of left behind subway passengers through archived data and video image processing. Transp. Res. Part C Emerg. Technol. 118, 102727. https://doi.org/10.1016/j.trc.2020.102727 http://www.sciencedirect.com/science/article/pii/S0968090X20306422.

Skinner, H.A., 1962. The origin of medical terms. Acad. Med. 37, 160.

Skog, L., 2014. Spatial Analysis and Modeling for Health Applications. Ph.D. thesis. KTH Royal Institute of Technology.

Skog, L., Linde, A., Palmgren, H., Hauska, H., Elgh, F., 2014. Spatiotemporal characteristics of pandemic influenza. BMC Infectious Dis. 14, 1–14.

Stefanos, P., Mamoulid, S.E., Robinson, M., Nettlerd, E., Tuells, J., Bergsaker, M.A.R., Heijbel, H., Varwood, J., on Standardization of Attitudinal Studies in Europe, V.W.G., et al., 2010. Tracking parental attitudes on vaccination across european countries: the vaccine safety, attitudes, training and communication project (vacsat). Vaccine 28, 5731–5737.

Sumner, A., Hoy, C., Ortiz-Juarez, E., et al., 2020. Estimates of the Impact of COVID-19 on Global Poverty. United Nations University World Institute for Development Economics Research.

Sun, X., Wandelt, S., Zhang, A., 2020. How did COVID-19 impact air transportation? A first peek through the lens of complex networks. J. Air Transp. Manage. 89, 101928. https://doi.org/10.1016/j.jairtraman.2020.101928 http://www.sciencedirect.com/science/article/pii/S0969699720305111.

Sun, X., Wandelt, S., Zhang, A., 2021. On the degree of synchronization between air transport connectivity and covid-19 cases at worldwide level. Transp. Policy 105. https://doi.org/10.1016/j.tranpol.2021.03.005.

Takahashi, T., Ellington, M.K., Wong, P., Israelow, B., Lucas, C., Klein, J., Silva, J., Mao, T., Oh, J.E., Tokuyama, M., et al., 2020. Sex differences in immune responses that underlie covid-19 disease outcomes. Nature 588, 315–320.

Tang, X., Wu, C., Li, X., Song, Y., Yao, X., Wu, X., Duan, Y., Zhang, H., Wang, Y., Qian, Z., et al., 2020. On the origin and continuing evolution of sars-cov-2. Natl. Sci. Rev. 7, 1012–1023.

Tian, H., Liu, Y., Li, Y., Wu, C.H., Chen, B., Kraemer, M.U., Li, B., Cai, J., Xu, B., Yang, Q., et al., 2020. An investigation of transmission control measures during the first 50 days of the covid-19 epidemic in china. Science 368, 638–642.

Vásquez, A., 2004. Causal tree of disease transmission and the spreading of infectious diseases. In: Discrete Methods in Epidemiology, pp. 163–179.

Vespignani, A., Tian, H., Dye, C., Lloyd-Smith, J.O., Eggo, R.M., Shrestha, M., Scarpino, S.V., Gutierrez, B., Kraemer, M.U., Wu, J., et al., 2020. Modelling covid-19. Nature Reviews Physics 2, 279–281.

Wang, Y., Zhang, D., Liu, Y., Dai, B., Lee, L.H., 2019. Enhancing transportation systems via deep learning: A survey. Transp. Res. Part C Emerg. Technol. 99, 144–163. https://doi.org/10.1016/j.trc.2018.12.004 http://www.sciencedirect.com/science/article/pii/S0968090X18304108.

World Health Organization, 2020a. Novel Coronavirus (2019-nCoV) Situation Report-1. URL https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200121-sitrep-1-2019-ncov.pdf.

World Health Organization, 2020b. WHO Coronavirus Disease (COVID-19) Dashboard. URL https://covid19.who.int/table.

World Health Organization, 2020c. WHO Director-General’s opening remarks at the media briefing on COVID-19. URL https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19-11-march-2020.

Zaki, A.M., van Boheemen, S., Bestebroer, T.M., Osterhaus, A.D., Fouchier, R.A., 2012. Isolation of a Novel Coronavirus from a Man with Pneumonia in Saudi Arabia. New Engl. J. Med. 367, 1814–1820. doi: 10.1056/NEJMoa1211721, arXiv:https://doi.org/10.1056/NEJMoa1211721. pMID: 23075143.

Zlojutro, A., Rey, D., Gardner, L., 2019. A decision-support framework to optimize border control for global outbreak mitigation. Sci. Rep. 9, 1–14.

Zou, L., Ruan, F., Huang, M., Liang, L., Huang, H., Hong, Z., Yu, J., Kang, M., Song, Y., Xia, J., et al., 2020. SARS-CoV-2 viral load in upper respiratory specimens of infected patients. N. Engl. J. Med. 382, 1177–1179.