A Network-Based Stochastic Epidemic Simulator: Controlling COVID-19 With Region-Specific Policies

Askat Kuzdeuov, Daulet Baimukashev, Aknur Karabay, Bauyrzhan Ibragimov, Almas Mirzakhmetov, Mukhamet Nurpeiissov, Michael Lewis, and Huseyin Atakan Varol, *Senior Member, IEEE*

***Abstract*—In this work, we present an open-source stochastic epidemic simulator calibrated with extant epidemic experience of COVID-19. The simulator models a country as a network representing each node as an administrative region. The transportation connections between the nodes are modeled as the edges of this network. Each node runs a Susceptible-Exposed-Infected-Recovered (SEIR) model and population transfer between the nodes is considered using the transportation networks which allows modeling of the geographic spread of the disease. The simulator incorporates information ranging from population demographics and mobility data to health care resource capacity, by region, with interactive controls of system variables to allow dynamic and interactive modeling of events. The single-node simulator was validated using the thoroughly reported data from Lombardy, Italy. Then, the epidemic situation in Kazakhstan as of 31 May 2020 was accurately recreated. Afterward, we simulated a number of scenarios for Kazakhstan with different sets of policies. We also demonstrate the effects of region-based policies such as transportation limitations between administrative units and the application of different policies for different regions based on the epidemic intensity and geographic location. The results show that the simulator can be used to estimate outcomes of policy options to inform deliberations on governmental interdiction policies.**

***Index Terms*—Stochastic epidemic simulator, epidemiology, compartmental models, epidemic control, network-based simulation, SEIR model, COVID-19.**

I. Introduction

CORONAVIRUS disease 2019 (COVID-19) has emerged as a global crisis that threatens to overwhelm public healthcare systems and disrupt the social and economic welfare of every country. The daily lives of billions have been impacted by the unprecedented social controls imposed by governments to inhibit the spread of the virus [1]–[3].

Manuscript received May 2, 2020; revised June 10, 2020; accepted June 23, 2020. Date of publication June 26, 2020; date of current version October 5, 2020. *(Corresponding author: Huseyin Atakan Varol.)*

The authors are with the Institute of Smart Systems and Artificial Intelligence (ISSAI), Nazarbayev University, Nur-Sultan 010000, Republic of Kazakhstan (e-mail: askat.kuzdeuov@nu.edu.kz; daulet.baimukashev@nu.edu.kz; aknur.karabay@nu.edu.kz; bauyrzhan.ibragimov@nu.edu.kz; almas.mirzakhmetov@nu.edu.kz; mukhamet.nurpeiissov@nu.edu.kz; mlewis@nu.edu.kz; ahvarol@nu.edu.kz).

Digital Object Identifier 10.1109/JBHI.2020.3005160

COVID-19 has spread rapidly amongst a globally susceptible population, with rates of propagation and mortality greater than the averages typically associated with influenza [4]. Initial projections of the global impact [5] indicate that it will likely become the most severe pandemic in over a century, dating to the influenza epidemic of 1918.

The situation is exacerbated by several unfortunate observations: the population has no prior exposure, carriers are contagious in pre-symptomatic states [6], there is no vaccine yet available, widespread testing for the virus began quite late due to lack of testing kits, reagents, and facilities, and there appears to be a broad variation of risk profiles based on population demographics and prior health history [7].

In this scenario, epidemiological models can be used to project the future course of the disease, and to estimate the impact of non-pharmaceutical interventions (NPIs) and related control measures that might be used to slow the contagion, and thereby provide time to enhance health care resources and develop effective immunological defenses such as vaccines.

One of the first epidemiological models, Susceptible-Infected-Recovered (SIR), was proposed by Kermack and McKendrick in 1927 [8]. In the SIR model, society is divided into three compartments. The first compartment, susceptible (S), contains individuals who are vulnerable and not yet infected. The second compartment, infected (I), is formed from susceptible individuals who become infected; in this state, they are capable of shedding the virus and spreading the disease. The last compartment, recovered (R) consists of previously infected individuals who have overcome the disease. The recovered individuals are presumed to have acquired some level of immunity to the disease, such that they have a lower probability of reinfection compared to susceptible individuals.

The experience with the SIR model has motivated many variations, with new compartmental states, such as SEIR, SEIRS, SIRS, SEI, SEIS, SI, and SIS [9]. Many of these models incorporate an additional compartment - exposed (E). In this state, individuals are infected but not yet infectious during a latent period. In general, the compartments and related state transitions are selected based on the characteristics of a specific disease and the purpose of the model.

The propagation of infectious diseases can be simulated using both deterministic and stochastic models [10]. A deterministic

epidemic model is formulated as a system of differential equations, while a stochastic epidemic model can be implemented using stochastic differential equations or discrete-time Markov processes [11]. Both methods have their advantages and limitations [12]. For instance, a deterministic model yields the same solution each time if the initial conditions and parameter values are not changed, which can be considered as a limitation due to the fact that epidemic propagation is inherently a stochastic phenomenon. The distinction of a stochastic model is that it contains at least one probabilistic element, i.e. the output of the model varies at some range even if the initial conditions and model parameters remain constant, which makes stochastic models more challenging to manage and demonstrate causal links.

The compartmental models, however, have several limitations based on simplifying assumptions that do not correlate well to actual viral propagation. For instance, models assume that recovery and infection rates are the same for all individuals in the population. Also, the models do not take population mobility into account. However, epidemic diseases now spread further and faster than before due to modern mass transportation systems by air, ground, and sea [13], [14]. Thus, it is important to consider population mobility in models to estimate the impacts of NPIs on interconnected regions.

In this work, we have developed and implemented a network-based stochastic epidemic simulator (leveraging our prior work [15]) which models cities and regions as nodes in a graph, and the edges between nodes representing transit links of roads, railways, and air travel routes to model the mobility of inhabitants amongst cities. In each node, the simulator runs a compartmental Susceptible-Exposed-Infectious-Recovered (SEIR) model, such that individuals can cycle through the four stages based on state transition probabilities. These probabilities are based on parameters such as the susceptible-to-exposed transition constant and the mortality rate, which can be influenced by age, gender, genetic profile, and health status. Also, each node includes demographics of population along with health care system capacity, in particular, the intensive care unit (ICU) availability, which serves as a negative impact multiplier when ICU capacity is exceeded. The SEIR model was calibrated using COVID-19 data as reported for the Lombardy region, Italy. Then, the calibrated model was used to simulate the spread of the COVID-19 in each region of Kazakhstan as a part of the network model.

The simulator can be used to estimate the extent and duration of an epidemic over time, and model the potential impact of NPI measures deployed to suppress or mitigate the spread of the virus. To the best of our knowledge, it is the first model, calibrated with actual COVID-19 data, that illustrates the effect of the transportation systems on the spread of the disease. A summary of our contributions are:

- We present a novel network-based stochastic epidemic simulator, calibrated with real COVID-19 data, for modeling dynamics of the spread of the COVID-19 among regions connected with transportation links. The simulator can switch from the network model to the compartmental SEIR model by simply disconnecting a target region from the network.
- We show that the model can simulate the spread of the epidemic between regions by applying region-specific policies such as quarantine measures over discrete time intervals. This feature allows us to more realistically replicate the spread of the epidemic, taking into account actual and projected control strategies.
- We included an additional Vaccination state (disabled by default) which can be enabled if and when an effective vaccine emerges.
- We made the source code of the simulator publicly available, and provide a user-friendly graphical user interface (GUI) so that non-programmers can easily run the simulator and comprehend the results.

The rest of the paper is organized as follows: In Section II, we present related works and the novelties of our work compared to those. Section III introduces the methodologies for the single-node epidemic simulator, the network model, and software development. In Section IV, we simulate a single node epidemic model using real COVID-19 data to fine-tune parameters of the model, then using the obtained parameters we simulate multiple scenarios for the network-based model of the COVID-19 epidemic in Kazakhstan. In Section V, we discuss the simulation results, and Section VI concludes the work.

II. Related Work

Many epidemic models for the COVID-19 have appeared in the literature recently [16]–[18]. For instance, an age-structured SEIR model, with a focus on maintenance of ICU resources, was developed to describe COVID-19 transmission in Ontario, Canada [16]. A deterministic stage-structured SEIR model was used to simulate the outbreak in Wuhan, China [17]. An extended SIR (eSIR) model with time-varying transmission rates was developed to evaluate the impact of intervention measures in Italy [18]. However, in these works, the role of the transportation networks was not addressed. We found no work in the literature on modeling the influence of the transportation systems, considering region-specific policies, on the propagation of the COVID-19 between regions. Therefore, we will compare our work with similar works developed for other epidemics.

For instance, the role of the airline transportation network in the prediction and predictability of global epidemics was investigated in [19]. The authors conclude that seasonal effects and geographical heterogeneity should be taken into account to make realistic predictions. Also, they assume that including other transportation systems such as railways and highways could be useful to predict longer periods. In our case, we ignore the effect of the weather on the spread of COVID-19 because it is not yet thoroughly investigated. However, we consider the geographical heterogeneity of the regions such as demography and ICU capacity. Also, we take into account the population mobility via railways and highways.

A SEIR model was implemented using Anylogic simulation software to examine the correlation between transportation and

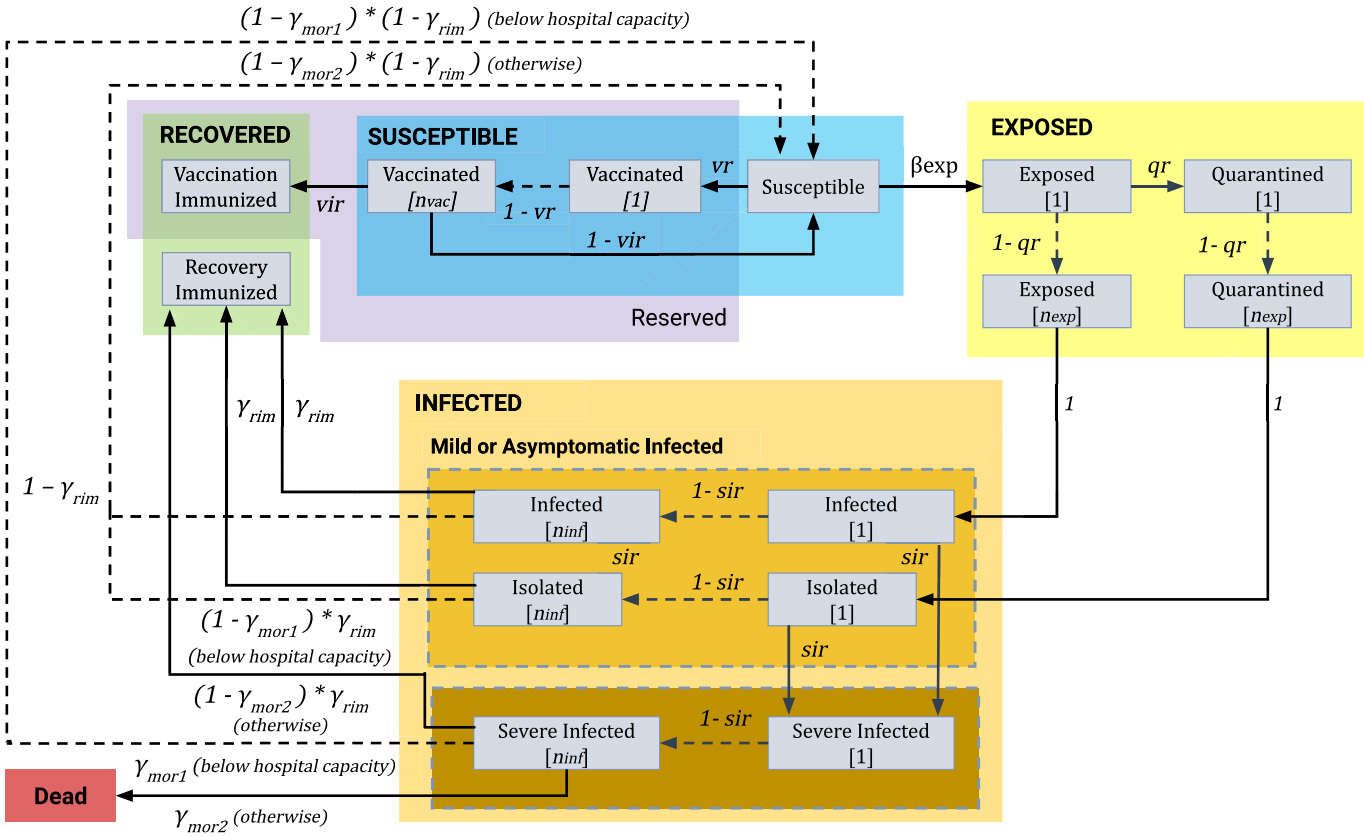

Fig. 1. The statechart of the SEIR epidemic node simulator.

infection of the H1N1 influenza virus [20]. The authors used the number of passenger inflows and outflows for each region. However, the arrival and departure points of passengers were not considered. Thus, the effect of traveling between specific regions on the spread of the disease was not investigated. In our case, a transition matrix contains information about the number of individuals and their departure points. This is necessary for simulating the effects of region-specific policies on the propagation of the disease in others.

The role of transportation networks in the spread of H1N1 influenza was investigated for mainland China [21]. The SEIR model was used to simulate the propagation of the epidemic in each province, and individuals moved between provinces with a fixed ratio. Authors conclude that they might have a bias in the results since the simulations did not consider the intervention strategies during travel and geographic factors. In our case, the model takes into account the intervention strategies over the corresponding time intervals (see Section IV).

TABLE I
LIST OF SIMULATION PARAMETERS AND THEIR DESCRIPTIONS

Parameter	Description
vr	Daily vaccination rate
vir	Vaccination immunization rate
β_{exp}	Susceptible to Exposed transition constant
qr	Daily quarantine rate (Ratio of Exposed individuals getting Quarantined)
sir	Daily severe infection rate (Ratio of the Infected and Isolated individuals getting Severe Infected)
hc	Hospital capacity
γ_{rim}	Infected to Recovery Immunized transition probability
γ_{mor1}	Severe Infected to Dead transition probability (if the hospital capacity is not exceeded)
γ_{mor2}	Severe Infected to Dead transition probability (if the hospital capacity is exceeded)
ε_{exp}	Transmission rate of Exposed compared to the Infected
ε_{qua}	Transmission rate of Quarantined compared to the Infected
ε_{sev}	Transmission rate of Severe Infected compared to the Infected
t_{exp}	Exposure period in days
t_{inf}	Infection period in days
t_{vac}	Vaccination immunization period in days
Δt	Sampling time, in days, for updating states

III. METHODOLOGY

A. Stochastic Epidemic Simulator for Single Nodes

In our simulator, the SEIR model consists of four superstates (Susceptible (S^S), Exposed (E^S), Infected (I^S) and Recovered (R^S)). Transitions amongst states occur in accordance with the statechart shown in Fig. 1. A description of parameters is given in the Table I. The Susceptible superstate (S^S) consists of two states: Susceptible (S) and Vaccinated (V). The daily vaccination rate vr dictates the transition between these states. The ratio of immunized people after vaccination vir is used to describe the portion of individuals going from the Vaccinated state (V) to the Vaccination Immunized (VI) or back to the Susceptible (S); however, it is not yet activated since there is no known vaccine for

COVID-19. In order to represent delay dynamics of epidemics, exposure t_{exp}, infection t_{inf}, and vaccination immunization t_{vac} periods are taken into account. The expected value (EV) of the transition from the Susceptible (S) to the Exposed (E) state is accomplished according to the parameter β_{exp}, as

$$EV_{S\to E[1]} = \frac{\beta_{exp}\cdot\Delta t\cdot N_S\cdot\lambda}{N} \tag{1}$$

where Δt is a simulation step size, N_S is the total number of individuals in the Susceptible state (S), N is the total population, and λ is the weighted population estimated as

$$\lambda = \sum_{i=1}^{n_{inf}} \boldsymbol{I}[i] + \epsilon_{sev}\sum_{i=1}^{n_{inf}} \boldsymbol{SI}[i] + \epsilon_{sev}\sum_{i=1}^{n_{inf}} \boldsymbol{Iso}[i] + \epsilon_{exp}\sum_{i=1}^{n_{exp}} \boldsymbol{E}[i] + \epsilon_{qua}\sum_{i=1}^{n_{exp}} \boldsymbol{Q}[i] \tag{2}$$

where the square brackets $[\cdot]$ are used as an indexing operator to access elements of the state vectors. n_{exp} and n_{inf} are the number of substates in the Exposed (E) and Infected (I) states. Iso and SI are the Isolated and Severe Infected states inside the Infected superstate (I^S), and Q is the Quarantined state in the Exposed superstate (E^S). The number of substates n_{exp} and n_{inf}, along with the number of substates n_{vac} in the Vaccination state (V), are calculated as

$$\begin{aligned} n_{exp} &= \lceil t_{exp}/\Delta t\rceil \\ n_{inf} &= \lceil t_{inf}/\Delta t\rceil \\ n_{vac} &= \lceil t_{vac}/\Delta t\rceil \end{aligned} \tag{3}$$

The transition from the Exposed (E) to the Quarantined (Q) state is governed by the daily quarantine rate qr. The transition rate from the Exposed superstate (E^S) to the Infected superstate (I^S) is equal to one. Because after the end of the incubation period t_{exp}, all individuals transfer from the Exposed (E) and Quarantined (Q) states to the Infected (I) and the Isolated (Iso) states, respectively.

The main difference of this model from our previous model [15] is the introduction of the Severe Infected (SI) within the Infected superstate (I^S). Individuals move from the Infected (I) and Isolated (Iso) states to the Severe Infected (SI) according to the severe infection rate sir.

The natural death rate is neglected in the Infected (I) and Isolated (Iso) states and the transition to the Dead state (D) occurs only through the Severe Infected state (SI). This is done to model a high number of asymptomatic and mild cases which carry the disease. The natural birth and death rates are ignored due to their negligible impact over the relatively short duration of the simulation. Individuals move from the Infected (I) and Isolated (Iso) states to the Recovery Immunized (RI) state according to the recovery immunization rate γ_{rim}, and the rest make a transition to the Susceptible state (S).

The mortality rate used in the model is divided into γ_{mor1} and γ_{mor2} depending on the hospital capacity: γ_{mor1} is applied when the number of people in the Severe Infected state (SI) is below the hospital capacity, otherwise the model uses γ_{mor2} mortality rate which is greater than γ_{mor1}. Finally, the individuals move from the Severe Infected state (SI) to the Recovery Immunized (RI) and Susceptible (S) states. In case, if the hospital capacity is not exceeded then the expected transition values are calculated as

$$EV_{SI[n_{inf}]\to S} = \boldsymbol{SI}[n_{inf}]\cdot(1-\gamma_{mor1})\cdot(1-\gamma_{rim}) \tag{4}$$

$$EV_{SI[n_{inf}]\to RI} = \boldsymbol{SI}[n_{inf}]\cdot(1-\gamma_{mor1})\cdot\gamma_{rim} \tag{5}$$

otherwise,

$$EV_{SI[n_{inf}]\to S} = \boldsymbol{SI}[n_{inf}]\cdot(1-\gamma_{mor2})\cdot(1-\gamma_{rim}) \tag{6}$$

$$EV_{SI[n_{inf}]\to RI} = \boldsymbol{SI}[n_{inf}]\cdot(1-\gamma_{mor2})\cdot\gamma_{rim} \tag{7}$$

This way, when an iteration of the simulation ends, the next state is estimated stochastically based on the expected transition values. The parameters of the model and an initial number of individuals in each state must be defined before running the simulator.

B. A Network Model for Epidemic Simulation

A network model, in the epidemiological context, is the representation of a country as a graph of interconnected nodes, where an individual node denotes an administrative unit of the country, such as a city or region, taking into account their respective population demographics and health care capacities. This is a convenient representation since measures during epidemics are usually taken for specific administrative units. The transportation connections between the regions can also be modeled as the edges of the network. Each node in the network model runs concurrently as a separate instance of the single-node SEIR model described in Section III-A. The internal parameters of each SEIR model can be changed independently. Thus, the network model allows for viral propagation, with corresponding state transitions, within and amongst the nodes. To model the network transitions open source data on the daily domestic air, rail, and highway travel between different regions are collected. For a network with M nodes, $T_{air}\in\mathbb{Z}^{M\times M}$, $T_{rail}\in\mathbb{Z}^{M\times M}$ and $T_{highway}\in\mathbb{Z}^{M\times M}$ represent the corresponding transition matrices for the mentioned means of transportation. The total transition matrix $T_{total}\in\mathbb{Z}^{M\times M}$ can be calculated using the Algorithm 1.

Our simulator allows disconnection of a particular node from the list of air, highway and railway transitions, to simulate the effects of quarantine policies. Also, to facilitate the fine-tuning of the transportation between the nodes, parameters traffic ratio tr and a leakage ratio lr are introduced. The traffic ratio is a real-valued coefficient in the range [0,1] by which the total transition matrix is multiplied. The default value of tr is 1, i.e. the transition between the nodes corresponds to the normal case. Then, by lowering the traffic ratio, the number of individuals transferring between the nodes can be decreased. This can be used to simulate the policies which throttle down the transportation between regions. The leakage ratio is also a real-valued coefficient in the range [0, 1] that allows minor transitions even though a node is disconnected from the network. By multiplying the leakage ratio with the transition number in the normal case, we impose some transitions to the disconnected node. In this

Algorithm 1: Calculate Total Transition Matrix.

input : $T_{air}, T_{rail}, T_{highway}, tr, lr$
output: T_{total}
$M \leftarrow 17$ // number of nodes
// find transition matrix for normal case
$T_{base} \leftarrow T_{air} + T_{rail} + T_{highway}$
// change node connections and update
// transition matrices based on GUI
$T'_{air}, T'_{rail}, T'_{highway} \leftarrow update\_matrix(T_{air}, T_{rail}, T_{highway})$
// find total transition matrix
// considering traffic ratio tr
$T_{total} \leftarrow (T'_{air} + T'_{rail} + T'_{highway}) * tr$
// find matrix entries with no transition
for $i = 0$ **to** M **do**
  for $j = 0$ **to** M **do**
    if $T_{total}[i, j] == 0$ **then**
      // add leakage transition
      $T_{total}[i, j] \leftarrow T_{base}[i, j] * lr$
    end
  end
end
return T_{total}

way, we take into account cases where individuals cross the quarantined borders illegally or for transporting essentials.

To simulate the network transition between nodes, at each sampling time of the network simulator ΔT, the randomly sampled population is transferred from one node to another according to the transition matrix. It is important to note that the sampling times ΔT used for the network and Δt for the node are different ($\Delta T > \Delta t$). The expected value of transition from state K of node i to the corresponding state K of node j is given by the following equation:

$$E_{K_i \to K_j} = \frac{T_{ij}}{n_i} \cdot s_i^K \tag{8}$$

where T_{ij} is the population transfer from node i to node j, n_i is the population of node i, s_i^K is the number of individuals in the state K of node i. Some of the states such as Quarantined and Severe Infected are excluded from the transfer between the nodes, i.e. $K \notin \{Q, SI, Iso, D\}$.

C. Simulator Implementation

The code for the simulator was written in the Python programming language. In order to run the SEIR model in all regions in parallel, we used the *multiprocessing* module of Python. A GUI for the simulator (see Fig. 2) was developed using interactive visualization library *Bokeh* [22], which is available under the BSD license. It utilizes web-browsers to visualize data which simplifies online deployment and enables platform independence. The source code[1] of our simulator was uploaded to GitHub under the BSD license.

[1]github.com/IS2AI/COVID-19-Simulator

TABLE II
LOMBARDY COVID-19 TIMELINE

Day	Date	Death	Confirmed cases	Event
0	1/1/2020	unk.	1	The first confirmed case [24].
54	24/2/2020	6	172	The COVID-19 data repository was launched [23].
61	2/3/2020	38	1,254	"Red zone" in Lombardy.
67	8/3/2020	267	4,189	Whole Lombardy locked down.
70	11/3/2020	617	7,280	Bars, restaurants are closed.
81	22/3/2020	3,456	27,206	Factories and all nonessential productions are closed.
151	31/5/2020	16,112	88,968	Date of last data available.

The components of the GUI are illustrated in Fig. 2. The number of individuals in each state for the selected region and for the whole country are shown on the top side (1) in real-time. Also, the starting date of the simulation and the current date are provided here. The epidemic dynamics versus time are visualized in the upper left plot (2). On the right side of this plot, the epidemic heat map (3) is displayed. Also, the GUI allows the user to repeatedly save simulation results and parameters of the model to a CSV file (6); the saved parameters of the model can be reloaded (7, 8) to continue the simulation from a prior point. Also, it is possible to reset the parameters to default values (4). Simulation parameters can be selected for each node (9) and for the whole network (10). The transition matrix can be adjusted using checkboxes (11) and slider bars to adjust traffic and leakage ratios. The resulting transition matrix is shown at the bottom part of the interface (12).

IV. Simulations and Results

A. Single Node Epidemic Simulation of Lombardy (Italy)

In order to test the capability of our simulator and fine-tune its parameters, we first focused on the simulation of the Lombardy region in Italy. We chose Lombardy since it is the epicenter of the COVID-19 outbreak in Italy, its epidemic timeline is well-established, and the Italian government has shared comprehensive daily epidemic data since 24 February 2020 [23], thus facilitating the calibration of model assumptions against actual events and data. The population of Lombardy is just over 10 million people, therefore, we use this number to define the initial number of susceptible individuals in our model. We start the simulation on 1 January 2020 based on the results presented in [24]. We set the number of exposed individuals to 10 on this date, as this figure yields accurate model state populations going forward. In order to tune parameters of the model, we use only the reported death cases for the Lombardy region for the period between 24 February and 31 May 2020. We do not use the total number of infected and recovered cases because the official numbers might be under-reported due to asymptomatic and mild infected cases [25], [26]. Also, we take into account the timeline of events and government policy decisions that would impact the simulation (see Table II).

The model parameters are summarized in Table III. The parameter β_{exp} was tuned according to the events described

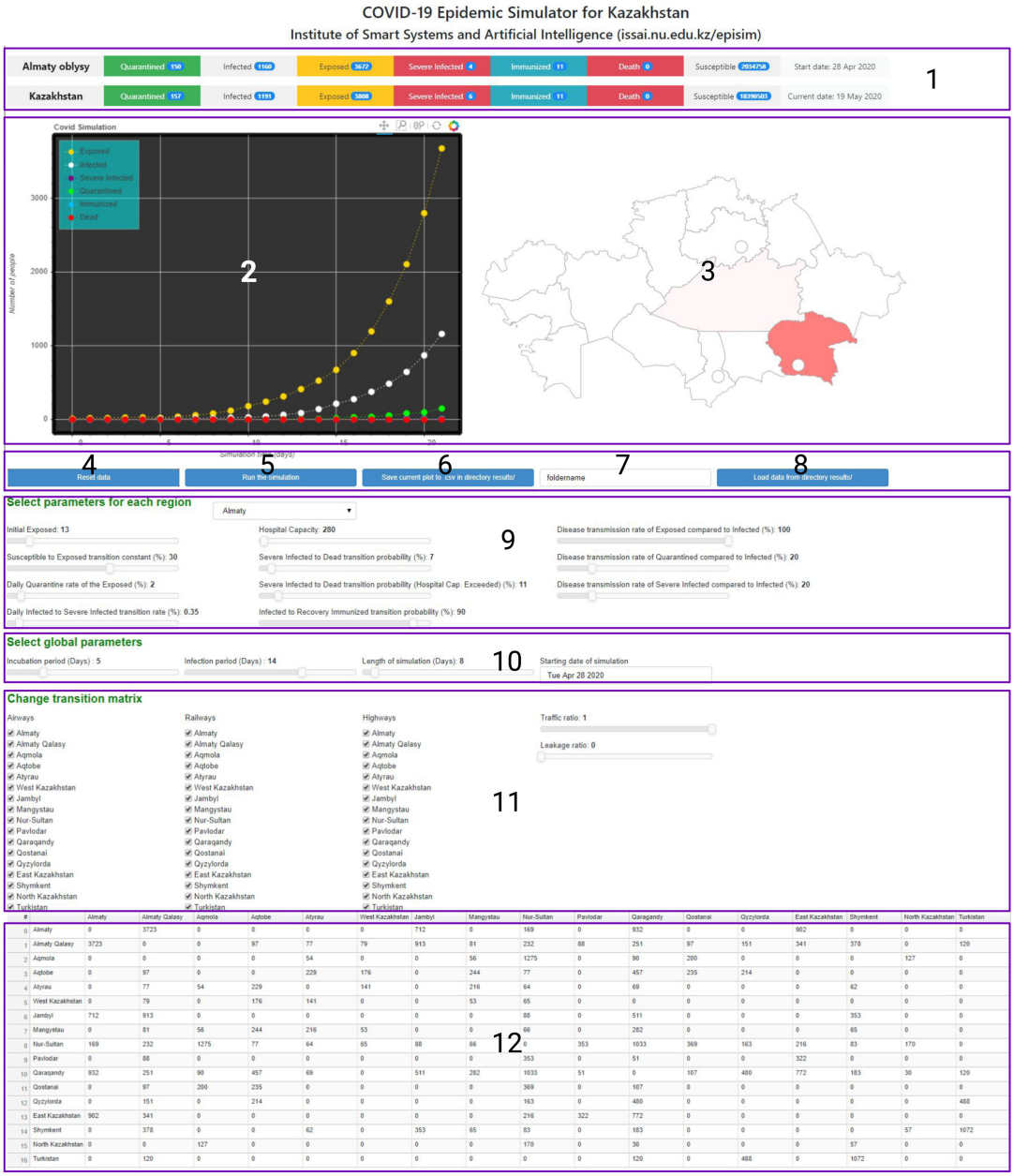

Fig. 2. The graphical user interface (GUI) of the simulator: 1 - Current values of states for a selected region and the whole country, 2 - Plot showing the time evolution of the states for a selected node, 3 - Interactive map of the country (in this case Republic of Kazakhstan), 4 - Button to reset current parameters, 5 - Button to start the simulation, 6 - Button for saving results and model parameters, 7 - Text box for inserting the path to save files, 8 - Button for loading a saved simulation, 9 - Sliders to set the parameters of a selected region, 10 - Sliders for choosing the global simulation parameters, 11 - Checkboxes to configure the transition matrix, and 12 - Transition matrix.

in Table II. The model was simulated for 180 days with the sampling time $\Delta t = 1/24$ days.

Due to the stochastic nature of the simulator, the model was simulated 100 times and the averaged results were obtained. The results of the simulation are shown in Fig. 3 with the standard deviations for the Infected and Dead states. The figure shows that the model closely fits the reported number of deaths for the given period of time: the simulator predicted 16,110 death cases on 31 May 2020 while the reported number was 16,112. Also, we observe that the epidemic reaches its peak around 26 March 2020 where the total number of infected individuals is 69,022 while the reported number was 34,889. The main source of such difference might be due to asymptomatic or mild infected cases which did not meet the testing profile and thus were not registered. Also, if we look at the reported data, new daily confirmed cases increased significantly starting from 19 March 2020 (2,171 cases) and remained high until 28 March 2020 (2,117 cases). After this period, the new daily cases started to gradually decrease. Thus, we can assume that the model estimated the peak of the spread of COVID-19 accurately. Also,

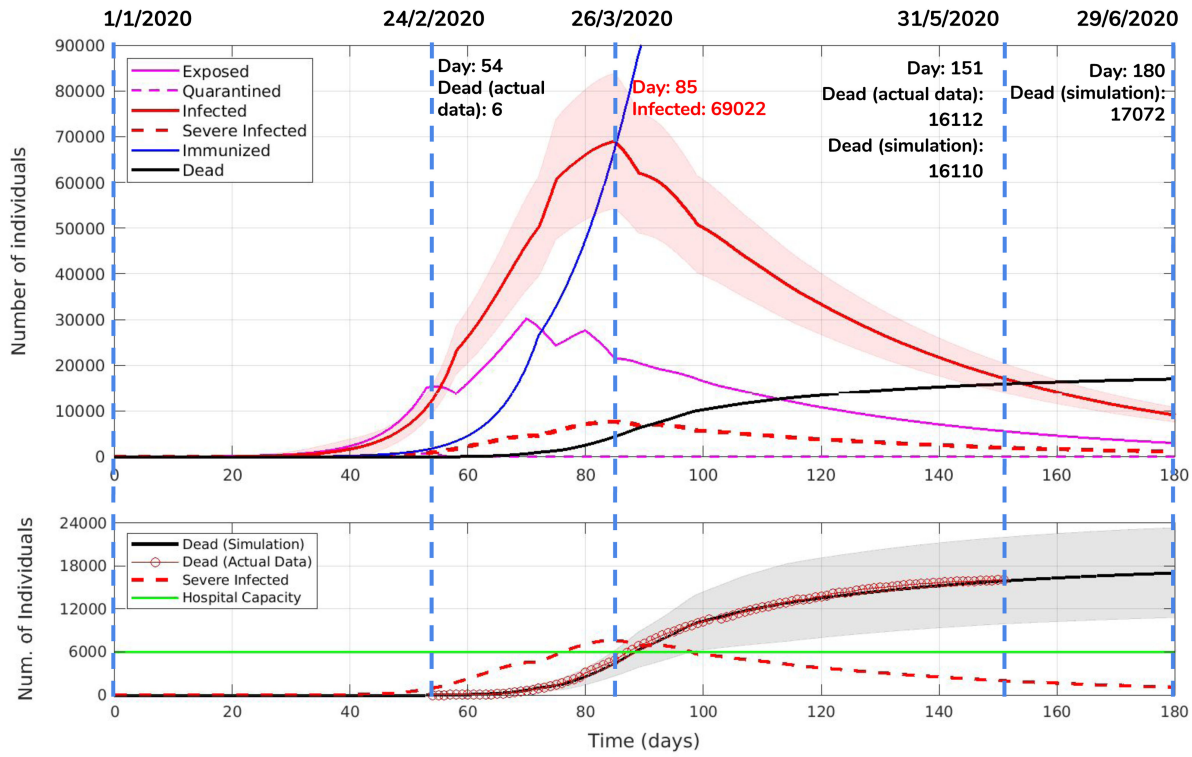

Fig. 3. The averaged results of 100 simulations for the Lombardy region. Upper plot shows the states of the epidemic simulation versus time. The vertical dashed lines indicate events from the timeline in the Table II. One standard deviation around the average Infected state curve is shaded. Bottom plot shows the number of Severe Infected individuals along with the hospital capacity and also compares the number of deaths in the simulation to the actual number of deaths attributed to COVID-19. One standard deviation around the average Dead state curve is shaded.

TABLE III
SINGLE NODE SIMULATION PARAMETERS FOR LOMBARDY

qr	sir	hc	γ_{rim}	γ_{mor1}	γ_{mor2}
0.02	0.015	6,000	0.9	0.19	0.44
β_{exp} $(Days)$	ε_{exp}	ε_{qua}	ε_{sev}	t_{exp}	t_{inf}
0.2 (0-55) 0.1 [55-71) 0.067 [71-82) 0.05 [82-180)	0.7	0.3	0.3	5	14

TABLE IV
NAMES, ABBREVIATIONS, POPULATION AND HOSPITAL CAPACITY OF THE NODES

Node	Name	Abbr.	Population	Hospital capacity
1	Almaty	AAO	2,039,376	280
2	Almaty Qalasy	AAK	1,854,556	2,395
3	Aqmola	AQM	738,587	895
4	Aqtobe	AQT	869,603	600
5	Atyrau	ATY	633,801	650
6	West Kazakhstan	WKZ	652,314	250
7	Zhambyl	ZHA	1,125,297	725
8	Mangystau	MAN	678,224	100
9	Nur-Sultan	NST	1,078,362	885
10	Pavlodar	PAV	753,804	425
11	Qaragandy	QAR	1,378,554	1,670
12	Qostanay	QOS	872,736	300
13	Qyzylorda	QYZ	794,165	465
14	East Kazakhstan	EKZ	1,378,504	1,420
15	Shymkent	SMK	1,011,511	1,505
16	North Kazakhstan	NKZ	554,519	380
17	Turkistan	TUR	1,981,747	300

according to the simulation results, we forecast that the epidemic might continue until the end of June and the number of total deaths might exceed 17,000 in Lombardy.

B. Network-Based Simulation of the COVID-19 in Kazakhstan

1) Model Parameters and Validation: As noted, we tested the model's assumptions against prior epidemic experience and then customized the model to simulate the spread of COVID-19 in the Republic of Kazakhstan. We performed model calibration and validation by replicating the observed COVID-19 numbers in the country.

Epidemics usually start in transportation hubs with the import of the disease from another country and propagates to other regions via transportation links amongst regions. The first confirmed instances of COVID-19 patients in Kazakhstan were detected in the largest cities, Nur-Sultan and Almaty, and then the disease spread to other regions [27]. These two cities serve as transportation hubs as highlighted in Fig. 4.

Table IV lists all nodes with the respective abbreviations, population numbers, and hospital capacities for Kazakhstan. The transition matrix between nodes representing the domestic movements was estimated based on government records and available statistical data (see Table V), and captures the end destination numbers, not accounting for transit points. For instance, if

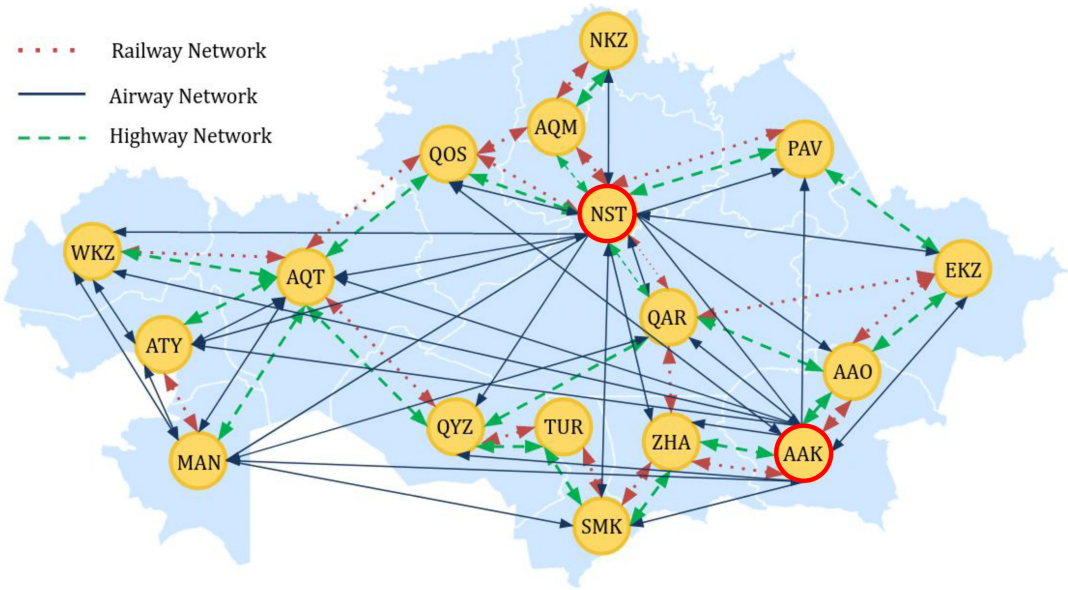

Fig. 4. Kazakhstan as a network: Nodes representing the administrative divisions (regions and big cities) of the country are overlaid on the map. The edges of the network represent the highway, railway, and airway connections between different administrative divisions.

TABLE V
THE TOTAL TRANSITION MATRIX BETWEEN THE NODES WHICH IS THE SUM OF THE DOMESTIC AIR, HIGHWAY AND RAIL TRANSITION MATRICES

Node	AAO	AAK	AQM	AQT	ATY	WKZ	ZHA	MAN	NST	PAV	QAR	QOS	QYZ	EKZ	SMK	NKZ	TUR
AAO	0	7,447	0	0	0	0	1424	0	338	0	1,865	0	0	1805	0	0	0
AAK	7,447	0	0	195	154	158	1827	163	465	176	503	195	303	683	756	0	240
AQM	0	0	0	0	109	0	0	113	2,551	0	180	400	0	0	0	254	0
AQT	0	195	0	0	459	352	0	448	154	0	914	471	429	0	0	0	0
ATY	0	154	109	459	0	283	0	431	128	0	139	0	0	0	124	0	0
WKZ	0	158	0	352	283	0	0	106	130	0	0	0	0	0	0	0	0
ZHA	1,424	1,827	0	0	0	0	0	0	176	0	1,023	0	0	0	706	0	0
MAN	0	163	113	488	431	106	0	0	133	0	565	0	0	0	130	0	0
NST	338	465	2,551	154	128	130	176	133	0	706	2067	738	326	433	167	340	0
PAV	0	176	0	0	0	0	0	0	0	706	103	0	0	645	0	0	0
QAR	1,865	503	180	914	139	0	1,023	565	2,067	103	0	214	960	1545	366	60	240
QOS	0	195	400	471	0	0	0	0	738	0	214	0	0	0	0	0	0
QYZ	0	303	0	429	0	0	0	0	326	0	960	0	0	0	0	0	977
EKZ	1,805	683	0	0	0	0	0	0	433	645	1,545	0	0	0	0	0	0
SMK	0	765	0	0	124	0	0	706	130	167	0	366	0	0	0	114	2,144
NKZ	0	0	254	0	0	0	0	0	340	0	60	0	0	0	114	0	0
TUR	0	240	0	0	0	0	1,384	0	0	0	240	0	977	0	2,144	0	0

one travels from the capital to another region via an intermediate city, only the departure and arrival nodes were considered in the transition matrix.

The timeline, introduced in Table VI, of the governmental policies related to COVID-19 have been used to dynamically adjust model parameters. Since the first confirmed-positive cases are dated 13 March 2020 in Nur-Sultan and Almaty cities, we started the simulation from 1 March 2020 with 10 exposed individuals in each of these two cities. The simulation parameters are summarized in Table VII. The node and network sampling times were set as $\Delta t = 1/24$ and $\Delta T = 1/2$ days. The leakage ratio was set to 0.1 to accommodate the opening of airway

TABLE VI
KAZAKHSTAN COVID-19 SIMULATION SCENARIO BASED ON THE TIMELINE FROM 1 MARCH TO 31 MAY 2020 [27]

Day	Date	Death	Confirmed cases	Event
0	1/3/2020	0	0	Start of the simulation.
12	13/3/2020	0	3	The first confirmed cases.
18	19/3/2020	0	49	"Red zone" in NST and AAK.
21	22/3/2020	0	56	NST and AAK locked down.
26	27/3/2020	0	150	"Red zone" in 6 other regions.
30	31/3/2020	2	340	The whole country locked down.
72	1/5/2020	31	5,200	End of the national emergency state.
93	31/5/2020	78	10,858	Date of the last data available

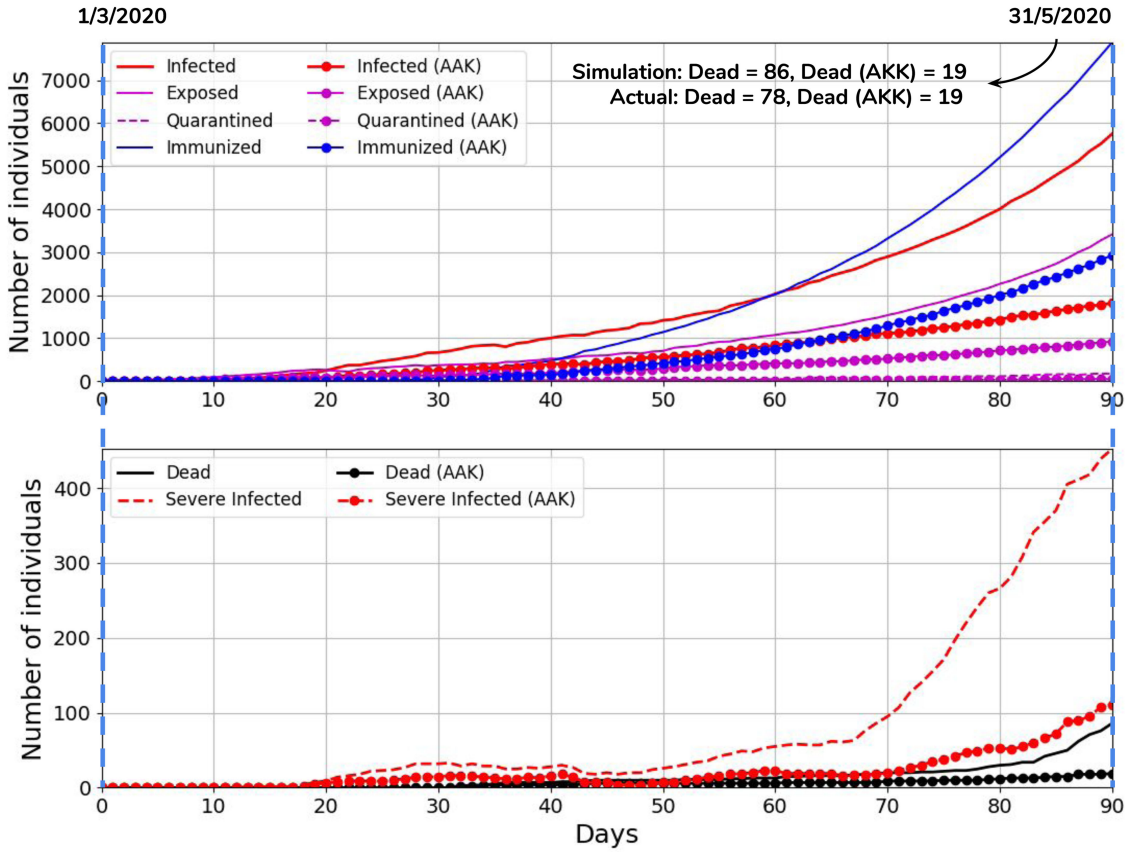

Fig. 5. Simulation results for Almaty city (AAK) and whole Kazakhstan from 1 March 2020 to 31 May 2020. Upper plot shows the number of Infected, Exposed, Quarantined and Immunized states versus time. The bottom plot shows the number deaths and severe infected versus time.

TABLE VII
NETWORK-BASED SIMULATION PARAMETERS FOR KAZAKHSTAN

t_{exp}	t_{inf}	sir	γ_{rim}	γ_{mor1}	γ_{mor2}
5	14	0.0125	0.9	0.2	0.4
β_{exp} for KZ & AAK, NST $(Days)$	qr	tr	ε_{exp}	ε_{qua}	ε_{sev}
0.27 & 0.27 [0-12)		1			
0.25 & 0.20 [12-18)		1			
0.20 & 0.12 [18-21)	0.02	0.8	0.7	0.3	0.3
0.17 & 0.10 [21-26)		0.8			
0.12 & 0.08 [26-31)		0.7			
0.10 & 0.08 [32-52)		0			
0.11 & 0.09 [52-72)		0			
0.16 & 0.10 [72 - 93]		0.3			

transportation after 11 May 2020 once the national emergency was lifted.

The simulation results are shown in Fig. 5. On 31 May 2020, the actual number of COVID-19 deaths in the country was 78 while the predicted number of deaths is 86. For the infected state, the officially reported and simulation numbers are 5,032 and 6,830, respectively. The discrepancy might be due to the stochastic nature of the simulator and limited testing which might result in a lower number of officially registered cases [27]. As for the Almaty city, the simulation and reported statistics both conclude 19 dead by 31 May 2020. Overall, this suggests that the model was calibrated accurately and can be utilized to simulate future scenarios.

C. COVID-19 Control Strategies

1) Four Strategies for Kazakhstan: After establishing the initial conditions for Kazakhstan, and calibrating the state transitions, the next step was to model outcome scenarios for four variations of policy intervention: 1) complete quarantine, 2) going back to normal life, 3) "New Normal," and 4) augmenting the "New Normal" strategy with technological measures to detect infection zones and facilitate individual exposure detection and contact tracing.

The New Normal strategy introduces measures such as an increase in hospital capacity, unhindered supply of necessary drugs, an effective social distancing of vulnerable people, increased testing, and persistent observation of hygienic precautions and moderate social distancing by the general population.

The introduction of technological monitoring, such as identification of infection zones, and automated notifications to individuals whenever they have visited an infection zone or come into physical proximity with an infected individual, would increase the ability to detect and trace infections. In this scenario, we used much higher quarantine rates due to the presumed greater ability to trace the interaction of individuals in the exposed and susceptible categories. Fig. 6 shows the number of infected and dead versus time for these four strategies simulated from 31 May. Table VIII summarizes the key parameters used in the simulations.

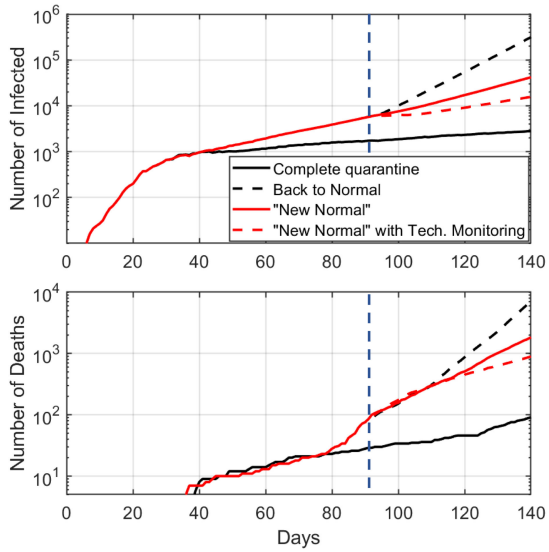

Fig. 6. Simulation results showing infected and deaths versus time for the four control strategies.

TABLE VIII
PARAMETERS FOR THE FOUR CONTROL STRATEGIES

Control Strategy	β_{exp}	tr	qr
Complete Quarantine	0.10	0	0.02
Back to Normal	0.16	1	0.02
New Normal	0.12	1	0.07
New Normal with Technological Monitoring	0.11	1	0.14

The results illustrate the effectiveness of the complete quarantine approach, and the impact of technological measures, compared to the New Normal strategy. Although the quarantine is demonstrably the most effective, it is not sustainable long-term due to extremely negative social and economic impact. As for the second strategy - back to normal lifestyle - simulation shows that it is also impractical due to an exponential second wave of epidemic growth in the country, thus effectively losing all benefits accrued from prior interventions. From these considerations, the optimal policy would be a hybrid approach that balances the effectiveness of quarantine with measures to minimize the b_{exp} and increasing quarantine rate qr using high-tech monitoring.

2) Region Specific Policies: To illustrate the effect of the region-specific policies, a variation of the "New Normal" scenario was considered. After a month of "New Normal" (from 1 July to 31 July 2020), six nodes were disconnected from the transportation networks and their quarantine rate was increased (see Table IX). These nodes are two cities with the largest number of infected (Nur-Sultan and Almaty Qalasy), two regions surrounding these cities (Aqmola and Almaty), Qaragandy region that is connected to both regions, and West Kazakhstan, a far region with the smallest number of infected.

The strategy of disconnecting the nodes from the transportation network is effective in preventing the spread of the infection. In the case of the "New Normal" with a full transition matrix, the number of deaths was 3,117. The "New Normal" enhanced with the region-specific policies resulted in 2,164 deaths for the same period.

TABLE IX
PARAMETERS FOR THE REGION SPECIFIC INTERVENTION SIMULATION SCENARIO

Node	β_{exp}	qr
Almaty city (AAK)	0.09	0.09
Nur-Sultan city (NST)	0.08	0.09
Almaty region (AAO)	0.10	0.07
Akmola region (AQM)	0.10	0.07
Qaragandy region (QAR)	0.10	0.07
West Kazakhstan region (WKZ)	0.10	0.07
Other regions	0.16	0.07

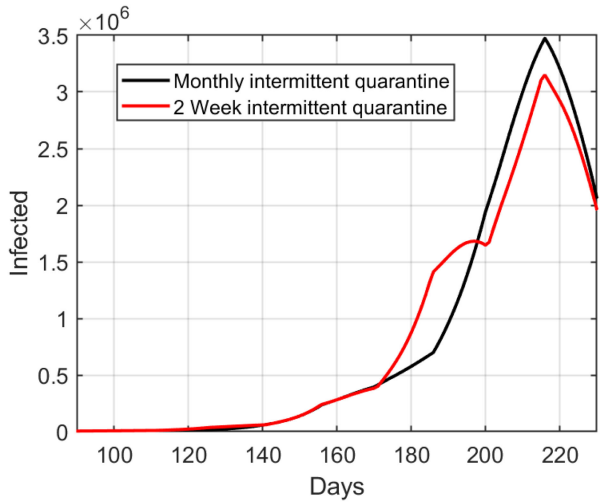

Fig. 7. Number of infected versus time for intermittent quarantine strategy simulations with monthly and weekly intervals.

3) Intermittent Quarantine Strategy: From the four different simulation scenarios, the complete quarantine resulted in the smallest number of infected and dead. But the social and economic aspects of this policy makes it not sustainable. Therefore, we tried to analyze the case of switching quarantine on and off to evaluate its efficiency in slowing down the infection spread. We considered an intermittent quarantine regime, where β_{exp} and tr values were 0.16 and 1 for the "quarantine off" period, and 0.1 and 0 for the "quarantine on" period respectively. Simulations for two different quarantine on and off periods have been performed: one month and two weeks. The simulation results in Fig. 7 illustrate that the number of infected individuals at the peak of the epidemic is higher for a longer period (one month) thus puts more pressure on the healthcare system.

4) Effects of Transportation Limitations: Our simulator incorporates domestic population mobility using the transition matrix. The inclusion of mobility makes the model more realistic, with more nuanced suppression policies. To illustrate the effect of transportation limitations on the infection spread, we simulated the validation scenario with four different traffic ratios (0, 0.33, 0.66 and 1) starting from 31 May 2020. The peak number of infected increases with the increasing traffic ratio as shown in the Fig. 8.

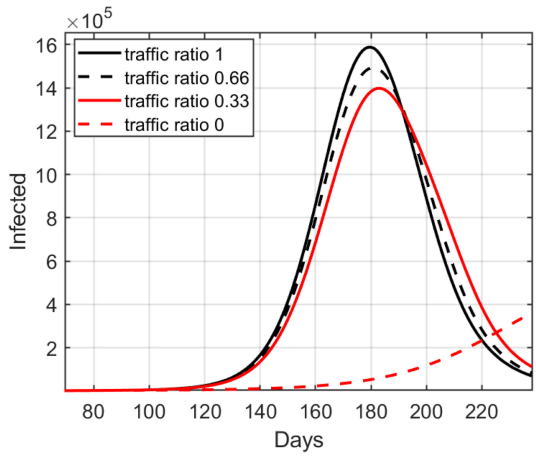

Fig. 8. Effect of transportation limitations on the epidemic spread.

The traffic ratio of zero exemplifies complete stop of transportation between regions, which might not be feasible for both practical and economic reasons.

V. Discussion

The network-of-nodes design of the stochastic epidemic simulator allows granular adjustments to the model parameters at the level of individual regions (nodes), and the interactive controls allow for testing policy interventions at discrete time intervals, also, unique to each region. The inclusion of mobility transitions between regions also allows for more precise simulation of events.

The calibration of the model presents a conundrum, in that if we fit the parameters of the model to the reported data on confirmed cases, and set the state transition rates based on those reports, then the simulator outcomes will be dependent on the relative accuracy of the reported data, and factors such as physical limits on the number of tests that can be conducted could skew the outcomes.

In the case of COVID-19, it is likely that the number of infected individuals, and the number of deaths attributed to it, are higher than the number of laboratory-confirmed cases, given the relatively uneven rates of testing across different populations, and wide variations in the percentage of positive tests, along with the variations in mortality rates, and the concurrent excess mortality observed in many of the reporting zones. Perhaps the single greatest factor is that most countries only test if a patient meets screening criteria, with explicit symptoms, but exposed patients can remain asymptomatic for approximately five days, and a non-trivial percentage of those infected may not seek medical care, due to light symptoms, and the infection may not be recognized or registered as COVID-19 due to sensitivity and specificity of rapid tests available on the market [28].

We have run many scenarios, and, often, the projections of the simulator ran higher than the reported figures. This was an initial concern, until updated reporting inclusive of testing rates, positive rates, mortality, and excess mortality were taken into account, which suggested that the preliminary reports were simply incomplete. Thus, if we run the simulator tightly, calibrating and updating based only on the reported confirmed case data, then we risk essentially modeling for those with severe enough symptoms that they require medical care, but by doing so the model will only capture a portion of the impact and skew the population sizes for the Susceptible and Recovered states. The significance of this outcome becomes greater over time, as it determines whether and when a population reaches herd-immunity levels, and so too, the duration and extent of NPI policies.

In order to address this concern, it is necessary to establish a more rigorous testing regimen, in accordance with standard testing protocols, to accurately estimate the size of each population state. In this case, the advantage of the node-based approach lies in the opportunity to adjust the policies at the regional level, and thereby reduce the potentially unnecessary economic and social disruption of a country-wide lockdown, and strategically smooth the pressure on hospitals in different regions.

Our model can be also used to simulate metapopulation models, introduced by Colizza and Vespignani [29], to simulate the spread of the epidemic in metropolitan cities. The metapopulation is composed of a heterogeneous network of subpopulations, connected by migration processes. Each subpopulation consists of susceptible, infected, and recovered individuals, and they can migrate from one subpopulation to another via connections between subpopulations. The metapopulation models are based on the simple assumption that the considered area is a highly fragmented environment which might not be true for metropolitan cities with dense populations. Therefore, the accuracy of the model will highly depend on the detailed knowledge of transportation infrastructures and movement patterns in the considered city.

VI. Conclusion

We have designed and implemented an open-source network-based stochastic epidemic simulator that models the movement of a disease through the SEIR states of a populace. The simulator incorporates interactive controls of state transition probabilities and modification of environmental factors such as health care capacity and mobility amongst regions.

The simulator was calibrated with recent data on COVID-19 in the Lombardy region of Italy. The simulator was then configured to model the COVID-19 outbreak in Kazakhstan, using empirical observations and authoritative data on initial conditions. Afterward, the simulator was used to evaluate scenarios to minimize strain on the health care system while reducing negative social and economic impact by balancing the containment of disease spread and the imposition of less stringent social controls on a localized basis.

On one extreme, the simulator indicates that a lack of mitigation will yield the unmitigated disaster. On the other extreme, persistent suppression (via quarantines and strict social distancing) will curtail disease propagation and reduce mortality, but also cause economic harm, and instigate social distress.

The projections of the simulator suggest that going forward, it will be necessary to maintain some level of social controls guided by a comprehensive testing regimen so as to constrain propagation while minimizing social and economic impact, until such time as more widespread immunity amongst the population (so-called "herd" immunity) is gradually achieved, or a vaccine is introduced.

References

[1] M. Chinazzi *et al.*, "The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak," *Sci.*, vol. 368, no. 6489, pp. 395–400, 2020. [Online]. Available: https://science.sciencemag.org/content/early/2020/03/05/science.aba9757

[2] M. Lazzerini and G. Putoto, "COVID-19 in Italy: Momentous decisions and many uncertainties," *Lancet Global Health*, vol. 8, no. 5, pp. e641–e642, May 2020. [Online]. Available: https://doi.org/10.1016/S2214-109X(20)30110-8

[3] C. J. Wang, C. Y. Ng, and R. H. Brook, "Response to COVID-19 in Taiwan: Big data analytics, new technology, and proactive testing," *J. Amer. Med. Assoc.*, vol. 323, no. 14, pp. 1341–1342, Apr. 2020. [Online]. Available: https://doi.org/10.1001/jama.2020.3151

[4] M. Biggerstaff, S. Cauchemez, C. Reed, M. Gambhir, and L. Finelli, "Estimates of the reproduction number for seasonal, pandemic, and zoonotic influenza: A systematic review of the literature," *BMC Infectious Diseases*, vol. 14, no. 1, pp. 480–499, 2014. [Online]. Available: https://doi.org/10.1186/1471-2334-14-480

[5] N. Ferguson *et al.*, "Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand," 2020. [Online]. Available: https://www.imperial.ac.uk/media/imperial-college/medicine/sph/ide/gida-fellowships/Imperial-College-COVID19-NPI-modelling-16-03-2020.pdf

[6] L. Ferretti *et al.*, "Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing," *Sci.*, vol. 368, no. 6491, 2020, Paper eabb6936. [Online]. Available: https://science.sciencemag.org/content/early/2020/04/09/science.abb6936

[7] F. Zhou *et al.*, "Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: A retrospective cohort study," *Lancet*, vol. 395, no. 10229, pp. 1054–1062, 2020. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0140673620305663

[8] W. O. Kermack and A. G. McKendrick, "A contribution to the mathematical theory of epidemics," *Proc. Roy. Soc. London. Ser. A*, vol. 115, no. 772, pp. 700–721, 1927. [Online]. Available: http://www.jstor.org/stable/94815

[9] H. W. Hethcote, "The mathematics of infectious diseases," *SIAM Rev.*, vol. 42, no. 4, pp. 599–653, Dec. 2000. [Online]. Available: https://doi.org/10.1137/S0036144500371907

[10] F. Brauer, "Mathematical epidemiology: Past, present, and future," *Infectious Disease Modelling*, vol. 2, no. 2, pp. 113–127, 2017. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S2468042716300367

[11] L. J. Allen and A. M. Burgin, "Comparison of deterministic and stochastic SIS and SIR models in discrete time," *Math. Biosci.*, vol. 163, no. 1, pp. 1–33, 2000. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0025556499000474

[12] P. Renard, A. Alcolea, and D. Ginsbourger, *Stochastic Versus Deterministic Approaches*. Hoboken, NJ, USA: Wiley, 2013, ch. 8, pp. 133–149. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/9781118351475.ch8

[13] A. J. Tatem, D. J. Rogers, and S. I. Hay, "Global transport networks and infectious disease spread," *Adv. Parasitol.*, vol. 62, pp. 293–343, 2006, Art. no. 16647974[pmid]. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/16647974

[14] A. Browne, S. St-Onge Ahmad, C. R. Beck, and J. S. Nguyen-Van-Tam, "The roles of transportation and transportation hubs in the propagation of influenza and coronaviruses: A systematic review," *J. Travel Med.*, vol. 23, no. 1, 2016, Paper tav002. [Online]. Available: https://doi.org/10.1093/jtm/tav002

[15] H. A. Varol, "MOSES: A Matlab-based open-source stochastic epidemic simulator," in *Proc. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2016, pp. 2636–2639.

[16] A. R. Tuite, D. N. Fisman, and A. L. Greer, "Mathematical modelling of COVID-19 transmission and mitigation strategies in the population of Ontario, Canada," *Can. Med. Assoc. J.*, vol. 192, no. 19, pp. E497–E505, 2020. [Online]. Available: https://www.cmaj.ca/content/early/2020/04/09/cmaj.200476

[17] K. Prem *et al.*, "The effect of control strategies to reduce social mixing on outcomes of the COVID-19 epidemic in Wuhan, China: A modelling study," *Lancet Public Health*, vol. 5, no. 5, pp. e261–e270, May 2020. [Online]. Available: https://europepmc.org/articles/PMC7158905

[18] J. Wangping *et al.*, "Extended SIR prediction of the epidemics trend of COVID-19 in Italy and compared with Hunan, China," *Frontiers Med.*, vol. 7, pp. 169–175, 2020. [Online]. Available: https://www.frontiersin.org/article/10.3389/fmed.2020.00169

[19] V. Colizza, A. Barrat, M. Barthélemy, and A. Vespignani, "The role of the airline transportation network in the prediction and predictability of global epidemics," *Proc. Nat. Acad. Sci.*, vol. 103, no. 7, pp. 2015–2020, 2006. [Online]. Available: https://www.pnas.org/content/103/7/2015

[20] J. Park, J. Jang, and I. Ahn, "Epidemic simulation of H1N1 influenza virus using GIS in South Korea," in *Proc. Int. Conf. Inf. Commun. Technol. Convergence*, 2017, pp. 58–60.

[21] Y. Zhang, Y. Zhang, and Z. Liu, "The role of different transportation in the spreading of new pandemic influenza in mainland China," in *Proc. Int. Conf. Geoinformat.*, 2011, pp. 1–6.

[22] Bokeh, "The Bokeh Visualization Library," 2019. [Online]. Available: https://docs.bokeh.org/en/latest/, Accessed: Apr. 27, 2020.

[23] P. del Consiglio dei Ministri Dipartimento della Protezione Civile, "Dati COVID-19 Italia," [Online]. Available: https://github.com/pcm-dpc/COVID-19, Accessed: Apr. 27, 2020.

[24] D. Cereda *et al.*, "The early phase of the COVID-19 outbreak in Lombardy, Italy," Mar. 2020, *arXiv:2003.09320*.

[25] R. Li *et al.*, "Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus SARS-CoV-2," *Sci.*, vol. 368, no. 6490, pp. 489–493, 2020. [Online]. Available: https://science.sciencemag.org/content/368/6490/489

[26] M. Gatto *et al.*, "Spread and dynamics of the COVID-19 epidemic in Italy: Effects of emergency containment measures," *Proc. Nat. Acad. Sci.*, vol. 117, no. 19, pp. 10 484–10 491, 2020. [Online]. Available: https://www.pnas.org/content/117/19/10484

[27] Kazinform, "Official situation with coronavirus in Kazakhstan," [Online]. Available: https://www.coronavirus2020.kz, Accessed: Apr. 27, 2020.

[28] D. F. Gudbjartsson *et al.*, "Spread of SARS-CoV-2 in the Icelandic population," *New England J. Med.*, Apr. 2020. [Online]. Available: https://doi.org/10.1056/NEJMoa2006100

[29] V. Colizza and A. Vespignani, "Epidemic modeling in metapopulation systems with heterogeneous coupling pattern: Theory and simulations," *J. Theor. Biol.*, vol. 251, no. 3, pp. 450–467, 2008. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0022519307005991