European Journal of Operational Research xxx (xxxx) xxx

Contents lists available at ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

Innovative Applications of O.R.

Predicting the outbreak of epidemics using a network-based approach

Saikat Das[a], Indranil Bose[b,*], Uttam Kumar Sarkar[a]

[a] Indian Institute of Management Calcutta, Diamond Harbour Road, Joka, Kolkata 700104, India
[b] Indian Institute of Management Ahmedabad, Vastrapur, Ahmedabad 380015, India

ARTICLE INFO

Article history:
Received 1 March 2022
Accepted 13 January 2023
Available online xxx

Keywords:
OR in medicine
Covid-19 pandemic
Epidemic model
Networks
Simulation

ABSTRACT

The spread of epidemics is a common societal problem across the world. Can operational research be used to predict such outbreaks? While equation-based approaches are used to model the trajectory of epidemics, can a network-based approach also be used? This paper presents an innovative application of epidemic modelling through the design of both approaches and compares between the two. The network-based approach proposed in this paper allows implementing heterogeneity at the level of individuals and incorporates flexibility in the variety of situations the model can be applied to. In contrast to the equation-based approach, the network-based approach can address the role of individual differences, network properties, and patterns of social contacts responsible for the spread of epidemics but are much more complex to implement. In this paper, we simulated the spread of infection at the beginning of Covid-19 (Coronavirus disease 2019) using both approaches. The results are showcased using empirical data for eight countries. Sophisticated measures, including partial curve mapping, are used to compare the simulated results with the actual number of infections. We find that the plots generated by the network-based approach match the empirical data better than the equation-based approach. While both approaches can be used to predict the spread of infections, we conclusively show that the proposed network-based approach is better suited with its ability to model the spread of epidemics at the level of an individual. Hence, this can be a model of choice for epidemiologists who are interested to model the spread of an epidemic.

1. Introduction

The outbreak of infectious diseases like SARS, H1N1, and Ebola have been frequent occurrences in recent years. What has been uncommon till the outbreak of Covid-19 (Coronavirus disease 2019) is the scale and magnitude of its spread. Covid-19, which emerged first in December 2019 in Wuhan, China, has affected millions and has caused damage to life and livelihood worldwide. Understanding the spread of infection in a complex system such as society is difficult and accurate forecasting of an epidemic is particularly challenging (Hofman et al., 2017; Jasny & Stone, 2017). It is important to model the early spread of an epidemic since such a model enables us to understand the scale, and is necessary for estimating the facilities required to control the spread of disease in the future (Lotfi et al., 2022; J. W. Taylor & Taylor, 2023). In this context, it is also noted that the growth pattern of this infection varied across the countries (Wilinski & Szwarc, 2021). As shown in Fig. 1, during December 2019 to February 2020, Covid-19 began as an epidemic in China and started spreading to other parts of the world through Europe.

Given the catastrophic impact of Covid-19 and the relevance of operational research to address global health issues using its problem-solving techniques (Silal, 2021), it is logical to ask how we can use the innovative applications of operational research to fight the spread of such an epidemic? In this paper, our first research question is:

RQ1: How can we realistically model the spread of the infection at the onset of an epidemic?

There are existing approaches to study the spread of infectious diseases. However, these approaches either focus on a specific region (Renardy et al., 2020) or a country (Alrasheed et al., 2020) and are limited by their assumptions. For example, well-known epidemic models often make assumptions such as fixed transmission rate of infection which do not hold in the current context (J. T. Chang & Kaplan, 2023). There are studies that analyse the spread of epidemics across multiple countries (Appadu et al., 2021) but they primarily rely on techniques that forecast macro-level outcomes rather than micro-level interactions. This creates a research gap and provides a scope for creating a bottom-up approach that can

* Corresponding author.
E-mail address: indranil_bose@yahoo.com (I. Bose).

https://doi.org/10.1016/j.ejor.2023.01.021
0377-2217/

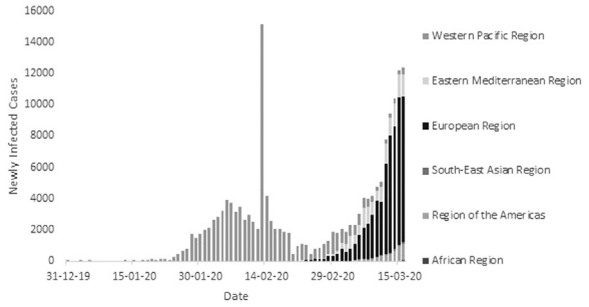

Fig. 1. Spread of Covid-19 at the beginning of the epidemic across the globe.

be applied to study macro-level outcomes based on micro-level interactions and which can be tested for multiple scenarios through simulations (Gupta et al., 2021; Ma & Nakamori, 2005). We fulfil the gap in this paper by first implementing a top-down equation-based approach to model the spread of an epidemic using homogenous parameters values at an aggregate population level. Then we propose a bottom-up network-based approach using heterogeneous values of parameters at an individual level. The parameters used to describe the contagiousness of the epidemic and the mechanism of its spread are discussed in detail. This leads us to our next research question:

RQ2: What would be the impact of varying the parameters of interest in the equation- and network-based approaches on the predicted trajectory of the epidemic?

In modelling the spread of epidemics, multiple input parameters are used. The choice of the value of these parameters, such as the size of the population, the fraction of infected individuals at the beginning, basic reproductive number to describe the contagiousness of the epidemic, and the type of underlying contact network structure, play a vital role in determining its spread. In this regard, we performed simulations using the equation- and the network-based approaches and studied the effect of varying the parameters of interest. We observed the changes in the predicted trajectory of number of infections in response to the variation of the values of parameters for each approach. In our third research question we ask:

RQ3: How can we determine the best fit model based on a comparison between the trajectory of the spread of infections?

To determine the best fit model, we compared the results of the equation- and network-based approaches using empirical data from different countries. For this purpose, we first collected data about the trajectory of newly infected cases for eight countries. These countries are compared using relevant country-level attributes. We formulated a dissimilarity index to assess the similarities between the countries. We used this knowledge to determine if the Covid-19 infections of countries which are similar or dissimilar in terms of their dissimilarity index followed the same pattern for their spread of the epidemic. These countries are further grouped into four clusters based on the visual similarity between their trajectory of newly infected cases. To determine the extent of this visual similarity, we used quantitative measures to compare the trajectory of newly infected cases under various scenarios. We compared the similarity between the patterns of the curves using partial curve mapping (Jekel et al., 2019). A model is considered best fit if it matched the empirical data consistently based on the proposed measures. We found that the trajectories generated by the network-based approach matched the empirical data more closely than the equation-based approach.

This paper contributes to the development and implementation of a novel network-based approach and compares its performance with that of the equation-based approach under different scenarios. We contribute by modelling heterogeneity at the level of an individual in the proposed network-based approach. Second, we incorporate flexibility by running simulations with different contact network structures, different values of parameters describing the infection, and individual attributes. Third, we simulate different scenarios and compare the results with empirical data for eight countries. Apart from contributing to the development of a novel network-based approach, the findings from our research help us identify the spread of infections at the initial stage of an epidemic.

The rest of this paper is structured as follows. Section 2 describes the literature review. The following methodology section describes the assumptions and the techniques of simulation for both the equation- and the network-based approaches. The fourth section is dedicated to the description of data used in implementing the simulations using various parameters. Section 5 reports the results and compares them with empirical data. Section 6 discusses the implications and limitations of this study. Finally, Section 7 summarizes the contribution of this study and concludes the paper.

2. Literature review

Epidemic models are widely used across multiple disciplines (Adly et al., 2020; Bozzani et al., 2021; Camacho et al., 2020), but are less explored in operational research (Pazoki & Samarghandi, 2021; Yaesoubi & Cohen, 2011). Data-driven research using techniques of operational research is relevant for the study of the outbreak of Covid-19, and as such operational researchers are now focussing on ways to fight the epidemic (Choi, 2021; Farahani et al., 2023). Existing epidemic models have been used to forecast aggregate outcomes like the number of infections (Nikolopoulos et al., 2021). The methods employed by the existing models can be grouped under mathematical, computational, and machine learning approaches. Mathematical models that are easier to understand and require low computational power are dominant (Duan et al., 2015), whereas advanced machine learning models have been gaining popularity in recent years.

Mathematical models are the earliest approaches used in epidemic modelling. They are well-established and have been used for modelling the spread of Covid-19, and many other infectious diseases (Brauer et al., 2019; Capasso, 2008; Grave et al., 2021; Martcheva, 2015). Forecasting methods such as time-series, ARIMA, exponential smoothing have been used as well and ARIMA has often outperformed the others (Petropoulos & Makridakis, 2020). However, forecasting the spread of infection in a society can be challenging without observing the system's evolution. Under mathematical models, the most commonly used method is ordinary differential equations (ODEs). Such models have been in great demand since the outbreak of Covid-19 (Grave et al., 2021; Martcheva, 2015). However, such models are limited by their consideration of variations over time and not space and homogenous treatment of individuals. To address the variation between individuals, these models divide the population into subgroups based on the individual's age, infectivity, and occupation (Duan et al., 2015) but are limited in their capability to represent the spread of infection in detail. Finally, these models are also highly dependant on model assumptions and fitting techniques (Alahmadi et al., 2020; Vytla et al., 2021) and so they do not reveal the dynamic dependency of parameters on the epidemic (Masum et al., 2022). For example, these models assume a fixed rate of transmission of an infection which is seldom the case in reality (Chang & Kaplan, 2023). Thus, although equation-based approaches are the natural choice of researchers to capture macro-level dynamics of an epidemic at a low computational cost, these models may not be the best choice to understand how the infection spreads. Machine learning mod-

els are gaining popularity in recent years. A study (Ribeiro et al., 2020), that focussed on comparing the performance of a mathematical model based on ARIMA with a machine learning model based on support vector machines, found the machine learning counterpart to have higher accuracy (Masum et al., 2022). Extant studies have demonstrated the superiority of deep learning techniques such as recurrent neural networks in accurately predicting the spread of infection. However, they do not reveal the transmission mechanism of an infection (Alahmadi et al., 2020).

We may look at another set of methods that may be better in exploring the dynamics behind the spread of infection. Computational models that explore the spread of infection at a micro-level are increasingly used to study epidemic outbreaks (Duan et al., 2015). Such models like the metapopulation model provide a detailed representation of realities (Duan et al., 2015). This model has the advantage of describing the spread of infection spatially across regions. However, it assumes well-mixed, homogenous subpopulations and is limited in explaining the spread of infection. On the other hand, agent-based models are a promising and well-known bottom-up approach under computational models that model each individual or agent in a population and their interactions defined by some rules. They can incorporate the heterogeneity at the level of individuals and their interactions through micro-level analysis. These models delineate the stochastic nature of the spread of infection (Duan et al., 2015). However, they are much more complex to understand and implement. Under the computational approach, there exists another type known as network-based approaches, which can handle heterogeneity at the individual level and can model the spread of infection in a population (Duan et al., 2015; Kiss et al., 2017). Compared to mathematical models, a network-based approach can represent the heterogeneous environment in which an infection spreads by controlling the parameters of nodes and links. Using simulations, these models can explore how the infection spreads and how the network evolves over time. However, most of them consider unweighted networks, thereby losing sight of the interaction patterns (Duan et al., 2015). Therefore, the classification of extant studies based on a single dimension (i.e., method of analysis) is not straightforward. Some recent studies (Du et al., 2021; Hunter et al., 2020; Miranda et al., 2021) have combined and contrasted multiple methods by proposing hybrid models.

Epidemic models can also be further classified based on whether they have used commercial software (Aggarwal et al., 2020), custom-built simulation tools (Appadu et al., 2021), or standardized techniques (Alenezi et al., 2021) for implementation. However, irrespective of their choice of software, they generally lack a flexible model that can be used to model various scenarios by selecting different underlying contact network structures. In this regard, a hybrid simulation modelling approach (Brailsford et al., 2019) may be useful. The underlying epidemic model that is used in most of these studies divides the population into compartments. The standard compartmental model (Brauer et al., 2019; Capasso, 2008; Kermack & McKendrick, 1927; Martcheva, 2015; Treibert, 2021) i.e., Susceptible-Infected-Removed (SIR) assumes that individuals in the population under study can be categorized into one of the compartments S, I, or R. There are many variations to the standard compartmental model based on the number and description of compartments. However, all rely on the premise of dividing the population into compartments and studying the transitions between them. The models used to study epidemics can also be classified into deterministic models (Alenezi et al., 2021; Shapiro et al., 2021) and stochastic models (Yaesoubi & Cohen, 2011; Zhang et al., 2020). Deterministic models such as those based on ODEs are top-down and primarily focused on macro-level analysis. They are useful for predicting aggregate outcomes but do not provide insights about how the infection is transmitted from one individual to another. Stochastic approaches, such as the network-based approach, are bottom-up. They are less common but are better suited to realistically capture the transmission of an infectious disease (Zhang et al., 2020). The literature on epidemic models is vast (Lu & Borgonovo, 2023) with high-quality papers getting published on a variety of research problems associated with the Covid-19 pandemic (Farahani et al., 2023). Our study aims to address the inherent gaps in the literature by developing a network-based approach that effectively captures heterogeneity at the level of individuals and can be broadly applied across various situations. Table 1 provides a glimpse of existing methods, some key references, the key contributions, gaps in existing studies, and explains how the current research aims to fill those gaps.

3. Methodology

In this section we discuss a standard SIER compartmental model for studying the spread of epidemics using an equation-based approach. In our study, we select SEIR over SIR because there exists an incubation period for Covid-19. Our choice of SEIR over SIR is further strengthened by the results of a recent study (Alenezi et al., 2021) that showed SEIR is better suited than SIR to predict infections for Covid-19. This discussion is followed up by the design and implementation of the proposed network-based approach.

3.1. Studying the spread of the epidemic using the equation-based approach

Although there can be many variations of the equation-based approach (Basnarkov, 2021; Gwizdałła, 2020), we study a representation that retains the properties of an ODE and can be compared with the network-based approach. The equation-based approach is defined below:

On day t, $S(t)$, $E(t)$, $I(t)$, and $R(t)$ denote the number of people in susceptible, exposed, infected and recovered states respectively. Those in state R are infected earlier and are assumed to have either recovered or died. If N denotes the size of the population, then on any given day t,

$$N = S(t) + E(t) + I(t) + R(t) \tag{1}$$

The rate at which S transitions to E by coming in contact with infected individuals is denoted as β, the rate at which E transitions to I after spending an incubation period is denoted as δ, and the rate at which I transitions to R depending on the number of days an individual can spread the disease before either recovery or death is denoted as γ. The following relationships can be used to define δ and γ.

$$\delta = \frac{1}{incubation\ period} \tag{2}$$

$$\gamma = \frac{1}{number\ of\ days\ an\ infected\ person\ can\ spread\ the\ disease} \tag{3}$$

The basic reproductive number R_0 that denotes the total number of individuals an infected person infects can be defined as:

$$R_o = \beta/\gamma \tag{4}$$

The value of β, can be derived from R_0 and γ. The SEIR model is expressed by the following ODEs (Please refer to Appendix A for further details):

$$dS/dt = -\beta * I * S/N \tag{5}$$

$$dE/dt = \beta * I * S/N - \delta * E \tag{6}$$

Table 1
A glimpse of extant literature on epidemic models.

Methods of Analysis	Key Reference	Key Contribution	Gap	Contribution of Current Research
Forecasting	(Ding et al., 2021)	Provides long-term prediction and analysis of epidemic dynamics	Study is specific to South Africa	It models the spread of infection by considering variations over time and space for micro-level interactions. It incorporates heterogeneity at the level of network, infection, and individuals. It adds flexibility in the choice of input parameters, including the selection of the underlying network structure. It allows a comparison of the simulated outcome of model scenarios with empirical data for multiple countries using sophisticated measures in an uncertain situation like the beginning of an epidemic when limited information is available.
	(Appadu et al., 2021)	Conducts multi-country analysis using forecasting methods	Does not focus on micro-level interactions	
ODEs	(Gebremeskel et al., 2021)	Studies a compartmental epidemic model with sensitivity analysis	Study is specific to Ethiopia	
Agent-based	(Ajelli et al., 2010)	Compares agent-based and metapopulation stochastic model for a pandemic event in Italy	Difficulty in gathering datasets for most regions of the world	
		Proposes agent-based modelling using computational simulation of the pandemic in Australia	Does not explicitly model the underlying contact network	
Network-based	(Alrasheed et al., 2020)	Provides a contact network-based approach that captures realistic social dynamics	Model is specific to Saudi Arabia	
	(Renardy et al., 2020)	Proposes a network based on synthetic population and models of disease progression	The underlying contact network is static	

Table 2
Relationships defining the membership of a node to a compartment.

Compartment	Description
Susceptible, S	*No infection i.e.,* $VL_i = 0$
Exposed, E	*Not infectious i.e.,* $VL_i > 0$ *but* $VL_i <= IL_i$
Infected, I	*Infect others i.e.,* $VL_i > IL_i$ *and* $VL_i \leq VUL_i$
Removed, R	$VL_i > VUL_i$ *or after remaining infectious for* d_R *consecutive days*

Note: The value of VL_i, $IL_i$and VUL_i is normalized between 0 and 1.

Table 3
Differences between the equation-based approach and the network-based approach.

Characteristics	Equation-based	Network-based
Complexity	Simple	Complex
Availability of dynamic model	Common	Rare
Execution time per run* ($N = 10{,}000$)	Fast (Few seconds)	Slow (Few minutes to several hours)
Application	Homogenous population	Heterogenous population
Network structure	Not explicit	Can support any type
Individual level attributes	Not applicable	Can support any attribute
Nature of output	Deterministic	Stochastic
Number of inputs	Few	Many
Number of test-cases	Relatively smaller	Relatively larger

Note: * The execution time is dependant on the processing capability of the platform where the model is executed.

$$dI/dt = \delta * E - \gamma * I \tag{7}$$

$$dR/dt = \gamma * I \tag{8}$$

3.2. Simulation of the equation-based approach

To simulate the equation-based approach described in the previous section we choose a variety of parameters. The output is sensitive to the choice of input parameters. The equation-based approach accepts the size of the population N, the number of infected $I(0)$ and exposed $E(0)$ at the beginning, incubation period $1/\delta$, number of days an infected person can spread the disease $1/\gamma$ and the basic reproductive number R_0. The equation-based approach is top-down and assumes individuals in a population are homogenous and interactions are implicit (Edoh & Maccarthy, 2018). In general, ODEs are used to model the spread of epidemics in the equation-based approach, and the results are deterministic. The equation-based approach is simple to use but have its limitations. It is sensitive to the choice of parameters determining the probability of infection and the heterogeneity of the population. Except for few studies (Gwizdałła, 2020; Miranda et al., 2021), this approach fails to address any difference arising from the underlying contact structure. In this study we compare two different models that use separate values of R_0 under the equation-based approach and have the same initial conditions.

3.3. Studying the spread of the epidemic using the network-based approach

The spread of an epidemic in a networked environment is shown in Fig. 2. This figure illustrates a simple network with nine individuals represented by the nodes and connected by the links. On day 1, node 1 represents an infectious individual, and its two neighbouring nodes 2 and 3 that are exposed. The rest of the individuals in the network, denoted as nodes 4 to 9, who are not in direct contact with the infectious node, are susceptible to infections in future. On day 2, one of the exposed neighbours, i.e., node 2 becomes infectious and its neighbouring node 4 is now exposed. With the progress of time, the disease spreads through the network. An infected individual either dies or recovers and becomes disconnected from the network like node 1 on day 3.

The following paragraphs elaborate how the transmission of the epidemic is modelled realistically in the network-based approach. The process consists of describing the model assumptions, defining the model parameters, creating the contact network, configuring the parameters, and implementing the logic for updating the network. Although the network-based approach is similar to the homogenous compartmental models, it considers an individual to be different from others in their ability to withstand the virus. It is assumed that this variation arises from the difference in the level of immunity and exposure to the virus. The spread of infection in a connected network environment is described at the level of the network, individual nodes, and links.

The network level parameters are taken as inputs from the user at the beginning of the simulation. These parameters include the number of nodes (n), the type of network (T), days to gain recovery (d_R), load reduction factor (l_R), and the fraction of nodes that are considered infected (f). The parameters may include network characteristics such as degree distribution and the probability of link formation depending on the type of network. The attributes of an individual i, in the network are represented by node level

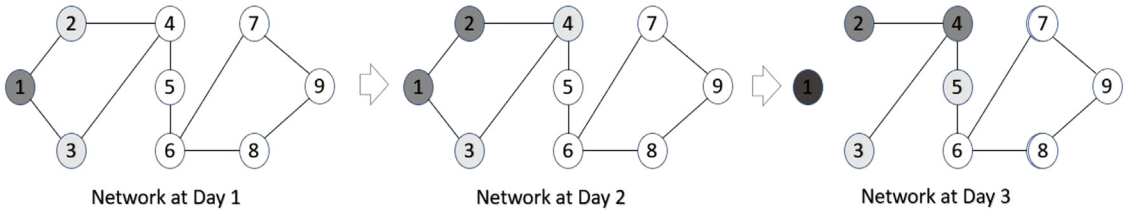

Fig. 2. Transmission of infections during an epidemic on a network.

parameters. These parameters correspond to viral load (VL_i), immunity level (IL_i), upper limit of viral load (VUL_i,) and days infected (d_i). A link represents the contact between two individuals *i* and j. The characteristics of the link is captured by the link level parameters. In this model, $weight_{i,j}$ is the only link level parameter. The value of $weight_{i,j}$ is set between 0 and 1 at the time of network configuration and is directly proportional to the probability of transmission of infection between the connecting nodes. The $weight_{i,j}$ is a normalized value based on the nature of contact between the connecting nodes. Appendix B provides further details about the assumptions of the network-based approach. The membership of an individual node to one of the four compartments is based on the relationships summarized in the following table.

It must be noted that in the proposed approach, an individual getting sick may get re-infected even after recovery depending on the value of node level attributes and their relationship.

3.4. Simulation of the network-based approach

In this paper, the network-based approach is simulated using synthetic contact networks. The proposed model is stochastic. The initial conditions set at the beginning of the simulation play an important role in determining the spread of the epidemic. The network-level parameters are set to values that are taken as inputs from the user. The link-level parameters are randomly assigned based on an algorithm following a uniform distribution within a pre-defined range. A combination of user input and algorithm-based assignment is used to set the initial values of the node level parameters. In case of the node level parameters, a minimum threshold value of immunity, min_{IL} and the upper limit of viral load min_{VUL}, are taken as user inputs. The individual values of VL_i, IL_i and VUL_i are randomly assigned to each node by the algorithm following a uniform distribution. In this study, we compare between two different models under the network-based approach by varying only the contact network structure and keeping the remaining input parameters unchanged. The simulation runs in a loop such that at each iteration, the logic to update the network is executed once. In the first stage, the algorithm creates the underlying contact network structure and configures its properties as specified by the user. The algorithm iterates over each node and link to set the attributes at this stage. The second stage starts by accepting the duration to simulate the spread of disease as an input from the user. The logic for updating the network (Please refer to Appendix C for details) in order to simulate the spread of the epidemic executes within a loop. The user can simulate the spread over successive periods to observe how the network evolves with time.

3.5. Difference between the two approaches

To conclude the discussion on methodology, we present a summary of the differences between the two approaches used in this study.

The comparison between the two approaches reveals that the choice of the approach depends on multiple factors. Both approaches can be used to model the spread of an epidemic. The equation-based approach works on a macro level and is simpler, faster, and easier to implement. On the other hand, the network-based approach works at an individual level and is preferable for studying a heterogeneous population. In modelling the spread of an epidemic, the nature of the contact network plays an important role, and this can be investigated using the network-based approach.

4. Numerical experimentation

4.1. Data consolidation and pre-processing

In the proposed network-based approach the network structure can be generated using synthetic generators or by accepting inputs to define any specific network structure from the user. There are several software packages available for network generation, analysis and visualization (Camacho et al., 2020). However, the choice of the programming tool based on suitable criteria (Fumagalli et al., 2019) is important to build the simulation model. In this study, the Python programming language is used to create, simulate, and analyse the network-based and equation-based approaches. To create and manipulate networks we used the NetworkX package (Hagberg et al., 2008). This package allows the creation of a network from scratch as well as by using synthetic generators. The synthetic generators are library functions defined under the package that accepts predefined inputs and returns network structure based on those inputs. The integrate library under scipy.integrate (*SciPy Documentation: Scipy.Integrate.Odeint*, 2020) is used for defining and solving equation-based approaches in Python. Similarly, other libraries in Python are used for feature selection and for calculating additional measures to compare the similarity between curves. The code used for modelling and analysis is written and executed on the Google Colaboratory cloud servers (Google, 2018, 2021). Appendix D provides details about the choice of software, hardware and synthetic generators that are used in implementing the simulations.

The initial value of the parameters used in the network-based approach can also be set using default parameters or can be taken as user input. Similarly, to run the simulation using the equation-based approach, the initial values of the parameters need to be set. The implementation of the network-based approach requires data for network generation and the values of various parameters. We searched open-access public datasets on Covid-19, and the dataset used by Appel et al., (2020) is found to be suitable for validation of the two approaches. This dataset shows the number of individuals infected by Covid-19 for various countries across the world. It also contains country-wise data on development index, demographics, health conditions, hospital facilities, etc. We have denoted the date on which the number of newly infected cases in a country reached 20 at the beginning of the spread of Covid-19. The choice of 8 different countries is made from different regions of the world. Please refer to Appendix E for details on the choice of these countries. The daily number of newly infected cases is considered for 60 days starting from the beginning of the epidemic for Australia (AUS), South Korea (KOR), Germany (DEU), Iran (IRN), Spain (ESP),

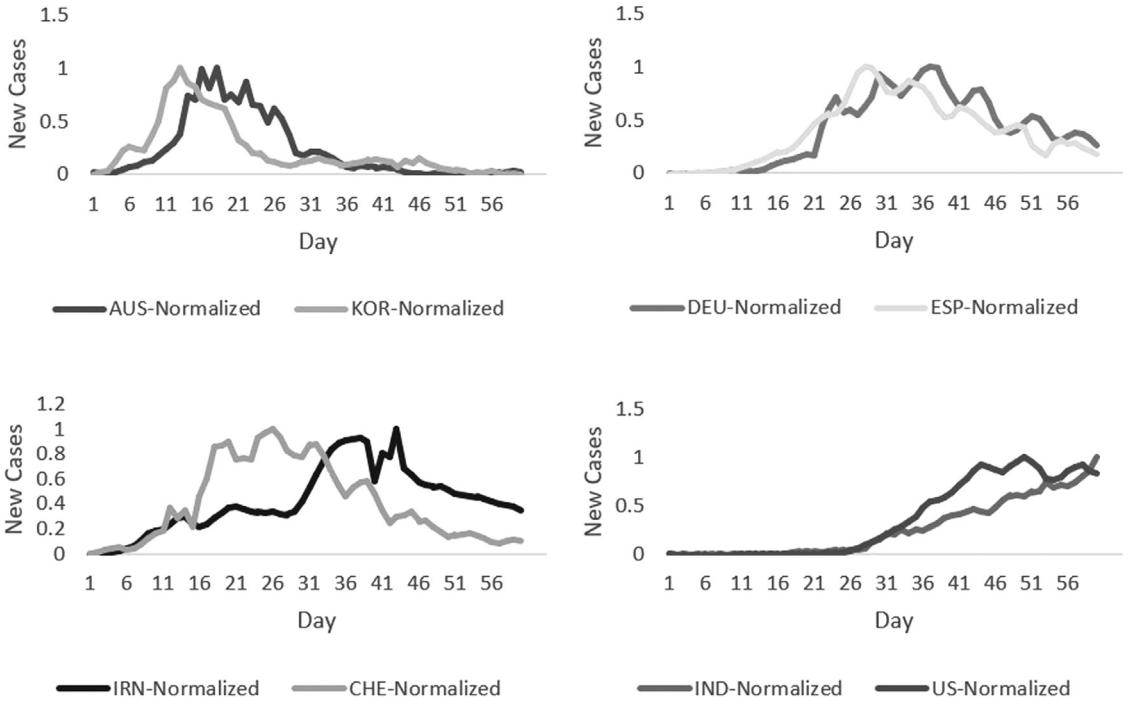

Fig. 3. Similarity in newly infected cases of Covid-19 for different countries.

Switzerland (CHE), India (IND) and United States (US). Although there exist studies (Wilinski & Szwarc, 2021) with a longer duration of analysis, we have chosen 60 days for this study as we are interested in predicting the spread of newly infected cases at the beginning of an epidemic. The choice is consistent with earlier studies (Appadu et al., 2021) on short, medium, and long-term predictions of an infectious disease. The choice of 60 days is also important as the contact network underlying the population may change beyond this time. We calculated a moving average of the data with an interval of 3 days to remove anomalies due to missing values and human errors. The data is then normalized within a range from 0 to 1 to compare the patterns in the outbreak trajectories. Based on similarity in patterns we divided the countries into four groups, as shown in Fig. 3.

4.2. Similarity between countries in terms of country level attributes

We calculated a dissimilarity index to compare between countries. Using extant literature (Atalan, 2020; Kadi & Khelfaoui, 2020; Zádori et al., 2020) we identified various attributes that are considered responsible for the spread of Covid-19. The relevant country-level attributes included size of population, density of population, median age of population, gross domestic product per capita, cardiovascular death rate, prevalence of diabetes, number of hospital beds per thousand people, life expectancy, human development index, and average stringency index. To decide about the attributes for calculating the dissimilarity index, we implemented a feature selection process. In this process, we computed the correlation between all the attributes and removed correlated ones based on a threshold correlation value. The process is repeated for different threshold values. At a threshold of 0.65 representing moderate correlation, we found that all attributes other than the size of the population and stringency index were dropped. The stringency index is a composite measure of a government's response at the country level to control the spread of Covid-19 with a value from 0 to 100 (Hale et al., 2021). Please refer to Appendix F for further details on the calculation of the dissimilarity index.

A lower value of this index depicts greater similarity between countries. According to the calculated value of this index, AUS is most similar to DEU and CHE, with values 0.06 and 0.08, respectively. However, their trajectories for newly infected cases of Covid-19 shown in Fig. 3 does not match. On a similar note, DEU is most similar to CHE and KOR, with the dissimilarity index value of 0.13 and 0.15 respectively, but their trajectories do not match. Thus, the country-level attributes such as the size of the population and stringency index can be used to group countries together although the pattern of the spread of newly infected cases remains dissimilar.

4.3. Data for simulations of the equation-based approach

The input parameters in the case of the equation- and the network-based approaches are not the same. This poses a challenge in selecting the input parameters for the simulation. We selected the same values during the simulation of both approaches for the common input parameters. In contrast, the values of the remaining parameters are selected based on earlier studies. Previous studies (Burda, 2020; Alenezi et al., 2021) have shown that the value of R_0 plays an interesting role in the spread of an epidemic. In this study, we varied the value of R_0 as mentioned in Section 3.2 to generate two scenarios under the equation-based approach and observed the outcomes. These two scenarios corresponded to the value of $R_0 = 15$ and 2.5 respectively. The value of R_0 is varied because earlier studies (Burda, 2020; Renardy et al., 2020) on Covid-19 have used values around 2.5. However, Covid-19 has some strains that are highly infectious and so a value of 15 is more appropriate to model the spread of highly infectious disease like measles (S. L. Chang et al., 2020). Since the equation-based approach is deterministic, the simulation is executed once for each scenario with the choice of initial parameters as shown in Table 4. To begin the simulation, we entered the population size N of 10,000, the number of initial infections $I(0)$ as 20 and number of exposed initially, $E(0)$ as 400. The value of the remaining parameters for the equation-based approach is selected as shown in Table 4.

Table 4
The parameters used for simulation of the equation-based approach.

Parameter	Value	References
Incubation period, $1/\delta$	5	(Burda, 2020; Renardy et al., 2020)
Number of days an infected person can spread the disease,$1/\gamma$	10	(Burda, 2020)
Basic reproductive number, R_0	15, 2.5	(Chang, Piraveenan, et al., 2020; Burda, 2020),

Table 5
Choice of initial parameters of the two models using the equation-based approach.

Highly Contagious Model A1	Moderately Contagious Model A2
$N = S(t) + E(t) + I(t) + R(t) = 10{,}000$	$N = S(t) + E(t) + I(t) + R(t) = 10{,}000$
$\delta = 1/5$	$\delta = 1/5$
$1/\gamma = 10$	$1/\gamma = 10$
$R_0 = 15$	$R_0 = 2.5$
$\beta = R_0 * \gamma = 1.5$	$\beta = R_0 * \gamma = 0.25$
$S0, E0, I0, R0 = N\text{-}420, 400, 20, 0$	$S0, E0, I0, R0 = N\text{-}420, 400, 20, 0$

Table 5 shows a comparison between the two model scenarios. In summary, model scenarios A1 and A2 are equation-based approaches that have the same choice of initial parameters except the value of the parameter of interest R_0.

4.4. Data for simulations of the network-based approach

To simulate the spread of Covid-19 at the country level, we assumed there are multiple clusters of a population where the disease spreads. In this study, we consider 10 such clusters with a size of 1000 where the infection begins. Earlier studies have used similar initial sizes of clusters (Basnarkov, 2021; Kim et al., 2021). The total size of the population is 10,000. It is kept the same at the beginning of simulations for each scenario under the equation- and network-based approaches to prevent the size of the network (Gwizdałła, 2020) from affecting the result. It is also assumed that the number of infected individuals in each cluster at the beginning of the simulation is 2. It corresponded to our choice of 20 newly infected cases as the beginning of the spread of Covid-19. In addition, the clusters are assumed to be in different locations of a country, and separated from each other. Thus, each cluster is represented by a separate network, and it is assumed that the contact networks across the different clusters are of the same type. The simulation is repeated 10 times with the same initial conditions to mimic the spread of the disease. The nature of the underlying contact network in a country is not known and so we simulated two distinct types of contact network structure. It considered them as two different scenarios under the network-based approach. The choice of contact network types used in the simulation is consistent with earlier studies (Dong et al., 2019; Edoh & Maccarthy, 2018; Gwizdałła, 2020; Jorritsma et al., 2020; Kim et al., 2021). The choice of the values of input parameters helped to create two different types of networks and the small-world network-based approach is labelled as B1 and the preferential attachment network-based approach is labelled as B2. Table 6 shows a comparison between the two models based on different underlying contact network structure. Please refer to Appendix G for details on the type of network structures used.

In a nutshell, model scenarios B1 and B2 are network-based approaches that differ in the underlying network structure. After creating the synthetic networks for the two scenarios, the values of the parameters that are fixed at the beginning of the simulation are shown in Table 7.

The value of f is set to 0.002 so that at the beginning of the simulation any 2 nodes in the network of 1000 nodes are infectious. The choice of this value is based on the fact that in a networked environment, it is better to presume that the infection starts spreading when the number of infected is more than one. A larger value is not considered as we studied the spread of the epidemic in its early stage. The choice of the value of l_R, and d_R are based on guidance (NCIRD, 2021) indicating that the time to recovery from Covid-19 even for critical adult patients is within 20 days. The minimum threshold value of immunity is kept at 0.5. The minimum threshold of the upper limit of viral load is kept above the minimum immunity threshold at 0.7. During each run of the simulation, the viral load of nodes is updated 60 times to mimic the growth of the Covid-19 infection over 60 days.

5. Results

In this section we report the impact of varying the parameters of interest on the number of infections. First, we plot the outcome of the equation- and network-based approaches with the given choice of parameters. We demonstrate the effect of varying the parameter R_0 for the equation-based approach and show the results as the highly contagious equation-based approach A1 and the moderately contagious equation-based approach A2 respectively. Next, we present the results of the two different scenarios obtained by varying the type of the underlying network in the network-based approach and denote them as the small-world network-based approach B1 and the preferential attachment network-based approach B2 respectively. In the subsequent sections, we compare the results of these models with the empirical data using different measures in order to determine the models with the best fit.

5.1. Outcome of the approaches

5.1.1. Equation-based approach

Fig. 4 shows the variation of infections for highly contagious equation-based approach A1 and the moderately contagious equation-based approach A2, by varying the value of R_0 and keeping all other parameters constant. In the plots, the x-axis denotes the day in progression of infection and the y-axis denotes the number of newly infected individuals on that day normalized between 0 and 1.

In Fig. 4, models A1 and A2 are obtained using the values of $R_0 = 15$, and 2.5 respectively. It is observed that with a decrease in the value of R_0 the curve shifts to the right and the epidemic continues beyond the time period of 60 days. On the other hand, for a higher value of R_0 the epidemic reaches its peak and dies down faster. Although, it may seem counter-intuitive that reducing the transmission rate is key to controlling the disease based on sensitivity analysis performed on the value of R_0 in an earlier study (Gebremeskel et al., 2021), our result points to an interesting aspect of epidemic models. While $R_0 = 2.5$ is in the range of values used in earlier studies (Burda, 2020; Renardy et al., 2020) on Covid-19, $R_0 = 15$ corresponds to the spread of highly infectious disease like measles (S. L. Chang et al., 2020). It is interesting to note that given a population of 10,000, the equation-based approach reaches a saturation point in which majority of the population gets infected and is subsequently removed earlier due to a higher value of R_0.

5.1.2. Network-based approach

The network-based approach models the spread of epidemic realistically by incorporating the characteristics of the entire population, disease at the level of an individual, and underlying contact network structure through user-defined input parameters. To demonstrate how this approach can be applied to various situations, we vary the underlying contact network structures

Table 6
Properties of the two models using the network-based approach.

Model	Type of Network	Parameter	Value	Links	Average Degree
Small-world network-based approach B1	Newman–Watts–Strogatz small-world graph: NWSG	Size	1000	10,000	20
		K	10		
		Probability of link creation	1		
Preferential attachment network-based approach B2	Barabási–Albert preferential attachment model.: BAG	Size	1000	9900	19.8
		Number of links to preferentially attach	10		

Note: The value of K is such that each node is joined with its K nearest neighbours forming a ring topology.

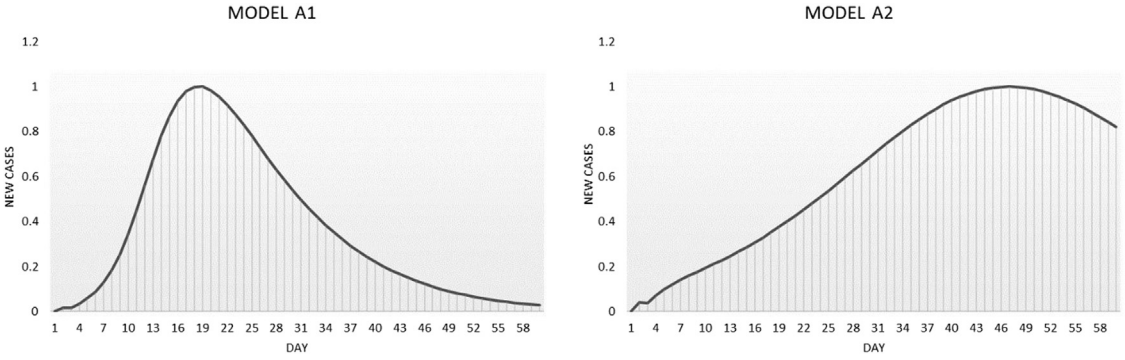

Fig. 4. Simulation results for models A1 and A2 using the equation-based approach.

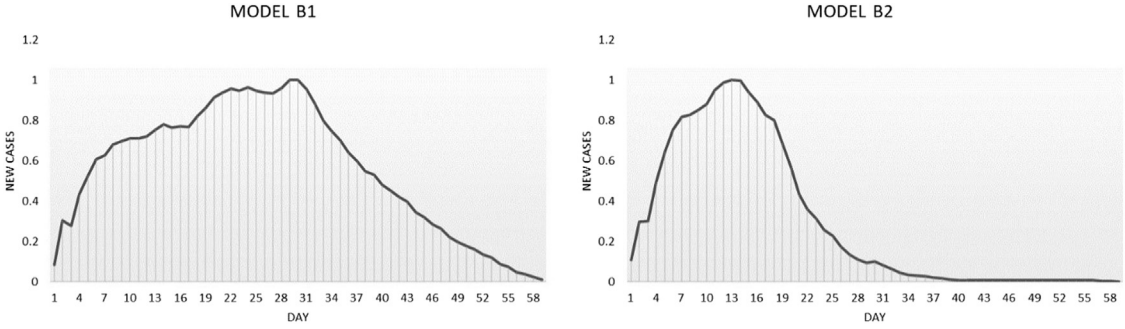

Fig. 5. Simulation results based for models B1 and B2 using the network-based approach.

Table 7
Values of various parameters.

Parameter	Value
Fraction infected, f	0.002
Load reduction, l_R	0.05
Number of days to gain recovery, d_R	20
Minimum threshold of immunity, min_{IL}	0.5
Minimum threshold of upper limit of viral load, min_{VUL}	0.7
Duration of simulation, t (in days)	60

Note: The simulation is repeated 10 times for given initial conditions.

keeping all other input parameters constant. We plot the results of the simulation using the small-world network-based approach B1 and the preferential attachment network-based approach B2, as shown in Fig. 5. The plots show how the spread of the epidemic varies with the underlying contact network structure.

In order to investigate the sensitivity of the output of the proposed model with respect to other input parameters, we vary the value of min_{VUL} between 0.5 to 0.9 for each of the above scenarios. In the preferential-attachment based model B2, we observe no significant difference. However, some changes are observed for the small-world network-based approach B1.

5.2. Comparison between the trajectory of the outbreak and actual infections

In order to compare between the outputs corresponding to the approaches and identify to what extent the outputs match the data corresponding to the eight different countries, a number of similarity measures are used. These include comparing the peak, calculating the error based on sum-of-squares based difference, and correlation between the slope of the curves. Furthermore, sophisticated measures like partial curve mapping, dynamic time warping, and curve length approach are used.

5.2.1. Comparison based on peak

To compare the simulation results with the empirical data, we first compare the day when the peak is reached. The comparison is made by identifying the day on which the maximum normalized value of the newly infected case is registered over the period of 60 days for each country. Please refer to Appendix H for details on how the actual number of cases are normalized and the peak is calculated for each country. It is observed that model B2 exactly matches KOR in terms of the day on which the peak is reached. We also find that the highly contagious model A1 matches AUS with a difference of 1 day and the small-world network-based

Table 8
Calculation of the sum-of-square based difference corresponding to AUS.

Day	1	2	3	4	5	6	7	8	9	10	11	12	13	14
A1	0.00	0.02	0.02	0.04	0.06	0.09	0.13	0.18	0.26	0.35	0.45	0.56	0.68	0.78
A2	0.00	0.04	0.04	0.07	0.10	0.12	0.14	0.16	0.18	0.19	0.21	0.23	0.25	0.27
B1	0.09	0.30	0.28	0.43	0.53	0.61	0.63	0.68	0.70	0.71	0.71	0.72	0.75	0.78
B2	0.11	0.30	0.30	0.49	0.64	0.76	0.82	0.83	0.85	0.88	0.95	0.99	1.00	1.00
AUS	0.02	0.01	0.01	0.02	0.04	0.07	0.09	0.11	0.13	0.18	0.23	0.30	0.38	0.74
(A1-AUS) ^2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.02	0.03	0.05	0.07	0.09	0.00
(A2-AUS) ^2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.22
(B1-AUS) ^2	0.00	0.09	0.07	0.17	0.23	0.29	0.29	0.32	0.32	0.28	0.23	0.18	0.14	0.00
(B2-AUS) ^2	0.01	0.08	0.08	0.22	0.36	0.47	0.53	0.51	0.51	0.49	0.52	0.48	0.38	0.07

Note: All values are rounded to two decimal places.

approach B1 matches ESP and CHE with a difference of 1 and 3 days, respectively. In terms of predicting when the peak is reached, the network-based approaches (model B1 and model B2) yield better results for the given dataset. One limitation of matching the peak is that while the peak of the trajectory of newly infected cases for a scenario and country may match, the pattern of their growth and decline may be different.

5.2.2. Comparison based on sum-of-square based difference

To further compare between trajectories, we calculate the sum-of-squares-based difference between the normalized values corresponding to data points for 60 days for each of the 32 pairs, i.e., a combination of the 4 model scenarios and 8 countries. Table 8 provides a snapshot of the calculation of the difference between the normalized values corresponding to the 4 scenarios and AUS for the first 14 days. The detailed table with calculation for all countries is shown in Appendix I.

In Table 8, the value 0.00 for the cell corresponding to the row (A1-AUS) ^2 and column 1 denotes the square of the difference between the normalized values of model A1 and AUS on day 1. Although not shown in Table 8, this calculation is repeated for AUS for 60 days. Similarly, it is repeated for the other 7 countries for 60 days as well. Finally, the sum of the differences for each scenario-country pair is taken, and a similarity score is calculated to identify the best fit. The sum-of-square of difference is divided by the maximum possible sum, i.e., 60, as the difference at a particular cell can take a maximum value of 1 to calculate the score. The result which is a number between 0 and 1 is a measure of the difference. It is subtracted from 1 and multiplied by 100 to obtain a percentage value to convert the result to a similarity score.

$$Similarity\ Score = \{1 - (Sum\ of\ square\ based\ difference/60)\} * 100$$

In this way we obtain a similarity score for each model-country pair and identify the model which fits a country best. Table 9 provides a snapshot of the scores for the 8 countries corresponding to each of the models.

From Table 9 we find that the highly contagious equation-based approach A1 matches AUS and model B2 matches KOR the best with a score of 97.13 and 96.36, respectively. This measure using sum-of-square of difference, despite being simple to use, has its limitations. It is not reliable as it can provide a high score even when the model's output does not match the empirical data visually, as in the case of IND and US. For example, if we calculate the score for a model that predicts all newly infected cases from day 1 to day 60 as zero, the similarity score obtained is 85.85 for AUS and 88.23 for KOR. Similarly, the trajectory of model A2 does not match IND and US when inspected visually, but has a high similarity score of 84.36 and 90.73, respectively.

5.2.3. Comparison based on correlation between slopes

To quantify the similarity between the plots we examine the correlation between the corresponding slope of each curve. The correlations between the slope of the curves of selected countries and models are presented in Table 10.

It is observed from Table 10 that the value of correlation for the country-model pairs is low and may not be suitable to identify the best fit. If we take a cut-off of 0.5, only the (KOR, model B2) pair is above the cut-off.

Table 9
Similarity scores using the sum-of-square based difference.

Model-Country	Sum of Square of Differences	Normalized Difference	Similarity Score
A1 - AUS	1.72	0.03	97.13
A2 - AUS	25.07	0.42	58.21
B1 - AUS	9.31	0.16	84.48
B2 - AUS	6.07	0.10	89.88
A1 - KOR	5.24	0.09	91.27
A2 - KOR	25.26	0.42	57.91
B1 - KOR	11.64	0.19	80.59
B2 - KOR	2.18	0.04	96.36
A1 - DEU	13.14	0.22	78.11
A2 - DEU	5.68	0.09	90.53
B1 - DEU	11.74	0.20	80.43
B2 - DEU	24.60	0.41	59.00
A1 - IRN	11.95	0.20	80.09
A2 - IRN	4.41	0.07	92.64
B1 - IRN	11.86	0.20	80.23
B2 - IRN	19.65	0.33	67.24
A1 - ESP	8.60	0.14	85.67
A2 - ESP	8.41	0.14	85.98
B1 - ESP	7.99	0.13	86.68
B2 - ESP	20.54	0.34	65.77
A1 - CHE	2.91	0.05	95.14
A2 - CHE	14.30	0.24	76.17
B1 - CHE	3.84	0.06	93.60
B2 - CHE	15.12	0.25	74.81
A1 - IND	19.10	0.32	68.16
A2 - IND	9.39	0.16	84.36
B1 - IND	24.55	0.41	59.08
B2 - IND	21.09	0.35	64.85
A1 - US	25.45	0.42	57.58
A2 - US	5.56	0.09	90.73
B1 - US	28.57	0.48	52.39
B2 - US	29.18	0.49	51.36

Note: All values are rounded to two decimal places.

Table 10
Correlation between the slope of curves for country-model pairs.

Country	Model A1	Model A2	Model B1	Model B2
AUS	0.43	−0.08	0.08	0.15
KOR	0.38	0.00	0.12	0.58
DEU	−0.06	0.25	0.16	−0.14
IRN	0.02	0.22	0.07	0.00
ESP	0.05	0.23	0.19	−0.24
CHE	0.29	0.06	0.21	0.01
IND	−0.21	−0.33	−0.19	−0.07
US	−0.28	0.26	−0.28	−0.05

Table 11
Calculated values of the additional measures of similarity.

Measure	Model/ Country	AUS	KOR	DEU	IRN	ESP	CHE	IND	US
PCM	A1	4.15	5.35	7.44	9.73	5.25	3.82	23.22	23.23
	A2	27.37	28.93	8.02	7.46	10.88	16.32	6.63	3.20
	B1	6.25	8.35	5.71	9.20	4.12	3.04	20.39	18.49
	B2	4.37	3.01	11.03	12.47	8.70	8.10	31.33	32.56
DF	A1	0.38	0.67	0.87	0.83	0.77	0.66	0.98	1.00
	A2	0.99	0.95	0.62	0.50	0.79	0.85	0.61	0.54
	B1	0.82	0.90	0.76	0.66	0.68	0.61	1.00	0.95
	B2	0.73	0.59	0.98	0.99	0.91	0.83	1.00	1.00
AREA	A1	8.12	12.52	23.82	23.39	19.50	10.54	27.84	33.65
	A2	31.62	33.82	14.20	11.77	16.82	23.06	21.05	14.93
	B1	19.79	20.70	23.30	24.90	17.40	10.15	34.07	38.02
	B2	12.55	7.53	35.40	30.72	31.58	26.01	30.56	37.13
CL	A1	2.15	2.86	3.97	5.04	2.80	2.09	7.18	7.07
	A2	4.97	5.15	3.48	2.64	4.03	3.81	3.13	2.33
	B1	2.47	2.91	2.79	3.88	2.31	1.60	5.90	5.56
	B2	3.13	1.95	6.61	7.96	5.11	4.84	9.78	9.97
DTW	A1	8.13	12.53	23.93	23.55	19.58	10.58	28.33	34.05
	A2	32.02	34.24	14.49	12.02	17.16	23.44	21.14	14.94
	B1	19.84	20.75	23.47	25.11	17.53	10.24	34.61	38.48
	B2	12.61	7.71	35.58	30.94	31.72	26.11	31.11	37.60

5.2.4. Comparison based on additional measures

The measures discussed so far have their limitations and this leads us to search for additional similarity measures between curves. Jekel et al. (2019) have identified five such measures, PCM (partial curve mapping), area method, DF (discrete Fréchet) distance, CL (curve length), and DTW (dynamic time warping). These measures use a combination of distance, area, and arc length to measure the similarity between curves and can be used to identify the best fit.

The PCM method calculates the similarity based on arc length and the area between the shorter and longer curve. The area method, on the other hand, finds the mismatch between curves based on the area determined by constructing quadrilaterals between the curves. The DF method is another measure of similarity based on a walking dog analogy. The CL method calculates deviations between the corresponding values of points on both curves that are compared. Similarly, the DTW method calculates the distance between each point of both the curves that are compared. However, it determines the optimal path with the smallest cumulative distance to measure the similarity between curves. Compared to the similarity measures based on peak, sum-of-square-based distance, and correlation between slopes discussed in the previous section, the additional measures based on features of a curve are likely to be more reliable. Appendix J provides further details about the additional measures. In each of these similarity measures, a smaller value is considered better. We calculate the value for each of the 5 measures for the 32 model-country pairs, as shown in Table 11.

Table 11 tells us which scenario fits the country best for a given measure. It is clear from the values that the highly contagious equation-based model A1 fits AUS the best when PCM is considered. The best fit between the model and the country can be identified for each measure. It is found during the analysis that not all values of the measures are acceptable for the comparison. For example, in case of IND, although a simple visual inspection reveals that none of the models match reality, Table 11 shows moderately contagious model A2 as the best fit when DF is used as the measure.

5.2.5. Best fit approach based on the measures

Table 12 summarizes the best fit between the trajectory of newly infected cases corresponding to the 4 models and the 8 countries studied in this paper. The curves are compared in terms of the similarity measures that are reliable, as discussed in the previous sections. The best fit is considered if the similarity measure meets the cut-off.

Table 12
Comparison between the scenarios to determine best fit for each measure.

Measure (cut-off)	Equation-based Approach		Network-based Approach	
Model	A1	A2	B1	B2
Peak (<= 3 days)	AUS	–	ESP, CHE	KOR
Correlation (> 0.5)	–	–	–	KOR
PCM (<= 5)	AUS	US	ESP, CHE	KOR
DF (<=0.6)	AUS	IRN, US	–	KOR
Area (<= 10)	AUS	–	–	KOR
CL (<= 2.5)	AUS	US	ESP, CHE	KOR
DTW (<= 10)	AUS	–	–	KOR

It is observed that in terms of the similarity measures with given cut-offs, the models considered under the network-based approach are better in predicting the pattern of newly infected cases in KOR, ESP, and CHE. On the other hand, the models considered under the equation-based approach are better in predicting the infection for AUS, US, and IRN. A closer inspection reveals that the network-based approach is more consistent in its performance as model B2 matches KOR across all the measures. Similarly, the small-world network-based approach B1 matches ESP and CHE for measures using Peak, PCM, and CL, whereas models A1 and A2 under the equation-based approach are less consistent in matching countries. It shows that underlying contact network structure may play an important role in the mechanism behind the spread of infection. The findings are encouraging from an academic and practical point of view since it suggests that well-informed decisions can be made in future crises by engaging appropriate modelling approach depending on the specific context.

6. Implications and limitations of research

6.1. Academic implications

The study in this paper has several implications. As for academic implications, this study distinguishes itself by implementing and comparing two contrasting epidemic approaches, using the

top-down equation-based approach and the bottom-up network-based approach. Most of the earlier studies have focused on either of the two approaches. However, we believe in this study a comparison between the two approaches better explains why the network-based approach is needed. Second, while some earlier researchers have used multi-country data, they have used top-down approaches for analysis. This study demonstrates the application of the network-based approach and compares the results with the empirical data for 8 countries from different regions of the world. Third, in this study we introduce the comparison of the trajectory of curves based on eight different measures. 'While literature pertaining to material models have used the measures, PCM, DF, Area, CL, and DTW for comparing between curves, they are new to the literature on epidemic modelling. Finally, we introduce heterogeneity in the choice of networks, infection parameters, and individual-level attributes, which are unique and have not been addressed in earlier studies, to the best of our knowledge.

6.2. Practical implications

The findings of this paper have practical implications as well. First, the proposed network-based approach can help policymakers and health service providers identify the disease's spread at its initial stage and plan accordingly. Second, being a bottom-up individual-based approach, it can be customized to identify individuals who are most vulnerable to contracting the disease and can be used for the early detection of clusters of severe cases. It can benefit authorities who can customize this approach to impose localised restrictions. They can also use this approach to prioritize who would be vaccinated first, not just based on age but also based on the overall vulnerability of an individual to an infectious disease. Finally, although this study has utilized the two most relevant types of networks as underlying contact network structures, the proposed network-based simulation can also generate a wide variety of network structures. We recommend that the practitioner should be aware of the variation in underlying contact network structure while considering the different model scenarios. Although, we do not have sufficient information to suggest whether a particular country has the same type of network structure as another, the close match between the model scenarios B2 and KOR is particularly encouraging. It is interesting to note that an earlier study investigating the underlying contact network structure in a population using sample from KOR had shown it to be indeed scale-free (Kim et al., 2021). In future, if data of a more realistic network structure is available, practitioners can use this approach to create a network based on its properties to study the spread of infections during an epidemic.

6.3. Limitations of proposed approach

The network-based approach proposed in this study also has a set of limitations. First, scaling up the network-based approach is expensive in terms of time and space. This study has executed the proposed approach using limited processing capacity by dividing the population into 10 clusters and simulating these clusters separately. A state-of-the-art distributed simulation approach (S. J. Taylor, 2019) can be used to address this limitation. Second, it uses a network that has a semi-static nature. As a result, the analysis period of this study is fixed at 60 days, within which the network allows only deletion of nodes and links. Future research can include the option to add new individuals to the contact network due to birth and emigration by adding nodes and links to the given network at specific time intervals. Third, in this study we do not consider vaccinated individuals since the study is focused on the early period of Covid-19 when the possibility of vaccinated individuals did not arise. However, the proposed network-based approach can be modified to incorporate immunity acquired through vaccination by manipulating the value of the node level attribute IL_i, that determines the immunity of a node.

Since neither the equation- or the network-based approaches are found to be superior for modelling all possible scenarios, another direction for future researchers could be to create a hybrid approach that switches between the equation- and network-based approaches. The switching can be based on a specific country or an available scenario to capture the advantages of both approaches.

7. Conclusion

A global catastrophe like Covid-19 necessitated the study of epidemics using operational research techniques. Researchers and policymakers are interested in knowing the mechanism behind the transmission of an epidemic and investigating its relationship with patterns of social interaction. It is important to apply the network-based approach to anticipate the trajectory at the beginning of the outbreak for a country. The network-based approach allows modelling the heterogeneity amongst individuals and interaction patterns not addressed by the equation-based approach. Unlike the traditional well-mixed compartmental models, the network-based approach does not assume the population to be homogenous even within a compartment. In this paper, the individuals are not only distinguished by their level of infection denoted by the concept of viral load but also by their level of immunity and the upper limit of the viral load that they can withstand. A complex yet realistic relationship that is conceptualized while developing the proposed approach helps to determine whether an individual infects others, recovers, or is removed from the network. In future, the network-based approach can be used to model policy interventions, such as lockdown, social distancing, and vaccination, by manipulating the node and link-level parameters of the network.

Operational research is called the 'science of better' (Mingers, 2007; Nikolopoulos, 2021). Consequently, researchers working on epidemic models use operational research techniques to search for a better model. However, a single model may not be relevant in all contexts. Some papers have reported accurate forecasts employing several time-series, epidemiological, machine learning, and deep learning methods (Nikolopoulos et al., 2021; Petropoulos & Makridakis, 2020). A closer inspection reveals that models based on macro-level analysis are not directly comparable and may be complementary to methods based on micro-level analysis. Macro-level analyses are simpler, faster, and suitable for specific practical use like what-if scenarios.

In contrast, micro-level analyses are far more complex and slower to execute but more relevant in accommodating individual differences. Researchers find methods with accurate predictions useful. Some may additionally need a flexible model to study an epidemic's emerging behaviour, which can be applied to various situations and yet can capture micro-level details realistically. This paper aims not to perform an exhaustive numerical comparison of these methods but to provide a viable alternative of a flexible and realistic approach to model epidemics, that the researchers may prefer under some situations. It highlights that researchers may prefer one approach over another depending on the context. Operational research, with its bouquet of techniques, can lead the way in epidemic modelling, where different approaches can co-exist and can be a part of a decision support toolkit for epidemic modelling.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.ejor.2023.01.021.

References

Adly, A. S., Adly, A. S., & Adly, M. S. (2020). Approaches based on artificial intelligence and the internet of intelligent things to prevent the spread of COVID-19: Scoping review. *Journal of Medical Internet Research, 22*(8), 1–15. https://doi.org/10.2196/19104.

Aggarwal, L., Goswami, P., & Sachdeva, S. (2020). Multi-criterion intelligent decision support system for COVID-19. *Applied Soft Computing, 101*, Article 107056. https://doi.org/10.1016/j.asoc.2020.107056.

Ajelli, M., Gonçalves, B., Balcan, D., Colizza, V., Hu, H., Ramasco, J. J., et al., (2010). Comparing large-scale computational approaches to epidemic modeling: Agent-based versus structured metapopulation models. *BMC Infectious Diseases, 10*(1), 1–13. https://doi.org/10.1186/1471-2334-10-190.

Alahmadi, A., Belet, S., Black, A., Cromer, D., Flegg, J. A., House, T., et al., (2020). Influencing public health policy with data-informed mathematical models of infectious diseases: Recent developments and new challenges. *Epidemics, 32*, Article 100393.

Alenezi, M. N., Al-Anzi, F. S., & Alabdulrazzaq, H. (2021). Building a sensible SIR estimation model for COVID-19 outspread in Kuwait. *Alexandria Engineering Journal, 60*(3), 3161–3175. https://doi.org/10.1016/j.aej.2021.01.025.

Alrasheed, H., Althnian, A., Kurdi, H., Al-Mgren, H., & Alharbi, S. (2020). COVID-19 spread in Saudi Arabia: Modeling, simulation and analysis. *International Journal of Environmental Research and Public Health, 17*(21), 1–24. https://doi.org/10.3390/ijerph17217744.

Appadu, A. R., Kelil, A. S., & Tijani, Y. O. (2021). Comparison of some forecasting methods for COVID-19. *Alexandria Engineering Journal, 60*(1), 1565–1589. https://doi.org/10.1016/j.aej.2020.11.011.

Appel, C., Beltekian, D., Gavrilov, D., Giattino, C., Hasell, J., Macdonald, B. et al. (2020). *Data on COVID-19 (coronavirus) by Our World in Data*. https://github.com/owid/covid-19-data/tree/master/public/data

Atalan, A. (2020). Is the lockdown important to prevent the COVID-19 pandemic? Effects on psychology, environment and economy-perspective. *Annals of Medicine and Surgery, 56*, 38–42.

Basnarkov, L. (2021). SEAIR Epidemic spreading model of COVID-19. *Chaos, Solitons & Fractals, 142*, Article 110394. https://doi.org/10.1016/j.chaos.2020.110394.

Bozzani, F. M., Vassall, A., & Gomez, G. B. (2021). Building resource constraints and feasibility considerations in mathematical models for infectious disease: A systematic literature review. *Epidemics, 35*, Article 100450. https://doi.org/10.1016/j.epidem.2021.100450.

Brailsford, S. C., Eldabi, T., Kunc, M., Mustafee, N., & Osorio, A. F. (2019). Hybrid simulation modelling in operational research: A state-of-the-art review. *European Journal of Operational Research, 278*(3), 721–737. https://doi.org/10.1016/j.ejor.2018.10.025.

Brauer, F., Castillo-Chavez, C., & Feng, Z. (2019). *Mathematical models in epidemiology*: Vol. 32. Springer.

Burda, Z. (2020). Modelling excess mortality in covid-19-like epidemics. *Entropy, 22*(11), 1–21. https://doi.org/10.3390/e22111236.

Camacho, D., Panizo-LLedot, Á., Bello-Orgaz, G., Gonzalez-Pardo, A., & Cambria, E. (2020). The four dimensions of social network analysis: An overview of research methods, applications, and software tools. *Information Fusion, 63*, 88–120. https://doi.org/10.1016/j.inffus.2020.05.009.

Capasso, V. (2008). *Mathematical structures of epidemic systems*: Vol. 97. Springer Science & Business Media.

Chang, J. T., & Kaplan, E. H. (2023). Modeling local coronavirus outbreaks. *European Journal of Operational Research, 304*(1), 57–68. https://doi.org/10.1016/j.ejor.2021.07.049.

Chang, S. L., Piraveenan, M., & Prokopenko, M. (2020). Impact of network assortativity on epidemic and vaccination behaviour. *Chaos, Solitons & Fractals, 140*, Article 110143. https://doi.org/10.1016/j.chaos.2020.110143.

Choi, T.-. M. (2021). Fighting against COVID-19: What operations research can help and the sense-and-respond framework. *Annals of Operations Research*, 1–17. https://doi.org/10.1007/s10479-021-03973-w.

Ding, W., Wang, Q.-. G., & Zhang, J.-. X. (2021). Analysis and prediction of COVID-19 epidemic in South Africa. *ISA Transactions, 124*, 182–190. https://doi.org/10.1016/j.isatra.2021.01.050.

Dong, X., Liu, Y., Wu, C., & Lian, Y. (2019). The topology of scale-free networks with an S-shaped nonlinear growth characteristic. *Chaos Solitons & Fractals, 121*, 137–148. https://doi.org/10.1016/j.chaos.2019.02.007.

Du, M., Sai, A., & Kong, N. (2021). A data-driven optimization approach for multi-period resource allocation in cholera outbreak control. *European Journal of Operational Research, 291*(3), 1106–1116. https://doi.org/10.1016/j.ejor.2020.09.052.

Duan, W., Fan, Z., Zhang, P., Guo, G., & Qiu, X. (2015). Mathematical and computational approaches to epidemic modeling: A comprehensive review. *Frontiers of Computer Science, 9*(5), 806–826.

Edoh, K., & Maccarthy, E. (2018). Network and equation-based models in epidemiology. *International Journal of Biomathematics, 11*(3), 1–16. https://doi.org/10.1142/S1793524518500468.

Farahani, R. Z., Ruiz, R., & Van Wassenhove, L. N. (2023). Introduction to the special issue on the role of operational research in future epidemics/pandemics. *European Journal of Operational Research, 304*(1), 1–8. https://doi.org/10.1016/j.ejor.2022.07.019.

Fumagalli, L., Polenghi, A., Negri, E., & Roda, I. (2019). Framework for simulation software selection. *Journal of Simulation, 13*(4), 286–303. https://doi.org/10.1080/17477778.2019.1598782.

Gebremeskel, A. A., Berhe, H. W., & Atsbaha, H. A. (2021). Mathematical modelling and analysis of COVID-19 epidemic and predicting its future situation in Ethiopia. *Results in Physics, 22*, Article 103853. https://doi.org/10.1016/j.rinp.2021.103853.

Google. (2018). *Colaboratory: Frequently asked questions*. https://research.google.com/colaboratory/faq.html

Google. (2021). *Welcome to Colaboratory*. https://colab.research.google.com/notebooks/intro.ipynb

Grave, M., Viguerie, A., Barros, G. F., Reali, A., & Coutinho, A. L. (2021). Assessing the spatio-temporal spread of COVID-19 via compartmental models with diffusion in Italy, USA, and Brazil. *Archives of Computational Methods in Engineering, 28*(6), 4205–4223. http://arxiv.org/abs/2102.07208.

Gupta, A., Wang, R., & Lu, Y. (2021). Addressing systemic risk using contingent convertible debt – A network analysis. *European Journal of Operational Research, 290*(1), 263–277. https://doi.org/10.1016/j.ejor.2020.07.062.

Gwizdałła, T. (2020). Viral disease spreading in grouped population. *Computer Methods and Programs in Biomedicine, 197*(January), Article 105715. https://doi.org/10.1016/j.cmpb.2020.105715.

Hagberg, A., Schult, D. A., & Swart, P. J. (2008). Exploring network structure, dynamics, and function using NetworkX. In *Proceedings of the 7th Python in Science Conference* (pp. 11–15).

Hale, T., Angrist, N., Goldszmidt, R., Kira, B., Petherick, A., Phillips, T., et al., (2021). A global panel database of pandemic policies (Oxford COVID-19 Government Response Tracker). *Nature Human Behaviour, 5*(4), 529–538. https://doi.org/10.1038/s41562-021-01079-8.

Hofman, J. M., Sharma, A., & Watts, D. J. (2017). Prediction and explanation in social systems. *Science (New York, N.Y.), 355*(6324), 486–488.

Hunter, E., Mac Namee, B., & Kelleher, J. (2020). A hybrid agent-based and equation based model for the spread of infectious diseases. *Journal of Artificial Societies and Social Simulation, 23*(4), 14. https://doi.org/10.18564/jasss.4421.

Jasny, B. R., & Stone, R. (2017). Prediction and its limits. *Science (New York, N.Y.), 355*(6324), 468–469. https://doi.org/10.1126/science.355.6324.468.

Jekel, C. F., Venter, G., Venter, M. P., Stander, N., & Haftka, R. T. (2019). Similarity measures for identifying material parameters from hysteresis loops using inverse analysis. *International Journal of Material Forming, 12*(3), 355–378. https://doi.org/10.1007/s12289-018-1421-8.

Jorritsma, J., Hulshof, T., & Komjáthy, J. (2020). Not all interventions are equal for the height of the second peak. *Chaos, Solitons and Fractals, 139*, Article 109965. https://doi.org/10.1016/j.chaos.2020.109965.

Kadi, N., & Khelfaoui, M. (2020). Population density, a factor in the spread of COVID-19 in Algeria: Statistic study. *Bulletin of the National Research Centre, 44*(1), 1–7.

Kermack, W. O., & McKendrick, A. G. (1927). A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society of London (Series A), 115*(772), 700–721.

Kim, Y., Ryu, H., & Lee, S. (2021). Effectiveness of intervention strategies on MERS-CoV transmission dynamics in South Korea, 2015: Simulations on the network based on the real-world contact data. *International Journal of Environmental Research and Public Health*, (7), 18. https://doi.org/10.3390/ijerph18073530.

Kiss, I. Z., Miller, J. C., & Simon, P. L. (2017), *Mathematics of epidemics on networks* p. 31. Cham: Springer.

Lotfi, R., Kheiri, K., Sadeghi, A., & Babaee Tirkolaee, E. (2022). An extended robust mathematical model to project the course of COVID-19 epidemic in Iran. *Annals of Operations Research*. https://doi.org/10.1007/s10479-021-04490-6.

Lu, X., & Borgonovo, E. (2023). Global sensitivity analysis in epidemiological modeling. *European Journal of Operational Research, 304*(1), 9–24.

Ma, T., & Nakamori, Y. (2005). Agent-based modeling on technological innovation as an evolutionary process. *European Journal of Operational Research, 166*, 741–755. https://doi.org/10.1016/j.ejor.2004.01.055.

Martcheva, M. (2015). *An introduction to mathematical epidemiology*: Vol. 61. Springer.

Masum, M., Masud, M. A., Adnan, M. I., Shahriar, H., & Kim, S. (2022). Comparative study of a mathematical epidemic model, statistical modeling, and deep learning for COVID-19 forecasting and management. *Socio-Economic Planning Sciences, 80*, Article 101249.

Mingers, J. (2007). Operational research: The science of better? *Journal of the Operational Research Society, 58*(5), 683–686.

Miranda, J. G. V., Silva, M. S., Bertolino, J. G., Vasconcelos, R. N., Cambui, E. C. B., Araújo, M. L. V., et al., (2021). Scaling effect in COVID-19 spreading: The role of heterogeneity in a hybrid ODE-network model with restrictions on the intercities flow. *Physica D. Nonlinear Phenomena, 415*, Article 132792. https://doi.org/10.1016/j.physd.2020.132792.

NCIRD. (2021). *Interim guidance on duration of isolation and precautions for adults with COVID-19*. National Center for Immunization and Respiratory Diseases (NCIRD). https://www.cdc.gov/coronavirus/2019-ncov/hcp/duration-isolation.html#print

Nikolopoulos, K. (2021). We need to talk about intermittent demand forecasting. *European Journal of Operational Research, 291*(2), 549–559.

Nikolopoulos, K., Punia, S., Schäfers, A., Tsinopoulos, C., & Vasilakis, C. (2021). Forecasting and planning during a pandemic: COVID-19 growth rates, supply chain disruptions, and governmental decisions. *European Journal of Operational Research, 290*(1), 99–115. https://doi.org/10.1016/j.ejor.2020.08.001.

Pazoki, M., & Samarghandi, H. (2021). Word-of-mouth and estimating demand based on network structure and epidemic models. *European Journal of Operational Research, 291*(1), 323–334. https://doi.org/10.1016/j.ejor.2020.09.004.

Petropoulos, F., & Makridakis, S. (2020). Forecasting the novel coronavirus COVID-19. *PloS one, 15*(3), Article e0231236.

Renardy, M., Eisenberg, M., & Kirschner, D. (2020). Predicting the second wave of

COVID-19 in Washtenaw County, MI. *Journal of Theoretical Biology, 507*, Article 110461. https://doi.org/10.1016/j.jtbi.2020.110461.

Ribeiro, M. H. D. M., da Silva, R. G., Mariani, V. C., & dos Santos Coelho, L. (2020). Short-term forecasting COVID-19 cumulative confirmed cases: Perspectives for Brazil. *Chaos, Solitons \& Fractals, 135*, Article 109853.

SciPy documentation: Scipy.integrate.odeint. (2020). Community, The SciPy. https://docs.scipy.org/doc/scipy/reference/generated/scipy.integrate.odeint.html

Shapiro, M. B., Karim, F., Muscioni, G., & Augustine, A. S. (2021). Adaptive susceptible-infectious-removed model for continuous estimation of the COVID-19 infection rate and reproduction number in the United States: Modeling study. *Journal of Medical Internet Research, 23*(4), e24389. https://doi.org/10.2196/24389.

Silal, S. P. (2021). Operational research: A multidisciplinary approach for the management of infectious disease in a global context. *European Journal of Operational Research, 291*(3), 929–934. https://doi.org/10.1016/j.ejor.2020.07.037.

Taylor, J. W., & Taylor, K. S. (2023). Combining probabilistic forecasts of COVID-19 mortality in the United States. *European Journal of Operational Research, 304*(1), 25–41.

Taylor, S. J. (2019). Distributed simulation: State-of-the-art and potential for operational research. *European Journal of Operational Research, 273*(1), 1–19. https://doi.org/10.1016/j.ejor.2018.04.03.

Treibert, S. M. (2021). *Mathematical modelling and nonstandard schemes for the corona virus pandemic*. Springer.

Vytla, V., Ramakuri, S. K., Peddi, A., Srinivas, K. K., & Ragav, N. N. (2021). Mathematical models for predicting COVID-19 pandemic: A review. *Journal of Physics: Conference Series, 1797*(1), 12009.

Wilinski, A., & Szwarc, E. (2021). A classification of countries and regions by degree of the spread of coronavirus based on statistical criteria. *Expert Systems with Applications, 172*, Article 114654. https://doi.org/10.1016/j.eswa.2021.114654.

Yaesoubi, R., & Cohen, T. (2011). Generalized Markov models of infectious disease spread: A novel framework for developing dynamic health policies. *European Journal of Operational Research, 215*(3), 679–687. https://doi.org/10.1016/j.ejor.2011.07.016.

Zádori, N., Váncsa, S., Farkas, N., Hegyi, P., & Erőss, B. (2020). The negative impact of comorbidities on the disease course of COVID-19. *Intensive Care Medicine, 46*(9), 1784–1786.

Zhang, Y., You, C., Cai, Z., Sun, J., Hu, W., & Zhou, X.-. H. (2020). Prediction of the COVID-19 outbreak in China based on a new stochastic dynamic model. *Scientific Reports, 10*(1), 1–10. https://doi.org/10.1038/s41598-020-76630-0.