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

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
SoftwareX 14 (2021) 100673

Original software publication

Toolbox to simulate and mitigate COVID-19 propagation

Zahrate El Oula Frihi a, Julian Barreiro-Gomez c,b, Salah Eddine Choutri c,b,∗, Hamidou Tembine c,b

a Lab. of Probability and Statistics (LaPS), Department of Mathematics, Badji-Mokhtar University, B.P.12, 23000, Annaba, Algeria
b NYUAD Research Institute, New York University Abu Dhabi, PO Box 129188, Abu Dhabi, UAE
c Learning & Game Theory Laboratory (L&G-Lab), New York University Abu Dhabi, PO Box 129188, Abu Dhabi, UAE

Article history:
Received 16 June 2020
Received in revised form 18 October 2020
Accepted 2 February 2021

MSC:
60J27
60-04

Keywords:
COVID-19
Markov chain
Simulation
Toolbox

ABSTRACT

In this paper, we present an age-and-migration-dependent Markov-chain-based model for the propagation of the COVID-19, together with a user-friendly MatLab-based toolbox. Furthermore, we present examples to simulate the propagation of the virus internationally and locally (in a particular country).

© 2021 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

1. Motivation and significance

COVID-19 is a new disease that was discovered in December 2019 in humans. It is believed that the virus causing the disease jumped from a wild animal to the human body in a local wildlife meat market in Wuhan in China. Since then, the disease started to spread rapidly, from human to human, in China and neighboring countries such as South Korea. The epidemic was declared as pandemic by World Health Organization (WHO) on 11 March 2020, after reaching tens of countries and leaving hundreds of deaths.

As of writing the world counts thousands of deaths and hundreds of thousands of infections in 180 countries and territories. The continents with the biggest number of infections are Asia (mainly China, Iran and South Korea), Europe (mainly western Europe) and the American continents (mainly USA and Brazil).

In the absence of a fully vaccinated population, countries try to control and slow down the spread of the pandemic by reducing the international mobility and imposing quarantine for infected people or newly arriving persons. Some countries imposed a total lockdown such as Italy and Spain.

The rapid international spread of the pandemic and the policies applied by countries to combat it, inspired us to suggest, in

https://doi.org/10.1016/j.softx.2021.100673
2352-7110/© 2021 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).
this paper, a model for the propagation of COVID-19 via Markov chains over a graph (network) representing the map of the world. Mathematical modeling of infectious diseases has been an important tool in studying the propagation of a number of diseases. Epidemic models aim to answer questions such as: is there a possibility to have a big outbreak of a certain infectious disease? how many infections will we have? how many people should be vaccinated to prevent an outbreak?

The deterministic SIR model is one of the simplest epidemic models where the propagation dynamics of the disease can be defined by 2 differential equations (see e.g. [1]), one for the number of susceptible people and the other for the infected people. The population is divided into three classes: susceptible, infected and removed (recovered or dead). In this type of model individuals can gain a permanent immunity. There exist other variants of SIR model such as SIS (susceptible–infected–susceptible) and SEIR (susceptible–exposed–infected–removed). All previous mentioned models assume that people mix homogeneously and the community is closed (no births, immigration, emigration and deaths). The closure assumption means that the total size of the population is constant, and it is equal to the sum of the number of susceptible, infected and removed people, which makes the derivation of an equation for the number of removed people trivial, given the equations for susceptible and infected people.

The output of the model relies on the assumption that people mix uniformly with each other. However, that might not be sufficient to obtain good results. For example, if we consider an epidemic outbreak in a school (small community), it seems reasonable to assume some uncertainty in the final number of infected people. Moreover, even if the basic reproduction number is larger than one and the community is large but the outbreak is initiated by few infected people, it can be possible, by chance, that the epidemic never takes off [2]. For extensions to more realistic stochastic models, see e.g. [3].

Many stochastic models of infectious diseases are based on continuous- or discrete-time Markov chains. In these models, the states of the Markov chains are the number of people that are susceptible, infected, ... etc. The major issues with these approaches are the big size of the state space and the large number of the associated Kolmogorov's differential equations describing the spread of the disease. For continuous-time Markov chain models, see e.g. [4–7]. For discrete-time models, see e.g. [8–10]. Another type of epidemic models is the network epidemic model [11]. This model is based on an undirected graph that describes the social structure in the community where the epidemic spreads. The nodes represent individuals and the edges represent the connection between individuals. Most of network-based models consider the propagation as a stochastic process or a mean-field approximation of it. The network epidemic models have other applications such as a malware propagation over computer networks, see e.g. [12,13].

1.1. A network Markov chain model

We model the spread of COVID-19 in the whole world via Markov chains over an undirected graph $G = (N, E)$, where $N = \{1, 2, \ldots, n\}$ is the set of nodes that are connected to each other by a set of edges $E$. The $n \times n$ adjacency matrix of the graph is given, for any $i,j \in N$, by: $[A]_{ij} = 1$ if $i,j \in E$, 0 otherwise. The graph represents the map of the world and each node represents an individual, the state of each individual is described by a Markov chain. Each Markov chain has a discrete state space $E = \{S, I, R, D\}$ where $S$ is susceptible, $I$ is infected, $R$ is recovered and $D$ is dead. The chain jumps from a state to another following certain transition probabilities. The infection occurs, with a probability $\alpha \in [0,1]$, when a susceptible individual interacts with an infected one from his neighbor $N$. Any infected individual (or node) can recover with probability $\gamma$. The recovered persons do not gain a permanent immunity as they can be reinfected with probability $\beta$ (see Fig. 1).

Our Markov chain model takes in consideration the age distribution in each continent. The transition probability from the state $I$ to the state $D$ (the probability $\delta$) is affected by the age. Old infected people have higher probability of death. Therefore, we set 3 different transition probabilities $\delta_1$, $\delta_2$ and $\delta_3$ for the different age categories (0,40), (40,70) and 70+, respectively. The sum of the three previous transition probabilities equals $\delta$. The age distributions for each continent is generated (by default) via normal distributions with suitable parameters. It is possible to change to other distributions depending on the continent. Furthermore, our model incorporates long distance interactions that can happen by traveling inside the same country/continent or to a different country/continent. We do that by considering a wider set of neighbors $N'$. Each person $i \in N$ has her own set of local neighborhood $N_i$ and a set of global neighborhood $N'_i$, for simplicity we dropped the subscripts.

Next, we introduce the proposed toolbox in order to simulate the propagation of the virus depending on the age and migration constraints between regions.

2. Toolbox description

The toolbox we propose is based on MatLab. It is mainly composed of six functions, which are explained briefly below. For details, refer to the associated toolbox documentation.

2.1. Construction of the continents data

The function `construct_continents(I,resolution_grid)` is in charge of generating the required data for each (populated) continent/region on the map of the world. This function takes as input the image of the world (or a country) $I$ and the chosen resolution $resolution_grid$ of its discretization (the number of nodes associated). As an output, the function returns six clusters (corresponding to each continent) composed of three elements: the number of nodes corresponding to the specific continent, a matrix of $x$–$y$ coordinates of each node specific to the continent and the corresponding filtered continent from the input image.

2.2. Initial visualization of the sets $S$, $I$, $R$, $D$, $C$, and $D$ for the continents

The function `show_map(S,I,R,D,C,color)` plots the discretized map by using the nodes constructed by the function `construct_continents`. This functions take as inputs: the set of susceptible people $S$, the set of infected people $I$, the set of recovered people $R$, the set of dead people $D$, the set of coordinates of the nodes corresponding to the map of the world $C$ and the color of the nodes corresponding to susceptible people. As an output, this function generates the plot of the map with the characteristics selected in the inputs.
2.3. Computation of a neighborhood set around a node

The function \texttt{neighbor(agent, radius, xpos, ypos)} allows computing a set around a certain node. This function is fundamental in order to compute the set of susceptible nodes around a sick node to propagate the virus. It takes the following inputs: a given node/case called \texttt{agent}, a scalar value called \texttt{radius} which determines the size of the neighborhood of a node (high values for international travels), \texttt{xpos} corresponding to all x coordinates of all the nodes on the map of the world, and \texttt{ypos} corresponding to all y coordinates of all the nodes on the map of the world.

This function returns a unique output which is a set of nodes corresponding to the neighborhood set of \texttt{agent}.

2.4. Generation of the age distribution

The function \texttt{show_age(C, age1, age2, age3)} plots the map of the world and showing the distribution of three age ranges defined by \texttt{age1}, \texttt{age2}, and \texttt{age3}. After inserting the distribution of ages for each continent (there are three ranges of age), this function plots the map of the world and assigning each node an age accordingly.
Fig. 4. Propagation of the virus in Colombia with migration constraints among the 32 departments (without quarantine).

Fig. 5. Propagation of the virus in Colombia with migration constraints among the 32 departments (with quarantine).

2.5. Computation of the flight connections

The function `flight_connections(A,S_NorthAmerica, ..., S_Oceania)` computes the set of nodes belonging to the regions to which a node can travel.

This function takes as input the set of nodes (susceptible people) corresponding to each continent and an adjacency matrix $A \in \{0, 1\}^{6 \times 6}$ describing the possible intercontinental travel. As outputs, the function returns set of nodes to which an agent belonging to a certain continent can move to another. Each output set corresponds to a continent.

2.6. Computation of the virus propagation

The function `update(S,I,R,D,xpos,ypos, flight_connections,p,Age)` updates the situation of the propagation of the virus after each iteration. This is done by updating all the input sets for each continent. The input `flight_connections` is the set of agents that can be infected from other continent because of flights and connectivity of continents, whereas $p$ is the set of transition probabilities $p = [p1, p2, p31, p32, p33, p4, p5]$ of the underlying Markov chain. As outputs, the sets of susceptible, infected, recovered and dead are updated.

3. Illustrative examples

We present two illustrative examples, the first is a simulation of the propagation of the virus on the map of the world, which is divided into six different continents, the simulation is done with and without intercontinental connections, see 2 and 3. The second example considers a unique country divided into several regions. Let the country be Colombia, which is divided into 32 departments. Therefore, we consider not only the different transition rates for the departments, but also the migration constraints among them.

Fig. 4 shows the propagation of the virus during a year without quarantine whereas Fig. 5 shows the propagation of the virus during a year with quarantine and no allowed domestic flights.

4. Impact

The MatLab toolbox can help in raising research questions about the efficiency of the lockdown and the closure of aerial
spaces. Using the toolbox one can simulate the spatial propagation of the disease and compare different lockdown strategies. Furthermore, our algorithm can be easily extended to take into account finer division of the map of the world. One can divide each continent into countries, each country into regions and each region into zones and then find out best lockdown policies by managing the interconnection between all of them.

This toolbox simplifies the simulation process of the propagation of COVID-19 (or any other viruses) and made the simulation more realistic by taking into account the age distribution in each continent as well as long distance travels. Our research group at Learning & Game Theory Laboratory uses the toolbox to generate simulations and study the propagation of COVID-19. Moreover, the lab is working on extending it to including different divisions and adding more features such as hospital capacity in each region, economic situation,... etc.

5. Conclusions

We have presented a Markov chain-based model for the propagation of COVID-19 with migration constraints given by a connected graph representing the allowed air traffic. Furthermore, the model takes into account an age-dependent mortality since it has been observed that the elderly people are more prone to death by the virus than younger people. The model is simple and can be easily tuned by means of the transition probabilities of the Markov chain. Besides, we have presented a user-friendly toolbox for the simulation of the propagation throughout multiple regions. The regions can be different continents, countries and/or states in a country. We have presented several examples explaining the use of the toolbox and have also provided the open source codes, which are accessible to any researcher. The goal is that the contribution presented in this article can be used, not only to motivate the research of the propagation of the COVID-19, but also as a model that can be used as a reference in the design of optimal controllers.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

This material is based upon work supported by Tamkeen under the NYU Abu Dhabi Research Institute grant CG002, and by U.S. Air Force Office of Scientific Research under grant number FA9550-17-1-0259.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.softx.2021.100673.

References

[1] Kermack WO, McKendrick AG. A contribution to the mathematical theory of epidemics. Proc R Soc A, Containing papers of a mathematical and physical character 1927;115(772):700–21.
[2] Britton Tom. Stochastic epidemic models: a survey. Math Biosci 2010;225(1):24–35.
[3] Allen LJ. An introduction to stochastic epidemic models. In: Mathematical epidemiology. Springer; 2008, p. 81–130.
[4] Jacquez JA, O’Neill P. Reproduction numbers and thresholds in stochastic epidemic models i. homogeneous populations. Math Biosci 1991;107(2):161–86.
[5] Jacquez JA, Simon CP. The stochastic SI model with recruitment and deaths I. Comparison with the closed SIS model. Elsevier; 1993.
[6] Nåsell I. Stochastic models of some endemic infections. Math Biosci 2002;179(1):1–19.
[7] Keeling MJ, Ross JV. On methods for studying stochastic disease dynamics. J R Soc Interface 2008;5(19):171–81.
[8] Abbey H. An examination of the reed-frost theory of epidemics. Human Biol 1952;24(3):201.
[9] Castillo-Chavez C, Yakubu A. Discrete-time SIS models with complex dynamics. Nonlinear Anal TMA 2001;47(7):4753–62.
[10] Yaesoubi R, Cohen T. Generalized Markov models of infectious disease spread: A novel framework for developing dynamic health policies. European J Oper Res 2011;215(3):679–87.
[11] Newman MJ. The structure and function of complex networks. SIAM Rev 2003;45(2):167–256. http://dx.doi.org/10.1137/S003614450342480.
[12] Tempine H. Mean-field-type games in engineering. AIMS Electron Electr Eng 2017;1:18. http://dx.doi.org/10.3934/ElectrEng.2017.1.18.
[13] Garetto M, Gong W, Towsley D. Modeling malware spreading dynamics. In: IEEE INFOCOM 2003. Twenty-Second annual joint conference of the IEEE computer and communications societies (IEEE Cat. No. 03CH37428). 3, IEEE; 2003, p. 1869–79.