Physica A 490 (2018) 1513–1521

Contents lists available at ScienceDirect

Physica A

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

Solving the patient zero inverse problem by using generalized simulated annealing

Olavo H. Menin [a,*], Chris T. Bauch [b]

[a] Instituto Federal de Educação, Ciência e Tecnologia de São Paulo, Rua Pedro Vicente, 625, 01109-010, São Paulo, SP, Brazil
[b] University of Waterloo, 200 University Avenue West, Waterloo, ON, N2L 3G1, Canada

HIGHLIGHTS

- The problem of finding the source of an epidemic outbreak is addressed.
- The main disease natural histories, SI, SIS, SIR and SIRS, are considered.
- The performance of the Generalized Simulated Annealing algorithm is assessed.
- Good accuracy is obtained even only partial data of the population are available.

ARTICLE INFO

Article history:
Received 10 March 2017
Received in revised form 27 May 2017
Available online 18 September 2017

Keywords:
Epidemic models
Index case
Infection disease
Stochastic optimization

ABSTRACT

Identifying patient zero – the initially infected source of a given outbreak – is an important step in epidemiological investigations of both existing and emerging infectious diseases. Here, the use of the Generalized Simulated Annealing algorithm (GSA) to solve the inverse problem of finding the source of an outbreak is studied. The classical disease natural histories susceptible–infected (SI), susceptible–infected–susceptible (SIS), susceptible–infected–recovered (SIR) and susceptible–infected–recovered–susceptible (SIRS) in a regular lattice are addressed. Both the position of patient zero and its time of infection are considered unknown. The algorithm performance with respect to the generalization parameter $\tilde{q}_v$ and the fraction ρ of infected nodes for whom infection was ascertained is assessed. Numerical experiments show the algorithm is able to retrieve the epidemic source with good accuracy, even when ρ is small, but present no evidence to support that GSA performs better than its classical version. Our results suggest that simulated annealing could be a helpful tool for identifying patient zero in an outbreak where not all cases can be ascertained.

1. Introduction

Recent decades have witnessed a prodigious growth in both the number and quality of publications on infectious disease modeling [1]. Such models have been an invaluable tool to describe epidemic process and aid decision-makers in adopting suitable strategies to control and eliminate infectious diseases [2]. In general, epidemic modelers have focused on predicting how the infection spreads through the population using available information such as the type of infectious pathogens, the infection rate, and the social structure of the population, among others [3,4].

* Corresponding author.
E-mail address: olavohmenin@ifsp.edu.br (O.H. Menin).

http://dx.doi.org/10.1016/j.physa.2017.08.077
0378-4371/

A relevant issue in epidemiological investigations of outbreaks concerns inferring the infected individual who is the source of an outbreak ('patient zero', or equivalently, the index case), using information obtained from later stages in the outbreak. This is sometimes referred to as the 'patient zero' inverse problem. The use of the expression 'patient zero' came from the early 1980s when a Canadian air steward was placed in the center of a sexual network of HIV-infected individuals [5]. The air steward was long thought to be patient zero of the United States HIV/AIDS epidemic. However, recent research has rewritten the established understanding of origin of the US HIV/AIDS epidemic and absolved the unfortunate Canadian flight attendant [6]. The mis-identification of patient zero in this example underscores the need for accurate methodologies to identify the source of an epidemic outbreak.

Since that time, the term 'patient zero' has been adopted to refer not only to the source of an infectious disease outbreak but also for rumor dissemination and computer virus propagation [7–9]. Patient zero in the Hong Kong SARS outbreak infected 47 healthcare workers in the hospital he was admitted to, after contracting the SARS virus on a trip to visit family in Guangdong province [10]. SARS is highly symptomatic and very dangerous, therefore case ascertainment rates were relatively high in the SARS outbreaks. However, some infectious diseases such as influenza exhibit a higher rate of asymptomatic infections and therefore a lower case ascertainment rate. This means that methods for identifying patient zero should be robust to imperfect case ascertainment.

Solving the patient zero problem brings all familiar challenges of the inverse problems since it does not satisfy the Hadamard criteria for a well-posed problem: (1) existence, (2) uniqueness and (3) stability of the solution. In spite of these difficulties there are several techniques to deal with inverse problems [11,12] and one them is to treat them as optimization problems. Is this case, one must minimize an objective function which evaluates the discrepancy between two set of data, one measured experimentally and another predicted by a mathematical model.

Previous mathematical models have addressed the patient zero problem with various methods. Most of them consider a single type of disease natural history, tree-like networks and known spread time [7,8,13]. Here our goal is to assess the use of the Generalized Simulated Annealing (GSA) algorithm in addressing the patient zero inverse problem. Based on the classical Simulated Annealing (SA) [14], GSA is a stochastic optimization algorithm, proposed originally by Tsallis and Stariolo [15], that has been applied in different problems such as protein folding [16], spin phase transition in the Ising model [17,18] and X-ray spectrum reconstruction [19]. Basically, candidate solutions are generated iteratively following the generalized Gaussian distribution and are accepted or not according to a Metropolis-like criterion [20]. The stochasticity of both the generation of solutions and the acceptance criterion are controlled by the "temperature", which is gradually decreased through a predefined cooling schedule.

Computational simulations of epidemic processes on a lattice were used to evaluate GSA performance mainly with respect to the parameters $\tilde{q}_v$, which generalizes the visiting distribution and the cooling schedule, and ρ, the fraction of the nodes for which infection status can be ascertained during a 'snapshot' in time. Our algorithm was designed to be fast and robust since it receives as input data only the epidemic state of the nodes at the exact time in which the 'snapshot' is taken and therefore does not depend on any information about the earlier stages of the outbreak. Furthermore, it can operate under different epidemic models. Indeed, the standard infection natural histories susceptible–infected (SI), susceptible–infected–susceptible (SIS), susceptible–infected–recovered (SIR) and susceptible–infected–recovered–susceptible (SIRS) in a regular lattice were investigated [21]. Finally, the algorithm tackles the problem when not only the location of patient zero but also their time of initial infection are considered unknown and must be inferred. Results show that the algorithm can find the location of the patient zero with good accuracy even when only about 10% of the nodes can have their infection status ascertained during the 'snapshot'. Moreover the simulation outcomes diverge from the literature with respect to the best value for the parameter $\tilde{q}_v$.

In the following Section 2 we formulate the epidemic spread model and define the objective function to be minimized. Then, in Section 3 we review the main features of the GSA algorithm. In Section 4 we describe our numerical experiments and discuss our results, and in Section 5 we summarize the paper with concluding comments.

2. Epidemic model and the objective function

In order to solve the patient zero inverse problem using the optimization approach, we must consider two steps: (1) the forward problem, which is a model to simulate the epidemic spread under some guess for the location of patient zero and the time of its initial infection; (2) the global minimization of an objective function that confronts two set of data, one obtained from the actual epidemic spread and another generated numerically by the forward model. As is widely known, in a well-mixed population with no demography, the SI, SIS, SIR and SIRS models can be described by the equations

$$\frac{dS}{dt} = -\beta SI + \nu I + \eta R, \tag{1}$$

$$\frac{dI}{dt} = \beta SI - (\nu + \gamma)I, \tag{2}$$

$$\frac{dR}{dt} = \gamma I - \eta R, \tag{3}$$

$$S + I + R = 1, \tag{4}$$