Physica A 389 (2010) 4755–4768

Contents lists available at ScienceDirect

Physica A

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

Discovering network behind infectious disease outbreak

Yoshiharu Maeno *

Social Design Group, Tokyo, Japan

ARTICLE INFO

Article history:
Received 31 March 2010
Received in revised form 11 June 2010
Available online 16 July 2010

Keywords:
Epidemiological compartment model
Meta-population network model
Maximal likelihood estimation
Severe Acute Respiratory Syndrome
Stochastic differential equation

ABSTRACT

Stochasticity and spatial heterogeneity are of great interest recently in studying the spread of an infectious disease. The presented method solves an inverse problem to discover the effectively decisive topology of a heterogeneous network and reveal the transmission parameters which govern the stochastic spreads over the network from a dataset on an infectious disease outbreak in the early growth phase. Populations in a combination of epidemiological compartment models and a meta-population network model are described by stochastic differential equations. Probability density functions are derived from the equations and used for the maximal likelihood estimation of the topology and parameters. The method is tested with computationally synthesized datasets and the WHO dataset on the SARS outbreak.

1. Introduction

When the epidemiologists at a public health agency detect a signal of an infectious disease outbreak, they rely heavily on mathematical models of disease transmission in estimating the rate of transmission, predicting the direction and speed of the spread, and figuring out an effective measure to contain the outbreak. Many of the models formulate stochasticity and spatial heterogeneity, which are of great interest recently. The spatial heterogeneity ranges from the uneven probabilities of contacts between the individuals in communities [1,2], dependence of the strength of the demographical interactions between cities on the distance [3], to nation-wide or world-wide inhomogeneous geographical structures [4,5].

A Monte-Carlo stochastic simulation is widely used to understand the influence of the spatial heterogeneity on stochastic spreading. In such a simulation, accuracy and reproducibility of the input demographical knowledge such as the amount of traffic between cities have great impacts on the reliability of the output pattern of the movement of pathogens and their hosts. But, in studying world-wide epidemics, just a collection of regular airline routes and aircraft capacities does not always present the transportation network which results in the real chain of transmission. Some routes are influential decisively, but the others are not. Examples are found in the spread of Severe Acute Respiratory Syndrome (SARS) from Asia to the world in 2003. There were not any cases in Japan in spite of the heavy traffic there from Asian countries. Many patients appeared in Canada earlier than the United States to which airlines connect Asian countries much more densely. Here arises an interesting question. Inversely, is it possible to learn the effectively decisive transportation network by observing how the disease spreads, reinforce the demographical knowledge on the network, and import the acquired knowledge into the mathematical model? This is an inverse problem similar to the network tomography [6,7].

In this study, a statistical method is presented to discover the effectively decisive topology of a heterogeneous network and reveal the parameters which govern stochastic transmission from a dataset on the early growth phase of the outbreak. The dataset consists of either the number of infectious persons or the number of new cases per an observation interval. The method is founded on a mathematical model for a stochastic reaction–diffusion process [8]. The population in the

* Corresponding address: Social Design Group, 1-6-38F Sengoku, Bunkyo-ku, Tokyo 112-0011, Japan.
E-mail address: maeno.yoshiharu@socialdesigngroup.com.

0378-4371/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.physa.2010.07.014

model is described by a set of Langevin equations. The equations are stochastic differential equations which include rapidly fluctuating and highly irregular functions of time. Probability density functions and likelihood functions are derived from the equations analytically, and used for the maximal likelihood estimation of the topology and parameters. The method is tested with a number of computationally synthesized datasets and the World Health Organization (WHO) dataset on the SARS outbreak in March through April in 2003.

2. Problem

2.1. Stochastic model

The model in this study is a special case of a stochastic reaction–diffusion process. The model is a combination of standard epidemiological SIR or SIS compartment models and a meta-population network model [9]. The meta-population network model [10] sub-divides the entire population into distinct sub-populations in N geographical regions. Movement of persons occurs between the sub-populations while the epidemiological state transitions (infection and recovery) occur in a sub-population. A sub-population is randomly well-mixed. Heterogeneity is present between sub-populations.

The geographical regions are represented by nodes n_i ($i = 0, 1, \ldots, N-1$). The movement is parameterized by a matrix $\boldsymbol{\gamma}$ whose i-th row and j-th column element γ_{ij} is the probability at which a person moves from n_i to n_j per a unit time. A person remains at the same node at the probability of $1-\sum_{j=0}^{N-1}\gamma_{ij}$. Generally, $\gamma_{ij} = \gamma_{ji}$ does not hold. By definition, $\gamma_{ii} = 0$. It is often confirmed empirically that a simple law relates a network topology to the movement [11]. The topology is specified by a neighbor matrix $\boldsymbol{l}$. The transportation between two regions is represented by a pair of unidirectional links. If a pair of links is present between n_i and n_j, $l_{ij} = l_{ji} = 1$. If absent, $l_{ij} = l_{ji} = 0$. By definition, $l_{ii} = 0$. In the experiments in Section 4, an empirically confirmed law $\gamma_{ij} = \Gamma_{ij}(\boldsymbol{l})$ is postulated, and the topology and the probability of movement are treated interchangeably.

The SIR compartment model [12] is a behavioral extreme where immunity is life-long. The state of a person changes from a susceptible state (S), through an infectious state (I), to a recovered state (R). In contrast, the immunity does not occur in the SIS compartment model. The state of a recovered patient goes back to S. The parameter α represents the probability at which an infectious person contacts a person and infect the person per a unit time. If the contacted person is susceptible, the number of the infectious persons increases by 1. The effective rate of infection by a single infectious person is the product of α and the proportion of the susceptible persons within the population. The parameter β represents the probability at which an infectious person recovers per a unit time. These parameters are constants over subpopulations and time. The basic reproductive ratio r is defined by $r = \alpha/\beta$ [13].

Movement, infection and recovery are Markovian stochastic processes governed by γ_{ij}, α, and β.

2.2. Time evolution of spread

In a stochastic process, even if the initial condition is known, there are many possible trajectories which the process might go along. A set of these possible trajectories is a statistical ensemble. The change in the population is described by a set of Langevin equations [14]. A Langevin equation is a stochastic differential equation [15]. The microscopic continuous time evolution of a system is obtained by adding a fluctuation (a stochastic term) to the known macroscopic time evolution of the system.

The quantity $S_i(t)$ is the number of susceptible persons at a node n_i at time t. $I_i(t)$ is the number of infectious persons. $R_i(t)$ is the number of recovered persons. The change in $I_i(t)$ ($i = 0, 1, \ldots, N-1$) is given by Eq. (1) [16]. It is a set of N stochastic differential equations.

$$\begin{aligned}\frac{\mathrm{d}I_i(t)}{\mathrm{d}t} &= \frac{\alpha S_i(t) I_i(t)}{S_i(t)+I_i(t)+R_i(t)} - \beta I_i(t) + \sum_{j=0}^{N-1}\gamma_{ji} I_j(t) - \sum_{j=0}^{N-1}\gamma_{ij} I_i(t) \\ &\quad + \sqrt{\frac{\alpha S_i(t) I_i(t)}{S_i(t)+I_i(t)+R_i(t)}}\,\xi_i^{[\alpha]}(t) - \sqrt{\beta I_i(t)}\,\xi_i^{[\beta]}(t) + \sum_{j=0}^{N-1}\sqrt{\gamma_{ji} I_j(t)}\,\xi_{ji}^{[\gamma]}(t) - \sum_{j=0}^{N-1}\sqrt{\gamma_{ij} I_i(t)}\,\xi_{ij}^{[\gamma]}(t). \end{aligned} \tag{1}$$

Stochastic terms $\boldsymbol{\xi}(t) = (\xi_i^{[\alpha]}(t), \xi_i^{[\beta]}(t), \xi_{ij}^{[\gamma]}(t))$ are rapidly fluctuating and highly irregular functions of time. The number of terms is $M = N^2 + N$ (N terms for infection, N terms for recovery, and $N(N-1)$ terms for movement). The functional forms of individual elements $\xi_a(t)$ ($a = 0, 1, \ldots, M-1$) are not known. Their statistical property is the Gaussian white noise which satisfies Eq. (2) through (4).

$$\langle \xi_a(t) \rangle_{\text{ensemble}} = 0 \tag{2}$$

$$\langle \xi_a(t)\xi_b(u) \rangle_{\text{ensemble}} = \delta_{ab}\delta(u-t) \tag{3}$$

$$\langle \xi_a(t)\xi_b(u)\xi_c(v)\cdots \rangle_{\text{ensemble}} = 0. \tag{4}$$