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.
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.

© 2010 Elsevier B.V. All rights reserved.

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 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
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, …, N – 1). The movement is parameterized by a matrix \( w \) whose \( i \)-th row and \( j \)-th column element \( w_{ij} \) 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 \( 1 - \sum_{j=0}^{N-1} w_{ij} \). Generally, \( w_{ij} = w_{ji} \) does not hold. By definition, \( w_{ii} = 0 \). 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 \( L \). 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 \( w_{ij} = L_{ij} \) 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 \( \alpha \) 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 \( \alpha \) and the proportion of the susceptible persons within the population. The parameter \( \beta \) 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 \( w_{ij} \), \( \alpha \), and \( \beta \).

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, …, N – 1) is given by Eq. (1) [16]. It is a set of N stochastic differential equations.

\[
\frac{dI_i(t)}{dt} = \frac{\alpha S_i(t)I_i(t)}{S_i(t) + I_i(t) + R_i(t)} - \beta I_i(t) + \sum_{j=0}^{N-1} \gamma_{ij}I_j(t) - \sum_{j=0}^{N-1} \gamma_{ji}I_i(t) + \sqrt{\frac{\alpha S_i(t)I_i(t)}{S_i(t) + I_i(t) + R_i(t)}} \xi_i^{[\alpha]}(t) - \sqrt{\beta I_i(t)S_i^{[\beta]}(t)} + \sum_{j=0}^{N-1} \sqrt{\gamma_{ij}I_j(t)S_i^{[\gamma]}(t)} - \sum_{j=0}^{N-1} \sqrt{\gamma_{ji}I_i(t)S_j^{[\gamma]}(t)}.
\]

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

\[
\langle \xi_a(t) \rangle_{\text{ensemble}} = 0
\]

\[
\langle \xi_a(t) \xi_b(u) \rangle_{\text{ensemble}} = \delta_{ab} \delta(u - t)
\]

\[
\langle \xi_a(t) \xi_b(u) \xi_c(v) \cdots \rangle_{\text{ensemble}} = 0.
\]
In these equations, $\delta(t)$ is a Dirac’s delta function, and $\delta_{ab}$ is a Kronecker’s delta symbol. The ensemble average of a variable $x$ is $\langle x \rangle_{\text{ensemble}}$. Eq. (3) means that there is no correlation at different times and between different terms. Eq. (4) means that the third and higher order moments vanish.

In most cases, the outbreak is contained before the spread reaches equilibrium. In the early growth phase of the outbreak, $I_t \ll S_t$ and $R_t \ll S_t$ hold true. The first term of the right side of Eq. (1) is independent of $S_t$ and $R_t$ because $S_t / (S_t + I_t + R_t) \approx 1$. The resulting equation is Eq. (5).

$$\frac{dl_t(t)}{dt} = \alpha l_t(t) - \beta l_t(t) + \sum_{j=0}^{N-1} \gamma_{j} d_{j}(t) - \sum_{j=0}^{N-1} \gamma_{j} l_{j}(t) + \sqrt{\alpha l_t(t)\xi_{i}^{[\alpha]}(t)} - \sqrt{\beta l_t(t)\xi_{i}^{[\beta]}(t)} + \sum_{j=0}^{N-1} \sqrt{\gamma_{j} d_{j}(t)\xi_{i}^{[\gamma]}(t)} - \sum_{j=0}^{N-1} \sqrt{\gamma_{j} l_{j}(t)\xi_{i}^{[\gamma]}(t)}.$$ (5)

The cumulative number of new cases until time $t$ is represented by $J_j(t)$ ($i = 0, 1, \ldots, N - 1$). The rate of increase in $J_j(t)$ equals to the first term of Eq. (5). That is, $\alpha l_t(t)$. The time evolution of $J_j(t)$ is given by Eq. (6).

$$\frac{dj_t(t)}{dt} = \alpha l_t(t) + \sqrt{\alpha l_t(t)\xi_{i}^{[\alpha]}(t)}.$$ (6)

The total number of the infectious persons at time $t$ is given by $I(t) = \sum_{i=0}^{N-1} I_i(t)$. Its time evolution is given by Eq. (7).

$$\frac{dl(t)}{dt} = \alpha l(t) - \beta l(t) - \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \sqrt{\alpha l(t)\xi_{i}^{[\alpha]}(t)} - \sum_{i=0}^{N-1} \sqrt{\beta l(t)\xi_{i}^{[\beta]}(t)}.$$ (7)

The total cumulative number of new cases until time $t$ is given by $J(t) = \sum_{i=0}^{N-1} J_i(t)$. Its time evolution is given by Eq. (8).

$$\frac{dj(t)}{dt} = \alpha l(t) + \sum_{i=0}^{N-1} \sqrt{\alpha l(t)\xi_{i}^{[\alpha]}(t)}.$$ (8)

2.3. Definition of problem

The problem is to discover the network topology $I$ (or $\gamma$) and reveal the transmission parameter $r$ (or $\alpha$ and $\beta$) from a given dataset $l_i(t_d)$ ($i = 0, 1, \ldots, N - 1, d = 0, 1, \ldots, D - 1$) or $J_i(t_d)$ ($i = 0, 1, \ldots, N - 1, d = 0, 1, \ldots, D - 1$). The dataset $l_i(t_d)$ is the time sequence of the number of infectious persons. The dataset $J_i(t_d) = J_i(t_{d+1}) - J_i(t_d)$ is the time sequence of the number of new cases between observations. An observation is made at every node $n_i$ ($i = 0, 1, \ldots, N - 1$) at times $t_d$ ($d = 0, 1, \ldots, D - 1$). The time interval between observations is $\Delta t = t_{d+1} - t_d$. For example, a bundle of the daily reports on cases from hospitals is a dataset $\Delta J_i(t_d)$ where $\Delta t = 1$ day. Other information is not known. That is, nothing is known about $S_i(t_d)$, $R_i(t_d)$, nor the initial condition which could identify the index case (the first patient from whom the infectious disease has spread).

3. Method

3.1. Likelihood function

Various techniques of statistical inference can be applied once the likelihood function is obtained analytically. The likelihood function is the conditional probability of the obtained dataset as a function of the unknown parameters of a parameterized statistical model. The conditional probability becomes noticeably large if the value of the parameters is close to the true value. For example, maximal a posteriori estimation is used to find the parameters which maximize the posterior distribution. In this study, the problem is solved by maximal likelihood estimation. The Langevin equation (5) through (8) are solved by obtaining the moments of probability variables at time $t$ so that the probability density functions and logarithmic likelihood functions can be derived, rather than by calculating the trajectories of time-dependent variables for a given functional form of $\xi_{i}(t)$ [4]. Four logarithmic likelihood functions $L^{[1]}(\theta), L^{[2]}(\theta), L^{[1]}(\theta)$, and $L^{[2]}(\theta)$ are derived for given datasets $l_i(t_d), I_l(t_d), \Delta J_i(t_d)$, and $\Delta J_i(t_d)$ respectively under the unknown parameters $\theta = [\gamma, \alpha, \beta]$.

Appendix A presents the procedure to solve the Langevin equations through a Fokker–Planck equation [17]. Appendix B summarizes the formula for the time evolution of $m^{[1]}(t|\theta), m^{[1]}(t|\theta)$ (row vectors whose $i$-th element is the mean of $I_i, J_i$) and $v^{[1]}(t|\theta), v^{[1]}(t|\theta)$ (matrices whose $i$-th row and $j$-th column element is the covariance between $I_i$ and $J_i$). Appendix C summarizes the formula for the time evolution of $m^{[2]}(t|\theta), m^{[2]}(t|\theta)$ (the mean of $l, J$) and $v^{[2]}(t|\theta), v^{[2]}(t|\theta), v^{[2]}(t|\theta)$ (the variance of $I$, covariance between $l$ and $J$, variance of $J$).
3.1.1. Case 1: for given \( I(t_d) \)

The logarithmic likelihood function \( L^{[I]}(\theta) \) is determined by \( I(t_d) \). If the third and higher order moments are ignored, the probability density function \( p(I, t_{d+1}|\theta) \) is a multi-variate Gaussian distribution with the mean \( m^{[I]}(t_{d+1}|\theta) \) and covariance \( v^{[I]}(t_{d+1}|\theta) \) in Eq. (9). This is the probability of \( I = \{I_0, \ldots, I_{n-1}\} \) at \( t = t_{d+1} \) given \( \theta \). \( X^T \) is a transpose of a matrix \( X \).

\[
p^{[I]}(I, t_{d+1}|\theta) = \frac{\exp \left( \frac{1}{2} (I - m^{[I]}(t_{d+1}|\theta)) v^{[I]}(t_{d+1}|\theta)^{-1} (I - m^{[I]}(t_{d+1}|\theta))^T \right)}{\sqrt{(2\pi)^N \det v^{[I]}(t_{d+1}|\theta)}}.
\]

(9)

\( L^{[I]}(\theta) \) is the logarithm of a product of the probability of the individual observation \( I(t_{d+1}) \) at \( t = t_{d+1} \). It is given by Eq. (10).

\[
L^{[I]}(\theta) = \sum_{d=0}^{D-2} \log p^{[I]}(I(t_{d+1}), t_{d+1}|\theta).
\]

(10)

If \( \Delta t \) is small, the formula for the moments become simpler. The exact formula for \( m^{[I]}(t|\theta) \) and \( v^{[I]}(t|\theta) \) are expanded in terms of \( \Delta t \), and the second and higher order terms are ignored. If the data \( I(t_d) \) at \( t = t_d \) is reliable completely, \( m^{[I]}(t_d|\theta) = I(t_d) \) and \( v^{[I]}(t_d|\theta) = 0 \). The moments after \( \Delta t \) are given by Eqs. (11) and (12). The moments at \( t = t_{d+1} \) depend only on the data at \( t = t_d \).

\[
m^{[I]}_i(t_{d+1}|\theta) \approx I(t_d) + \left\{ (\alpha - \beta - \sum_{k=0}^{N-1} \gamma_k) I(t_d) + \sum_{j=0}^{N-1} \gamma_j I_j(t_d) \right\} \Delta t
\]

(11)

\[
v^{[I]}_{ij}(t_{d+1}|\theta) \approx \left\{ (\alpha + \beta + \sum_{k=0}^{N-1} \gamma_k) I(t_d) + \sum_{k=0}^{N-1} \gamma_k I_k(t_d) \right\} \delta_{ij} - \gamma_i I_i(t_d) - \gamma_j I_j(t_d) \right\} \Delta t.
\]

(12)

Similarly, the logarithmic likelihood function \( L^{[2]}(\theta) \) is determined by \( I(t_d) \). \( I(t_d) \) can be calculated from the given dataset \( I(t_d) \). The probability density function \( p^{[2]}(I, t_{d+1}|\theta) = p^{[2]}(I, t_{d+1}|\alpha, \beta) \) is a Gaussian distribution with the mean \( m^{[2]}(t_{d+1}|\theta) \) and variance \( v^{[2]}(t_{d+1}|\theta) \). It does not depend on \( \gamma \). \( L^{[2]}(\theta) = L^{[2]}(\alpha, \beta) \) is the logarithm of a product of the probability of individual observation \( I(t_{d+1}) \) at \( t = t_{d+1} \). It is given by Eq. (13).

\[
L^{[2]}(\alpha, \beta) = \sum_{d=0}^{D-2} \log p^{[2]}(I(t_{d+1}), t_{d+1}|\alpha, \beta).
\]

(13)

Again, if \( \Delta t \) is small, the formula for the moments become simpler. They are given by Eqs. (14) and (15).

\[
m^{[I]}(t_{d+1}|\theta) \approx I(t_d) + (\alpha - \beta) I(t_d) \Delta t
\]

(14)

\[
v^{[I]}(t_{d+1}|\theta) \approx (\alpha + \beta) I(t_d) \Delta t.
\]

(15)

3.1.2. Case 2: for given \( \Delta J(t_d) \)

The logarithmic likelihood function \( L^{[I]}(\theta) \) is determined by \( \Delta J(t_d) \). The probability density function \( p^{[I]}(J, t_{d+1}|\theta) \) is a multi-variate Gaussian distribution in the same functional form as Eq. (9) with the mean \( m^{[I]}(t_{d+1}|\theta) \) and covariance \( v^{[I]}(t_{d+1}|\theta) \). The functional form of \( L^{[I]}(\theta) \) is also the same as Eq. (10). \( J(t_{d+1}) \) can be calculated from the dataset by \( J(t_{d+1}) = J(t_d) + \sum_{\alpha=0}^{D-1} \Delta J(t_d) \). The first observation \( \Delta J(t_d) \) may include all the known cases at that time. Then, \( J(t_{d+1}) \) can be deleted from the above formula. However, there is a big difference from \( L^{[I]}(\theta) \). The probability at \( t = t_{d+1} \) is not determined by the data \( J(t_d) \) because the moments of \( J \) at \( t = t_{d+1} \) depends only on \( J \) at \( t = t_d \) whose value is not known. Such an approximation as Eq. (14) or (15) is not correct. Thus, the exact formula for \( v^{[I]}(t|\theta) \) as a function of \( t \) must be evaluated to calculate the value of \( L^{[I]}(\theta) \).

Similarly, the logarithmic likelihood function \( L^{[2]}(\theta) \) is determined by \( J(t_d) \). \( J(t_d) \) can be calculated from the given dataset \( \Delta J(t_d) \). The probability density function \( p^{[2]}(J, t_{d+1}|\theta) = p^{[2]}(J, t_{d+1}|\alpha, \beta) \) is a Gaussian distribution with the mean \( m^{[2]}(t_{d+1}|\theta) \) and variance \( v^{[2]}(t_{d+1}|\theta) \). The functional form of \( L^{[2]}(\theta) = L^{[2]}(\alpha, \beta) \) is the same as Eq. (13). The approximation in Eqs. (14) and (15) is not correct either. Thus, the exact formula for \( v^{[2]}(t|\theta) \) as a function of \( t \) must be evaluated to calculate the value of \( L^{[2]}(\alpha, \beta) \).

3.2. Estimation procedure

Theoretically, every formula in the following can be applied in estimating \( \gamma \) directly, as well as estimating \( I \) and obtaining \( \gamma \) by the law \( \gamma_j = I_j - I \). But, the estimation of \( N(N-1)/2 \) binary parameters \( I \) tends to be more robust than that of \( N(N-1) \) continuous parameters \( \gamma \). The binary parameters are suitable for reliable combinatorial optimization by means of well-established numerical algorithms and computational implementations. Thus, the estimation of \( I \) is detailed here and demonstrated in Section 4.
3.2.1. Case 1: for given \( I_d(t_d) \)

The procedure for the estimation from \( I_d(t_d) \) is presented. The problem is solved by dividing it to two sub-problems and solving them sequentially, rather than by searching the maximal likelihood estimators \( \hat{\alpha} \), \( \hat{\beta} \), and \( \hat{I} \) simultaneously. The first sub-problem is to obtain \( \hat{\alpha} \) and \( \hat{\beta} \) by solving Eq. (16).

\[
\hat{\alpha}, \hat{\beta} = \arg \max_{\alpha, \beta} L^{[2]}(\alpha, \beta).
\]

The estimators are given by Eqs. (17) and (18) where \( \Delta I(t_d) = I(t_{d+1}) - I(t_d) \).

\[
\hat{\alpha} = \frac{1}{2\Delta t} \left\{ \frac{1}{D} \sum_{d=0}^{D-1} \frac{\Delta I(t_d)}{I(t_d)} - \frac{1}{D} \left( \frac{1}{D} \sum_{d=0}^{D-1} I(t_d) \right)^2 + \frac{1}{D} \sum_{d=0}^{D-1} \frac{\Delta I(t_d)}{I(t_d)} \right\}
\]

\[
\hat{\beta} = \frac{1}{2\Delta t} \left\{ \frac{1}{D} \sum_{d=0}^{D-1} \frac{\Delta I(t_d)}{I(t_d)} - \frac{1}{D} \left( \frac{1}{D} \sum_{d=0}^{D-1} I(t_d) \right)^2 - \frac{1}{D} \sum_{d=0}^{D-1} \frac{\Delta I(t_d)}{I(t_d)} \right\}
\]

The second sub-problem is to obtain the maximal likelihood estimator \( \hat{I} \) using the obtained values of \( \hat{\alpha} \) and \( \hat{\beta} \). They are obtained by solving Eq. (19).

\[
\hat{I} = \arg \max_I L^{[1]}(\hat{\alpha}, \hat{\beta}, \hat{I}(I)).
\]

Eq. (19) cannot be solved analytically. There are \( 10^{14} \) possible topologies for \( N = 10 \), and \( 10^{57} \) for \( N = 20 \). Simulated annealing [18] is a powerful meta-heuristic algorithm to solve such a combinatorial global optimization problem. A candidate of parameters \( I \) is generated randomly near the present value of \( I \). The parameters are updated from \( y_{ij} = \Gamma_y(I) \) to \( \Gamma_y(I') \) according to the probability \( p(s) \) in Eq. (20) in the s-th step (\( s = 0, 1, \ldots \)) of iterations.

\[
p(s) = \min \left( \exp \left( \frac{L^{[1]}(\hat{\alpha}, \hat{\beta}, \hat{I}(I)) - L^{[1]}(\hat{\alpha}, \hat{\beta}, \hat{I}(I'))}{kT(s)} \right), 1 \right).
\]

\( T(s) \) is the annealing temperature in the s-th step. A typical cooling schedule is \( T(s) = 1/ \log(s + 1) \). Since \( O(T) = 1 \), the scaling constant \( k \) is selected as an appropriate value whose order is the same as that of \( L^{[1]} \).

3.2.2. Case 2: for given \( \Delta I_d(t_d) \)

The procedure for the estimation from \( \Delta I_d(t_d) \) is presented. Again, the problem is divided to two sub-problems. The first sub-problem is to solve Eq. (21). The quantity \( I(0) \) is the initial value of the number of infectious persons, which appears in the formula for the mean and variance of \( I \) in Appendix C. It is not the same as the known \( J(t_0) \), but an unknown parameter.

\[
\hat{\alpha}, \hat{\beta}, \hat{I}(0) = \arg \max_{\alpha, \beta, I(0)} L^{[2]}(\alpha, \beta, I(0)).
\]

Simulated annealing uses the probability \( p(s) \) in Eq. (22) for the update of a candidate. An alternative means to solve Eq. (21) is such a function maximization algorithm as a BFGS quasi-Newton method [18].

\[
p(s) = \min \left( \exp \left( \frac{L^{[2]}(\alpha, \beta, I(0)) - L^{[2]}(\alpha', \beta', I(0)' \)}{kT(s)} \right), 1 \right).
\]

A great difficulty in maximizing \( L^{[1]}(\theta) \) is encountered in solving the second sub-problem. The very complex formula for \( y_{ij}^{[1]}(I) \) to obtain the value of \( L^{[1]}(\theta) \) is not tractable even numerically unless \( N \) is very small. An approximation is introduced to convert this problem to the computationally tractable second sub-problem in 3.2.1. The value of \( I_d(t_d) \) is approximately obtained from the value of \( \Delta I_d(t_d) \) by Eq. (23), which use the already obtained value of \( \hat{\alpha} \). Eq. (19) is solved with the converted values of \( I_d(t_d) \) instead of maximizing \( L^{[1]}(\theta) \) directly.

\[
I_d(t_d) \approx \frac{\Delta I_d(t_d)}{\hat{\alpha} \Delta t}.
\]

Eq. (23) is a discrete time approximation of Eq. (6) for small \( \Delta t \). This relationship holds true for the mean values of \( I_d(t_d) \) and \( \Delta I_d(t_d) \). But the variance of \( I_d(t_d) \) is overestimated by neglecting the stochastic term \( \sqrt{\alpha I_d(t_d)^2} \). Because of the
approximation, the estimation from $\Delta l_j(t_d)$ would be more erroneous than that from $l_i(t_d)$. The estimation errors are demonstrated in Section 4.

4. Experiment

4.1. Computationally synthesized dataset

A number of test datasets are synthesized by numerical integration [15] of a Langevin equation (1) for random network topologies and transmission parameters. The network is a Erdős–Rényi model in a combination of $N$ and the average nodal degree $(k_i)$. The nodal degree of a node $n_i$ is given by $k_i = \sum_{j=0}^{N-1} l_{ij}$. The probability at which $l_{ij} = 1$ is $(k_j)/(N-1)$.

It is postulated that the total number of persons who moves from $n_i$ to $n_j$ per a unit time is proportional to $\sqrt{k_i k_j}$ if a link is present. This law is known valid generally for the world-wide airline transportation network [11]. It is also postulated that the initial population $P_i(0) = S_i(0) + l_i(0) + R_i(0)$ of a node $n_i$ is proportional to the total number of persons who outgoes from the node per a unit time. Consequently, $y_{ij}$ is determined as a function of $l$ by Eq. (24). The fraction of persons who outgoes per a unit time is a constant $\gamma$ over the network. This is an additional unknown parameter in solving Eq. (19). The law in Eq. (24) is used in discovering the network topology by Eq. (19) as well as synthesizing the datasets computationally.

$$y_{ij} = r_j l_i \sqrt{k_i k_j} \gamma.$$  \(24\)

$P_i(0)$ is given by Eq. (25). The total population is set to $P = 10^6 N$ in the experiment.

$$P_i(0) = \frac{N-1}{\sum_{i=0}^{N-1} l_i \sqrt{k_i k_j}} P.$$  \(25\)

The estimation error of the basic reproductive ratio is defined by Eq. (26). It is a relative absolute deviation from the true value.

$$E_i = \frac{|\hat{\beta} - \beta|}{\beta}.$$  \(26\)

The estimation error of the topology is defined by Eq. (27). It is the fraction of links whose presence or absence is estimated wrongly.

$$E_l = \frac{\sum_{i<j} |\hat{l}_{ij} - l_{ij}|}{N(N-1)/2}.$$  \(27\)

Fig. 1 illustrates an example of the network topology estimated from a computationally synthesized dataset. The graph [A] shows the dataset $l_i(t_d)$ with $\Delta t = 1$ and $D = 100$ when $r = 2$. The drawing [B] shows the topology with $N = 10$ and $(k_i) = 3$ to synthesize the dataset. The index cases appear at $n_0$. The network includes a core sub-structure consisting of $n_0$, $n_2$, $n_4$, $n_5$, and $n_9$. It is nearly a clique where every node is connected to every other node. Links are present except for the one between $n_3$ and $n_9$. The drawing [C] shows the topology estimated from the dataset. The error is $E_l = 0.18$. The core is discovered correctly. The ability of the method is surprising in distinguishing the only pair of nodes where the link is absent. The links from the core to $n_1$ and $n_7$ are discovered. Although the method identifies that $n_3$, $n_6$, and $n_9$ do not belong to the core, but form the stubs (dead ends) from the core, it fails to estimate how they are connected to each other and the core.

The number of cases is the smallest at these nodes. The movements of infectious persons to and from them are so infrequent that the analysis on them is not so reliable.

The estimation error $E_l$ of the method in this study is compared with those of a naive estimation and a mere random guess. The naive estimation relies on the correlation between nodes. When an infectious person moves from $n_i$ to $n_j$, $l_i$ decreases and $l_j$ increases by one simultaneously. Thus, intuitively, the negative correlation of the change in $l_i$ and $l_j$ is the signal of the presence of a link. The correlation $\rho_{ij}$ between $n_i$ and $n_j$ is calculated by Eq. (28) where $\Delta l_i(t_d) = l_i(t_{d+1}) - l_i(t_d)$.

$$\rho_{ij} = \frac{D-2}{\sum_{d=0}^{D-2} \Delta l_i(t_d) - \frac{1}{N} \sum_{k=0}^{N-1} \Delta l_k(t_d)} \left( \Delta l_i(t_d) - \frac{1}{N} \sum_{k=0}^{N-1} \Delta l_k(t_d) \right).$$  \(28\)

Note that $\sum_{k=0}^N \Delta l_k(t_d)/N$ is not the average over the time sequence for $n_k$, but the average over the nodes at $t = t_d$. This formula is supposed to exclude the positive correlation because of the common growing trends in $l$ (dependent on


Fig. 1. Example of the network topology estimated from a computationally synthesized dataset. [A]: dataset $I_i(t_d)$ with $\Delta t = 1$ and $D = 100$ when the basic reproductive ratio is $r = 2$ ($\alpha = 0.067, \beta = 0.033, \gamma = 0.1$). Individual curves represent the nodes. [B]: random network topology with $N = 10$ and $\langle k \rangle = 3$ to synthesize the dataset in [A]. The index cases appear at $t_0$. At $t = t_{f0}, I_2 > I_4 > I_0 > I_3 > I_7 > I_0 > I_6 > I_4 > I_7$. [C]: network topology estimated from the dataset in [A].

$\langle \alpha - \beta \rangle I_i \Delta t$ in Eq. (11). The naive estimation predicts $l_{ij} = 1$ if $\rho_{ij} < 0$. The random guess is the worst bound of estimation. The number of links whose presence or absence is predicted wrongly obeys a binomial distribution. The mean and standard deviation of $E_i$ are 0.5 and 0.0745 for $N = 10$, and 0.5 and 0.0256 for $N = 20$ theoretically.

Fig. 2 shows $E_i$ for various values of the normalized average degree $\langle k \rangle / (N - 1)$ ($\langle k \rangle = 2, 3, 4$ for the number of nodes $N = 10$) when $r = 2$, and $l_i(t_d)$ with $\Delta t = 1$ and $D = 100$ is given as a dataset. For small $\langle k \rangle$, the naive estimation does not work at all. As $\langle k \rangle$ increases, $E_i$ of the method increases and that of the naive estimation decreases. For large $\langle k \rangle$, the naive estimation becomes less erroneous than the random guess. But, it never surpasses the method in this study. The initially heterogeneous node-to-node distribution of infectious persons relaxes more quickly in the networks having more links. For example, the standard deviation of $l_i(t_{f0})$ is about 400 for $\langle k \rangle = 2$, and about 300 for $\langle k \rangle = 4$, while the mean for both $\langle k \rangle$ is about 550 ($\sim I_0(0) \exp((\alpha - \beta)D \Delta t)/N$). The growing trends also become more homogeneous. Under such homogeneity, the negative correlation implies the movements between nodes directly. The naive estimation may be a substitute if assuming homogeneous distribution during observation are well-grounded.

Fig. 3 shows the estimation errors $E_i$ and $E_E$ for various values of the normalized average degree $\langle k \rangle / (N - 1)$ ($\langle k \rangle = 2, 3, 4$ for $N = 10$ and $\langle k \rangle = 3, 6, 9$ for $N = 20$), $N$, and $r$ when $l_i(t_d)$ with $\Delta t = 1$ and $D = 100$ is given as a dataset. The findings are as follows.

- As $\langle k \rangle$ increases, $E_i$ increases from around 0.2 to 0.4. Although $E_i$ (the average ± standard deviation) remains less than 0.5 within the range of the experimental conditions here, the estimation comes close to a mere random guess ($E_i \rightarrow 0.5$) for the dense network limit ($\rightarrow$ a complete graph). Discerning the presence or absence of links becomes more difficult as the spread goes on over more links in parallel and reaches more nodes along more possible routes. On the other hand, $E_E$ does not change largely as $\langle k \rangle$ changes.
Fig. 3. Estimation errors $E_i$ and $E_r$ for various values of the normalized average degree $\langle k \rangle/(N-1)$ ($\langle k \rangle = 2, 3, 4$ for $N = 10$ and $\langle k \rangle = 3, 6, 9$ for $N = 20$), the number of nodes $N$, and the basic reproductive ratio $r$ when $I(t_d)$ with $\Delta t = 1$ and $D = 100$ is given as a dataset. The initial condition is $I_0(0) = 200$ and $I_i(0) = 0$ for all $i \neq 0$. The individual plots show the mean and standard deviation over trials for 100 different random networks. [A]: $E_i$ for $r = 2 (\alpha = 0.067, \beta = 0.033, \gamma = 0.1)$, [B]: $E_i$ for $r = 4 (\alpha = 0.08, \beta = 0.02, \gamma = 0.1)$, [C]: $E_i$ for $r = 6 (\alpha = 0.086, \beta = 0.014, \gamma = 0.1)$, [D]: $E_r$ for $r = 2$ (the same as [A]), [E]: $E_r$ for $r = 4$ (the same as [B]), [F]: $E_r$ for $r = 6$ (the same as [C]).

- As the network becomes larger, $E_i$ increases and $E_r$ decreases. As the model becomes more complex (the number of links $O(N^2)$ becomes larger compared to the amount of data $O(ND)$), there may appear more optimal or sub-optimal topologies. Choosing a unique right answer becomes more difficult from such similar candidates. On the other hand, the central limit theorem guarantees that the fluctuation decreases as the network becomes larger because $\alpha$ and $\beta$ are estimated from the sum of $N$ probability variables ($\langle I(t_d) \rangle = \sum_{i=0}^{N-1} I_i(t_d)$).
- As $r$ increases, $E_i$ increases (but the difference between $r = 4$ and $r = 6$ is very small) and $E_r$ increases from around 0.1 to 0.35. The observations cannot track down the rapid reproduction of patients when $r \Delta t$ is large.

The dependence of the errors on $\gamma$ in Eq. (24) is investigated. The errors increase from $E_i = 0.2$ and $E_r = 0.092$ for $\gamma = 0.1$ in [a] of Fig. 3[A] to $E_i = 0.26$ and $E_r = 0.10$ for $\gamma = 0.2$, and $E_i = 0.32$ and $E_r = 0.099$ for $\gamma = 0.4$. The accuracy of estimation is limited when many persons move between nodes in both directions because of large $\gamma$. The dependence of the errors on $\Delta t$ is investigated. The errors increase from $E_i = 0.2$ and $E_r = 0.092$ in [a] of Fig. 3[A] to $E_i = 0.31$ and $E_r = 0.18$ if the observations are made four times less frequently ($\Delta t = 4, D = 25$). But they are improved only slightly to $E_i = 0.18$ and $E_r = 0.074$ if the observations are made 4 times more frequently ($\Delta t = 0.25, D = 400$). A small time interval between observations is relevant to accurate estimation. The errors are investigated for various initial population distributions $P_i(0)$. If the population is a thousandth ($P = 10^3N$), $E_i = 0.23$ and $E_r = 0.091$ for $N = 10$, $\langle k \rangle = 2$, and $r = 2$ when $I_0(0) = 20$. If $P_i(0) \propto \sum_{j=0}^{N-1} j_i(k_i k_j)^4$ rather than $\sum_{j=0}^{N-1} j_i \sqrt{k_i k_j}$ in Eq. (25), $E_i = 0.25$ and $E_r = 0.09$ when the
Fig. 4. Estimation errors $E_l$ and $E_r$ for various values of $\langle k_i \rangle / (N - 1)$, $N$, and $r$ when $\Delta_i(t_d)$ with $\Delta t = 1$ and $D = 100$ is given as a dataset. The experimental conditions are the same as those for Fig. 3.

population is $P = 10^6 N$. In this case, the population ranges in vastly diverse scales. The ratio of the population of the most populated node to that of the least populated node is $P_{\text{max}}(0)/P_{\text{min}}(0) \approx 2000$ while $P_{\text{max}}(0)/P_{\text{min}}(0) \approx 7$ in case of Eq. (25). $E_l$ is not affected by the distribution. $E_l$ increases when much less populated nodes are present. But $E_l$ still remains small.

Fig. 4 shows the estimation errors $E_l$ and $E_r$ for various values of $\langle k_i \rangle / (N - 1)$, $N$, and $r$ when $\Delta_i(t_d)$ with $\Delta t = 1$ and $D = 100$ is given as a dataset. The experimental conditions are the same as those for Fig. 3. The findings are as follows.

- The dependency of $E_l$ and $E_r$ on $\langle k_i \rangle$, $N$, and $r$ in Fig. 4 is similar to those in Fig. 3.
- The absolute value of errors tends to increase. For example, $E_l = 0.31$ in Fig. 4 is much larger than $E_l = 0.2$ in Fig. 3 under the same experimental conditions $\langle k_i \rangle = 2$, $N = 10$, and $r = 2$. In contrast, the increase in $E_r$ is relatively small. $E_r$ is nearly the same when $r = 6$. The deterioration in estimating the topology, therefore, seem to result from the influence of the approximation in Eq. (23).

As a summary of the experiments, the estimation is particularly reliable ($E_l \approx 0.2$ and $E_r \approx 0.1$) when $I(t_d)$ for a slow reproduction over a sparse network is used as an input. Such performance cannot be achieved by the naive estimation. The estimated topology from $\Delta_i(t_d)$ is more erroneous than that from $I(t_d)$ by as much as 50%.
4.2. SARS dataset

SARS is a respiratory disease in humans caused by the SARS corona-virus. The epidemic of SARS appears to have started in the Guangdong Province of south China in November 2002. SARS spread from the Guangdong Province to Hong Kong in early 2003, and eventually nearly 40 countries around the world by July. WHO archives the cumulative number of reported probable cases of SARS.\(^1\) The dataset in the archive had been updated nearly every day since March 17. It is a time sequence dataset \(J_i(t_d)\) with \(\Delta t = 1\) day. In this study, the target geographical regions are those where five or more cases had been reported in a month since March 17. They include Canada (CAN), France (FRA), United Kingdom (GBR), Germany (GER), Hong Kong (HKG), Malaysia (MAS), Taiwan (ROC), Singapore (SIN), Thailand (THA), United States (USA), and Vietnam (VIE). Mainland China is not included because no data is available in some periods and no data outside of Guangdong Province is reported in other periods.

Fig. 5 shows the date when the first patient appeared and the propagating wavefront of the spread. It is almost certain that neither FRA nor MAS are the origin of the outbreak. But, nobody can tell the chain of transmission among CAN, HKG, SIN, GBR, ROC, and USA in just two days from March 17 to 19 reliably. The wavefront is not as informative as anticipated. Such an naive gleaning verifies the obvious series of events at best.

The estimated transmission parameters are \(\hat{\alpha} = 0.18\) and \(\hat{\beta} = 0.13\). The basic reproductive ratio is \(\hat{r} = 1.4\). According to the field-based medical case studies, the basic reproductive ratio (except for super-spreading events \(^{19}\)) was \(r = 2.7\) in February and went down to \(r = 1\) in late March in Hong Kong \(^{20}\), and \(r = 7\) in February and \(r = 1\) in early March in Singapore \(^{13}\). The decrease of \(r\) is due to the quarantine, hospitalization and public awareness starting to take effects after WHO issued a world-wide alert on March 12. But the spread of SARS was still going on world-wide. The value slightly greater than 1 seems reasonable at that stage.

It is postulated that the law in Eq. (24) holds true in analyzing the SARS dataset. The topology which achieves the largest value of the logarithmic likelihood function among many trials is chosen. This is efficient in rejecting the local maximum to which simulated annealing may converge. The trials use different random number sequences to generate nearby candidates \(l'\) in Eq. (20). Fig. 6 shows the estimated topologies \(\hat{l}\). The topology [A] is the most likely (the largest value of the likelihood \(L = -9985\)). The best 30 trials out of 300 trials converge to [A]. It includes 11 links \((\langle k_i \rangle / (N - 1) = 0.2)\). The topology [B] is the second most likely \((L = -9998)\). The next 5 trials converge to [B]. It include 13 links \((\langle k_i \rangle / (N - 1) = 0.24)\). The topology [C] is the third most likely \((L = -10012)\). The next 29 trials converge to [C]. It include 14 links \((\langle k_i \rangle / (N - 1) = 0.25)\). About 20% of the trials converge to either [A] or [C].

The sub-structures common in all of [A], [B], and [C] are a star from HKG to CAN, ROC, and SIN, another star from USA to GBR, MAS, and VIE, and a link between the centers of these stars (HKG and USA). A triangle between USA, VIE, and THI appears in [A] and [C]. The likelihood seems sensitive to the topological whereabouts of GER and FRA given these common sub-structures as a core of the network. This may happen to make [B] a tall but narrow peak in the landscape of the likelihood, which simulated annealing sometimes fail to discover. In addition to these sub-structures, a few remarkable points are seen in these topologies. The nodes SIN and ROC are stubs where \(k_i = 1\). The role of SIN is not so relevant in spreading the disease despite the fact that the number of cases was more than 100 in the middle of April. The nodes CAN and USA have links to distant geographical regions, and USA is a hub \((k_i = \text{the largest})\). They are relevant intermediate spreader nodes. The links around GER are not stable among the three topologies. The number of cases in some European countries is too small to draw a reliable conclusion.

The estimated topologies are not meant to reproduce the trajectories of individual patients’ movement, but rather demonstrate some demographical interactions within the macroscopic world-wide transportation behind the SARS outbreak. Nevertheless, the sub-structures mentioned above seem to be consistent with the following publicly known series of events on some individual patients’ microscopic movements.

- Two of the index patients in Toronto in Canada, three of the index patients in Singapore, and another three of the index patients in the United States stayed in a hotel in Hong Kong where a Chinese nephrologist, who had treated many patients

\(^1\) World Health Organization, cumulative number of reported probable cases of SARS, http://www.who.int/csr/sars/country/en/index.html (2003).
Fig. 6. Estimated topologies from the WHO dataset on the cases of SARS in Canada (CAN), France (FRA), United Kingdom (GBR), Germany (GER), Hong Kong (HKG), Malaysia (MAS), Taiwan (ROC), Singapore (SIN), Thailand (THI), United States (USA), and Vietnam (VIE) from March 17 through April 17. [A]: the most likely topology. [B]: the second most likely topology. [C]: the third most likely topology.

in Guangzhou and became infected, was staying in late February.\(^2\) This event implies the links from HKG to CAN, SIN, and USA forms a chain of transmission in the early growth phase of the SARS outbreak.

- A garment manufacturer from the United States became infected during their stay in Hong Kong on the way to Hanoi in Vietnam, showed symptoms there, and was evacuated to a hospital in Hong Kong\(^2\). An Italian physician, who treated him at a hospital in Hanoi, showed symptoms in Bangkok in Thailand where he would attend a conference in early March. These events imply that the interactions among HKG, USA, VIE, and THI are present potentially, which could result in another chain of transmission.

The WHO dataset is not of perfectly reliable quality. Particularly, the data on mainland China is of poor quality, and cannot be used in this study. Even the individual number of cases which was reported from the other local governments may not be accurate. The number of cases is highly fluctuating and seems noisy. Data in a city-level resolution, rather than nation-level, would be necessary for accurate estimation when large countries like USA play an important role as a spreader. It is surprising that, in spite of these limitations, the method reproduces some characteristics of the network over which SARS spread from Hong Kong to Southeast Asia and North America.

5. Conclusion

The method presented in this study solves an inverse problem to discover the effectively decisive network and reveal the transmission parameters from the observation \(I_i(t_d)\) or \(\Delta J_i(t_d)\) on the spread of an infectious disease. The findings with test datasets are that the estimation is particularly successful when the topology is sparse and reproduction is slow, and that the estimation from \(\Delta J_i(t_d)\) is more erroneous than that from \(I_i(t_d)\). The network topology discovered from a seemingly noisy dataset on the SARS outbreak reproduces some characteristic patterns of the spread from Hong Kong to Southeast Asia and North America. So far, there has been a great deal of effort in obtaining a complete picture of how an infectious disease did and will spread from the found pieces of an epidemiological jigsaw puzzle. The method presents new pieces from a viewpoint of macroscopic transportation. These pieces can be put together with the pieces found in the conventional field-based medical case studies on the individual patients’ microscopic movements.

The method can be extended to apply to a more practical situation. The experimental condition in this study is an extreme where nothing but \(I_i(t_d)\) or \(\Delta J_i(t_d)\) is known and no informative prior knowledge is available. If some demographical statistics on the traffic between cities or the findings on the past contacts between individual patients are available, the consequent posterior distribution enables more comprehensive Bayesian inference. Another extension is to employ more complicated but realistic epidemiological compartment models. The latent period (infected but not infectious) and hospitalization are relevant for some diseases. The dependence of recovery on time \((\beta \neq \text{constant})\) is realistic for other diseases. Strictly speaking, the time interval from infection to recovery and from a movement to another obeys appropriate probability density functions. Analytical treatment of the stochastic process with these effects tends to be considerably difficult. The estimation may count on such a numerical method as a Markov-chain Monte-Carlo sampling.

In addition to the extensions to the method, it is sometimes critical to gather such a larger dataset as a collection of multiple independent time sequences starting from different index cases. Less erroneous estimation may be possible even for the regions whose population is small or where the number of cases is small. Such a dataset is not available for SARS, but possibly for influenza which spreads around the world in seasonal epidemics. Understanding the landscape of the likelihood functions is essential in identifying the requisites for a dataset, a network topology, and a mathematical model of disease transmission to make the inverse problem well-posed and stabilize the solution. This remains the challenge for the future.

---

2 SARS Expert Committee (Hong Kong), SARS in Hong Kong: from experience to action, http://www.sars-expertcom.gov.hk/english/reports/reports/reports_fullrpt.html (2003).
Appendix A. Probability density function

A generic form of a Langevin equation for multiple time-dependent variables \(x_i(t)\) is given by Eq. (29). The fluctuations \(\xi_i(t)\) are stochastic terms.

\[
\frac{dx_i(t)}{dt} = \mu_i(x_0(t), \ldots, x_{N-1}(t)) + \sum_{a=0}^{M-1} \sigma_{ia}(x_0(t), \ldots, x_{N-1}(t))\xi_a(t).
\] (29)

Eq. (29) can be solved by deriving the probability density function \(p(x, t)\) for probability variables \(x = (x_0, \ldots, x_{N-1})\) at time \(t\). The time evolution of \(p(x, t)\) is given by the Fokker–Planck equation in Eq. (30).

\[
\frac{\partial p(x, t)}{\partial t} = -\sum_{i=0}^{N-1} \frac{\partial}{\partial x_i} A_i(x)p(x, t) + \frac{1}{2} \sum_{i,j=0}^{N-1} \frac{\partial^2}{\partial x_i \partial x_j} B_{ij}(x)p(x, t).
\] (30)

The coefficients \(A_i\) and \(B_{ij}\) are given by Eqs. (31) and (32).

\[
A_i(x) = \mu_i(x) \tag{31}
\]

\[
B_{ij}(x) = \sum_{a=0}^{M-1} \sigma_{ia}(x)\sigma_{ja}(x). \tag{32}
\]

The mean (the first order moment) of \(x_i\) at \(t\) is given by \(m_i(t) = \langle x_i \rangle_t = \int xp(x, t)dx\). The time evolution of \(m_i(t)\) is given by Eq. (33). It is derived by multiplying Eq. (30) by \(x\) and partial integration under the condition where \(p\) and \(\partial p/\partial x\) decay more rapidly than \(A_i\) and \(B_{ij}\) near the boundary of the domain of \(x\).

\[
\frac{dm_i(t)}{dt} = \langle A_i(x) \rangle_t. \tag{33}
\]

The covariance (the second order moment) between \(x_i\) and \(x_j\) at \(t\) is given by \(v_{ij}(t) = \langle x_i x_j \rangle_t - m_i(t)m_j(t)\). The time evolution of \(v_{ij}(t)\) is given by Eq. (34). Derivation is similar to that for Eq. (33).

\[
\frac{dv_{ij}(t)}{dt} = \langle B_{ij}(x) \rangle_t + \langle x_i A_j(x) \rangle_t + \langle A_i(x) x_j \rangle_t. \tag{34}
\]

Higher order moments can be obtained recursively as a solution of the differential equations which include the calculated lower order moments.

Appendix B. Moments of \(I_i\) and \(J_j\)

Eq. (35) through (39) are the differential equations for the time evolution of the first and second order moments of \(I_i\) and \(J_j\). The symbols \(m^{[I]}(t|\theta)\), \(m^{[J]}(t|\theta)\) are the row vectors whose \(i\)-th element is the mean of \(I_i\), \(I_j\), and \(v^{[I]}(t|\theta), v^{[J]}(t|\theta)\) are the \(N \times N\) matrices whose \(i\)-th row and \(j\)-th column element is the covariance between \(I_i\) and \(I_j\). \(I_i\) and \(J_j\) are the unknown network topology and transmission parameters represented by a symbol \(\theta = \{\gamma, \alpha, \beta\}\).

\[
\frac{dm^{[I]}(t|\theta)}{dt} = m^{[I]}(t|\theta)a^T \tag{35}
\]

\[
\frac{dm^{[J]}(t|\theta)}{dt} = \alpha m^{[I]}(t|\theta) \tag{36}
\]

\[
\frac{dv^{[I]}(t|\theta)}{dt} = av^{[I]}(t|\theta) + v^{[I]}(t|\theta)a^T + \langle B_i \rangle \tag{37}
\]

\[
\frac{dv^{[J]}(t|\theta)}{dt} = av^{[J]}(t|\theta) + \alpha(v^{[I]}(t|\theta) + c(t)) \tag{38}
\]

\[
\frac{dv^{[I]}(t|\theta)}{dt} = \alpha(v^{[J]}(t|\theta) + v^{[I]}(t|\theta)^T + c(t)). \tag{39}
\]
Definitions of the $N \times N$ matrices $a$, $B$, and $c$ which appear in Eq. (35) through (39) are given by Eq. (40) through (42).

$$a_{ij} = \left( \alpha - \beta - \sum_{k=0}^{n-1} \gamma_k \right) \delta_{ij} + \gamma_{ji}$$  \hspace{1cm} (40)

$$B_{ij} = \left\{ \left( \alpha + \beta + \sum_{k=0}^{N-1} \gamma_k \right) J_i + \sum_{k=0}^{N-1} \gamma_k l_k \right\} \delta_{ij} - \gamma_{ij} l_i - \gamma_{ji} l_j$$  \hspace{1cm} (41)

$$c_{ij}(t) = \delta_{ij} m^{ij}(t|\theta).$$  \hspace{1cm} (42)

Eq. (43) through (47) are the solutions. $E$ is a unit matrix.

$$m^{[ii]}(t|\theta) = I(0) \exp(a^T t)$$  \hspace{1cm} (43)

$$m^{[ij]}(t|\theta) = I(0) \{ \alpha(a^T)^{-1} \exp(a^T t) - \alpha(a^T)^{-1} + E \}$$  \hspace{1cm} (44)

$$v^{[ii]}(t|\theta) = \int_0^t \exp(a(t-t'))(B) \exp(a^T(t-t')) dt'$$  \hspace{1cm} (45)

$$v^{[ij]}(t|\theta) = \int_0^t \alpha \exp(a(t-t'))(v^{[ii]}(t'|\theta) + c(t')) dt'$$  \hspace{1cm} (46)

$$v^{[ii]}(t|\theta) = \int_0^t \alpha(v^{[ij]}(t'|\theta) + v^{[ij]}(t'|\theta)^T + c(t')) dt'$$

$$= \int_0^t \alpha^2 \int_0^t \exp(a(t'-t''))(v^{[ii]}(t''|\theta) + c(t'')) dt'' \exp(a^T(t'-t'')) dt' + \alpha c(t') dt'.$$  \hspace{1cm} (47)

### Appendix C. Moments of $I$ and $J$

Eq. (48) through (52) are the differential equations for the time evolution of the first and second order moments of $I$ and $J$. The symbols $m^{[ii]}(t|\theta), m^{[ij]}(t|\theta)$ are the mean of $I,J$, and $v^{[ii]}(t|\theta), v^{[ij]}(t|\theta), v^{[ij]}(t|\theta)$ are the variance of $I$, covariance between $I$ and $J$, variance of $J$.

$$\frac{dm^{[ii]}(t|\theta)}{dt} = (\alpha - \beta)m^{[ii]}(t|\theta)$$  \hspace{1cm} (48)

$$\frac{dm^{[ij]}(t|\theta)}{dt} = \alpha m^{[ij]}(t|\theta)$$  \hspace{1cm} (49)

$$\frac{dv^{[ii]}(t|\theta)}{dt} = 2(\alpha - \beta)v^{[ii]}(t|\theta) + (\alpha + \beta)m^{[ij]}(t|\theta)$$  \hspace{1cm} (50)

$$\frac{dv^{[ij]}(t|\theta)}{dt} = (\alpha - \beta)v^{[ij]}(t|\theta) + \alpha(v^{[ii]}(t|\theta) + m^{[ij]}(t|\theta))$$  \hspace{1cm} (51)

$$\frac{dv^{[ij]}(t|\theta)}{dt} = \alpha(2v^{[ii]}(t|\theta) + m^{[ij]}(t|\theta)).$$  \hspace{1cm} (52)

Eq. (53) through (57) are the solutions.

$$m^{[ii]}(t|\theta) = I(0) \exp(\alpha - \beta) t$$  \hspace{1cm} (53)

$$m^{[ij]}(t|\theta) = I(0) \left( \frac{\alpha}{\alpha - \beta} \exp(\alpha - \beta) t - \frac{\beta}{\alpha - \beta} \right)$$  \hspace{1cm} (54)

$$v^{[ii]}(t|\theta) = I(0) \frac{\alpha + \beta}{\alpha - \beta} \exp(2(\alpha - \beta) t - \exp(\alpha - \beta) t)$$  \hspace{1cm} (55)

$$v^{[ij]}(t|\theta) = I(0) \left\{ \frac{\alpha(\alpha + \beta)}{(\alpha - \beta)^2} \exp(2(\alpha - \beta) t - \frac{2\alpha \beta}{(\alpha - \beta)^2} (\alpha - \beta) t) \exp(\alpha - \beta) t \right\}$$  \hspace{1cm} (56)

$$v^{[ij]}(t|\theta) = I(0) \left[ \frac{\alpha^2(\alpha + \beta)}{(\alpha - \beta)^3} \exp(2(\alpha - \beta) t - \frac{4\alpha^2 \beta}{(\alpha - \beta)^2} (\alpha - \beta) t) \exp(\alpha - \beta) t - \frac{\alpha \beta (\alpha + \beta)}{(\alpha - \beta)^3} \right].$$  \hspace{1cm} (57)
References

[1] D.W. Walker, D. Allingham, H.W.J. Lee, M. Small, Parameter inference in small world network disease models with approximate Bayesian computational methods, Physica A 389 (2010) 540–548.
[2] M. Small, C.K. Tse, D.M. Walker, Super-spreaders and the rate of transmission of the SARS virus, Physica D 215 (2006) 146–158.
[3] M.J. Keeling, S.P. Brooks, C.A. Gilligan, Using conservation of pattern to estimate spatial parameters from a single snapshot, Proceedings of the National Academy of Sciences USA 101 (2004) 9155–9160.
[4] C.E. Dangerfield, J.V. Ross, M.J. Keeling, Integrating stochasticity and network structure into an epidemic model, Journal of the Royal Society Interface (2009) doi:10.1098/rsif.2008.0410.
[5] S. Riley, Large-scale spatial-transmission models of infectious disease, Science 316 (2007) 1298–1301.
[6] Y. Maeno, Node discovery problem for a social network, Connections 29 (2009) 62–76.
[7] M.G. Rabhat, M.A.T. Figueiredo, R.D. Nowak, Network inference from co-occurrences, IEEE Transactions on Information Theory 54 (2008) 4053–4068.
[8] A. Baroncelli, M. Catanzaro, R. Pastor-Satorras, Bosonic reaction–diffusion processes on scale-free networks, Physical Review E 78 (2008) 016111.
[9] M. Simões, M.M.T. da Gama, A. Nunes, Stochastic fluctuations in epidemics on networks, Journal of the Royal Society Interface 5 (2008) 555–566.
[10] V. Colizza, A. Vespignani, Invasion threshold in heterogeneous meta-population networks, Physical Review Letters 99 (2007) 148701.
[11] A. Barrat, M. Barthélemy, R. Pastor–Satorras, A. Vespignani, The architecture of complex weighted networks, Proceedings of the National Academy of Sciences USA 101 (2004) 3747–3752.
[12] M.J. Keeling, J.V. Ross, On methods for studying stochastic disease dynamics, Journal of the Royal Society Interface 5 (2008) 171–181.
[13] M. Lipsitch, et al., Transmission dynamics and control of severe acute respiratory syndrome, Science 300 (2003) 1966–1970.
[14] L. Hufnagel, D. Brockmann, T. Geisel, Forecast and control of epidemics in a globalized world, Proceedings of the National Academy of Sciences USA 101 (2004) 15124–15129.
[15] P.E. Kloeden, E. Platen, Numerical Solution of Stochastic Differential Equations, Springer, 1992.
[16] V. Colizza, A. Barret, M. Barthélemy, A. Vespignani, The role of the airline transportation network in the prediction and predictability of global epidemics, Proceedings of the National Academy of Sciences USA 103 (2006) 2015–2020.
[17] N.G. van Kampen, Stochastic Processes in Physics and Chemistry, Elsevier, 2007.
[18] W.H. Press, S.A. Teukolsky, W.T. Vetterling, B.P. Flannery, Numerical Recipes: The Art of Scientific Computing, Cambridge University Press, 2007.
[19] R. Fujie, T. Odagaki, Effects of superspreaders in spread of epidemic, Physica A 374 (2007) 843–852.
[20] S. Riley, et al., Transmission dynamics of the etiological agent of SARS in Hong Kong: impact of public health interventions, Science 300 (2003) 1961–1966.
[21] K.T. Greenfield, China Syndrome, HarperCollins Publishers, 2006.