2004 International Symposium on Nonlinear
Theory and its Applications (NOLTA2004)
Fukuoka, Japan, Nov. 29 - Dec. 3, 2004

Modeling the SARS Outbreak in Hong Kong with Small World or Scale Free Networks

Michael Small† and Chi Kong Tse‡

Department of Electronic and Information Engineering, Hong Kong Polytechnic University, Hong Kong
Email: †ensmall@polyu.edu.hk, ‡cktse@ieee.org

Abstract—The SARS outbreak in Hong Kong during 2003 exhibited several "super-spreader events" (SSEs) which cannot be modeled well with standard homogeneous SIR or SEIR type models. We propose an alternative model structure using either small-world or scale-free structure of inter-node connections and disease transmission only along the network links. Such structure naturally models the SSEs and provides simulations quantitatively similar to the true dynamics.

1. Introduction

Two characteristic features have been observed during the SARS outbreak in Hong Kong in 2003 [1, 2]: (i) so-called super-spread events (SSEs), in which a single individual initiates a large number of cases; (ii) persistent transmission within the community. Two widely cited SSEs were observed early in the epidemic and have been the subject of much attention: at the Amoy Gardens housing estate and at the Prince of Wales hospital. Moreover, epidemiological studies [3, 4] have found that in Hong Kong, (i) the mean incubation period was 6.4 days (range 2 to 10) [5]; (ii) the duration between onset of symptoms and hospitalization was 3 to 5 days; (iii) the mean number of individuals infected by each case during the initial phase of the epidemic (excluding SSEs) was 2.7 [2].

Standard deterministic SIR (susceptible-infected-removed) models of the spread of infectious diseases [6] make several serious assumptions. Recently, both small-world (SW) and scale-free (SF) networks have been observed in many areas of natural and physical science, including social relationships [7, 8]. In such areas, this new model structure has unveiled a rich range of behaviors. We apply these methods to the modeling of the spread of SARS in Hong Kong [9, 10], transmission is only allowed to occur along a limited number of direct links between individuals. By doing this, we will avoid one of the most flawed assumptions of standard Susceptible-Infected-Removed (SIR) models: a homogeneous fully connected populous. The SIR model assumes that all individuals are susceptible to the disease and all suffer an equal, small positive probability of contracting the virus. This homogeneous model leads to a continuous and smooth inter-day distribution of infections. Irregularities about this are usually attributed to random variation and non-stationarity in the model parameters.

In the next section we describe our model and study its behavior. In the subsequent section we provide some numerical simulations and summarize our results.

2. The Model

In the following subsections we define our model structure (Section 2.1) and derive some analytic results concerning the likelihood of a widespread outbreak (Section 2.2).

2.1. Model Topology

Our aim is to accurately mimic the qualitative features of the SARS epidemic with the simplest (fewest parameters) model. As in [10], we propose four distinct states. Individuals can be susceptible (S), prone (P), infected (I), or removed (R). *Susceptible individuals* are those that are capable of being infected, *prone individuals* are infected but not infectious, *infected individuals* are infected and infectious and, finally, *removed individuals* are those that are no longer either infected or capable of being infected.

Infected individuals can cause susceptible individuals, to whom they are linked, to become prone with some probability (p_1 or p_2). By *infection* we mean the transition from the susceptible to prone state. Infected individuals can cause their immediate neighbors to become infected with probability p_1, long range links cause infection with probability p_2. Prone individuals become infected with probability r_0 and finally, infected individuals become removed with probability r_1.

Just as in the SIR model we do not distinguish fatalities from recoveries: in either case the individuals are assumed to have acquired immunity.

In our model we explicitly model the geographical structure of the population. We include both "local" and "non-local" links. Because of common transmission of SARS within specific housing estates and districts in Hong Kong, and the (both real and perceived) risk of transmission at places of employment (primarily hospitals and schools) or other public areas, we model these two types of transmission separately. The geographical arrangement of nodes represents the residence of each individual. So, by "local" transmission, we mean only transmission within a family unit (i.e. residents of a single flat), or between adjacent flats.

Hence "non-local" transmission refers to transmission between non-family members due to the mixing of individuals in public spaces. In the context of the SARS outbreak in Hong Kong, this would include transmission within hospitals, schools and public spaces. Under our model, we expect SSE to occur through a single node with a large number of non-local connections.

The population of N nodes are arranged in a regular grid, of side length L ($L^2 = N$) and each node is connected directly to n_1 immediate neighbors. An infected individual will infect each of its n_1 neighbors (provided they are still susceptible) with probability p_1. Furthermore, each node has n_2 non-local (i.e. long distance) links. These are links to nodes that are geographically remote from one another, infection occurs along these pathways with probability p_2. For each node i the number $n_2^{(i)}$ is fixed and so are the links to it's $n_2^{(i)}$ remote neighbors.

To achieve a small-world model structure, the number $n_2^{(i)}$ is chosen to be proportional to a decaying exponential $f_X(x) \propto e^{-\frac{x}{\mu}}$ with parameter μ proportional to the expected (average) number of links to remote nodes. For scale free structure the distribution of links is required to follow the fatter-tailed, power law distribution.

It is the inclusion of non-local links with a random number of links that can give rise to the network's SW (and, in other cases, not considered here, SF) structure. In this paper we assign an exponentially decaying probability distribution to any number of links, and (for uni-directional links) this is sufficient to generate the necessary SW properties. A SF network requires a power law distribution of the number of links, which can consequently lead to more nodes with many more links.

It is worth considering that for the model we present here, the links between nodes are *uni*-directional. That is, infection only spreads in one direction. Clearly, the true network of social interaction consists of *bi*-directional links. But for the purposes of simulating disease transmission, unidirectional links are sufficient. The consequence of this is that it becomes easier to generate the small-world (and elsewhere the scale-free) network.

Finally, for each simulation we seed the model with one initial infection.

2.2. Behavior

The epidemic will be contained if the rate of infection is lower than the rate of removal. Intuitively, provided $(n_1 p_1 + \mu p_2) \gg r_1$ one would expect the disease to become endemic, conversely, if $(n_1 p_1 + \mu p_2) \ll r_1$ the disease will be contained. In what follows we study this condition more precisely.

Moreover, with this model we can analytically compute the probability of an outbreak being self-terminating. For a single infectious node the probability of no further infections on *a given day* is given by

$$P_{\mathrm{no1}} = \frac{(1-p_1)^{n_1}(e^{\frac{1}{\mu}} - 1)}{e^{\frac{1}{\mu}} - 1 + p_2}. \tag{1}$$

Hence the probability of *no further infections* from this node is given by

$$P_{\mathrm{none}} = \frac{P_{\mathrm{no1}} r_1}{1 - P_{\mathrm{no1}}(1 - r_1)} \tag{2}$$

provided $|P_{\mathrm{no1}}(1-r_1)| < 1$. Upon substitution of equation (1) into (2) we find that

$$P_{\mathrm{none}} = \frac{r_1(1-p_1)^{n_1}}{1 - (1-r_1)(1-p_1)^{n_1} + p_2 / \left[e^{\frac{1}{\mu}} - 1\right]}. \tag{3}$$

Equation (3) is the probability of no infections from a given individual and is therefore a weak lower bound on the probability of no general outbreak.

Now, let us denote the probability of no further infections occurring given that there are k infectious nodes by

$$P^k = P_{\mathrm{none}}^k$$

where for notational convenience we will drop the subscript on P_{none}. Treating infections as discrete events (i.e. they occur one at a time), we have that $(1-P^k)$ is the probability of at least one further infection from k infectious nodes. The probability that the epidemic will terminate is

$$P_{\mathrm{safe}} = \sum_{m=0}^{\infty} P^{m+1} \prod_{n=1}^{m} (1 - P^n), \tag{4}$$

where $P = P_{\mathrm{none}}$ is given by equation (3) [11].

Because of the assumption that infections occur individually, and sequentially, the derivation leading to equation (4) is only an approximation. The exact probability of no general outbreak can be obtained from using a branching process method [12]. However, the numerical distinction does not appear to be very large [11].

Although equation (4) can be easily computed, it is not in a form which is immediately amenable for further analysis. However, since $P_{\mathrm{safe}} \geq P_{\mathrm{none}}$ it is clear that $\mu \left[1 - (1-p_2)e^{-\frac{1}{\mu}}\right] \gg 1$ will make $P_{\mathrm{safe}} \approx 0$. Hence, either $\mu \gg 1$ or $p_2 \approx 1$ will lead to widespread infection (as expected). Differentiating (4) with respect to $(1-p_1)^{n_1}$ it is easy to verify that P_{safe} is a monotonic function of both p_1 and n_1. One can therefore observe that $P_{\mathrm{safe}} \approx 0$ if $p_1 \approx 1$ or $n_1 \gg 1$.

To take this analysis further, we now consider the rate of transmission. Let $P(t)$, $I(t)$, and $R(t)$ be the number of prone, infected and removed individuals at time t (in days). Suppose that the number of susceptible individuals $S(t) \gg R(t) + I(t) + P(t)\ \forall t$. Then, assuming that the population is seeded with a single infectious individual, the

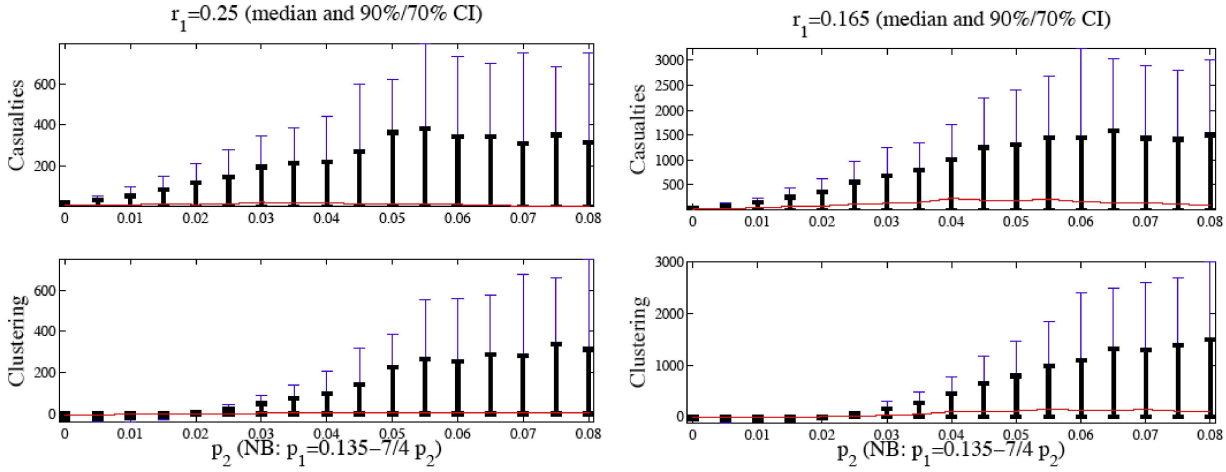

Figure 1: Unconstrained growth of the infectious population. The upper panels show the number of individuals infected after 50 days; the lower plots show the number of distinct clusters detected after the same time, with parameters set at $p_1 = 0.135 - \frac{7}{4}p_2$, $r_0 = 0.1$, $n_1 = 4$ and $\mu = 7$. The left hand plots are for $r_1 = 0.25$ (i.e. no nosocomial transmission) and the right panels are for $r_1 = 0.165$ (a mean infection period of 6 days). The results are median, 70% and 90% confidence intervals from 1000 simulations.

state of the epidemic after t days is given by

$$\begin{bmatrix} R(t) \\ I(t) \\ P(t) \end{bmatrix} = \begin{bmatrix} 1 & r_1 & 0 \\ 0 & (1-r_1) & r_0 \\ 0 & n_k & (1-r_0) \end{bmatrix}^t \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = PD^tP^{-1}[0\ 1\ 0]^T. \quad (5)$$

The corresponding eigenvalues are given by

$$\lambda_1 = 1,$$
$$\lambda_{2,3} = 1 - \frac{r_0 + r_1}{2} \pm \sqrt{\frac{1}{4}(r_0 - r_1)^2 + n_k r_0},$$

and it follows that the system has a marginally stable focus (i.e. the epidemic will terminate) if $|\lambda_{2,3}| < 1$, i.e.

$$n_k < r_1, \quad (6)$$
$$n_k r_0 < (2 - r_0)(2 - r_1). \quad (7)$$

Therefore the epidemic is controllable provided $n_k = n_1 p_1 k + \mu p_2 < r_1$. The left hand side of this inequality is the rate of infection and the left hand side is the rate of removal, as expected. In fact, this results is exactly analogous to the equivalent result for the continuous SIR model. Moreover,

$$\max_{i=1,2,3} |\lambda_i| = 1 - \frac{r_0 + r_1}{2} - \sqrt{\frac{1}{4}(r_0 - r_1)^2 + n_k r_0}. \quad (8)$$

Computationally, we can see that as r_0 or n_k increases, the rate of growth of the epidemic also increases. Conversely, as r_1 increases the rate of growth decreases. This is as one would expect as increasing r_1 will decrease the number of infectious individuals while increasing either r_0 and n_k increases this quantity.

3. Computation

In the following subsections we confirm the preceding relationships and numerically explore the behavior of our models under a variety of conditions. Following [9, 10], we take: $L = 2700$; $r_0 = \frac{1}{7.4}$; $r_1 = \frac{1}{4}$; $n_1 = 4$; $\mu = 7$; and, $p_1 = 0.135 - \frac{7}{4}p_2$. Note that because we have the possibility of P to I transition after zero days $r_0 = \frac{1}{7.4}$ rather than $\frac{1}{6.4}$. This does not have a significant effect on our results, and is merely a computational convenience.

3.1. Epidemic Growth

Now from equation (8) we can deduce that the rate of growth is significantly less than exhibited in the data, or rates of infection significantly greater [11]. Even for reasonable variation of d and the average number of secondary infections, we obtain similar results. Hence, we conclude that the assumption of no nosocomial transmission is inconsistent with the observed data. Increasing the average infectious time to 6 days gives a substantially higher rate of infection: consistent with the observed data. Moreover this observation is confirmed computationally in Fig. 1

From Fig. 1, we see that only with $r_1 \geq 0.165$ do we obtain results for which the true data is not statistically atypical. Moreover, this result is robust to moderate changes of the other relevant parameters.

3.2. Simulations

Finally, we provide simulations of the Hong Kong epidemic and demonstrate results consistent with the observed data. We initiate the model with a single infected individual and a relatively low removal rate r_1. Figure 2 depicts our results.

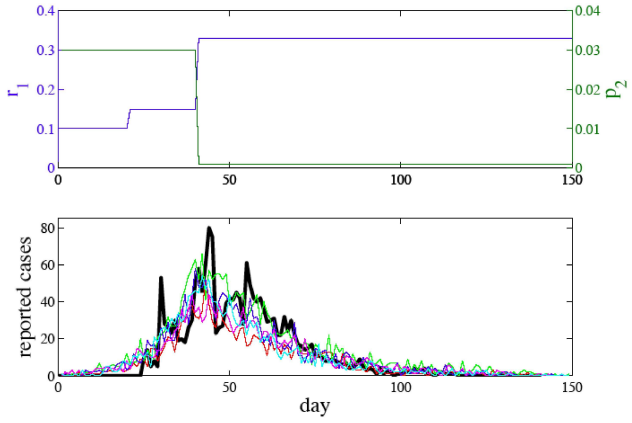

Figure 2: Model simulations. The top panel shows the change in parameters r_1 and p_2 with time (all other parameters are constant: $p_1 = 0.08$, $n_1 = 4$ and $\mu = 7$). The bottom plot shows five model simulations and the true SARS data for Hong Kong. The five model simulations were selected to ensure that a "full" outbreak occurred (a total number of infections greater than 1000). The true data is plotted as a heavy solid line.

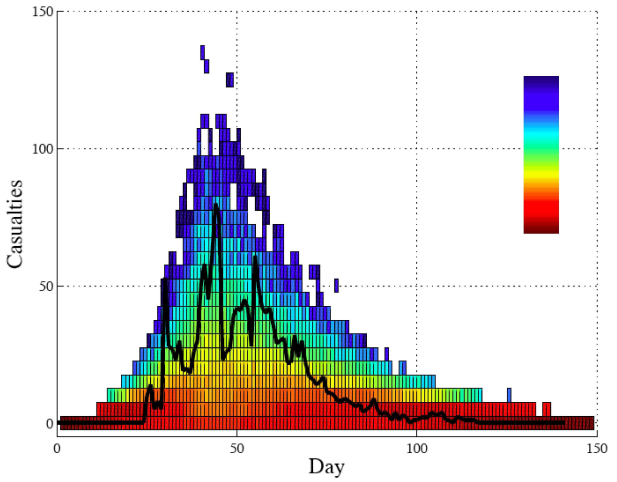

Figure 3: Probability distribution of infection dynamics. The probability distribution of the daily number of infections for 1000 simulations of the model in Fig. 2 are shown on logarithmic scale. Blue represents low probability, while red represents high probability of a particular infection tally for any number of days after onset.

We can see from Fig. 2 that many of the features of the true data are reproduced well in the simulations. However, two important aspects of the simulations are not sufficiently similar to the simulations. Firstly, the initial spreading of the disease is exponential rather than the single SSE observed in the real data. This can be overcome by simply altering the distribution of non-local links. Secondly, the magnitude of the SSEs in the simulations is somewhat smaller than the largest SSEs in the data. The initial SSE in the data cannot be modeled well by our simulations, except, by chance. Therefore, to achieve similar initial events we would expect that we would have to execute many simulations (and choose only those which suit our purpose), or simply build the SSE into the model. Neither of these approaches are desirable. We prefer the simpler model structure shown in Fig. 2.

Figure 3 shows the probability distribution for the daily number of infections. We found that the probability of infecting fewer than 20 people was approximately 0.18 while the probability of infecting more than 1000 was 0.27. One can see that, with respect to these gross statistics, the true situation for Hong Kong (1755 casualties) is quite typical.

From these simulations we cane therefore conclude that with effective control measures in place the likelihood of a significant outbreak is low.

Acknowledgments

This work was supported by a Hong Kong Polytechnic University Research Grant (no. A-PE46) and a Hong Kong Research Grants Council Competitive Earmarked Research Grant (no. PolyU 5235/03E).

References

[1] Severe Acute Respiratory Syndrome Expert Committee, "Report of the severe acute respiratory syndrome expert committee," Technical report, Hong Kong Department of Health, 2 October 2003.

[2] S. Riley *et al.*, "Transmission dynamics of the etiological agent of SARS in Hong Kong: impact of public health interventions," *Science*, 300:1961–1966, 2003.

[3] T.-W. Wong *et al.*, "Cluster of SARS among medical students exposed to single patient in Hong Kong," *Emerging Infectious Diseases*, 10, 2004.

[4] M. Chan-Yeung and R.-H. Xu, "SARS: epidemiology," *Respirology*, 8:S9–S14, 2003.

[5] C.A. Donnelly *et al.*, "Epidemiological determinants of spread of causal agent of severe acute respiratory syndrome in Hong Kong," *Lancet*, May 7 2003.

[6] J.D. Murray, *Mathematical Biology*, Biomathematics Texts, vol. 19, 2nd edition, Springer, 1993.

[7] D.J. Watts, *Six Degrees: The Science of a Connected Age*, W.W. Norton & Company, 2003.

[8] D.J. Watts and S.H. Strogatz, "Collective dynamics of 'small-world' networks," *Nature*, 393:440–442, 1998.

[9] M. Small and C.K. Tse, "Small world and scale free model for transmission of SARS," *Int. J. Bifurc. Chaos*, to appear.

[10] M. Small and C.K. Tse, "Plausible models for propagation of the SARS virus," *IEICE Trans. Fund. Electron. Comm. and Comp. Sci.*, E87-A, 2004.

[11] M. Small and C.K. Tse, "Clustering model for transmission of the SARS virus: Application to epidemic control and risk assessment," to appear.

[12] O. Diekmann and J.A.P. Heesterbeek, *Mathematical Epidemiology of Infectious Diseases*. Wiley, Chichester, 2000.