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.
Clustering model for transmission of the SARS virus: application to epidemic control and risk assessment

Michael Small*, C.K. Tse

Department of Electronic and Information Engineering, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong

Available online 26 January 2005

Abstract

We propose a new four state model for disease transmission and illustrate the model with data from the 2003 SARS epidemic in Hong Kong. The critical feature of this model is that the community is modelled as a small-world network of interconnected nodes. Each node is linked to a fixed number of immediate neighbors and a random number of geographically remote nodes. Transmission can only propagate between linked nodes. This model exhibits two features typical of SARS transmission: geographically localized outbreaks and “super-spreaders”. Neither of these features are evident in standard susceptible-infected-removed models of disease transmission. Our analysis indicates that “super-spreaders” may occur even if the infectiousness of all infected individuals is constant. Moreover, we find that nosocomial transmission in Hong Kong directly contributed to the severity of the outbreak and that by limiting individual exposure time to 3–5 days the extent of the SARS epidemic would have been minimal.

© 2005 Elsevier B.V. All rights reserved.

PACS: 05.45.—a; 89.75.—k; 0.2.50.—r; 87.10. + e

Keywords: Severe acute respiratory syndrome; Nonlinear dynamics; Disease transmission; Epidemiological methods; Small world networks

*Corresponding author. Tel.: +852 2766 4744; fax: +852 2362 8439. E-mail address: ensmall@polyu.edu.hk (M. Small).
1. Severe acute respiratory syndrome (SARS)

The transmission of SARS in Hong Kong [1] (see Fig. 1) may be characterized by two apparently unusual features [2]: so-called super-spread events (SSE), in which a single individual initiates a large number of cases; and persistent transmission within the community. Two notable SSE were observed early in the epidemic and have been widely reported: at the Amoy Gardens housing estate and at the Prince of Wales hospital. Epidemiological studies [3,4] have found that in Hong Kong: the fatality rate was approximately 17% (compared to 11% globally); the mean incubation period was 6.4 days (range 2–10) [5]; the duration between onset of symptoms and hospitalization was 3–5 days; and the mean number of individuals infected by each case during the initial phase of the epidemic was 2.7 [2]. Like any epidemic, an initial exponential outbreak is evident in the SARS infection data for Hong Kong [5], but this initial explosion was soon tempered. In this paper, we present an alternative

Fig. 1. Daily reported (upper panel) and revised (lower panel) SARS infection data for Hong Kong. At the end of the epidemic, 105 days after the first recorded case, a total of 1755 confirmed infections had been identified in Hong Kong. The reported data are the daily number of hospital admission of SARS cases, as published in the South China Morning Post (and other local media) during the outbreak. Subsequent to the epidemic (several months later) a revised set of data was released by the local government health authority. These are the “revised” data in the lower panel. As we are unsure of the revision process use, and we believe it to include strong assumptions about the etiological agent involved, we analyze only the reported data in this paper.
explanation for the occurrence of SSE. We show that SSE may occur, as a result of social structure, even if the etiological agent is equally infectious in all patients.

Standard mathematical models of the spread of infectious diseases [6] are well known and widely applied. According to such models, individuals in the community can be classified as either susceptible to the disease (S), infected with the disease (I) or removed (and therefore immune to the disease) (R). Classically, transition between these three groups may then be governed by deterministic difference (or differential) equations and the transmission dynamics may exhibit only exponential growth or decay subject to population limitations. For a disease such as SARS the change in transmission dynamics are usually modelled by non-stationarity in the model parameters [7].

In this paper, we apply small world (SW) models to simulation of the spread of SARS in Hong Kong. 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.

For social networks, SW models are characterized by the property that every individual is connected to almost every other individual through a short chain of mutual acquaintances: the so-called “six degrees of separation” [8]. This effect has been observed in a large number of social models: joint authorship of academic publications, professional acquaintances, email exchanges, spread of computer viruses [9] and even co-starring in Hollywood movies [10]. For infectious diseases that are spread through close personal contact, a similar model of propagation is likely.

2. The model

Our aim is to accurately mimic the qualitative features of the SARS epidemic with the simplest (fewest parameters) model. We propose four distinct states. Individuals can be susceptible (S), prone (P), infected (I), or removed (R). These four categories are similar to, but distinct from the SEIR type model of infection widely used to model diseases with a significant incubation period [11]. As is usual, the class S represents individuals that are susceptible to the infectious agent, but are not yet infected. Category R represents individuals that are permanently incapable of either transmitting or acquiring the disease: typically, this represents those that have recovered from the infection (and are assumed to be immune), those that have died, and those that are either isolated or quarantined. The I class represents individuals that are infected and infectious, whereas P represents those that are infected but not yet infectious (similar to E in the standard formulation).

The transmission path (state transition graph) is depicted in Fig. 2. 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
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.

We fix the population $N$ and assume that there are no other births or deaths from any other cause. 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 $p_1$, and three other nodes (hashed) 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.

We fix the population $N$ and assume that there are no other births or deaths from any other cause. 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 $n_2$ long distance links (see Fig. 2). These are links to nodes that are geographically remote from one another, infection occurs along these pathways with probability $p_2$. For each node $i$

\[^1\text{The grid has periodic boundary conditions, that is, nodes (1,1) and (L,L) are neighbors.}\]
the number $n_2^{(i)}$ is fixed and so are the links to its $n_1^{(i)}$ remote neighbors. The number $n_2^{(i)}$ is chosen to follow a discrete exponentially decaying distribution $f_X(x) = \frac{1}{C} e^{-x/\mu}$ with parameter $\mu$ proportional to the expected (average) number of links to remote nodes, and parameter $C = 1/(1 - e^{-1/\mu})$ ensures that $f_C$ is a probability distribution function. It is the inclusion of long distance links with a random number of links that gives rise to the network’s SW structure. Moreover, the same long distance links can also cause the network to exhibit scale free (SF) properties. By SF we mean that logarithm of the probability of a node having $n$ links is inversely proportional to $\log(n)$ (very large number of links occur relatively frequently). Whether the induced network of actual infections follows a SF distribution remains to be tested [7] and is not the focus of this paper.

Finally, for each simulation we seed the model with one initial infection. We expect that computational simulations of this network will show that infection will spread locally, just as SARS spread within particular geographical regions of Hong Kong. Moreover, the system may also exhibit long-range infection, as a single individual may infect individuals in distant communities. Occasionally, individuals will infect a large number of other individuals exactly as was observed at the start of the SARS epidemic in Hong Kong (a SSE).

As stated, our model has seven parameters: $L$, $n_1$, $\mu$, $p_1$, $p_2$, $r_1$, and $r_2$. To simulate the population of Hong Kong we set $L = 2700$ ($N = L^2 = 7,290,000$). We arbitrarily choose $n_1 = 4$ (this choice will be partially vindicated later, but this number also slightly exceeds the average household size in Hong Kong $^3$). From the available data we know that the average incubation period between infection and becoming symptomatic is 6.4 days [5]. With transition probability $r_0$ the number of days in the prone state is the result of a series of independent Bernoulli trials with a mean $1/r_0$ and therefore follows a geometric distribution $f_X(x) = (1 - p)^{x-1}p$ (see Fig. 3). Similarly, the time before hospitalization (and presumably quarantining and therefore removal) is 3–5 days. Suppose that the average is 4. In our model, the number of days prior to hospitalization also follows a geometric distribution with mean $1/r_1$. Fig. 3 illustrates the distribution of the model parameter $n_2$ and the probability distribution of the number of days in state P and I.

Hence the only free parameters are $\mu$, $p_1$ and $p_2$. Without active control we also know that the average number of new infections per case is 2.7 [2]. In this state, each infectious individual will infect, on average, $n_1 p_1 + E(n_2^{(i)}) p_2$ new individuals every day. Since $E(n_2^{(i)}) = \mu$ and we suppose that the time before hospitalization is 5 days (on the upper end of the documented range) we have

\[ n_1 p_1 + \mu p_2 \approx 0.54. \]
Hence \( p_1/C25 1 = n_1 \left( 0.54 - 0.25\mu p_2/m \right) \). We can, therefore, parameterize the model behavior in terms of only the SW network parameters \( \mu \) and \( p_2 \), subject to \( \mu p_2 < 0.54 \).

Now, \( n_1 \) and \( n_2 \) represent the number of interactions an individual has each day. Therefore, the sum \( n_1 + n_2 \) is a lower bound on the number of active acquaintances. The reason \( n_1 + n_2 \) is only a lower bound, is that \( n_1 + n_2 \) is the number of links that are sufficiently intimate to support transmission of the virus. In reality, some links would be closer than others and would be far more likely to lead to transmission. We make the simplification that transmission will occur with probability \( p_1 \) on all the \( n_1 \)-links and probability \( p_2 \) along all the \( n_2 \)-links. Informally, \( p_1 \) is the probability of transmission between members of the same household in a 24 h period, \( p_2 \) is the probability of transmission between acquaintances over one day. These two probabilities are not necessarily the same.
Therefore $n_2$ may be defined as the number of individuals with this transmission probability. Experimental evidence suggests that any two individuals in continental North America (in 1967) are connected by (on average) no more than six links [12]. Ignoring mutual acquaintance, and assuming a population of 200,000,000 one can deduce that each individual has approximately $\log(2 \times 10^8)/\log 6 \approx 10.667$ unique acquaintances. The problem of mutual acquaintances complicates things slightly, but this will not have an significant effect on the result. Moreover, if we set $n_1 = 4$ (the horizontal and vertical neighbors only) then the probability of two nodes having mutual acquaintances is only $n^2 = N$. Hence, in our simulations, we take $n_1 = 4$ and $E(n_2) = \mu = 7$. Note that the choice of $n_1$ and $n_2$ is not critical, what is more important is the infection probability $p_1 n_1 + p_2 n_2$. In fact, the variation in control (such as quarantining of suspected infected individuals) can be equally modelled as a reduction in $n_1$ and $n_2$. In this model we aim to keep the parameter changes as simple as possible and achieve the same result by changing only $p_2$.

We summarize the model parameters as follows:

$$L = 2700,$$
$$r_0 = \frac{1}{14} \approx 0.135135,$$
$$r_1 = \frac{1}{4} = 0.25,$$
$$n_1 = 4,$$
$$\mu = 7,$$ 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 $r_0 = \frac{1}{14}$ rather than $\frac{1}{54}$. This does not have a significant affect on our results, it is merely a computational convenience. Finally, we have only one free parameter $p_2$. In what follows, we will investigate the effect of time dependent changes in $p_2$ (i.e., $p_2(t)$). Therefore we constrain the constant $p_1 = 0.135 - 0.25p_2(0)$.

### 3. Computation

Our reasoning in the preceding section has left us with a single unknown parameter $p_2$. We start by computing multiple simulations for various values of $p_2$ in the range $[0, 0.08]$ with $p_1$ determined by Eq. (1). Our aim is to identify parameter values $p_{1,2}$ that exhibit dynamics most like those observed in Hong Kong. We start by simulating only the uncontrolled case (i.e., both $p_1$ and $p_2$ are constant). For each set of parameter values we simulate 1000 random realizations starting from a single infected individual until either: (i) the disease dies out (no I or P individuals remain);

---

4In future work we plan to explore further the effect of variation in $n_1$ and $n_2$. 
(ii) the disease becomes uncontrollable (more than 3000 I and P individuals); or (iii) the disease persists for 50 days.

Fig. 4 illustrates estimates on the number of casualties, and likelihood of outbreak for $r_1 = 0.25$ and $p_2 \in [0, 0.08]$. From Fig. 4 we see that, consistent with our expectations, for large $p_2$ the scale of the outbreak increases. Moreover outbreaks are most likely when there is clustering (i.e., $p_2 > 0$). However, we also note that the total number of

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{graph.png}
\caption{Expected model behavior for $r_1 = 0.25$. The top panel shows the distribution on the number of casualties during the first 50 days of transmission. The middle plot shows the number of clusters exhibited by the model over 50 days. The bottom plot shows the probability of an uncontrollable outbreak (defined as more than 3000 infected (I and P) individuals, or infection lasting more than 50 days); and the expected total number of casualties in the cases for which the disease was controlled (i.e., self-terminated). The upper two panels show median and 70% and 90% confidence intervals for 1000 simulations for each value of $p_2$. The bottom panel are mean and incidence rate estimated from the same 1000 simulations.}
\end{figure}
infections, even in the worst case is significantly lower than for the true data. In comparison the number suspected cases in Hong Kong 50 days after the first reported case was 1589. From the revised data for Hong Kong, the number of identified SARS cases 50 after the index patient was hospitalized was 1161. Hence, even with external control measures the true data is significantly larger than these model simulations.

Because we assume that the time between onset of symptoms (entering state I) and isolation (entering state R) is 3–5 days, Fig. 4 represents simulations from our model under the assumption that hospitalization implies complete isolation (hospitalization occurred in Hong Kong after 3–5 days). However, Fig. 4 clearly underestimates the extent of transmission and we are therefore forced to abandon this assumption. For the extent of transmission observed in Hong Kong, significant transmission must have occurred after hospitalization.

In other words, the least reliable assumption of our model was that hospitalization represented isolation and that isolation would therefore occur after 3–5 days. Hence, the data depicted in Fig. 4 excludes the possibility of nosocomial infection. We can modify our model and include this possibility by decreasing \( r_1 \). In Fig. 5 we show equivalent results for \( r_1 = 0.16 \) (corresponding to removal after a mean of about 6 days). Fig. 5 illustrates that with a much lower removal rate, infection levels similar to those seen in Hong Kong during 2003 are observed. Moreover, from Fig. 5 it is clear that large scale transmission occurs when models exhibit a large number of clusters. Hence quarantine, travel restrictions and closure of schools and work places are all effective control strategies for this epidemic. However, such measures need to be implemented early.

We now explore more closely variation in \( r_1 \) for fixed values of \( p_{1,2} \). We fix \( p_2 = 0.04 \) and \( p_1 = 0.065 \) and estimate the expected casualty rate for various removal rates \( r_1 < 0.33 \) is given in Fig. 6. As should be expected, the model exhibits a sensitive dependence on \( r_1 \). But more significantly, widespread outbreak is likely only if individuals cannot be isolated within a few days. Furthermore, we see that the rate of transmission of the disease is inversely proportional to \( r_1 \) (and therefore directly proportional to the expected number of days prior to removal of an individual).

Finally, we now provide simulations of the Hong Kong epidemic by varying the control parameter \( r_1 \). We initiate the model with a single infected individual and a relatively low removal rate \( r_1 \). As the disease progresses we increase \( r_1 \) (after 20 and 40 days) to reflect improved governmental control measures. After 20 days the first cases of SARS were isolated in Hong Kong. After 40 days it had become clear that a serious health problem was evolving and improved isolation measures (quarantine and school closure) were introduced. We also decrease \( p_2 \) to reflect the changing health practices of the community. For the sake of simplicity we introduce only a single step change in \( p_2 \) at the peak of the epidemic. This corresponds to increased public awareness of appropriate hygiene measures and a corresponding drop in the transmission rate. Fig. 7 presents our results.

It is apparent from Fig. 7 that the simulations and the data share many common features. The approximately exponential growth prior to 45 days and exponential decay subsequent to this is a direct result of the selected parameters \((r_1, p_2)\). More significantly, all simulations exhibit “bursty” data consistent with localized
outbreaks resulting from SSEs, apart from peaks at day 45 (or shortly thereafter) all extremum result from the stochastic nature of the model.

The parameter values selected in Fig. 7 are arbitrary, and rather simplistic. Obviously the true situation would include a more irregular and gradual change in parameters. However, we have found that this simple situation is sufficient to
reproduce the salient features observed in the data. Given the model described here, the probability of infecting fewer than 20 people was approximately 0.205, while the probability of infecting more than 1000 was 0.263. One can see that the true situation for Hong Kong (1755 casualties) is quite typical\textsuperscript{5} of this model. Nonetheless, there is

\textsuperscript{5}From 1000 simulations, 110 exhibited a larger number of casualties.
a very large variation on the number of casualties. However, in all cases the parameter values of Fig. 7 provided effective control of SARS transmission after approximately 150 days. It is therefore clear that with effective control measures in place the likelihood of a significant outbreak is low.

4. Discussion

The standard susceptible–infected–removed (SIR) model [6] assumes that at any time all susceptible (S) individuals have an equal small probability of becoming infected (I), and that any infected individual may become removed (R): either through death, isolation, or recovery and henceforth immunization. However, by imposing a SW-SF structure on the system we simulated the effect of propagation by acquaintance: one can only become infected after contact with an infected individual [13,14]. Furthermore, the explicit modelling of associations allows more effective simulation of control measures such as quarantining (or school closure). Finally, localized outbreaks (SSE) can be more realistically modelled with the inhomogeneous SW-SF model structure. The implication of this is that the SSE do not necessarily require highly contagious individuals, only a highly connected individual, and, highly connected individuals occur naturally in any community. In other words, high rates of infection for certain individuals are not necessarily related to the action of the disease, only to their degree of social connectedness. Therefore we conclude that a highly variable infection rate is not necessarily a significant factor in the transmission of SARS.

Significantly, our simulations indicate that nosocomial transmission of SARS was a significant factor during the early part of the epidemic. Without hospital-based transmission the rate of disease transmission is substantially lower than the observed behavior. Hence, isolation of patients as soon as symptoms are identified is key to containing future outbreaks of SARS. However, our analysis is far from complete. A more systematic analysis of the effect of hospitalization—possibly including separate categories in the compartmental state model—is required [15]. Such work has already been done for the SARS outbreak in Toronto [15], and this needs to be extended to the larger population in Hong Kong.

Our simulations indicate that exponential growth of the epidemic was only prevented by improved isolation (higher \( r_1 \) and lower \( p_2 \)). It is not clear whether this improved behavior was a result of changing public practice (higher \( r_1 \)) or simply changing weather conditions (lower \( p_2 \)).

It is usual to model changing control of a disease by changing the removal rate \( r_1 \). In addition to this our models suggest that improved quarantine and public health procedures could be modelled by decreasing \( n_2 \) (or \( \mu \)). It is clear from our investigation that \( n_2 > 0 \) is necessary for widespread contamination. The physical meaning of \( n_2 \) is clear, and further analysis with this model may provide a more detailed indication of the efficacy of control measures (as belatedly implemented in Hong Kong during 2003) such as closure of schools and travel restriction. In a
companion paper we derive analytic expressions for the expected behavior of the model [16].

Acknowledgements

This work was supported by a Hong Kong University Grants Council Competitive Earmarked Research Grant (No. PolyU 5235/03E) and a Hong Kong Polytechnic University Direct Allocation (No. A-PF95).

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. http://www.sars-expertcom.gov.hk/english/reports/report.html.
[2] 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.
[3] G.W.K. Wong, D.S.C. Hui, Severe acute respiratory syndrome (SARS): epidemiology diagnosis and management, Thorax 58 (2003) 558–560.
[4] Moira Chan-Yeung, Rui-Heng Xu, SARS: epidemiology, Respirology 8 (2003) S9–S14.
[5] C.A. Donnelly, et al., Epidemiological determinants of spread of causal agent of severe acute respiratory syndrome in Hong Kong, Lancet (2003) http://image.thelancet.com/extras/03art4453-web.pdf.
[6] J.D. Murray, Mathematical Biology, Biomathematics Texts, vol. 19, second ed., Springer, Berlin, 1993.
[7] M. Small, Pengliang Shi, C.K. Tse, Plausible models for propagation of the SARS virus, IEICE Trans. Fund. Electron. Commun. Comput. Sci. E87-A (2004) 2379–2386.
[8] D.J. Watts, Six Degrees: The Science of a Connected Age, Norton, New York, 2003.
[9] R. Pastor-Satorras, A. Vespignani, Epidemic spreading in scale-free networks, Phys. Rev. Lett. 86 (2001) 3200–3203.
[10] J.J. Collins, C.C. Chow, It’s a small world, Nature 393 (1998) 409–410.
[11] O. Diekmann, J.A.P. Heesterbeek, Mathematical Epidemiology of Infectious Diseases, Wiley, Chichester, 2000.
[12] S. Milgram, The small world problem, Psychol. Today 2 (1967).
[13] L.S. Tsimring, R. Huerta, Modelling of contact tracing in social networks, Physica A 325 (2003) 33–39.
[14] R. Huerta, L.S. Tsimring, Contact tracing and epidemic control in social networks, Phys. Rev. E 66 (2002) 056115-1-4.
[15] G. Webb, M. Blaser, H. Zhu, S. Ardel, J. Wu, Critical role of nosocomial transmission in the Toronto SARS outbreak, Math. Biosci. Eng. 1 (2004) 1–13.
[16] M. Small, C.K. Tse, Small world and scale free model of transmission of SARS, International Journal of Bifurcations and Chaos (2005), to appear.