

# Forecast and control of epidemics in a globalized world

L. Hufnagel<sup>†</sup>, D. Brockmann, and T. Geisel

Max-Planck-Institut für Strömungsforschung, Bunsenstrasse 10, 37073 Göttingen, Germany; and Kavli Institute for Theoretical Physics, University of California, Santa Barbara, CA 93106

Edited by Robert May, University of Oxford, Oxford, United Kingdom, and approved August 26, 2004 (received for review December 15, 2003)

The rapid worldwide spread of severe acute respiratory syndrome demonstrated the potential threat an infectious disease poses in a closely interconnected and interdependent world. Here we introduce a probabilistic model that describes the worldwide spread of infectious diseases and demonstrate that a forecast of the geographical spread of epidemics is indeed possible. This model combines a stochastic local infection dynamics among individuals with stochastic transport in a worldwide network, taking into account national and international civil aviation traffic. Our simulations of the severe acute respiratory syndrome outbreak are in surprisingly good agreement with published case reports. We show that the high degree of predictability is caused by the strong heterogeneity of the network. Our model can be used to predict the worldwide spread of future infectious diseases and to identify endangered regions in advance. The performance of different control strategies is analyzed, and our simulations show that a quick and focused reaction is essential to inhibiting the global spread of epidemics.

The application of mathematical modeling to the spread of epidemics has a long history and was initiated by Daniel Bernoulli's (1) work on the effect of cowpox inoculation on the spread of smallpox in 1760. Most studies concentrate on the local temporal development of diseases and epidemics. Their geographical spread is less well understood, although important progress has been achieved in a number of case studies (2–4). The key question, as well as difficulty, is how to include spatial effects and quantify the dispersal of individuals. This problem has been studied with some effort in various ecological systems, for instance, in plant dispersal by seeds (5). Today's volume, speed, and nonlocality of human travel (Fig. 1), as well as the rapid worldwide spread of severe acute respiratory syndrome (SARS) (Fig. 2), demonstrate that modern epidemics cannot be accounted for by local diffusion models that are applicable only as long as the mean distance traveled by individuals is small compared to geographical distances. These local reaction–diffusion models generically lead to epidemic wavefronts, which were observed, for example, in the geotemporal spread of the Black Death in Europe from 1347 to 1350 (6–10).

Here we focus on mechanisms of the worldwide spread of infectious diseases. Our model consists of two parts: a local infection dynamics and the global traveling dynamics of individuals similar to the models investigated in ref. 11. However, both constituents of our model are treated on a stochastic level, taking full account of fluctuations of disease transmission, latency, and recovery on the one hand and of the geographical dispersal of individuals on the other. Furthermore, we incorporate nearly the entire civil aviation network.

## Local Infection Dynamics

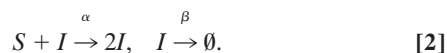
In the standard deterministic systemic inflammatory response (SIR) model for infectious diseases, a population with  $N$  individuals is categorized according to its infection status: susceptibles ( $S$ ), infectious ( $I$ ), or recovered and immune ( $R$ ) (6, 12). The dynamics that specifies the flow among these categories is given by

$$ds/dt = -\alpha sj, \quad dj/dt = \alpha sj - \beta j, \quad [1]$$

where  $s = S/N$  and  $j = I/N$  denote the relative number of susceptibles and infecteds, respectively. The relative number of recovered individuals  $r = R/N$  is obtained by conservation of the entire population, i.e.,  $r(t) = 1 - j(t) - s(t)$ , and  $\tau = \beta^{-1}$  is the average infectious period. The key quantity describing the infection is the basic reproduction number  $\rho_0 = \alpha/\beta$ , which is the average number of secondary infections transmitted by an infectious individual in an otherwise uninfected population. If  $\rho_0 > 1$  and the initial relative number of susceptibles are greater than a critical value  $s_c = 1/\rho_0$ , an epidemic develops ( $dj/dt > 0$ ). As the number of infected individuals increases, the fraction of susceptibles  $s$  decreases, and thus the number of contacts of infected individuals with susceptibles decreases until  $s = s_c$ , when the epidemic reaches its maximum and subsequently decays.

The above SIR model incorporates the underlying mechanism of transmission and recovery dynamics and has been able to account for experimental data in a number of cases. However, transmission of and recovery from an infection are intrinsically stochastic processes, and the deterministic SIR model does not account for fluctuations. These fluctuations are particularly important at the beginning of an epidemic when the number of infecteds is very small.

In this regime, a probabilistic description must be used. Schematically, the stochastic infection dynamics is given by



The first reaction reflects the fact that an encounter of an infected individual with a susceptible results in two infecteds at a probability rate  $\alpha$ ; the second indicates that infecteds are removed (recover) at a rate  $\beta$  and effectively disappear from the population. The quantity of interest is the probability  $p(S, I; t)$  of finding a number  $S$  of susceptibles and  $I$  infecteds in a population of size  $N$  at time  $t$ . Assuming that the process is Markovian on the relevant time scales, the dynamics of this probability is governed by the master equation (13)

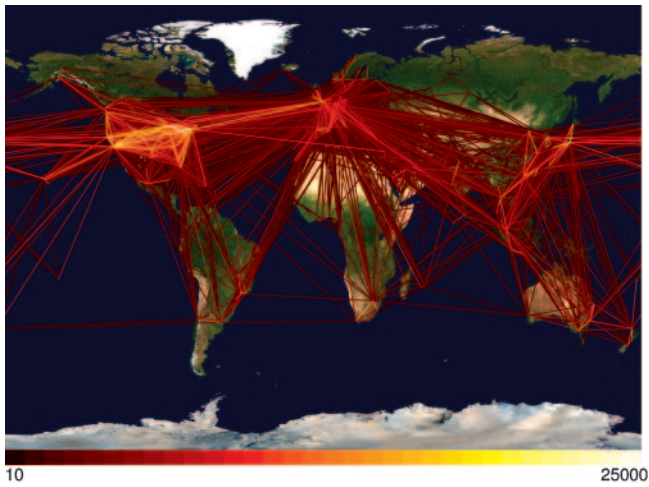
$$\begin{aligned} \partial_t p(S, I; t) = & \frac{\alpha}{N} (S+1)(I-1) p(S+1, I-1; t) \\ & + \beta(I+1) p(S, I+1; t) \\ & - \left( \frac{\alpha}{N} SI p - \beta I \right) p(S, I; t). \end{aligned} \quad [3]$$

This paper was submitted directly (Track II) to the PNAS office.

Abbreviations: SARS, severe acute respiratory syndrome; SIR, systemic inflammatory response; WHO, World Health Organization.

<sup>†</sup>To whom correspondence should be addressed. E-mail: lars@chaos.gwdg.de.

© 2004 by The National Academy of Sciences of the USA



**Fig. 1.** Global aviation network. A geographical representation of the civil aviation traffic among the 500 largest international airports in >100 different countries is shown. Each line represents a direct connection between airports. The color encodes the number of passengers per day (see color code at the bottom) traveling between two airports. The network accounts for >95% of the international civil aviation traffic. For each pair  $(i, j)$  of airports, we checked all flights departing from airport  $j$  and arriving at airport  $i$ . The amount of passengers carried by a specific flight within 1 week can be estimated by the size of the aircraft (We used manufacturer capacity information on >150 different aircraft types) times the number of days the flight operates in 1 week. The sum of all flights yields the passengers per week, i.e.,  $M_{ij}$  in Eq. 7. We computed the total passenger capacity  $\Sigma M_{ij}$  of each airport  $j$  per week and found very good agreement with independently obtained airport capacities.

In addition to this dynamics, one must specify the initial condition  $p(S, I; t = t_0)$ , which is typically assumed to be a small but fixed number of infecteds  $I_0$ , i.e.,  $p(S, I; t = t_0) = \delta_{I, I_0} \delta_{S, N - I_0}$ .

The relation of the probabilistic master equation 3 to the deterministic SIR model (1) can be made in the limit of a large but finite population, i.e.,  $N \gg 1$ . In this limit, one can approximate the master equation by a Fokker–Planck equation by means of an expansion in terms of conditional moments [Kramers–Moyal expansion (13); see supporting information, which is published on the PNAS web site]. The associated description in terms of stochastic Langevin equations reads

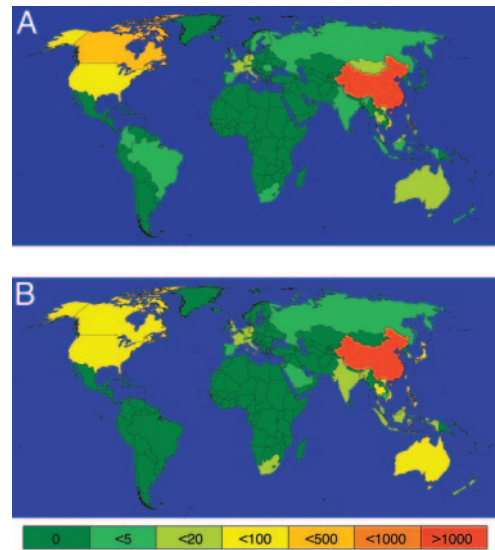
$$ds/dt = -\alpha sj + \frac{1}{\sqrt{N}} \sqrt{\alpha sj} \xi_1(t) \quad [4]$$

$$dj/dt = \alpha sj - \beta j - \frac{1}{\sqrt{N}} \sqrt{\alpha sj} \xi_1(t) + \frac{1}{\sqrt{N}} \sqrt{\beta j} \xi_2(t). \quad [5]$$

Here, the independent Gaussian white noise forces  $\xi_1(t)$  and  $\xi_2(t)$  reflect the fluctuations of transmission and recovery, respectively. Note that the magnitude of the fluctuations are  $\propto 1/\sqrt{N}$  and disappear in the limit  $N \rightarrow \infty$ , in which case Eqs. 1 are recovered. However, for large but finite  $N$ , a crucial difference is apparent: Eqs. 4 and 5 contain fluctuating forces, and  $N$  is a parameter of the system. A careful analysis shows that even for very large populations (i.e.,  $N \gg 1$ ), fluctuations play a prominent role in the initial phase of an epidemic outbreak and cannot be neglected. For instance, even when  $\rho_0 > 0$ , a small initial number of infecteds in a population may not necessarily lead to an outbreak that cannot be accounted for by the deterministic model.

### Dispersal on the Aviation Network

As individuals travel around the world, the disease may spread from one place to another. To quantify the traveling behavior of



**Fig. 2.** Global spread of SARS. (A) Geographical representation of the global spreading of probable SARS cases on May 30, 2003, as reported by the WHO and Centers for Disease Control and Prevention. The first cases of SARS emerged in mid-November 2002 in Guangdong Province, China (17). The disease was then carried to Hong Kong on the February 21, 2003, and began spreading around the world along international air travel routes, because tourists and the medical doctors who treated the early cases traveled internationally. As the disease moved out of southern China, the first hot zones of SARS were Hong Kong, Singapore, Hanoi (Vietnam), and Toronto (Canada), but soon cases in Taiwan, Thailand, the U.S., Europe, and elsewhere were reported. (B) Geographical representation of the results of our simulations 90 days after an initial infection in Hong Kong. The simulation corresponds to the real SARS infection at the end of May 2003. Because our simulations cannot describe the infection in China, where the disease started in November 2002, we used the WHO data for China.

individuals, we have analyzed all national and international civil flights among the 500 largest airports by passenger capacity.<sup>‡</sup> This analysis yields the global aviation network shown in Fig. 1; further details of the data collection are compiled in the supporting information. The strength of a connection between two airports is given by passenger capacity, i.e., the number of passengers that travel a given route per day.

We incorporate the global dispersion of individuals into our model by dividing the population into  $M$  local urban populations labeled  $i$  containing  $N_i$  individuals. For each  $i$ , the number of susceptible and infected individuals is given by  $S_i$  and  $I_i$ , respectively. In each urban area, the infection dynamics is governed by master equation 3.

Stochastic dispersal of individuals is defined by a matrix  $\gamma_{ij}$  of transition probability rates among populations

$$S_i \xrightarrow{\gamma_{ij}} S_j \quad I_i \xrightarrow{\gamma_{ij}} I_j, \quad i, j = 1, \dots, M, \quad [6]$$

where  $\gamma_{ii} = 0$ . Along the same lines as presented above, one can formulate a master equation for the pair of vectors  $\mathbf{X} = \{S_1, I_1, \dots, S_M, I_M\}$ , which defines the stochastic state of the system. This master equation is provided explicitly in Eq. 3, supporting information.

To account for the global spread of an epidemic via the aviation network, one needs to specify the matrix  $\gamma_{ij}$ . Because the

<sup>‡</sup>Data on flight schedules and airport information are available from OAG Worldwide Limited, London ([www.oag.com](http://www.oag.com)), and the International Air Transport Association, Geneva ([www.iata.org](http://www.iata.org)).

**Table 1. A comparison of the SARS case reports provided by the WHO and the results of our simulation for all countries with a reported case number  $\geq 4$**

Country	WHO		Simulation			
	5/20/2003	5/30/2003	Average	$\eta$	Minimum	Maximum
Hong Kong	1,718	1,739	1,951	0.35	1,373.9	2,770.4
Taiwan	383	676	318.2	0.55	184.0	550.3
Singapore	206	206	136.6	0.68	69.4	268.7
Japan	—	—	60.4	0.84	26.6	137.0
Canada	140	188	41.8	0.94	16.4	106.6
U.S.	67	66	65.9	0.84	28.4	152.7
Vietnam	63	63	49.2	0.86	20.7	116.3
Philippines	12	12	30.0	0.97	6.2	50.7
Germany	9	10	14.4	1.1	4.8	43.1
The Netherlands	—	—	5.9	1.09	2.0	17.6
Bangladesh	—	—	10	1.15	3.2	31.6
Mongolia	9	9	—	—	—	—
Italy	9	9	5.3	1.02	1.9	14.6
Thailand	8	8	35.4	0.89	14.5	86.8
France	7	7	7.6	1.09	2.6	22.6
Australia	6	6	27.0	1.05	10.1	72.5
Malaysia	7	5	17.7	1.05	6.2	50.7
United Kingdom	4	4	16.7	1.04	5.9	47.0

The expected number of infecteds predicted by our model is estimated by the average <1,000 realizations of the stochastic model. The epidemic was simulation for  $t = 90$  days after the initial outbreak of SARS in Hong Kong (February 19, 2003), yielding a simulation end of May 20, 2003. The range designated by columns 6 and 7 was computed by means of the fluctuation measure  $\eta$  (see text), which is approximately one for all countries.

global exchange of individuals among urban areas is carried out by airborne travel, one can estimate the probability rate matrix  $\gamma_{ij}$  by  $t$ . We assume that an individual remains in an urban area for some time before traveling to another region. A flight  $j \rightarrow i$  is chosen according to the weights

$$w_{ij} = M_{ij} / \sum_i M_{ij} \quad [7]$$

where  $M_{ij}$  is the number of passengers per unit time that depart from an airport in region  $j$  and arrive at an airport in region  $i$ . The matrix  $w$  accounts for the overall connectivity of the aviation network as well as for the heterogeneity in the strength of the connections. Denoting the typical time period, individuals remain at  $i$  by  $\tau_i$ ; the matrix  $\gamma_{ij}$  is expressed in terms of  $w_{ij}$  according to  $\gamma_{ij} = w_{ij}/\tau_j$ . If we assume that each airport is surrounded by a catchment area with a population  $N_i$ , the typical time individuals remain at  $i$  is given by  $\tau_i = N_i/\sum_j M_{ji}$ . If the capacity of airport  $i$  reflects the need of the associated catchment area (i.e.,  $N_i \propto \sum_j M_{ji}$ ), the waiting times  $\tau_i$  are identical for all  $i$ , i.e.,  $\tau_i = \tau = \gamma^{-1}$ , which implies  $\gamma_{ij} = \gamma w_{ij}$ . In our model, the global rate  $\gamma$  is a free parameter. To verify its validity, we apply our model to the SARS outbreak. The rate  $\gamma$  can be computed from the ratio of the number of infected individuals in Hong Kong to the number of infected individuals outside Hong Kong, which is provided by the World Health Organization (WHO) data.<sup>§</sup> For the local infection dynamics, we use a simple extension of the above stochastic SIR model: The categories  $S$ ,  $I$ , and  $R$  are completed by a category  $L$  of latent individuals who have been infected but are not yet infectious themselves, accounting for the latency of the disease. In our simulations, individuals remain in the latent or infectious stage for periods drawn from the delay distribution provided in figure 2 in ref. 14. In our simulation, we chose

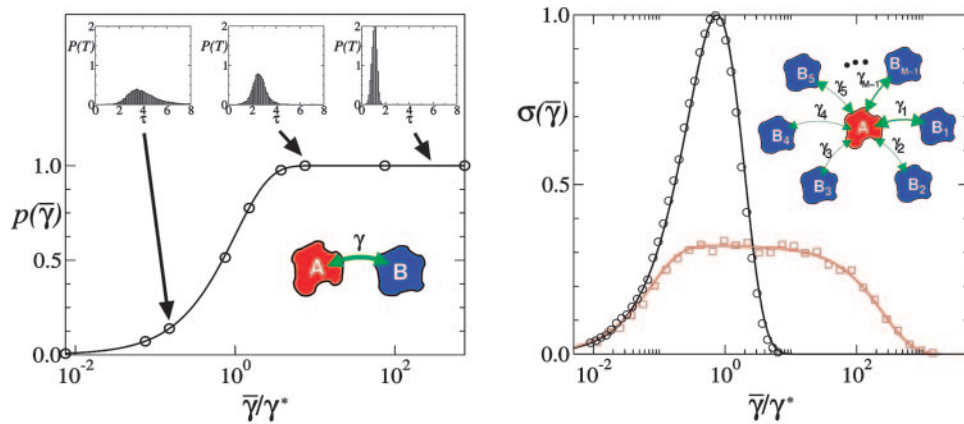
random infection times, the distribution of which is known for SARS (14). In a realistic simulation, the basic reproduction number  $\rho_0$  cannot be assumed to be constant over time. Successful control measures, for instance, generally decrease  $\rho_0$ . We chose a time-dependent  $\rho_0(t)$ , as provided by refs. 15 and 16.

### Results of Simulations

Fig. 2 depicts a geographical representation of the results of our simulations. Initially, an infected individual was placed in Hong Kong. For this initial condition, we simulated 1,000 realizations of the stochastic model and computed the mean value  $\langle I(t) \rangle$  of the number of infecteds at each node  $i = 1, \dots, M$  of the network. Because the size of catchment areas varies on many scales, the fluctuation range is best quantified by the means of the relative variance of  $z = \log I$ , i.e.,  $\eta = \sqrt{\langle z^2 \rangle - \langle z \rangle^2} / \langle z \rangle$ . In our simulations, we computed this measure for every  $i$  of the network. Fig. 2 shows the prediction of our model for the spread of SARS at  $t = 90$  days after the initial outbreak in Hong Kong (February 19, 2003), corresponding to the May 20, 2003 outbreak. The results of our simulations are in remarkable agreement with the worldwide spread of SARS as reported by the WHO (compare Fig. 2): There is an almost one-to-one correspondence between infected countries as predicted by the simulations and the WHO data.

Also, the orders of magnitude of the numbers of infected individuals in a country agree (Table 1). Although for most countries the cases reported by the WHO lie within the fluctuation range, two deviations between the reported cases and the predictions of the simulation are apparent: Our simulations predict a relatively high number of SARS cases in Japan (between 26.6 and 137.0). However, the Japanese government reported no confirmed case (only five suspected cases) of SARS in Japan as of May 30, 2003. How a single realization may deviate from the expectation can be seen from the difference between the simulation and the reported cases in the U.S. and Canada. The simulations show that on average the U.S. should have a higher number of SARS cases than Canada, although the opposite was reported by the WHO. The impact of the inherent

<sup>§</sup>Data on the SARS outbreak are available from the WHO, Geneva ([www.who.int/csr/sars/en](http://www.who.int/csr/sars/en)), and the Center for Disease Control and Prevention, Atlanta ([www.cdc.gov/ncidod/sars](http://www.cdc.gov/ncidod/sars)).



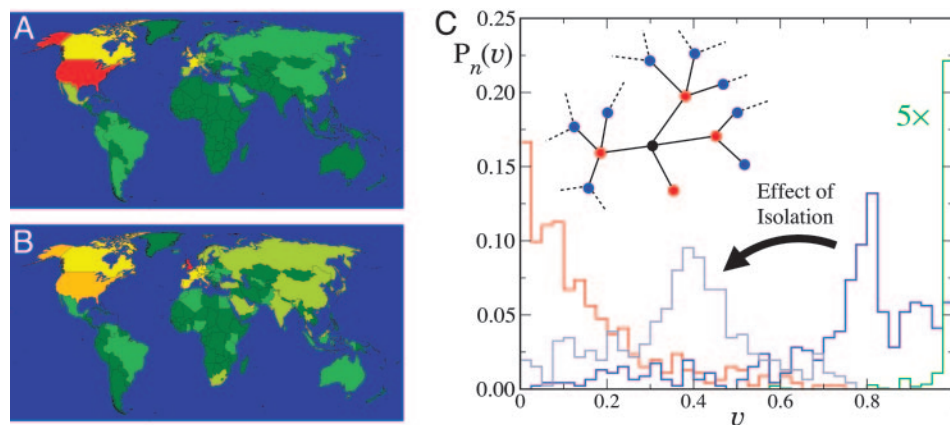
**Fig. 3.** Demonstration of the impact of fluctuations on inhomogeneous networks. Two confined populations with exchange of individuals are shown. In each population, the dynamics is governed by the SIR reaction scheme (6). Individuals travel from one population to the other at a rate  $\gamma$ . Parameters are  $N_A = N_B = 10,000$ ,  $R_0 = 4$ , and an initial number of infecteds  $I_0 = 20$  in population A. (Left) The probability  $p(\gamma)$  of an outbreak occurring in population B as a function of transition rate  $\gamma$ . (Insets) Histograms of the time lag  $T$  between the outbreaks in A and B for those realizations for which an outbreak occurs in B. The circles are results of the simulations of 100,000 realizations; the solid curve is the analytic result of Eq. 8. (Right) A star-shaped network with a central population A connected to  $M - 1$  populations  $B_1, \dots, B_{M-1}$  with rates  $\gamma_1, \dots, \gamma_{M-1}$ . The cumulated variance (see text) for a star network with 32 populations is depicted as a function of the average transmission rate  $\bar{\gamma}$ . Two cases are exemplified: equal rates (circles) and distributed rates according to Eq. 10 with  $\gamma_{\max}/\gamma_{\min} \approx 1,000$  (squares). The solid lines show the analytical results given by Eqs. 8 and 9. Parameters are  $N_A = N_B = 10,000$ ,  $R_0 = 4$ , and an initial number of infecteds  $I_0 = 20$  in population A. The numerical values are obtained by calculating the variance of the fluctuations of 100 different realizations of the epidemic outbreak for each  $\bar{\gamma}$ .

stochasticity of the infection and traveling dynamics is discussed below.

### The Impact of Fluctuations

Bearing in mind the low number of infections and the small value of  $\rho_0$  for SARS, the high degree of predictability, i.e., the low impact of fluctuations on the network level, is rather surprising, especially because our simulations take into account the full spectrum of fluctuations of disease transmission, recovery, and dispersal, and that the system evolves on a highly complex network. Naively, one expects that dispersal fluctuations between two given populations are amplified as the epidemic spreads globally, and that no prediction can be made. To clarify this important point, consider the system of two confined populations A and B, which exchange individuals as depicted in

Fig. 3. For simplicity, we assume that both populations have the same size (i.e.,  $N_A = N_B = N$ ), and individuals traverse at a rate  $\gamma$ . Now assume that initially a small number of infecteds  $I_0$  is introduced to population A without any infecteds contained in B. For a sufficiently high number of infecteds in A, an epidemic occurs. For  $\gamma > 0$ , infecteds are introduced to B, and a subsequent outbreak may occur in B after a time lag  $T$ . Fig. 3 depicts the results of simulations for two populations with  $n = 10,000$  and  $\rho_0 = 4$ . Various realizations of the time course  $I_A(t)$  and  $I_B(t)$  of the epidemic in both populations were computed. The initial number of infecteds in the population was  $I_A(t = 0) = I_0 = 20$ . Fig. 3 Left depicts the probability  $p(\gamma)$  of an outbreak occurring in population B as a function of the transition rate  $\gamma$ . For large enough rates, the probability is nearly unity, because a sufficient number of infecteds is introduced to B. For very low



**Fig. 4.** Inhomogeneity of the aviation network and control strategies. (A and B) Geographical representation of the results of two simulations of hypothetical SARS outbreaks 90 days after an initial infection in New York (A) and London (B) for the same parameters and color code as in Fig. 2. (C) Impact and control of epidemics. The probability  $p_n(v)$  of having to vaccinate a fraction  $v$  of the population to prevent the epidemic from spreading, if an initial infected individual is permitted to travel  $n = 1$  (red), 2 (blue), and 3 (green) times. The probability  $p_n(v)$  is estimated by placing the infected individual on a node  $i$  (black dot) of the network. The fraction  $v_i$  associated with node  $i$  is given by the number of susceptibles in a subnetwork that can be reached by the infected individual after  $n = 1, 2$  and 3 steps. Histogramming  $v_i$  for all nodes  $i$  yields an estimate for  $p_n(v)$ . The light-blue curve depicts the strong impact of isolating only 2% of the largest cities after an initial outbreak ( $n = 2$ ) and is to be compared to the blue curve.

rates  $\gamma$ , no infecteds are introduced to  $B$  during the time span of the epidemic in  $A$ , and thus  $p(\gamma) \rightarrow 0$  as  $\gamma \rightarrow 0$ . For intermediate values of  $\gamma$ , the probability  $p(\gamma)$  is neither one nor unity, and the time course in population  $B$  cannot be predicted with certainty. The function  $p(\gamma)$  is given by

$$p(\gamma) = 1 - \exp(-\gamma/\gamma^*), \quad [8]$$

where the critical rate  $\gamma^*$  is a function of the parameters  $\rho_0$  and  $N$ . Fig. 3 *Insets* depict histograms of the time lag  $T$  for those realizations for which an outbreak occurred in  $B$ . Each histogram corresponds to a different transition rate  $\gamma$ . The smaller  $\gamma$ , the higher the variability in  $T$ . Note that even in a range in which  $p(\gamma) \approx 1$ , the time lag  $T$  is still a stochastic quantity with a high degree of variance.

Consequently, the introduction of stochastic exchange of infected individuals leads to a lack of predictability in the time of onset of the initially uninfected population. In light of the analysis of two populations, the predictability in the case of SARS on the aviation network seems even more puzzling.

The situation changes drastically in networks that exhibit a high degree of variability in the rate matrix  $\gamma_{ij}$ . Clearly, this is the case for the aviation network. Consider the simple network depicted in Fig. 3. Each population contains  $N$  individuals. A central population  $A$  is coupled to a set of  $M-1$  surrounding populations  $B_1, \dots, B_{M-1}$ . Assume that initially a number of infecteds  $I_0$  is introduced to the central population  $A$ , such that an outbreak occurs. The entire set of rates  $\{\gamma_{ij}\}_{j=1, \dots, M-1}$  determines the behavior in the surrounding populations. If all rates  $\gamma_{ij}$  are identical and very small, we expect no infection to occur in the  $B_j$ ; for large enough  $\gamma_{ij}$ , an outbreak will occur in every  $B_j$ . In the aviation network, however, transition rates are distributed on many scales, and the response of the network to a central outbreak depends on the statistical properties of this distribution denoted by  $q(\gamma)$ . To quantify the reaction of the network, we introduce for each surrounding population a binary number  $\xi_j$  with  $j = 1, \dots, M-1$ , which is unity if an outbreak occurs in  $B_j$  and zero if it does not. According to Eq. 8, for a given rate  $\gamma$ , this quantity is a random number with a conditional probability density  $p(\xi_j|\gamma) = (1 - p(\gamma)) \delta(\xi_j) + p(\gamma) \delta(\xi_j - 1)$ . The variability of the network is thence quantified by the cumulative variance per population, and we define

$$\sigma = \frac{4}{M-1} \sum_i \text{var}(\xi_i) = \int d\gamma p(\gamma)(1 - p(\gamma))q(\gamma) \quad [9]$$

as a measure for the uncertainty of the network response. If, for example,  $q(\gamma) = \delta(\gamma - \bar{\gamma})$ , i.e., all transition rates are identical and equal to  $\bar{\gamma}$ , then  $\sigma(\bar{\gamma}) = 4p(\bar{\gamma})(1 - p(\bar{\gamma}))$ , which is unity for  $p(\bar{\gamma}) = 1/2$ . Comparing with Eq. 8, we see that when  $\gamma = \bar{\gamma}^* \log 2$ , the system with identical transition rates  $\gamma_{ij} = \bar{\gamma}$  exhibits the highest degree of unpredictability when the rates are of the order of the critical rate defined by Eq. 8. The function  $\sigma(\bar{\gamma})$  is shown in Fig. 3.

Now, assume rates  $\gamma_{ij}$  are drawn from a distribution

$$q(\gamma) = \frac{1}{\log(\gamma_{\max}/\gamma_{\min})} \frac{1}{\gamma} \quad \gamma_{\max} \leq \gamma \leq \gamma_{\min}, \quad [10]$$

which implies a high degree of variance within the interval  $[\gamma_{\min}, \gamma_{\max}]$  (i.e.,  $\gamma_{ij}$  is distributed uniformly on a logarithmic scale). This

high variability in rates drastically changes the predictability of the system. Inserting into Eq. 9 yields  $\sigma(\bar{\gamma})$  for strongly distributed rates. In Fig. 3, this function is compared to a system of identical transition rates. On the one hand, for intermediate values of  $\gamma \approx \gamma^*$ , the predictability is much higher than in the system of identical rates. This is a rather counterintuitive result. Despite the additional randomness in transition rates, the degree of determinism is increased.

### Control Strategies

Fig. 4 exemplifies how our model can be used to predict endangered regions if the origin of a future epidemic is located quickly. Fig. 4 depicts simulations of the global spread of SARS at  $t = 90$  days after hypothetical outbreaks in New York and London, respectively. Despite the worldwide spread of the epidemic in each case, the degree of infection of each country differs considerably, which has important consequences for control strategies.

Vaccination of a fraction of the population reduces the fraction of susceptibles and thus yields a smaller effective reproduction number  $\rho$ . If a sufficiently large fraction is vaccinated,  $\rho$  drops below 1, and the epidemic becomes extinct. The global aviation network can be used to estimate the fraction of the global population that needs to be vaccinated to prevent the epidemic from spreading. Fig. 4 demonstrates that a quick response to an initial outbreak is necessary if global vaccination is to be avoided. Also depicted is the probability  $p_n(v)$  of having to vaccinate a fraction  $v$  of the population if an infected individual is randomly placed in one of the cities and permitted to travel  $N = 1, 2$ , or 3 times. For the majority of originating cities, the initial spread is regionally confined, and thus a quick response to an outbreak requires only a vaccination of a small fraction of the population. However, if the infected individual travels twice, the expected fraction  $\langle v \rangle$  of the population that needs to be vaccinated is considerable (74.58%). For  $n = 3$ , global vaccination is necessary.

As a reaction to a new epidemic outbreak, it might be advantageous to impose travel restrictions to inhibit the spread. Here we compare two strategies: (i) the shutdown of individual connections and (ii) the isolation of cities. Our simulations show that an isolation of only 2% of the largest cities already drastically reduces  $\langle v \rangle$  (with  $N = 2$ ) from 74.58% to 37.50% (compare the blue and light-blue curves in Fig. 4). In contrast, a shutdown of the strongest connections in the network is not nearly as effective. To obtain a similar reduction of  $\langle v \rangle$ , the top 27.5% of connections would need to be taken off the network. Thus, our analysis shows that a remarkable success is guaranteed if the largest cities are isolated as a response to an outbreak.

In a globalized world, with millions of passengers traveling around the world week by week, infectious diseases may spread rapidly around the world. We believe that a detailed analysis of the aviation network represents a cornerstone for the development of efficient quarantine strategies to prevent diseases from spreading. Because our model is based on a microscopic description of traveling individuals, our approach may be considered a reference point for the development and simulation of control strategies for future epidemics.

We thank E. Bodenschatz for critical reading of the manuscript and stimulating discussions. This research was supported in part by National Science Foundation Grant PHY99-07949.

- Bernoulli, D. (1760) *Mém. Math. Phys. Acad. R. Sci. Paris*, 1–45.
- Keeling, M. J., Woolhouse, M. E. J., Shaw, D. J., Matthews, L., Chase-Topping, M., Haydon, D. T., Cornell, S. J., Kappey, J., Wilesmith, J & Grenfell, B. T. (2001) *Science* **294**, 813–817.
- Smith, D. L., Lucey, B., Waller, L. A., Childs, J. E & Real, L. A. (2002) *Proc. Natl. Acad. Sci. USA* **99**, 3668–3672.
- Keeling, M. J., Woolhouse, M. E. J., May, R. M., Davies, G & Grenfell, B. T. (2003) *Nature* **421**, 136–142.

- Bullock, J. M., Kenward, R. E & Hails, R. S., eds. (2002) *Dispersal Ecology*, 42nd Symposium of the British Ecological Society (Blackwell, London).
- Murray, J. D. (1993) *Mathematical Biology* (Springer, Berlin).
- Langer, W. L. (1964) *Sci. Am.* **2**, 114–121.
- Noble, J. V. (1974) *Nature* **250**, 726–729.
- Grenfell, B. T., Bjornstad, O. N & Kappey, J. (2001) *Nature* **414**, 695–696.
- Mollison, D. (1991) *Math. Biosci.* **107**, 255–287.
- Rvachev, L. A. & Longini, I. M., Jr. (1985) *Math. Biosci.* **75**, 3–22.

12. Anderson, R. M & May, R. M. (1991) *Infectious Diseases of Humans* (Oxford Univ. Press, Oxford).
13. Gardiner, C. W. (1985) *Handbook of Stochastic Methods* (Springer, Berlin).
14. Donnelly, C. A., Ghani, A. C., Leung, G. M., Hedley, A. J., Fraser, C., Riley, S., Abu-Raddad, L. J., Ho, L.-M., Thach, T.-Q., Chau, P., *et al.* (2003) *Lancet* **361**, 1761–1766.
15. Riley, S., Fraser, C., Donnelly, C. A., Ghani, A. C., Abu-Raddad, L. J., Hedley, A. J., Leung, G. M., Ho, L.-M., Lam, T.-H., Thach, T. Q., *et al.* (2003) *Science* **300**, 1961–1966.
16. Lipsitch, M., Cohen, T., Cooper, B., Robins, J. M., Ma, S., James, L., Gopalakrishna, G., Chew, S. K., Tan, C. C., Samore, M. H., *et al.* (2003) *Science* **300**, 1966–1970.
17. Centers for Disease Control (2003) *Morbid. Mortal. Wkly. Rep.* **52**, 241.