Available online at www.sciencedirect.com
ScienceDirect
Mathematical Biosciences 203 (2006) 301–318
Mathematical Biosciences
www.elsevier.com/locate/mbs

Semi-empirical power-law scaling of new infection rate to model epidemic dynamics with inhomogeneous mixing

Phillip D. Stroud *, Stephen J. Sydoriak, Jane M. Riese, James P. Smith, Susan M. Mniszewski, Phillip R. Romero

Los Alamos National Laboratory, MS F-607, Los Alamos, NM 87545, USA

Received 22 September 2005; received in revised form 11 January 2006; accepted 20 January 2006
Available online 15 March 2006

Abstract

The expected number of new infections per day per infectious person during an epidemic has been found to exhibit power-law scaling with respect to the susceptible fraction of the population. This is in contrast to the linear scaling assumed in traditional epidemiologic modeling. Based on simulated epidemic dynamics in synthetic populations representing Los Angeles, Chicago, and Portland, we find city-dependent scaling exponents in the range of 1.7–2.06. This scaling arises from variations in the strength, duration, and number of contacts per person. Implementation of power-law scaling of the new infection rate is quite simple for SIR, SEIR, and histogram-based epidemic models. Treatment of the effects of the social contact structure through this power-law formulation leads to significantly lower predictions of final epidemic size than the traditional linear formulation.

Keywords: Epidemic model; Epidemic dynamics; Power law; Homogeneous mixing

* Corresponding author. Tel.: +1 505 667 6654; fax: +1 505 665 5283.
E-mail address: Stroud@LANL.gov (P.D. Stroud).

0025-5564/$ - see front matter
doi:10.1016/j.mbs.2006.01.007

1. Introduction

1.1. Thesis

In traditional modeling of disease epidemic dynamics, the expected number of new infections per day per infectious person is taken to be proportional to the fraction of the population that is susceptible. This linear formulation implements the homogeneous-mixing assumption, by which each susceptible person is equally likely to become the next victim. In real social structures, however, some people have a greater chance to receive and transmit disease than others. In an epidemic, highly connected people tend to be infected earlier than less-connected people. As an epidemic progresses, not only are there fewer susceptible people, but those that remain tend to have fewer social contacts. Recent literature suggests a power law formulation for the degree distribution of the social contact structure. We have found that epidemic models can practically incorporate inhomogeneous mixing by taking the number of new infections per day per infectious person to scale as a power (greater than one) of the fraction of the population that is susceptible.

A large-scale simulation, in which the social contact structure of a large urban population is implemented with unprecedented fidelity, enables computation of epidemic dynamics with a realistic social contact structure. Synthetic populations have been constructed for three major US metropolitan areas, and epidemics of avian-related influenza have been simulated on these populations. From these simulations, the number of new infections per day per infectious person is found to scale not linearly with the fraction of the population that is susceptible, but as a power law. High-fidelity simulation can be used to determine the scaling exponent semi-empirically.

The power-law mixing formulation enables traditional epidemic models to capture the effects of social contact structures on epidemic dynamics in an empirical way. This power-law formulation gives significantly different epidemic dynamics than the traditional linear scaling. In particular, for a given basic reproductive number, power-law scaling predicts much less severe epidemics than linear scaling model. This semi-empirical power-law formulation can easily be implemented into existing epidemic models to account for the social contact structure within populations.

1.2. New infections per day in epidemic modeling

From the outset of modern epidemic modeling, the number of new infections per unit time per infectious person has been modeled as proportional to the fraction of the population that remains susceptible [1–3]. This linear scaling is a consequence of the homogenous mixing assumption, wherein every susceptible individual in the population is assumed to have an equal chance of becoming the next victim. Under the homogeneous mixing assumption, the number of new infections per unit time at time t, $q(t)$, is given by

$$q = \alpha IS/P, \tag{1}$$

where $S(t)$ is the number of susceptible persons, $I(t)$ is the number of infectious persons, P is the number of people in the initial population, and the coefficient α is the average number of disease transmissions per day per infectious person that would be expected if the entire population were susceptible.

The homogeneous mixing formulation of the new infection rate is used in traditional few-component *stock-flow* models (e.g., susceptible-infectious (SI), susceptible-infectious-susceptible (SIS), susceptible-infectious-removed (SIR), and susceptible-incubating-infectious-removed (SEIR) formulations), as well as in stock-flow models employing many components to represent disease states, treatment status, and victim behaviors [4,5]. It is also used in histogram-based time-binned disease progression models [6]. Traditional stochastic epidemic models also incorporate the homogeneous mixing assumption [7], making use of the probabilistic interpretation of Eq. (1), in which $q\,\mathrm{d}t$ is the probability that a new infection occurs in $\mathrm{d}t$ about t.

A generalized infection rate of the form

$$q = \alpha I(S/P)^{\nu} \tag{2}$$

with scaling power ν greater than one, provides a semi-empirical formulation in which the homogeneous mixing assumption is relaxed. The actual functional dependence of the new infection rate on the susceptible fraction will obviously be more complicated than Eq. (2). Nevertheless, the use of Eq. (2), with the exponent determined by high-fidelity simulation or historic epidemiology in actual cities, can give much better epidemic dynamic modeling than the homogeneous mixing model. Herd immunity can be incorporated in a straight-forward way:

$$q = \alpha I(S_0/P)(S/S_0)^{\nu}, \tag{3}$$

where S_0 is the initial number of susceptible individuals, but for simplicity, the following analysis will assume fully susceptible populations, i.e., $S_0 = P$.

The power-law scaling of Eq. (2) has appeared in the literature [8], but the exponent ν was restricted to a value less than 1; the exponent was written as 1-b, and b was called the *safety-in-numbers-power*. The formulation was unrelated to social contact structure, nor was data used to place a value on the exponent. The power-law formulation of Eq. (2) has been used in an SIR model [9], but the only result reported was that with an exponent of $\nu = 2$, an epidemic threshold occurs at $\alpha\tau_I = 1$, where τ_I is the average duration of the infectious period.

The coefficient α is related to the basic reproductive number, or can be obtained in alternative ways. The reproductive number, R_0, is the mean number of disease transmissions per index case that would be expected if the population was entirely susceptible [3]. If the disease is uniformly infectious over an infectious period lasting τ_I days, the coefficient would be

$$\alpha = R_0/\tau_I. \tag{4}$$

For many diseases, infectiousness depends on how long a person has been infectious. Defining $i(\tau)$ as the mean number of new cases per day transmitted by a person who has been infectious for τ days (assuming a completely susceptible population), the coefficient would be the average infectiousness of all infectious persons:

$$\alpha = \int_0^{\tau_I} \mathrm{d}\tau I(t,\tau) i(\tau)/I(t), \tag{5}$$

where $I(t)$ is the number of infectious people at time t; $I(t,\tau)\,\mathrm{d}\tau$ is the number of people that have been infectious for time in $\mathrm{d}\tau$ about τ at time t, and τ_I is the maximum infectious period. α will depend weakly on how fast the epidemic is growing or decreasing, which in turn depends on time, or on the susceptible fraction. This dependence is eliminated here by examining a disease model in which infectiousness is constant over the infectious period so that Eq. (4) can be used.

Alternatively, α could be formulated as a product of the number of contacts per person, times the transmission probability per day per contact [10].

1.3. Estimate of the epidemic size with power-law scaling

The epidemic size, F, also known as the attack rate, is defined as the total number of people infected during the epidemic divided by the initial population:

$$F = \int_0^\infty \mathrm{d}t q(t)/P. \tag{6}$$

For homogeneous mixing, with the number of new cases per day given by Eq. (1), the epidemic size satisfies the well-known relation [2]

$$F_H = 1 - \mathrm{e}^{-R_0 F_H}. \tag{7}$$

This result for epidemic size is not affected by the presence or absence of an incubation stage, or by the distribution of infectious stage durations. For $R_0 > 1$, F takes a non-zero value and an epidemic is said to occur.

For the number of new cases per day given by the power-law formulation of Eq. (2), a derivation along the lines of that leading to Eq. (7) obtains the new result that the epidemic size satisfies the relation:

$$F_I = 1 - [1 + (v-1)R_0 F_I]^{-1/(v-1)}. \tag{8}$$

For $v = 2$, the epidemic size is simply $F_{v=2} = 1 - 1/R_0$.

Eqs. (7) and (8) are relations that define the epidemic size as an implicit function of R_0, which can be evaluated with a few Newton–Raphson iterations. Fig. 1 shows the epidemic size as a func-

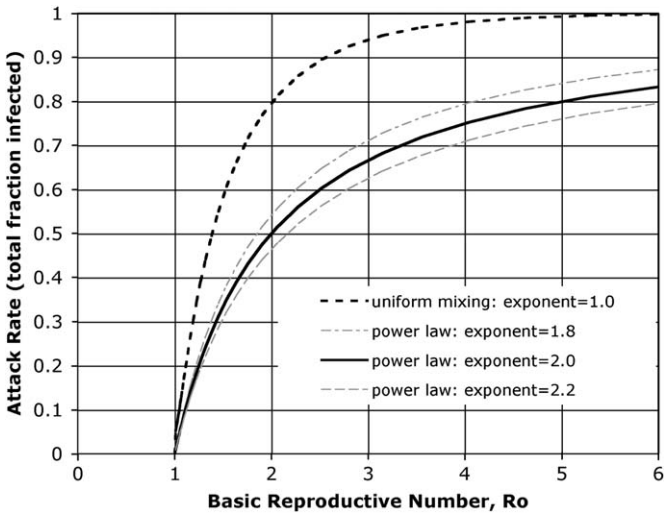

Fig. 1. Epidemic size (i.e., the fraction of the initial population that gets infected during an epidemic) vs. the basic reproductive number, for the homogeneous mixing model (exponent = 1.0) and for the power-law model.

tion of the basic reproductive number R_0, for homogeneous mixing and for the power-law formulation with scaling exponents of 1.8, 2.0, and 2.2. For a given R_0, the power-law formulation leads to a smaller epidemic size than the traditional epidemic formulation would predict. For a scaling power of $v = 2$, the inhomogeneous-mixing model predicts a 50% epidemic size for $R_0 = 2$, while the homogeneous-mixing model predicts 80% epidemic size for the same reproductive number. The homogeneous-mixing model would obtain the same 50% epidemic size with a much less contagious disease (i.e., with $R_0 = 1.38$).

For small epidemics, the epidemic size found by expansion of Eq. (7) or Eq. (8) is $F \sim 2(R_0 - 1)/v$, where $v = 1$ gives the linear homogeneous-mixing formulation. For small epidemics, infecting under a few percent of the population, the power-law formulation predicts an epidemic size of about half (i.e., $1/v$) as large as that predicted by the linear model, for the same R_0.

1.4. Studies of epidemic dynamics with inhomogeneous mixing

Several factors contribute to inhomogeneous mixing whereby some susceptible individuals are more likely than others to become the next victim. The susceptibility (i.e., the likelihood of becoming infected, after receiving a specified dose) can vary from individual to individual. Highly susceptible individuals tend to become infected earlier, leaving the remaining population to be composed of less and less susceptible persons as the epidemic progresses. Although variation in susceptibility is usually modeled with a log-normal distribution, an analysis of epidemic dynamics with an assumed power-law distribution of susceptibility has been conducted [11], but global results for new case rate are not given.

Another potential contributor is the spatial distribution of the infectious source. An analysis of epidemic dynamics with a fractal spatial distribution of the infectious source [12] did not report the scaling of new infections per day per infectious person with respect to the susceptible fraction of the population.

A third mechanism contributing to inhomogeneous mixing is that each individual belongs to a finite number of mixing groups, each of finite size. People in infected mixing groups are much more likely to become the next victim than those not belonging to infected groups. One approach to implementing mixing-group structure into an epidemic dynamics model is to compute disease transmission among hundreds or thousands of individuals, each having been assigned to several pre-constructed mixing groups (representing household, school, work, hospitals, casual contacts, etc.) [13,14]. Each mixing group allows a specified number of persons to interact, and is characterized by a transmission probability per day per fellow-mixing-group-member ranging from high for households, to moderate for schools, to low for neighborhoods. While such an approach could generate the number of new cases per infectious person per day as a function of the remaining susceptible fraction of the population, that result has not been reported. A result that was reported [13] was that for an R_0 value of 3.2 (characterizing smallpox in a fully susceptible population), the expected value of the epidemic size computed for populations of 2000 with a household-school-neighborhood social connection structure is 63%. This epidemic size is much smaller than the 95.2% that would be predicted with the homogeneous-mixing model, but agrees with the corresponding result for a power-law formulation with $v = 2.285$.

In a cellular automata approach to modeling local mixing within a spatial framework, hundreds or thousands of individuals are placed on a 2D lattice, and transmission can occur only between

neighboring individuals [15]. In the absence of long-range connections, outbreaks are highly localized on the 2D grid. Each individual belongs to a mixing group consisting of themselves and eight neighbors. The number of new cases per day per infectious person has not been reported from 2D cellular automata models.

A fourth mechanism contributing to inhomogeneous mixing is that different individuals have differing numbers of social contacts. Recent literature suggests a power law degree distribution, in which the number of people having k contacts is proportional to $k^{-\gamma}$, with γ in the range of 2–3 [16–19]. A minimum number of connections, and an exponential roll-off factor is generally imposed to avoid divergences associated with high and low values of k. Systems of coupled epidemic dynamic equations have been formulated wherein groups consisting of all individuals with k contacts are described by separate dynamic equations [20–23]. These models assume that disease is transmitted and received in proportion to the number of contacts that a person has. Results have been obtained to relate the epidemic size to the contact distribution parameters and the transmission probability per connection. Such analyses have not derived a formulation like Eq. (2), but make such a formulation plausible. A large scale simulation with 7 290 000 individuals connected with a power-law distribution of contacts per person has been used to simulate the dynamics of SARS-like epidemics [24], but the susceptible fraction in the reported runs never drops significantly below unity.

Bond-percolation analysis (in which individuals are connected by a contact graph, each link has a probability of being active, and statistics are evaluated on the sizes of the resulting subgraphs) have been conducted with a power-law distribution of number of contacts [25,26]. The percolation approach does not give epidemic dynamics, but it does obtain epidemic thresholds and the epidemic size as a function of the average transmission probability on a contact. Although the reported results were not directly comparable with the results of this study, further bond-percolation analysis may corroborate the scaling of epidemic size with the power-law exponent.

One more mechanism contributing to inhomogeneous mixing is the variation in the proximity and duration of contacts, and the variation in the ventilation in the room where the contact occurs. These mechanisms have not been analyzed in the literature.

2. Simulated epidemics on realistic social contact structures

2.1. EpiSims

EpiSims is an epidemic simulation tool that explicitly represents every person in a city, and every place within the city where people interact [19,27]. Each person in the simulation is created according to actual demographic distributions drawn from census and other data, so that the synthetic population has the correct demographics, e.g., age distribution, household statistics, residential population density, etc.

A city is represented physically by a set of locations. The city's road network is represented as a graph of road segments and intersections. EpiSims treats each road segment as a location. Each census block (the 8.5 million US census blocks contain on average 34 residents) of the 2000 US census is mapped geographically to one or more road segments, to obtain the number of households, and the number and demographic make-up of the residents associated with the location

representing that road segment. Business directory databases (e.g., Dunn & Bradstreet) provide all business addresses, work type, and number of employees, so that businesses can be assigned to locations. Other data sources are used to map schools, colleges, stores, and places of social recreation to locations. Some businesses and households are treated as separate locations, and some are aggregated to road segment locations.

Each person in the simulation is created with a schedule of activities that they undertake throughout the day. The schedules specify the starting and ending time, the type, and the location of each activity. There are eight types of activity: home, work, shopping, visiting, social recreation, carpooling, school, and college; plus a ninth activity designated other. Information about the time, duration, and location of activities is obtained through surveys.

Each location has one or more *rooms* where the various types of activities take place. Infectious diseases can be transmitted between persons that occupy the same room at the same time. For some activity categories, e.g., home or work, a person does that activity in the same room each day, with the same other people. For other activity categories, such as shopping, the person is assigned to a randomly selected room at the location designated in her schedule. The activity schedules and the locations are statistically the same as those of actual people. An urban mobility simulation [28] computes the travel time between activities, accounting for distance, roads and traffic.

The mixing-group structure, the social contact degree distribution, and the distribution of contact durations emerge empirically from an EpiSims simulation. They are generated by a synthetic population that engages in activities at locations that are statistically the same as those of the actual population. There is no assumption of power-law degree distributions, nor any other theoretical distribution of number of contacts per person.

Synthetic populations have been constructed for the greater metropolitan areas of Portland, Chicago and Los Angeles. The Los Angeles consolidated metropolitan statistical area includes the five counties of Los Angeles, Riverside, San Bernardino, Orange, and Ventura. The combined population, taken from the 2000 US Census, is represented in EpiSims by 16 106 535 individuals. Los Angeles is physically represented as having 562 452 locations, each corresponding to a segment of road between intersections, or to a specific business address. The population of the Chicago CMSA is represented in EpiSims by 8 868 976 individuals, and Portland is modeled with 1 615 864 individuals. The average household size is smaller in Portland than in the other cities, and it also has a disproportionately large population of young adults.

2.2. The influenza disease model

A disease model of pandemic influenza has been constructed for EpiSims based on review literature [29–32]. Influenza is modeled as having a non-contagious incubation stage, followed by an infectious symptomatic stage. The incubation stage sojourn time distribution is described by the half-day histogram $\{0, 0.12, 0.18, 0.259, 0.238, 0.13, 0.07, 0.003\}$, giving respectively the fraction of cases that incubate for a period of between 0 and 0.5 days, 0.5 and 1.0 days, etc. before transitioning to the infectious stage. The average incubation time is 1.9 days. The infectious stage sojourn time distribution is described by the histogram $\{0, 0, 0, 0, 0.005, 0.125, 0.16, 0.205, 0.205, 0.12, 0.08, 0.06, 0.04\}$, giving the fraction of cases that are infectious for 0 to 0.5 days, 0.5 to 1.0 days, etc. The average case is infectious for $\tau_I = 4.1$ days.

The baseline infectiousness (i.e., the probability per hour that a particular susceptible person will become infected, given that there is one symptomatic infectious adult or senior in the same room at the same time) is 0.00285 transmissions per hour. This baseline infectiousness was selected to give an epidemic that infects about 25% of the population. The probability that a susceptible person becomes infected during a visit to a room depends on: how many infectious persons co-occupy the room, how long each contact lasts, the type of activity, and the infectiousness category of the infectious person [19]. Symptomatic children and preschoolers have twice the baseline infectiousness, i.e., 0.0057 transmissions per contact-hour.

The EpiSims influenza model takes that 50% of adults and seniors, 75% of students, and 80% of pre-schoolers will stay at home if they become symptomatic with influenza. These people can then transmit disease only to household members or visitors. The EpiSims influenza model takes that 33.3% of infections are sub-clinical. The sub-clinical manifestation is only half as infectious as the symptomatic manifestations (i.e., 0.001425 transmissions per contact-hour for sub-clinical adults and seniors, 0.00285 for sub-clinical children). Persons with sub-clinical manifestations will continue their normal activities.

Each epidemic starts with ~200 randomly selected individuals becoming infected at time = 0. While EpiSims is designed to explore public health responses, this study examines the scenario that neither vaccine nor effective antiviral treatments are available.

2.3. Simulation results for moderate influenza epidemics

The EpiSims simulation output is a list of all disease-related events, including each new infection, change of location of infected persons, and disease stage transition. By post-processing the event file, the number of new cases can be evaluated for any time interval, and the number of infectious persons can be evaluated for any time. The number of new cases per day per infectious person is computed by dividing the total number of new cases that occur in a 24 hour period by the average number of infectious persons over the same interval. Also, the fraction of the initial population that remains susceptible can be extracted at times of interest.

The number of new cases per day generated in the EpiSims simulation is shown in Fig. 2, for the baseline Los Angeles scenario. Also, the number of currently incubating and infectious people are shown. The epidemic grows from 202 sick persons on day 0 to a peak of 387 902 sick persons on day 128. By day 251, the number of sick persons has dropped to 358. A total of 3 813 253 people get infected during the epidemic, representing a epidemic size of 23.68% of the initial population.

Early in the epidemic, before any significant drop in the number of susceptible individuals, the number of new cases per day grows exponentially, with a growth rate of roughly 7.0% per day. Thus it would require about 75 days for the epidemic to grow from a single case to 200 sick persons.

The spatial distribution of infected persons is illustrated in Fig. 3 at the peak of the simulated Los Angeles epidemic (day 128). Each bar in Fig. 3 indicates the number of infected people at the corresponding location on the map. At 10 a.m., there are 387 901 infected people, at 102 128 locations. By 10 p.m. of day 128, many of the infected people have gone to their household location, so that there are at that time 170 564 locations having at least one infected person.

For the baseline untreated pandemic influenza in Los Angeles, Fig. 4(a) shows the number of new cases per day per infectious person, as a function of the fraction of the initial population that

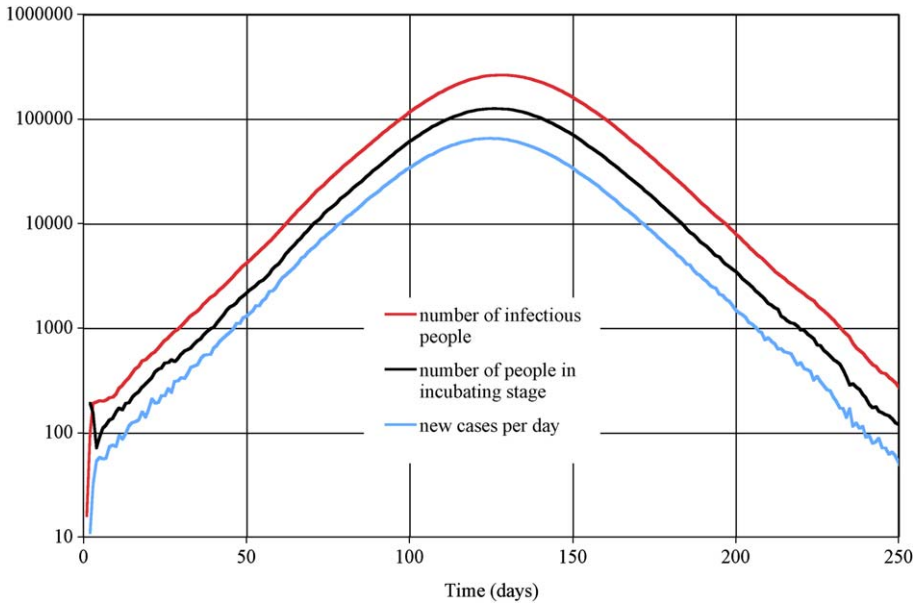

Fig. 2. EpiSims simulation results for an influenza epidemic in Los Angeles, showing the current number of people in the incubating and infectious stages, and the number of new cases infected per day.

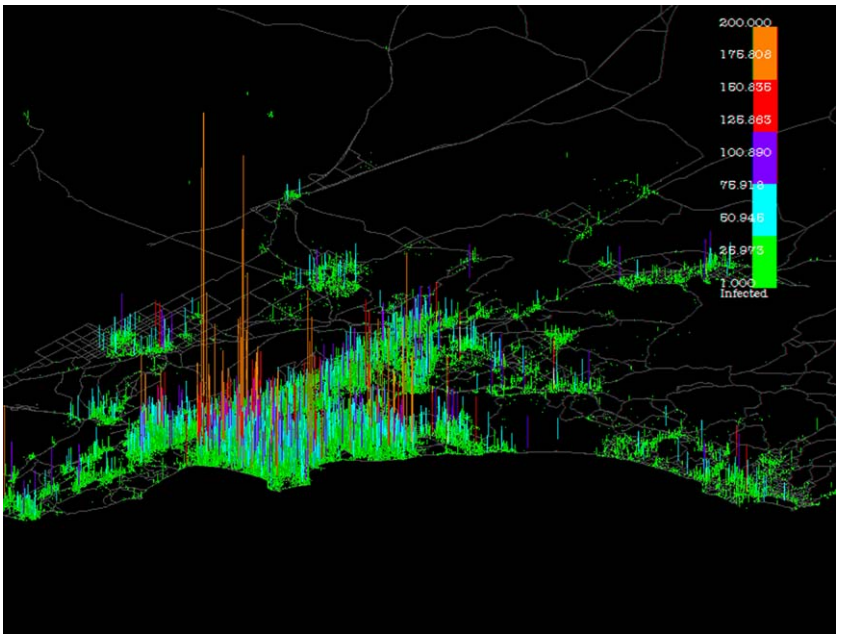

Fig. 3. The spatial distribution of the simulated epidemic in Los Angeles greater metropolitan area, at 10 a.m. of day 128. The view looking north-east extends from Oxnard on the left nearly to San Diego on the right. The height and colour of the bars indicate the number of infected people at a location.

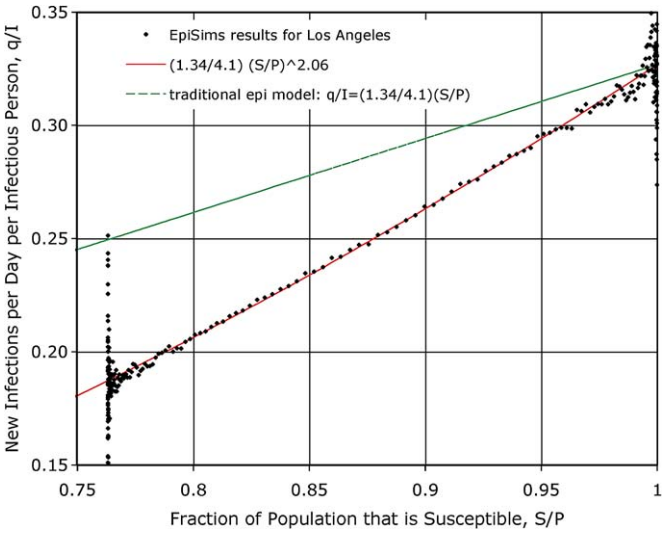

Fig. 4(a). The number of new cases per day per infectious person, as a function of the fraction of the initial population that is still susceptible, from EpiSims simulation of influenza epidemic in Los Angeles, using baseline infectiousness values.

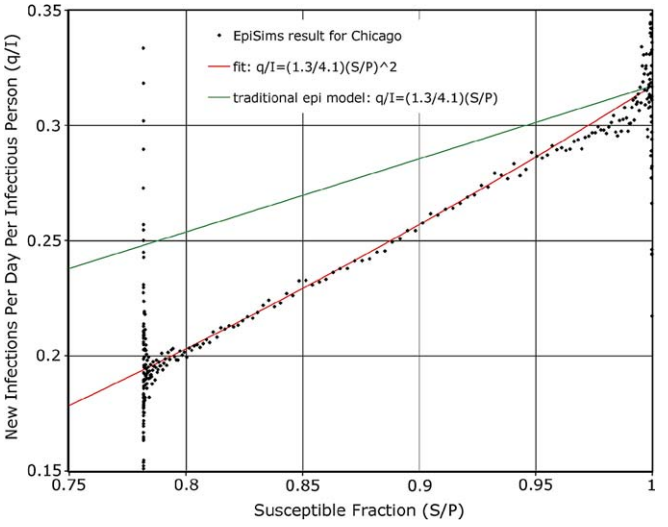

Fig. 4(b). The number of new cases per day per infectious person, as a function of the fraction of the initial population that is still susceptible, from EpiSims simulation of influenza epidemic in Chicago, using baseline infectiousness values.

remains susceptible. The large fluctuations correspond to the beginning and end of the epidemic, when there are relatively few new cases per day. Very early in the epidemic, while the entire population is susceptible, there are $\alpha = 0.327$ new cases per day per infectious person. Since the disease model implements a uniform infectiousness over a 4.1 day average infectious period, the basic reproductive number is $R_0 = \alpha\tau_I = 1.34$. The best power-law fit to the EpiSims simulation is found

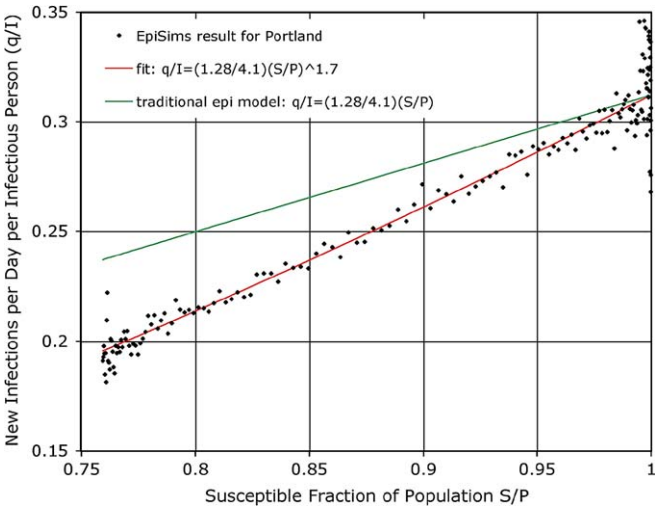

Fig. 4(c). The number of new cases per day per infectious person, as a function of the fraction of the initial population that is still susceptible, from EpiSims simulation of influenza epidemic in Portland, using infectiousness values of 20% above baseline.

to be $q/I = (1.34/4.1)(S/P)^{2.06}$, which is also plotted on Fig. 4(a). Finally, the homogeneous-mixing new case rate per infections person, $q/I = (1.34/4.1)(\mathrm{S/P})$, is also plotted to show how poorly it agrees with the simulation results.

The number of new cases per day per infectious person is shown as a function of the susceptible fraction in Fig. 4(b) for the EpiSims simulation of an influenza epidemic in Chicago. The disease model is identical to that used for the Los Angeles simulation. For Chicago, the best power-law fit is found to be $\alpha = R_0/\tau_I = 1.30/4.1$ and $v = 2.0$. The power-law fit with these parameters is also shown in Fig. 4(b).

The same information is shown in Fig. 4(c) for the EpiSims simulation of an influenza epidemic in Portland. The disease model is identical to that used for the Los Angeles simulation, except that the infectiousness values have been increased by 20% in order to obtain a epidemic size of around 25%. For Portland, the best power-law fit is found to be $\alpha = R_0/\tau_I = 1.30/4.1$ and $v = 1.7$. The power-law fit with these parameters is also shown in Fig. 4(c).

3. Power-law formulation of epidemic models

3.1. Inhomogeneous mixing in SIR epidemic models

The semi-empirical treatment of inhomogeneous mixing can be implemented into the traditional SIR epidemic model:

$$\mathrm{d}S/\mathrm{d}t = -q; \quad \mathrm{d}I/\mathrm{d}t = q - I/\tau; \quad q = (R_0/\tau)I(S/P)^{v}. \tag{9}$$

The epidemic curve obtained by numerical integration of the SIR equations is shown in Fig. 5 for parameters matching the EpiSims simulation of the baseline influenza epidemic in Los Angeles

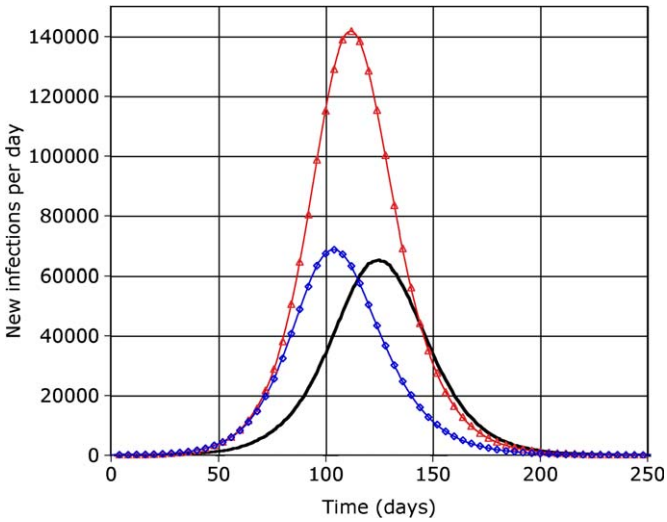

Fig. 5. Linear and power-law scaling implementations of the traditional SIR formulation, for an influenza epidemic in Los Angeles. The epidemic curve from the EpiSims simulation is shown as a solid black line. The SIR model with homogeneous mixing is shown as the thin red line with triangular markers, while the SIR model with power-law mixing is shown as the thin blue line with diamond markers.

($P = 16\,106\,525$, $I(0) = 202$, $\tau = 4.1$ days, $R_0 = 1.34$). The SIR model epidemic curves are given for homogeneous mixing with $v = 1$, and power-law scaling with $v = 2.06$. The power-law scaling provides a much better fit to the detailed simulation than does the linear model. Because the SIR formulation ignores the incubation stage, the epidemic rises faster, peaks sooner, and ends earlier than would be expected with proper treatment of incubation and infectious stage transition dynamics. Nevertheless, the power-law SIR formulation gives a 24.7% epidemic size (in fairly good agreement with the EpiSims result of 23.68%), and produces much better overall agreement than the SIR model with homogeneous mixing.

3.2. Inhomogeneous mixing in SEIR epidemic models

The semi-empirical treatment of inhomogeneous mixing can be implemented directly into the SEIR epidemic model:

$$\mathrm{d}S/\mathrm{d}t = -q; \quad \mathrm{d}E/\mathrm{d}t = q - E/\tau_E; \quad \mathrm{d}I/\mathrm{d}t = E/\tau_E - I/\tau_I; \quad q = (R_0/\tau_I)I(S/P)^v. \tag{10}$$

The epidemic curve obtained by numerical integration of the SEIR equations is shown in Fig. 6 for parameters matching the EpiSims simulation of the baseline influenza epidemic in Los Angeles ($P = 16\,106\,525$, $I(0) = 202$, $\tau_E = 1.9$ days, $\tau_I = 4.1$ days, $R_0 = 1.34$). The SEIR model epidemic curves are given for homogeneous mixing with $v = 1$, and power-law scaling with $v = 2.06$. Because the SEIR formulation imposes an exponential distribution of incubation and infectious stage sojourn times, many people remain sick for much longer than would be expected with proper treatment of incubation and infectious stage transition dynamics. The SEIR model predicts an

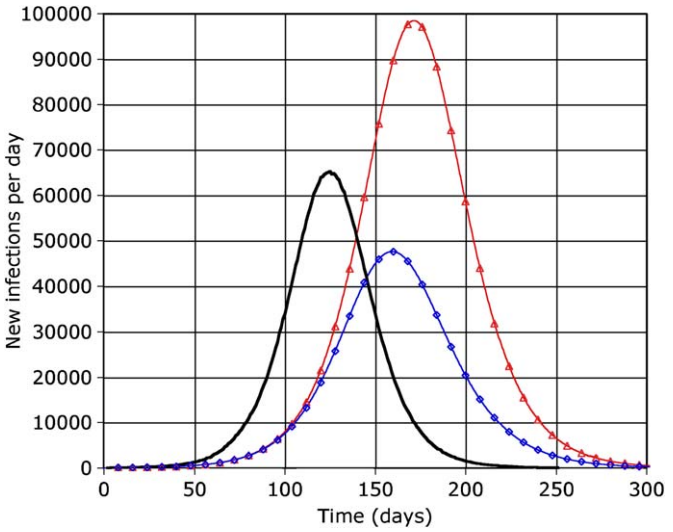

Fig. 6. Linear and power-law scaling implementations of the traditional SEIR formulation, for an influenza epidemic in Los Angeles. The epidemic curve from the EpiSims simulation is shown as a solid black line. The SEIR model with homogeneous mixing is shown as the thin red line with triangular markers, while the SEIR model with power-law mixing is shown as the thin blue line with diamond markers. (For interpretation of the references in colour in this figure legend, the reader is referred to the web version of this article.)

epidemic that peaks later than it should. Nevertheless, the power-law SEIR formulation gives a 24.7% epidemic size (in fairly good agreement with the EpiSims result of 23.68%), and produces much better overall agreement than the SEIR model with homogeneous mixing.

3.3. Inhomogeneous mixing in histogram-based epidemic models

Histogram-based epidemic models use distinct components to identify people that have been incubating or infectious for specified time intervals, compute transitions from data-derived sojourn-time histograms, and obtain epidemic curves by discrete time integration or by discrete event simulation. The power-law formulation can be implemented into a histogram-based model as follows:

$$E_{ij} = E_{i-1,j-1}(1 - g_{j-1/2}), \quad \text{where } g_{j-1/2} = \frac{c_{j+1} - c_{j-1}}{c_j + c_{j-1}} \quad \text{and} \quad c_{j-1} = 1 - \sum_{j'=1}^{j} h_{j'-1/2}. \tag{11}$$

$$I_{i0} = \sum_{j=0}^{j_{\max}} E_{i-1,j} g_{j-1/2} \tag{12}$$

$$I_{ik} = I_{i-1,k-1}(1 - G_{k-1/2}), \quad \text{where } G_{k-1/2} = \frac{C_{k+1} - C_{k-1}}{C_k + C_{k-1}} \quad \text{and} \quad C_{k-1} = 1 - \sum_{k'=1}^{k} H_{k'-1/2}, \tag{13}$$

$$N_i = S_{i-1}(1 - \mathrm{e}^{-q_{i-1}T/S_{i-1}}) \quad \text{for } v = 1 \tag{14a}$$

and

$$N_i = S_{i-1}\left[1 - (1 + (v-1)q_{i-1}T/S_{i-1})^{-1/(v-1)}\right] \quad \text{for } v > 1, \tag{14b}$$

$$q_i = (S_{i-1}/P)^v \sum_{k=0}^{k_{\max}} I_{i-1,k}\alpha_k, \tag{15}$$

$$E_{i0} = N_i, \tag{16}$$

$$S_i = S_{i-1} - N_i, \tag{17}$$

where T is the time step, e.g., 0.5 days; S_i is the number of susceptible individuals at time iT; E_{ij} is the number of individuals at time iT that have been incubating for between jT and $(j+1)T$; I_{ik} is the number of individuals at time iT that have been infectious for between kT and $(k+1)T$; $h_{j-1/2}$ is the expected fraction of infected individuals that incubate for longer than $(j-1)T$ but less than jT; $H_{k-1/2}$ is the expected fraction of cases that are infectious for longer than $(k-1)T$ but less than kT; N_i is the number of new infections that occur between $(i-1)T$ and iT; α_k is the average number of new cases per infectious person per day, assuming the entire population is susceptible and the infectious person has been infectious for time kT.

Fig. 7 shows the epidemic curve obtained from the EpiSims simulation of an influenza epidemic in Los Angeles, the histogram-based epidemic model with homogeneous mixing ($v = 1$), and the

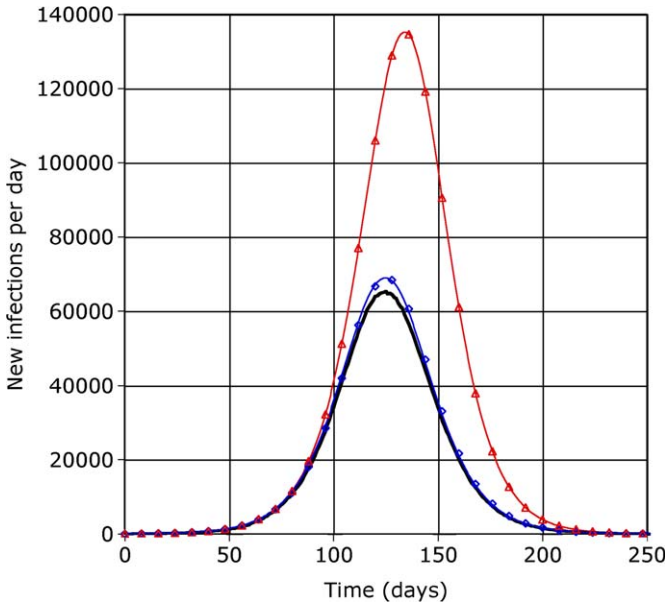

Fig. 7. New cases per day, from the EpiSims simulation of an influenza epidemic in Los Angeles, from the histogram-based epidemic model with linear scaling, and from the histogram-based epidemic model with power-law scaling (with $v = 2.06$, $R_0 = 1.34$). The epidemic curve from the EpiSims simulation is shown as a solid black line. The histogram model with homogeneous mixing is shown as the thin red line with triangular markers, while the histogram model with power-law mixing is shown as the thin blue line with diamond markers. (For interpretation of the references in colour in this figure legend, the reader is referred to the web version of this article.)

histogram-based epidemic model with power-law scaling ($v = 2.06$). All three models use the same initial population (16 106 535), the same number of initial index cases (202), a basic reproductive rate of 1.34 transmissions per index case, equivalent to 0.327 transmissions per day per infectious person, and the same incubation and infectious stage duration histograms.

For all three epidemic curves, the early part of the epidemic is characterized by exponential growth in the new cases per day, with a growth rate of 7% per day. The histogram-based epidemic model with power-law scaling gives an epidemic curve in much better agreement with the EpiSims simulation than does the same model with linear scaling. The epidemic size obtained by the EpiSims simulation is 23.68%. The histogram-based epidemic model with power-law scaling exponent of 2.06 gives a epidemic size of 24.9%. Since the epidemic size for this exponent is given analytically to be 24.70%, the numerical error associated with the histogram-based half-day timestep integrator is seen to be small. The histogram-based epidemic model with linear scaling, however, gives a epidemic size of 46.5%.

The peak new case rate in the EpiSims simulated epidemic in Los Angeles is 65 278 new cases per day, occurring on day 125. The histogram-based epidemic model with power-law scaling exponent of 2.06 gives a peak of 68 942 new cases per day, occurring on day 125. The histogram-based epidemic model with linear scaling, however, gives a much higher peak of 135 194 new cases per day, which occurs on day 134.

4. Severe epidemics

The epidemics described so far are moderate, in the sense that if an index case infects an average of 1.34 persons total, his contact groups will not all become sick within a generation or two. For severe epidemics, in which each index case might infect an average of four or more persons, some of the contact groups (such as his household or workplace) may become completely infected, or nearly so. The secondary cases will then have fewer susceptible contacts, and the average new cases per day per infectious person will drop, relative to that in mild or moderate epidemics.

An EpiSims simulation was conducted under the identical conditions as the Los Angeles epidemic described above, except that the infectiousness values were all increased by a factor of five (i.e., 0.01425 transmissions per hour per contact for symptomatic adults and seniors and for sub-clinical children, 0.0285 for symptomatic children, and 0.007125 for sub-clinical adults and seniors). Since the baseline infectiousness corresponds to an R_0 of 1.34, the fivefold increase in infectiousness gives an R_0 of 6.7. The EpiSims simulated severe epidemic grows exponentially with a exponential growth coefficient of 0.415 per day, until approximately 100 000 new cases per day is attained. The epidemic reaches a peak of 1 268 620 new cases per day some 10.5 days later. The last case occurs 41 days after the peak, by which time 83.2% of the initial population has been infected.

Fig. 8 shows the new cases per day per infectious person for the severe epidemic in Los Angeles. It also shows the power-law formulation of Eq. (2) with $\alpha = R_0/\tau_I = 6.7/4.1 = 1.634$ and $v = 2.06$. This figure illustrates several phenomena that were also observed in EpiSims simulations of severe epidemics in Chicago and Portland. Within two generations of the initial infections, the number of new cases per day per infectious person drops from R_0/τ_I to about 0.70 R_0/τ_I (i.e., starting at 1.634 and dropping to 1.13). We attribute this to saturation of some of the contact groups. The new cases per day per infectious person remains below the power-law formulation until it drops to

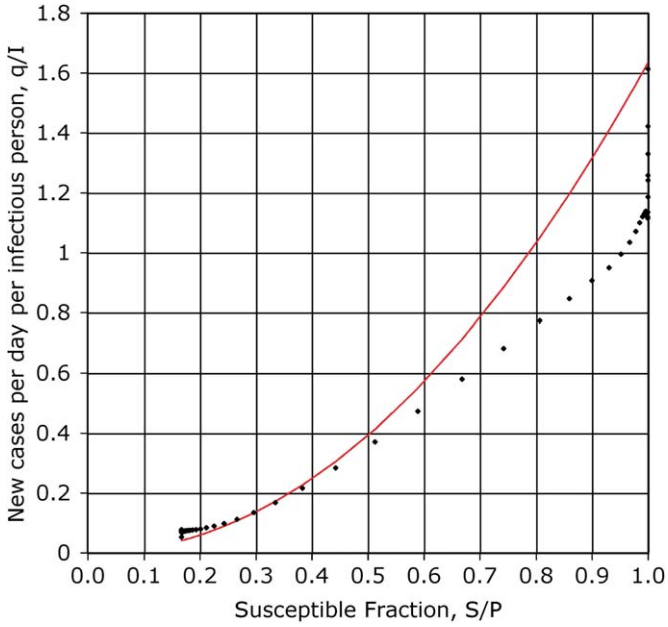

Fig. 8. The number of new cases per day per infectious person for a severe epidemic. The diamonds show the EpiSims simulation daily tabulation. The solid line is the power-law formulation, Eq. (2), with $R_0 = 1.34 \times 5 = 6.7$, $\tau_I = 4.1$ days, and $\nu = 2.06$.

around $1/\tau_I$, by which time it rejoins the power-law formulation. Since the epidemic peaks around the time that the new cases per day per infectious person is equal to $1/\tau_I$, and since roughly half of the cases occur before the peak, the power-law formulation ought to provide a moderately reliable estimate of the epidemic size, even for severe epidemics. It will however predict an earlier peak than the high fidelity simulation. The histogram-based epidemic model using the power-law formulation with $\alpha = 1.634$ and $\nu = 2.06$ predicts a epidemic size of 86.1%. The histogram-based epidemic model using the traditional linear scaling would predict 100% epidemic size.

5. Conclusions

The high-fidelity simulation EpiSims has been used to observe epidemic dynamics on social contact structures that statistically match actual populations of large metropolitan areas, in terms of where and with whom people live, work, play, drive, shop, and attend school. These contact structures emerge from millions of synthetic individuals engaging in realistic daily schedules, sometimes occupying locations simultaneously with other people, thus creating potential for disease transmission. For mild epidemics, the number of new infections per day per infectious person was observed to scale as a power of the susceptible fraction of the population. This result was seen in synthetic populations emulating three different cities, over a range of disease parameters. This power-law dependence is attributed to inhomogeneous mixing: some individuals are more likely to acquire and transmit disease than others.

For Los Angeles, the scaling power was found to be 2.06. The appropriate scaling power depends on the social contact structure of the city of interest. Mild epidemics in Chicago were best fit with a scaling power of 2.0. Portland, however, was best fit with a scaling power of 1.7, possibly due to its smaller average household size and disproportionately large number of young adults. It was also found that for a given disease infectiousness, an index case in Portland would transmit only 83% as many secondary cases on average relative to index cases in either Los Angeles or Chicago.

A semi-empirical formulation has been developed to account for the impact of inhomogeneous mixing in epidemic models. The number of new cases per day per infectious person, which has traditionally been modeled as proportional to the susceptible fraction of the population, is modeled instead as proportional to the susceptible fraction raised to a power. The scaling power is extracted from the results of high-fidelity simulation. This semi-empirical formulation has been implemented into SIR, SEIR, and histogram-based epidemic models. For mild to moderate epidemics, a histogram-based epidemic model with the semi-empirical power-law formulation of inhomogeneous-mixing is able to generate epidemic curves that are quite faithful to the high-fidelity simulation, once the scaling exponent has been discovered. In SIR and SEIR epidemic models with the power-law formulation, the resulting epidemic curves still give timing errors due to other approximations in these models, but the resulting predictions of epidemic size are much improved. Furthermore, a very simple relation has been obtained for the expected epidemic size, given the scaling power and the disease parameters.

Treatment of the effects of the social contact structure through the power-law formulation leads to significantly lower predictions of final epidemic size than the traditional linear formulation. In a range of epidemic size of great importance for public health planning, proper treatment of inhomogeneous mixing might lead to a epidemic size of 25%, while models assuming homogeneous mixing would predict 50% epidemic size for the same disease parameters.

Some interesting phenomena have been observed in high-fidelity simulation of severe epidemics. In particular, rapid saturation of contact groups causes a significant reduction in the number of new infections per day per infectious person. The semi-empirical formulation, while not capturing the early dynamics of severe epidemics, nevertheless generates fairly accurate estimates of their final epidemic size.

References

[1] W.O. Kermack, A.G. McKendrick, A contribution to the mathematical theory of epidemics, Proc. Royal Soc. London, Series A CXV (1927) 700.
[2] N.T.J. Bailey, The Mathematical Theory of Infectious Diseases and its Applications, second ed., Charles Griffin, London, 1975.
[3] R. Anderson, R.M. May, Infectious Diseases of Humans: Dynamics and Control, Oxford University, Oxford, UK, 1991.
[4] E.H. Kaplan, D.L. Craft, L.M. Wein, Emergency response to a smallpox attack: the case for mass vaccination, Proc. Natl. Acad. Sci. 99 (2002) 10935.
[5] S. Del Valle, H. Hethcote, J.M. Hyman, C. Castillo-Chavez, Effects of behavioral changes in a smallpox attack model, Math. Biosci. 195 (2005) 228.
[6] M.I. Meltzer, I. Damon, J.W. LeDuc, J.D. Millar, Modeling potential responses to smallpox as a bioterrorist weapon, Emerg. Infect. Dis. 7 (2001) 959.

[7] S.A. Bozzette, R. Boer, V. Bhatnagar, J.L. Brower, E.B. Keeler, S.C. Morton, M.A. Stoto, A model for a smallpox-vaccination policy, New Engl. J. Med. 348 (2003) 416.
[8] N.C. Severo, Generalizations of some stochastic epidemic models, Math. Biosci. 4 (1969) 395.
[9] V.J. Haas, A. Caliri, M.A.A. Da Silva, Temporal duration and event size distribution at the epidemic threshold, J. Bio. Phys. 25 (1999) 309.
[10] M. Kretzschmar, S. van den Hof, J. Wallinga, J. van Wijngaarden, Ring vaccination and smallpox control, Emerg. Infect. Dis. 10 (2004) 832.
[11] P. Sabatier, P.M. Guigal, D.M. Dubois, Fractals and epidemic process, Int. J. Comput. Anticipat. Syst. 1 (1998) 135.
[12] M.I. Meltzer, The potential use of fractals in epidemiology, Prev. Vet. Med. 11 (1991) 255.
[13] M.E. Halloran, I.M. Longini Jr., A. Nizam, Y. Yang, Containing bioterrorist smallpox, Science 298 (2002) 1428.
[14] J. Epstein, D. Cummings, S. Chakravarti, R. Singa, D. Burke, Toward a containment strategy for smallpox bioterror: an individual-based computational approach, Center on Social and Economic Dynamics working paper number 31, December 2002. Available from: <www.brook.edu/es/dynamics/papers/bioterrorism.pdf>.
[15] B. Eidelson, I. Lustik, VIR-POX: An agent-based analysis of smallpox preparedness and response policy, J. Artif. Soc. Social Simulat. 7 (2004). Available from: <http://jass.soc.surrey.ac.uk/7/3/6.html>.
[16] S.H. Strogatz, Exploring complex networks, Nature 410 (2001) 268.
[17] R. Albert, A.L. Barabasi, Statistical mechanics of complex networks, Rev. Mod. Phys. 74 (2002) 47.
[18] A.L. Barabasi, E. Bonabeau, Scale-free networks, Sci. Am. (May) (2003) 50.
[19] S. Eubank, H. Goclu, A. Kumar, M. Marathe, A. Srinivasan, Z. Totoczkal, N. Wang, Modelling disease outbreaks in realistic urban social networks, Nature 429 (2004) 180.
[20] R.M. May, A.L. Lloyd, Infection dynamics on scale-free networks, Phys. Rev. E 64 (2001) 066112.
[21] Y. Moreno, R. Pastor-Satorras, A. Vespignani, Epidemic outbreaks in complex heterogeneous networks, Eur. Phys. J. B 26 (2002) 521.
[22] R. Pastor-Satorras, A. Vespignani, Epidemic dynamics in finite size scale-free networks, Phys. Rev. E 65 (2002) 035108.
[23] M. Barthelmy, A. Barrat, R. Pastor-Satorras, A. Vespignani, Velocity and heirarchical spread of epidemic outbreaks in scale-free networks, Phys. Rev. Lett. 92 (2004) 178701.
[24] M. Small, C.K. Tse, Modeling the SARS outbreak in Hong Kong with small world or scale free networks, in: Proc. Int. Symp. on Nonlinear Theory and its Applications, Fukuoka, Japan, November 29–December 3, 2004.
[25] M.E.J. Newman, The spread of epidemic disease on networks, Phys. Rev. E 66 (2002) 016128.
[26] Y. Moreno, A. Vasquez, Eur. Phys. J. B 31 (2003) 265.
[27] C.L. Barrett, S.G. Eubank, J.P. Smith, If smallpox strikes Portland, Sci. Am. 292 (March) (2005) 41.
[28] L. Smith, R. Beckman, D. Anson, K. Nagel, M. Williams, TRANSIMS: Transportation analysis and simulation system, Compendium of Papers of 5th Natl. Conf. on Transportation Planning Methods, April 1995.
[29] I.M. Longini, M.E. Halloran, Strategy for distribution of influenza vaccine to high-risk groups and children, Am. J. Epidemiol. 161 (2005) 303.
[30] I.M. Longini, M.E. Halloran, A. Nizam, Y. Yang, Containing pandemic influenza with antiviral agents, Am. J. Epidemiol. 159 (2004) 623.
[31] I.M. Longini, A. Nizam, S. Xu, K. Ungchusak, W. Hanshaoworakul, D.A.T. Cummings, M.E. Halloran, Containing pandemic influenza at the source, Science 10 (2005) 1126-1.
[32] J.M. Hyman, T. LaForace, Modeling the spread of influenza among cities, in: H.T. Banks, C. Castillo-Chavez, (Eds.), Bioterrorism: Mathematical Modeling Applications in Homeland Security, SIAM Series: Frontiers in Applied Mathematics, vol. 28, 2003 (Chapter 10).