Available online at www.sciencedirect.com
ScienceDirect
Social Networks 29 (2007) 539–554
SOCIAL NETWORKS
www.elsevier.com/locate/socnet

Mixing patterns between age groups in social networks

S.Y. Del Valle [a,*], J.M. Hyman [b], H.W. Hethcote [c], S.G. Eubank [d,1]

[a] *D-3, Systems Engineering & Integration (MS B262), Los Alamos National Laboratory, Los Alamos, NM 87545, USA*
[b] *T-7, Mathematical Modeling & Analysis (MS B284), Los Alamos National Laboratory, Los Alamos, NM 87545, USA*
[c] *Department of Applied Mathematical & Computational Sciences, University of Iowa, 14 McLean Hall, Iowa City, IA 52242, USA*
[d] *CCS-5, Discrete Simulation Sciences (MS M997), Los Alamos National Laboratory, Los Alamos, NM 87545, USA*

Abstract

We present a method for estimating transmission matrices that describe the mixing and the probability of infection between age groups. Transmission matrices can be used to estimate age-dependent forces of infection in age-structured, compartmental models for the study of infectious diseases. We analyze the social network generated by the synthetic population of Portland and extract mixing patterns. Our results show that the mixing within the population consists of two groups, children and adults. Children interact most frequently with other children close to their own age, while adults interact with a wider range of age groups and the durations of typical adult contacts are shorter than typical contacts between children. Furthermore, the transmission matrix shows that children are more likely to acquire infection than adults.

Keywords: Infectious diseases; Mathematical models; Network epidemiology; Contact patterns; EpiSimS; WAIFW matrix

1. Introduction

A major determinant in understanding the spread of diseases is our lack of data on the mixing patterns in the population. It is important to appropriately account for the formation of contacts to accurately understand disease spread and develop control measures. Correctly accounting for

* Corresponding author. Tel.: +1 505 665 9286.
E-mail address: sdelvall@lanl.gov (S.Y. Del Valle).
[1] Present address: Virginia Bioinformatics Institute, Virginia Tech, Blacksburg, VA 24061, USA.

0378-8733/$ – see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.socnet.2007.04.005

the mixing patterns in a population may be crucial to accurately predict the path of a disease and thus where the outbreak could be intercepted most effectively.

Mathematical models can estimate the likelihood of a disease outbreak based on the *basic reproduction number*. The *basic reproduction number* ($\Re_0$) is the average number of secondary cases produced by a "typical" infectious individual during its infectious period (Diekmann et al., 1990). The rate at which infectious individuals spread the disease depends on the number of adequate contacts (the contacts that will result in infection) between infecteds and susceptibles. Several studies have shown that averaging the mixing patterns of heterogeneous populations can cause $\Re_0$ to decrease or remain unaltered (Isham and Mdley, 1996). Thus, if we determine the mixing patterns in the population, we can obtain better estimates of $\Re_0$. This result can help modelers predict the severity of an outbreak and the best means of containing it.

Because of the recognition that heterogeneous contact patterns govern sexually transmitted diseases (STDs), several techniques have been developed to incorporate heterogeneous mixing in mathematical models for STDs. However, not much effort has been channeled into incorporating heterogeneous mixing for other infectious diseases. Numerous models have studied the effects of different mixing functions or mixing matrices in the form of compartmental models for STDs (Anderson et al., 1990; Blythe and Castillo-Chavez, 1989; Hyman and Stanley, 1989; Hyman and Li, 1997; Hyman et al., 1999; Knolle, 2004). Some of the techniques developed to incorporate non-random mixing into epidemic models include restricted mixing (Jacquez et al., 1988), proportionate mixing (Hethcote and Van Ark, 1987; Nold, 1980), preferred mixing (Hethcote and Yorke, 1984), selective mixing (Koopman et al., 1989), and non-proportionate mixing (Anderson and May, 1991). These techniques involve defining an $n \times n$ matrix, the elements of which represent adequate contacts between individuals in age group i and age group j. However, these matrices require knowledge of the forces of infection, the mixing structure, and the steady states of the endemic disease. The forces of infection are usually estimated using serological data, but these data are often not available for many diseases.

Survey studies of mixing patterns can be useful tools in understanding disease spread. Several studies have reported data on mixing patterns from different populations and their impact on the spread of sexually transmitted diseases (Aral et al., 1999; Ghani and Garnett, 1998; Laumann and Youm, 1999; Youm and Laumann, 2002), however, only a limited number of studies have reported data on mixing patterns that could lead to the transmission of airborne infections (Beutels et al., 2006; Edmunds et al., 1997, 2006; Wallinga et al., 2006).

Edmunds et al. (1997) studied a sample of 65 individuals and estimated contact patterns that could lead to the spread of airborne infections. They concluded that older adults mix with themselves and all other age groups at the same rate, and that younger adults do not. They also found that people have a different mixing pattern during the weekend than on weekdays. However, some of the limitations of this study are the sample size, the lack of quantification of duration of contact, and the fact that all the participants were adults, even though a great number of diseases are transmitted by children.

A recent study by Wallinga et al. (2006) showed that school-aged children and young adults are more likely to become infected and transmit disease to others than the rest of the population. The also estimated age-specific transmission parameters using data from self-reported surveys. However, this study is limited by recall bias, the lack of quantification of duration of contact, and their definition of contact, which ignores casual contacts (i.e. people in the same room or bus). Nevertheless, their study is a step forward towards understanding the impact of social contacts on disease transmission.

A review article by Wallinga et al. (1999) discussed the use of networks in developing contact patterns and the spread of airborne infectious diseases. They noted that more studies are needed to better understand contact patterns to predict disease spread. A commentary by Halloran (2006) also discusses the need of more empirical studies to estimate transmission parameters for models with complex mixing structures.

Any realistic model for the spread of infectious diseases must take into account the mechanism of its transmission, the heterogenities of risk between individuals, the spatial and local nature of interactions among the population, and the probability of transmission per contact (Isham and Mdley, 1996; Grenfell and Harwood, 1997; Wallinga et al., 1999). However, the standard compartmental models assume homogeneously mixing populations and thus ignore many of these important properties that are crucial for the modeling of human disease-transmission. Therefore, while traditional compartmental models have proved to be useful in developing theoretical epidemiological insights, the complex nature of the mixing patterns in the population and transmission routes could have severe implications.

Diseases are often spread through social contact, thus, contact information is a key to controlling an epidemic. Here we apply tools from social network epidemiology (Meyers et al., 2004; Newman, 2002; Zaric, 2002; Read and Keeling, 2003) to extract mixing patterns that may be responsible for the spread of disease. In social network modeling, the structure of a community is represented by a graph consisting of nodes, representing people, and edges, representing contact between two people. Social networks allow for more realistic representation of populations and their social contact structure.

We use a social network generated by the synthetic population of Portland, Oregon, to determine mixing patterns between age groups. We analyze the simulated movement of these individuals and determine the likelihood of infection. A transmission matrix is estimated based on the mixing patterns observed in the population. This matrix can be used in mathematical models to determine age-dependent forces of infection.

2. Realistic social networks

Computer simulation techniques have increasingly been used to model the spread of infectious diseases within human populations (Ackerman et al., 1984; Adams et al., 1999; Halloran et al., 2002). One of the earliest attempts was made by Ackerman et al. (1984), who introduced a discrete-time, stochastic model for influenza and illustrated the impact of different vaccination strategies. Their influenza model was based on a structured population of 1000 individuals in five age groups with subgroup mixing in families, neighborhoods, preschool play groups, schools, and total community mixing. Halloran et al. (2002) used a similar approach to develop a discrete-time, stochastic simulation model of smallpox spread to compare the effectiveness of various intervention strategies in a community of 2000 people. In contrast to Ackerman's model, Halloran et al. (2002) used the 2000 US Census data to obtain the average age distribution of the population and approximate household sizes. Adams et al. (1999) developed an elaborate stochastic simulation model named GERMS to study the spread of sexually transmitted diseases. Their model estimated in detail the contact patterns for partnership formation and the probability of infection. However, one of the main drawbacks of all these models is the fact that the user inputs the contact patterns and thus the results could be biased by these inputs. EpiSimS (Epidemic Simulation System) allows the contact patterns to emerge from the simulation and it is thus a further evolution of these pioneer attempts.

EpiSimS is a discrete event stochastic simulation model used for the spread of disease in large urban populations (Barrett et al., 2005; Chowell et al., 2003; Del Valle et al., 2006; Eubank,

2002; Eubank et al., 2004, 2006; Stroud et al., 2006a,b). The original EpiSimS model was based on the city of Portland, Oregon, in which the simulated movement of more than 1.6 million individuals was constructed. Each individual in the simulation was instantiated according to actual demographic distributions drawn from census data, so that the synthetic population had the correct demographics, e.g. age distribution, household statistics, population density, etc.

In EpiSimS a region is represented physically by a set of road segment locations and a set of business locations. EpiSimS treats each road segment as two locations (one for each side of the road) and the midpoint of each road segment is geo-located. The 2000 US Census tabulates the population and number of households in each of 8.5 million census blocks using 100% of the surveys. Each census block maps geographically to several road segments (typically the four road segments that go around a city block). EpiSimS distributes accordingly the census block residential population and number of households to these road segment locations. The city of Portland was mapped into 181,230 locations such as households, schools, workplaces, and shopping centers. Information on the number of actual physical locations in Portland were obtained from the Enhanced TIGER data set.

The US Census tabulates demographic and economic characteristics at each of 200,000 Census block groups, using a 5% sample of the surveys. Block group data are tabulated so that only marginal distributions can be extracted (e.g. the distribution of household income). The US Census also tabulates demographic and economic characteristics for each of 65,443 census tracts, again using a 5% sample of the surveys, from which joint distributions can be extracted (e.g. the joint distribution of household income and household size). A statistical procedure called iterative proportional fitting is used to create a synthetic population that matches the marginal distributions at the block group level, while retaining the demographic correlation structure of the Census tract tabulations. It creates the joint distribution, matching the marginal distributions by taking samples from the partial set of full records. The result is a set of households and individuals geographically distributed with correct demographics, statistically indistinguishable from the real population.

For each household in the synthetic population, an appropriate activity pattern is selected from the surveys according to a procedure called binary tree matching on the chosen demographics. For example, the population could be split into households with income above or below $50 K/year, split again according to those with and without children under 5, and so on. In each of these final categories there will be a set of survey households appropriate to the synthetic household, and one is chosen randomly. Surveys are chosen by household not individual, to insure that a household's activities are correctly correlated. Each individual's schedule specifies the starting and ending time, the type, and the location of each assigned activity. There are eight types of activities: home, work, shopping, visiting, social recreation, passenger server (e.g. bus driver), school, and college; plus a ninth activity designated other. Information about the time, duration, and location of activities is obtained from the National Household Transportation Survey.

Publicly available land use data describes the number of users of a location and the type of activity being done. Travel time between any two locations is a simple function of distance and travel mode (i.e. car, walk, bike, etc.), from an external table of typical travel times by time of day, or from a micro-simulation accounting for the interaction of all travelers plans through the day. The surveys are studied to determine which demographic fields optimally classify important characteristics, such as the total time spent per day in each type of activity.

From these three components (synthetic population based on census data, business locations based on business directory data, and activity schedules based on the National household Transportation Survey data), EpiSimS computes which individuals are together at the same location at the same time. The number of people at various locations and at various times varies widely,

from zero up to many thousands. Not every pair of individuals who happen to be at the same location at the same time will be close enough to transmit disease. In EpiSimS, each location is partitioned into one or more rooms where the various types of activities take place. Disease transmission events can only occur between individuals that occupy the same room at the same time. For example, a school location will be sub-divided into classrooms. The households on a city block are represented as a single geo-located location, which is divided into separate rooms, each representing individual households. Each individual belongs to a household, and the sublocation model puts each individual in the correct room when they are at home. 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. However, for other activity categories, such as shopping, the person is assigned to a randomly selected room at the location designated in the schedule.

Each activity type has a room size parameter, specified in the configuration file. The room size parameter represents the typical maximum mixing group size in activities of that type. The room size parameter is used in combination with a partition file to calculate how many rooms there are for each activity type at each location. The partition file is created by pre-processing the schedule file into an hour-by-hour synopsis of how many people are engaged in each activity at each location. The number of rooms is computed as the maximum hourly occupants at a location for a given activity type, divided by the room size parameter for that activity type. The parameter values used in the configuration file for this analysis are, school: 8, college: 12, shop: 3, work: 13, and social recreation: 6. Some rooms may contain individuals engaging in different activities: a teacher is assigned to each school room, visitors are temporarily added to households, and individuals with an activity assigned as "other" are added to random non-home rooms.

The simulation starts by creating a network represented as a bipartite graph, G_{PL}, where P describes people and L locations (Fig. 1). Each person and location is represented as a vertex in the network and the edges represent movement of individuals between locations. Each edge in G_{PL} has different weights, which represent duration in each location. The simulation keeps

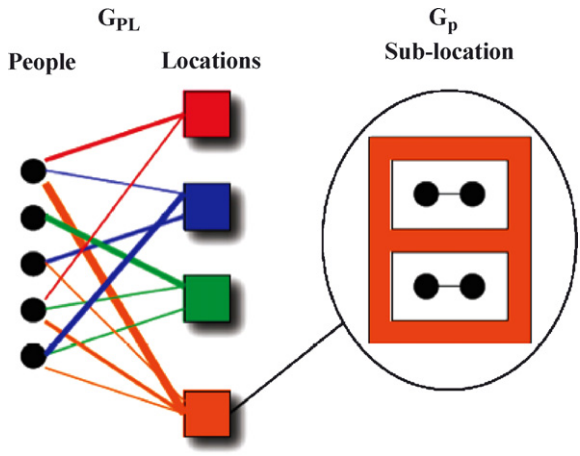

Fig. 1. Schematic representation of EpiSimS social contact network. A bipartite graph G_{PL} with two types of vertices representing five people and four locations. The edges connecting the people and the locations represent movement of individuals between locations throughout the day. The thickness of the edges represents the weight or time spent at different locations. A location is divided into rooms and each contact between two people in a room is represented by an edge. The result is G_P, a person–person social network.

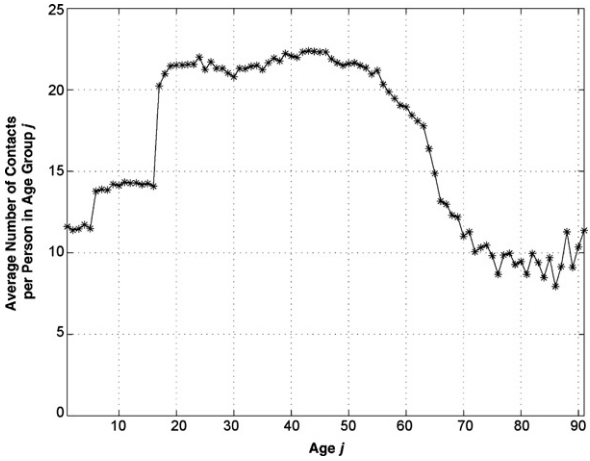

Fig. 2. Daily average number of contacts per person in age group j. The average number of contacts per person is defined by dividing the total number of contacts made by age group j by the size of age group j.

track of every single individual on a second-by-second basis and is therefore able to determine the contacts, including identities of those in contact, the location, the duration of the contact, and the nature of the activity where the contact took place.

EpiSimS integrates all this information into a computer model in order to provide estimates of physical contact patterns for large human populations. Because we are interested in human contact patterns and their impact on the spread of infections diseases, we focus on G_P, a graph containing only people (Fig. 1). The simulation generates time dependent person–person social contact networks based on the sequence of activities each person carries out throughout the day.

We begin by analyzing some of the properties of the social contact network, G_P, generated by the synthetic population of Portland over 1 day. The G_P graph consists of 1,615,860 nodes (e.g. people) and 13,143,900 edges (e.g. connections). The G_P graph is not fully connected but has a giant component of 1,591,010 people and an average degree of 16.52. We estimated the average number of contacts per person in each age group (Fig. 2). In general, older adults have fewer contacts than children and middle aged adults. The average number of contacts generated by the synthetic population of Portland, Oregon is consistent with previous studies (Edmunds et al., 1997; Wallinga et al., 2006). The average number of people contacted per person can give us an estimate of how many secondary cases can potentially acquire infection from one index case. Note that the properties of this network will most likely change in the presence of an infectious disease due to isolation measures and changes in behavior by the affected population (Del Valle et al., 2005).

Although the parameter values used to estimate the average number of people per room were obtained from different data sources, there is still some uncertainty in their values. For example, we used the average student–teacher ratio to determine the average classroom size in Portland, Oregon. However, the student–teacher ratio varies widely across Portland and thus, you may loose some of the heterogeneity that is present in the population. Therefore, we performed sensitivity analysis to determine the effects of changes in the "sub-location" model on the simulation results. Our simulation results (Robbins et al., 2006) show that the average degree (average number of contacts per person) is sensitive to variations in size of the sub-location. That is, when the number of people present in each room increases, the average degree increases and vice versa. However, we

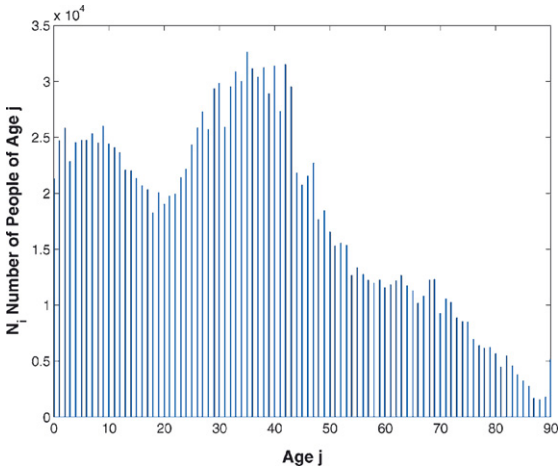

Fig. 3. Age distribution of the synthetic population for the city of Portland. The population consists of 1,615,860 individuals of ages ranging from 0 to 90 years. The population is described by a double-hump distribution with mean of 34.37 and median of 33.

found that the spread of the disease is most sensitive to changes in the size of work sub-locations. This might be due to the population makeup, since over 60% of the population are between 19 and 64. We also found that the mixing structure is robust to changes in the sub-location model. That is, there are always two blocks of mixing and the weak coupling between children and adults.

2.1. Population

EpiSimS uses a synthetic population that resembles the real population of Portland, Oregon in the course of carrying out their daily activities over one randomly chosen day. Fig. 3 shows a histogram of the age distribution of the population. Portland is somewhat unusual because of the disproportionately large population of young adults, resulting in a double-hump distribution.

2.2. Daily number of contacts

A contact was defined as being present in the same sub-location (e.g. room) which is relevant for the spread of airborne infections. Nevertheless, each contact has a weight or duration (described in the next section), which in turns modifies the probability of transmission.

We estimated the total number of contacts between age groups, C_{ij}, by using the G_P social contact network. The total number of contacts between age groups for the social network of Portland is given in Fig. 4. Notice that this matrix is symmetric, so that $C_{ij} = C_{ji}$ for all i and j. That is, the number of contacts that people of age i have with people of age j is the same as the number of contacts that people of age j have with people of age i.

We also determined the average number of contacts per day of a person in age group i with people in age group j, by dividing the total number of contacts C_{ij}, by the total population size N_i in age group i. The resulting $n \times n$ matrix is defined as γ_{ij}.

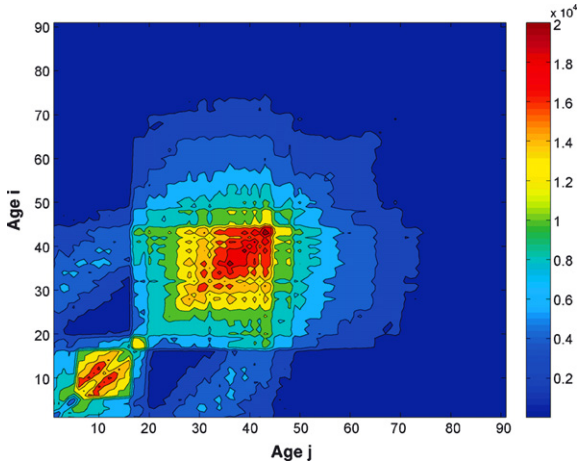

Fig. 4. The total number of contacts between age groups in EpiSimS. The contact rates are defined by the elements of the $n \times n$ matrix, C_{ij}, where C_{ij} represents the total number of contacts of all people of age i with people of age j per day. Observe two blocks of mixing and a weak coupling between adults and children.

2.3. Duration of contacts

Although duration of contact plays an important role in the spread of infectious diseases, to our knowledge, this mechanism has not been analyzed in the literature. The duration of each contact is as important as the number of contacts. The duration is defined as the total length that two people spent together in the same sub-location. If a person has contact with the same person several times a day, all the contact durations of the multiple encounters are added up and the total aggregated length makes the final contact duration. See Heesterbeek and Metz (1993) for theoretical explanations of the effects of contact duration on epidemic spread. In the simulation, people have an average of 16 contacts per day and Fig. 5 shows the distribution of these contact

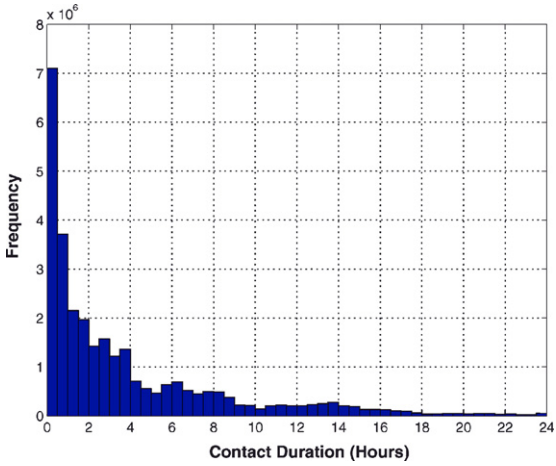

Fig. 5. The distribution of contact duration among people in Portland, Oregon. There are many short-duration contacts representing casual interactions. Structure related to the daily activities of the population is evident.

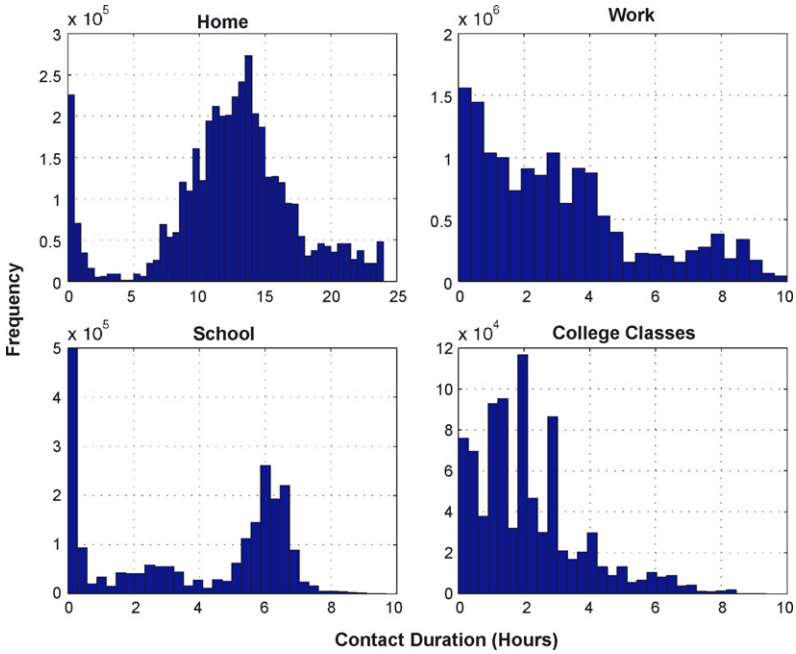

Fig. 6. The distribution of the average contact duration in each of the four top-level activity categories. The average and standard deviation for these contacts are given in Table 1.

Table 1
Breakout of average duration per contact by activity category

Activity category	Average duration	Standard deviation
Home	12 h 24 min	5 h 8 min
School	3 h 47 min	2 h 40 min
Work	3 h 4 min	2 h 29 min
College	2 h 8 min	1 h 37 min
Social recreation	59 min	58 min
Visiting	37 min	49 min
Other	33 min	1 h 8 min
Shopping	30 min	36 min
Passenger server	6 min	12 min

durations. There are many short-duration contacts (four contacts per day per person are less than 30 min) representing casual interactions. Structure related to the daily activities of the population (e.g. 6 h school related activities and 8 h work shifts) is readily apparent.

Fig. 6 shows the distributions of contact durations in each of the four top-level EpiSimS activity categories. Table 1 shows the average contact duration by activity category. The longest contacts occur at home, with an average contact duration of over 12 h, followed by school and work (our results are consistent with Hamermesh et al., 2005 and Aguiar and Hurst, 2006).

Given the duration of each contact, we estimated the total duration of all contacts between age groups, D_{ij}. Notice that this matrix is also symmetric, so that $D_{ij} = D_{ji}$ for all i and j. Then, we estimated the average duration of contact in hours that someone of age i has with someone of age j, T_{ij}, by dividing the total duration of all contacts by the total number of contacts. That is,

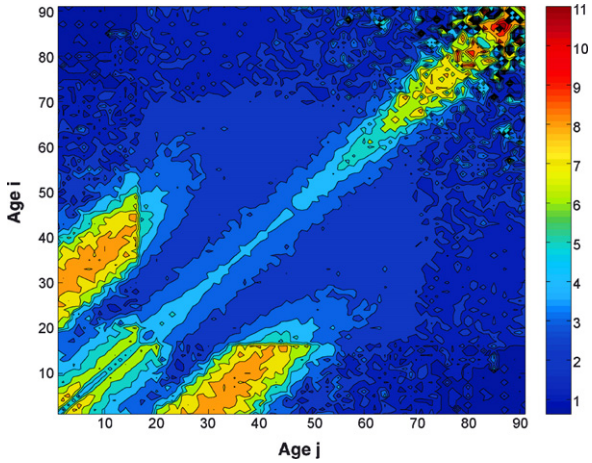

Fig. 7. The average duration of contact per day in hours between people of age i and age j, T_{ij}. The two most striking features of T_{ij} are that people interact with others of similar age and children also interact with their parents.

$T_{ij} = D_{ij}/C_{ij}$, where D_{ij} is the total duration of all contacts between people of age i and people of age j, and C_{ij} is the total number of contacts between people of age i and people of age j. The symmetric contact matrix T_{ij} is shown in Fig. 7.

3. Estimating the transmission matrix

Here we demonstrate how the social contact network generated by EpiSimS can be used to estimate a transmission matrix. The force of infection λ_i is the relative rate at which susceptibles of age i acquire infection. For heterogeneous mixing, the forces of infection reflect the age-related differences in the degree of mixing and contact, within and among age groups.

Empirical evidence of age-related differences in λ_i have been documented for several childhood infections by Anderson and May (1982, 1991) and Grenfell and Anderson (1985), Farrington et al. (2001), and Shkedy et al. (2006). They estimated forces of infection as a function of age using serological data or records of case notifications. These studies suggest that the age-related differences in the force of infection are important factors in modeling infectious diseases. Their results show that for human diseases the force of infection λ_i, tends to increase with age up to about 5–15 years, and then to decrease in later years. Farrington et al. (2001) also showed how to estimate the basic reproduction number using the forces of infection for different diseases obtained from serological survey data.

The standard method used in mathematical models to take account of age-dependent mixing patterns of the population is to define a transmission matrix or WAIFW (who acquires infection from whom) matrix (Anderson and May, 1991; Kanaan and Farrington, 2005). The WAIFW matrix describes how individuals mix with other age groups. The elements of the WAIFW matrix, β_{ij}, represent the rate at which an infective of age j will infect a susceptible of age i. However, this technique requires knowledge of the forces of infection (obtained from serological data and often not available), the mixing structure, and the steady states of the endemic disease (Anderson and May, 1991). Nevertheless, the pre-judgment of the mixing structure may be unrealistic, so that the data leads to mixing matrices with negative entries. Therefore, there is a great need to develop new methods to estimate age-dependent forces of infection.

We use the social network of the synthetic population of Portland to estimate a transmission matrix and, consequently, age-dependent forces of infection. We assume that the population is demographically divided into different age groups that can progress through various infection stages (Hyman et al., 1999). For this model, we consider 91 age groups and m infection stages.

We define the force of infection λ_i as the rate of disease transmission from infected people in all age groups to susceptibles in age group i. That is, λ_i is the sum of the rate of disease transmission from all infection stages in all age groups for age groups $1 \leq j \leq 91$ and infection stages $1 \leq k \leq m$, to the susceptible group, S_i. This means that a susceptible person in age group i can get infected by a person in any infection stage in any age group. Thus

$$\lambda_i = \sum_{j=1}^{91} \sum_{k=1}^{m} \lambda_{ijk}(t). \tag{1}$$

where λ_{ijk} is the rate of disease transmission from the infected people I_{jk} in infection stage k of age group j to the susceptibles in age group i. We calculate λ_{ijk} in (1) as the product of the number of contacts per unit time that each individual in age group i has with age group j; the probability of disease transmission per contact between an infected in stage k of age group j and a susceptible in age group i; and the fraction of those contacts that are infected. That is

$$\begin{array}{lccc} & \text{(Number of)} & \text{(Probability of)} & \text{(Fraction of)} \\ \lambda_{ijk} = & \text{(Contacts per)} & \text{(Disease Transmission)} & \text{(Contacts that)} \\ & \text{(Unit Time)} & \text{(per Unit Time)} & \text{(are Infected)} \end{array}$$

In terms of the EpiSimS data, we can define the force of infection λ_{ijk} as the product of the average number of contacts, γ_{ij}; the probability of disease transmission, which is the product of the susceptibility (α_i) of a susceptible in age group i, the infectivity (ξ_{jk}) of an infective in stage k of age group j, and the probability of transmission P_{ij} based on the average duration of contacts between age groups i and j and the fraction of contacts that are infected. That is

$$\lambda_{ijk} = (\gamma_{ij}(t)) \quad (\alpha_i \xi_{jk} P_{ij}) \quad \left(\frac{I_{jk}(t)}{N_j(t)}\right), \tag{2}$$

where I_{jk} is the number of people in infection stage k of age group j and N_j is the size of age group j. Let σ be the mean number of transmission events per hour of contact between fully infectious and fully susceptible people. For events that occur randomly in time, the number of occurrences in a period of time of length t obeys a Poisson probability law with parameter σt. Thus, the probability of no occurrences in time interval t is $e^{-\sigma t}$ and the probability of at least one occurrence is $1 - e^{-\sigma t}$. Using the mean duration T_{ij} of contacts between a person in age group i with people in age group j, we assume that the probability of transmission in this time interval T_{ij} is given by

$$P_{ij} = 1 - e^{-\sigma T_{ij}}. \tag{3}$$

We can fit this model to past epidemics to estimate the transmissibility parameter, σ. That is, if we know $\Re_0$ for a particular disease, we can estimate σ (Chowell et al., 2007). Using Eq. (3) with $\sigma = 0.2$ and the average durations of contact per pair, T_{ij} we obtain the probability of transmission based on average duration of contacts for all age groups, P_{ij} (Fig. 8). Because of the larger average duration of contact among people of the same age, the probabilities of

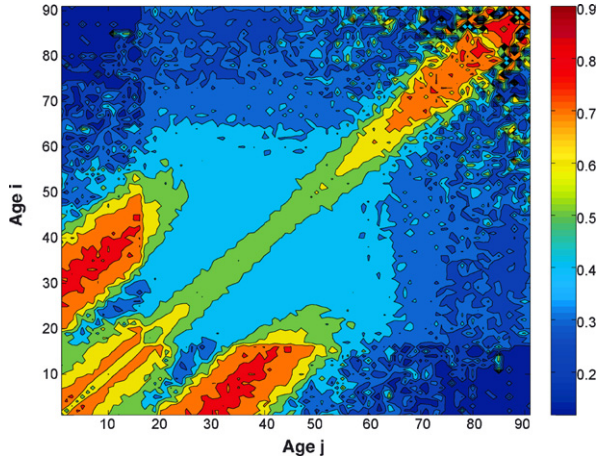

Fig. 8. P_{ij}, probability of transmission based on duration of contacts between a susceptible in group i and an infected in group j. Notice that there is a high probability of transmission along the diagonal for all age groups.

transmission are high along the diagonal for all age groups. Furthermore, observe a weak coupling between middle aged adults and young individuals, possibly due to child-parent duration of contact.

Lastly, we define β_{ij} as the rate of disease transmission between a susceptible in age i with people in age j, which is the product of the average number of contacts, the susceptibility, the infectivity, and the probability of disease transmission, that is, $\beta_{ij} = \gamma_{ij} \times \alpha_i \times \xi_{jk} \times P_{ij}$. Using γ_{ij}, $\alpha_i = 1$, $\xi_{jk} = 1$, and P_{ij} (Fig. 8), we estimate a transmission matrix β_{ij} (Fig. 9), for the given social network. For simplicity, we assume in this example that all age groups are equally susceptible ($\alpha_i = 1$), and that all infected individuals are equally infectious ($\xi_{jk} = 1$), regardless of the infection stage or age group. The transmission matrix, β_{ij} shows the rate at which an infected person in age group j will infect a susceptible person in age group i. The transmission matrix in

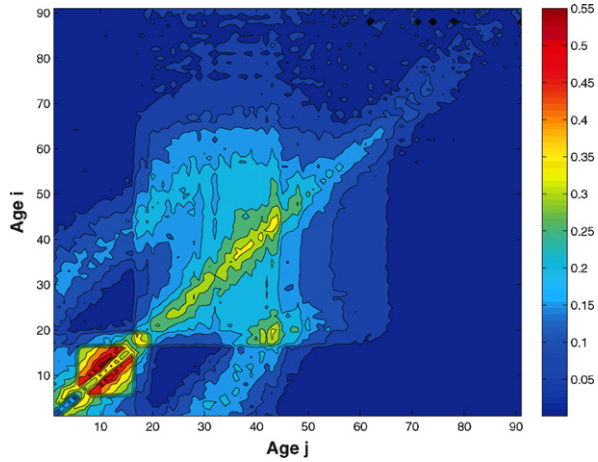

Fig. 9. Transmission matrix β_{ij} (WAIFW). Observe that children and teenagers are more likely to acquire infection than the rest of the population.

Table 2
Transmission matrix (WAIFW) of the daily number of adequate contacts per person between the aggregated age groups

Age	0–4	5–12	13–19	20–29	30–39	40–49	50–59	60–69	70–90
0–4	1.23	1.94	0.72	1.23	1.57	0.41	0.18	0.09	0.07
5–12	1.17	3.37	1.52	0.51	1.63	0.79	0.16	0.08	0.06
13–19	0.58	2.01	1.91	1.06	1.67	1.99	0.61	0.27	0.16
20–29	0.6	0.41	0.65	2.50	2.29	2.01	1.12	0.54	0.26
30–39	0.62	1.07	0.79	1.80	2.86	1.93	0.95	0.51	0.24
40–49	0.19	0.60	1.15	1.93	2.36	2.53	1.11	0.53	0.28
50–59	0.12	0.22	0.61	1.87	2.06	1.9	1.57	0.67	0.31
60–69	0.10	0.14	0.33	1.09	1.32	1.09	0.77	0.99	0.45
70–90	0.06	0.10	0.12	0.49	0.57	0.54	0.35	0.40	0.92

Note that the table is not symmetric because the average time a person of age i spends with a person of age j is not the same as the time a person of age j spends with a person of age i. For example, many adults live in households without children, but children do not live in households without adults.

Fig. 9 exhibits two blocks of mixing, young individuals (< 20 years) and adults (> 20 years). This matrix is consistent with the prevailing opinion that the probability of disease transmission between children is high. Adults are described as likely to acquire infection from a wider range of age groups, mainly middle aged groups.

Moreover, we partition the population into nine categories: 1–5, 6–12, 13–18, 19–30, 31–40, 41–50, 51–60, 61–80, and 81–91, and aggregate the elements of each category using the data that generated Fig. 9. Table 2 shows the average values for the transmission rates for the aggregated age groups. The transmission rates represent the daily probability that an infected person of age j will transmit the disease to a fully susceptible population.

The transmission matrix β_{ij} can be used to obtain the forces of infection needed in a mathematical model with age structure. In order to estimate the force of infection for an specific disease, one would need to estimate the susceptibilities (α_i) for each age group, the infectivities (ξ_{jk}) and the transmissibility parameter (σ) for the disease.

4. Conclusions

Contact patterns play an important role in determining the progression of epidemics. We have introduced a method for obtaining useful information on the mixing patterns of a social contact network, which might lead to the spread of airborne infections. We argue that mathematical models that use contact matrices based on social networks will be better able to capture age-specific infection patterns of infectious diseases than models that use transmission parameters based on homogeneous mixing or ad hoc assumptions.

Estimating forces of infection is crucial when using models for specific infectious diseases. The forces of infection determine the rate of disease transmission and are based on the age-related differences in the degree of mixing and contacts within the population. We used the average number of contacts and a probability distribution based on the average duration per contact to estimate a transmission matrix. With the appropriate specification of disease-related parameters of susceptibility and infectivity, this matrix can be used to estimate age-dependent forces of infection for any disease.

Our results show that in general there are two main blocks of mixing within the population: young individuals (< 20 years) and adults (> 20 years). Furthermore, we observe a weak

coupling between children and middle aged adults, probably due to child-parent contacts. The transmission matrix in Fig. 9 shows that school children are more likely to become infected than the rest of the population. This may be due to long duration of contacts children have with other children at school. In contrast, adults interact with a wider range of age groups, but their duration of contact is shorter.

Our study is limited because the current version of EpiSimS does not stratify classrooms by age and therefore the probability of having a contact with any age group in each school is the same for all children. However, children attending elementary schools mix more in their classrooms with other school children of their own age than with children of other ages. The spread of many childhood diseases is governed by the pattern of contact among children and therefore it is important to incorporate realistic mixing patterns. While recognizing some of the limitations in the current EpiSimS simulation model, EpiSimS represents a potentially powerful resource in the face of an actual outbreak.

For mathematical models of infectious diseases to be useful in guiding public health policy, they must consider age-dependent forces of infection. Individual behavior is crucial for the spread of infectious diseases and predicting disease spread is difficult. Therefore, new techniques such as the one developed here are needed as alternative tools when aggregate behavior cannot be applied to the population. The adequate-contact matrix developed is useful in providing estimates of the age-dependent forces of infection for mathematical models. However, much more needs to be known about the interactions between people that lead to infection before it will be possible to accurately predict an epidemic.

Acknowledgments

We acknowledge the work of all the members of the EpiSimS team and we thank the editor and referee for their valuable comments. This research has been supported through the Mathematical Modeling and Analysis Group and the Systems Engineering and Integration Group at Los Alamos National Laboratory under the Department of Energy contract DE-AC52-06NA25396.

References

Ackerman, E., Elveback, L.R., Fox, J.P., 1984. Simulation of infectious disease epidemics. Charles Thomas, Illinois.

Adams, A.L., Koopman, J.S., Chick, S.E., Yu, P.J., 1999. GERMS: An epidemilogic simulation tool for studying geographic and social effects on infection transmission. Proceedings of the 1999 Winter Simulation Conference 2, 1549–1556.

Aguiar, M., Hurst, E., 2006. Measuring trends in leisure: The allocation of time over five decades. Federal Reserve Bank of Boston 06–2, 1–61.

Anderson, R.M., May, R.M., 1982. Population dynamics of infectious diseases: Theory and applications. Chapman & Hall, London.

Anderson, R.M., Gupta, S., Ng, W., 1990. The significance of sexual partner contact networks for the transmission dynamics of HIV. Journal of Acquired Immune Deficiency Syndromes 3, 417–429.

Anderson, R.M., May, R.M., 1991. Infectious diseases of humans. Oxford University Press, Oxford.

Aral, S.O., Hughes, J.P., Stoner, B., Whittington, W., Handsfield, H.H., Anderson, R.M., Holmes, K.K., 1999. Sexual mixing patterns in the spread of gonococcal and chlamydial infections. American Journal of Public Health 89, 825–833.

Barrett, C.L., Eubank, S.G., Smith, J.P., 2005. If smallpox strikes Portland. Scientific American 292, 54–61.

Beutels, P., Shkedy, Z., Aerts, M., Van Damme, P., 2006. Social mixing patterns for transmission models of close contact infections: exploring self-evaluation and diary-based data collection through a web-based interface. Epidemiology and Infection 134 (6), 1158–1166.

Blythe, S.P., Castillo-Chavez, C., 1989. Like-with- like preference and sexual mixing models. Mathematical Biosciences 96, 221–238.

Chowell, G., Hyman, J.M., Eubank, S., Castillo-Chavez, C., 2003. Laws for the movement of people between locations in a large city. Physical Review E 68, 1–7.

Chowell, G., Nishiura, H., Bettencourt, L.M.A., 2007. Comparative estimation of the reproduction number for pandemic influenza from daily case notification data. Journal of The Royal Society Interface 4, 155–166.

Del Valle, S., Hethcote, H., Hyman, J.M., Castillo-Chavez, C., 2005. Effects of behavioral changes in a smallpox attack model. Mathematical Biosciences 195, 228–251.

Del Valle, S.Y., Stroud, P.D., Smith, J.P., Mniszewski, S.M., Riese, J.M., Sydoriak, S.J., Kubicek, D.A., 2006. EpiSimS: Epidemic Simulation System. Los Alamos National Laboratory Unclassified Report, 06-6714.

Diekmann, O., Heesterbeek, J.A.P., Metz, J.A.J., 1990. On the definition and the computation of the basic reproduction ratio $\Re_0$ in models for infectious diseases in heterogeneous populations. Journal of Mathematical Biology 28, 365–382.

Edmunds, W.J., O'Calaghan, C.J., Nokes, D.J., 1997. Who mixes with whom? A method to determine the contact patterns of adults that may lead to the spread of airborne infections. Proceedings of the Royal Society of London B: Biological Sciences 264, 949–957.

Edmunds, W.J., Kafatos, G., Wallinga, J., Mossong, J.R., 2006. Mixing patterns and the spread of close-contact infectious diseases. Emerging Themes in Epidemiology 3, 10.

Eubank, S., 2002. Scalable, efficient epidemiological simulation. Proceedings of the 2002 ACM Symposium on Applied Computing, 139–145.

Eubank, S., et al., 2004. Modeling disease outbreaks in realistic urban social networks. Nature 429, 180–184.

Eubank, S., Anil Kumar, V.S., Marathe, M.V., Srinivasan, A., Wang, N., 2006. Structure of social contact networks and their impact on epidemics. AMS-DIMACS Special Volume on Epidemiology.

Farrington, C.P., Kannaan, M.N., Gay, N.J., 2001. Estimation of the basic reproduction number for infectious diseases from age-stratified serological survey data. Applied Statistics 50, 251–283.

Ghani, A.C., Garnett, G.P., 1998. Measuring sexual partner networks for transmission of sexually transmitted diseases. Journal of the Royal Statistical Society: Series A 161, 227–238.

Grenfell, B.T., Anderson, R.M., 1985. The estimation of age related rates of infection from case notifications and serological data. Journal of Hygiene 95, 419–436.

Grenfell, B.T., Harwood, J., 1997. Containing bioterrorist smallpox. (Meta)population dynamics of infectious diseases 12, 395–399.

Halloran, M.E., Longini, I.M., Nizam, A., Yang, Y., 2002. Containing bioterrorist smallpox. Science 298, 1428–1432.

Halloran, M.E., 2006. Invited commentary: Challenges of using contact data to understand acute respiratory disease transmission. American Journal of Epidemiology 164 (10), 945–946.

Hamermesh, D.S., Frazis, H., Stewart, J., 2005. Data Watch: The American Time Use Survey. Journal of Economic Perspectives 19, 221–232.

Heesterbeek, J.A.P., Metz, J.A.J., 1993. The saturating contact rate in marriage and epidemic models. Journal of Mathematical Biology 31, 529–539.

Hethcote, H.W., Van Ark, J.W., 1987. Epidemiological models for heterogeneous populations: Proportionate mixing, parameter estimation, and immunization programs. Mathematical Biosciences 84, 85–118.

Hethcote, H.W., Yorke, J.A., 1984. Gonorrhea transmission dynamics and control. Springer-Verlag, New York.

Hyman, J.M., Li, J., 1997. Disease transmission models with biased partnership selection. Applied Numerical Mathematics 24, 379–392.

Hyman, J.M., Li, J., Stanley, E.A., 1999. The differential infectivity and staged progression models for the transmission of HIV. Mathematical Biosciences 155, 77–109.

Hyman, J.M., Stanley, E.A., 1989. The effect of social mixing patterns on the spread of AIDS. In: Lecture notes in biomathematics, mathematical approaches to problems in resource management and epidemiology. Springer, Berlin, pp. 190–219.

Isham, V., Mdley, G., 1996. Models for infectious human diseases: Their structure and relation to data. Cambridge University Press, Cambridge.

Jacquez, J.A., Simon, C.P., Koopman, J.S., Sattenspiel, L., Perry, T., 1988. Modeling and analyzing HIV transmission: The effect of contact patterns. Mathematical Biosciences 92, 119–199.

Kanaan, M.N., Farrington, C.P., 2005. Matrix models for childhood infections: A Bayesian approach with applications to rubella and mumps. Epidemiology and Infection 133, 1009–1021.

Knolle, H., 2004. A discrete branching process model for the spread of HIV via steady sexual partnerships. Mathematical Biology 48, 423–443.

Koopman, J.S., Jacquez, J.A., Park, T.S., 1989. Selective contact within structured mixing with an application to HIV transmission risk from oral and anal sex. In: Lecture notes in biomathematics, mathematical and statistical approaches to AIDS epidemiology. Springer-Verlang, New York, pp. 317–348.

Laumann, E.O., Youm, Y., 1999. Racial/ethnic group differences in the prevalence of sexually transmitted diseases in the United States: A network explanation. Sexually Transmitted Diseases 26 (5), 250–261.

Meyers, L.A., Pourbohloul, B., Newman, M.E.J., Skowronski, D.M., Brunham, R.C., 2004. Network theory and SARS: Predicting outbreak diversity. Journal of Theoretical Biology 232, 71–81.

Newman, M.E.J., 2002. Spread of epidemic disease on networks. Physical Review E 66, 1–11.

Nold, A., 1980. Heterogeneity in disease-transmission modeling. Mathematical Biosciences 52, 227–240.

Read, J.M., Keeling, M.J., 2003. Disease evolution on networks: the role of contact structure. Proceedings of the Royal Society of London B: Biological Sciences 270, 699–708.

Robbins, D., Del Valle, S., Mniszewski, S., 2006. Effects of population mixing in Portland. Los Alamos National Laboratory Unclassified Report, 06-5937.

Shkedy, Z., Aerts, M., Molenberghs, G., Beutels, Ph., Van Damme, P., 2006. Modelling age-dependent force of infection from prevalence data using fractional polynomials. Statistics in Medicine 25, 1577–1591.

Stroud, P.D., Sydoriak, S.J., Riese, J.M., Smith, J.P., Mniszewski, S.M., Romero, P.R., 2006a. Semi-empirical power-law scaling of new infection rate to model epidemic dynamics with inhomogeneous mixing. Mathematical Biosciences 203, 301–318.

Stroud, P.D., Del Valle, S.Y., Mniszewski, S.M., Riese, J.M., Sydoriak, S.J., 2006b. Pandemic influenza impact analysis report: Simulation of disease spread and intervention effectiveness. Los Alamos National Laboratory Unclassified Report, 06-7966.

Wallinga, J., Edmunds, W.J., Kretzschmar, M., 1999. Perspective: Human contact patterns and the spread of airborne infectious diseases. Trends in Microbiology 7, 372–377.

Wallinga, J., Teunis, P., Kretzschmar, M., 2006. Using data on social contacts to estimate age-specific transmission parameters for respiratory-spread infectious agents. American Journal of Epidemiology 164 (10), 936–944.

Youm, Y., Laumann, E.O., 2002. Social network effects on the transmission of sexually transmitted diseases. Sexually Transmitted Diseases 29 (11), 689–697.

Zaric, G.S., 2002. Random vs. nonrandom mixing in network epidemic models. Health Care Management Science 5, 147–155.