Dynamic Contact Patterns and Social Structure in Realistic Social Networks

Sara Y. Del Valle[1*], Phillip D. Stroud[1], Susan Mniszewski[2]

[1]Systems Engineering & Integration, Los Alamos National Laboratory,
MS K488, Los Alamos, NM 87545
[2]Information Sciences, Los Alamos National Laboratory,
MS B265, Los Alamos, NM 87545

[*]Corresponding author: Tel.: +1 505 665 9286, e-mail: sdelvall@lanl.gov

ABSTRACT

A number of agent-based and network models have been proposed for the analysis of the spread of infectious diseases, but most of these models are inherently static and do not explicitly incorporate the social structures of real populations. We use an agent-based model that captures geospatially varying demographic characteristics, travel patterns of individuals, and contact opportunities at households, work locations, schools, social recreational venues, hospitals, restaurants, shops, and other locations. The population consists of 19 million synthetic individuals interacting in a landscape of 6.4 million households and 938 thousand business places, geolocated on 700 thousand road segments, representing a six-county region of Southern California (Los Angeles, Orange, Riverside, San Bernardino, Ventura, and San Diego counties). Individual and household demographic characteristics are specified at the most finely granulated level that can be extracted from U.S. Census data. We simulate the social contact structure by computing every person to person interaction in space and time, using statistical models of the geographic distribution of population demographics, households, mixing places, and the movement of individuals as they undertake their daily activities.

We show some interesting results about how the degree of mixing depends on the age of the individuals and their primary non-household activity. We also show how the relative magnitude of contact rates between individuals during weekdays and weekends affects the contact structure dynamics. Models that use realistic populations are better able to capture the true dynamics of disease spread and may be useful in guiding public health policy. The simulations here can help improve current models of disease transmission and may provide an insight into patterns of interactions between real individuals.

INTRODUCTION

A social network can be represented as a graph of relationships and interactions within a population. Many natural and man-made systems can be represented by graphs composed of nodes (such as individuals, airports, or web pages) that are connected by edges representing social or physical relationships (such as friendship, airline routes, or hypertext links) (Strogatz, 2001). Analyzing the interactions of a social network can yield a deeper and more accurate understanding of the system.

The most critical part in epidemic modeling lies in how the process of human contact that underlies transmission is represented. Traditional models (Anderson and May, 1991; Hethcote, 2000) represent communities as compartments of identical individuals mixing uniformly and randomly—in other words, these models assume that each individual in the population is in contact with every other individual. Because of the recognition of the complex nature of disease spread a new type of modeling has emerged. The new paradigm is social networks and agent-based models (Watts and Strogatz, 1998; Pastor-Satorras and Vespignami, 2001; Newman, 2002; Germann et al. 2006; Ferguson et al. 2006; Stroud et al. 2007). Epidemic models on social networks differ from compartmental models in that transmission is a stochastic process and it occurs between discrete individuals, which allow these models to preserve demographic and epidemiological differences in disease transmission. However, the problem with many models is that the contact networks are approximated to be static, thus excluding realistic changes that occur in the social contact network structure in response to the events of interest.

Disease spread has been modeled in artificial societies that consist of individuals residing in scale-free networks (Pastor-Storras and Vespignani, 2001) and small-world networks (Watts and Strogatz, 1998). The former characterized by degree distributions that follow a power-law, and the latter characterized by high levels of clustering and global connectivity. However, these methods are not intended to capture realistic contact networks. Other techniques for building realistic social networks include: tract-to-tract worker flow (Germann et al. 2006), Landscan data (Ferguson et al. 2006), and survey studies (Edmunds et al. 1997; Wallinga et al. 2006; Stroud et al. 2007).

German et al. (2006) developed an individual-based model, called EpiCast, to analyze the potential spread of pandemic influenza in the United States. EpiCast is a simulation in which the synthetic population consists of 140,500 identical structured communities of 2,000 individuals each. Each community has 855 households, four neighborhoods, one high school, one middle school, four elementary schools, and some playgroups for preschoolers. Individuals mix at home each night, and in a mixing group (work or school) during the day. Each of 43,323 census tracts in the United States are assigned one or more of these communities, depending on the number of people who live in the census tract. U.S. Census tract-to-tract worker flow data is used to move workers from their tract of residence at night to their tract of work during the day. Because each community is modeled with an identical distribution of household size and identical values of other demographic measures, spatial variation in disease severity due to spatially varying demographics is not seen.

Another approach for simulating disease spread in massive artificial societies is satellite imagery through the use of Landscan data. The Landscan data uses a worldwide spatial grid of 30 arc-second cells (approximately 1 km square), and satellite imagery and other data to estimate the number of people who occupy each grid cell. Ferguson et al. (2006) developed a model using Landscan data to construct synthetic households, schools, and workplaces and used the assumption that household size and age distributions did not vary geographically. Population movement was computed to match commuting distance

distributions and long-range travel patterns. The model observed spatial variation in the timing of the pandemic peak but did not report spatial variation due to demographics.

Edmunds et al. (1997) collected observational data on contacts of 65 individuals and estimated their social network. They concluded that the contact network of this study group could be approximated by a normal distribution. Wallinga et al. (2006) conducted a larger study consisting of 1,813 individuals and collected information on social contacts and analyzed age-specific contact patterns. These types of studies provide valuable information on the rates and patterns of mixing that may influence the spread of infectious diseases. However, this methodology may not be applicable for large-scale studies.

The model presented here attempts to overcome these limitations by direct simulation of the spread of infection through an agent-based generated network. Three components in this agent-based model, not present in other agent-based models (Germann et al. 2006, Ferguson et al. 2006), are important in capturing more precisely the pattern of observed realistically mixing populations (Del Valle et al. 2007). First, a seasonality component represents changes in contact rates, which occur due to school-term and vacation patterns (Edmunds et al. 1998). This seasonal input or forcing is implemented by school and partial work closures associated with school term, holiday times, and weekday and weekends contact patterns. Second, spatially varying demographics are constructed to be consistent with each U.S. Census block group where the synthetic population is matched. Spatial variation is important for capturing severity and timing of disease spread and appropriately forecasting and preparing for future epidemics. Third, classrooms are stratified by age, which is consistent with mixing patterns observed in school grounds. Children attending elementary schools mix more in their classrooms with other school children of their own age than with children of other ages.

Here, we show how to use survey studies to generate social networks and extract realistic mixing patterns. We then analyze the simulated movement of this population and characterize the emergent contact distribution at different social settings and timings. Finally, we show how these factors affect the dynamics of disease spread.

METHODOLOGY

Construction of the Artificial Society

In a real population, the course of an epidemic in space and time is a stochastic process engendered by the transmission opportunities that are created by physical proximity of individuals as they go about their daily activities. EpiSimS (Epidemic Simulation System) (Stroud et al. 2007) creates a virtual abstraction of the unknowable space-time web of social contacts by making software agents to represent each individual and simulating the movement and activities of each agent. The unique approach taken in EpiSimS is to construct the attributes and behaviors of each agent in a way that captures the maximum amount of relevant information from massive data sets about real people and places (primarily the U.S. Census (USCB, 2000), the Dun & Bradstreet business

3

database (D&B, 2003), the National Household Travel Survey (NHTS) (USDOT, 2003), and the NAVTEQ road database (NAVTEQ, 2004)). The methodology used to construct the synthetic population from census data, to assign activities to individuals based on household activity surveys, and to assign locations to activities, has been described and characterized in detail (Beckman et al. 1995). Additional details about the artificial society used in this analysis are presented here.

The Synthetic Individuals

A synthetic population consisting of 18,828,569 agents was constructed to match the actual population of six southern California counties (Los Angeles, Orange, San Diego, Riverside, San Bernardino, and Ventura). The U.S. Census subdivides these six counties into 3,978 census tracts, which are further subdivided into 12,226 census block groups. 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. In each block group, the synthetic population matches the actual population in several statistical measures: the number of residents, the number of households, the householder's age distribution, the household size and membership distribution, the household income distribution, number of workers, and the number of vehicles. There are a total of 6,345,751 households, of which 1,455,712 are agents living alone, 1,532,985 are married agents with young agents living at home, and the rest are various combinations of agents of various ages. The EpiSimS synthetic population represents the most finely granulated characterization of individuals and household structures that can be extracted from the U.S. Census data. This is a unique feature of the EpiSimS approach, in contrast to other large-scale simulations (Germann et al. 2006, Ferguson et al. 2006) in which the distribution of household structure is independent of location.

The synthetic population of southern California represents only individuals reported as household residents in the 2000 U.S. Census. It does not represent the 2.11% of the population reported as living in group quarters (for example, jails, dorms, nursing-care facilities). It is not clear to what extent the synthetic population captures the activity patterns of the undocumented population, which made up an estimated 6.5% of the California population in 2000 (USINS, 2003). The artificial society does not include visiting tourists, and does not explicitly treat guests in hotels or travelers in airports.

Activity Schedule

The survey data (USDOT, 2003) EpiSimS uses to generate the sequence of activities that each agent undertakes each day contains the time that the person started traveling to their next activity, the time spent in transit, the mode of transportation, the time that the person arrived at the activity location, and the amount of time the person engaged in that activity. In addition to being at home, activities include work, shopping, eating in a

restaurant, attending school or college, visiting a doctor, riding in or driving a carpool (designated by the NHTS activity *serve passenger*), social recreation, visiting another household, and an NHTS activity category designated *other*.

The NHTS is a comprehensive household survey of both daily and long-distance travel. It collected surveys from about 26,000 nationally representative households consisting of approximately 60,000 individuals, which document about a quarter-million daily trips and 45,000 long-distance trips. Data were collected from March 2001 to May 2002. Households were asked about all trips (daily travel) they took on a specific randomly assigned day, including how far they traveled, what they did when they got there, and how long they stayed there. For each synthetic household, EpiSimS identified a subset of the NHTS surveys that came from actual households with similar composition and demographics to the synthetic household. One of these surveys was selected at random and the reported activity schedule of that actual household was applied to the members of the synthetic household.

This approach provides a much richer artificial society than simpler day-night 12-hour time step models (Germann et al. 2006, Ferguson et al. 2006). As in the household surveys used to construct the activity schedules, some synthetic individuals work evenings or nights and spend the day at home.

To determine the weekday schedules, we used non-holiday weekdays from the NHTS household survey and assigned these activities to our individual agents. To account for absenteeism, we allow 84% of the workers to go to work on weekdays and 35% on weekends (USDOL, 2008). Whereas 95% of students go to school on each simulated weekday, which is consistent with school absenteeism (USDOE, 2008), none attend school on weekends. The percentage of workers and students who stay home each day is randomly chosen to allow for variability within the simulation. When an adult or a child is chosen to take the day off from work or school (due to absenteeism), respectively, the model assumes that these individuals will remain home for as long as they were supposed to be at work or school, which is dictated by their itineraries.

Locations

EpiSimS represents each business, school, restaurant, office, and shop that has a business address listed in the Dun & Bradstreet business directory database as a location. Each business is characterized by its geographic location, the type of business that is conducted there (from its standardized industry classification, or SIC, code), and the number of workers who are employed there. There are 938,000 separate business locations represented in the southern California landscape.

The actual physical road network consists of road segments and intersections. NAVTEQ (2004) provides the latitude and longitude of the end-points of every road segment (that is, stretch of road between two intersections) in the United States, updated quarterly. The road network in the six southern California counties has 700,000 road segments. EpiSimS maps each business address to a road segment (sometimes combining rural road

segments) and places that business location at the latitude and longitude of the road segment center. Each road segment is mapped to a census block. The U.S. Census data provides the number of households in each census block. EpiSimS apportions this number of households onto the associated road segments. The numbers of students attending each school are drawn from the National Center of Educational Statistics (), along with the school address and the range of grades.

Assignment of Activities to Locations

EpiSimS uses a two-stage gravity algorithm to assign a location to each non-household activity of each individual. The gravity model is widely used in traffic analysis and has been described in detail (Voorhees, 1956; FHWA, 1978; Martin and McGuckin, 1998). Each individual has an anchor activity. For workers, the anchor is their work activity; for students, it is their school activity; otherwise, it is the place of residence. The first stage of the gravity algorithm assigns a location to the anchor activity of each worker and student.

The EpiSimS implementation of the gravity algorithm in effect sets the probability that worker i works at location j to be proportional to

$$e^{\gamma N_j} e^{-\beta d_{ij}} / d_{ij},$$

Eq. 1-1

where d_{ij} is the travel distance in meters from the residence of worker i to work location j, and N_j is the number of workers that are employed at location j. The coefficient values $\gamma=0.377$ and $\beta=0.000209$ have been fit to ensure that the number of workers assigned to each work location matches the Dun & Bradstreet data, and that the distribution of commute distances from home to work matches the distribution extracted from the NHTS data. The second stage gravity model that assigns locations to non-anchor activities follows a similar formulation, except the distance is replaced by the sum of the distance from the anchor activity to the non-anchor activity plus the distance from the non-anchor activity to the place of residence.

Sub-location modeling

EpiSimS partitions each location into sub-locations to model how people interact within a location. The sub-locations are intended to represent individual classroom mixing groups within a school location, shops within a shopping center, or workgroups within a business location. Furthermore, EpiSimS has the capability to model transmission between students at the same school who are not classmates, workers at the same business address who are not in the same work group, and individuals who live in different households on the same city block.

The number of sub-locations at each location is computed by dividing the location's peak occupancy by the appropriate mixing group size. The mean workgroup size varies by SIC code. Two data sources were used to estimate the mean workgroup size by SIC code. Yee and Bradford (1999) conducted a survey to determine the average worker density in the

workplace, quantified as workers per square feet, by SIC code. The U.S. Department of Energy's Energy Information Administration conducted an extensive survey of commercial building usage, including workers per building, floors per building, and enterprises per building, by SIC code (Michaels, 2003). The mean workgroup size was computed as the average from the two data sources (normalizing the worker density data) and ranges from 3.1 for transportation workers to 25.4 for health service workers.

RESULTS

The Social Contact Network

The dynamics of disease spread are determined in large part by the dynamic structure of the social contact network. EpiSimS does not impose a social contact network in the population but allows it to emerge from the individual activity patterns of millions of individuals. More than 80 million activities occur each day. The simulation keeps track of every individual as they travel between locations 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. Thus, time dependent person-to-person social contact networks based on the sequence of activities each person carries out throughout the day can be extracted from the simulation.

Weekday vs Weekend Contacts

The topology of the graph determines the population's connectivity and consequently how effectively an infectious disease can be disseminated on it. Although much work has been done to map and analyze disease spread on static networks, little is known about the impact of dynamic networks. Here we take a step towards filling this gap: we have constructed an agent-based simulation model in which the agents have weekday and weekend schedules, which has enabled us to characterize the impact of dynamic contacts.

Two individuals that are present in the same sub-location (e.g. room or mixing group) at the same time are taken to be in contact for purposes of modeling the spread of airborne infections. The strength of those connections varies depending on the duration of contact and the type of activity occurring in the shared location.

Figure 2 shows the degree distribution of the population for the weekend and weekday schedules, specifying how many individuals have a given number of contacts. The number of contacts a person has each day ranges between 0 and 149 (mean 12.8) and 0 and 209 (mean 18.1) for the weekend and weekday, respectively. Neither simple contact structure models (as in fully mixed contact groups of specified size) nor analytic contact structure models (as in power-law distributions) capture the structure seen in the emergent social contact structure.
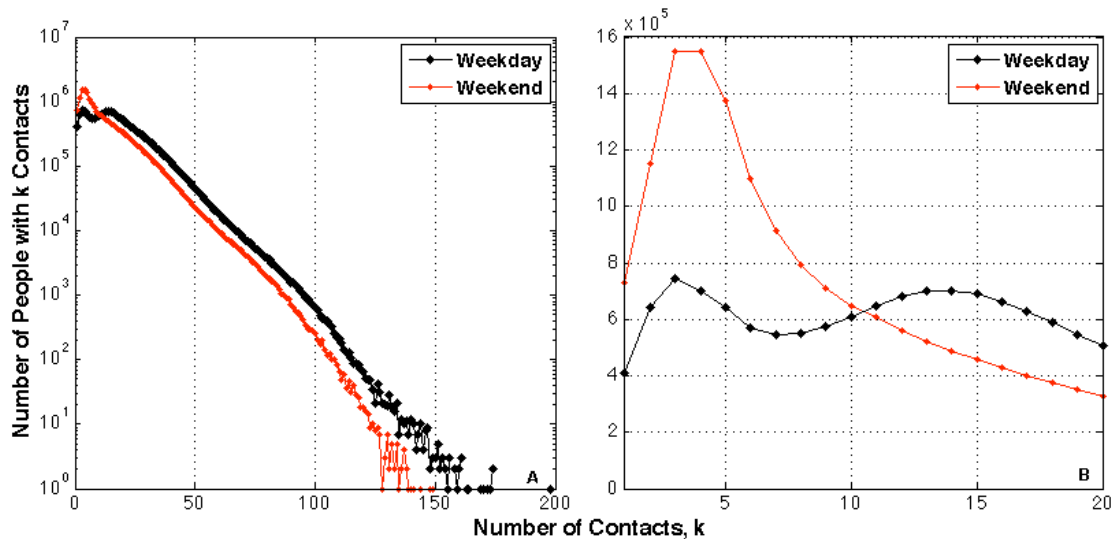
FIGURE 2. Weekday and weekend contact distribution of the population of Southern California. Note that the distributions show an exponential decay behavior. **A.** The degree distribution of weekday and weekend contacts in log-linear scale, showing how many people have a given number of contacts per day. **B.** The degree distribution of weekday and weekend contacts in linear scale showing the first 20 contacts.

Figure 3 shows the degree distribution of contacts per person (per day) in the three top-level social settings and aggregated over the remaining six activities for weekdays and weekends. With the exception of household contacts, it is clear from observation of figure 3 that the distribution of the context of contacts varies from weekday to weekend. During the weekend, adults tend to have fewer work related contacts, and children do not have school related contacts. However, household related contacts do not vary from weekday to weekend, although the time spent at home by members of the household might change. The contact distribution for the remaining of the activities is moderately affected by the weekday and weekend schedules. Tables 1 and 2 show the average number of contacts for weekdays and weekends by activity category. The largest average number of contacts occurs at college and while shopping for both weekdays and weekends. Note that although the number of classes offered at colleges during the weekends is less than the weekdays, we chose not to reduce the college activities on the weekends in order to account for the social contacts college students have on the weekends. With the exception of work, school, and other, the average number of contacts does not vary much from weekday to weekend.
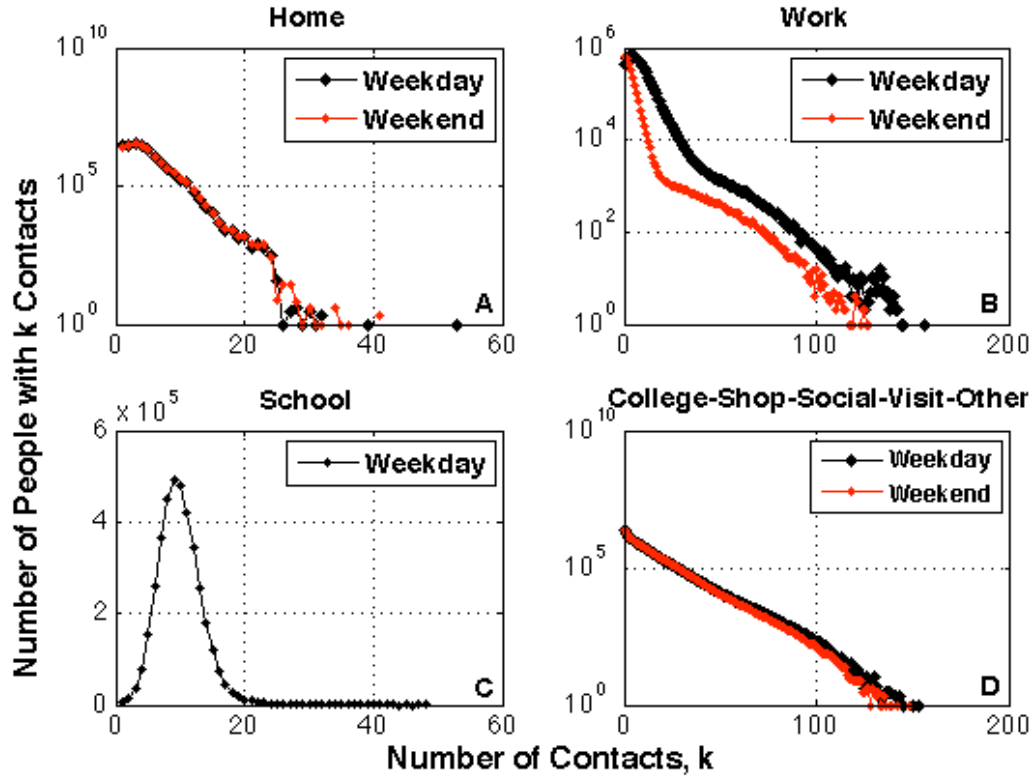
FIGURE 3. Weekday and weekend contact distribution of the population of Southern California for the four top-level activities. **A.** The household degree distribution of weekday and weekend contacts in log-linear scale, showing how many people has a given number of contacts per day. **B.** The work degree distribution of weekday and weekend contacts in log-linear scale. **C.** The school degree distribution of weekday contacts in linear scale. We assume that schools are closed on the weekends, consequently, there are no contacts generated at schools on the weekends. **D.** The degree distribution of weekday and weekend contacts in log-linear scale for college, shopping, social, visit, other, and passanger activities.

Table 1. Breakout of average, standard deviation, minimum, and maximum average number of contact by activity category for a weekday schedules

Activity	Mean	STD	Min	Max
Home	3.3	2	0	53
School	9.9	3.3	1	48
Work	7.8	6.3	1	156
College	21.2	9.9	1	81
Visit	3.4	2.5	1	75
Other	6.3	6.8	1	130
Social	7.8	5.8	1	55
Shop	14.8	12.1	1	153
Carpool	2.2	1.8	1	35

Table 2. Breakout of average, standard deviation, minimum, and maximum average number of contact by activity category for a weekend schedule.

Activity	Mean	STD	Min	Max
Home	3.3	2.1	0	41
School	0	0	0	0
Work	3.7	4.5	1	127
College	21.4	9.6	1	69
Visit	3.8	2.7	1	72
Other	4.3	5.3	1	114
Social	7.8	5.8	1	55
Shop	14.1	11.6	1	148
Carpool	2.1	1.7	1	35

Weekday vs Weekend Contacts by Age

Figure 4 shows the average number of contacts per person per representative weekday or weekend by age, aggregated for all activities and for the top three activity categories. The average number of contacts ranges between 9 and 16 (mean 13.6) for all age groups during the weekend and between 12 and 21 (mean 18) for all age groups during the weekday. In general, school-aged children tend to have more contacts than adults during the weekday; however, the roles are reversed during the weekend. The average number of contacts of older adults (> 60) varies moderately between weekdays and weekends, which might be due to the fact that their activities do not change much from day to day. For household related contacts, the average number of contacts decreases with age, such that older individuals (> 50) have fewer contacts than younger individuals (< 50). Work related contacts decreased two-fold for most working adults during the weekend. The reason why children appear under the work related contacts is because working adults interact with children at schools and day care centers, which are considered workplaces. The average number of 'other' related contacts decreased dramatically from weekend to weekday for all age groups.
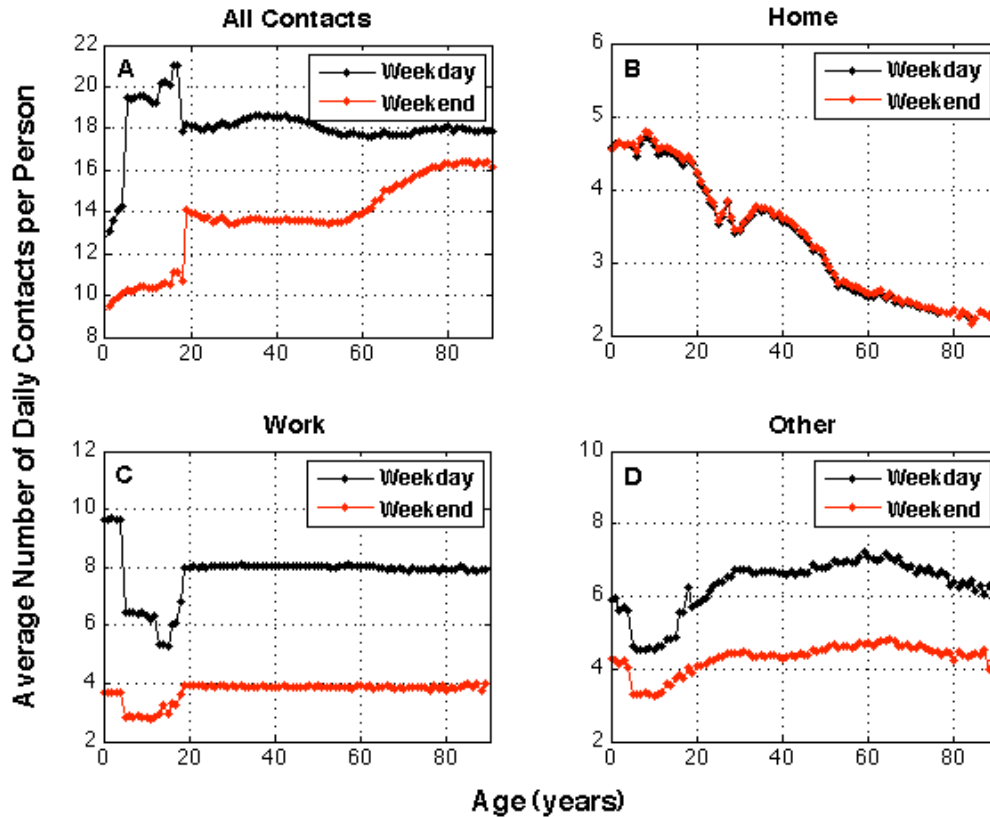
FIGURE 4. Weekday and weekend contact distribution of the population of Southern California as a function of age, aggregated over all activities and for the three top-level activities. **A.** The degree distribution of weekday and weekend contacts as a function of age aggregated over all activities. **B.** The household degree distribution of weekday and weekend contacts as a function of age. **C.** The work degree distribution of weekday contacts. **D.** The 'other' degree distribution of weekday and weekend contacts as a function of age. This social setting corresponds to the NHTS activity category designated as 'other'.

The average number of contacts per student per representative school day is shown in Figure 5. EpiSimS assigns students to classroom mixing groups with other kids of the same calendar age (Figure 5). In real classrooms, there will typically be a mix of six and seven year olds in first grade. In EpiSimS, however, six year olds are only with six year olds. For the simulations presented here, the classroom mixing group size parameter is set to 10, so that 10 kids of the same age are placed in a room with a teacher. On each successive school day, students are assigned to a randomly selected classroom of the appropriate age group so that they do not mix with exactly the same students each day.
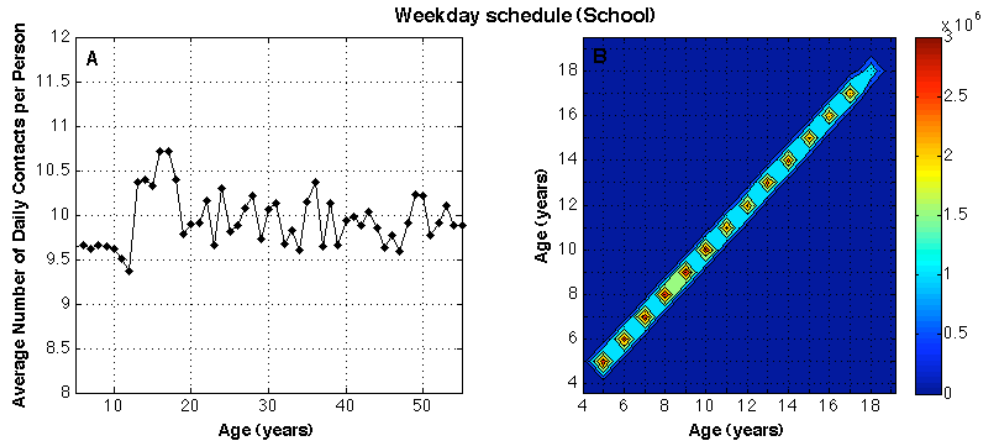
FIGURE 5. The distribution of contacts by age among students in southern California. **A.** The school degree distribution of weekday contacts as a function of age. **B.** The total number of school related contacts between age groups. The color indicates the simulated number of contacts between persons of each age pair.

Weekday vs Weekend Contact Duration

Although duration of contact plays a crucial role in the transmission 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.

Figure 6 shows the distribution of contact durations for the weekday and weekend schedules. There are many short-duration contacts (6+ contacts per weekday per person last less than 30 minutes) representing casual interactions. Structure related to weekday activities of the population (e.g. 6 hours school related activities and 8 hours work shifts) is readily apparent. The average contact duration per person is 3.5 hours for both weekdays and weekends; however, the standard deviation is 5 hours for the weekend schedule and 3 hours for the weekday schedule.
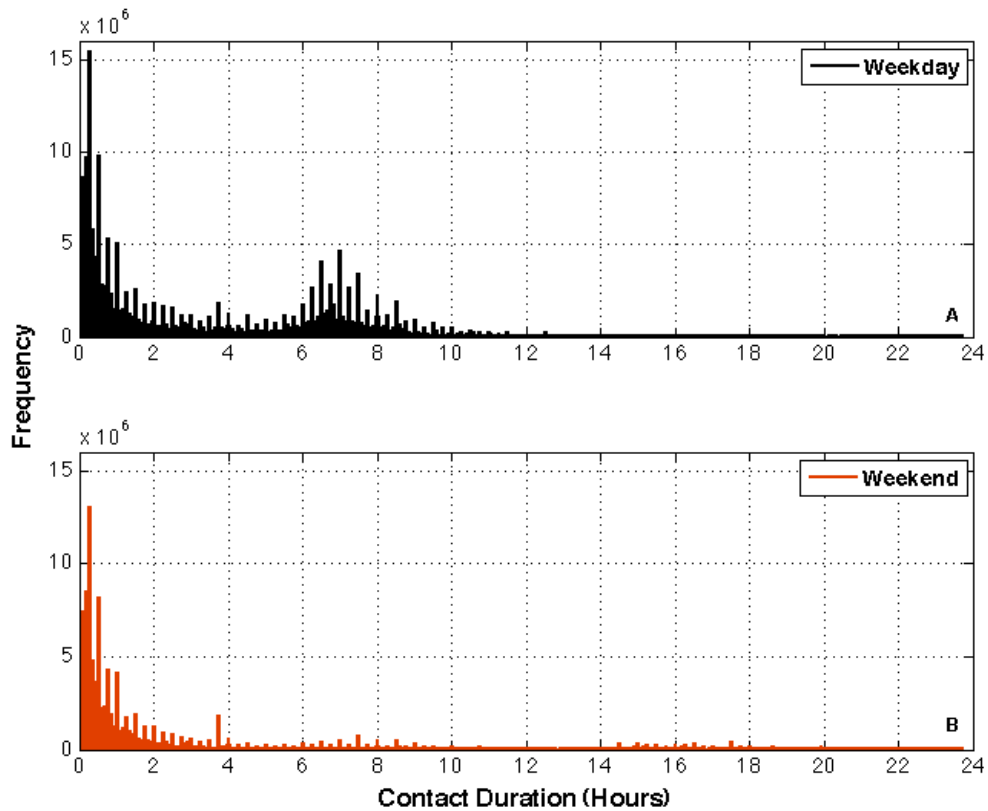
FIGURE 6. The distribution of contact duration among people in southern California. **A.** The distribution of weekday contact duration. **B.** The distribution of weekend contact duration.

Table 3 and 4 show the average contact duration by activity category for the weekday and weekend schedule, respectively. For the weekday, the longest contacts occur at home, with an average contact duration of more than 8 hours, followed by school and work with an average duration of contact of 6 and 4.5 hours, respectively. During the weekend, the longest contacts occur at home, with an average contact duration of almost 11 hours. Note that there are no school contacts because schools are closed on the weekends.

Table 3. Breakout of average, standard deviation, minimum, and maximum duration per contact by activity category for a weekday schedule.

Activity	Mean	STD	Min	Max
Home	8 hrs 11 min	3 hrs 41 min	1 sec	23 hrs 45 min
School	5 hrs 59 min	1 hr 49 min	1 sec	15 hrs 15 min
Work	4 hrs 33 min	2 hrs 48 min	1 sec	19 hrs 55 min
College	2 hrs 25 min	1 hr 57 min	5 sec	14 hrs 25 min
Visit	1 hr 23 min	1 hr 23 min	1 sec	17 hrs 10 min
Other	1 hr 2 min	1 hr 21 min	1 sec	17 hrs 48 min
Social	55 min	59 min	2 sec	19 hrs 25 min
Shop	23 min	29 min	2 sec	17 hrs 10 min

Carpool	11 min	25 min	2 sec	9 hrs 39 min

Table 4. Breakout of average, standard deviation, minimum, and maximum duration per contact by activity category for a weekend schedule.

Activity	Mean	STD	Min	Max
Home	10 hrs 57 min	5 hrs 53 min	1 sec	23 hrs 45 min
School	0	0	0	0
Work	3 hrs 8 min	2 hrs 38 min	1 sec	17 hrs 25 min
College	2 hrs 27 min	1 hr 57 min	5 sec	14 hrs 25 min
Visit	1 hr 36 min	1 hr 37 min	1 sec	20 hrs 55 min
Social	55 min	59 min	2 sec	19 hrs 15 min
Other	49 min	1 hr 8 min	1 sec	14 hrs 30 min
Shop	23 min	28 min	2 sec	18 hrs 45 min
Serve	11 min	25 min	2 sec	10 hrs 9 min

Impact of Social Network on Disease Spread

In order to gain further insight into the effects of weekday and weekend contact patterns for the social network of southern California, we analyzed the impact of pandemic influenza spread. The disease model considered for this experiment is from the influenza model by Longini et al. (2004), as described by Stroud et al. (2004).

Disease progression is characterized by 14 disease states. A susceptible individual who becomes infected progresses through a sequence of disease states, beginning with non-infectious incubation followed by a pre-symptomatic infectious stage. From there, an individual can become symptomatic-infectious, or asymptomatic-infectious. The asymptomatic-infectious passes through a less-infectious stage and then recovers. The symptomatic-infectious splits into two levels of severity: some continue their activities, some stay home. Those who continue their activities pass through a less-infectious stage on their way to recovery. Those symptomatics who stay home split into manifestations with and without severe complications such as pneumonia that would require hospitalization. Non-circulating symptomatics will either die or progress through a convalescent stage on their way to recovery. The duration of each state is a stochastic variable, with distributions of sojourn times matched to case history distributions (Longini et al. 2004).

For this experiment we considered two scenarios: a model in which individuals have a continuous schedule and another where individuals have a schedule consisting of weekdays, weekends, and holidays. The continuous schedule model averages weekdays and weekends to get a representative day (during which 5/7 of the population engage in their weekday activities, and 2/7 engage in their weekend activities). Also, holidays and summer schedules are averaged in with school year schedules. The result of this approximation is that on a representative day, roughly half of the students are in school, since about half the days in a year are school days. In contrast, the weekday/weekend

schedule model allows 95% of the students in school for 5 days followed by no students for 2 days and 84% of the working population in work activities for 5 days followed by 35% for 2 days. We then seeded the simulation with the same 102 infected individuals for both scenarios.

Figure 7 shows the results of pandemic influenza spread on a static and dynamic contact network. Student dismissals and work closures on weekends and holidays have a big impact on the epidemic spread. As more school dismissals and/or workplace closures are added to the scenario, the social contact network thins out and thus it takes longer for the disease to spread around southern California. Note that the clinical attack rate drops from 35.6% for the static network to 29.3% for the dynamic network. Accurate predictions of these quantities can be critical to efficiently plan allocation of limited resources.

Our analysis suggest that the common measure of the transmissibility of a disease known as the effective reproduction number (R_0), which is the average number of secondary cases produced by a typical infectious individual during its infectious period, is very different for these two approaches. Although these procedures are quantitatively similar, they produce completely different social dynamics, which in turn alter the spread of the disease. The topology of the network has a great influence in the overall behavior of the spread of the epidemic. Our findings show that the true mixing patterns of humans are far more complex than as usually represented by simple networks. Thus, our results show that simulation assumptions can have severe implications for modeling human disease spread on realistic social networks.
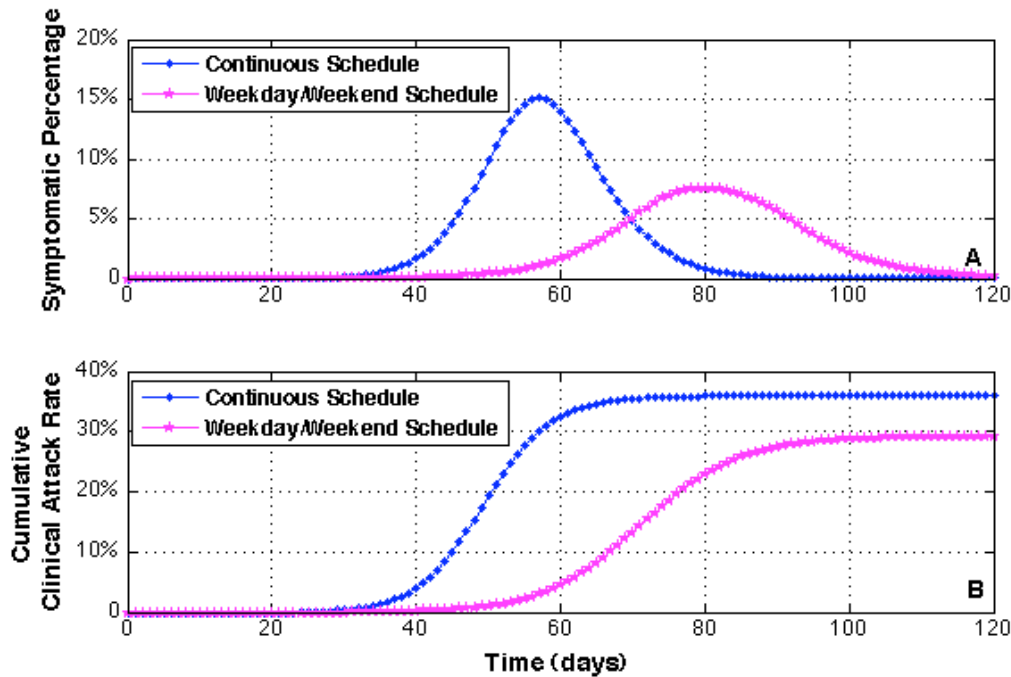

FIGURE 7. Epidemiological predictions of pandemic influenza on two classes of networks: static (continuous) and dynamic (weekday/weekend) contact networks. **A.** The

current percentage of the southern California population that is symptomatic, showing the impact of different mixing assumptions on disease spread. **B.** The cumulative percentage of the population that has developed symptoms.

CONCLUSION

The structures of human contact networks play a crucial role in the transmission of many infectious diseases, since infection is often spread by contact between infective and susceptible individuals. There are different approaches used to generate social networks. We used a survey-based model to generate the social contact network of nearly 19 million individuals and showed the different mixing patterns for weekdays and weekends. We showed that mixing assumptions could have a significant effect on predictions of influenza spread. In particular, we showed that the topology of the network could have severe implications on the reproduction number (R_0), which it is typically used to determine whether or not an epidemic occurs and if so, its severity.

We argue that models that use dynamic contact networks and realistic mixing patterns are better able to capture the dynamics of infectious-disease transmission than models that use static and homogeneous mixing assumptions. The emergent degree distribution of the baseline social network is in agreement with contact patterns observed in small convenience samples (Edmunds et al. 1999; Wallinga et al. 2006). The simulations suggest that dynamic contact networks can break paths of transmission and prolong the spread of an infectious disease.

Capturing realistic mixing patterns can have a profound influence in the predictions of future disease spread and the resources needed to contain an outbreak. It is clear from the results of this study that different assumptions can give different answers to the same questions. Thus, if the model predictions are being used to guide public health policy, it is essential to use realistic mixing assumptions. The simulations here are useful in providing estimates of the effects of dynamic mixing patterns for future epidemic guidelines.

REFERENCES

Anderson, R.M., May, R.M., 1991. Infectious Diseases of Humans. Oxford University Press, Oxford.

Beckman, R.J., baggerly, K.A., McKay, M.D. 1995. Creating Synthetic Baseline Populations. Transportation Research A 30A(64), 415-429.

Del Valle, S.Y., Hyman, J.M., Hethcote, H.W., Eubank, S.G., 2007. Mixing Patterns Between Age Groups in Social Networks. Social Networks 29, 539-554.

D&B (Dun & Bradstreet), 2003. Business Directory Database Available: http://www.dnb.com/.

Edmunds, W.J., O'Callaghan, 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 B 264, 949-957.

Ferguson, N.M., Cummings, D.A.T., Fraser, C., Cajka, J.C., Cooley, P.C., Burke, D.S., 2006. Strategies for Mitigating an Influenza Pandemic. Nature 442, 448-452.

FHWA (Federal Highway Administration), 1978. Quick-Response Urban Travel Estimation Techniques and Transferable Parameters. NCHRP Report 187. Available: http://nationalacademies.org/trb/bookstore.

Germann, T.C., Kadu, K., Longini, I.M., Macken, C.M., 2006. Mitigation Strategies for Pandemic Influenza in the United States. Proceedings of the National Academy of Sciences of the United States of America 103, 5935-5940.

Hethcote, H.W., 2000. The Mathematics of Infectious Diseases. SIAM Review 42, 599-653.

Longini, I.M., Halloran, M.E., Nizan, A., Yang, Y., 2004. Containing Pandemic Influenza with Antivirals Agents. American Journal of Epidemiology, 159(7), 623-633.

Martin, W.A., McGuckin, N.A., 1998. Travel Estimation Techniques for Urban Planning. NCHRP Report 365. Available: http://nationalacademies.org/trb/bookstore.

Michaels J. 2003. Commercial Buildings Energy Consumption Survey. Available: http://www.eia.doe.gov/emeu/cbecs/cbecs2003/detailed_tables_2003/detailed_tables_2003.html

NAVEQ, 2004. United States Road Network Data. Available: http://www.navteq.com.

Newman, M.E.J. 2002. The Spread of Epidemic Disease on Networks. Physical Review E 66, 016128.

Pastor-Satorras, R., Vespignami, A., 2001. Epidemic Spreading in Scale-Free Networks. Physical Review Letters 86, 3200-3203.

Strogatz, S.H., 2001. Exploring Complex Networks. Nature 410, 268-276.

Stroud, P., Del Valle, S., Sydoriak, S., Riese, J., Mniszewski, S., 2007. Spatial Dynamics of Pandemic Influenza in a Massive Artificial Society. Journal of Artificial Societies and Social Simulation 10 (4), 9.

USCB (United States Census Bureau), 2000. Decennial Census. Available: http://factfinder.census.gov.

USINS (United States Immigration and Naturalization Service), 2003. Estimates of the Unauthorized Immigrant Population Residing in the United States: 1990 to 2000. Available: www.dhs.gov/xlibrary/assets/statistics/publications/Ill_Report_1211.pdf.

USDOE (United States Department of Education), 2008. Available: http://www.ed.gov/.

USDOL (United States Department of Labor), 2008. Available: http://www.dol.gov/.

USDOT (United States Department of Transportation), 2003. Highlights of the 2001 National Household Travel Survey. BTS03-05. Available: http://www.bts.gov/publications/highlights_of_the_2001_national_household_travel_survey/.

Voorhees, A.M., 1956. A General Theory of Traffic Movement. 1955 Proceedings, Institute of Traffic Engineers, New Haven, CT.

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.

Watts, D.J., Strogatz, S.H., 1998. Collective Dynamics of 'Small-World' Networks. Nature 393, 440-442.

Yee, D., Bradford J., 1999. Employment Density Study. Canadian METRO Council Technical Report, April 6, 1999.