Disease Dynamics in a Dynamic Social Network

Claire Christensen[1](*cpc146@phys.psu.edu)*,
István Albert[3](iua1@psu.edu),
Bryan Grenfell[2,4](grenfell@psu.edu), and
Réka Albert[1,2](*ralbert@phys.psu.edu*),

[1]Department of Physics,
[2]Center for Infectious Disease Dynamics, and
[3]The Huck Institutes for the Life Sciences,
The Pennsylvania State University, University Park PA 16802, USA
[4]Fogarty International Center, National Institutes of Health, Bethesda, MD 20892-2220, USA

Abstract

We outline a learning algorithm for the development of a realistic, evolving social network (a city) into which a disease is introduced. We compare the results of simulations in populations spanning two orders of magnitude to prevaccine era measles data for England and Wales and demonstrate that our simulations are able to capture the quantitative and qualitative features of epidemics in populations as small as 10,000 people. In addition, we show the utility of network simulation in concurrently probing contact network dynamics and disease dynamics, and we suggest that our suite of algorithms can be extended to the study of less well-documented diseases.

1. Introduction

An area of fundamental importance to epidemiology lies in determining which characteristics of an underlying population are most important in dictating the manner in which a disease will spread through that population. That is, what is it about the demographics and connectivity of an underlying contact network that creates the most prominent features of the landscape in which the disease travels, and moreover how do *changes* to that landscape affect the dynamic behavior of the disease? An extensive body of work has explored these fundamental questions, demonstrating, for example, that a global demographic as simple as population size can have profound effects on epidemic occurrence, leading to large, regular epidemics with few fadeouts in populations above a certain threshold size, while populations whose size is below this threshold value experience only small, chaotic epidemics [1]. The large amount of work on childhood diseases such as measles, has shown that dynamic trends in the underlying social structure--for example, the change in aggregation among school children with the onset and end of the school term--create dynamic trends in epidemic profiles (epidemics tend to occur when children aggregate [1-3]); in addition, [1, 2] have proven that an increase in birthrate can cause a shift in epidemic periodicity from biennial to annual in populations above a certain threshold .

To date much of the mathematical modeling explaining epidemiological data has employed fully mixed compartmental models [1, 2, 4, 5]. In the simplest of these models—the basic SIR model—the population is assumed to consist of individuals who are either susceptible (*S*), infected/ infectious (*I*), or recovered/immune (*R*). The population is assumed to mix fully, and interactions are governed by coupled differential equations such that the rate of change in the number of susceptible individuals is proportional to $-\beta SI$, where β is the contact rate; in addition, if individuals recover at a rate γ per unit time, then the rate of change in the number of infected individuals is proportional to $\beta SI - \gamma I$ (see, for example, the supplement in [3]). This basic SIR model is quite successful in reproducing and explaining the real-world behaviors of large and even some intermediate-sized populations for which the assumption of full mixing is a good approximation. In fact, a handful of the model's more complex variants capture some of the nuances of more complicated populations for which full mixing is *not* a fair assumption [6, 7]. However, as useful as fully mixed compartmental models are for large populations, their utility tends to be drastically reduced as

population size decreases and as population structure (i.e. heterogeneity in connectivity) increases, since the assumption of full mixing fails to hold and/or the mathematics needed to describe the heterogeneities in mixing becomes intractable.

Thus, a new avenue in epidemic modeling involving the *in-silico* simulation of both the contact network in which the disease will propagate and the propagation of the disease, itself, has arisen to address some of the shortcomings of deterministic modeling [8, 9]. Simulating contact networks and computationally exploring disease propagation in these networks affords us many opportunities that are not available from solutions to deterministic equations, alone. We can, for example, directly probe the relationships between the topology of contact networks and disease dynamics at multiple scales and for multiple populations with varying topologies. Furthermore, we can simulate realistic and detailed underlying topologies by incorporating relevant statistical data and can allow this topology *to evolve* according to rates collected from statistical data for real societies, then observing *directly*, for any population size, how the dynamics of the topology drives the dynamics of disease.

Here, we outline the essential features of a computational program that allows us both to simulate an underlying social network that is a conglomerate of family networks, work networks, school and preschool networks and individuals and that grows and changes according to salient real-world statistical rates. The program allows us to track both the dynamics of the population and the dynamics of a disease propagating through this population. Simulations have been generated for measles, and we present a comparison of our findings to data found in [1-3], demonstrating agreement in long-term epidemic profiles for a range of population sizes from 10,000 individuals to ~250,000, and providing unique insight into how and why the social topology of the contact network influences the propagation of the disease through the population.

2. Simulating social networks and disease dynamics

Our network simulation algorithm is essentially a learning system contact network, meaning, first, that the individuals comprising the underlying social population are represented by vertices (nodes) in a network, and second, that their most salient social interactions (familial, working, and (pre)school) are indicated by edges; furthermore, the social network "learns" how to grow and change over time from statistical data, such as age distributions, birth rates, marriage rates, etc. In addition, a disease algorithm propagates a disease with specific transmission and recovery parameters through the evolving social network. Thus, the social network algorithms are interlinked with the disease algorithm, and two dominant timescales are adopted in the simulations: a yearly timescale for "slow" social processes—i.e. marriages, formation of work groups and (pre)school groups—and a weekly timescale for "fast" or distributed social processes—i.e. births, deaths, immigration-- and for (most) disease updates. It is important to understand that while statistical data is abundant for *node-related* quantities in social networks, it is almost non-existent in regard to social *edges* (*connectivities*) in (large) social networks. It is therefore necessary to establish logical rules that are based on observation and "reverse-engineering" of the social underpinnings of social institutions (e.g. families, workplaces, (pre)schools) to account for how and why people are connected (in terms of having "social edges" between them) in a population. In the following subsections we will briefly describe both some of the major social network algorithms capable of producing a simple population model as well as disease algorithms (tailored for childhood diseases, such as measles) included in our simulations.

2.1 Dominant social processes

2.1.1 Basic demographics and social processes: births, deaths, age distribution and immigration

For a simulated population of size N, an initial age distribution is adapted from the vital

statistics of New York City [10], which has been interpolated in order to determine values for each age between 0 and 95 in an arbitrary population. The age of each individual in the population is incremented yearly, and for simplicity, all individuals age simultaneously. During each yearly time step, women (roughly half the total population) between the ages of 15 and 45 are eligible to have children, and the number of children born to women of each age *i* in this range is determined according to documented fertility rates (a rate per 1000 women of age *i*) [10]. The week at which each baby will be added to the population is randomly chosen, as are the women who will become new mothers. Similarly, the number of people of age *i* who will die in the current year is determined according to the documented rate per 1000 people of age *i* [10]. Individuals are randomly selected for death (unless their age exceeds 95, in which case they are automatically removed from the population), and the week of their removal is randomly determined.

Finally, to allow an average annual population growth rate of roughly 0.6%, close to what is observed in many large western cities [10, 11], but to maintain a stable age distribution, immigration rates were adapted from [11], and were interpolated in order to find a rate for each age between 0 and 95. The number of people of age *i* who are added to (or subtracted from) the population at time *t* is equal to the product of the number of people of age *i-1* at time *t-1* and the immigration rate for age *i:*

$$n_i(t) = r_i n_{i-1}(t-1) .$$

It is required that individuals under the age of 18 leave the population with a family group, and to the extent that it is possible to satisfy the immigration-by-age distribution, entire family groups are moved out of the population if any family member is randomly chosen as an emigrant. The week at which an individual will enter or leave the population is randomly determined; however, if the individual is moving with a family, his or her entire family group will move at the same time.

2.1.2 Marriages and family graphs

Family groups form the first of four broad classes of social subnetwork within the larger simulated population. In these simulations, a *family group* must contain at least one, and no more than two (one male and one female) adults, and any number of children. To simulate marriage, approximately 54% of the population over the age of 18 will be paired with a person of the opposite gender at any given time [10]. Marriage is not necessarily correlated with childrearing in our simulations, since single women can have children. Each individual in a family group is connected to every member of the immediate family (the graph is fully-connected) and remains so until the child(ren) turn 18; at this point only the mother and father will remain connected. Because it occurs very infrequently, children whose parents die are not reassigned to new families. The average family size is 3-4.

2.1.3 Work graphs

Work groups comprise the second of the four broad classes of social subnetwork within the larger simulated population. Individuals have the option to enter the workforce at age 18, but must exit from the workforce at age 65. Each year, based on (un)employment statistics [12] approximately 8% of current 18-year olds will remain unemployed, and all others will enter the workforce. The initial number of workplaces is approximately 1% of the total population size [12], and is allowed to grow over time. Each initial workplace is a hub-and-spoke graph—in essence, a boss and employees. As new workers are added to the workplaces, they are attached uniformly randomly to a minimum of three and a maximum of all other workers in the workplace. For childhood diseases, where clustered populations of adults do not greatly affect the transmission of the disease, it is acceptable to update the work groups only once per year (as opposed to weekly updates), for computational efficiency; in the data we present, such is the case.

2.1.4 Preschool graphs

A large preschool network, consisting of children between the ages of 2 and 5, forms the third of the four social subnetworks within the larger population. Children in this age group are randomly connected to between 0 and $\frac{10000xN_{2-5}}{N}$, where N_{2-5} is the number of children between the ages of 2 and 5 and where N is the total number of people in the population.

2.1.5 School graphs

School groups form the final, and in the case of measles, arguably the most important class of social subnetwork within the larger simulated population. All children between the ages of 6 and 18 are included in the school subnetwork. This subnetwork consists of fully-connected, age-specific classes of maximum size equal to 40, that are, in turn, interconnected randomly (from a truncated normal distribution) according to the following rule: children between the ages of 6 and 13 can have a maximum of 0.5*(size of their class) connections to other classes in this age range; children between the ages of 14 and 18 can have a maximum of 0.75*(size of their class) connections to other classes in this age range; all children have at least .25*(size of their class) connections to other classes. Each year, 5% of school edges are severed from the beginning of April until the end of June, and again from the beginning of September until the beginning of December to represent a lower transmission season during which children might spend more time outdoors, in less proximity to one another. In addition, all school edges are severed from the beginning of July through the end of August, to simulate summer vacation. Only from December through March, when children are forced to be in enclosed spaces with one another due to colder weather, are all the school edges in place. With this scheme of seasonal edge removal and reinstitution which is also applied to the preschool subpopulation, we are able to approximate school-term forcing [2, 3].

2.2 Disease processes

2.2.1 Initial immunity by age

Although an initial immunity profile (i.e. what portion of the population, by age, has initial immunity to the disease) will likely have to be tailored on a by-disease basis, we again use measles as an illustration of how such a profile is acquired and implemented in our simulation. The immunity profile (by age) was adapted from [4] and is a set of input parameters at time $t=0$. The susceptibility of immigrants is assumed to be higher (3.5%) than it is in the native population (1%) [11].

2.2.2 Loss of maternally-acquired immunity

All newborns are assumed to be immune to the disease for (in the case of measles) up to six months after the week of birth. The specific week during this six month period at which maternally-acquired immunity will be lost is randomly chosen from a Gaussian distribution with mean equal to 12 weeks and sigma equal to two weeks.

2.2.3 Spread and recovery

Disease can spread from an infected individual to a susceptible individual if there is an edge (contact) between the two. The probability of infection is also dependent on an age-specific transmission rate, β, that is obtained from a matrix whose (i,j) entry represents a contact between an infected individual in age group i and a susceptible individual in age group j. The values in the matrix used for these simulations were derived from [5], using their mixing matrix structure with five basic age groups ([0,4], [5,9], [10,14], [15,19], [20,95]). However, the rates for children between the ages of 0 and 6 have been linearized to allow for a more natural contact-transmission process between older children in elementary school and their younger siblings. Although the values in [5] were reported for pertussis, this disease is similar enough to measles to be applicable to measles simulations; a comparison of plots of the mean ages at first infection for measles data and for our simulations using pertussis data shows a marked

peak in the 5-6 year-old category (see discussion in . ∮3.2.3).

2.2.4 Maintaining epidemics: "sparking"

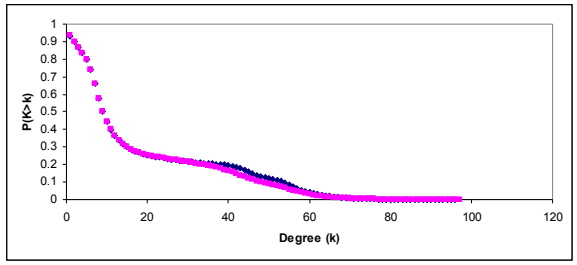
Particularly in the case of smaller populations, where fadeouts (periods during which no one is infected, and after which further epidemics would be impossible without reintroduction of the disease) are frequent, to ensure that there is always some chance of infection at each weekly update of the disease algorithm, we introduce a *sparking* process, whereby, with probability $P(t) \sim \ln(N(t))$, where *N(t)* is the total population size at time step *t,* a susceptible individual will become infectious without contact with an already infected individual in the population. The implicit assumption is that the "spark" has had contact with an infected individual from outside the native population, has become infected, and has introduced the disease into the native population. The spark is chosen at random from all current susceptibles in the population.

3. Results

3.1 Social network topology

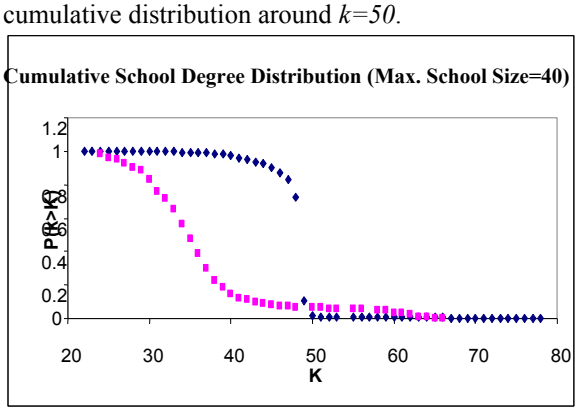
One of the most important topological quantities dictating the manner in which a disease will spread from one individual to others is the degree (*k*) – the number of edges adjacent to a node—that each individual has. Remarkably, for our simulations, when the entire network is looked at in terms of a *cumulative degree distribution* (the probability that a given individual will have degree higher than degree *K*), regardless of the total population size, the degree distributions for the full populations are quite similar (Fig. 1). All networks exhibit nodes with between 0 and 100 connections (some of these may be multiple connections to the same person), and the majority of nodes in the network do not have degree much greater than ~40, since the probability of having *k*>40 is only about 20%.

Figure 1: Cumulative degree distribution for populations of size $\sim 10^5$ (blue) and $\sim 10^4$ (pink).



The shallow slope of the region between *k*=22 and *k*=40 is due predominantly to (pre)school connectivity and secondarily to work connectivity. The region of the cumulative distribution to the right of *k*=40 is separated from the region to the left by a segment that becomes increasingly switch-like as population size grows. In fact, the preschool subgraph actually masks a higher-degree separation point at *k=50*, a point that is a function of school connectivity. It is *this* higher-degree feature that becomes increasingly switch-like as population size grows and that therefore seems to have the dominant effect on topology as population size changes. Specifically, the cumulative school degree distribution is switch-like in appearance for large populations, but decreases more gradually for smaller populations (Fig. 2), and can be superimposed on the transition region between *40<k<55* in the total cumulative degree distribution if the preschool subgraph is excised from the network.

Figure 2: Cumulative school degree distribution. The probability, P(K>k), is plotted for each degree in an averaged set of typical simulated networks of $\sim 10^5$ (blue) and $\sim 10^4$ (pink) nodes. The switch-like behavior of the cumulative distribution for the larger graph is a reflection of a delta function-like non-cumulative distribution around *k=50*.



The difference in behavior between the school degree distribution of large populations and that of small populations is a result of a combination of factors. Not only will large populations have far more schools (classes) than small populations, but these classes are also far more likely to be filled to capacity (40 students). Therefore the average school-aged child will have more school links in a larger population than will a comparably-aged child in a smaller population. In addition, because both the average class size, as well as the number of classes are greater in the larger population, the average inter-school degree is much higher in the larger population than it is in the smaller population ($< k_{inter}^{large} > \approx 12$ and $< k_{inter}^{small} > \approx 3$, respectively). The net effect of population size on school degree is to cause the school degree distribution (non-cumulative) to become more and more sharply peaked around k=50, during the season of highest connectivity (transmissivity); during seasons of lower, but non-zero connectivity, the distribution shifts slightly to the left.

3.2 Comparison of simulated measles dynamics to observed dynamics

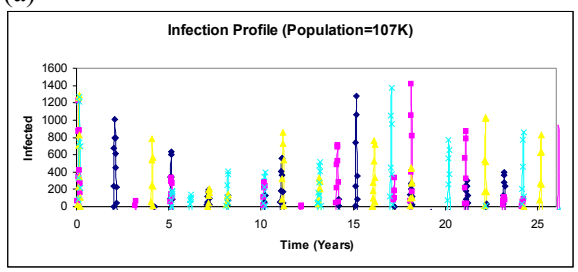
A major emphasis of our simulations has been to run extensive tests and comparisons between the output from the simulations and comparable features or plots of real-world data for a well-studied disease—namely, measles—the idea being that thorough quality control at this stage will make the algorithms applicable to diseases that have not previously been studied. To that end, in this section, we report comparisons for both a small ($\sim 10^4$) and large ($\sim 10^5$) simulated population, demonstrating not only that their dominant epidemic features— for example, the average number of individuals infected during a given epidemic (<I(t)>), the interepidemic period (I_{inter}), and the epidemic duration (τ)—are in excellent agreement with real data, but also that our simulations capture more subtle, yet measurable features of (measles) epidemics that relate topological dynamics to disease dynamics.

3.2.1 Infection profiles, interepidemic periods, and epidemic lengths

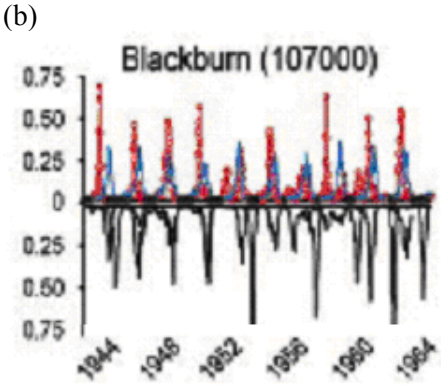
The infection profiles of a sampling of simulated populations of starting size *N=107,*000 over a time span of 25 years (Fig. 3a) agree nicely with data from [2] for the city of Blackburn, also with population of 107,000 (Fig. 3b). Similarly, profiles for a simulated population 10 times smaller (Fig. 3c) concur with data from [2] for the city of Teignmouth (Fig. 3d), whose population is comparable.

Figure 3. (a,c) Infection profiles for 5 simulated populations of starting size, *N=107,000* and *N=10,700* over 25 years. Counts are aggregated monthly. (b,d) Infection profiles for Blackburn (population 107,000) and Teignmouth (population 10,700) over the period 1944-1964 (red lines). Counts are aggregated monthly. Copyright by the Ecological Society of America [2].

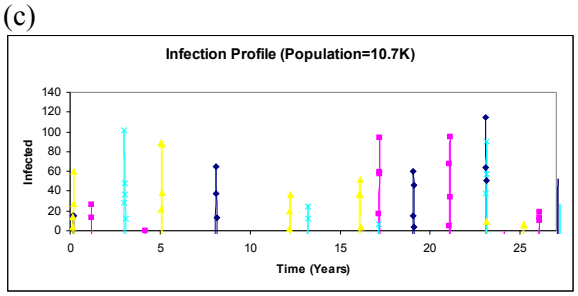
(a)



(b)



(c)



(d)

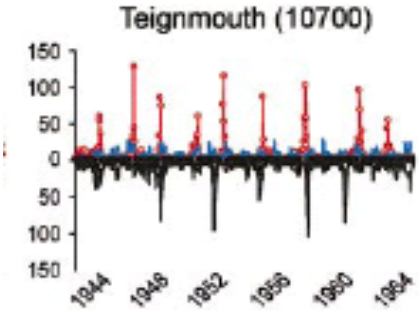


Table 1 compares the average number of infected individuals (per unit time), the interepidemic period, and the epidemic duration for the simulated populations and the sister data from [2], and demonstrates that simulation and data agree quantitatively, as well as qualitatively.

Table 1: Simulation versus real data: comparison of simulated epidemic features in large and small populations.

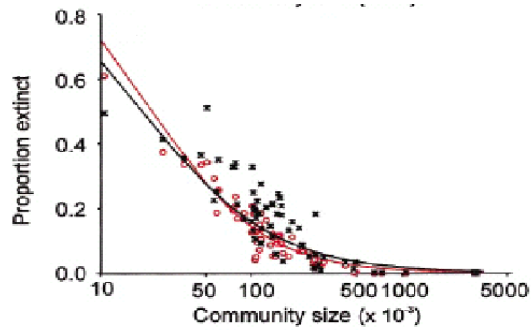
Data Set	<I(t)>	$<T_{inter}>$	<τ>
Blackburn	~500	1-2 years	41 weeks
107,000	~ 413	2.02 ± 1 years	40 ± 2 weeks
Teignmouth	~81	1-4 years	38 weeks
10,700	~ 88	4 ± 3.31 years	40 ± 2 weeks

3.2.2 Time spent in fadeout as a function of population size

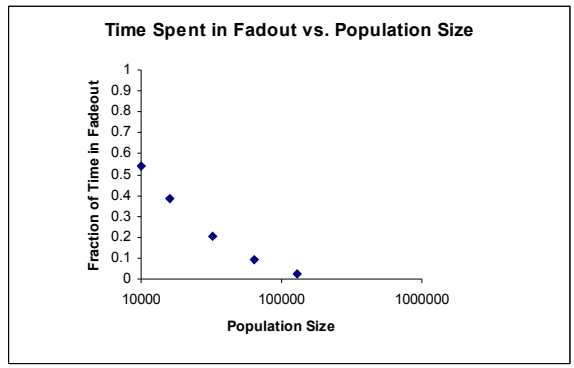
It is well documented that the amount of time a population spends in fadeout—i.e. without any infected individuals—should decrease as the size of the population increases, in the manner depicted in (Fig. 4a). We find a similar trend for population sizes between $N \sim 10^4$ and $N \sim 10^5$ (Fig. 4b).

Figure 4: (a) Proportion of time a community is in fadeout, as a function of community size. Copyright by the Ecological Society of America [2]. (b) Average proportion of time communities of size 10^3, 1.6×10^3, 3.2×10^3, 6.4×10^3, and 12.8×10^3 spend in fadeout. The data points are averaged values for five simulations at each population size.

(a)



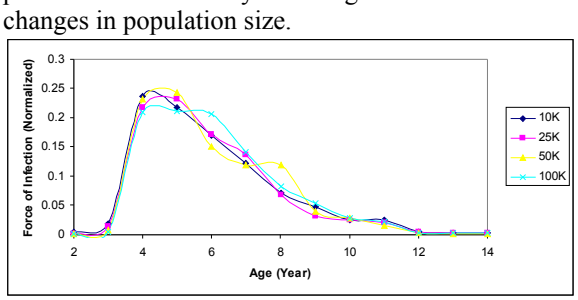
(b)



3.2.3 Force of infection by age cohort (λ)

The interplay between topological dynamics and disease dynamics in our simulations produces a force of infection profile—i.e. the by-age likelihood *(λ)* of acquiring an infection—that not only peaks between 5 and 6 years of age, but that is also independent of population size (Fig. 5). Similar trends have been noted in [4, 5], and thus serve as benchmarks for measles simulations.

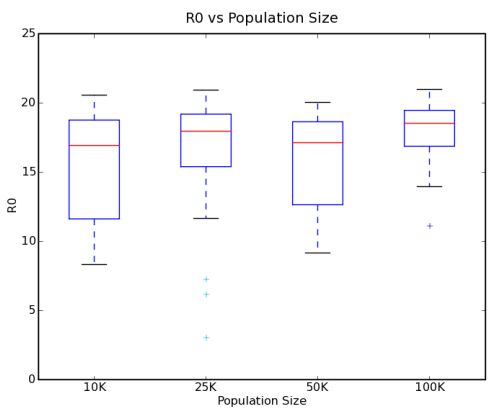
Figure 5: Force of infection by age cohort. Line colors differentiate among population sizes, keyed in the legend to the right of the graph. The distributions peak between 5 and 6 years of age and are robust to changes in population size.



3.2.4 Basic reproductive ratio (R_o) and population size

As was found in [1], our simulations demonstrate an R_o value ($R_0 = 17.6 \pm .75$) that is independent of population size and that falls within the range of generally-accepted reproductive ratio values for measles, $14 < R_o < 18$ [1] . Furthermore, the data plotted in (Fig. 6) are aggregate values from time series data, which, itself, shows little variation from one year to the next, suggesting that R_o is not affected by the underlying dynamics of the social network (on an annual timescale).

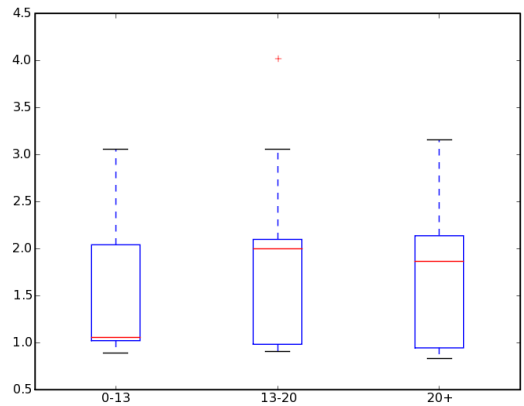
Figure 6: Basic reproductive ratio as a function of population size. For each population size, the box plot has been aggregated over 10 simulations, each spanning 30 years.



3.2.5 Baby booms and epidemic period

As has been well documented in [2], a sudden and dramatic change in the birth rate of a population, such as that experienced in England and Wales in the late 1940's, will cause the period attractor for measles epidemics in large populations to shift from biennial to annual. Over a ten year period (beginning at year 3 and terminating at year 13 of the simulations) we increased the overall birthrate by 30% and monitored the epidemic period during this simulated "baby boom"; between years 13 and 20 and once more between years 20 and 30 we again monitored the epidemic period. The simulations suggest that the median epidemic period is strongly annual during the baby boom era, but quickly returns to a biennial attractor once the boom has terminated. The biennial attractor stabilizes as time progresses (Fig. 7).

Figure 7: Median epidemic periods in a population of size $>N=10^5$ during a simulated baby boom (years 3-13), following the termination of the boom (years 13-20) and long after the termination of the boom (years 20-30). During the baby boom the birthrate has been increased by 30%.

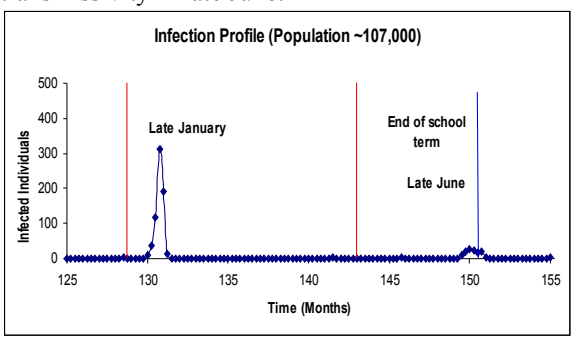


3.2.6 Seasonal dynamics of recurrent epidemics

Recent work by Stone *et al.* [3] has shown that for seasonally-driven diseases, such as measles, seasonal changes—and therefore, the time of year at which an outbreak begins or ends—can play a crucial role in determining whether an outbreak will develop into a full-scale epidemic, or whether it will be curtailed and result in a "skip" [3]. Stone *et al.* demonstrate that, in general, if an epidemic peaks early in a given year (during months 0-3), the following year will usually experience a skip; that is, there will either be no epidemic, or the epidemic will occur late and be curtailed by the changing of seasons from high transmissibility to low transmissibility. As can be seen in (Fig. 8), we observe similar trends in our simulations: for example, we see epidemics that peak early in the high-transmissibility season of one year, followed by a late outbreak, which is ultimately curtailed in the following year, due to the decrease in transmissibility with the onset of summer vacation. Furthermore, we find that in all cases for which a potential epidemic has been interrupted by a change in seasonal

transmissibility, we can clearly differentiate skips from continued, but decreased epidemics, since for the former, the number of susceptible individuals continues to increase immediately following the epidemic peak, while for the latter, this is not the case. Additionally, the mathematical criterion put forth in [3] to differentiate these seasonally-induced skips from seasonally-weakened epidemics (namely that if a skip occurs, the fraction of the population that is infected should be less than or equal to the local per capita replenishment rate), is satisfied for each occurrence in our simulations.

Figure 8: An example of seasonally-driven dynamics: the early-peaking outbreak in the first year (red lines indicate the beginnings of years; blue line indicates end of term) causes a late-breaking epidemic in the following year. Although this outbreak begins to gain footing, it is curtailed by the end of the school term, and the subsequent drop in transmissivity in late June.



4. Discussion

Synthesizing the contact network data and some of the information gleaned from our disease dynamics simulations provides a layer of insight into the interrelationship between the dynamics of the social contact network and the dynamics of the disease propagating on that network that cannot be readily obtained via compartmental modeling.

4.1.1 R_0 and school subgraph topology

Recall that the cumulative degree distribution of school-aged children becomes increasingly switch-like around $k=50$ as population size grows. We previously noted that the reason for this effect is a combination of fuller classrooms and more interclass connections in larger populations. Simply put, a child attending a school in a large city is more likely to be in a full or overcrowded classroom, and will probably contact more people in the course of his or her day in the hallways and lunchroom, on the playground and bus, etc., than will the average student in a small, rural setting. Thus, there is *heterogeneity* in connectivity that differentiates urban schools from rural schools. However, we surprisingly still see an essentially constant value of R_0 across population sizes. Implicitly, compartmental modeling has assumed *homogeneity* in connectivity between schools in large populations and schools in small populations as the root cause behind the constant value of R_0 across population sizes [2]. The results of our simulations suggest that this may not be the case, and that, instead, the *mean age at first infection* may have more to do with the independence of R_0 across population sizes. We find that regardless of population size, infection is most likely to gain initial footing in a classroom of six-year olds, in part because this is where transmission rates are highest. In smaller populations, where the interclass links are sparse and where the number of classes within a given age group tends to be small, infection of one class will not necessarily lead to infection of more classes—it will not lead to "school-hopping"—and therefore the mean age at first infection in smaller populations will generally show a strong bias towards six years of age (as is indicated by the peak in the force of infection profile at age six). On the other hand, in larger populations, we observe a "school-hopping" effect in epidemic dynamics, whereby if one classroom in a large population becomes infected, the high density of interclass links ensures that other classes of the same age will also immediately become infected. Since the transmission rate among six-year olds dominates other transmission rates, we again see a strong biasing of the mean age at first infection towards six years of age. Using the approximate relation $R_0 = L/\langle a \rangle$ [2], where L is the life expectancy of the population and where $\langle a \rangle$ is the mean age at first infection, a simple explanation for

why R_0 remains independent of population size lies in the fact that as population size grows, the mean age at first infection does not deviate—it is (on average) six years old, regardless of population size. We also find that transmissibility between elementary-aged children and high school-aged children is sufficiently low so as to prevent jumping of the disease between elementary schools and high schools in *both* small and large populations, even though the density of elementary-high school links is higher in larger populations. Thus, unless an outbreak initiates in a high school, it will almost certainly be restricted to elementary-aged children, ultimately preventing the R_0 value from changing with population size.

4.1.2 A topological metric for epidemic prediction

For different population sizes, a peak at a high degree in a specialized degree distribution or combination of distributions at time *t* or at time *t-1* proves to be a strong predictor of a global epidemic in our simulated networks at time *t* (or at the following time point). By defining *nonrecovered intraclass* and *interclass edges* as those edges that connect nodes of a school graph to infected or susceptible schoolchildren within and between classes, we can form the respective degree distributions that change in time both as the disease propagates through the school networks and as the topology of the school network changes due to social evolution.

We introduce a metric—the *predictive strength* of characteristic distribution, *i*—that allows us to determine the likelihood of observing a large-scale epidemic in the network at time *t* based on the observation that we have noted a peak in the degree distribution of, for example, nonrecovered intraclass edges at time *t*. If E_{Max} is the maximum size of all observed epidemics over many simulations, then the predictive strength (S_i) of a characteristic edge distribution *i* can be expressed as:

$$Si = \int_{E_{0\,\mathrm{Major}}}^{E_{Max}} P(E)dE$$

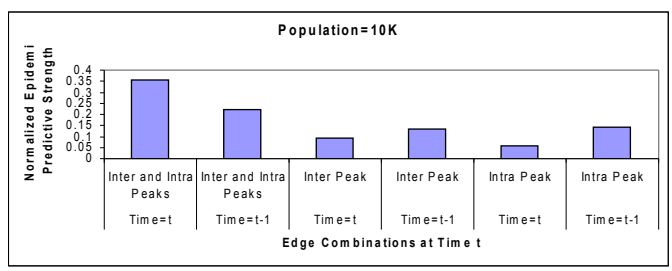
where *P(E)* is the probability of seeing an epidemic of size greater than *E*, and where the integration is over all (simulated) epidemic sizes of at least half the observed maximum epidemic size. Different characteristic edge distributions can then also be compared using a *relative predictive strength* metric:

$$\overline{s_i} = \frac{s_i}{\sum_{i=1}^{n} s_i}$$

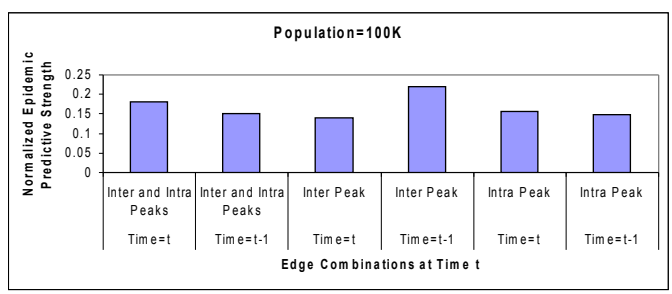
Analysis indicates a clear difference in characteristic behavior between small (N=10^4) populations and larger (>N=10^5) populations: while a peak in the mixed inter/intraclass nonrecovered edge distribution is clearly the strongest predictor of same-year epidemics in small populations (Fig. 9a), large populations exhibit a tie between this condition and a peak in the interclass nonrecovered edge distribution during the *previous* year (Fig. 9b).

Figure 9: Relative predictive strength of characteristic edge distributions in forecasting large-scale epidemics. (a) In a small population (N=10^4); (b) In a large population (>N=10^5).

a)



b)



The explanation for this difference in predictive behavior is largely topological.

The most likely indicator of an epidemic for small populations necessitates a set of topological conditions that is *rare* in small populations (it implies both very full, susceptible classrooms, and a dense abundance of interclass links terminating in a mixture of susceptibles and infecteds), but when these conditions do occur, a "chain-reaction" type large-scale epidemic is likely to propagate in a small network, just as it would in a large network where such conditions occur frequently, and lead to regular epidemics (and are therefore good epidemic predictors). However, large epidemics can also be achieved in large networks if there is a high density of interlinks terminating in infecteds at the previous time point; since there are always *more* large susceptible pockets (classrooms) in larger populations than there are in smaller populations, a protracted epidemic can result from this condition.

Conclusions

Computer simulation of detailed, evolving social networks and disease dynamics promises to bring to light many questions, such as the previously-mentioned R_0 curiosity, that may prove extremely valuable to epidemiological modeling, because such simulation affords us the opportunity to explore (on multiple scales) changes in topology concurrently with disease dynamics. Our results for a fairly simple set of algorithms reproduce the dominant dynamic trends for measles in populations that span two orders of magnitude, suggesting that the algorithms could be successfully tailored to less well-studied diseases.

References

[1] Bornstad, O., N., Finkenstadt, B. F., and Grenfell, B. T.: 'Dynamics of Measles Epidemics: Estimating scaling of transmission rates using a time series SIR model', Ecological Monographs, 2002, 72, pp. 169-184.

[2] Grenfell, B. T., Bjornstad, O. N., and Finkenstadt, B. F.: 'Dynamics of measles epidemics: scaling noise, determinism, and predictability with the TSIR model', Ecological Monographs, 2002, 72, pp. 185-202.

[3] Stone, L., Olinky, R., and Huppert, A.: 'Seasonal dynamics of recurrent epidemics', Nature, 2007, 446, pp. 532-536.

[4] Edmunds, W. J., Gay, N. J., Kretzschmar, M., and Pebody, R. G.: 'The pre-vaccination epidemiology of measles, mumpes and rubella in Europe: implications for modelling studies', Epidemiology and Infection, 2000, 125, pp. 635-650.

[5] Grenfell, B. T. and Anderson, R. M.: 'Pertussis in England and Wales: An Investigation of Transmission Dynamics and Control by Mass Vaccination', Proceedings of the Royal Society of London. Series B, Biological Sciences, 1989, 236, pp. 213-252.

[6] Glass, K., Kappey, K., and Grenfell, B. T.: 'The effect of heterogeneity in measles vaccination on population immunity', Epidemiol Infect, 2004, 132, pp. 675-83.

[7] Hanski, I. and Ovaskainen, O.: 'The metapopulation capacity of a fragmented landscape', Nature, 2000, 404, pp. 755-8.

[8] Ancel Meyers, L., Pourbohloul, B., and Newman, M. E. J.: 'Network theory and SARS: Predicting outbreak diversity', Journal of Theoretical Biology, 2005, 232, pp. 71-81.

[9] Toroczkai, Z. and Guchu, H.: 'Proximity Networks and Epidemics', arXiv:physics/0701255v1, 2007.

[10] 'Vital Statistics of New York State 1998 Tables', in 'Information for a Healthy New York, Health, N. Y. S. D. o., Ed., 1998.

[11] Scardamalia, R.: 'The Face of New York-- The Numbers', 2001.

[12] 'U.S. Department of Labor Bureau of Labor Statistics'.