The Functional Form of an Epidemic in a Real-World Contact Network

Lisa Brouwers & Fredrik Liljeros

Abstract

Epidemics initially grow exponentially in large populations with random homogeneous mixing. In a clustered network, where there is an increased probability that an individual's contacts will also have contact with each other, the initial speed of the outbreak will be slower than in a random network. If the level of clustering is high enough, the initial growth will be polynomial instead of exponential. In this paper we simulate the spread of an infectious disease in a highly clustered contact network, where the contacts are family and workplace contacts. The contact network is extracted from official governmental data on nine million Swedes. The experiment shows that, in spite of the high level of clustering, the simulated spread is almost exponential.

1 Introduction

The initial speed of an epidemic can be measured by the number of new infections per time unit. One policy consideration where information on the growth rate of an epidemic is of interest is whether an epidemic can be stopped by targeted vaccination or whether mass vaccination is required. Highly contagious diseases, such as measles, often grow exponentially in a susceptible population. This type of growth is characterized by a slow growth in the beginning, followed by an explosion of new infections, see Figure 1. For a disease that spreads exponentially, public health officials must be particularly alert to stop an epidemic from developing. Three main factors affect the course of an epidemic: (1) how infectious the disease is; (2) the time a person remains infectious; and (3) how many susceptible persons a carrier meets during the infectious period (Giesecke 2002). It has been demonstrated that extremely simplified assumptions about the contact structure in a population are often sufficient to create very precise transmission models (Andersson and May 1992, Diekmann and Heesterbeek 2000). A common simplification of the contact structure is to assume that all individuals are equally likely to meet during a certain time period, a principle referred to as random homogeneous mixing. In such models, all persons have more or less the same number of contacts, and each person's contacts are randomly chosen from the entire population. The probability that a contact is already infected is therefore very low during the early stages of an outbreak in a large population. Note that such models also can be represented as random networks, where the nodes are persons and the edges indicate that a contact has occurred.

The number of infected persons at time is determined by the rate at which

an infectious individual transmits the disease per unit of time, referred to as the transmission rate k, and the number of infectious persons at time t, referred to as I. The speed of new infections during the early stages of an outbreak can then be expressed by a differential equation,

$$\frac{dI}{dt} = kI \tag{1}$$

This differential equation has the solution,

$$I(t) = e^{kt} \tag{2}$$

The deterministic model described by Equation 1 predicts that the number of infected persons will rise exponentially ad infinitum. Naturally, this prediction is unrealistic since populations are finite. Sooner or later the number of infected persons will be large enough to slow down the rate of new infections, a phenomenon called *global saturation*. We focus, however, only on the early stages of an epidemic, when global saturation has not yet been achieved.

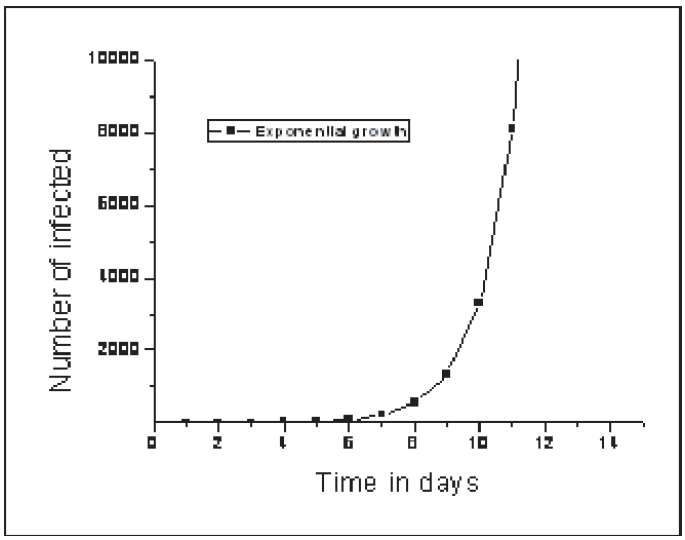

Figure 1: Illustration of an exponential growth process

2 Polynomial Growth Rate

A striking difference between human contact networks and random networks is that in human contact networks people tend to interact in different social arenas. They also tend to interact with almost all other persons in the

same arena, for example at workplaces and in households. It is characteristic of contacts within such stable social settings that the different contacts one individual has are likely to have contacts with each other as well (Newman 2003). In network research, this phenomenon is called clustering or transitivity (Scott 2000, Wasserman and Faust 1994).

A high level of clustering in a contact network has recently been suggested to slow down the initial growth from an exponential to a polynomial functional form, where is a constant (cf. Szendroi and Csanyi (2004). The rationale for this is that a local degree of saturation will set in rapidly in a clustered network because it is highly probable that the contacts of infected individuals are already infected. A similar tendency, that is, slower initial growth, has also been detected in spatially structured contact networks (Gastner and Newman 2004). The differences between exponential and polynomial growth of a disease are shown in Figures 3 and 4. The purpose of this research is to investigate whether the high level of clustering in a realistic human contact network will decelerate the initial spread from an exponential functional form to a polynomial functional form.

Here, a relevant objection could be that a polynomial growth rate is not always "better", that is, slower, than exponential growth. A rapid polynomial function does not necessarily have to be slower initially than slow exponential function (even though exponential growth always outpaces polynomial growth in the long run. Polynomial growth in a highly clustered

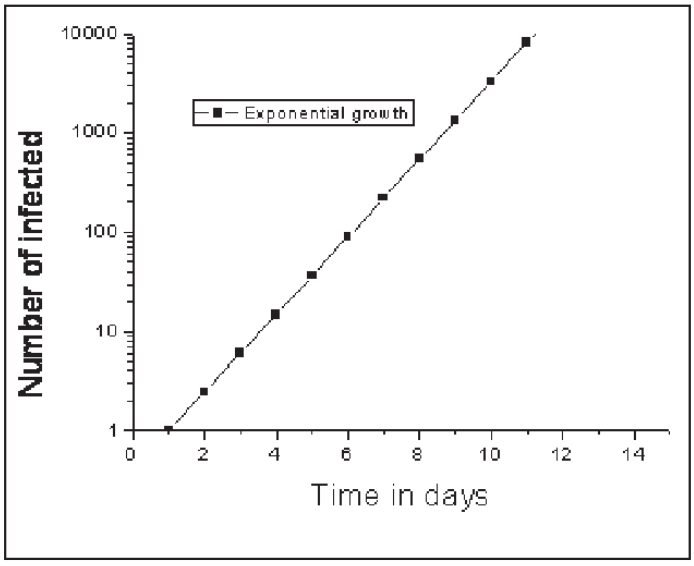

Figure 2: An exponential growth process in a graph with alogarithmic y-axis. An exponential trajectory takes the form of a straight line in such graphs.

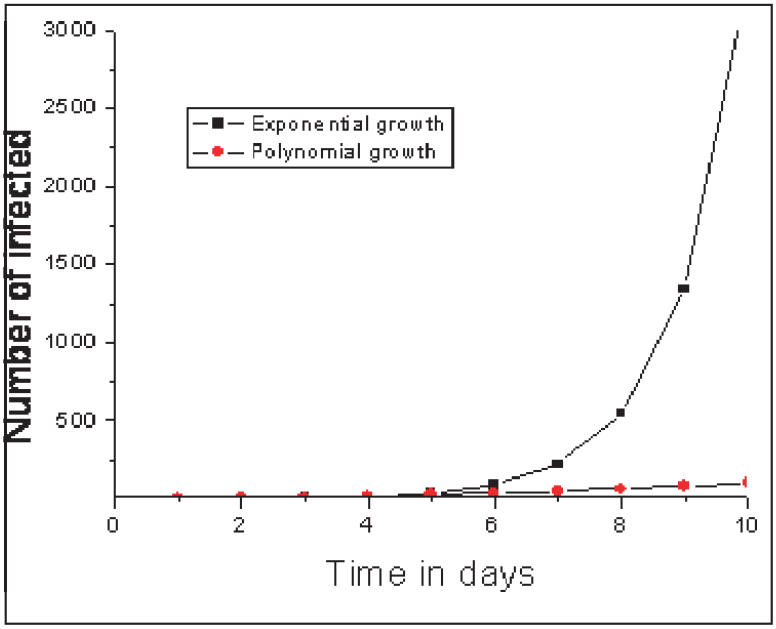

Figure 3: Exponential and polynomial growth in two outbreaks with similar growth rates the five first days, the first 10 days.

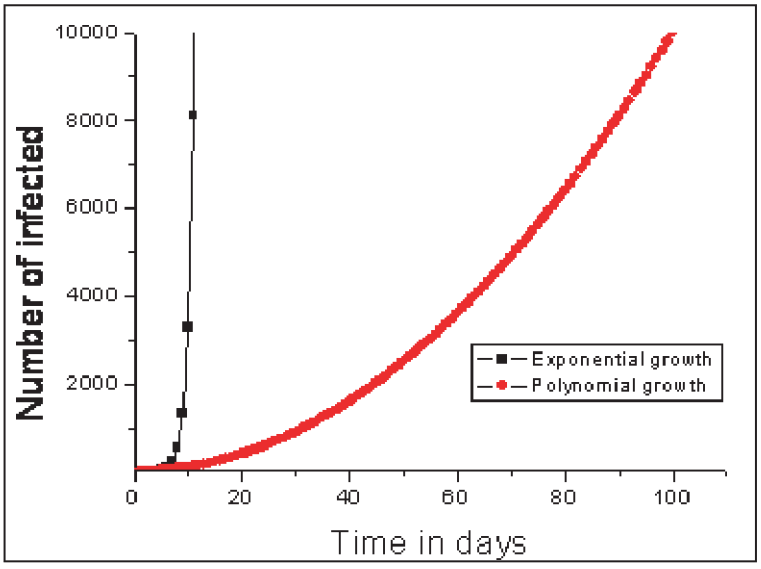

Figure 4: Exponential and polynomial growth in two outbreaks with similar growth rates the five first days, the first 100 days.

network will, however, be slower than exponential growth in a random network with a similar degree of distribution. We can therefore be certain that if the level of clustering is high enough to yield polynomial growth, it will be slower than an exponential growth in a similar network. Another possible objection is that, because it is nearly impossible to distinguish a polynomial growth process with large α (roughly $\alpha \geq 4$) from an exponential growth process during the short time interval under study, we cannot be sure that the growth is indeed reduced to a polynomial functional form in a clustered network. Such a high α value is, however, not realistic for most infectious diseases because it implies that the first infectious individual must infect 15 other persons during the first day of the infectious period. This can be compared with the spread of measles, for which it has been estimated that in a totally susceptible western population, an infectious individual on average infects a total of 15 other persons during the entire 6-8 days the person is infectious (cf. Giesecke (2002), Heymann (2004)). If measles followed a polynomial growth function, we would expect it to have an exponent $\alpha \approx 1.6$ to predict 16 cases day 6.

he remainder of this article is organized as follows. We describe the experiment by first briefly discussing the disease that is transmitted and its parameter values. We then present the MicroPox model and the experiment. After this we present the results of the simulation. The article concludes with a discussion of the results.

3 The Disease

We have chosen to model an outbreak of a disease that resembles smallpox (Heymann 2004). This decision is motivated by the fact that smallpox is known to be spread mainly through close contacts, and that it is therefore reasonable to assume that the contact network is of greater importance than it would be for a highly contagious disease like measles. Smallpox has also gained a great deal of publicity lately as a potential bioweapon used by terrorists (Henderson 1999). As a consequence, the disease has gained much attention among epidemiological modelers (Bozzette et al. 2003, Eubank et al. 2004, Kaplan et al. 2002). After September 11th, a number of simulation models evaluating potential smallpox policy interventions were published. Kaplan et al. elaborated one of the first models that investigated policy interventions against smallpox (Kaplan et al. 2002). From experimental simulation results that showed an apparently exponential progress of the outbreak, they concluded that mass vaccination would be a better alternative than targeted vaccination to hinder an epidemic. The success of

targeted vaccination-in this case to vaccinate only persons who have been in contact with an infected person-depends on the speed with which the disease spreads in relation to the speed with which the contacts of the infected persons are traced, referred to by Kaplan et al. as the "race to trace." The faster the disease spreads, the harder it is to control it by tracking down the infected persons, and to isolate and vaccinate them. Even as the model gained acceptance for the level of detail and realism with which the post-outbreak vaccination and contact tracing was represented, a number of its assumptions were also criticized as being simplistic (Halloran et al. 2002). The main critique was the underlying assumption of homogeneous mixing; in the model by Kaplan *et al.*, it was assumed that all persons met all other persons during each time unit. In response to this, several modelers have developed models that in various respects claim to be less simplistic (c.f. Halloran et al. (2002)).

We have, as far as possible, chosen to represent the disease that is transmitted in the model using the same parameter values as Kaplan et al. use, since it makes it possible to compare the results. A drawback of this approach is that Kaplan et al. make some unrealistic assumptions about when a person becomes infective and about the behavioral response, for instance assuming that all persons go to work during the entire infective period. Our results will therefore be a poor prediction of a real smallpox outbreak, but should instead be seen as representing the first stages of an outbreak of any moderately infectious disease that is transmitted through close contact in a socially structured network.

4 The MicroPox Model and the Experiment

MicroPox is a microsimulation model designed to represent the Swedish contact structure realistically enough to provide good estimates of the path of an epidemic and of the effect of various interventions such as ring vaccination and mass vaccination. For the purpose of this work, we will not use all features of the model, but only what is required to simulate the spread of the disease in households and at workplaces. The model is a microsimulation model, meaning that all 8 861 393 Swedes (size of the Swedish population when the dataset was collected) are represented in the model. The level of clustering is calculated as the average of all fractions between the "actual number of contacts between the neighbors" and the "maximum possible number of contacts between the neighbors" for each node (Watts and Strogatz 1998). In the contact network, extracted from governmental data, the level

of clustering is as high as 0.927.[1] A unique feature of the model is that it uses governmental registry data about where each person works, with whom the person works, and with whom the person lives, which makes it possible to extract a contact network that shows the work and family contacts. The contact network deterministically depicts the contacts between persons. A day in the simulation model is divided into day and night. During the first hour of the day, persons with a job go to work. If they are unemployed or retired, they stay at home. Schools and kindergartens are not are included in this experiment, since we do not have data on these "workplaces" for children. Children are therefore assumed to spend the days at home in this model. All persons return home after work and spend the night there with the family. Since transmission only occurs where persons are collocated, the transmission process proceeds at dwellings and workplaces. A simulation is run for a number of days. Two groups of parameters are set before the start of a simulation. We present the parameters together with the values that were used in the experiment.

- Simulation parameters
  - Number of days (100)
  - Number of replications (100)
- Model parameters
  - Population (8 861 393 persons)
  - Number of initially infected persons (30)
  - Transmission probability, TrP per place: (0.5 at dwellings, 0.1 at workplaces)
  - Maximum number of contacts NoC per place: (20 at dwellings, 25 at workplaces)
  - Incubation period (7 - 19 days from infection), see (Brouwers 2005) for distribution
  - Infectious period (3 days)

Transmission of the disease is done in the following way. The first step in the transmission process is to count the number of infectious members at each place. A place's current infection risk depends on:

1. the predefined transmission probability for that type of place, TrP

[1] The network used in these experiments has a slightly lower value since the number of contacts at workplaces has an upper limit, NoC.

2. the number of infectious people at that place

3. the maximum number of possible contacts at that specific place, NoC

The risk that infectious individual(s) at a certain place will infect a susceptible individual can be written as

$$P(infected|place)_t = 1 - (1 - TrP_{place} \cdot SW_{stage} \cdot \delta)^{NoInf}$$

where

$$\delta = \{ \begin{smallmatrix} 1 & :NoM \leq NoC \\ \frac{NoC}{NoM} & :NoM > NoC \end{smallmatrix} \quad (3)$$

The term δ is used to decrease the number of contacts at large workplaces. If the number of people at a workplace exceeds the value of the parameter maximum number of contacts, NoC for that type of place, the current infection risk from each single infectious individual is reduced by the proportion of the total number of members, NoM at the place.

5 Results

In Figure 5, the dynamic evolution of the average number of new infections is plotted on a graph in which the y-axis is logarithmic; pure exponential growth would generate a straight line, as indicated by the red line. In Figure 6, the same plot is displayed on a graph where both axes are logarithmic. Pure polynomial growth would generate a straight line, as indicated by the red line for which the value is three. Graph 5 shows clearly that the functional form of the growth of the simulated epidemic is close to exponential. Graph 6 gives further support for this interpretation. The error bars in Figures 5 and 6 show the upper and lower interval for the 95% of the simulations that are closest to the mean value. The functional form of the growth becomes steeper and steeper each day; it does not follow the polynomial form indicated by the red line.

The small initial deviations in growth during the first 30 days are likely to be explained by the stochastic incubation time (7-14 days), resulting in waves that can be separated initially but that later are blurred. A small tendency toward a sub exponential growth may possibly be observed in Figure 5.

6 Discussion

The results show that, in spite of the high level of clustering, the functional form of a moderately infectious disease will be close to exponential during

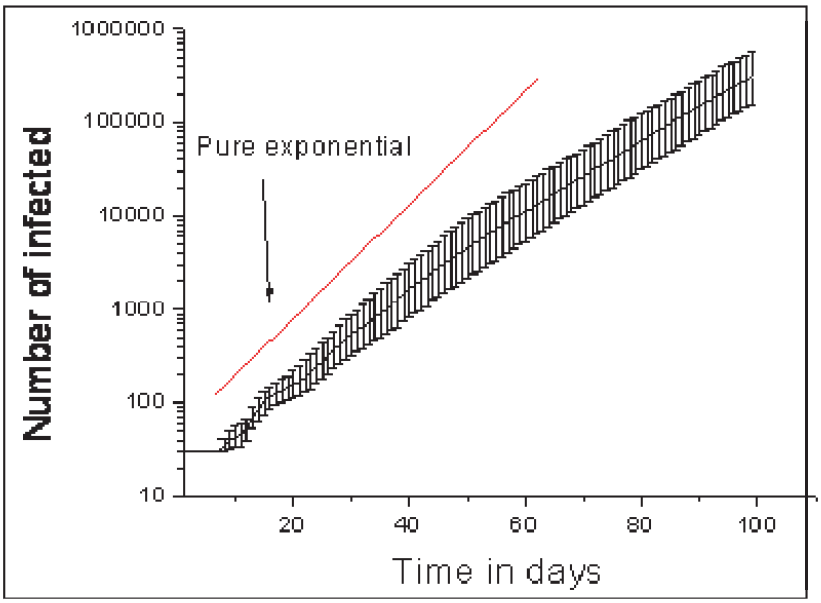

Figure 5: Evolution of average number of infected plotted on a graph with logarithmic y-axis.

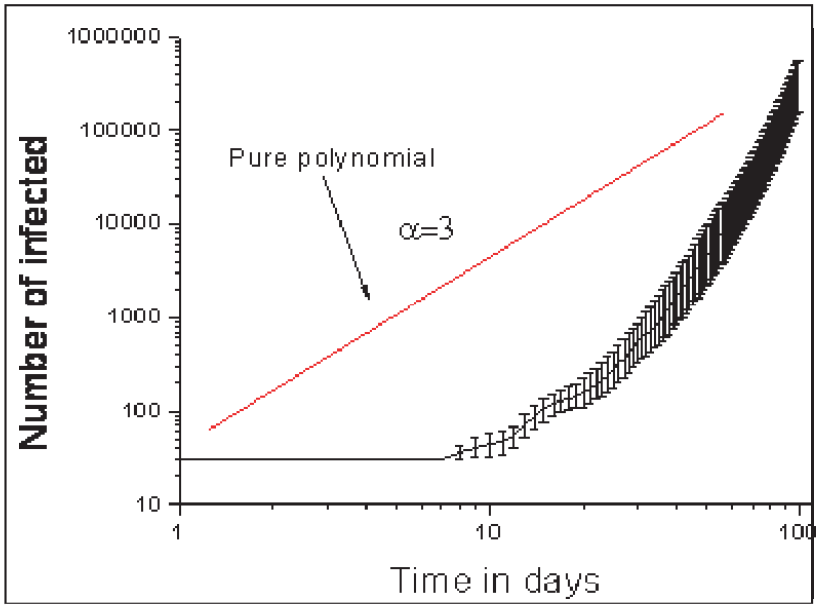

Figure 6: Evolution of average number of infected plotted on a graph with logarithmic x-axis and y-axis.

the early stages (before global saturation slows down spread). The results are to some extent surprising, considering how highly clustered the network is. A possible explanation is that the model assumes random homogeneous mixing at workplaces with more than 25 employees. If the number of members is larger than NoC (25 in this experiment), risk is reduced (see Equation 3). This was true for 62% of all working persons and for 32% of the total population. Another possible explanation is that the contact structure under study is a small-world network as suggested by Watts and Strogatz Watts and Strogatz (1998). They have shown that it is often enough to add a small fraction of contacts between a set of randomly selected pairs of nodes to give an ordered network an average shortest path length[2] similar to that of a random network with similar number of nodes and links. A disease would therefore be transmitted more or less as rapidly in a small-world network as in a random network. In this model, such "additional random" connections could be persons who live far from their workplaces, who therefore tie together different parts of the network. For policy purposes it is important to use transmission models that represent the contact structure realistically. This experiment shows that, in Sweden and for a disease that is spread in a similar way, the high level of clustering will not be able to slow down the speed of the outbreak enough to give it a polynomial form.

[2]To calculate the average shortest path of a network: for all possible node pairs, identify the shortest path between them, and then calculate the average of these paths for the entire network.

References

Andersson, R. M. and May, R. M.: 1992, *Infectious Diseases of Humans:Dynamics and Control*, Oxford Univ Press, Oxford.

Bozzette, S. A., Boer, R., Bhatnagar, V., Brower, J. L., Keeler, E. B., Morton, S. C. and Stoto, M. A.: 2003, A model for a smallpox-vaccination policy, *N Engl J Med* **348**(5), 416–425.

Brouwers, L.: 2005, MicroPox: A Large-Scale and Spatially Explicit Microsimulation Model for Smallpox Planning, *in* V. Ingalls (ed.), *The Proceedings of the 15th International Conference on Health Sciences Simulation*, Society for Modeling and Simulation International (SCS), San Diego, California, pp. 70–76.

Diekmann, O. and Heesterbeek, J.: 2000, *Mathematical Epidemiology of Infectious Diseases: Model Building, Analysis and Interpretation*, John Wiley & Sons Ltd., Chichester.

Eubank, S., Guclu, H., Kumar, V., Marathe, M., A.Srinivasan, Z.Toroczkai and N.Wang: 2004, Modelling disease outbreaks in realistic urban social networks, *Nature: Letters to Nature* **429**(6988), 180–184.

Gastner, M. and Newman, M.: 2004, The Spatial Structure of Networks. con-mat/0407680.

Giesecke, J.: 2002, *Modern Infectious Disease Epidemiology*, 2nd edn, Arnold, London, chapter 11, Mathematical Models for Epidemics, pp. 119–132.

Halloran, M., Longini, I., Nizam, A. and Yang, Y.: 2002, Containing bioterrorist smallpox, *Science* **298**(5597), 1428–1432.

Henderson, D. A.: 1999, The looming threat of bioterrorism, *Science* **283**, 1279.

Heymann, D.: 2004, *Control of Communicable Diseases Manual*, American Public Health Association, Washington.

Kaplan, E. H., Craft, D. L. and Wein, L. M.: 2002, Emergency response to a smallpox attack—the case for mass vaccination, *PNAS* **99**(16), 10935–10940.

Newman, M.: 2003, Properties of Highly Clustered Networks, *Physical Review* **68**(2).

Scott, J.: 2000, *Social Network Analysis: A Handbook*, SAGE, London.

Szendroi, B. and Csanyi, G.: 2004, Polynomial Epidemics and Clustering in Contact Networks, *Proc. R. Soc. Lon. B (Suppl. Biology Letters)* **271**, 356–366.

Wasserman, S. and Faust, K.: 1994, *Social Network Analysis : Methods and Applications*, Cambridge University Press, Cambridge.

Watts, D. and Strogatz, S.: 1998, Collective Dynamics of 'small-World' Networks, *Nature* **393**(6684), 440–442.