Los Alamos National Laboratory
Rensselaer Polytechnic Institute
University of Maryland

Modeling Disease Outbreaks in Realistic Urban Social Networks[†]

S. Eubank[1], H. Guclu[2], V. S. A. Kumar[1], M. V. Marathe[1], A. Srinivasan[3], Z. Toroczkai[4], and N. Wang[5]

[1] *Basic and Applied Simulation Science Group,* ***Los Alamos National Laboratory****, MS M997, Los Alamos, New Mexico 87545, USA*
[2] *Department of Physics, Applied Physics and Astronomy,* ***Rensselaer Polytechnic Institute*** *110 8th Street, Troy, New York 12180-3590, USA*
[3] *Department of Computer Science and Institute for Advanced Computer Studies,* ***University of Maryland****, College Park, Maryland 20742, USA*
[4] *Centre for Nonlinear Studies and Complex Systems Group,* ***Los Alamos National Laboratory****, MS B258, Los Alamos, New Mexico 87545, USA*
[5] *Department of Computer Science,* ***University of Maryland****, College Park, Maryland 20742, USA .*

Abstract

Most mathematical models for the spread of disease use differential equations based on uniform mixing assumptions or *ad hoc* models for the contact process. Here we explore the use of dynamic bipartite graphs to model the physical contact patterns that result from movements of individuals between specific locations. The graphs are generated by large-scale individual-based urban traffic simulations built on actual census, land-use and population-mobility data. We find that the contact network among people is a strongly connected *small-world*-like graph with a well-defined scale for the degree distribution. However, the locations graph is *scale-free*, which allows highly efficient outbreak detection by placing sensors in the hubs of the locations network. Within this large-scale simulation framework, we then analyze the relative merits of several proposed mitigation strategies for smallpox spread. Our results suggest that outbreaks can be contained by a strategy of targeted vaccination combined with early detection without resorting to mass vaccination of a population.

Introduction

The dense social-contact networks characteristic of urban areas form a perfect fabric for fast, uncontrolled disease propagation. How can an outbreak be contained before it becomes an epidemic, and what disease surveillance strategies should be implemented? Recent studies, under the assumption of homogeneous mixing, make the case for mass vaccination in response to a smallpox outbreak. With different assumptions, it has been shown that mass vaccination is not required. Here we present a highly resolved agent-based simulation tool (**EpiSims**), which combines realistic estimates of population mobility, based on census and land-use data, with parameterized models for simulating the progress of a disease within a host and of transmission between hosts. EpiSims is based on the Transportation Analysis and Simulation System (**TRANSIMS**) developed at Los Alamos National Laboratory, which produces estimates of social networks based on the assumption that the transportation infrastructure constrains people's choices about where and when to perform activities. The case study we present is a model of Portland, Oregon, USA, but the approach is broadly applicable. People, in the course of carrying out their daily activities (such as work, study or shopping), move between several locations, both exposing themselves to infectious agents within these locations and transporting those agents between locations. We represent these processes by a social contact network, which can be represented as a bipartite graph, G_{PL}, as shown by the example in Fig. 1a. For Portland, G_{PL} has about 1.6 million vertices, with a giant component of about 1.5 million people and 180,000 locations. The degree distribution of the people vertices in G_{PL}, that is, the number of people Q_j^{PL} who visited j different locations, is shown in Fig. 2a. It has a sharp peak near the average value of about four different locations, followed by a fast, exponentially decaying tail. The degree distribution for the location vertices in G_{PL} is very different, as shown in Fig. 2b. This is the number of locations M_i^{PL} having i different visitors during the day. The distribution has a power-law tail with an exponent of about -2.8.

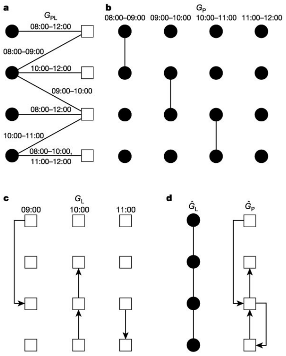

Figure 1 An example of a small social contact network. **a**, A bipartite graph G_{PL} with two types of vertex representing four people (P) and four locations (L). If person p visited location l, there is an edge in this graph between p and l. Vertices are labeled with appropriate demographic or geographic information, edges with arrival and departure times. **b**, **c**, The two disconnected graphs G_P and G_L induced by connecting vertices that were separated by exactly two edges in G_{PL}. **d**, The static projections $\hat{G}_P$ and $\hat{G}_L$ resulting from ignoring time labels in G_P and G_L. People are represented by filled circles, and locations by open squares.

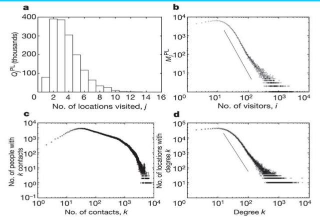

Figure 2 Degree distributions for the estimated Portland social network. **a**, The number of people Q_j^{PL} who visited j different locations in the bipartite people–locations graph G_{PL}. **b**, The number of locations M_i^{PL} in G_{PL} that are visited by exactly i different people. The slope of the straight-line graph is -2.8. **c**, The number of people who have k neighbors in the static people–contact graph $\hat{G}_P$ on log–log scale. **d**, The in and out degree distributions of the locations network G_L. The slope of the straight-line graph is -2.8.

The Model

For many infectious diseases, transmission occurs mainly between people who are collocated (simultaneously in the same location), and spread is due mainly to people's movement. Hence we look at two natural projections of G_{PL} obtained by drawing an edge between all pairs of vertices distance two from each other on the bipartite graph, as illustrated in Fig. 1b, c. The result is two disconnected graphs: G_P, containing only people vertices, and G_L, containing only locations. In G_P, the edges are labeled with the sets of time intervals during which the people were collocated. For simplicity, however, we consider $\hat{G}_P$, a static projection of the time-resolved G_P, obtained by discarding time labels, as shown in Fig. 1d. Figure 2c shows the degree distribution of $\hat{G}_P$ for the Portland network. The other important projection of the bipartite graph is the locations network G_L. If there is at least one person traveling from location l_1 directly to l_2 during the day, the two vertices corresponding to locations l_1 and l_2 are connected by a directed edge in G_L from l_1 to l_2 that indicates whether the person is traveling in or out of the location. The in and out degree distributions for the locations network are superimposed in Fig. 2d. The power-law decay evident there shows that $\hat{G}_L$ is a scale-free network with an exponent of -2.8.

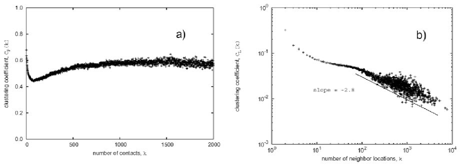

Figure 3 Clustering coefficients by degree for a) the people contact network, and b) the locations network (after discarding the direction of edges in the latter)

Results

Measurements of the average clustering coefficient for $\hat{G}_P$ yield $C_P \approx 0.48$, and for $\hat{G}_L$, $C_L \approx 0.04$, both much larger than the roughly 10^{-6} of an Erdös–Rényi random graph with the same number of vertices and average degree. This, together with the degree distribution and its small diameter (about 6), suggests that the people-contact graph is more like a *small-world* graph than a random graph. The clustering coefficient versus degree shown in Fig. 3 indicate that the locations network $\hat{G}_L$ is an empirical example of a hierarchical scale-free structure. It is natural to consider estimation schemes for global topological measures, such as expansion. Informally, the higher the expansion, the quicker is the spread of any phenomenon (such as disease, gossip or data). We estimated an expansion value of about 2 for $\hat{G}_P$ by random sampling, indicating that the people-contact graph is extremely connected. An immediate consequence is that, as for an assortatively mixed network, $\hat{G}_P$ cannot be shattered by removing (by means of vaccination or quarantine) a small number of high-degree vertices. To verify this, we have computed the size of the giant component—the maximum number of people at risk for disease introduced by a single person—when all vertices of degree more than k are removed. A unique giant component persists even when all vertices of degree 11 and higher are removed, as shown in Fig. 4a. Thus, attempting to shatter the contact graph by vaccinating the most gregarious people in a population would essentially be equivalent to mass vaccination. Similarly, we show in Fig. 4b, c that closing the most-visited locations—or vaccinating everyone who visits them—does not shatter the induced people-contact graph until large fractions of the population have been affected.

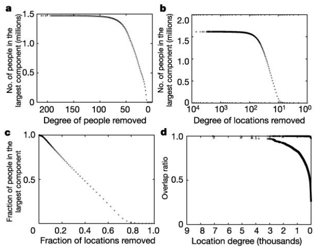

Figure 4 Shattering and covering the people–contact graph. In **a** we remove (by vaccination or quarantine) all people with degree k and higher from the bipartite graph G_{PL}. In **b** and **c** we remove all locations with degree k or higher from G_{PL} and monitor the size of the largest connected component in the static people–contact graph induced by the remaining bipartite graph. **d**, Overlap ratios by degree. The lower curve shows the cumulative overlap ratio by degree, which is the overlap ratio for locations having degree k or less. The upper curve shows the overlap ratio for locations having degree exactly k.

Can epidemics be stopped without resorting to mass vaccination? Alternatives rely on early detection and efficient targeting. Here we introduce the overlap ratio, another non-local property of the graph that is crucial to early detection. Consider an idealized situation in which sensors at a location can detect whether any person there is infected. The feasibility of early detection depends on the number of sensors required to cover the population. This problem is equivalent to finding the minimum dominating set. The overlap ratio by degree is shown in Fig. 4d. Clearly, not many people visit more than one high-degree location, which implies that the high-degree location vertices form a near-optimal dominating set. With high probability, early identification could be accomplished by using sensors placed at locations with the highest degree.

There is not yet a consensus on models of smallpox. We have designed a model that captures many features on which there is widespread agreement and allows us to vary poorly understood properties through reasonable ranges. We studied the sensitivity of the number of casualties to three factors: mitigation efforts, delay in implementing mitigation efforts, and whether people move about while infectious. We simulated a passive (do nothing) '*baseline*' and three active responses: mass vaccination covering 100% of the population in 4 days ('*mass*'); targeted vaccination and quarantine with unlimited resources ('*targeted*'); and the same targeted response, using only half as many contact tracers and vaccinators ('*limited*'). For a movie showing the spatial spread of disease under two different response strategies, see [*]. Figure 5 compares the efficacy of these strategies. For each strategy we plot (on a logarithmic scale) the ratio of the cumulative number of deaths by day 100 to the number initially infected. The absolute numbers are less important than the rank and relative sizes of gaps between the points. Also shown are the effects of delays of 4, 7 or 10 days in implementing the response. For each of the responses including the baseline, we allowed infected people to isolate themselves by withdrawing to the home. This could be due either to the natural history of the disease, which incapacitates its victims, or to actions taken by public health officials encouraging people to stay home. The results are grouped according to time of withdrawal to the home: (1) *early*, in which everyone withdraws before becoming infectious; (2) *late*, in which everyone withdraws about 24 h after becoming infectious; and (3) *never*, in which everyone carries on their daily activities unless they die. The extreme cases are unrealistic but are shown here because they demonstrate the existence of a clear transition. In this study, time of withdrawal to the home is by far the most important factor, followed by delay in response. This indicates that targeted vaccination is feasible when combined with fast detection. Ironically, the actual strategy used is much less important than either of these factors.

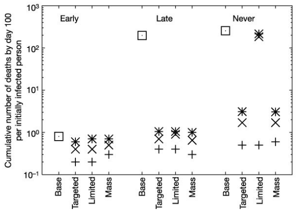

Figure 5 Cumulative number of deaths per number of initial infected, for the case of a smallpox outbreak in downtown Portland, under a number of different response strategies: squares, no vaccine; stars, 10-day delay; multiplication signs, 7-day delay; plus signs, 4-day delay.

Acknowledgments

We thank G. Korniss, G. Istrate and the Fogarty International Center at the National Institutes of Health for useful discussions, and acknowledge the work of all the members of the TRANSIMS and EpiSims team. The EpiSims project is funded by the National Infrastructure Simulation and Analysis Program (NISAC) at the Department of Homeland Security. The TRANSIMS project was funded by the Department of Transportation. H.G. was supported in part by the National Science Foundation (Division of Materials Research) and Z.T. by the Department of Energy.

Summary

Assessments of how best to respond to bio-terrorist attack have come up with conflicting results in the matter of smallpox vaccination. Is mass vaccination vital? Or can targeted vaccination of mobile at-risk individuals be effective? Our work suggests that, if the smallpox release is detected promptly and the population retreats home quickly, targeted vaccination can do the job in an urban situation. This work involved the EpiSims epidemiological simulation system, a derivative of the TRANSIMS system produced at the Los Alamos National Laboratory to simulate regional traffic movements. The traffic grid is a good proxy for a social network as it is transport infrastructure that constrains people's choices about where to go, and when to go there.

References and Contact

[†] Eubank et al., Nature **429**, 180 (2004)

[*] For supplementary information:

http://www.nature.com/nature/journal/v429/n6988/suppinfo/nature02541.html

Contact: S.E. (eubank@lanl.gov)