Controlling disease spread on networks with incomplete knowledge

B. Dybiec*

Institute of Physics, Jagellonian University, 30-059 Kraków, Poland

A. Kleczkowski[†] and C. A. Gilligan[‡]

Department of Plant Sciences, University of Cambridge, Cambridge CB2 3EA, England

(Received 19 January 2004; revised manuscript received 1 October 2004; published 28 December 2004)

Models for control of highly infectious diseases on local, small-world, and scale-free networks are considered, with only partial information accessible about the status of individuals and their connections. We consider a case when individuals can be infectious before showing symptoms and thus before detection. For small to moderately severe incidence of infection with a small number of nonlocal links, it is possible to control disease spread by using purely local methods applied in a neighborhood centered around a detected infectious individual. There exists an optimal radius for such a control neighborhood leading to the lowest severity of the epidemic in terms of economic costs associated with disease and treatment. The efficiency of a local control strategy is very sensitive to the choice of the radius. Below the optimal radius, the local strategy is unsuccessful; the disease spreads throughout the system, necessitating treatment of the whole population. At the other extreme, a strategy involving a neighborhood that is too large controls the disease but is wasteful of resources. It is not possible to stop an epidemic on scale-free networks by preventive actions, unless a large proportion of the population is treated.

I. INTRODUCTION

The last two decades have seen several large-scale epidemics of international importance, including human, animal, and plant epidemics. Notable among these are SARS [1], foot-and-mouth disease [2], Dutch elm disease [3], citrus canker [4], sudden oak death [5], and rhizomania [6]. Apart from wide geographic range, massive losses, and large costs for attempted containment, the epidemics have several factors in common. They all spread on complicated networks with a mixture of short- and long-range links that are often difficult or impossible to identify, despite great effort invested in tracking contacts [2]. There is also incomplete knowledge about the epidemiological status of individuals (humans, animals, herds or farms, fields or plants). An infected individual can initially go undetected or untreated, while spreading the disease to other individuals.

In this paper, we compare the efficiency of *local* control strategies for stopping the spread of a disease on networks with a mixture of local and global links and cryptic infection. The topologies range from purely local spread on a one-dimensional lattice, typical of epidemics of coastal sea mammals [7], and a two-dimensional lattice, through "small-world" networks [8] with predominantly local spread and a small proportion of global links, to scale-free networks [9]. We assume that in attempts to control the epidemic, knowledge about the network structure is limited to local links only. The local links are much easier to define and track for a real-life epidemic (as they are largely determined by geo-graphical location) than long-range (occasional) links. The control of large outbreaks also raises the question of economic aspects of disease treatment [10]. Traditionally, successful control was understood in terms of reducing the number of individuals affected by the disease, regardless of cost. This might be justified when the treatment is cheap, but often the epidemic must be stopped at a manageable cost and with limited resources [11]. Here we adopt a simple approach to optimize a control strategy for a linear sum of the numbers of individuals that have been through the disease and the numbers that have been treated.

We also assume that upon infection, each individual enters a presymptomatic state in which it can infect other individuals, but does not trigger any control action. We compensate for the lack of knowledge by extending the control of the disease to a larger, but still local, neighborhood. Even when there are no long-range links, the preventive measures need to cover an extended neighborhood so that it contains all potentially infected, but not yet symptomatic, individuals. When long-range links are included, it is more likely that the disease has already been transmitted outside the ring before detection. However, we show that if the control ring is large enough and there are only few shortcuts, there is still a possibility of stopping the spread of disease.

II. MODEL

We consider three basic topological structures, a one-dimensional ring (SW1D) and a two-dimensional regular lattice (SW2D), both with the addition of shortcuts leading to small world topologies [8], and a scale-free topology (SF) [9]. The disease spreads on the full network, including any local and global links. The control actions can only follow a

---

*Electronic address: bartek@th.if.uj.edu.pl
†Electronic address: adamk@mathbio.com
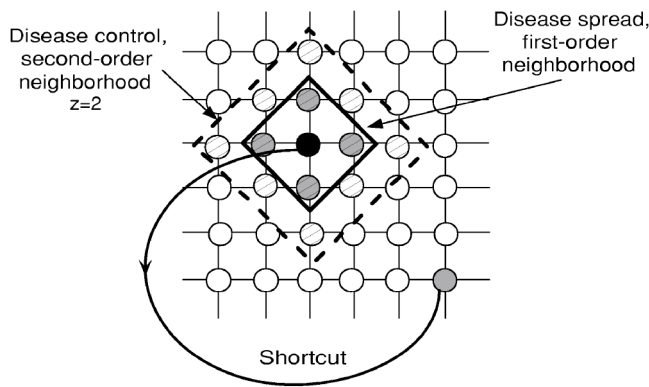‡Electronic address: cag1@cam.ac.uk

FIG. 1. SW2D topology: In this example, a symptomatic individual (black circle) is in contact with its four nearest neighbors on the disease network (marked by a solid-line square) and to one node connected by a shortcut, all marked by gray circles. The control is limited to the 12 neighbors in up to the second-order neighborhood, i.e., eight second-order neighbors and four first-order neighbors, and the symptomatic individual itself, on a treatment network (marked by hatched circles and a broken-line square), $z=2$.

subset of those links, and in particular for the SW1D and SW2D topologies we assume that the subset contains local links only (Fig. 1); see below for details of the SF network.

For the SW2D topology (Fig. 1), the starting point is a two-dimensional regular lattice with periodic boundary conditions, with the first-order neighborhood spanning four nearest neighbors, the second-order neighborhood including eight neighbors, etc. The SW1D topology is constructed in the same way, taking a ring (1D chain with periodic boundary conditions) as a starting point. For compatibility with the SW2D topology, we define the first-order neighborhood as spanning four nearest neighbors (two on the left and two on the right). Higher-order neighborhoods are then defined iteratively (with four neighbors in each order). For SW1D and SW2D, the disease can spread locally within the first-order neighborhood and along additional links (shortcuts) leading to individuals not in the first-order neighborhood. The individuals that are located in the local neighborhood up to a fixed order, z, can be treated by control actions ($z \geq 1$). The control network excludes all additional long-range links.

The SF topology for the disease spread is created following the Barabási-Albert algorithm [9]. To the initial fully connected core of size C, additional nodes are added as follows. Every new node is connected to C older nodes, which are chosen according to a preferential attachment rule, i.e., the probability that the new node is connected to the node i is equal to $k_i / \Sigma k_i$, where k_i is the number of links already attached to a node i. We assume that disease can spread to all nodes connected to an infected node by a direct link only, on a full (infection) network. Incomplete knowledge of links for controlling the epidemics is modeled by considering a subset of the full network (control network), with the first C_I ($C_I \leq C$) links from the nodes outside the core included only. Every node from the core can be connected to up to C_I or $C-1$ (whichever is smaller) nodes with smaller numbers. Such a mechanism guarantees that the control network is not

disconnected and that for $C_I=C$, the control network is equivalent to the infection network. The control algorithm allocates neighbors by following all links to a given order, z, on the control network. Infection and control networks for the SF topology correspond to epidemic and control neighborhoods for the SW1 and SW2 topologies.

For the epidemiological part of the model, we choose a SIR model [12] modified so that it includes pre- and post-symptomatic individuals (both contributing to the spread of infection) and distinguishes between recovered and treated nodes. We assume that all nodes in a network are occupied. The initial state is a mixture of the majority of susceptible individuals, with the addition of a few (0.5% of the total population) infectious (and symptomatic) individuals. This corresponds to the initial stage of an epidemic at the moment when decisions about controlling the outbreak need to be made. A susceptible individual **S** can be infected with probability p by any infectious or symptomatic individual in its infection neighborhood. Upon successful infection, it moves to an infected and presymptomatic class **I**. When the disease is diagnosed, an individual moves to the detected class **D** with probability q. A detected individual can subsequently move to the recovered class **R** (with probability r of a spontaneous recovery) or can trigger a control measure (with probability v). Control affects all individuals within the control neighborhood (including the one that has triggered it). Neither **R** nor **V** can recover or become reinfected. The transitions and their probabilities are summarized in the following graph:

$$
\begin{array}{ccc}
 & & \mathbf{R} \\
 & & {\scriptstyle r}\nearrow \\
\mathbf{S} \xrightarrow{p} \mathbf{I} \xrightarrow{q} \mathbf{D} & & . \\
 & & {\scriptstyle v}\searrow \\
 & & \mathbf{V}
\end{array}
\tag{1}
$$

We denote the total number of nodes by N and the number of susceptible nodes by S, infected by I, detected by D, recovered by R, and treated by V.

Transition probabilties were calculated and the state of the system updated at each step, with a fixed (and small) time increment. The simulation loop was performed until $t = \mathrm{T}_{max}$ or $I(t) + D(t) = 0$, whichever occurred sooner. An arbitrary value of T_{max} was selected to allow all simulations to finish due to lack of infectious and detected individuals. For a given set of parameters, 2500 realizations were simulated, for the total number of nodes equal to 2500. The sample was subdivided into 50 groups, and R and V computed for each group. Average optimal radius and its variability were then estimated based on data for each of the 50 groups (for details, see below). Different sizes of disease spread neighborhood, different initial conditions, larger sizes of the network, and larger sample sizes were explored as well, but they did not change the results qualitatively.

III. RESULTS

We define a severity index, $X \equiv R(\infty) + V(\infty)$, representing a linear combination of the severity of an outbreak at the end of an epidemic measured by the number of affected individuals (R) and the costs of treating it (the number of treated individuals, V). We varied all parameters, except r. The treatment range z (order of the treatment neighborhood) varied between 1 and 15. The rate of a spontaneous recovery was small, $r=0.01$, giving $R(\infty)/V(\infty) \lesssim 0.037$, so X is dominated by the number of treated individuals, $V(\infty)$.

First, we consider the case with no shortcuts and concentrate on the role of undetected infectious individuals in designing control strategies. The general shape of the severity index, $X=R(\infty)+V(\infty)$, as a function of the range of the control neighborhood, z, and the probability of infection p (Fig. 2), is very similar for all networks. For fixed and not very small p, the severity index is high for small values of z, decreases rapidly when z reaches an optimal value $z=z_c$, and subsequently increases again. Thus, for highly infectious diseases (large p) or when the treatment neighborhood is small (small z), the control is unsuccessful, and almost all individuals are either affected by the disease or need to be treated preventively. In contrast, for large z, the epidemic is stopped very quickly, but X increases with z as more and more individuals are treated. For small p, the control is successful for the SW1D and SW2D topologies, leading to small values of z_c and X. For the SF topology, the severity index is very high for all values of p and z, and the minimum is not well defined (similarly for SW1D and SW2D networks with many shortcuts, see below).

X and z_c both increase with increasing probability of infection, p (Fig. 3), and decrease with increasing probability of detection, q. The SW1D topology is easiest to control and produces less severe epidemics than the SW2D topology. For the SF networks, we need to extend the control to almost the whole population for all but the lowest values of p (or for $q \simeq 1$). The difference is primarily due to the actual size of the neighborhood of a given order in different topologies. For example, a single control neighborhood of order $z=8$ includes 1.3% of all nodes in the SW1D topology, 5.8% in the SW2D topology, and about 48% in the SF topology (for $N=2500$, $C=5$, $C_I=1$; for SF this proportion is even higher for larger C_I/C ratios). Increasing the probability for detection decreases z_c and X of the disease at this optimum, but the decrease is small and quickly settles on an asymptote.

Addition of shortcuts makes the local control less efficient as the disease may jump to new locations outside the control ring (Fig. 1). The X dependence on p and z is similar to that in Fig. 2, but the surfaces are shifted upwards and minima become less pronounced. If the proportion of long-range links is small, it is still possible to control the disease by local preventive treatment. The qualitative behavior of z_c and X does not change for the proportion of shortcuts up to 1.3% of the total number of links (Fig. 4), although the severity increases slightly due to more infections being produced before the disease is stopped. When the number of shortcuts is larger, z_c becomes large, X to N, and the disease escapes control. The behavior becomes then similar to the SF net-
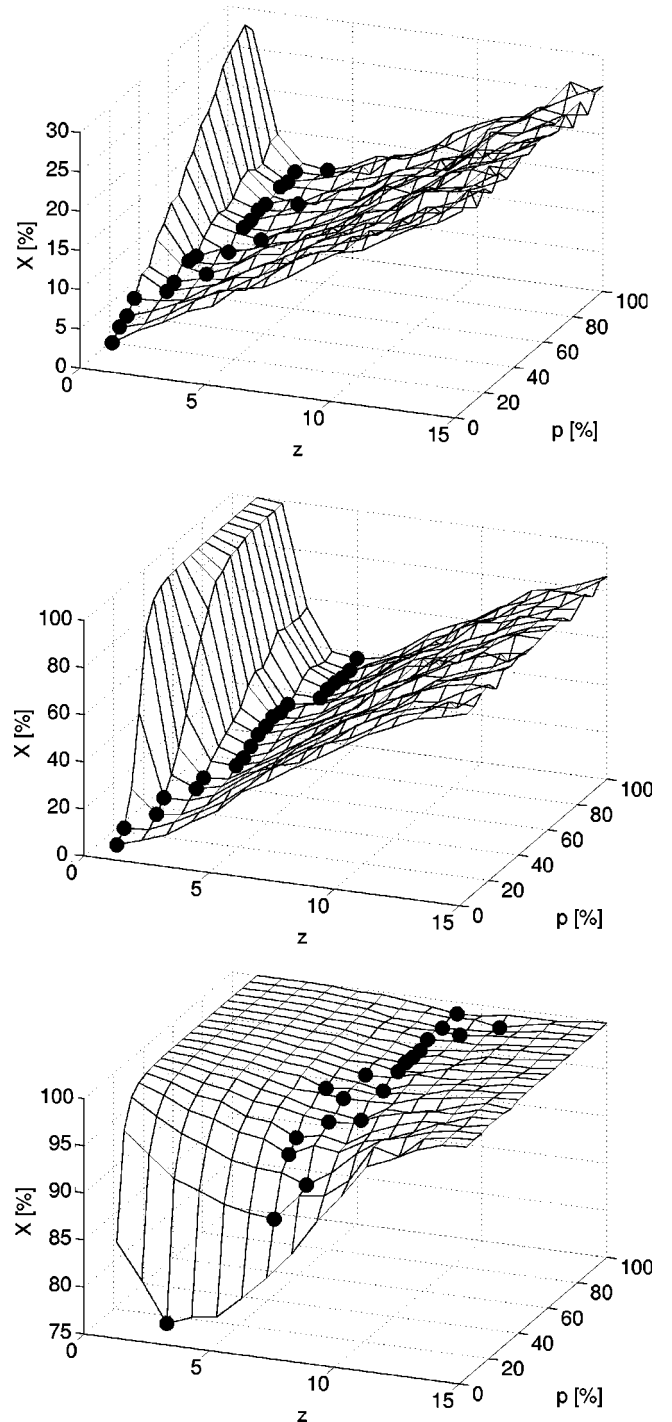

FIG. 2. Severity index $[X=R(\infty)+V(\infty)]$ as a function of the treatment neighborhood, z, and probability of infection, p, for (from top to bottom) the SW1D, SW2D, and SF topology (average over 50 replicates). Large dots show the optimal range of the control neighborhood, z_c, for a given probability p. There were no shortcuts for the SW1D and SW2D networks. Other parameters are $q=0.5$, $r=0.01$, $v=0.1$, $C=5$, $C_I=1$. X is an average value over 50 simulation runs.

works (Fig. 4), where most individuals need to be treated unless p is very small. For the SF network, $z_c \approx 8$ (not shown in Fig. 4 for clarity) and about 48% of the population is treated in a single action.
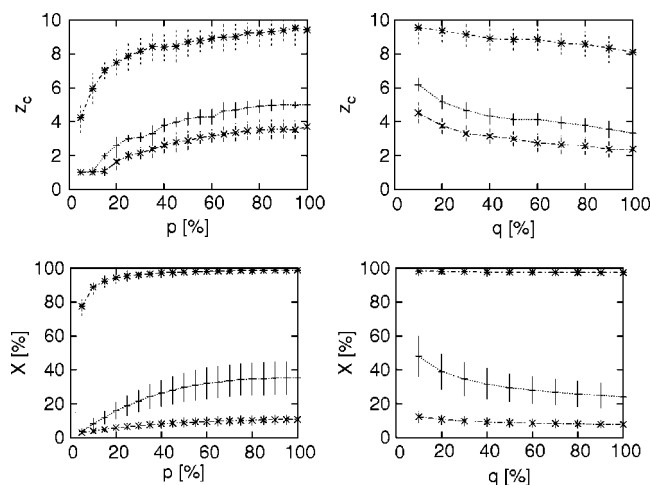
FIG. 3. Optimal range of a control neighborhood, z_c, is shown here as a function of the probability of infection, p, and the probability for a transition to the detected state, q, for networks without shortcuts. In addition, the severity index, X, is shown as a function of p and q, corresponding to a minimal severity index under the optimal local strategy. $\times$, SW1D; $+$, SW2D; $*$, SF. Other parameters are $q=0.5$ (left panel), $p=0.5$ (right panel), $r=0.01$, $v=0.1$, $C=5$, $C_I=1$. The results for the probability to treat, v, are similar to the results for q and are not shown here. Mean value of z_c and the standard deviation were estimated by averaging over 50 values of z_c for each group.

IV. DISCUSSION

A successful disease control strategy involves a combination of local and global, reactive, and preventive actions [2]. However, models of disease spread used to design appropriate control often use mean-field approximations [2] and assume perfect knowledge of both the status of each individual (pre- and postsymptomatic individuals) [2] and the structure of links [12,13]. Obtaining this information can be very expensive and time-consuming, particularly for networks with nonlocal links. It might be necessary to know not only the geographical location of new cases (so that they and their immediate neighbors can be treated) but also all possible connections that can span the whole population. For authorities faced with a large-scale epidemic, the collection of such data might be difficult. Our results show that in some cases it is possible to control epidemics by actions that are purely local and therefore do not require extensive contact tracking, by extending the diameter of the control ring. However, the costs of such treatment can build up very quickly as larger
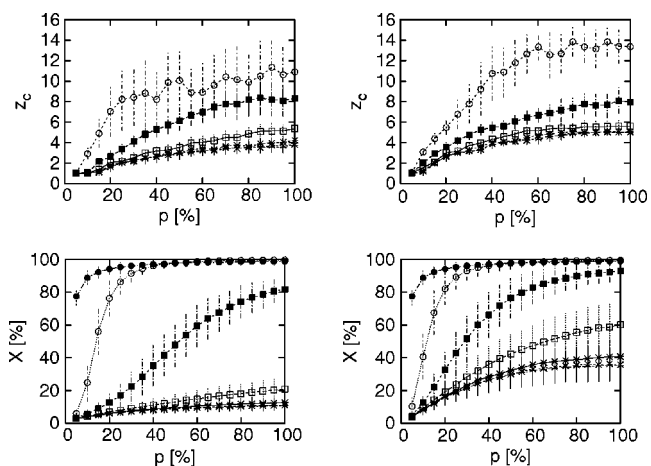
FIG. 4. Optimal range of the control neighborhood, z_c, and the severity index, X, are shown as a function of the probability to infect p for the SW1D (left column) and SW2D (right column) networks with shortcuts. Other parameters are $q=0.5$, $r=0.01$, $v=0.1$. Different symbols denote an increasing number of shortcuts: $+$, 0; $\times$, 3 (0.06% of total number of links in the control network); $*$, 15 (0.3%); $\square$, 63 (1.3%); $\blacksquare$, 255 (5.1%); and $\circ$, 1023 (20.5%) shortcuts. $\bullet$, in the severity plots, corresponds to the SF network. Details of simulations as in Fig. 3.

rings are needed with the increasing proportion of shortcuts (Fig. 4). An alternative strategy would be to increase our knowledge of network topology, primarily targeting highly connected individuals [13].

There is a clear distinction between the case when the local control works and the disease can be stopped early, and when it does not work and we need to treat most of the population. This is reflected in an existence of the optimal control radius, z_c. Making the ring of control even a fraction larger might lead to a dramatic increase in the efficiency of the treatment (Fig. 2). The existence of a critical ring size might suggest that the epidemic spread with local control is equivalent to percolation. The equivalence has been shown for systems with a global control [14], but for a local control this requires further studies.

[1] C. Dye and N. Gay, Science **300**, 1884 (2003).
[2] M. J. Keeling, M. E. J. Woolhouse, D. J. Shaw, L. Matthews, M. Chase-Topping, D. T. Haydon, S. J. Cornell, J. Kappey, J. Wilesmith, and B. T. Grenfell, Science **294**, 813 (2001); L. Matthews, D. T. Haydon, D. J. Shaw, M. E. Chase-Topping, M. J. Keeling, and M. E. J. Woolhouse, Proc. R. Soc. London,

Ser. B **270**, 1659 (2003).
[3] J. Swinton and C. A. Gilligan, Philos. Trans. R. Soc. London, Ser. B **351**, 605 (1996).
[4] T. R. Gottwald, G. Hughes, J. H. Graham, X. Sun, and T. Riley, Phytopathology **91**, 30 (2001).
[5] D. M. Rizzo, M. Garbelotto, J. M. Davidson, G. W. Slaughter,

and S. T. Koike, Plant Dis. **86**, 205 (2002).

[6] A. J. Stacey, J. E. Truscott, M. J. C. Asher, and C. A. Gilligan, Phytopathology **94**, 209 (2004).

[7] J. Swinton, J. Harwood, B. T. Grenfell, and C. A. Gilligan, J. Anim. Ecol. **67**, 54 (1998).

[8] D. J. Watts and S. H. Strogatz, Nature (London) **393**, 440 (1998); D. J. Watts, *Small Worlds* (Princeton University Press, Princeton, NJ, 1999); A. Kleczkowski and B. T. Grenfell, Physica A **274**, 355 (1999); C. Moore and M. E. J. Newman, Phys. Rev. E **62**, 7059 (2000).

[9] R. Albert and A. L. Barabási, Rev. Mod. Phys. **74**, 47 (2002).

[10] C. A. Gilligan, in *Battling Resistance to Antibiotics and Pesticides: An Economic Approach*, edited by R. Laxminarayan

(Resources for the Future, Washington, 2003), pp. 221–243.

[11] T. Hanslik, P. Y. Boelle, M. Schwarzinger, F. Carrat, K. A. Freedberg, A. J. Valleron, and A. Flahault, Vaccine **21**, 3614 (2003); H. C. Tuckwell, T. Hanslik, A. J. Valleron, and A. Flahault, Epidemiol. Infect. **130**, 419 (2003).

[12] R. M. Anderson and R. M. May, *Infectious Diseases of Humans: Dynamics and Control* (Oxford University Press, Oxford, 1991).

[13] R. Pastor-Satorras and A. Vespignani, Phys. Rev. E **65**, 036104 (2002); Z. Dezső and A. L. Barabási, *ibid.* **65**, 055103 (2002).

[14] J. Wallinga, Oikos **74**, 377 (1995); H. M. Bröker and P. Grassberger, Physica A **267**, 453 (1999).