UNIVERSITY OF STIRLING

Searching for the optimal control strategy of epidemics spreading on different types of networks

Katarzyna Oleś

Doctor of Philosophy

Supervisor: **Dr Adam Kleczkowski**

Stirling, 2014

Acknowledgements

During the PhD studies and preparation of this work I have received much help and support from many people. I would like to thank my PhD advisors, Prof. Ewa Gudowska - Nowak and Dr Adam Kleczkowski for many enlightening discussions and support during the whole period of my PhD studies.

I acknowledge the hospitality of Department of Computing Sciences and Mathematics at the University of Stirling during four semesters of my studies.

I kindly appreciate the hospitality of the Prof. James P. Gleeson (MACSI, Department of Mathematics and Statistics, University of Limerick, Ireland) where I spent two weeks and the financial support of the "Exploring the Physics of Small Devices (EPSD)" of European Science Foundation (ESF).

Contents

List of publications

1. A. Kleczkowski, K. Oles, E. Gudowska-Nowak, C.A. Gilligan, Searching for the most cost-effective control strategy for controlling epidemics spreading on regular and small-world networks. Journal of the Royal Society Interface, January 7 (2012) 9:158-169;

2. K. Oleś, E. Gudowska-Nowak, A. Kleczkowski, Understanding disease control: influence of epidemiological and economical factors. PLoS ONE (2012) 7(5): e36026. doi:10.1371/journal.pone.0036026;

3. K. Oleś, E. Gudowska-Nowak, A. Kleczkowski, Efficient control of epidemics spreading on networks: Balance between treatment and recovery. PLoS ONE (2013) 8(6): e63813. doi:10.1371/journal.pone.0063813;

4. K. Oleś, E. Gudowska-Nowak, A. Kleczkowski, Cost-benefit analysis of epidemics spreading on clustered random networks. Acta Physica Polonica B (2014) 45(1): 103-120

1 Introduction

This thesis arises from an agreement between University of Stirling with Jagiellonian University to offer me joint PhD studies in the field of Mathematics and its Applications to Physics of Complex Systems. The collaboration between the Department of Computing Science and Mathematics with the Marian Smoluchowski Institute of Physics and Mark Kac Center for Complex Systems Research has been extended in the form of a programme of International PhD studies in Physics of Complex Systems (MPD) supported by a grant from the Polish Foundation for Science.

The primary motivation of my research were the serious epidemic outbreaks of human [1], animal [2] and plant [3, 4] diseases and the strong need to design an effective way of controlling them.

The underlying assumption of such strategies is the wide availability and low economic or social cost of treatment as the form of preventive vaccination or therapy [5]. However, these assumptions are not true in many cases, particularly for large outbreaks like cholera [6], AIDS [5], severe acute respiratory syndrome (SARS) [1] or foot-and-mouth disease (FMD) [2]. Therefore, there is a need for a 'marriage of economics and epidemiology' [5] in designing effective strategies for control of disease [7]. Key to this approach is the realization that an optimal policy does not necessarily result in curing any individual in the population regardless of costs. Instead, it might be acceptable to tolerate some lower level of disease persistence in situation when the costs of eradication are prohibitively high [8].

Epidemiological modelling plays an important role because it explains a range of crucial issues:

- Estimation of the scale of the epidemic;
- Prediction of how far the disease could spread;

- Design of the effective ways of controlling the outbreaks.

In successful modelling all these tasks need to be achieved, even though in many cases it is impossible to observe the whole process and measure the relevant parameters [9]. Despite these uncertainties the mathematical modelling can be used to design effective control measures. Control scenarios can be designed to lead to the lowest overall cost of the epidemic outbreak [10, 8, 11] and a number of studies have used network models to address this issue [9, 12, 13, 14].

My research has concentrated on mathematical modelling of the spread of the epidemic on different types of networks that represent a map of contacts between individuals in a population through which the disease can be transmitted. The aim of my simulations was to search for the most optimal control strategy to stop the epidemic outbreak when economic factors were considered. The analysis of both epidemiological and economic parameters allowed me to find conditions under which different control scenarios are the most cost-effective. Moreover, my research can be used to predict optimal control strategy even with incomplete knowledge about emerging disease, based on economic analysis only.

2 Literature review - the scope of the studies

The effectiveness of such factors like improved sanitation, antibiotics, and vaccination programs made us believe that infectious diseases could be eliminated from the environment [15]. However, diseases not only have continued to be major issue in developing countries but also infectious disease pathogens adapt and evolve and new infectious diseases have emerged [16]. The human immunodeficiency virus (HIV), which is the etiological agent for acquired immunodeficiency syndrome (AIDS), was discovered in 1981 and has become an important sexually transmitted disease throughout the world [15]. Diseases such as plague, cholera, and hemorrhagic fevers (Bolivian, Ebola, Lassa) cause occasional outbreaks.

Mathematical models have become very important and indispensable tools in analysing the spread and designing the control of infectious diseases. In order to formulate mathematical models, assumptions, variables, and parameters must be predefined and clarified. As a result, epidemiological modelling provides much crucial information in designing epidemic spread, like thresholds, basic reproduction numbers or contact numbers. Moreover, with the help of computer simulations it is a useful tool for building and examining theories, answering specific questions, determining sensitivities to changes in parameter values, and estimating key parameters from data [15].

Most mathematical models assume that all organisms in the population stay in contact with each other and the probability of infecting any individuals is equal. In the real world, each individual has contacts only with some fraction of the total population and the number of interactions between organisms can vary from one person to another [17]. That is the reason that we incorporate network topology mimicking the pattern of contact in the system. And this is a crucial element in epidemiological modelling.

The study of networks has been rooted in several fields ranging from social sciences [18] to mathematical graph theory [19, 20] and complex systems [21, 22]. Social sciences mostly pay attention to the reason behind the connections rather than to the properties of the network structure itself. Investigations of complex networks have been used to describe evolution of ideas and innovations in societies [18], and observed social dynamics can be understood through analysis of the social networks that underlie them [20]. Research has been concentrated mainly on the nature of connections, particularly properties such as a symmetry (whether a relationship between A and B implies a relationship between B and A) and transitivity (whether the friend of a friend is a friend) [23, 24]. Additionally, many simple as well as complex measures of the importance of individuals can be derived: number of connections of each individual or the number of paths between other actors in which an individual features [20, 23].

In contrast, graph theory has provided a variety of quantitative measures. 'Adjacency matrix' describes the connections within a population and with its help other important characteristics such as the average number of contacts per individual, average path length (the distance between two randomly chosen nodes), clustering (group of individuals fully connected), and percolation threshold (critical fraction of nodes that must be connected in order to create a continuous path of nearest neighbours from one side to another) have been applied in epidemiology [20, 25].

Several forms of networks have been studied for modelling disease transmission: random [26], regular [27], small-world [28, 29] and scale-free [30, 31]. These network models can be defined in terms of how individuals are distributed in space and how connections are formed.

Alternative approaches like moment closure models, household models or meta-population models can be used to addressed epidemiological problems.

Moment closure approximations are used for non-linear stochastic population in order to provide its analytic approximation, insight into model behaviour and to validate results from simulations [32, 33]. However this methods is inadequate for the cases with skewed population distribution [32].

Metapopulation models assume that the distribution of the species can be described as a system of local populations [34]. Subpopulation may turnover as a result of extinction and then recolonization. Infection is equivalent to colonization, and death or recovery of the host is equivalent to local extinction. The balance between decolonisation and extinction affects the patch occupancy.

In order to model a realistic system where population can mix heterogeneously the household model can be adapted. In such models population is partitioned into subgroups - households [35] which may differ in rate of e.g. infectiousness, detectability or recovery. This methods allows also to incorporate vaccination and isolation strategies, based upon the appearance of diagnosed individuals in households.

The choice of network based model or mean field approximation can be made upon consideration of the nature of the disease and how it spreads at the level of nodes, the mechanism of contact and the infectious period [36]. For the random graph and mean field model there is an equivalency in results. In contrast, different results can be obtain for sparsely connected networks.

Network analysis plays an important role in linking qualitative and quantitative measures of epidemic progress. In addition, many ways of control, such as e.g. contact tracing [37, 38] or ring vaccination [39, 40], can only be accurately modelled by means of network theory.

3 General overview of modelling and results

3.1 Epidemiological model

Most early mathematical studies of disease propagation make the assumption that populations are "fully mixed", meaning that all infective individuals are equally likely to spread the disease [41, 42, 15]. In the limit of a large population size this assumption allows us to write down nonlinear differential equations denoting the evolution of e.g. numbers of infective individuals in time. Resulting solutions provide means to understand such measures as typical sizes of outbreaks, and make it possible to predict under what circumstances the epidemics occurred [17].

The epidemiological model that has been used in this work is an extended SIR (Susceptible-Infected-Removed) model to account for pre-symptomatic and symptomatic stages [12]. Initially all individuals are susceptible (**S**), except for fixed small number of infected pre-symptomatic (**I**) individuals (0.01%, 0.1% or 1% proportion of the total number), located randomly throughout the population.

Each individual is in contact with a fixed number of neighbours and the disease can be transmitted from/to each of them. Details of the spatial arrangement and size of the neighbourhood are given below. With probability f per single contact with either an infected individual (**I**) or the detected individual (**D**), the disease is passed to a susceptible individual (**S**) that becomes an infectious but pre-symptomatic individual (**I**). Subsequently the infected individual displays symptoms and the transition to a symptomatic state (**D**) occurs with probability q.

A symptomatic individual is assumed to be still infectious, but can spontaneously become removed (**R**) with probability r and cease to pass on infection. Alternatively, it can also trigger a control event, with probability v. Thus, at

each time step, the detected individual stays in the same class with probability $(1-r)(1-v)$. This mechanism accounts for possible delays and imperfections in detection of disease symptoms – any individual can show symptoms but not be treated until after a number of steps.

The treatment event is a combination of two processes. Firstly, a detected individual is treated and moves to the treated class (**V**). Secondly, all individuals except removed (i.e. **S**, **I** or **D**) in the control neighbourhood (see below) are also treated. This process enables the health control authorities to capture individuals in the class **I** that do not show symptoms and all detected individuals (**D**) that are still waiting for treatment. In addition, it creates a zone around the focus of infection in which there are no susceptible individuals. Neither **V** nor **R** individuals can become infected again. The population has a constant number of individuals N, so that $N = S + I + D + V + R$.

3.2 Network models

The subject of my research have been networks of various types : regular, small-world, and random with different level of clustering. In regular networks, I assume that individuals are located at nodes of a square lattice that represent geographical distribution of hosts, for example. On this lattice, I define a local neighbourhood of order z as a von Neumann neighbourhood in which I include z shells and $\Phi(z) = 2z(z+1)$ individuals, excluding the central one. Accordingly, $z = 0$ corresponds to a single individual, which means that this individual is not in contact with anyone, $z = 1$ corresponds to 4 nearest neighbours while $z = \infty$ corresponds to the whole population in the limit of infinite size of the system.

For the small world model a fixed number of long range links has been added to the regular network described above. Those links span the whole

population, but otherwise behave like local links.

In random networks a constant number of contacts for each node have been chosen randomly from the whole population. Along links either the pathogen is transmitted or the control process is triggered.

Although these kinds of networks can successively be used for modelling emerging diseases and their control, they describe ideal situations. For real-life contact structures, the more adequate description might be provided by random clustered networks. These networks exhibit a certain proportion of fully connected subgraphs in the form of cliques. Each vertex (representing an individual) can be a part of a c-clique, i.e. a group of c individuals that are fully connected, or can be a single node (i.e. a member of a 1-clique). Nodes which are members of a c-cliques have $c-1$ edges linking them with the neighbours within the same clique. For a random node with k connections to other vertices in the network there are additional $k-c+1$ edges outside the clique. In my work, I restrict the attention to random graphs in which all nodes have the same degree k. Random clustered networks are described by the joint probability $\gamma(k,c)$ that a randomly chosen vertex has degree k and is a member of a c-qlique [43]. In turn, the local clustering coefficient for a node is defined as a fraction of pairs of neighbours of this node which are also neighbours of each other. The degree-dependent clustering (or clustering spectrum c_k) is the average of the local clustering coefficient over the class of all nodes of degree k.

A separate network structure is used to model spread of the pathogen, which can only be passed to individuals that are in infected neighbourhood, z_{inf}. Another is created for the control process in neighbourhood of order z in order to find the optimal size of treatment (or culling) z_c, which, depending on economic factors, may differ from infected neighbourhood. Infection can be passed to all neighbours within the range described by $z = z_{inf}$. As

the spread of disease involves asymptomatic individuals, there could be some infectious organisms beyond the immediate neighbourhood of a detected individual. Thus the control process typically needs to be applied to a larger neighbourhood and I denote by z the range of the control neighbourhood.

3.3 Economic model

The effectiveness of a control strategy is found by considering the severity of the disease outbreak and its financial implications. Two types of costs can be distinguished during the epidemics. Firstly, the costs associated with removed individuals (e.g. hospitalisation, absence from work, loss of production) can be estimated by the total number of individuals that have caught the infection and have gone through the disease but have never been treated, i.e. $R(t = \infty)$. Costs spent on preventive treatment (vaccination, culling) are calculated by considering the final number of individuals that have been treated, i.e. $V(t = \infty)$. The cost-effectiveness of different control strategies can be quantified by the total cost obtained by

$$X = c_1 R(t = \infty) + c_2 V(t = \infty) \tag{1}$$

- c_1 - a unit cost associated with each *removed* individual (**R**),
- c_2 - a unit cost associated with each *treated* individual (**V**).
- $R(t = \infty)$ and $V(t = \infty)$ are counted at the end of a single simulation run.

Without loss of generality the assumption that $c_1 = 1$ and $c_2 = c$ is true, which leads to an conclusion that the relative cost of treatment, c, is the main control parameter. (The parameter describing the cost of treatment is called

as "c" in the first three publications, whereas in the fourth paper it is denoted by "a").

The effective strategy is equivalent to the minimal value of the total cost, X. In simulation, the minimisation of the X has been achieved by sweeping through different values of control neighbourhood size, z, while keeping other parameters constant. Once z is set, the disease evolves on networks and at the end of epidemic outbreaks ($t = \infty$) the value of X is computed in the stationary state. This operation is repeated 100 times in order to yield the average values of z denoted by z_c and X described by X_c along with their standard deviations. z_c corresponds to the minimum of X, so that

$$\min_{-1 \leq z \leq z_{max}} \quad X(z, t = \infty) = X_c(z_c, t = \infty) \quad . \tag{2}$$

4 Results

4.1 Structure of my Thesis

I am presenting four interconnected papers. Paper 1 formulates the basic model and explores dependence of optimal control size, z_c on both probability of disease spread, f and treatment cost, c. It also introduces regular, small-world and random networks as well as a mean-field model. Paper 2 is mainly devoted to the relationship between optimal control range, z_c and epidemiological factors such as probability of disease spread, f, probability of detection, q, probability of spontaneous recovery, v and the size of infected neighbourhood, z_{inf}. Dependence on recovery/removal rate, r and comparison of two similar epidemiological models are the subjects of the Paper 3. Finally, Paper 4 introduces clustered networks and analyses the influences of level of clustering and node degree on optimal control size, z_c.

4.2 Principal results

The main goal of my studies has been to search for the optimal control strategy of controlling epidemics when taking into account both economical and social costs of the disease. Three control scenarios emerge with treating the whole population (global strategy, GS), treating a small number of individuals in a well-defined neighbourhood of a detected case (local strategy, LS) and allowing the disease to spread unchecked (null strategy, NS). The choice of the optimal strategy is governed mainly by a relative cost of palliative and preventive treatments. Although the properties of the pathogen might not be known in advance for emerging diseases, the prediction of the optimal strategy can be made based on economic analysis only.

The details of the local strategy and in particular the size of the optimal

treatment neighbourhood weakly depends on disease infectivity but strongly depends on other epidemiological factors (probability of detection, spontaneous recovery). The required extent of prevention is proportional to the size of the infection neighbourhood, but this relationship depends on time till detection and time till treatment in a non-nonlinear (power) law.

The spontaneous recovery also affects the choice of the control strategy. I have extended my results to two contrasting and yet complementary models, in which individuals that have been through the disease can either be treated or not. Whether the removed individuals (i.e., those who have been through the disease but then spontaneously recover or die) are part of the treatment plan depends on the type of the disease agent. The key factor in choosing the right model is whether it is possible - and desirable - to distinguish such individuals from those who are susceptible. If the removed class is identified with dead individuals, the distinction is very clear. However, if the removal means recovery and immunity, it might not be possible to identify those who are immune. The models are similar in their epidemiological part, but differ in how the removed/recovered individuals are treated. The differences in models affect choice of the strategy only for very cheap treatment and slow spreading disease. However for the combinations of parameters that are important from the epidemiological perspective (high infectiousness and expensive treatment) the models give similar results. Moreover, even where the choice of the strategy is different, the total cost spent on controlling the epidemic is very similar for both models.

Although regular and small-world networks capture some aspects of the structure of real networks of contacts between people, animals or plants, they do not include the effect of clustering noted in many real-life applications [44, 45]. The use of random clustered networks in epidemiological modelling takes an important step towards application of the modelling framework to

realistic systems. Network topology and in particular clustering also affects the applicability of the control strategy.

4.3 Key results for paper 1

A. Kleczkowski, K. Oleś, E. Gudowska-Nowak, C.A. Gilligan, Searching for the most cost-effective control strategy for controlling epidemics spreading on regular and small-world networks. Journal of the Royal Society of Interface, January 7 (2012) 9:158-169;

- Taking into account relative costs of treatment and illness, three main control strategies emerge: treating a large number of individuals (global strategy, GS), treating a proportion of individuals in a well-defined neighbourhood of a detected case (local strategy, LS), refrain from treatment (null strategy, NS).

- Destruction of local interactions, either by addition of long-range (small-world) links or by inclusion of many initial foci, expands the range of costs for which the null strategy (NS) is most cost-effective. The global strategy (GS) emerges for the case when the cost of prevention is much lower than the cost of treatment. Then there is a substantial non-local component in the disease spread.

- In the mean-field case only two optimal solutions are possible: to treat the whole population if the cost of the vaccine is low or to refrain from control if cost is expensive.

- The basic reproduction ratio, R_0 (the expected number of secondary cases produced by a typical infected individual during its entire infectious period), does not depend on the rate of responsive treatment and the disease always invades.

- The properties of the pathogen of emerging diseases may not be known in advance. The choice of the strategy (GS, LS or NS) can be made

based on economic analysis only. The main influence on the option of the control scanario comes from value of relative cost of treatment to the cost of infection, c.

4.4 Key results for paper 2

K. Oleś, E. Gudowska-Nowak, A. Kleczkowski, Understanding disease control: influence of epidemiological and economical factors. PLoS ONE (2012) 7(5): e36026. doi:10.1371/journal.pone.0036026;

- The local strategy (LS, treating susceptible or infectious individuals in well defined neighbourhood of certain size) matches the scale of epidemic with the scale of control.
- The details of the local strategy and in particular the size of the optimal treatment neighbourhood weakly depends on disease infectivity but is strongly influenced by the other epidemiological factors, like probability of detection or spontaneous recovery.
- The required extent of prevention is proportional to the size of the infection neighbourhood.
- The control neighbourhood size depends on time till detection and time till treatment, however this relationship is nonlinear but follows power law.
- The optimal size of control neighbourhood is highly sensitive to the relative cost, particularly for inefficient detection and control application.

4.5 Key results for paper 3

K. Oleś, E. Gudowska-Nowak, A. Kleczkowski, Efficient control of epidemics

spreading on networks: Balance between treatment and recovery. PLoS ONE (2013) 8(6): e63813. doi:10.1371/journal.pone.0063813;

- Two epidemiological models have been compared. In many real life situations it is difficult to examine the immunisation to the particular pathogen. The differences in the models correspond to the their part denoting gaining immunisation.
- The differences in models affects the choice of the strategy in the situation when treatment is very cheap and when disease spreads slowly.
- From the epidemiological point of view, in the crucial scenario (high infectiousness and expensive treatment) both models predict very similar results.
- Even where the choice of the strategy differs, the total cost spent on controlling the epidemic is at the same level for both models.

4.6 Key results for paper 4

K. Oleś, E. Gudowska-Nowak, A. Kleczkowski, Cost-benefit analysis of epidemics spreading on clustered random networks. Acta Physica Polonica B (2014) 45(1): 103-120

- In order to adapt mathematical modelling to real life application the model of random clustered networks has been used.
- Network topology and in particular clustering also affects the applicability of the control strategy.
- The networks characteristics such as average path length or local clustering coefficient appears to play the most important role. The larger average path length, the larger the interval for which LS is optimal.

- The proportion of individuals in cliques affects the local coefficient of clustering. With higher density of cliques in networks and with greater value of clustering coefficient, the range of the treatment costs, for which control scenario is optimal, increases.

5 Discussion and possible extensions

Mathematical epidemiology has now evolved into a separate area of population dynamics that is parallel to mathematical ecology. Epidemiology models are now used to combine complex data from various sources in order to study equally complex outcomes. Mathematical models are used in comparing, planning and optimising the whole range of processes: detection, prevention, therapy, control scenarios, making general forecasts, and estimating the uncertainty in predictions [46, 15]. Moreover, the incomplete knowledge of the newly emerged disease or the way it is transmitted through the system does not prevent modellers to propose successful control options even at the beginning of the epidemics. Incorporating economic factors into designing control strategies results in a very powerful tool for authorities that need to decide whether and how resources need to be allocated in order to stop the epidemics as quickly as possible and at a manageable cost.

Even though the mathematical models estimate results and predict scenarios we should be aware of their limitation and assumption on which they based.

- Network models assume the homogeneity of nodes, however in real life every individuals are unique and distinguishable.
- Epidemiological factors are fixed and do not change during epidemic.
- The efficacy of vaccination is assumed to be 100% efficient.
- Model allows to vaccinate infected (I and D) individuals.
- No fixed costs are considered only variable costs are included.

My research can be extended in several ways, and the most interesting as well as challenging appear to be:

- SIRS model: a model in which after some period of immunity to the disease, individuals become susceptible again and could catch a disease few times. The best examples are influenza and sexually-transmitted diseases.

- Dynamical networks: networks with connections that could change in time, e. g. describing the situation when the behaviour of a population can markedly change as a consequence of an outbreak of infection, which needs to be considered when designing interventions.

- Social networks: e.g. scale-free networks, more realistic network type, especially for modelling human diseases.

- Other economical circumstances, e.g. limited budget that could be spent on epidemic outbreaks, and the resources need to be allocated wisely presents situation to which the health authorities need to face up.

- Time dependent control: control, which size could change in time, may properly allocate resources and better adopt efforts to the actual scale of epidemics.

- Spread of more than one pathogen in the population: this extension describes very serious real-life scenario that could results in higher vulnerability for one diseases when individuals are affected by the other pathogen. Alternatively, an individual that catches one type of disease and eventually gains immunity from it, might also acquire immunity to the other one, even though has not been treated.

6 Author Contributions

Results presented in my thesis have been achieved by numerical methods with programmes written mainly in C, and by use of Matlab software.

I have developed myself code in C using Monte Carlo Methods to execute disease evolution of SIDRV model and control process in a neighbourhood of any size.

I used lattices with periodic boundary conditions to represent different network types such as regular, small-world and random by different way of choosing links between nodes. In regular networks, edges have been placed between the nearest neighbours that described geographical distribution of hosts. Small-world networks have been created on the basis of regular ones, by adding number of randomly chosen links that could span the whole lattices. In random networks all nodes have fixed number of connections that have been placed uniformly random on the lattice. In order to properly adjust control events and disease spread I have used two separate matrices corresponding to infected and control neighbourhood.

The large size of the system (represented by lattices of 200 by 200 nodes), puts special demands on memory. I run my programmes on the Jagiellonian University computer grid called "Shiva" cluster, as a single thread programmes. Afterwards, data mimicking evolution of the epidemic process were analysed. For that purpose I have designed and written a code in C which uses the economic model and takes into account the requirement of optimising control strategy for the costs ranging from 10^{-4} to 10^{3}.

During my collaborative visit in the group of Professor James Gleeson at the University of Limerick in Ireland, I was given a code in Matlab that generates random networks with different level of clustering. Working with Sergiej Melnik, I adopted the code in Matlab to generate random clustered

networks with the whole range of size of neighbourhood needed in control process and pathogen propagation.

Source code of all programs (except the Matlab code) is available on request.

1. *Searching for the most cost-effective control strategy for controlling epidemics priding an regular and small-world networks.*

    - Conceived and designed the experiments: AK KO.
    - Performed the experiments: KO.
    - Analysed the data: KO.
    - Contributed analysis tools: KO AK EGN.
    - Wrote the paper: KO AK CAG EGN.

2. *Understanding disease control: influence of epidemiological and economical factors.*

    - Conceived and designed the experiments: AK KO.
    - Performed the experiments: KO.
    - Analysed the data: KO EGN AK.
    - Contributed analysis tools: KO EGN AK.
    - Wrote the paper: KO EGN AK.

3. *Efficient control of epidemics spreading on networks: Balance between treatment and recovery.*

    - Conceived and designed the experiments: KO AK.
    - Performed the experiments: KO.
    - Analysed the data: KO EGN AK.

- Contributed analysis tools: KO EGN AK.
- Wrote the paper: KO EGN AK.

4. *Cost-benefit analysis of epidemics spreading on clustered random networks.*
   - Conceived and designed the experiments: KO JG EGN AK.
   - Performed the experiments: KO.
   - Analysed the data: KO EGN AK.
   - Contributed reagents/materials/analysis tools: KO EGN AK.
   - Wrote the paper: KO AK EGN.

References

[1] Dye, C ; Gay, N: Modeling the SARS epidemic. In: *Science* 300 (2003), Juni, Nr. 5627, S. 1884–1885

[2] Haydon, Daniel T. ; Kao, Rowland R. ; Kitching, R P.: The UK foot-and-mouth disease outbreakï£¡the aftermath. In: *Nature Reviews Microbiology* 2 (2004), Nr. 8, S. 675–681

[3] Gottwald, Tim R. ; Hughes, Gareth ; Graham, James H. ; Sun, Xiaoan ; Riley, Tim: The citrus canker epidemic in Florida: the scientific basis of regulatory eradication policy for an invasive species. In: *Phytopathology* 91 (2001), Nr. 1, S. 30–34

[4] Behlau, F ; Belasque, J ; Bergamin Filho, A ; Graham, JH ; Leite, RP ; Gottwald, TR: Copper sprays and windbreaks for control of citrus canker on young orange trees in southern Brazil. In: *Crop Protection* 27 (2008), Nr. 3, S. 807–813

[5] Gersovitz, Mark ; Hammer, Jeffrey S.: Infectious diseases, public policy, and the marriage of economics and epidemiology. In: *The World Bank Research Observer* 18 (2003), Nr. 2, S. 129–157

[6] Jeuland, Marc ; Lucas, Marcelino ; Clemens, John ; Whittington, Dale: A cost–benefit analysis of cholera vaccination programs in Beira, Mozambique. In: *The World Bank Economic Review* 23 (2009), Nr. 2, S. 235–267

[7] Woolhouse, Mark ; Chase-Topping, Margo ; Haydon, Daniel ; Friar, John ; Matthews, Louise ; Hughes, Gareth ; Shaw, Darren ; Wilesmith, John ; Donaldson, Alex ; Cornell, Stephen u. a.: Epi-

demiology: Foot-and-mouth disease under control in the UK. In: *Nature* 411 (2001), Nr. 6835, S. 258–259

[8] Forster, Graeme A. ; Gilligan, Christopher A.: Optimizing the control of disease infestations at the landscape scale. In: *Proceedings of the National Academy of Sciences of the United States of America* 104 (2007), März, Nr. 12, S. 4984–4989

[9] Cohen, Reuven ; Havlin, Shlomo ; Ben-Avraham, Daniel: Efficient immunization strategies for computer networks and populations. In: *PRL* 91 (2003), Nr. 24, S. 247901

[10] Kao, Rowland R.: The role of mathematical modelling in the control of the 2001 FMD epidemic in the UK. In: *Trends In Microbiology* 10 (2002), Nr. 6, S. 279–286

[11] Hollingsworth, T D. ; Klinkenberg, Don ; Heesterbeek, Hans ; Anderson, Roy M.: Mitigation strategies for pandemic influenza A: balancing conflicting policy objectives. In: *PLoS Computational Biology* 7 (2011), Nr. 2, S. e1001076

[12] Kleczkowski, Adam ; Oleś, Katarzyna ; Gudowska-Nowak, Ewa ; Gilligan, Christopher A.: Searching for the most cost-effective strategy for controlling epidemics spreading on regular and small-world networks. In: *Journal of The Royal Society Interface* 9 (2012), Nr. 66, S. 158–169

[13] Oleś, Katarzyna ; Gudowska-Nowak, Ewa ; Kleczkowski, Adam: Understanding disease control: influence of epidemiological and economic factors. In: *PLoS ONE* 7 (2012), Nr. 5, S. e36026

[14] Gómez-Gardenes, Jesús ; Echenique, Pablo ; Moreno, Yamir: Immunization of real complex communication networks. In: *The European*

Physical Journal B-Condensed Matter and Complex Systems 49 (2006), Nr. 2, S. 259–264

[15] Hethcote, Herbert W.: The mathematics of infectious diseases. In: *SIAM review* 42 (2000), Nr. 4, S. 599–653

[16] Jacquez, John A. ; Simon, Carl P. ; Koopman, James ; Sattenspiel, Lisa ; Perry, Timothy: Modeling and analyzing {HIV} transmission: the effect of contact patterns. In: *Mathematical Biosciences* 92 (1988), Nr. 2, S. 119–199

[17] Newman, Mark E.: Spread of epidemic disease on networks. In: *Physical Review E* 66 (2002), Nr. 1, S. 016128

[18] Leinhardt, Samuel: *Social networks: A developing paradigm*. Academic Press New York, 1977

[19] Bollobás, Béla: *Modern graph theory*. Bd. 184. Springer, 1998

[20] Keeling, Matt J. ; Eames, Ken T.: Networks and epidemic models. In: *Journal of the Royal Society Interface* 2 (2005), Nr. 4, S. 295–307

[21] Bar-Yam, Yaneer: Dynamics of complex systems. (2003)

[22] Haken: *Information and self-organization: a macroscopic approach to complex systems*. Springer, 2006

[23] Wasserman, Stanley: *Social network analysis: Methods and applications*. Bd. 8. Cambridge university press, 1994

[24] Karlberg, Martin: Testing transitivity in graphs. In: *Social Networks* 19 (1997), Nr. 4, S. 325–343

[25] Dorogovtsev, SN ; Mendes, JFF: *Evolution of networks: From biological nets to the Internet and WWW*. Oxford : Clarendon Press, 2003

[26] Erdős, Paul ; Rényi, Alfréd: On random graphs. In: *Publicationes Mathematicae Debrecen* 6 (1959), S. 290–297

[27] Newman, Mark: *Networks: an introduction*. Oxford University Press, 2009

[28] Watts, Duncan J. ; Strogatz, Steven H.: Collective dynamics of small-world networks. In: *Nature* 393 (1998), Nr. 6684, S. 440–442

[29] Moore, Cristopher ; Newman, Mark E.: Epidemics and percolation in small-world networks. In: *Physical Review E* 61 (2000), Nr. 5, S. 5678

[30] Barabási, Albert-László ; Albert, Réka ; Jeong, Hawoong: Mean-field theory for scale-free random networks. In: *Physica A: Statistical Mechanics and its Applications* 272 (1999), Nr. 1, S. 173–187

[31] Barabási, Albert-László ; Albert, Réka ; Jeong, Hawoong: Scale-free characteristics of random networks: the topology of the world-wide web. In: *Physica A: Statistical Mechanics and its Applications* 281 (2000), Nr. 1, S. 69–77

[32] Krishnarajah, Isthrinayagy ; Cook, Alex ; Marion, Glenn ; Gibson, Gavin: Novel moment closure approximations in stochastic epidemics. In: *Bulletin of mathematical biology* 67 (2005), Nr. 4, S. 855–873

[33] Keeling, MJ ; Rand, DA ; Morris, AJ: Correlation models for childhood epidemics. In: *Proceedings of the Royal Society of London. Series B: Biological Sciences* 264 (1997), Nr. 1385, S. 1149–1156

[34] Grenfell, Bryan ; Harwood, John: (Meta) population dynamics of infectious diseases. In: *Trends in Ecology & Evolution* 12 (1997), Nr. 10, S. 395–399

[35] Ball, Frank G. ; Knock, Edward S. ; OÕNeill, Philip D.: Control of emerging infectious diseases using responsive imperfect vaccination and isolation. In: *Mathematical biosciences* 216 (2008), Nr. 1, S. 100–113

[36] Green, Darren M. ; Kiss, Istvan Z. ; Kao, Rowland R.: Parameterization of individual-based models: comparisons with deterministic mean-field models. In: *Journal of theoretical biology* 239 (2006), Nr. 3, S. 289–297

[37] Dybiec, Bartłomiej: Random strategies of contact tracking. In: *Physica A: Statistical Mechanics and its Applications* 387 (2008), Nr. 19, S. 4863–4870

[38] Ma, Junling ; Driessche, P van d. ; Willeboordse, Frederick H.: The importance of contact network topology for the success of vaccination strategies. In: *Journal of theoretical biology* 325 (2013), Mai, S. 12–21

[39] Keeling, MJ ; Woolhouse, MEJ ; May, RM ; Davies, G ; Grenfell, BT: Modelling vaccination strategies against foot-and-mouth disease. In: *Nature* 421 (2002), Nr. 6919, S. 136–142

[40] Tildesley, Michael J. ; Savill, Nicholas J. ; Shaw, Darren J. ; Deardon, Rob ; Brooks, Stephen P. ; Woolhouse, Mark E. ; Grenfell, Bryan T. ; Keeling, Matt J.: Optimal reactive vaccination strategies for a foot-and-mouth outbreak in the UK. In: *Nature* 440 (2006), Nr. 7080, S. 83–86

[41] Bailey, Norman T. u. a.: *The mathematical theory of infectious diseases and its applications*. Charles Griffin & Company Ltd, 5a Crendon Street, High Wycombe, Bucks HP13 6LE., 1975

[42] Anderson, Roy M. ; May, Robert M. ; Anderson, B: *Infectious diseases of humans: dynamics and control*. Bd. 28. Wiley Online Library, 1992

[43] Gleeson, James P. ; Melnik, Sergey ; Hackett, Adam: How clustering affects the bond percolation threshold in complex networks. In: *Physical Review E* 81 (2010), Jun, S. 066114

[44] Danon, Leon ; Read, Jonathan M. ; House, Thomas A. ; Vernon, Matthew C. ; Keeling, Matt J.: Social encounter networks: characterizing Great Britain. In: *Proceedings. Biological sciences / The Royal Society* 280 (2013), August, Nr. 1765, S. 20131037. – ISSN 1471–2954

[45] Volz, Erik: Random networks with tunable degree distribution and clustering. In: *Physical Review E* 70 (2004), November, Nr. 5, S. 056115. – ISSN 1539–3755

[46] Hethcote, Herbert W.: Three basic epidemiological models. In: *Applied Mathematical Ecology*. Springer, 1989, S. 119–144

Publications

Downloaded from rsif.royalsocietypublishing.org on June 9, 2011

JOURNAL OF THE ROYAL SOCIETY Interface

Searching for the most cost-effective strategy for controlling epidemics spreading on regular and small-world networks

Adam Kleczkowski, Katarzyna Oles, Ewa Gudowska-Nowak and Christopher A. Gilligan

J. R. Soc. Interface published online 8 June 2011
doi: 10.1098/rsif.2011.0216

References This article cites 49 articles, 11 of which can be accessed free
http://rsif.royalsocietypublishing.org/content/early/2011/06/03/rsif.2011.0216.full.html#ref-list-1

P<P Published online 8 June 2011 in advance of the print journal.

Subject collections Articles on similar topics can be found in the following collections

biomathematics (239 articles)
computational biology (222 articles)

Email alerting service Receive free email alerts when new articles cite this article - sign up in the box at the top right-hand corner of the article or click **here**

Advance online articles have been peer reviewed and accepted for publication but have not yet appeared in the paper journal (edited, typeset versions may be posted when available prior to final publication). Advance online articles are citable and establish publication priority; they are indexed by PubMed from initial publication. Citations to Advance online articles must include the digital object identifier (DOIs) and date of initial publication.

To subscribe to *J. R. Soc. Interface* go to: **http://rsif.royalsocietypublishing.org/subscriptions**

This journal is © 2011 The Royal Society

Downloaded from rsif.royalsocietypublishing.org on June 9, 2011

JOURNAL OF THE ROYAL SOCIETY Interface

FirstCite® e-publishing

J. R. Soc. Interface
doi:10.1098/rsif.2011.0216
Published online

Searching for the most cost-effective strategy for controlling epidemics spreading on regular and small-world networks

Adam Kleczkowski[1,*], Katarzyna Oleś[1,2], Ewa Gudowska-Nowak[2] and Christopher A. Gilligan[3]

[1]*Department of Computing Science and Mathematics, University of Stirling, Stirling FK9 4LA, UK*
[2]*Marian Smoluchowski Institute of Physics, Mark Kac Center for Complex Systems Research, Jagellonian University, ulica Reymonta 4, 30–059 Kraków, Poland*
[3]*Department of Plant Sciences, University of Cambridge, Cambridge CB2 3EA, UK*

We present a combined epidemiological and economic model for control of diseases spreading on local and small-world networks. The disease is characterized by a pre-symptomatic infectious stage that makes detection and control of cases more difficult. The effectiveness of local (ring-vaccination or culling) and global control strategies is analysed by comparing the net present values of the combined cost of preventive treatment and illness. The optimal strategy is then selected by minimizing the total cost of the epidemic. We show that three main strategies emerge, with treating a large number of individuals (global strategy, GS), treating a small number of individuals in a well-defined neighbourhood of a detected case (local strategy) and allowing the disease to spread unchecked (null strategy, NS). The choice of the optimal strategy is governed mainly by a relative cost of palliative and preventive treatments. If the disease spreads within the well-defined neighbourhood, the local strategy is optimal unless the cost of a single vaccine is much higher than the cost associated with hospitalization. In the latter case, it is most cost-effective to refrain from prevention. Destruction of local correlations, either by long-range (small-world) links or by inclusion of many initial foci, expands the range of costs for which the NS is most cost-effective. The GS emerges for the case when the cost of prevention is much lower than the cost of treatment and there is a substantial non-local component in the disease spread. We also show that local treatment is only desirable if the disease spreads on a small-world network with sufficiently few long-range links; otherwise it is optimal to treat globally. In the mean-field case, there are only two optimal solutions, to treat all if the cost of the vaccine is low and to treat nobody if it is high. The basic reproduction ratio, R_0, does not depend on the rate of responsive treatment in this case and the disease always invades (but might be stopped afterwards). The details of the local control strategy, and in particular the optimal size of the control neighbourhood, are determined by the epidemiology of the disease. The properties of the pathogen might not be known in advance for emerging diseases, but the broad choice of the strategy can be made based on economic analysis only.

Keywords: epidemiological modelling; disease spread; stochastic modelling; epidemiological control

1. INTRODUCTION

Epidemiological modelling has long been used to design strategies to control disease outbreaks [1]. The underlying assumption of these strategies is the wide availability and low economic or social cost of treatment, be it in the form of preventive vaccination or therapy [2]. These assumptions are however not true in many cases, particularly for large outbreaks like cholera [3], AIDS [2], severe acute respiratory syndrome (SARS) [4] or foot-and-mouth disease [5]. There is, therefore, a need for a 'marriage of economics and epidemiology' [2] in designing effective strategies for control of disease [6]. Key to this approach is the realization that an optimal policy does not necessarily result in curing everybody in the population at any cost; it might instead be acceptable to tolerate some lower level of disease persistence if the costs of eradication are prohibitively high [7]. Several recent papers have combined epidemiological with economic constraints to identify optimal strategies for disease control or management [8–12]. Most of these studies, however, ignore the spatial components of disease spread and control

*Author for correspondence (ak@cs.stir.ac.uk).

Received 8 April 2011
Accepted 13 May 2011

This journal is © 2011 The Royal Society

Downloaded from rsif.royalsocietypublishing.org on June 9, 2011

while searching for an optimum strategy (see, however, Rowthorn *et al.* [13]). The spatial scale at which control is applied in relation to the spatial scale of the pathogen dispersal has been identified for many diseases, notably for plant diseases in which the spatial component of the location of the hosts plays a particular important role [14,15]. The relationship between the epidemic and control scales can however be affected by economic aspects of both disease and treatment. Simple network models, while capturing the essence of the topology of spread and control, offer a unique opportunity to analyse the relationship between the epidemic and control scales when there are cost constraints [6,16–20]. In this paper, we analyse a model for optimal control of disease spreading on regular and 'small-world' networks [6,20]. The importance of long-range transmissions in influencing the efficiency of control strategies has been shown for numerous major epidemics of human (e.g. SARS [4] and influenza [21–23]), animal (e.g. foot-and-mouth disease [5,24]) and plant diseases (e.g. citrus canker [25], sudden oak death [26] and rhizomania of sugar beet [14,15]).

There exist two broad strategies in response to a threat of an infectious disease. The authorities can implement control measures before the potential outbreak (e.g. a preventive vaccination [1]) or prepare a set of reactive measures, with a mixture of palliative care and control implemented only after the outbreak. In this paper, we consider the second case and assume that the outbreak has already started. A successful reactive control strategy needs to combine therapy (i.e. treatment of existing cases) with prevention against secondary cases (e.g. vaccination or culling) [2]. Treatment limited to individuals who are displaying symptoms is usually not enough to stop an outbreak, particularly if the disease includes a pre-symptomatic stage [27]. Thus, by the time a symptomatic individual is detected, the disease will have spread well beyond the original focus. Combination of a palliative with a preventive (although applied after the start of the outbreak) treatment allows the control to be more effective, if enough individuals are included in the population to catch all infectious individuals or to remove susceptible ones from the perimeter of the spreading focus [15]. However, such a strategy is also costly—it invariably leads to treating individuals that might never have been infected and become diseased even when no action were taken. If treatment is simple and cheap, this perhaps does not matter. The experience of large outbreaks of foot-and-mouth disease [28,29] and citrus canker [25] shows, however, that treatment cost may be very important. Thus, the process of designing the optimal strategy must involve in the first step the identification of all potential costs (including disease and control costs) and subsequently finding the right balance between them [3].

In this paper, we identify two main sources of costs associated with a disease outbreak and subsequent control [2]. These are the cost of untreated disease cases and the cost of treating individuals located around those cases (including the cost of surveillance needed to identify existing cases). If no preventive measure is taken, infection, and hence disease, spreads and many individuals become ill and either recover or die. This leads to direct costs associated with, for example, hospitalization and drugs that need to be administered and indirect costs associated with the loss of revenue owing to illness, and with death or incapacity of individuals. Such associated costs can be very high if the epidemic is severe and affects all or most of the population. The main objective of the preventive measures is to lower the total cost by investing in treatment or vaccination in the initial stages of the epidemic, with the hope that this will arrest the disease spread [30]. Control might, for example, involve a mass vaccination as early in the outbreak as possible, or continuous preventive vaccination [1,31,32]. Although there is a potentially large cost associated with such a strategy, the investment is seen as worthwhile if it leads to a significantly reduced number of infections owing to removal of susceptible individuals. Vaccination, culling or other forms of preventive treatment can also be targeted, by concentrating on individuals that exhibit disease symptoms or their neighbours, regardless of their status [5,27,33,34]. Such a form of 'ring vaccination' has been identified as a cost-effective measure, since it concentrates the effort where it is needed. The drawback of such strategies is that they require a detailed knowledge of the actual location of infected individuals and their contacts [17], and this might also involve costly surveillance schemes [35].

In this paper, we compare spatially targeted control strategies. We show that, depending on the relative cost of treatment and infection, a choice of three strategies arises: treating nobody (null strategy, NS), treating only selected individuals within a well-defined neighbourhood of each detected (symptomatic) individual (local strategy, LS) and treating as many individuals in the whole population as possible (global strategy, GS). We also show that the randomness of disease distribution in the initial phases of the epidemic plays a very important role in deciding which strategy to choose. This can result either from an initial distribution of disease foci or from topology of interactions. The details of the LS depend on the epidemiology but not on the economic parameters—it is the choice of the strategy that does depend on the relative costs. The 'bang–bang' strategy of either treating nobody or treating all individuals has been observed in non-spatial systems where control strategy varies over time [7,8,36], but to our knowledge not for a spatial control strategy.

2. MODEL

The spatial model that underlies this paper is an extension of the susceptible–infected–removed (SIR) model to account for pre-symptomatic spread [6,20]. We first introduce a spatial model in which control is applied locally in response to observed cases. Subsequently, we construct mean-field approximations for the spatial model.

2.1. Spatial model

For simplicity, we assume that individuals are located at nodes of a square lattice that represents the geographical distribution of hosts. On this lattice, we define a local neighbourhood of order z as a von Neumann neighbourhood in which we include z shells and $\phi(z) = 2z(z+1)$

Downloaded from rsif.royalsocietypublishing.org on June 9, 2011

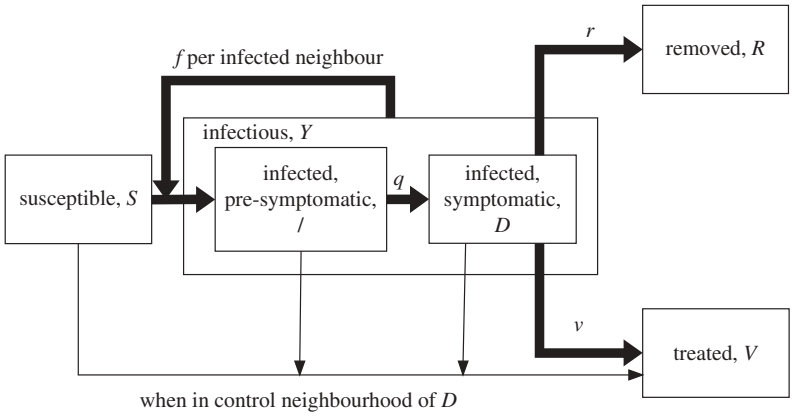

Figure 1. Block diagram illustrating transitions in the model considered in the paper. Thick lines represent transitions performed at each time step, whereas light lines represent transitions triggered by treatment.

individuals (excluding the central one). Thus, $z = 1$ corresponds to the four nearest neighbours, while $z = \infty$ corresponds to the whole population in the limit of infinite size of the system.

The epidemiological model is a version of an SIR model [1], modified to include pre-symptomatic and symptomatic stages of the illness and to account for detection and treatment (figure 1). All individuals are initially susceptible (S). The epidemic is initiated by the introduction of a few infected but pre-symptomatic (I) individuals. Each infectious (pre-symptomatic or symptomatic) individual is in contact with a fixed number of other individuals and infection is transmitted along these contact routes with probability f per contact. Upon successful infection, the susceptible individual moves to the pre-symptomatic class. Stochastic simulations are performed with a fixed time step so that each probability is interpreted as a hazard.

We consider two models for transmission: local-spread and small-world models. In the local-spread model, a fixed number of individuals is chosen in the nearest neighbourhood of order z_{inf} surrounding each susceptible individual. Each infected individual located within the neighbourhood contributes to the total hazard for this particular susceptible individual. We consider $z_{\mathrm{inf}} = 1$ with $\phi(z_{\mathrm{inf}}) = 4$ individuals in the infection neighbourhood, but the results are similar for other choices of z_{inf} . A small-world model [6,37] is similar to the local-spread model, but an additional number of non-local links is added randomly to the lattice of local interactions. These links can span the whole population and the probability of passing an infection along any of the long-range links is the same as for local links.

With a probability q each pre-symptomatic individual develops symptoms that can be detected (and hence moves to class D). Both pre-symptomatic and symptomatic individuals can infect susceptible individuals. At each time step, each symptomatic individual can move to a removed class (R) with a probability r or, if it does not recover, can trigger a treatment event with probability v. This process models delays in public health actions leading to preventive treatment (vaccination or culling). Each treatment event affects the central symptomatic individual and all susceptible S, pre-symptomatic I and symptomatic D (but not removed R) individuals located within a von Neumann neighbourhood of order z centred on a detected individual, as they move to the treated class, V. This represents a localized 'ring' treatment (vaccination or culling). For convenience, we extend the definition of z to include two cases: $z = -1$ describes a strategy in which no spatial control is applied, and $z = 0$ corresponds to a strategy in which the detected individual is treated only. Neither R nor V can infect or be re-infected any more. The number of individuals in each class is denoted by S , I , D , R and V , respectively, and $N = S + I + D + R + V$ is the total number.

2.2. Mean-field equations

The model without control can be described by the following set of mean-field equations:

$$\left.\begin{aligned} \frac{\mathrm{d}S}{\mathrm{d}t} &= \frac{-\beta\phi(z_{\mathrm{inf}})}{N} S(I+D), \\ \frac{\mathrm{d}I}{\mathrm{d}t} &= \frac{\beta\phi(z_{\mathrm{inf}})}{N} S(I+D) - qI, \\ \frac{\mathrm{d}D}{\mathrm{d}t} &= qI - rD \\ \text{and} \qquad \frac{\mathrm{dR}}{\mathrm{d}t} &= rD. \end{aligned}\right\} \tag{2.1}$$

The parametrization of the infection force by $\beta\phi(z_{\mathrm{inf}})$ allows a direct comparison of the simulations with the fully spatial model, although β can only cautiously be interpreted as an equivalent of f. If the control is just applied to the detected individual ($z = 0$), these individuals are removed at the rate v and the equation for D is modified by including a term $-vD$,

$$\frac{\mathrm{d}D}{\mathrm{d}t} = qI - rD - vD. \tag{2.2}$$

When $z > 0$, an additional number of individuals, $\phi(z)$, is selected for treatment. In the spatial model, those individuals are located in the neighbourhood of the

Downloaded from rsif.royalsocietypublishing.org on June 9, 2011

detected individual, but, in the mean-field approximation, the spatial information is lost. Thus, the corresponding number of individuals is selected at random from the population at each control event. As the control events occur at the rate $-vD$, the rate at which individuals are treated equals $-v\phi(z)D$. Out of these, a proportion of S/N individuals are susceptible, I/N individuals are pre-symptomatic and D/N are symptomatic (the control event does not distinguish between the state of the individuals subject to treatment, except for the removed class). Incorporating the relevant terms into equation (2.1) we obtain

$$\left.\begin{aligned}
\frac{\mathrm{d}S}{\mathrm{d}t} &= -\frac{\beta\phi(z_{\mathrm{inf}})}{N}S(I+D) - v\phi(z)D\frac{S}{N},\\
\frac{\mathrm{d}I}{\mathrm{d}t} &= \frac{\beta\phi(z_{\mathrm{inf}})}{N}S(I+D) - qI - v\phi(z)D\frac{I}{N},\\
\frac{\mathrm{d}D}{\mathrm{d}t} &= qI - rD - vD - v\phi(z)D\frac{D}{N},\\
\frac{\mathrm{dR}}{\mathrm{d}t} &= rD\\
\text{and}\qquad \frac{\mathrm{d}V}{\mathrm{d}t} &= v\phi(z)D\frac{S+I+D}{N} + vD.
\end{aligned}\right\} \tag{2.3}$$

2.3. Cost of treatment

From an economic point of view, the problem of designing an optimal control strategy can be viewed as a special case of a net present value test [38]. In this approach, the value of future benefits (reduction in the number of infection cases) is compared with the value of future and current costs (associated with a particular control strategy). The values are often discounted if the optimization horizon spans a longer period of time. For simplicity, we assume that the duration of an epidemic is short enough (e.g. within 1 year) so that no discounting is necessary. The strategy is decided at the beginning of the epidemic and is not changed over time. The economic outcome, on the other hand, is deferred until the end of the epidemic when costs are compared with gains. We also assume that there are no budget constraints and so the decision maker can spend as much as is necessary on controlling the disease within the prescribed strategy.

In this paper, we aim to minimize the total cost of the outbreak and we allocate costs to two groups. The first term representing the palliative cost is associated with individuals who are never treated and therefore spontaneously move into the removed class. This term is equal to $R(\infty)$ multiplied by a unit cost of treatment, c_1. The second term describes costs associated with treatment of susceptible and pre-symptomatic individuals aimed at prevention of further spread. For simplicity, we assume that this term also includes surveillance costs involving searching for and detection of infected (symptomatic) individuals as well as treatment of any symptomatic individuals (including the one that triggered the treatment event; figure 1). Thus, the second term is equal to $c_2 V(\infty)$, with c_2 being a unit cost of preventive treatment. These assumptions lead to the following general form for the total cost of the epidemic:

$$X = c_1 R(\infty) + c_2 V(\infty). \tag{2.4}$$

We are normally not interested in the absolute measure of X, but only intend to use it to compare different strategies. Thus, without loss of generality we can put $c_1 = 1$ and $c_2 = c$, so that $X = R + cV$ with c measuring the relative cost of treatment to infection [6,17]. The goal of the simulation is to find an optimal control strategy, identified here with a value of z (and denoted z_c) for the spatial model (and its mean-field approximations), that minimizes the total cost, X, with other parameters fixed. We call X the severity index, as it characterizes the combined severity of the epidemic including individuals that have been through the disease but were not treated (R) and individuals that have been treated both in response to their symptoms and preventively to halt the spread of the disease (V).

We consider two prevention strategies exemplifying our approach, preventive vaccination (or spraying) and culling (or destruction), for three complementary diseases, influenza [39–41], foot-and-mouth disease [24,27] and citrus canker [25], although our approach is more general. Attempts to control an influenza outbreak include preventive vaccination or treatment with anti-viral drugs [42]—a similar approach has been suggested for measles [43] and for Ebola [44]. For foot-and-mouth disease, both vaccination and preventive slaughter of animals on contiguous premises [24,45] have been used to control spread. Likewise, citrus canker can be controlled by early spraying with copper compounds on resistant varieties, but immediate and rapid destruction of infected trees is essential for controlling the spread [46]. The two exemplary treatments differ in costs associated with them. Vaccination (for influenza or foot-and-mouth disease) and preventive spraying (for citrus canker) are typically cheaper than loss of an individual owing to disease (foot-and-mouth disease, canker) or costs associated with inability to work or even hospitalization (influenza). Thus, for example, Weycker *et al.* [40] estimates the costs of influenza vaccine at $c_2 =$ US$6–24, with direct costs of infection at $c_3 =$ US$70 and indirect at US$351, leading to c ranging from 0.017 to 0.341 (see also Meltzer *et al.* [39]). Similar estimates can be obtained for rotavirus and hepatitis A [47,48], with $c = 0.01 - 0.85$. On the other hand, the cost of culling an animal or destroying a tree is typically comparable to or more expensive than the disease, as it includes not only the lost revenue associated with the culled animal or destroyed tree but also the labour associated with treatment; this leads to $c \gtrsim 1$.

2.4 Simulations

Simulations were performed on a lattice of 200 by 200 individuals with periodic boundary conditions. The size of the lattice was a compromise between numerical efficiency and small-size effects that we wanted to avoid. We performed simulations for other sizes and found no effect for sufficiently large lattices. We have considered a range of initial numbers of infected individuals, but

Downloaded from rsif.royalsocietypublishing.org on June 9, 2011

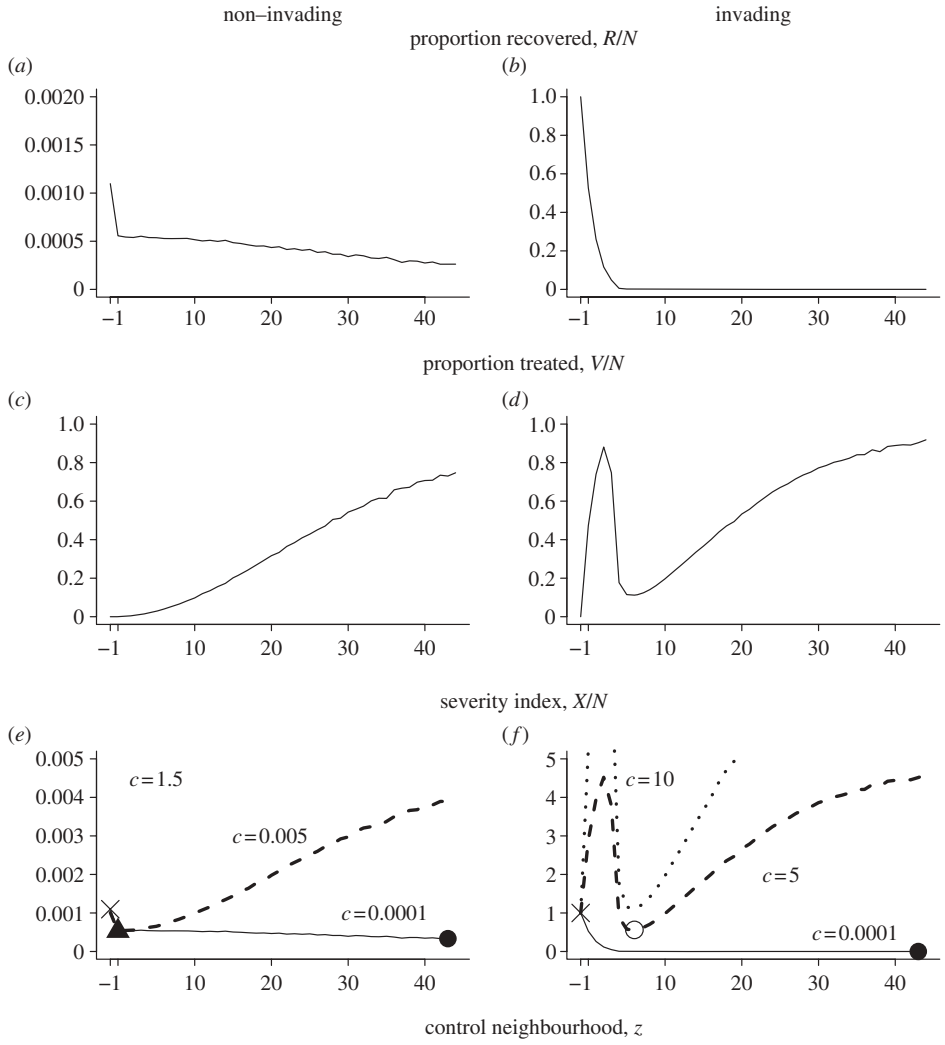

Figure 2. (*a,b*) The proportion of spontaneously removed individuals, $R(\infty)$; (*c,d*) treated individuals, $V(\infty)$, as a function of the control neighbourhood, z, for different values of f and c. The corresponding values of the severity index, X, are shown in (*e,f*). Points mark the global minimum values of X, X_c and the corresponding optimal control radius, z_c (treat all, filled circles; treat some, open circles; treat detected only, filled triangles; treat none, crosses). $f = 0.001$ in (*a,c,e*), representing a non-invasive disease; solid line and filled circles, $c = 0.0001$; dashed line and filled triangles, $c = 0.005$; dashed-dotted line and crosses, $c = 1.5$. $f = 0.5$ in (*b,d,f*), for an invasive disease; solid line and filled circles, $c = 0.0001$; dashed line and filled circles, $c = 5$; dashed-dotted line and crosses, $c = 10$. Initial condition is $I(0) = 40$ (0.1% of the total population). Other parameters: $z_{\mathrm{inf}} = 1$, $q = 0.5$, $r = 0.1$, $v = 0.1$. Note that $z = -1$ corresponds to no control whereas $z = 0$ represents controlling the detected individual only.

the results are shown for 40 initial foci (0.1% of the total population) and 400 initial foci (1% of the total population). Smaller numbers of initial foci led to too many cases in which disease died out without spreading, which affected the optimization procedure. Except when indicated otherwise, $z_{\mathrm{inf}} = 1$, $v = 0.1$, $r = 0.1$ and $q = 0.5$. Each simulation was run until $I(t) + D(t) = 0$ and X was computed at the end of the run. For the simulation model, the minimization of X is achieved by sweeping through different values of z while performing only a single simulation for each value of z. For such a sample, the actual minimal value of X and the corresponding value of z are found. This procedure is then repeated 100 times to yield average values of z_c and X_c and their standard deviations. Numerical solution of the differential equations was done using *R* [49].

3. RESULTS

The long-term behaviour of the spatial model in the absence of control ($z = -1$) is determined by f, the probability that infection is passed to a susceptible node from any of the four neighbours ($z_{\mathrm{inf}} = 1$). For small values of f, the disease quickly dies out, whereas, for large values of f, the pathogen and hence disease is highly contagious and spreads through the whole population when no treatment is applied, $X \simeq R(\infty) \simeq N$; compare figure 2*a,b*. The extreme cases of f are

Downloaded from rsif.royalsocietypublishing.org on June 9, 2011

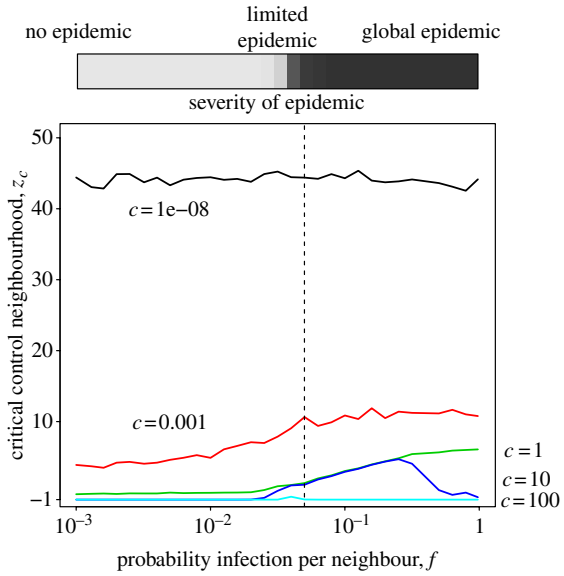

Figure 3. Critical value for the range of control neighbourhood, z_c, as a function of the probability of spread, f, for different values of the relative treatment cost, c. Other parameters: $q = 0.5$, $v = 0.1$, $r = 0.1$ and $z_{\inf} = 1$. The initial condition is $I(0) = 40$ (0.1% of the total population). The figure at the top shows the average final size of epidemics in the absence of any control, $R(\infty)$ (light grey corresponds to low values of $R(\infty)$ and dark grey to high values of $R(\infty)$). Vertical line at $f = 0.04$ corresponds approximately to the threshold for invasion. (Online version in colour.)

separated by a threshold for disease invasion, with an exact critical value of f depending on the spatial structure of the network and presence or absence of long-range links. For the simplest case of $z_{\inf} = 1$ and no long-range links, the transition occurs at $f = 0.04$ (figure 3); addition of the small-world links shifts the threshold towards the value of $f = 0.02$ that can be compared with the mean-field critical value of $\beta = 0.02$ associated with $R_0 = 1$ (for details of mean-field calculations see below and in particular equation (3.1)). When control is applied, $z \geq 0$, R/N declines monotonically with the order of the control neighbourhood, z, for both invasive and non-invasive diseases (figure 2*a*,*b*). Thus, the increased control effort leads to a reduction in the number of cases. However, the number of individuals treated, $V(\infty)$, increases at the same time (figure 2*c*,*d*). The increase is monotonic for a non-invading disease (figure 2*c*), but non-monotonic for an invading disease (figure 2*d*). $R(\infty)$ and $cV(\infty)$ are subsequently combined to form $X = R(\infty) + cV(\infty)$ (figure 2*e*,*f*). The special case of a vaccination that does not cost anything, $c = 0$ (not illustrated in figures), corresponds to $X = R(\infty)$ and leads to an optimal strategy of treating all (GS). If $c \neq 0$, various types of global minima can be obtained depending on the value of c and the shape of v.

First, consider a non-invasive disease (figures 2*a*,*c*,*e* and 3). If vaccination is cheap (small but finite c), X is dominated by $R(\infty)$ (the cost of an uncontrolled epidemic) and the minimum value of X occurs at $z_c = \infty$ corresponding to the GS of treating all individuals (GS) (figure 2*e*: filled circle). As the cost, c, increases, the minimum rapidly shifts to $z = 0$, corresponding to treating just the detected individual (a subset of the local strategy, LS; figure 2*e*: filled triangle). For very high values of c (thick line in figure 2*e*), the strategy shifts further to $z = -1$ when nobody is treated (the NS). The value of a critical control radius, z_c, depends strongly on c but only weakly on f for small values of f (figures 2*e* and 3).

As f increases and the epidemic character changes from non-invading to invading, $V(\infty)$ becomes a non-monotonic function of z (figure 2*d*). While for small values of c, the GS is still the best option (figure 2*f*), a new type of LS appears for moderate values of c, corresponding to the treatment within a well-defined region around each detected case. For a very high value of c, the minimum of X corresponding to a finite value of z disappears and the NS of treating nobody becomes optimal (figure 2*f*). The switch from GS to LS and subsequently to NS is clearly seen in figure 3, which also shows the relative independence of the choice of the optimal control strategy on f.

Thus, the choice of the optimal strategy is determined by two main factors: the infectiousness of the disease, f, and the relative cost of the treatment, c. The dependence on the rate of disease spread, f, is relatively weak for most values of c (figure 3). The values of the optimal control neighbourhood, z_c, cluster in two regions. For small c ($c < 10^{-4}$), z_c is independent of f and corresponds to a GS, $z_c \simeq 45$. For moderate c ($0.01 < c < 1$), z_c is below 10 (for the parameters discussed here) and slowly increases as the disease switches from non-invasive to invasive. For high values of c ($10 < c < 100$), the dependence on f is non-monotonic as z_c first increases and subsequently drops back to 0 (treat only detected individuals). Finally, for very high costs of treatment, $z_c = -1$ (refrain from treatment) for almost all values of f.

The economic aspects of the control determine three regions for c (figure 4). To illustrate the details of the behaviour, we assume that each untreated case (i.e. the individual in the removed class, R, at the end of the epidemic) costs £100. (We use arbitrary but realistic values here, to illustrate general principles rather than to focus on a particular disease.) We consider two contrasting cases for the cost of each treated individual (i.e. the individual in the treated class, V, at the end of the epidemic), £0.01 and £1000. We also assume that initially there are $I(0) = 40$ cases in a population of 40000. Consider first the costs of the NS, under which nobody is treated and so $X \simeq R(\infty)$. For the non-invasive disease (small f), the total cost is approximately £$100I(0) = 4000$, whereas for the invasive disease (high f) the total cost reaches £$100N = 4$ million. For the GS, we treat all individuals indiscriminately and as quickly as possible and so the cost is £$0.01N = 400$ for small c and £$1000N = 40$ million for large c, independent of f. Finally, for the LS, it is not possible to obtain a simple estimate of the cost as it depends on z and the effectiveness of prevention.

For the very cheap preventive treatment (e.g. costing £0.01, i.e. $c = 10^{-4}$), $cN < I(0)$, the cost of treating the whole population (GS, £400) is smaller than the cost associated with the infection of the initial cases

Downloaded from rsif.royalsocietypublishing.org on June 9, 2011

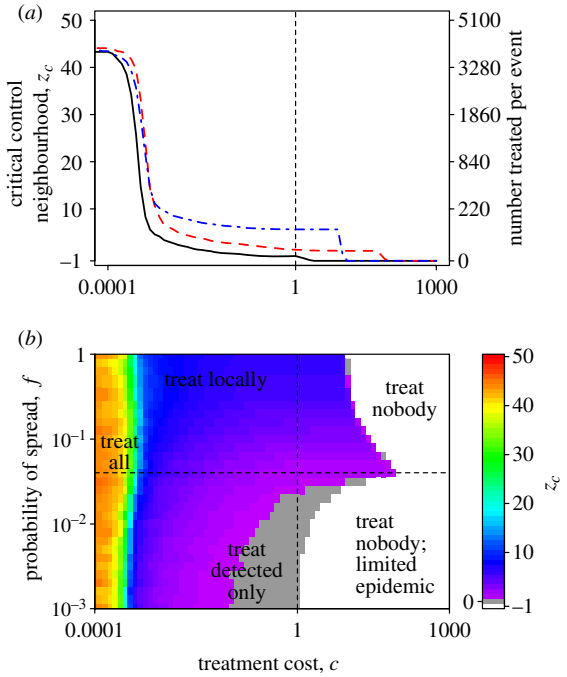

Figure 4. Critical value for the range of control neighbourhood, z_c, as a function of the relative treatment cost, c, for different values of the probability of spread f. (*a*) Examples of the dependence of the optimal control radius, z_c, on the treatment cost for various values of f (solid line, $f = 0.005$; dashed line, $f = 0.05$; dashed-dotted line, $f = 0.5$). The optimal control radius, z_c, is also shown in (*b*) as a function of both f and c. White colour represents $z_c = -1$ (NS), grey colour corresponds to treating detected individuals only ($z_c = 0$) and a range of colours represent $z_c > 0$. Other parameters: $z_{\mathrm{inf}} = 1$, $q = 0.5$, $v = 0.1$, $r = 0.1$. The initial condition is $I(0) = 40$ (0.1% of the total population). The vertical line is placed at $c = 1$ (equal costs of disease and treatment). The horizontal line at $f = 0.04$ separates invasive and non-invasive epidemics (cf. figure 3). (Online version in colour.)

(NS, £4000). Thus, for both invasive and non-invasive diseases, it is better to spend £400 and stop the epidemic immediately than to allow even the initial cases to go through the disease process (at a minimum cost of £4000). If the cost of treating the whole population is comparable to or higher than treating the initial cases, $c \geq I(0)/N$, the GS is no longer optimal. For high c and low f, if the treatment cost of just the few initial cases (£40 000) is significantly higher than the cost of allowing the epidemic to run to its completion (NS for low f, £4000), we expect the NS to be optimal. Similarly, the cost of the GS is high (£40 million) compared with the NS (£4 million) for large f and the NS is again optimal. The range of c between those two extremes is occupied by the LS with $z_c = 0$ (treat only detected individuals) for the non-invasive disease and $z_c < 10$ for the invasive disease (figure 4).

A remarkable feature of the LS is the stability of z_c as a function of c over a wide range of c and f, (figure 4*a*; see also figure 3). Interestingly, even in cases when the cost of the preventive treatment, c, exceeds the cost of uncontrolled disease ($c > 1$), the LS is still optimal for some combinations of c and f (even though the NS is optimal for very high values of c. The mechanism for this behaviour is related to spatial correlations in the spread of the disease. Consider a focus originating with a single pre-symptomatic but infectious individual. Infection subsequently spreads to its nearest neighbour and then to their neighbours, but the focus still remains undetected. It is only when the first individual in the group shows symptoms that the authorities might become aware of the infection (this individual is usually the original source of infection, but owing to the stochastic nature of the process it might also be another one). Further delay (represented by the finite value of v) before any responsive treatment (vaccination or culling) is applied leads to further expansion of the focus. Thus, with a high probability, we can expect pre-symptomatic but infectious individuals in the immediate neighbourhood of a detected one. The optimal local control strategy will aim at treatment of all such pre-symptomatic individuals, but without extending the control neighbourhood too far (which will lead to an unnecessary increase in costs). This is a similar mechanism to herd immunity [1], but local application makes it a very effective strategy. Thus, the epidemic can be stopped within a few steps, even though the rest of the population remains susceptible. We also note that the fewer the initial foci, the less effort is required to stop the outbreak in this case, and so we expect that the critical value of c determining the transition between the LS and the NS will increase with a decreasing number of initial foci. However, once the spatial correlation is destroyed, we expect the LS to be no longer efficient for any value of $c > 1$.

3.1. Destroying spatial structure

The spatial correlations can be destroyed either by introducing non-local spread, for example in the form of long-range links in a small-world model, or by increasing the number of initial foci. There is not much change in the behaviour for small f (cf. figure 4 with figure 5) where 30 per cent long-range links have been added to the model structure. In this case, it is still preferable to treat individuals locally for a broad range of c . However, as the disease becomes more infectious, the probability of it spreading via long-range links increases. In this case, the region of optimality for the GS extends to higher values of c, whereas the range of the NS extends to lower values of c until they merge at $c = 1$ for high f (cf. figure 4 with figure 5).

The effect of changing the number of non-local links is shown in figure 6, which is analogous to figures 4*a* and 5*a*, but for a smaller range of c and for a single value of $f = 0.98$. For 40 initial foci (0.1% of the total number) and the purely local spread, the switch between the LS and the NS occurs at approximately $c = 10$ (thin line in figure 6*a*) while $z_c \simeq 6$ for c below 10. The addition of 2 per cent links decreases the range of c for which the LS is still optimal but does not increase z_c (dashed line). However, the number of individuals treated preventively, $V(\infty)$, increases markedly compared with the purely local case (figure 6*b*). Addition of 30 per cent long-range links shifts the critical value

Downloaded from rsif.royalsocietypublishing.org on June 9, 2011

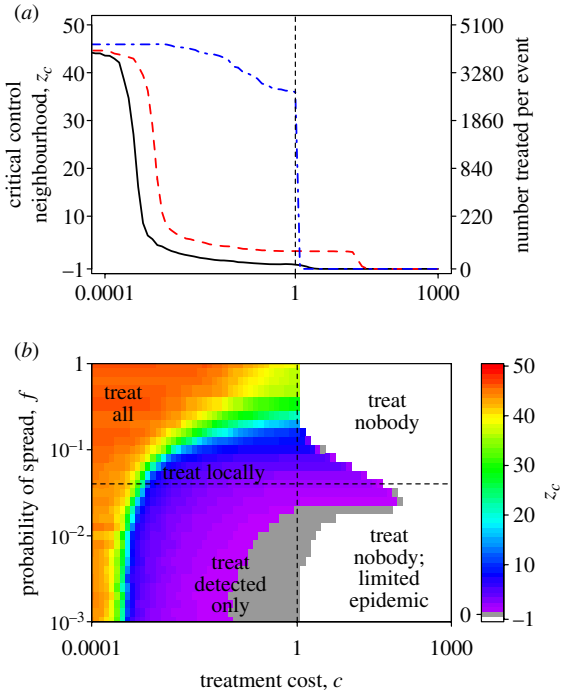

Figure 5. Critical value of the range of control neighbourhood, z_c, as a function of the relative treatment cost, c, and the probability of spread, f, but for 30% long-range links. Other parameters and labels as in figure 4. (Online version in colour.)

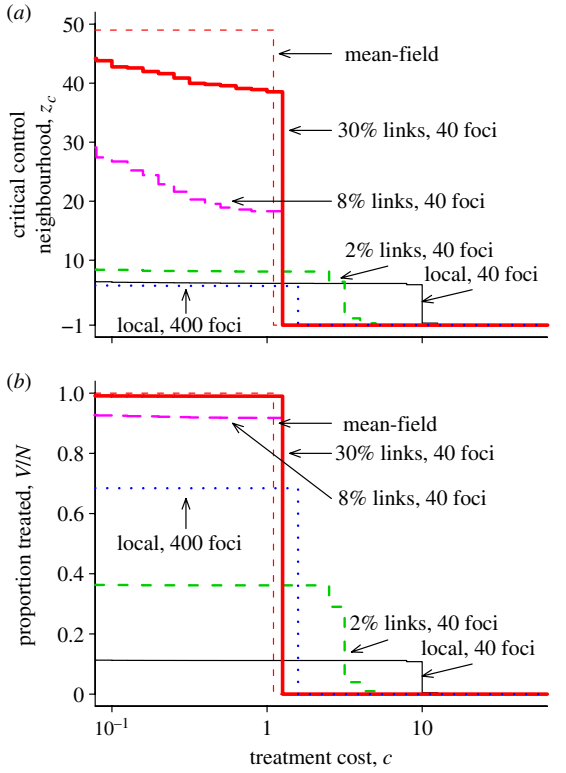

Figure 6. (*a*) The optimal control neighbourhood, z_c, and (*b*) the proportion of treated individuals, $V(\infty)/N$, as functions of the treatment cost, c, for local network, small-world network and mean-field models. For the number of initial foci $I(0) = 40$ the number of non-local links as a proportion of the total number of nodes is marked as follows: 0%, solid line; 2%, dashed line; 8%, thick dashed line; and 30%, thick line. The dashed-dotted line corresponds to the local spread (0% links), but the number of initial foci is increased to $I(0) = 400$. The thin-dashed line represents the simulation of the mean-field model. The rate of disease spread is $f = 0.98$; other parameters: $q = 0.5$, $r = 0.1$, $v = 0.1$. (Online version in colour.)

for c close to $c = 1$, while increasing the size of the control neighbourhood to $z_c \simeq 40$ (thick line in figure 6*a*). In this case, the proportion of the population that needs to be treated, $V(\infty)/N$, is also very high for $c < 1$ (figure 6*b*).

If the number of initial foci is increased without addition of long-range links, the effect on the critical value of c is similar to the addition of links, although there is no noticeable increase in z_c below the critical value (the dashed-dotted line in figure 6*a*). Thus, in each treatment event, we are still treating a small number of individuals. However, overall, we still need to treat a large proportion of individuals (figure 6*b*).

The change in z_c and $V(\infty)$ can be very rapid as long-range links are added to the system (figure 7*a*). This is reminiscent of the rapid transition associated with the small-world model in which addition of only a few links can drastically change the behaviour of the system [37]. If the preventive treatment (e.g. vaccination or culling) is even marginally more costly than allowing the disease to run without control, $c = 1.25$, the addition of 6 per cent of long-range links renders the LS inefficient (figure 7*a*). In this case it is best to refrain from any preventive treatment (and follow the NS), even if 4 per cent links still leave the LS optimal. The reason for this critical behaviour is clear from figure 7*b*. Consider the case of $c = 1$ for which it is optimal to treat locally even for a large number of non-local links. However, in this case, the proportion of treated individuals exceeds 50 per cent of the total population for 5 per cent or more of long-range links (marked by the arrow in figure 7*b*). This shows how critical it is to reduce the number of non-local links in the population [6,17], if local control strategies are applied.

3.2 Mean-field limit

With the increase in the number of non-local links, we are approaching the mean-field approximation (figure 6*a*). In this case, there are only two options for treatment. The GS is optimal for $c \leq 1$ and the NS is optimal for $c > 1$. This can be confirmed by the analysis of the mean-field equations (2.3). The responsive treatment in which the treatment rate depends on the current number of detected cases is not capable of controlling the invasion of the disease. When the basic reproduction ratio, R_0, is computed for equation (2.3), the result does not depend on the rate of treatment v,

$$R_0 = \beta\phi(z)\left(\frac{1}{q} + \frac{1}{r}\right). \tag{3.1}$$

Downloaded from rsif.royalsocietypublishing.org on June 9, 2011

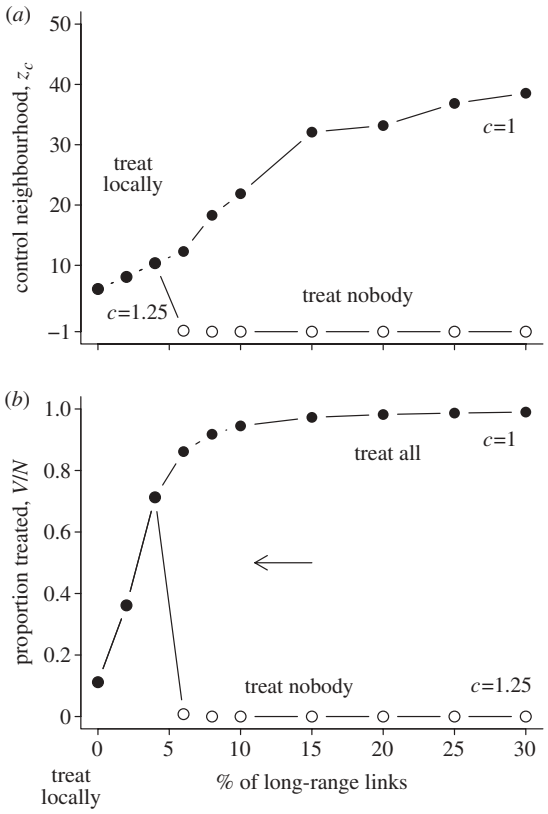

Figure 7. (*a*) The optimal control neighbourhood, z_c, and (*b*) the proportion of treated individuals, $V(\infty)/N$, for the small-world model with a varying number of long-range links and for two similar values of $c = 1$ and $c = 1.25$. Other parameters as in figure 6; $I(0) = 40$. The arrow marks the point at which $V/N = 0.5$ (see text).

In this formula, $\beta\phi(z)$ is the rate of infection, $1/q$ is the average time an infected individual spends before detection and $1/r$ is the average time a detected individual spends before spontaneous removal. As a consequence, the stability of the disease-free equilibrium ($I = 0$, $D = 0$) is unaffected by the control since, for low levels of infection ($I,D \ll N$), the control term is very small. Although as the number of cases increases, so does the control effort, but the dependence of the control rate on D means that the effort always follows the infection. Simulations show that the final number of treated individuals, $V(\infty)$, and the final number of spontaneously removed individuals, $R(\infty)$, are closely related in this case so that, if $c = 1$, $X = R(\infty) + V(\infty) = N$ independently of v. Thus, if it is cheaper to prevent the disease than to avoid treatment, $c \leq 1$, it is best to treat all individuals (GS). By contrast, if it is cheaper to refrain from treatment, $c > 1$, it is best not to treat anybody (NS). However, regardless of the applied control strategy, the whole population is affected either by the infection or by the control (figure 6*b*). The results agree with simulations of a stochastic spatial model in which individuals contact $\phi(z_{\mathrm{inf}})$ individuals randomly (not shown here).

4. DISCUSSION

The main objective of the 'optimal' control strategy is to stop the epidemic not only in the shortest possible time but also at a manageable cost. Faced with a large outbreak, health authorities need to decide quickly whether to build up a coordinated effort to vaccinate or to treat a large proportion of the population, despite often substantial costs involved [3]. In some cases, refraining from treatment might be a more cost-effective choice than to act. Such decisions are often very difficult, as they involve many unknown factors. Mathematical modelling is then used to provide help and guidance by, among others, pointing to factors that do or do not influence the final outcome of the control process. Among the factors that mainly influence the decision are the costs associated with both preventive control and the disease itself. While the first category can be estimated with a certain degree of accuracy, the second factor might be difficult to determine.

In this paper, we provide a systematic study of the choice of the optimal strategy for a range of diseases for which spread is either localized or not. We have identified three basic strategies, the GS (treat all), the local strategy (treat within a well-defined neighbourhood of any detected individual or treat just the detected individual) and the NS (do not preventively treat any individual). In the last case, the individuals can still be treated for disease symptoms, but no prevention is effected on the population.

The details of the LS (when it is applied) surprisingly do not depend strongly on the cost of treatment, although the decision whether to apply the control locally or globally (or not at all) does depend on the cost. Once we decide on application of the local control, it is the epidemiology and social network structure that determine the spatial extent of LS. The results presented here for the LS show that it is important to match the scale of control with the scale of the disease dispersal; see [14] for a practical application in matching scales for control with the inherent scale of spread for a crop disease at the landscape scale. There are, however, also cases when the balance of costs is an over-riding factor and it is necessary to treat all individuals as quickly as possible (GS) or to refrain from treatment (NS).

When the purely local structure of the disease spread is destroyed by an increase in the number of initial foci or by addition of long-range links, local control can still be applied. Dybiec *et al.* [6] found that, for a small number of links, the local strategy still works, but at a cost of an increased control neighbourhood. This is necessary to catch the pre-symptomatic individuals before they cause new foci to appear via long-range links. Interestingly, the case of $c = 1$ that was considered by Dybiec *et al.* [6] corresponds to the minimal impact of a small-world structure in the order of control neighbourhood. If the cost of treatment is only marginally higher than the cost associated with infection (i.e. $c > 1$), it might be more profitable to withhold treatment completely rather than to use local control strategies.

Our cost function, equation (2.4), is linear in R and V and we assume that the budget is unlimited. For

Downloaded from rsif.royalsocietypublishing.org on June 9, 2011

rapidly spreading epidemics, there might be a situation when the number of cases in a certain locality exceeds the maximum capacity of the control system (either health or veterinary care system). This leads to a rapid increase in costs per treated individual when compared with a small-size outbreak [50,51]. We extended our model to include nonlinear (quadratic) terms in either R and V, but there was no qualitative change compared with the linear cost function. In particular, increasing ε in $X = (R(\infty) + \varepsilon R^2(\infty)) + cV(\infty)$ shifts the curve in figure 4*a* horizontally towards the higher values of c (results not shown). Thus, the range of c for which the GS is optimal increases, whereas the range for which the NS is optimal decreases. This can be understood in terms of the penalty against outbreaks with a large value of r, leading to more strict criteria for control. The effect of including a nonlinear (quadratic) term in v is opposite, as the areas of the NS and LS shift in the direction of lower values of c. The critical value of z_c at the plateau is unaffected in both cases. Here we are penalizing against outbreaks leading to large spending on prevention and therefore are more likely to let the disease spread unchecked.

The critical control neighbourhood, z_c, and the resulting severity index, X, are very sensitive to the percentage of long-range links (figure 6). However, precise network structure and the actual number of long-range links are unlikely to be exactly known. In this case, the precautionary principle suggests to expect the worst case scenario and to either use the largest possible number of expected links, if known, or use the mean-field approximation corresponding to a large number of such links. In this case, the critical value of c is 1 and therefore the GS is optimal for $c \leq 1$ and the NS for $c > 1$.

The current work assumes that the time span of the potential epidemic is very short and so no discounting is applied. In addition, we assume that, once the strategy is decided at the start of the epidemic, the authorities continue with the implementation. Each of these assumptions can be relaxed. A general relationship between the cost and the epidemic variables can be written as

$$X = \int_0^\infty (F_1(I(t), D(t)) + F_2(D(t)) + F_3(D(t), V(t)))\mathrm{e}^{-\delta t}\mathrm{d}t, \tag{4.1}$$

where F_1 is a functional representing *responsive* costs, F_2 represents *surveillance* costs and F_3 corresponds to *prevention* costs while δ is a discounting factor. Under some simple assumptions on the functionals F_1, F_2 and F_3, we recover equation (2.4) if discounting is ignored and if the costs are only counted at the end of the epidemics. In general, however, the costs need to be evaluated as the epidemics unfold. Similarly, the radius of control neighbourhood z can change in time. This approach would require changes to the simulation procedure as it is no longer efficient to scan all possible values of z to search for z_c as done in this paper.

The model describes a single, relatively short outbreak of a disease that either kills the infected individuals or leads to complete immunity and also ignores influx of new susceptibles. Extension of the model to include recovery and/or re-infection (as in an SIS model) is planned for the future, but would require a different approach to cost calculations. We have also assumed that all social, economic and epidemiological parameters are fixed and well known in advance. This is not the case for emerging diseases. There is, therefore, a need to study the sensitivity of various control strategies to uncertainties in f, z_{inf} and the structure of the network. The long-term goal is to identify a selection of strategies that can be applied at the beginning of an emerging epidemic, even if we do not know the details of the disease, and then modified as the epidemic unfolds. However, the results of this paper suggest that if c can be reliably estimated in advance, we can decide between the overall control strategies (NS, LS or GS) even without knowing exactly what the value of f is for a given emerging disease.

Work carried out by K.O. and A.K. is supported by the International PhD Projects Programme of the Foundation for Polish Science within the European Regional Development Fund of the European Union, agreement no. MPD/2009/6. A.K. acknowledges funding by DEFRA, and C.A.G. gratefully acknowledges the support of a BBSRC Professorial Fellowship. The authors are grateful to Peter Dickinson for his involvement in the initial stages of the project. We are also very grateful to four anonymous referees who helped with manuscript revisions.

REFERENCES

1 Anderson, R. & May, R. 1991 *Infectious diseases of humans: dynamics and control*. Oxford, UK: Oxford University Press.
2 Gersovitz, M. & Hammer, J. 2003 Infectious diseases, public policy, and the marriage of economics and epidemiology. *World Bank Res. Obs.* **18**, 129–157. (doi:10.1093/wbro/lkg011)
3 Jeuland, M., Lucas, M., Clemens, J. & Whittington, D. 2009 A cost-benefit analysis of cholera vaccination programs in Beira, Mozambique. *World Bank Econ. Rev.* **23**, 235–267. (doi:10.1093/wber/lhp006)
4 Dye, C. & Gay, N. 2003 Modeling the SARS epidemic. *Science* **300**, 1884–1885. (doi:10.1126/science.1086925)
5 Woolhouse, M. *et al.* 2001 Epidemiology—foot-and-mouth disease under control in the UK. *Nature* **411**, 258–259. (doi:10.1038/35077149)
6 Dybiec, B., Kleczkowski, A. & Gilligan, C. 2004 Controlling disease spread on networks with incomplete knowledge. *Phys. Rev. E* **70**, 066145. (doi:10.1103/PhysRevE.70.066145)
7 Forster, G. A. & Gilligan, C. A. 2007 Optimizing the control of disease infestations at the landscape scale. *Proc. Natl Acad. Sci. USA* **104**, 4984–4989. (doi:10.1073/pnas.0607900104)
8 Barrett, S. & Hoel, M. 2007 Optimal disease eradication. *Environ. Dev. Econ.* **12**, 627–652. (doi:10.1017/S1355770X07003816)
9 Klein, E., Laxminarayan, R., Smith, D. L. & Gilligan, C. A. 2007 Economic incentives and mathematical models of disease. *Environ. Dev. Econ.* **12**, 707–732. (doi:10.1017/S1355770X0700383X)
10 Levin, S. A. 2007 Introduction: infectious diseases. *Environ. Dev. Econ.* **12**, 625–626. (doi:10.1017/S1355770X07003798)
11 Mbah, M. L. N., Forster, G. A., Wesseler, J. H. & Gilligan, C. A. 2010 Economically optimal timing for crop disease

Downloaded from rsif.royalsocietypublishing.org on June 9, 2011

control under uncertainty: an options approach. *J. R. Soc. Interface* **7**, 1421–1428. (doi:10.1098/rsif.2010.0056)
12 Mbah, M. L. N. & Gilligan, C. A. 2010 Optimization of control strategies for epidemics in heterogeneous populations with symmetric and asymmetric transmission. *J. Theoret. Biol.* **262**, 757–763. (doi:10.1016/j.jtbi.2009.11.001)
13 Rowthorn, R., Laxminarayan, R. & Gilligan, C. A. 2009 Optimal control of epidemics in metapopulations. *J. R. Soc. Interface* **6**, 1135–1144. (doi:10.1098/rsif.2008.0402)
14 Stacey, A., Truscott, J., Asher, M. & Gilligan, C. 2004 A model for the invasion and spread of rhizomania in the United Kingdom: implications for disease control strategies. *Phytopathology* **94**, 209–215. (doi:10.1094/PHYTO.2004.94.2.209)
15 Gilligan, C. A., Truscott, J. E. & Stacey, A. J. 2007 Impact of scale on the effectiveness of disease control strategies for epidemics with cryptic infection in a dynamical landscape: an example for a crop disease. *J. R. Soc. Interface* **4**, 925–934. (doi:10.1098/rsif.2007.1019)
16 Dybiec, B., Kleczkowski, A. & Gilligan, C. 2005 Optimising control of disease spread on networks. *Acta Phys. Pol. B* **36**, 1509–1526.
17 Kleczkowski, A., Dybiec, B. & Gilligan, C. A. 2006 Economic and social factors in designing disease control strategies for epidemics on networks. *Acta Phys. Pol. B* **37**, 3017–3026. (http://arxiv.org/abs/physics/0608141)
18 Dybiec, B. 2008 Random strategies of contact tracking. *Phys. A Stat. Mech. Appl.* **387**, 4863–4870. (doi:10.1016/j.physa.2008.04.027)
19 Dybiec, B. 2009 SIR model of epidemic spread with accumulated exposure. *Eur. Phys. J. B* **67**, 377–383. (doi:10.1140/epjb/e2008-00435-y)
20 Dybiec, B., Kleczkowski, A. & Gilligan, C. A. 2009 Modelling control of epidemics spreading by long-range interactions. *J. R. Soc. Interface* **6**, 941–950. (doi:10.1098/rsif.2008.0468)
21 Hollingsworth, T. D., Ferguson, N. M. & Anderson, R. M. 2007 Frequent travelers and rate of spread of epidemics. *Emerg. Infect. Dis.* **13**, 1288–1294.
22 Hsu, C.-I. & Shih, H.-H. 2010 Transmission and control of an emerging influenza pandemic in a small-world airline network. *Accident Anal. Prevent.* **42**, 93–100. (doi:10.1016/j.aap.2009.07.004)
23 Hosseini, P., Sokolow, S. H., Vandegrift, K. J., Kilpatrick, A. M. & Daszak, P. 2010 Predictive power of air travel and socio-economic data for early pandemic spread. *PLoS ONE* **5**, e12763. (doi:10.1371/journal.pone.0012763)
24 Kao, R. R. 2002 The role of mathematical modelling in the control of the 2001 FMD epidemic in the UK. *Trends Microbiol.* **10**, 279–286. (doi:10.1016/S0966-842X(02)02371-5)
25 Gottwald, T., Hughes, G., Graham, J., Sun, X. & Riley, T. 2001 The citrus canker epidemic in Florida: the scientific basis of regulatory eradication policy for an invasive species. *Phytopathology* **91**, 30–34. (doi:10.1094/PHYTO.2001.91.1.30)
26 Rizzo, D., Garbelotto, M., Davidson, J., Slaughter, G. & Koike, S. 2002 *Phytophthora ramorum* as the cause of extensive mortality of *Quercus spp.* and *Lithocarpus densiflorus* in California. *Plant Dis.* **86**, 205–214. (doi:UNSPD-2002-0115-01R)
27 Haydon, D., Kao, R. & Kitching, R. 2004 The UK foot-and-mouth disease outbreak—the aftermath. *Nat. Rev. Microbiol.* **2**, 675–681. (doi:10.1038/nrmicro960)
28 Kitching, R., Hutber, A. & Thrusfield, M. 2005 A review of foot-and-mouth disease with special consideration for the clinical and epidemiological factors relevant to predictive modelling of the disease. *Vet. J.* **169**, 197–209. (doi:10.1016/j.tvjl.2004.06.001)
29 Kitching, R., Thrusfield, M. & Taylor, N. 2006 Use and abuse of mathematical models: an illustration from the 2001 foot-and-mouth disease epidemic in the United Kingdom. *Rev. Sci. Tech. Office Int. Epizooties* **25**, 293–311.
30 Sander, B., Nizam, A., Garrison, L. P., Postma, M. J., Halloran, M. E. & Longini, I. M. 2009 Economic evaluation of influenza pandemic mitigation strategies in the United States using a stochastic microsimulation transmission model. *Value Health* **12**, 226–233. (doi:10.1111/j.1524-4733.2008.00437.x)
31 Rohani, P., Earn, D. & Grenfell, B. 2000 Impact of immunisation on pertussis transmission in England and Wales. *Lancet* **355**, 285–286. (doi:10.1016/S0140-6736(99)04482-7)
32 Keeling, M., Woolhouse, M., May, R., Davies, G. & Grenfell, B. 2003 Modelling vaccination strategies against foot-and-mouth disease. *Nature* **421**, 136–142. (doi:10.1038/nature01343)
33 Keeling, M., Tildesley, M., Savill, N., Woolhouse, M., Shaw, D., Deardon, R., Brooks, S. & Grenfell, B. 2006 Fmd control strategies—comment. *Vet. Record* **158**, 707–708. (doi:10.1136/vr.158.20.707)
34 Tildesley, M. J., Savill, N. J., Shaw, D. J., Deardon, R., Brooks, S. P., Woolhouse, M. E. J., Grenfell, B. T. & Keeling, M. J. 2007 Vaccination strategies for foot-and-mouth disease—reply. *Nature* **6**, E12–E13. (doi:10.1038/nature05605)
35 Gravenor, M., Papasozomenos, P., McLean, A. & Neophytou, G. 2004 A scrapie epidemic in Cyprus. *Epidemiol. Infect.* **132**, 751–760. (doi:10.1017/S0950268804002110)
36 Sethi, S. P. 1978 Optimal quarantine programmes for controlling an epidemic spread. *J. Oper. Res. Soc.* **29**, 265–268. (doi:10.1057/jors.1978.55)
37 Moore, C. & Newman, M. 2000 Epidemics and percolation in small-world networks. *Phys. Rev. E* **61**, 5678–5682. (doi:10.1103/PhysRevE.61.5678)
38 Varian, H. R. 2006 *Intermediate microeconomics*, 7th edn. New York, NY: Norton.
39 Meltzer, M., Cox, N. & Fukuda, K. 1999 The economic impact of pandemic influenza in the United States: priorities for intervention. *Emerg. Infect. Dis.* **5**, 659. (doi:10.3201/eid0505.990507)
40 Weycker, D., Edelsberg, J., Halloran, M. E., Longini, I. M., Nizam, A., Ciuryla, V. & Oster, G. 2005 Population-wide benefits of routine vaccination of children against influenza. *Vaccine* **23**, 1284–1293. (doi:10.1016/j.vaccine.2004.08.044)
41 Beach, R. H., Poulos, C. & Pattanayak, S. K. 2007 Farm economics of bird flu. *Can. J. Agric. Econ. Rev. Can. D Agroecon.* **55**, 471–483. (doi:10.1111/j.1744-7976.2007.00103.x)
42 Medlock, J. & Galvani, A. P. 2009 Optimizing influenza vaccine distribution. *Science* **325**, 1705–1708. (doi:10.1126/science.1175570)
43 Grais, R. F., De Radiguès, X., Dubray, C., Fermon, F. & Guerin, P. J. 2006 Exploring the time to intervene with a reactive mass vaccination campaign in measles epidemics. *Epidemiol. Infect.* **134**, 845–849. (doi:10.1017/S0950268805005716)
44 Legrand, J., Grais, R. F., Boelle, P. Y., Valleron, A. J. & Flahault, A. 2007 Understanding the dynamics of Ebola epidemics. *Epidemiol. Infect.* **86**, 610–621. (doi:10.1017/S0950268806007217)
45 Tildesley, M. J., Bessell, P. R., Keeling, M. J. & Woolhouse, M. E. J. 2009 The role of pre-emptive

Downloaded from rsif.royalsocietypublishing.org on June 9, 2011

culling in the control of foot-and-mouth disease. *Proc. R. Soc. B* **276**, 3239–3248. (doi:10.1098/rspb.2009.0427)

46 Behlau, F., Belasquejr, J., Bergaminfilho, A., Graham, J., Leitejr, R. & Gottwald, T. 2008 Copper sprays and windbreaks for control of citrus canker on young orange trees in southern Brazil. *Crop Prot.* **27**, 807–813. (doi:10.1016/j.cropro.2007.11.008)

47 Bilcke, J. & Beutels, P. 2009 Reviewing the cost effectiveness of rotavirus vaccination: the importance of uncertainty in the choice of data sources. *Pharmacoeconomics* **27**, 281–297. (doi:10.2165/00019053-200927040-00002)

48 Luyten, J. & Beutels, P. 2009 Costing infectious disease outbreaks for economic evaluation: a review for hepatitis A. *Pharmacoeconomics* **27**, 379–89. (doi:10.2165/00019053-200927050-00003)

49 Crawley, M. J. 2007 *The R book*. New York, NY: Wiley. (doi:10.1002/9780470515075)

50 Schoch-Spana, M. 2000 Implications of pandemic influenza for bioterrorism response. *Clin. Infect. Dis.* **31**, 1409–1413. (doi:10.1086/317493)

51 Caetano, M. A. L. & Yoneyama, T. 2001 Optimal and sub-optimal control in Dengue epidemics. *Optim. Control Appl. Methods* **22**, 63–73. (doi:10.1002/oca.683)

OPEN ACCESS Freely available online

PLoS one

Understanding Disease Control: Influence of Epidemiological and Economic Factors

Katarzyna Oleś[1,2]*, Ewa Gudowska-Nowak[1], Adam Kleczkowski[2]

1 M. Kac Complex Systems Research Center and M. Smoluchowski Institute of Physics, Jagiellonian University, Kraków, Poland, **2** Department of Computing Science and Mathematics, University of Stirling, Stirling, United Kingdom

Abstract

We present a model of disease transmission on a regular and small world network and compare different control options. Comparison is based on a total cost of epidemic, including cost of palliative treatment of ill individuals and preventive cost aimed at vaccination or culling of susceptible individuals. Disease is characterized by pre-symptomatic phase, which makes detection and control difficult. Three general strategies emerge: global preventive treatment, local treatment within a neighborhood of certain size and only palliative treatment with no prevention. While the choice between the strategies depends on a relative cost of palliative and preventive treatment, the details of the local strategy and, in particular, the size of the optimal treatment neighborhood depend on the epidemiological factors. The required extent of prevention is proportional to the size of the infection neighborhood, but depends on time till detection and time till treatment in a non-nonlinear (power) law. The optimal size of control neighborhood is also highly sensitive to the relative cost, particularly for inefficient detection and control application. These results have important consequences for design of prevention strategies aiming at emerging diseases for which parameters are not nessecerly known in advance.

Citation: Oleś K, Gudowska-Nowak E, Kleczkowski A (2012) Understanding Disease Control: Influence of Epidemiological and Economic Factors. PLoS ONE 7(5): e36026. doi:10.1371/journal.pone.0036026

Editor: Yamir Moreno, University of Zaragoza, Spain

Received October 3, 2011; **Accepted** March 26, 2012; **Published** May 9, 2012

Copyright: © 2012 Oleś et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: Project operated within the Foundation for Polish Science International Ph.D. Projects Programme co-financed by the European Regional Development Fund covering, under the agreement no. MPD/2009/6, the Jagiellonian University International Ph.D. Studies in Physics of Complex Systems. The project was carried out within the MPD programme "Physics of Complex Systems" of the Foundation for Polish Science and cofinanced by the European Development Fund in the framework of the "Innovative Economy Programme" and within the framework of "Exploring the Physics of Small Devices (EPSD)" of European Science Foundation (ESF). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: kas@cs.stir.ac.uk

Introduction

The network-based approaches are a common tool in epidemiological studies [1]. These individual-based methodologies allow incorporating the diverse patterns of interaction that underlie disease transmission and have been proved to capture topology of populations [2,3]. An interesting aspect of such studies, with an obvious goal to target spread of the disease, is identification of optimal strategies for the control of a disease under additional constraints [4–6]. Network modelling has been successfully used for many systems in order to design such control strategies [7]. However, there are only very few attempts to incorporate economic factors in such realistic models. Conversely, bioeconomic models usually ignore the spatial components of the disease spread [8–10].

In this paper we present a combined epidemiological and economic model to address the problem of optimization of disease control on networks with incomplete knowledge. Two main sources of costs can be associated with a disease outbreak and its control: the palliative cost associated with disease case and costs of measures aimed at preventing further cases [11,12]. The objective of preventive actions is to lower the total cost by investing e.g. in vaccination at the initial stages of the epidemic or culling of infected/susceptible individuals.

In our approach, we define a measure of the total cost of the epidemic (the severity index, X) and analyze the influence of the parameters on its minimum. Work so far has shown that it is possible in such models to find an optimal control strategy [12]. Three optimal control scenarios (Global Strategy (GS), Local Strategy (LS), Null Strategy (NS)) emerge from the cost-effectiveness analysis. However, the relationship between the details of the Local Strategy and the model parameters is still elusive [7,12]. Establishing such a relationship is an essential step in designing control strategies for emerging diseases and hence we have concentrated on this task in the paper. We investigate propagation of the disease in a small-world network. The basic topology represents a regular lattice, with additional long-range bonds between randomly chosen pairs of sites. Inclusion of shortcuts into a regular lattice enhances communication of the disease and causes proliferation of epidemics at locations far apart from the original infected source.

Our principal objective is to identify optimal strategies for eradication of the disease by determining the threshold size of the control neighborhood. In the proposed model, the neighborhood order z is introduced as a measure of either the distance that the disease can spread (epidemic neighborhood), or the spatial extension of the control measures in a single "event" (control neighborhood). To investigate how limited resources should be balanced between disease detection and eradication, we analyze combined effects of the average time until detection and the treatment rate on optimal control size of the neighborhood.

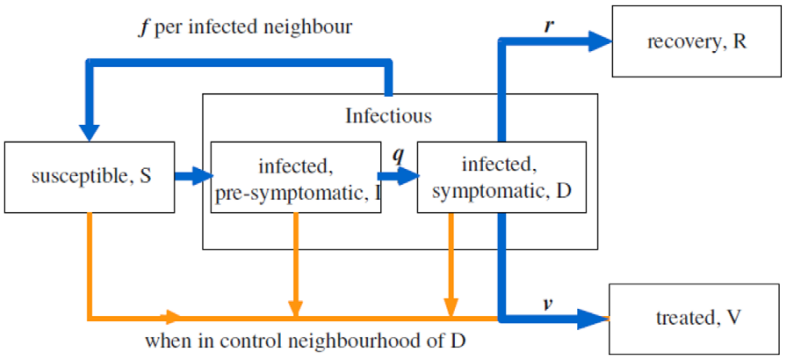

Figure 1. Block diagram illustrating transitions in the model: transitions performed at each time step (blue solid lines) and transitions triggered by treatment (orange thin lines).
doi:10.1371/journal.pone.0036026.g001

We have found that the scale of control matches the scale of dispersal of a disease and so the larger the infection neighborhood, the further the control has to be extended. This relationship can be approximated by a linear function which coefficients depend algebraically on the detection and treatment rates following a power law. Small change in the relative cost of preventive to palliative treatment may result in big changes in this relationship. Addition of small world links narrows the range where the scaling (power) law is valid but the scaling persists for small values of detection and treatment times.

Methods

Model

We assume that individuals are located at nodes of a regular (square) lattice that represents geographical distribution of hosts. On this lattice, we define a local neighborhood of order z as a von Neumann neighborhood in which we include z shells and $\phi(z)=2z(z+1)$ individuals, excluding the central one. Accordingly, $z=0$ corresponds to a single individual, which means that this individual is not in contact with anyone, $z=1$ corresponds to 4 nearest neighbors while $z=\infty$ corresponds to the whole population in the limit of infinite size of the system. For the small world model a fixed number of long range links has been added to the regular network described above. Those links span the whole population, but otherwise behave like local links.

The epidemiological model is a standard SIR (Susceptible-Infected-Removed) model [13], modified to include pre-symptomatic and symptomatic stages of the disease and to account for detection and treatment (cf. fig. 1). All individuals are initially susceptible (**S**) and the epidemic is initiated by introduction of several infected (**I**), pre-symptomatic individuals. Each of infected individuals (symptomatic and pre-symptomatic) stays in contact with a given (fixed) number of other individuals in its infection neighborhood of order z_{inf}. After infection, the susceptible individual moves first to infected, pre-symptomatic class, (**I**) compartments. It can further infect its neighbors with probability f per a contact but cannot be treated yet. As symptoms develop with probability q, individual moves to **D** class and can be detected. It is still infectious but can spontaneously recover with probability r and accordingly, move to a recovery class, (**R**) and cannot be further infected or treated.

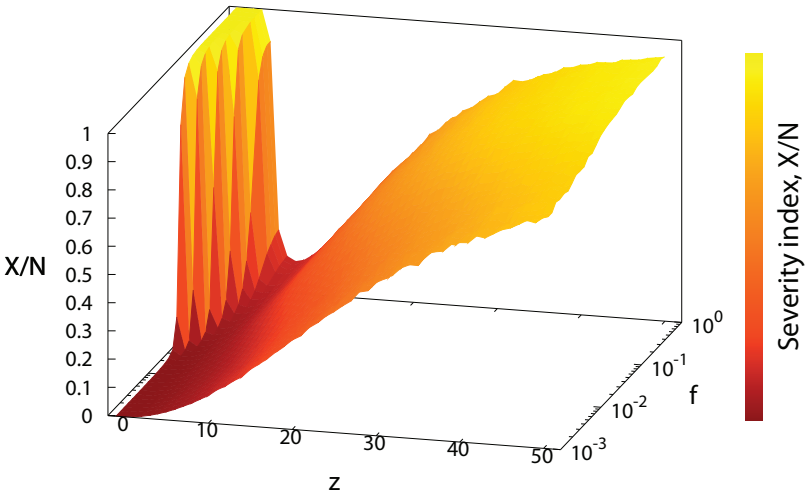

Figure 2. Severity index, X, as a function of the infection rate per contact f and the control neighborhood size z. Simulation parameters: $q=0.5$, $v,r=0.1$ with 40 initial foci and infected neighborhood size set to $z_{inf}=1$, cost $c=1$.
doi:10.1371/journal.pone.0036026.g002

Detection triggers the control process which becomes activated with probability v. In consequence, all individuals (except **R**) within control neighborhood of size z centered at the detected host, transfer to the treated class **V**. The order of control neighborhood z may be different from the order of infectious neighborhood z_{inf} and is typically larger. Accordingly, the group of individuals subject to the treatment is composed of at least one symptomatic and a mixture of susceptible and infected pre-symptomatic individuals. For convenience, we extend the definition of the neighborhood z to capture situations when no spatial control is applied ($z=-1$), or when the treatment is applied solely to the detected individual ($z=0$).

Numbers of individuals in each class are denoted by S, I, D, R and V, respectively with $N=S+I+D+R+V$ being the total constant number of individuals in the population.

In order to investigate the optimal control strategy, we need to compare value of future benefits (reduction of infection cases) with the value of future and current costs associated with a particular choice of measures in disease control and treatment. In this paper we allocate the costs to two groups:

$$X(z,t=\infty)=R(z,t=\infty)+cV(z,t=\infty). \quad (1)$$

The first term represents the palliative cost and is associated with individuals who are not treated and therefore spontaneously move into the **R** class. The second term describes costs associated with treatment of detected individuals and their neighbors and is assumed to be proportional to the number of treated individuals V. In the above formula c represents a cost of treatment relative to the cost of infection and z stands for the control neighborhood size.

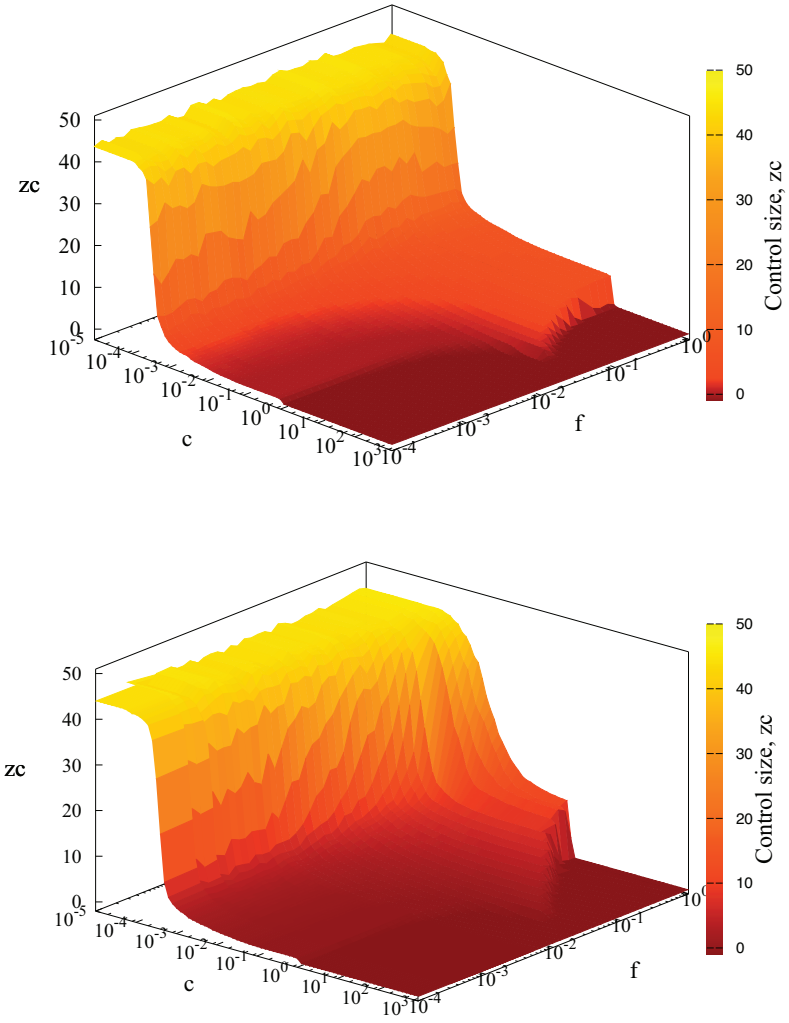

Figure 3. Control neighborhood size as a function of treatment cost *c* and infectiousness of the disease *f* for regular network and small world model. Simulation parameters: $q=0.5$, $v,r=0.1$, with 40 initial foci and $z_{inf}=1$. Control size $z_c>0$ represents local strategy (LS), $z_c=0$ corresponds to the strategy when only the detected individual is treated and $z_c\geq 30$ denotes GS (more than 99% of individuals are treated). Null strategy corresponds to $z_c=-1$. Top figure denotes results for disease spreading on regular networks, whereas bottom to small world model with inclusion of additional 2000 number of long range links (5%).
doi:10.1371/journal.pone.0036026.g003

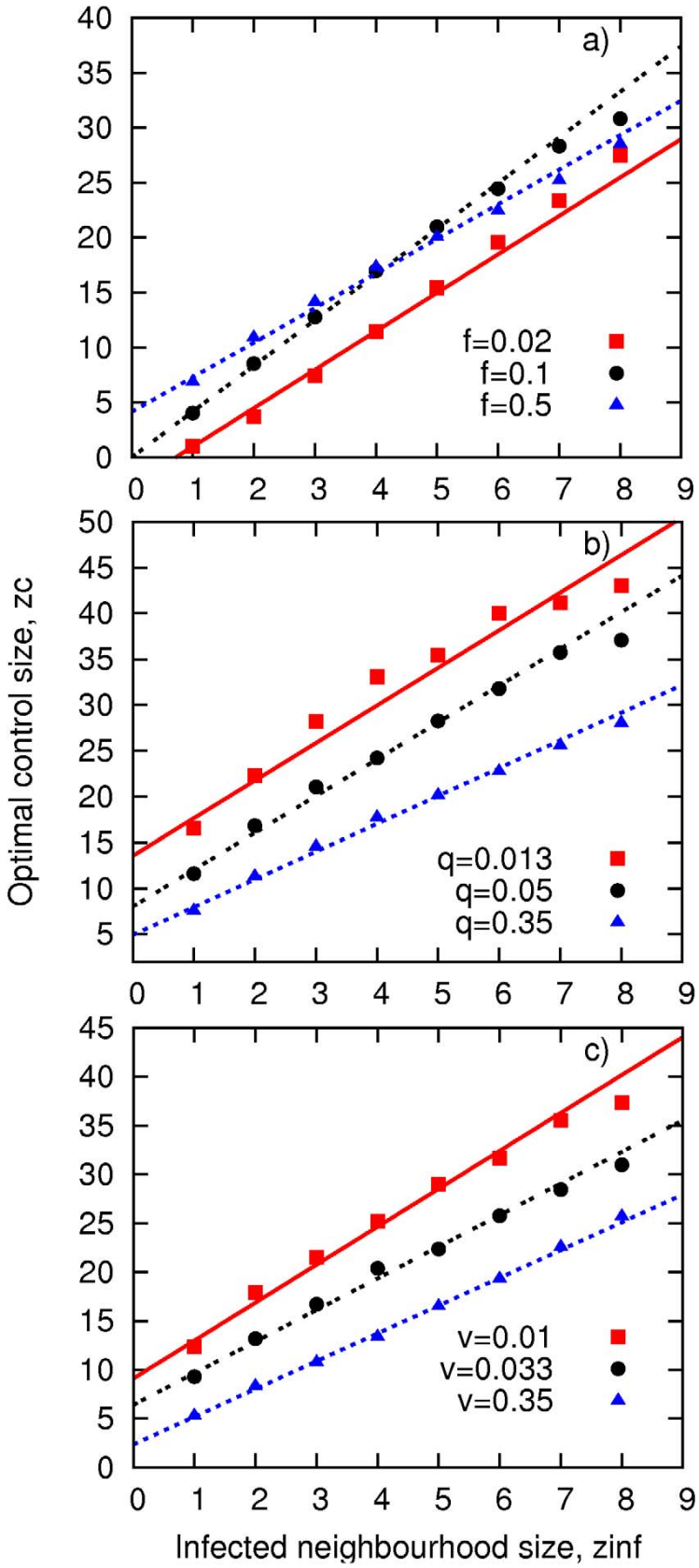
a)
f=0.02
f=0.1
f=0.5
b)
q=0.013
q=0.05
q=0.35
c)
v=0.01
v=0.033
v=0.35
Optimal control size, zc
Infected neighbourhood size, zinf

Figure 4. Relationship between z_c and z_{inf} for treatment cost $c=1$. Points mark the simulation results whereas lines correspond to fitted linear function $z_c = z_{inf} * a + b$. From top to bottom, the following sets of constant kinetic parameters have been assumed: (a) $q=0.5, v=0.1$, (b) $v=0.1, f=1$, (c) $q=0.5, f=1$. Errors (standard deviation from the mean) are too small to be visible.
doi:10.1371/journal.pone.0036026.g004

Both estimates of R and V are evaluated at the end of each simulation run ($t \to \infty$).

Simulations

Monte Carlo simulations have been performed on a regular grid of 200 by 200 cells with periodic boundary conditions with and without long-range links. This choice of size has been dictated by a trade off between numerical efficiency and avoidance of small-size effects which could influence results. Additional numerical tests proved the consistency of results for different system sizes [12].

Epidemics have been initiated by addition of 40 infected individuals to an otherwise susceptible population. Each simulation run has been continued until $I(t)+D(t)=0$ (i.e. up to the time when no further infection can occur). Subsequently the severity index X has been evaluated from the formula eq.(1). The optimal strategy is then determined by the minimal value of the severity index X_c. The corresponding value of z gives the optimal size of the control neighborhood, z_c (see fig. 2 for illustration). In the simulations, the minimization of the severity index is achieved by sweeping through different values of the control neighborhood size z, while keeping other parameters fixed. For each value of z only a single simulation has been performed. Collections of this results yield a dependence of X on z. A minimum value of X in this collection gives an estimate of X_c and the corresponding z gives an estimate of z_c. This procedure has been repeated 100 times to yield representative average values of z_c and X_c and their corresponding standard deviations.

Results

The long time ($t \to \infty$) behavior of the model in the absence of control (Null Strategy, NS, i.e. $z=-1$) is determined by the probability f of passing the infection to a susceptible node from any of its neighbors within the neighborhood size ranging from 4 ($z=1$) to 144 ($z=8$). For small f, the infection quickly dies out. Disease spreads invasively over the population for large f, when no control is applied, $X(z,\infty) \propto R(z,\infty) \simeq N$. When $z \geq 1$, the ratio R/N declines with the order of the control neighborhood. However, at the same time the number of treated individuals V increases, contributing to the total cost X, cf. eq.(1). For $c \neq 0$, $X(z)$ is either a monotonic function of z for small values of f or a non-monotonic function for highly contagious disease (large f), see fig. 2.

Three regions can be identified in the dependence of z_c on c and f, see fig. 3. For small values of c, Global Strategy (GS) is dominating, whereas for large c, it is best to refrain from treatment, Null Strategy (NS), fig. 3.

Although the location of the minimum of $X(z)$ varies with increasing f and c values (see figs. 2, 3), a relatively wide plateau region with an almost constant z_c develops for intermediate values of c and f and corresponds to the local strategy (LS), fig.3. The structure in fig. 3 is partially deformed by addition of long-range links, however, the plateaux persists for small values of f.

We have therefore focused on the plateaux region (LS) of z_c and have explored its dependence on epidemiological parameters: z_{inf}, q, v, with constant f and c. We have first explored dependence of z_c on the size of infection neighborhood for $c=1$, see fig. 4. The relationship can be accurately approximated by a linear function for a wide range of parameters, infectiousness f (fig.4a), the rate at which symptoms appear, q (fig.4b) and the treatment rate, v (fig.4c) for $z_{inf} \in [1,8]$.

As already seen in fig. 3, infectiousness f hardly affects the slope and intercept of the linear relationship, fig.4a. Increasing q and v causes the lines to shift towards lower values of z_c, with major changes in the intercept but slope only slightly affected (cf. fig.4b,c). In contrast, the relationship between z_c and q (or v) for fixed z_{inf} is non-linear. It is more convenient to consider $1/q$ instead of q as $\tau_q = 1/q$ has an interpretation of average time till detection of symptoms. Similarly, $\tau_v = 1/v$ can be interpreted as an average time till treatment.

Broadly speaking, z_c increases with τ_q and τ_v, fig. 5. This is consistent with the following mechanism. Consider a single infected but pre-symptomatic individual. The disease focus centered on it will spread until appearance of symptoms after time τ_q. Thus, the longer it takes to discover symptoms of the disease, the farther the disease would spread from its original focus. As a consequence, the infected area becomes larger and so does z_c. Similarly, the longer time from detection until treatment, the further the disease moves away from original focus. As a result, the control size grows with increasing treatment time.

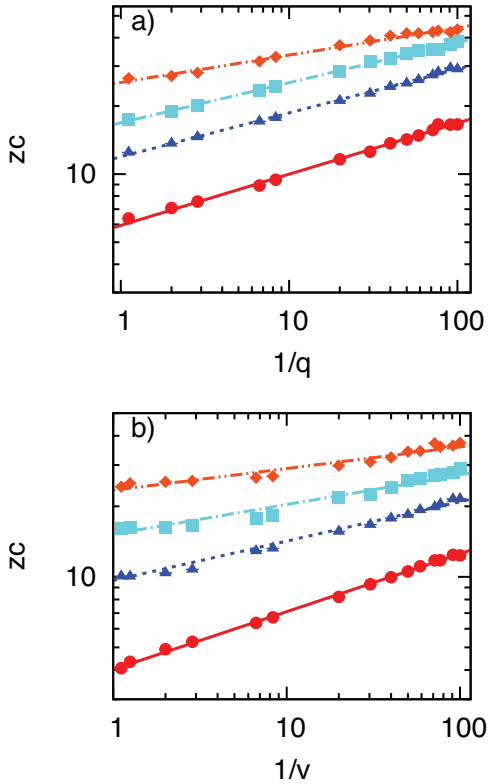

Figure 5. Relationship between z_c and τ_q in a) and τ_v in b). Points mark the simulation results and lines correspond to fitted functions: a): $z_c(\tau_q) = \alpha_q \tau_q^{\beta}$ and b): $z_c(\tau_v) = \alpha_v \tau_v^{\beta'}$ for red: $z_{inf}=1$, navy blue: $z_{inf}=3$, blue: $z_{inf}=5$, orange: $z_{inf}=8$.
doi:10.1371/journal.pone.0036026.g005

Intriguingly, it appears that z_c scales algebraically with τ_q (and with τ_v) following a power law: $z_c = \alpha_q \tau_q^{\beta}$ and $z_c = \alpha_v \tau_v^{\beta'}$ eq.(3) (see fig. 5) with exponents well below 1.

The exponents β, β' are similar for a range of z_{inf} within the plateaux regime of an optimal control radius of the epidemic, (see fig.3), i.e. for $z_{int} \in [1,8]$, $\beta \in [0.14,0.25]$ and $\beta' \in [0.10,0.27]$..

While fig. 5 is representative of results for $c \leq 1$, moving c just beyond $c = 1$ causes a dramatic change in the $z_c(\tau_v)$ dependence for large values of τ_q and τ_v, corresponding to detection and vaccination time comparable with duration of epidemics (approximately 10^4 time steps for large values of τ_v and τ_q). The control neighborhood z_c decays abruptly for increasing times τ_q, τ_v, as illustrated in fig. 6. This change is associated with very inefficient control (long time till detection, $\tau_q \gg 1$ and long time from detection to treatment, $\tau_v \gg 1$). If the cost of control is lower or equal to the cost of palliative care, it is still better to treat, even though we are not very efficient with treatment and most individuals are spontaneously removed. However, if the cost of vaccination is only marginally higher than the cost of untreated case, prevention is no longer cost-effective. We also note that it is only a combination of very long values of τ_q *and* τ_v that leads to a limited range of application of the scaling formulas ($z_c = \alpha_q \tau_q^{\beta}$ and $z_c = \alpha_v \tau_v^{\beta'}$).

The scaling region of z_c as a function of τ_q and τ_c also depends on c in a fashion reminiscent of fig. 3. For small values of c, Global Strategy of treating everybody is optimal regardless of the parameters, cf. fig. 3 with fig. 7. In contrast, Null Strategy is optimal for large c (figs. 3 and 7). The region where Local Strategy is optimal occupies the region near $c = 1$, but it becomes narrower when the disease is more infectious (fig. 3) or when the control is less efficient (for increasing values of τ_q (fig. 7a) and τ_v (fig. 7b). Within this region, z_c is given by scaling formulas. As seen before,

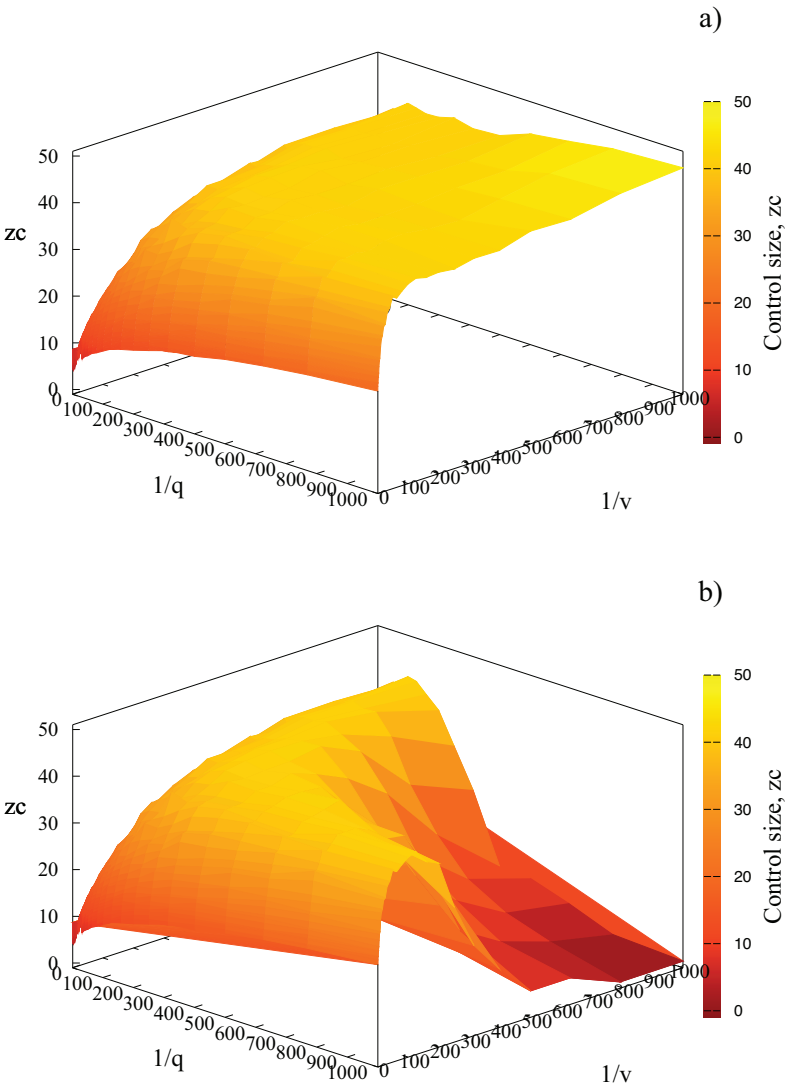

Figure 6. Control neighborhood size as a function of both detection time, τ_q, and recovery time, τ_v for $c = 1$ in a) and $c = 1.001$ in b). Simulation parameters: $f = 0.1$, $r = 0.1$, $z_{inf} = 1$, 40 initial foci.
doi:10.1371/journal.pone.0036026.g006

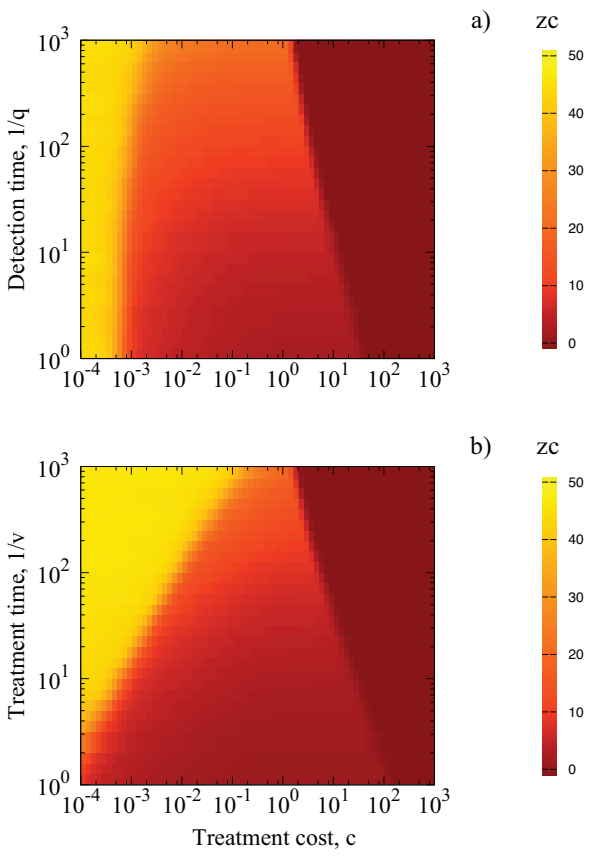

Figure 7. Control neighborhood size as a function of treatment cost c and detection time τ_q (a) and treatment time τ_v (b). Simulation parameters: $f=1$, $q=0.5$,v,$r=0.1$, with 40 initial foci and $z_{inf}=1$. Color borderlines between different regions indicate transition regions among various optimal strategies.
doi:10.1371/journal.pone.0036026.g007

$c=1$ is a special case asymptotically associated with a breakdown of LS for very large or very small f (fig. 3) and very large values of τ_q and τ_v (fig. 7).

The addition of long range links shitfs the optimal radius of control towards larger values, figs. 3, 8. The scaling behaviour (cf. fig.5) is characteristic for a regular network and changes when long-range bonds is added (see fig.8). With 400 random long-range contacts (corresponding to 1% of all links) the scaling relation between z_c and τ_q (τ_v) breaks down for detection (treatment) times exceeding 10. This is clearly indicated by deviation of the results from red bottom line (in fig.8) denoting simulation data for regular networks (the same as in fig. 5). Altogether, addition of small world links reduces the range of detection τ_q and treatment τ_v times for which the power law relationship is valid. This is caused by long range links allowing disease to escape from the local control. In contrast, if we are able to detect disease quicker, it has not much chance to escape and the disease spread is effectively short range. Consequently, the scaling can be observed for small values of detection and treatment times, τ_q,τ_v. In summary, with increasing degree of randomness of networks (larger number of links) not only the control radius rises but also the scaling disappears. Note that the dashed black line, $z_c=40$ in fig. 8, represents Global Strategy.

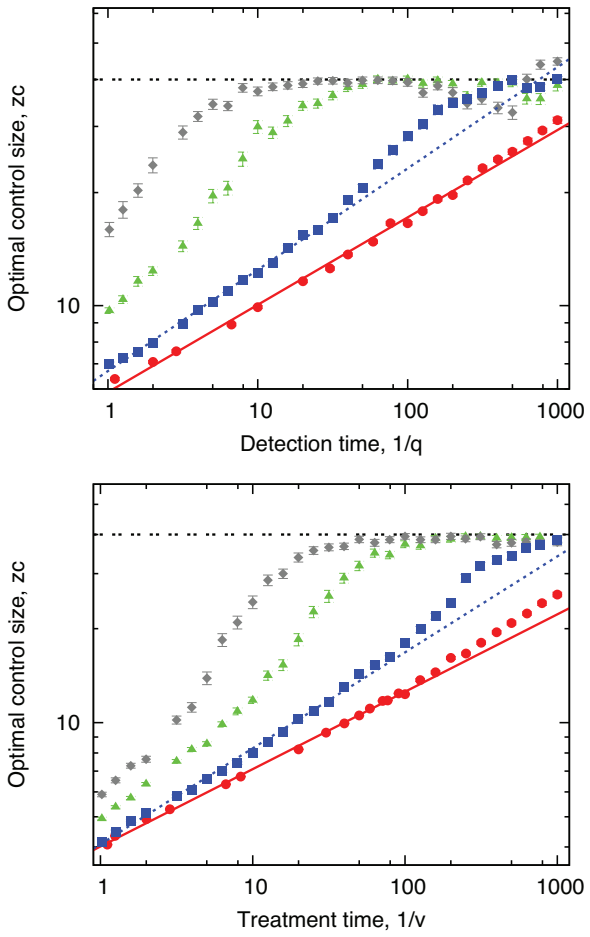

Figure 8. Relationship between z_c and τ_q on top and τ_v on bottom for regular and small world network with varying number of long-range links. Lines and simulation points from bottom: red: regular network, navy blue: small world with additional 1% of links, green: small world with additional 5% of links, grey: small world with additional 10% of links, dashed black: $z_c=40$ which corresponds to treating the whole population. Simulation parameters: $f=1$, v,$r=0.1$, $z_{inf}=1$, 40 initial foci, 1%,5%,10% of long-range links.
doi:10.1371/journal.pone.0036026.g008

Discussion

In order to design a successful strategy for controlling a disease we need to take into account not only epidemiological and social factors (including the topology of the social network of contacts and in particular z_{inf}), but also economic considerations. Some of these factors might be unknown or hard to estimate, particularly in real time as the epidemic unfolds. It is therefore crucial to understand the relationship between the optimal control strategy and parameters, for a wide range of possible values. It is even more important to establish those processes and parameters to which a selection of optimal strategy is not particularly sensitive, as this allows us to find strategies that can be designed in advance, even without knowing their actual values for a given emerging disease.

Regular networks have been traditionally used for modelling epidemic outbreaks of human, animal and plant diseases [14,15] and many variants of such an approach (with e.g. constant or

randomized probabilities of infection passed to neighbouring nodes on a grid) have been studied. However, an accumulated experimental evidence demonstrates that real systems rarely follow this kind of idealization being neither completely random nor located on regular lattices. Among other types of networks that have been the object of intense studies are the small-world and scale-free networks. In particular, the small world network with randomly chosen shortcuts between the nodes, is considered a model well extrapolating between extremes like regular and random network. It has been also preferentially used by modellers discribing outbreaks of disease starting simultaneously in different regions of the world (propagation of the SARS virus, [16]. Accordingly, in order to assess the occasional long distance dispersal of the disease, we have also considered small world links, representing e.g. random transport by wind or by plane.

In our previous paper we have shown that for a given set of z_{inf}, q and v, the broad choice of the strategy is determined by the relative cost of the treatment, c. For small values of c, GS is optimal, for large values of c, NS. Close to $c=1$, a LS dominates and the detailed value of the control neighborhood z_c depends on the epidemiological parameters, although not on f in a wide range. In this paper we extend this analysis to include other epidemiological parameters. In particular we show that the broad division between GS (for $c \ll 1$), NS (for $c \gg 1$) and LS (for $c \simeq 1$) holds for a wide range of parameters q and v (inverse of time to detection and inverse of time to treatment, respectively), fig. 7.

Three other key results emerge from our analysis. Firstly, it is very important to match scale of control to the scale of infection dispersal. This has already been seen in other papers [17], but this is the first time we show it for spatial control on networks in the presence of economic evaluation. However, we also show that the size of the control neighborhood is not just simply equal to the size of the infection neighborhood (see fig. 4 and compare the scale of horizontal and vertical axes). In the presence of pre-symptomatic individuals ($\tau_q \gg 0$) and in the face of delays associated with application of control ($\tau_v \gg 0$) we need to extend z_c well beyond z_{inf}. The relationship between z_{inf} and z_c is one of the key formulas for planning response to epidemics. It enables authorities to plan actions aiming at eradication of the disease by setting a sufficiently large – but not too large – zone of eradication around each detected case. Traditionally, such recommendations are based on the dispersal patterns of the disease, although increasingly simulation models are used. This procedure has led to establishment of the 1,900ft rule for citrus canker [18] whereby all citrus trees are cut down within this radius from every affected tree and the 3 km/10 km rule for foot-and-mouth disease [19].

However, our results show that the relationship between z_c and z_{inf} is non-trivial and in particular it involves non-linear functions of τ_q and τ_v. Although we are still far from being able to provide a formula relating z_c to all epidemiological parameters, our result stresses importance of using models to design control strategies [20].

We also show that $c=1$ is a special case. In particular, we show high sensitivity of z_c to changes in c for large values of τ_q and τ_v. Thus, if the symptom detection time (τ_q) and reaction time (τ_v) are both long, small change in c leads to very big changes in z_c, see fig. 6 and 7. Without knowing the exact value of c it is therefore very difficult to design the strategy in this case. Suppose we believe that $c>1$ and therefore we chose a small value of z_c based upon fig. 6b. However, if in reality $c \leq 1$ (although very close to 1), z_c should be close to 50 (fig. 6a). This shows the importance of knowing what the actual value of c is [12] estimated that for vaccination $c=0.01$–0.85, but can be larger than 1 for culling.

In this paper we have used regular and small wold networks to describe the topology of interaction between individuals. Addition of small world links into population narrows the range where the scaling (power law) relationship of z_c on τ_q and τ_v is valid but the scaling persists for small values of detection and treatment times.

Our studies can also be extended in other ways. The current work assumes relatively short overall time length of each epidemic and so no discounting is applied when the costs and benefits are estimated. We also assumed that the strategy is unchanged throughout the epidemic and that the network structure is static and relatively simple. Each of these assumptions can be relaxed. Discounting is often used in economics, but we expect for it to have a small impact on our results. Adapting the strategy to the current status of the epidemic often leads to a bang-bang solution [21], similar to our distinction between NS and GS.

Finally, a lot of attention have been recently given to non-local and random networks (small-world or scale-free networks) [12,22], to dynamic networks [23], and networks with random parameters [24]. Further extension of this work to include static and dynamic disorder is in progress.

Acknowledgments

We are very grateful to Bartek Dybiec for useful discussions.

Author Contributions

Conceived and designed the experiments: AK KO. Performed the experiments: KO. Analyzed the data: KO EGN AK. Contributed reagents/materials/analysis tools: KO EGN AK. Wrote the paper: KO EGN AK.

References

1. Newman M (2010) Networks: an introduction. Oxford Univ Pr.
2. Keeling M (2005) Models of foot-and-mouth disease. Proceedings of the Royal Society B: Biological Sciences 272: 1195.
3. Gastner M, Newman M (2006) Optimal design of spatial distribution networks. Physical Review E 74: 016117.
4. Barrett S (2003) Global disease eradication. Journal of the European Economic Association 1: 591–600.
5. Rowthorn R, Laxminarayan R, Gilligan C (2009) Optimal control of epidemics in metapopulations. Journal of the Royal Society Interface 6: 1135.
6. Ndeffo Mbah M, Gilligan C (2010) Optimization of control strategies for epidemics in heterogeneous populations with symmetric and asymmetric transmission. Journal of Theoretical Biology 262: 757–763.
7. Ferguson N, Donnelly C, Anderson R (2001) The foot-and-mouth epidemic in great britain: pattern of spread and impact of interventions. Science 292: 1155.
8. Klein E, Laxminarayan R, Smith D, Gilligan C (2007) Economic incentives and mathematical models of disease. Environment and Development Economics 12: 707–732.
9. Gersovitz M, Hammer J (2004) The Economical Control of Infectious Diseases*. The Economic Journal 114: 1–27.
10. Boccara N, Cheong K, Oram M (1994) A probabilistic automata network epidemic model with births and deaths exhibiting cyclic behaviour. Journal of Physics A: Mathematical and General 27: 1585.
11. Kleczkowski A, Dybiec B, Gilligan C (2006) Economic and social factors in designing disease control strategies for epidemics on networks. Arxiv preprint physics/ 0608141.
12. Kleczkowski A, Oleś K, Gudowska-Nowak E, Gilligan C (2011) Searching for the most cost-effective strategy for controlling epidemics spreading on regular and small-world networks. Journal of The Royal Society Interface.
13. Anderson R, May R (1991) Infectious diseases of humans: dynamics and control. Wiley Online Library.
14. Jeger M, Pautasso M, Holdenrieder O, MW S (2007) Modelling disease spread and control in networks: implications for plant sciences. New Phytologist 174: 279–297.
15. Shirley M, Rushton S (2005) The impacts of network topology on disease spread. Ecological Complexity 2: 287–299.
16. Small M, Tse C, Walker D (2006) Super spreaders and the rate of transmission of the SARS virus. Physica D 215: 146–158.

17. Gilligan CA, Truscott JE, Stacey AJ (2007) Impact of scale on the effectiveness of disease control strategies for epidemics with cryptic infection in a dynamical landscape: an example for a crop disease. Journal of the Royal Society Interface 4: 925–934.
18. Gottwald TR, Hughes G, Graham JH, Sun X, Riley T (2001) The citrus canker epidemic in Florida: the scientific basis of regulatory eradication policy for an invasive species. Phytopathology 91: 30–4.
19. DEFRA website. Available: http://animalhealth.defra.gov.uk/managing-disease/notifiable-disease/footand-mouth.html. Accessed 2012 Apr, 3.
20. Kao RR (2002) The role of mathematical modelling in the control of the 2001 FMD epidemic in the UK. Trends In Microbiology 10: 279–286.
21. Forster GA, Gilligan CA (2007) Optimizing the control of disease infestations at the landscape scale. Proceedings of the National Academy of Sciences of the United States of America 104: 4984–4989.
22. Dybiec B, Kleczkowski A, Gilligan CA (2004) Controlling disease spread on networks with incomplete knowledge. Physical Review E 70.
23. Vernon MC, Keeling MJ (2009) Representing the UK's cattle herd as static and dynamic networks. Proceedings of the Royal Society B-Biological Sciences 276: 469–76.
24. Taraskin SN, Ludlam JJ, Neugebauer CJ, Gilligan CA (2005) Extinction of epidemics in lattice models with quenched disorder. Physical Review E 72.

OPEN ACCESS Freely available online

PLOS ONE

Efficient Control of Epidemics Spreading on Networks: Balance between Treatment and Recovery

Katarzyna Oleś[1,2]*, Ewa Gudowska-Nowak[1], Adam Kleczkowski[2]

1 M. Kac Complex Systems Research Center and M. Smoluchowski Institute of Physics, Jagiellonian University, Kraków, Poland, **2** Department of Computing Science and Mathematics, University of Stirling, Stirling, United Kingdom

Abstract

We analyse two models describing disease transmission and control on regular and small-world networks. We use simulations to find a control strategy that minimizes the total cost of an outbreak, thus balancing the costs of disease against that of the preventive treatment. The models are similar in their epidemiological part, but differ in how the removed/recovered individuals are treated. The differences in models affect choice of the strategy only for very cheap treatment and slow spreading disease. However for the combinations of parameters that are important from the epidemiological perspective (high infectiousness and expensive treatment) the models give similar results. Moreover, even where the choice of the strategy is different, the total cost spent on controlling the epidemic is very similar for both models.

Citation: Oleś K, Gudowska-Nowak E, Kleczkowski A (2013) Efficient Control of Epidemics Spreading on Networks: Balance between Treatment and Recovery. PLoS ONE 8(6): e63813. doi:10.1371/journal.pone.0063813

Editor: Jesus Gomez-Gardenes, Universidad de Zarazoga, Spain

Received November 29, 2012; **Accepted** April 5, 2013; **Published** June 4, 2013

Copyright: © 2013 Oleś et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: Project operated within the Foundation for Polish Science International Ph.D. Projects Programme co-financed by the European Regional Development Fund covering, under the agreement no. MPD/2009/6, the Jagiellonian University International Ph.D. Studies in Physics of Complex Systems and was co-financed by the European Development Fund in the framework of the "Innovative Economy Programme" and within the framework of "Exploring the Physics of Small Devices (EPSD)" of European Science Foundation (ESF). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: kas@cs.stir.ac.uk

Introduction

Networks can provide a good representation of how individuals interact [1–3]. Despite many simplifications, models based upon network structures have successfully been used in many applications [4,5] including spread of rumours and news [3] and computer viruses [1]. A particularly important application of network models has been in epidemiology [6–10] of plant, animal and human pathogens [11–13]. Modelling in epidemiology plays an important role: It allows us to estimate the scale of the epidemic, to predict how far the disease could spread and to design effective ways of control. All these tasks need to be achieved despite the fact that in many cases we are not able to observe the whole process and/or measure all relevant parameters [14]. The state of individuals, whether they are susceptible, infected and pre-symptomatic, infected and symptomatic or recovered, is in particular often difficult to ascertain [15]. Despite these uncertainties it is possible to use modelling to design effective control measures leading to the lowest overall cost of the epidemic outbreak [16–19] and a number of studies have used network models to address this issue [14,20–23].

Economic and behavioural aspects influence the spread of disease and affect the choice of a control strategy. For instance, if the treatment does not cost anything, the best strategy is to control the whole population. Contrarily, for very expensive control measures it might be better to refrain from treatment at all. Optimisation of total disease costs, including palliative cost associated with disease cases and cost of appropriate control measures, leads to appearance of three basic strategies [20]: The Global Strategy (GS) whereby all individuals are treated regardless of their status can be contrasted with the Null Strategy (NS) when the public authorities completely refrain from preventive treatment and concentrate on palliative treatment of cases. The Local Strategy (LS) emerges for intermediate costs of treatment. In this case, not only detected symptomatic individuals are treated preventively, but the treatment includes also their neighbours.

The work so far has concentrated on the role of processes associated with disease spread on the broad choice of the treatment strategy [20] and on the details of the local strategy [21]. However, the spontaneous recovery also may affect the results and in the current paper we explore this dependence in detail.

We extend our results to two contrasting and yet complementary models in which we either treat individuals that have been through the disease or not. Whether the removed individuals (i.e. those who have been through the disease but then spontaneously recover or die) are part of the treatment plan depends on the type of the disease agent. The key factor in choosing the right model is whether it is possible – and desirable – to distinguish such individuals from those who are susceptible. If the removed class is identified with dead individuals, the distinction is very clear. However, if the removal means recovery and immunity, it might not be possible to identify those who are immune. For example, many people might not want to report that they have been through the infection, or the disease symptoms might be relatively mild. For animal diseases, immunological testing might be the only way to identify such individuals, but this leads to increased costs and test results might not be reliable. In other situations, we might know the status of the individual, but might not be able to target the treatment to susceptible and infected individuals. Plant and

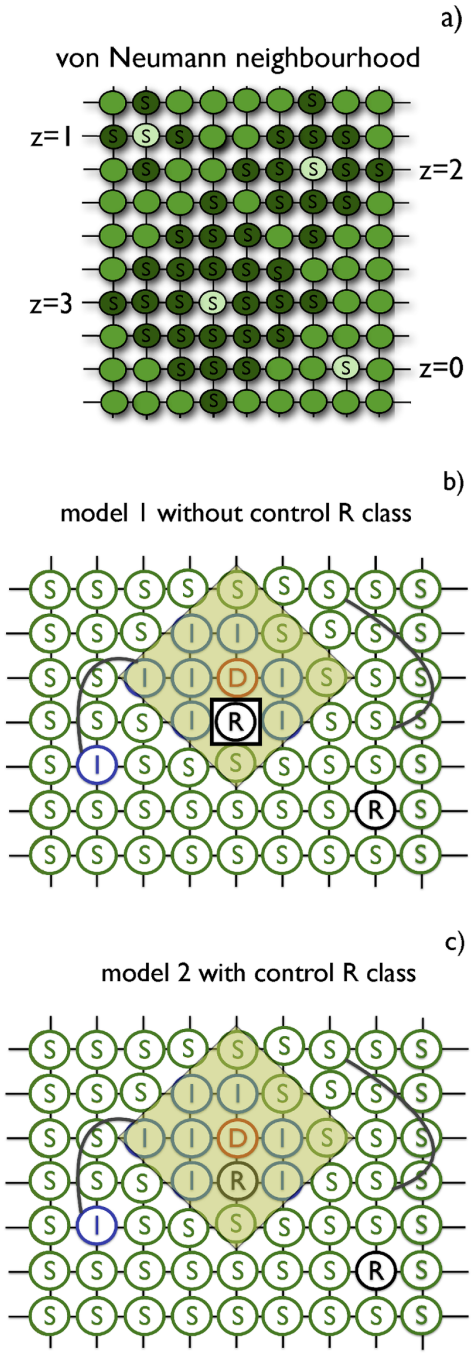

Figure 1. (a) Definition of the von Neumann neighborhood of different values of order z, as used in the simulations and analysis. (b) Illustration of spread of a disease (model 1) on a regular network with additional randomly chosen long-range links represented by curved lines (approximation of a small-world network). The applied control of radius z is centered on node **D** (yellow shaded area). Note that in model 1 the **R** individuals are excluded from the control and thus non-treated. (c) Representation of model 2: All individuals contained in the control neighbourhood of order z are preventively treated and moved to **V** class. In both models treatment does not take into account individuals connected by additional long-range links. S, I, D, R symbols stand for Susceptible, Pre-symptomatic, Symptomatic and Recovered, respectively. The order z of infection neighbourhood equals $z_{inf}=2$ in (b) and (c).
doi:10.1371/journal.pone.0063813.g001

crop diseases might serve as an example here, whereby it might be easier to treat the whole field regardless of whether some plants there are already immune to the disease.

Although such individuals do not contribute to the spread of the disease, the cost of treating them affects the economic side of the evaluation and therefore leads to changes in the design of the optimal strategy. We study this case in our paper and show that although there is a difference in the choice of the strategy (LS vs. GS) and the resulting number of treated individuals, there is only a small difference in the overall total cost of the epidemic.

Methods

We assume that individuals are located at nodes of a square lattice that represents geographical distribution of hosts, see fig. 1. On this lattice, we define a local infection neighbourhood of order z_{inf} as a von Neumann neighbourhood. In that neighbourhood $2z_{inf}(z_{inf}+1)+1$ individuals are included, involving the central one. We additionally define $z=0$ as corresponding to this central individual, which means that this individual is not in contact with anyone, while $z=\infty$ corresponds to the whole population, see fig. 1. To increase realism of our analysis, we also consider the small-world model [24,25] which adds a certain number of links among randomly chosen nodes, thus adding some long-range connections to the regular lattice ones [24]. Although the disease can spread along these long-range links, we assume that they are so difficult to identify that they are not included in any treatment strategy (see below).

The epidemiological SIDRV model is a standard SIR (Susceptible-Infected-Removed) model [26], modified to account for latent period and preventive and responsive treatment (fig. 2), see also [21]. Taking into consideration the latent period, the infectious class is now composed of two separate, pre-symptomatic and symptomatic classes (**S**, **I**, **D**, **R** and **V**, respectively). Number of individuals in each class is denoted by S, I, D, R, and V, respectively, and $N=S+I+D+R+V$ is the total constant number of individuals in the population.

Initially, all individuals are assumed to be susceptible (**S**). The epidemic is initiated by an introduction of few infected but pre-symptomatic (**I**) individuals, which are located randomly and uniformly over the whole network. Each infected individual is in contact with a fixed number of other individuals in its infection neighbourhood z_{inf}. These connections do not change during the epidemic. The disease is transmitted along these contact routes with probability f per contact. Upon a successful infection, the susceptible individual moves to the pre-symptomatic class.

Each infected pre-symptomatic individual moves to a symptomatic class (**D**) with probability q. Detected individuals still can infect other individuals. Subsequently, each detected individual can spontaneously move to a removed class (**R**) with probability r. However, detection also triggers a control event with probability v and subsequently a number of individuals selected from the von Neumann neighbourhood of order z centered at the detected individual move to a treated class (**V**); for details see below. Neither **R** nor **V** can infect or be re-infected any more.

According to the responsive treatment two versions of the SIDRV model have considered: (i) model 1 with control of all individuals in selected area except removed (**R** class), see fig. 1b,

and (ii) model 2 with control of all individuals in selected area regardless of their status (and thus including **R**), see fig. 1c.

The control event is localized within a von Neumann neighbourhood of order z centred on a symptomatic individual. The order of control neighbourhood, z, can be different than the order of the infection neighbourhood, z_{inf}, and is typically found larger. Thus, a group of individuals in the treatment neighbourhood consists of a mixture of susceptible, infected pre-symptomatic, infected symptomatic and recovered individuals (preventive treatment). We have extended the definition of control neighborhood size in order to include the situation when no control is applied, $z=-1$.

Simulations

All simulations have been performed on the lattice of 200 by 200 individuals with periodic boundary conditions. Simulations started with 40 initial infected foci, which corresponds to 0.1% of the total population.

Control size, z, has been varied, while other parameters (such as f, q, v, r, z_{inf},) have been kept constant. Each simulation has been run until $I(t)+D(t)=0$, which means that no infection can occur afterwards. At the end of the run all R and V individuals have been counted, yielding information about severity of the epidemic as well as effectiveness of the treatment involved.

Effectiveness of control strategies

The effective control strategy is found by taking into account severity of the epidemic and its financial implications. In order to quantify the effectiveness of different control strategies we introduce the severity index, X [15,20]. By seeking the minimum values of X, we find which strategy is optimal.

The severity index, X, includes two terms corresponding to the cost of infection and control. First term describes costs associated with death, absence in work, lower productivity etc., whereas second term includes costs of vaccine, quarantine, transport of drugs to infection foci, etc. We assume that X is a linear combination of number of individuals which have gone through disease and recovered (**R**) and treated individuals (**V**).

We measure X in units of a number of single infected individuals, so that:

$$X(z,t=\infty)=R(z,t=\infty)+cV(z,t=\infty) \quad . \tag{1}$$

Here c represents a cost of treatment relative to the cost of infection and z stands for the control neighbourhood size. Both $R(z,t=\infty)$ and $V(z,t=\infty)$ are counted at the end of a single simulation run.

Effective strategy is equivalent to the minimal value of X, which means that the epidemic is stopped at the manageable cost. In our simulation, the minimization of the severity index has been achieved by sweeping through different values of control neighbourhood size, z while keeping other parameters constant. Once z is set, we let the system evolve and then compute the value of X in the stationary state. We repeat this operation 100 times and then we denote with z_c and X_c the average values, of z and X, corresponding to the minimum of X, so that

$$\min_{-1\le z\le 50} X(z,t=\infty)=X_c(z_c,t=\infty) \quad . \tag{2}$$

Results

In the absence of control, the disease will either progress through the population until it exhausts a large part of initially susceptible population (for large values of the infection probability f) or it will quickly stop spreading (for small values of f). As control is applied in extended neighbourhood of radius z_c centred at a symptomatic individual, the number of recovered (**R**) individuals declines rapidly, see fig. 3a. Models 1 and 2 examined in this work differ in the way they treat or not treat the recovered class, **R**, cf fig. 1 We observe the same behaviour for both considered models (with and without treating **R** class). However, when we allow the control of **R** individuals (model 2), the proportion of recovered declines faster than in model 2, see fig. 3a (insert). The proportion of preventively treated individuals, **V**, in both models is similar for the whole range of control size, z. With increasing control neighbourhood, V(z) grows very quickly, then drops near $z=6$ and finally rises monotonically till $z\sim 50$ (fig. 3b). Combination of these two relationships, $R(z)$ and $V(z)$, according to eq(1), gives total

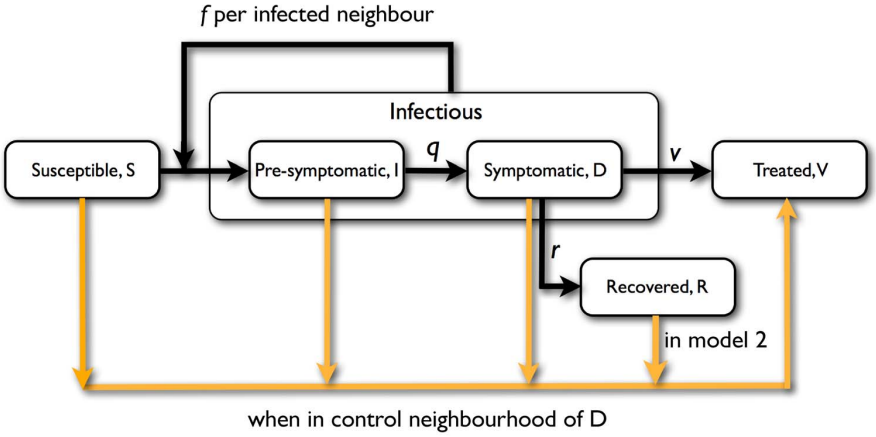

Figure 2. Model scheme of disease transition (black lines) and control (orange lines). In model 2 there is a possible transition between recovered (R) and treated (V) class when R-individual is in the control neighbourhood of any symptomatic D-individual.
doi:10.1371/journal.pone.0063813.g002

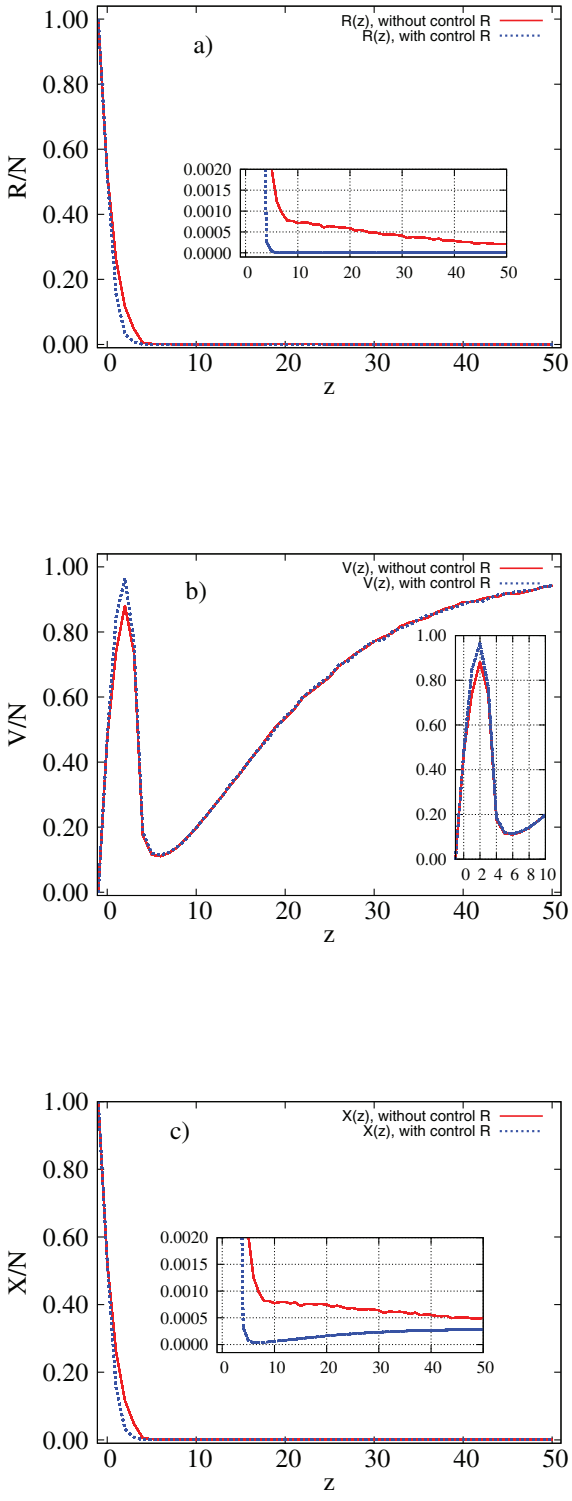

Figure 3. (a) The proportion of recovered individuals, R/N, (b) the fraction of treated (controlled) individuals, V/N and (c) the total cost of epidemic as a fraction of the system size, X/N, for $c=0.0003$ and various control sizes z. Red solid line: model 1; blue dotted line: model 2. Results of simulations with parameters $f=0.5$, $q=0.5$, $r=0.1$, $v=0.1$ and $z_{inf}=1$ performed on regular networks. Inserts show the relevant magnifications of the graph.
doi:10.1371/journal.pone.0063813.g003

cost of epidemic, X, as a function of z, see fig. 3c. For a very low treatment cost, e.g. $c=0.0003$, total cost of control of epidemic, X, is almost equal for both models, with difference less than 0.1%, see fig. 3c (insert). The choice of optimal strategies is different for model 1 (GS) than for model 2 (LS), although the corresponding X values are similar. In model 1 the minimal value of X corresponds to the highest value of control size, $z_c=50$ (GS), whereas in model 2, the minimum is identified with $z_c=6$, (LS) fig. 3c.

Regular networks – influence of recovery rate, r on control strategies

Increasing cost of treatment, c, decreases the optimal control neighbourhood, z_c. For very cheap control the optimal scenario is identified with $z_c\sim 45$ (GS) for model 1, regardless of the recovery rate, r (fig. 4a). The more expensive the treatment, the higher the total costs spent on controlling outbreaks. This leads to change in optimal strategy, see fig. 4a, b. We cannot afford the preventive control of the whole population (GS) and have to shift into treating in neighbourhood of symptomatic individuals. We observe that z_c rapidly decreases with increasing costs, especially for model 1. For intermediate values of c, z_c drops to ~ 10 depending on recovery rate, r. Higher recovery rate, r, results not only in a shorter plateaux for LS (see fig. 4a, b) but also moves the plateaux towards larger control size, z_c. As treatment becomes more expensive, second threshold is observed that describes change from LS to NS. Although for model 2 the global strategy is selected rather than the local one as for model 1 (fig. 4b, d) for the high values of recovery rate, r and low c, the total cost of epidemic, X_c, does not differ much between the two models, see fig. 5. The highest costs are associated with fast spreading diseases (large f) and expensive treatment (large c) for both models (upper right part of plots in fig. 5). Slow spreading disease does not significantly affect the budget for control regardless of treatment costs (lower part of plots in fig. 5) and model selected. For model 2 the global strategy is predominantly selected for high values of recovery rate r and at low c, in contrast to model 1 (fig. 4b, d) where the local strategy prevails. Despite these differences, the total cost of epidemic, X_c, does not differ between the two models, see fig. 5.

Regular networks – control strategies

Control size, z_c depends strongly on the cost of treatment, c, and on the infectiousness of the disease, f (fig. 6). For small f and c, both models suggest preventive control extended to the whole population (GS) (lower left part of each plot in fig. 6). In case of highly infectious disease and low treatment costs, model 1 predicts higher effectiveness of GS whereas model 2 selects LS as an optimal solution, upper left part of each plot in fig. 6. However, in both examined models the total cost of epidemic, X, is approximately the same, see fig. 3. As treatment cost, c, increases, LS becomes the most cost-effective strategy. LS changes to NS when c is of order 1 for small f and of order 10 for high f, regardless of the choice of the model or the exact value of r, compare fig. 6a, b with fig. 6c, d.

The main difference in selection of the optimal strategy occurs for small c. Changes in r affect only low c regions. Increasing r from 0.1 to 0.2 extends the region of validity of GS and moves it towards marginally larger values of c and high values of f, fig. 6c, d. This trend is continued for larger values of r, see fig. 4, and can

be associated with faster removal of individuals without triggering control events.

Small world networks – control strategies

Addition of small-world links does not change the behaviour for small f and c. However, there are substantial differences for large f and the effect differs for the two models. Introducing disorder into the topology by adding long-range links changes ranges of optimal strategy for both considered models, compare fig. 6a, b with fig. 7. In model 1 small number of links, e.g.6%, fig. 7a, extends GS when disease spreads fast and costs are higher. The small number of links 6% in model 2 does not change choice of control strategy, compare fig. 6b with fig. 7b, as in model 1 (top panel in fig. 7). Nonetheless, the total cost of epidemic remains almost the same. For large values of f, destroying spatial structure by adding 20% links results in only two effective strategies for highly infectious disease, GS for $c<1$ and NS otherwise, fig. 7c. The higher disorder (20% of long range links) in model 2, introduces GS when probability of spreading the epidemic, f, increases, fig. 7d.

Discussion

The goal in designing cost-effective control strategy is to stop the epidemic outbreak very quickly at a minimal possible cost. In order to achieve this by using the local strategy (LS) we need to catch in the preventive control neighbourhood as many infected but pre-symptomatic individuals and to form a fire-break by treating around the infection focus. The extend of control is a crucial factor; however, it is not obvious by how much we need to enlarge the neighborhood in which preventive treatment is applied. We need to balance epidemiological and economic aspects of disease spread and control [27]. When we extend prevention to the whole population we might be able to successfully protect population from epidemic outbreaks but we will need to spend a lot of resources. On the other hand, when we apply control to too small neighbourhood, we will spend a lot but the disease will still invade the whole population. Under some conditions an optimal solution emerges in between these two extremes and can be associated with the Local Strategy; in other cases the extreme solutions (Global Strategy and Null Strategy) are optimal. As we have already shown [20,21], the effective control neighbourhood can be chosen based on combined epidemiological and economic analysis.

The previous analyses [20,21] left three key questions unanswered. Firstly, should we treat individuals that are already immune? Although the answer clearly depends on the nature of the disease and the treatment, some general principles can be established. This depends on the relative – economic, social and medical – cost of the preventive treatment compared to the palliative care (when we just let the disease to run its natural course). Secondly, are our results stable with respect to structural changes of the model? We illustrate the stability by considering two versions of the same model, with and without treating recovered **R** individuals. Finally, it is the dependence of the results on the actual recovery rate, r. In real-life applications it is difficult to distinguish between individuals that have been through the disease and those who do not. It is therefore very important to check whether the model and the resulting policy implications are robust with respect to the potential uncertainties. We show that this is the case in general but also identify the region of the parameters when the two models have different behavior (small c, large f).

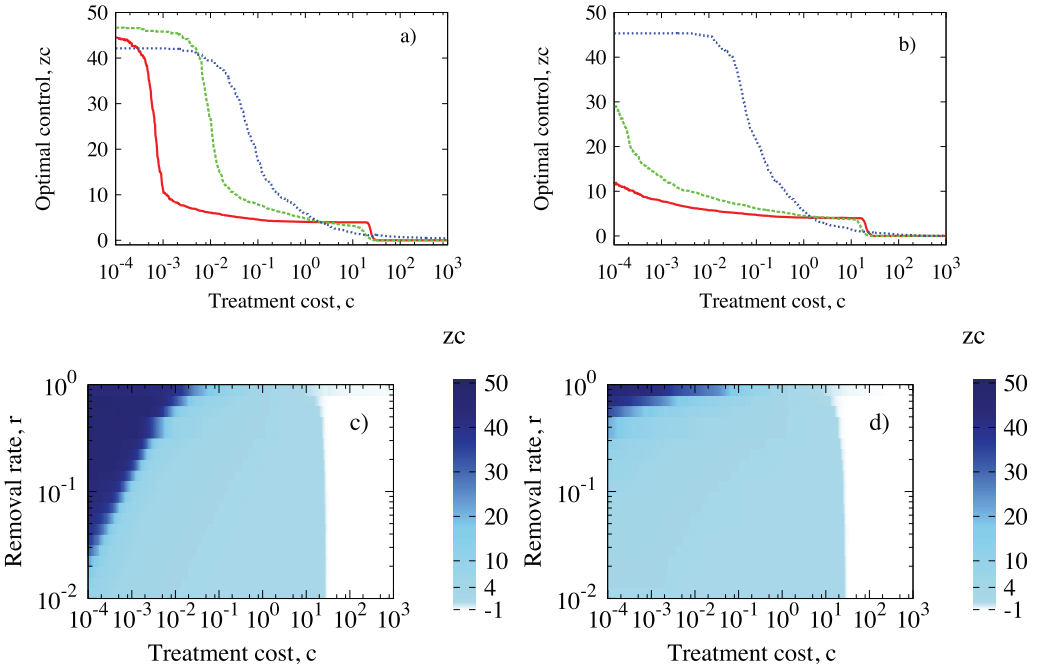

Figure 4. Control size z_c as a function of the treatment cost c ((a) and (b)) and as a function of the recovery rate, r, and the treatment cost, c ((c) and (d)) for model 1 (left column) and model 2 (right column). In (a) and (b) $r=0.10$ (red line), $r=0.63$ (green dashed line), $r=0.98$ (blue dotted line). All simulations done on regular networks with parameters $f=0.1$, $q=0.5$, $v=0.1$, $z_{inf}=1$.
doi:10.1371/journal.pone.0063813.g004

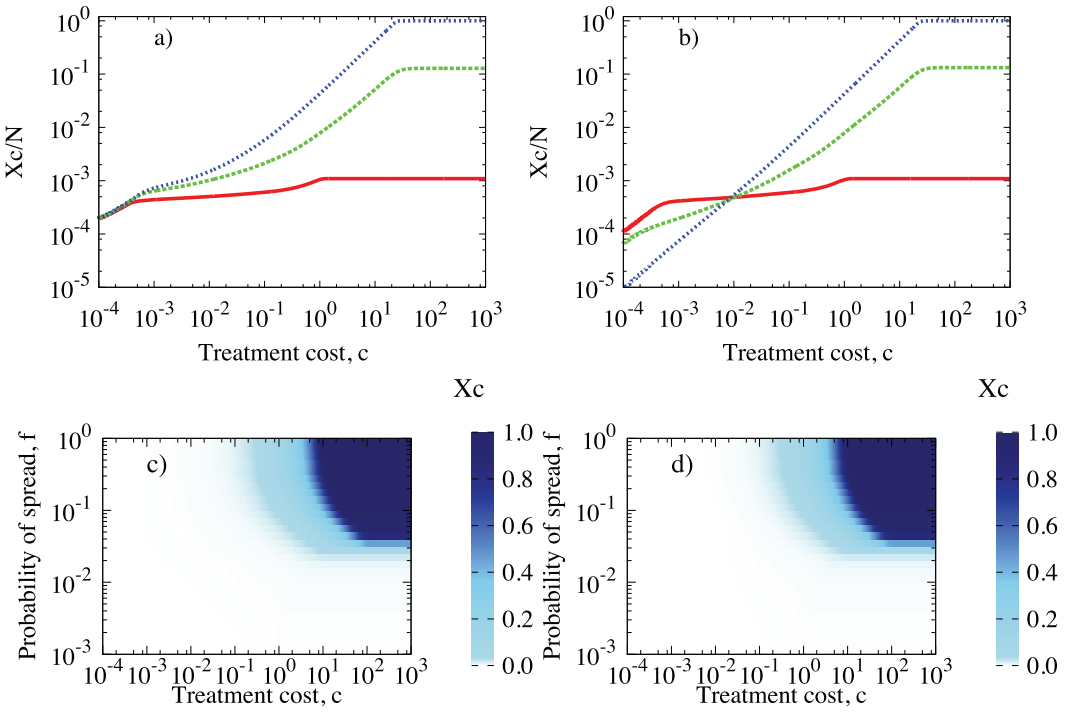

Figure 5. Total cost of epidemic at optimum, X_c, as a function of the treatment cost c ((a) and (b)) and as a function of both infectiousnes, f, and cost, c ((c) and (d)) for model 1 (left column) and model 2 (right column). In (a) and (b) $f=0.001$ (red line), $f=0.032$ (green dashed line), $f=0.1$ (blue dotted line). All simulations done with parameters $q=0.5$, $v=0.1$, $r=0.1$, $z_{inf}=1$. Disease spreading on regular networks.
doi:10.1371/journal.pone.0063813.g005

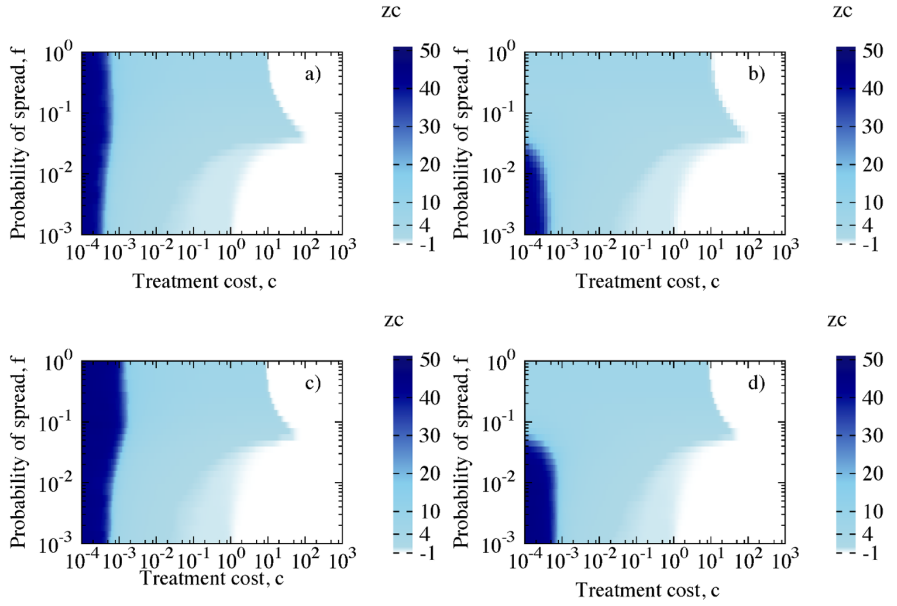

Figure 6. Control size, z_c, as a function of both infectiousness, f, and treatment cost, c, for model 1 (left column) and model 2 (right column). Simulation parameters for top panel ((a) and (b)): $r=0.1$; for bottom panel ((c) and (d)): $r=0.2$; other parameters: $q=0.5$, $v=0.1$, $I(0)=40$, $z_{inf}=1$. Disease spreading on regular networks.
doi:10.1371/journal.pone.0063813.g006

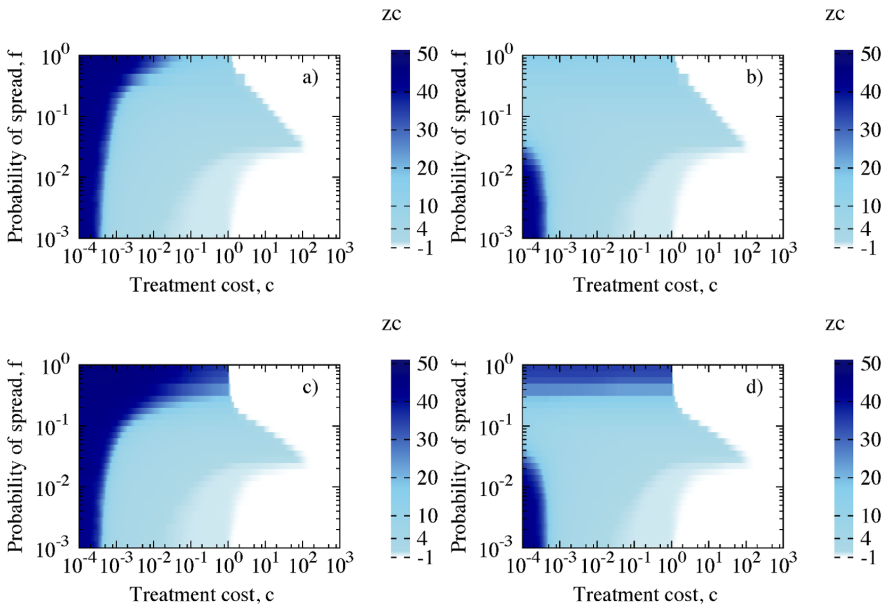

Figure 7. Control size, z_c as a function of both infectiousness, f, and treatment cost, c, for model 1 (left column) and model 2 (right column). Number of additional links with respect to all the ones that can be added, on the top panel ((a) and (b)): 6%, on the bottom panel ((c) and (d)) –20%. Other simulation parameters: $q=0.5$, $r=0.1$, $v=0.1$, $z_{inf}=1$. Disease spreading on small-world networks.
doi:10.1371/journal.pone.0063813.g007

Two contrasting cases can be distinguished in answer to the first question. If the treatment is costly and/or may lead to complications, the authorities might want to invest in testing individuals in order to find out who is and who is not naturally immune. This would identify individuals in the **R** class who then might not be offered the treatment. Contrariwise, if it is not immediately obvious what the actual status of the individual is and testing is expensive, lengthy or unreliable, the authorities might decide to treat all individuals regardless of their status. Our results from this paper suggest that the choice of the strategy depends on whether treatment includes or excludes **R** but the total budget spent on controlling epidemic remains similar for both models.

Secondly, in the most important region of parameter space, corresponding to expensive preventive treatment and a highly infectious disease, both models yield very similar scenarios (right part of fig. 4c, d). Thus, the results appear to be stable with respect to structural changes of the model. Where the difference is marked, for low c and high f, the models suggest a different choice of strategy (GS for model 1 and LS for model 2). However, we also found that in this case the economic outcome of either GS or LS is very similar (see fig. 3c).

Thirdly, the main effect of increasing r is to shift the boundary between the GS and LS for small c, rendering the GS less attractive as r decreases – and the infectious period increases. For model 2 (without treatment of **R**) the area of preference of GS over LS is limited to very small values of c. Thus, the longer the infectious period, the more likely the local strategy is to work. The boundary between LS and NS for large values of c remains unchanged.

Addition of long-range links enlarges the region of applicability of GS towards higher f and c for both models. The large number of randomly placed long-range links destroys spatial structure of spreading the pathogen and causes that it spreads mostly globally so that LS is no longer effective option of control the epidemic.

The results obtained in this paper can be used for those diseases for which spread is dominated by local transmission or by a mixture of local and long-range links. Examples include human (notably SARS [28] and influenza [29–31]), animal (foot-and-mouth disease [32]) and plant diseases (citrus canker [33], sudden oak death [34–36] and rhizomania of sugar beet [37,38]). Although our model assumes a simple network structure, we believe that the results can be generalised to more complex, but also more realistic networks, including social networks [31]. This work can also be extended in several ways. The most interesting will be the SIRS model, in which after some period of immunity to the disease individuals become susceptible again and could catch a disease few times; with influenza [29–31] and sexually-transmitted diseases [39,40] being the best examples.

Author Contributions

Conceived and designed the experiments: KO AK. Performed the experiments: KO. Analyzed the data: KO EGN AK. Contributed reagents/materials/analysis tools: KO EGN AK. Wrote the paper: KO EGN AK.

References

1. Newman M (2003) The structure and function of complex networks. SIAM Review 45: 167–256.
2. Keeling MJ, Rohani P (2007) Modeling Infectious Diseases in Humans and Animals. Princeton University Press, 408 p.
3. Newman M (2010) Networks: An introduction. Oxford Univ Press.
4. Aleksiejuk A, Holyst J, Stauffer D (2002) Ferromagnetic phase transition in Barabási-Albert networks. Physica A: Statistical Mechanics and its Applications 310: 260–266.
5. Bianconi G, Barabási A (2001) Bose-Einstein condensation in complex networks. Physical Review Letters 86: 5632–5635.

6. Andersson H (1999) Epidemic models and social networks. Mathematical Scientist 24: 128–147.
7. Keeling M, Eames K (2005) Networks and epidemic models. Journal of the Royal Society Interface 2: 295–307.
8. Pastor-Satorras R, Vespignani A (2001) Epidemic spreading in scale-free networks. Physical review letters 86: 3200–3203.
9. Moreno Y, Pastor-Satorras R, Vespignani A (2002) Epidemic outbreaks in complex heterogeneous networks. The European Physical Journal B-Condensed Matter and Complex Systems 26: 521–529.
10. Meloni S, Perra N, Arenas A, Gómez S, Moreno Y, et al. (2011) Modeling human mobility responses to the large-scale spreading of infectious diseases. Scientific reports 1.
11. Keeling M (2005) Models of foot-and-mouth disease. Proceedings of the Royal Society B: Biological Sciences 272: 1195.
12. Jeger M, Pautasso M, Holdenrieder O, Shaw M (2007) Modelling disease spread and control in networks: implications for plant sciences. New Phytologist 174: 279–297.
13. Ferguson N, Donnelly C, Anderson R (2001) The foot-and-mouth epidemic in Great Britain: pattern of spread and impact of interventions. Science 292: 1155.
14. Cohen R, Havlin S, Ben-Avraham D (2003) Efficient immunization strategies for computer networks and populations. Physical review letters 91: 247901.
15. Dybiec B, Kleczkowski A, Gilligan C (2004) Controlling disease spread on networks with incomplete knowledge. Physical Review E 70: 066145.
16. Kao R (2002) The role of mathematical modelling in the control of the 2001 FMD epidemic in the UK. Trends In Microbiology 10: 279–286.
17. Forster GA, Gilligan CA (2007) Optimizing the control of disease infestations at the landscape scale. Proceedings of the National Academy of Sciences of the United States of America 104: 4984–4989.
18. Klepac P, Laxminarayan R, Grenfell B (2011) Synthesizing epidemiological and economic optima for control of immunizing infections. Proceedings of the National Academy of Sciences 108: 14366–14370.
19. Hollingsworth TD, Klinkenberg D, Heesterbeek H, Anderson RM (2011) Mitigation Strategies for Pandemic Influenza A: Balancing Conflicting Policy Objectives. PLoS Computational Biology 7: e1001076.
20. Kleczkowski A, Oleś K, Gudowska-Nowak E, Gilligan C (2012) Searching for the most cost-effective strategy for controlling epidemics spreading on regular and small-world networks. Journal of The Royal Society Interface 9: 158–169.
21. Oleś K, Gudowska-Nowak E, Kleczkowski A (2012) Understanding disease control: influence of epidemiological and economic factors. PloS One 7: e36026.
22. Dybiec B, Kleczkowski A, Gilligan C (2005) Optimising control of disease spread on networks. Acta Physica Polonica B 36: 1509–1526.
23. Gomez-Gardenes J, Echenique P, Moreno Y (2006) Immunization of real complex communication networks. The European Physical Journal B-Condensed Matter and Complex Systems 49: 259–264.
24. Watts D (2003) Small worlds: the dynamics of networks between order and randomness. Princeton University Press.
25. Verdasca J, Telo da Gama M, Nunes A, Bernardino N, Pacheco J, et al. (2005) Recurrent epidemics in small world networks. Journal of Theoretical Biology 233: 553–561.
26. KermackWO, McKendrick AG (1927) A contribution to the mathematical theory of epidemics. Proceedings of the Royal Society of London Series A 115: 700–721.
27. Gersovitz M, Hammer J (2004) The Economical Control of Infectious Diseases. The Economic Journal 114: 1–27.
28. Dye C, Gay N (2003) Modeling the SARS epidemic. Science 300: 1884–1885.
29. Ferguson NM, Cummings DAT, Cauchemez S, Fraser C, Riley S, et al. (2005) Strategies for containing an emerging influenza pandemic in Southeast Asia. Nature 437: 209–214.
30. Halloran ME, Ferguson NM, Eubank S, Longini IM, Cummings DaT, et al. (2008) Modeling targeted layered containment of an influenza pandemic in the United States. Proceedings of the National Academy of Sciences of the United States of America 105: 4639–44.
31. Cauchemez S, Bhattarai A, Marchbanks TL, Fagan RP, Ostroff S, et al. (2011) Role of social networks in shaping disease transmission during a community outbreak of 2009 H1N1 pandemic influenza. Proceedings of the National Academy of Sciences of the United States of America 108: 2825–2830.
32. Woolhouse MEJ, Chase-Topping M, Haydon D, Friar J, Matthews L, et al. (2001) Epidemiology – Footand-mouth disease under control in the UK. Nature 411: 258–259.
33. Gottwald T, Hughes G, Graham J, Sun X, Riley T (2001) The citrus canker epidemic in Florida: the scientific basis of regulatory eradication policy for an invasive species. Phytopathology 91: 30–4.
34. Rizzo DM, Garbelotto M, Davidson JM, Slaughter GW, Koike ST (2002) Phytophthora ramorum as the cause of extensive mortality of quercus spp. and lithocarpus densiflorus in California. Plant Disease 86: 205–214.
35. Meentemeyer RK, Cunniffe NJ, Cook AR, Filipe JA, Hunter RD, et al. (2011) Epidemiological modeling of invasion in heterogeneous landscapes: spread of sudden oak death in California (1990-2030). Ecosphere 2.
36. Filipe JA, Cobb RC, Meentemeyer RK, Lee CA, Valachovic YS, et al. (2012) Landscape epidemiology and control of pathogens with cryptic and long-distance dispersal: sudden oak death in northern Californian forests. PLoS computational biology 8: e1002328.
37. Stacey AJ, Truscott JE, Asher MJC, Gilligan CA (2004) A model for the invasion and spread of rhizomania in the United Kingdom: Implications for disease control strategies. Phytopathology 94: 209–215.
38. Gilligan C, Truscott J, Stacey A (2007) Impact of scale on the effectiveness of disease control strategies for epidemics with cryptic infection in a dynamical landscape: an example for a crop disease. Journal of the Royal Society Interface 4: 925–934.
39. Wylie JL, Cabral T, Jolly AM (2005) Identification of networks of sexually transmitted infection: a molecular, geographic, and social network analysis. Journal of Infectious Diseases 191: 899–906.
40. Stoner BP, Whittington W, Hughes JP, Aral SO, Holmes KK (2000) Comparative epidemiology of heterosexual gonococcal and chlamydial networks: implications for transmission patterns. Sexually transmitted diseases 27: 215–223.

COST-BENEFIT ANALYSIS OF EPIDEMICS SPREADING ON CLUSTERED RANDOM NETWORKS

Katarzyna Oleś[a,b], Ewa Gudowska-Nowak[a], Adam Kleczkowski[b]

[a]Mark Kac Complex Systems Research Center
and
The Marian Smoluchowski Institute of Physics, Jagiellonian University
Reymonta 4, 30-059 Kraków, Poland
[b]Department of Computing Science and Mathematics, University of Stirling
Stirling FK9 4LA, United Kingdom

(Received December 23, 2013; revised version received January 10, 2014)

2554

We study, control of infectious disease epidemics spreading on random networks with different levels of clustering. We use Gleeson's *et al.*, *Phys. Rev.* **E80**, 036107 (2009) algorithm to create clustered networks in which a proportion of individuals are located in fully-connected cliques of certain size. A SIR model is extended to include delayed and imperfect detection of infectious individuals. We also include a combination of responsive (palliative) and preventive (vaccination) treatments and design cost-effective disease control strategies. Cost-benefit analysis is used in combination with epidemiological simulations to identify an optimal radius for a treatment centred upon the symptomatic individual. Three general control strategies occur depending on the relative cost of treatment and prevention. Network topology and, in particular, clustering also affects the applicability of the control strategy. The average path length appears to be more important; the range for thecontrol strategy is wider with the length, but the optimal radius of control also extends. As the proportion of individuals in cliques and therefore the coefficient of clustering is higher, the range of the costs for which control scenario is optimal is greater. This results have important consequences for designing disease control strategies that also satisfy economic optimality criteria.

DOI:10.5506/APhysPolB.45.????
PACS numbers: 05.50.+q, 87.18.–h, 07.05.Tp, 02.70.Uu

1. Introduction

The spread of many human [1–3], animal [4, 5] and plant [6, 7] epidemics can successfully be described by network models [8–12]. In this approach, individuals are represented as nodes on a network and their interactions by edges [13–15]. Analytical solutions arising from the graph theory [16, 17]

and percolation [18, 19] or simulations can be used to answer questions concerning the potential for a particular disease to invade the population and persist there [20, 21], the relationship between the network structure and rate of spread [22–24], the future course of an unfolding epidemic [25], and, finally, to assess control strategies that either prevent the disease from invading [26] or aim at its eradication [27–29]. Network models are particularly suitable for the latter task, as they allow to represent spatial aspects of the disease spread [30, 31] and, therefore, help in designing responsive and local control strategies that target particular individuals or their connections [32].

A successful disease control strategy should not only aim to stop the disease from spreading, but should achieve this at the lowest possible overall cost [31, 33–35], including both costs of the treatment as well as of the disease itself. In this approach, an optimal strategy is the one that minimises the total cost of the epidemic [31, 32, 35–37] with monetary as well as social costs included.

However, the task of identifying an optimal strategy is made complicated by a typical lack of information about the status of the individuals and their connectivity to others. We typically do not know whether a particular individual is already infected and infectious, unless symptoms are displayed and can be identified. For many diseases this lack of knowledge can be a serious problem [2, 4] as the disease can spread far before the first symptomatic individual is discovered. This makes responsive and local strategies difficult, as they depend on our ability to identify epidemic foci around which they are applied. Despite this problems, contact tracing [23, 38], "clean ring" strategies [39–41], and similar treatment and vaccination options either are used or are proposed to combat the disease spread. In these approaches, an observation of a symptomatic individual triggers an action which typically affects a number of individuals connected to the observed case. The inclusion of individuals is based upon a typical distance at which the disease can travel unobserved [3, 31, 34], although this relationship is not always clear [32]; this usually means treatment within a certain distance from the focus measured in an appropriate metric [7, 42, 43].

The ability of capturing the network structure is essential for successful epidemiological modelling of the kind studied here [22–24, 30, 37, 44]. For convenience and tractability, many models represent interactions between individuals as a regular network, possibly with addition of "small-world" interactions [10, 45, 46]. Alternatively, random network models including scale-free networks have been used [13, 14, 19, 26, 28]. However, there is a mounting evidence [47] that many real-life networks are not tree-like, but instead possess substantial degree of clustering [48]. Clustering (or transitivity) in a complex networks refers to the tendency of two neighbours of a given node to also be neighbours of each other, thus forming a triangle of

edges within the graph [18, 49]. It has been shown that presence of clustering increases the bond percolation threshold and affects the threshold behaviour of the epidemic spread [50] when networks with the same degree distribution and similar correlation structure are compared.

In this paper, we extend the results of our previous work [31, 32, 35] to more realistic clustered networks. We begin by briefly reviewing the epidemiological model used in our studies. We further apply a recently proposed model of embedded cliques [18, 49] to examine epidemics spreading in clustered random networks. We show that three broad control strategies can be identified, the Global Strategy (GS) whereby the location of treated individuals does not depend on their distance from the focus, the Null Strategy (NS) when it is more cost-effective not to treat anybody, and the Local Strategy (LS) which targets individuals located in the neighbourhood of the detected (symptomatic) individual. The choice of the strategy as well as the details of LS (the size of the treatment "ring") are shown to depend on the level of clustering in the network.

2. Model

Three elements form a description of our model. Firstly, we present the epidemiological scheme describing the progress of the disease in the individual and its spread to other individuals conditioned on a link existing between them. Secondly, we describe the structure of the network with contributing links that provide the potential for the spread of the disease. Finally, we describe the epi-economic framework in which we assess the cost and benefits of the control measures.

2.1. Epidemiological model

Epidemiological model that has been used in this work is an extended SIR (Susceptible-Infected-Removed) model to account of pre-symptomatic and symptomatic stages [31]. Initially, all individuals are susceptible (**S**), except of a fixed small number of infected pre-symptomatic (**I**) individuals (5 in the total population of 5 000), located randomly throughout the population.

Each individual is in contact with a fixed number of neighbours and the disease can be transmitted from/to each of them. Details of the spatial arrangement and size of the neighbourhood are given below. With probability f per single contact with either an infected individual (**I**) or the detected individual (**D**), the disease is passed to a susceptible individual (**S**) that becomes infectious but pre-symptomatic individual (**I**). Subsequently, the infected individual displays symptoms and the transition to a symptomatic state (**D**) occurs with probability q.

A symptomatic individual is assumed to be still infectious, but can spontaneously become removed (**R**) with probability r and cease to pass on infection. Alternatively, it can also trigger a control event, with probability v. Thus, at each time step, the detected individual stays in the same class with probability $(1-r)(1-v)$. This mechanism accounts for possible delays and imperfections in detection of disease symptoms — any individual can show symptoms but not be treated until after a number of steps.

The treatment event is a combination of two processes. Firstly, a detected individual is treated and moves to the treated class (**V**). Secondly, all individuals except removed (*i.e.* **S**, **I** or **D**) in the control neighbourhood (see below) are also treated. This process enables the health control authorities to capture individuals in the class **I** that do not show symptoms and all detected individuals (**D**) that are still waiting for treatment. In addition, it creates a zone around the focus of infection in which there are no susceptible individuals. Neither **V** nor **R** individuals can become infected again. The population has a constant number of individuals N, so that $N = \mathbf{S} + \mathbf{I} + \mathbf{D} + \mathbf{V} + \mathbf{R}$.

2.2. Network model

Interactions between individuals are captured by a network structure that exhibits a certain density of fully connected subgraphs in the form of cycles (termed otherwise cliques). Each vertex (representing an individual) can be a part of a c-clique, *i.e.* a group of c individuals that are fully connected, or can be a single node (*i.e.* a member of a 1-clique). Nodes which are members of a c-cliques have $c-1$ edges linking them with the neighbours within the same clique. For a random node with k connections to other vertices in the network, there are additional $k-c+1$ edges outside the clique. Here, we restrict our attention to random regular graphs, *i.e.* random graphs in which all nodes have the same degree k. Accordingly, each individual node simply connects to k other nodes (either single or in cliques).

Random clustered networks are described by the joint probability $\gamma(k,c)$ that a randomly chosen vertex has degree k and is a member of a c-clique [49]. In turn, the local clustering coefficient for a node is defined as a fraction of pairs of neighbours of this node which are also neighbours of each other. The degree-dependent clustering (or clustering spectrum c_k) is the average of the local clustering coefficient over the class of all nodes of degree k. The joint probability $\gamma(k,c)$ is represented by k by c matrix. In our paper, we consider random clustered networks where all vertices have the same degree ($k = 4$ or $k = 10$, see Fig. 1), and can be either a single node or part of a c-clique. The proportion of individuals in cliques is denoted by p. As an

example, the joint probability distribution $\gamma(k, c)$ that generates network with all nodes with degree $k = 4$, where p individuals are in 4-cliques and the rest $(1 - p)$ are single nodes is presented below:

$$\gamma(k, c) = \begin{matrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1-p & 0 & 0 & p \end{matrix} . \tag{1}$$

In practice, the algorithm by Gleeson *et al.* [49] works as follows. First, it generates a list of sizes of cliques in the network (in our model, the sizes are fixed). It then adds cliques directly into the adjacency matrix A_{ij} by selecting c nodes at random and connecting all nodes within the clique (by definition A_{ij} is 1 if the nodes i and j are connected, and 0 if not). A list of external stubs is also created which subsequently form inter-clique edges. Edges connecting cliques to other cliques, to individual points, and between individual points are then added to the adjacency matrix. Finally, self- and multi-connections are removed so that there is no more than one link connecting two different nodes. Each vertex can be a part of only one clique. Figure 1 shows three examples of different clustered networks.

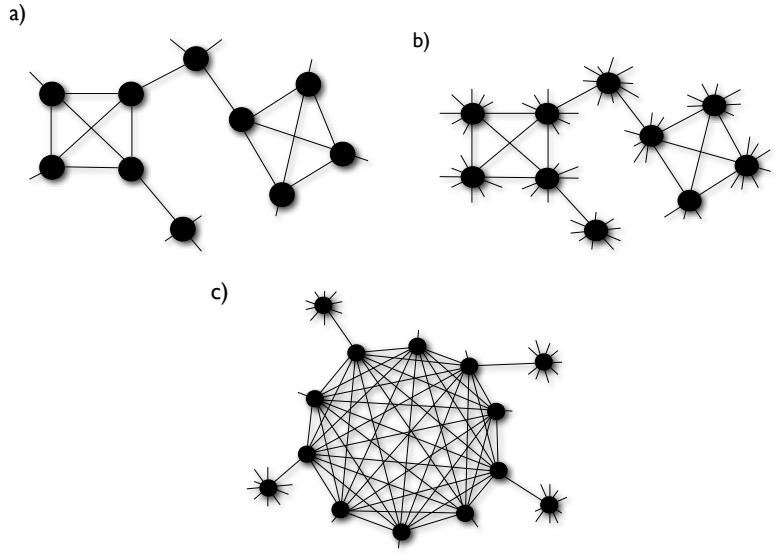

Fig. 1. Clustered random networks with a 4-cliques and a single nodes of degree 4 — network A (a), 4-cliques and a single nodes of degree 10 — network B (b), 10-clique and a single nodes of degree 10 — network C (c).

Infection and control neighbourhoods are defined iteratively. A neighbourhood $z = 1$ describes a set of k points which are connected to the central neighbour (note that each vertex has k connections). Then $z = 2$ extends this set to include all first-order neighbours of each neighbour from the set with $z = 1$. This procedure is then performed for higher-order neighbours. The zone $z = -1$ corresponds to an empty set (only applies to control), whereas $z = 0$ corresponds to the central individual only. Infection vicinity (characterised by z_{inf}) contains nodes to which disease can be transmitted (if the central node is infectious, either **I** or **D**), or from which the disease can be contracted (if the central node is susceptible, **S**). This neighbourhood is different to, and typically smaller than, the control neighbourhood (described by z).

The neighbourhoods naturally extend to cliques. In particular, if a control event is triggered by an individual that belongs to a c-clique, all individuals in this clique *and* at least one individual node that does not belong to any clique, are treated. If $z > 2$, than more cliques than one can be included in a single control event.

2.3. Network characteristics

Networks used in this paper can be characterized (among other measures) by the degree-dependent clustering coefficient, c_k, and by an average path length, L, see Table I. The degree-dependent clustering coefficient [18, 49] is given in terms of the sum

$$c_k = \sum_c \frac{\gamma(k,c)}{P_k} \frac{(c-1)(c-2)}{k(k-1)}\,, \tag{2}$$

where the degree distribution of the network (*i.e.* the probability that a random node has k neighbours) is obtained from the relation $P_k = \sum_{c=1}^{k+1} \gamma(k,c)$. The mean degree of the network is then $\langle k\rangle = \sum_k kP_k$. A node chosen at random from the set of all k-degree vertices is a member of a c-clique with probability $\gamma(k,c)/P_k$. Being a member of a c-clique, it is then a part of $\left(\begin{array}{c} c-1 \\ 2 \end{array}\right)$ triangles, so that its local clustering coefficient [18] is expressed by a fraction $\left(\begin{array}{c} c-1 \\ 2 \end{array}\right) \Big/ \left(\begin{array}{c} k \\ 2 \end{array}\right)$. The average path length [12] is defined by

$$L = \frac{\log N}{\log\langle k\rangle}\,, \tag{3}$$

where N is the number of nodes in the network, and $\langle k\rangle = k$ (in our work) stands for an average number of links per node. Increase in any of the

three parameters, k, c, and p results in increase of the clustering coefficient, representing increase in the proportion of individuals that are located in cliques. In contrast, the average length decreases when k increases from 4 to 10, and is not dependent on c and p (as in our network each node has exactly k links). However, the non-local properties of the neighbourhood control strategy means that all the clustering characteristics, c, k, p and c_k affect the optimal choice of control strategies.

TABLE I

Values of the parameters for networks used in the paper, ordered by a decreasing clustering coefficient. The last column lists an average path length. Note that $c \leq k$ always.

$k = \langle k \rangle$	c	p	c_k	L
10	10	0.75	0.6	4.7
4	4	0.75	0.375	7.8
10	10	0.25	0.2	4.7
4	4	0.25	0.125	7.8
10	4	0.75	0.05	4.7
10	4	0.25	0.0167	4.7

2.4. Economic model

The effectiveness of a control strategy is assessed in terms of a total "cost" associated with a disease outbreak when such a strategy is applied. In particular, we distinguish between two types of costs. Firstly, the costs associated directly with diseased individuals (*e.g.* palliative treatment, hospitalisation, absence from work, loss of production) can be estimated by the total number of individuals that have been through the disease throughout the outbreak, *i.e.* $R(t = \infty)$. Costs associated with preventive treatment (vaccination, culling) can be estimated by considering the final number of individuals in the V class, *i.e.* $V(t = \infty)$. Both approaches are possible because in our model there is no transition out of either R or V classes.

Thus, the total cost of the outbreak can be estimated by

$$X = a_1 R(t = \infty) + a_2 V(t = \infty)\,, \tag{4}$$

where a_1 is a unit cost associated with each *diseased* individual, while a_2 is a unit cost associated with each *treated* individual. Without loss of generality, we assume that $a_1 = 1$ and $a_2 = a$. The relative cost of treatment, a, is the main control parameter in our paper and varies between 10^{-4} (preventive treatment much cheaper than disease costs) to 10^3 (prevention much more expensive than disease). Although it is difficult to estimate this values for

real epidemics, values corresponding to $a = 0.017$–0.341 for influenza [51, 52] and $a = 0.01$–0.85 for rotavirus and hepatitis A [53, 54] can be found in literature. Even smaller values of a can be associated with diseases for which a vaccine is readily available and very cheap, *e.g.* measles ($a = 0.001$–0.01). However, when costs of developing, producing and administering a vaccine, including costs of delivery, are taken into account, a can exceed 1. In addition, culling animals or cutting trees, is also likely to bring a above 1.

In this context, we define the optimal strategy as a value of the treatment neighbourhood, z_c (which is typically larger than the infection neighbourhood, z_{inf}), for which the total cost, X is minimal (and then $X = X_c$). The optimisation is performed by fixing all parameters except control size, z, performing a single replicate of a simulated outbreak for a range of values of z. A minimum value of X, X_c, is then found for this series together with the associated neighbourhood, z_c. The whole process is then repeated 100 times to find the average values of X_c and z_c and their standard deviations. As a consequence of this procedure, the optimal control size, z_c does not need to be an integer (even though, the control size, z, is a discrete number) and in that way our results are illustrated in figures. However, in practice, the optimal control radius, z_c, will be rounded up due to the precautionary principle.

2.5. Simulation parameters

The population size is $N = 5\,000$. In this paper, we assess sensitivity of the optimal control strategy to changes in probability of disease spread, f, probability of symptoms development, q, and probability of treatment, v. Where not indicated otherwise, $f = 0.1$, $q = 0.5$, and $v = 0.1$. Other parameters are fixed; probability of spontaneous recovery, $r = 0.1$, infection neighbourhood, $z_{\text{inf}} = 1$ (*i.e.* k immediate neighbours are affected in one step). Initial number of infected (pre-symptomatic) individuals is $I(0) = 5$ (*i.e.* 0.1% of the population) and they are distributed randomly throughout the population.

To assess sensitivity of the results to network structure and clustering, we consider two levels of the number of links per node, $k = 4$ and $k = 10$; two levels of cluster sizes, $c = 4$ (for $k = 4$ and $k = 10$) and $c = 10$ (for $k = 10$); and three levels of the proportion of individuals in clusters, $p = 0$ (random network), $p = 0.25$ (25% individuals in clusters), and $p = 0.75$ (75% individuals in clusters). Note that $c \leq k$.

3. Results

As shown in our previous papers [31, 32, 35], the behaviour of system without control is characterised by a transition from limited, non-invasive

disease for small values of f to an invasive epidemic for larger f. As f tends to 1, all individuals in the population become infected. An addition of control allows the authorities to stop the disease spread even for high values of f, however, at the increased cost of treatment. There is, therefore, a trade-off between the costs of disease cases and preventive treatment [31]. If the treatment neighbourhood, z, is too small, the disease escapes control resulting in high values of R and, therefore. X. In contrast, if z is too big, treatment is wasted on healthy individuals which have no contact with infectious individuals (V and therefore X are large). As a result, a clear optimal value of z, z_c, appears, associated with the minimum of X, X_c. In the following, we analyse how the choice of optimal strategy represented by z_c changes with the relative cost of treatment, a, for different properties of the network (number of links per node, k, and size of the cluster, c) and the epidemiological parameters.

3.1. Effect of changing probability of spread, f

In absence of clustering, the network is identical to a random network. When the disease is invasive (for all f except the lowest one, $f = 0.01$), the only admissible control strategies are the Global Strategy (GS) whereby the control extends to all individuals in the population in one or very few steps, and the Null Strategy (NS) when it is optimal not to treat any individual, Fig. 2. GS is associated with control size $z_c \simeq 8$ (for node degree $k = 4$, almost all individuals are within distance of $z = 8$ from a random node and so will be treated in a single event) or $z_c \simeq 4$ (for $k = 10$). NS corresponds to $z_c = -1$ as no individual is treated — not even the infected one (see above

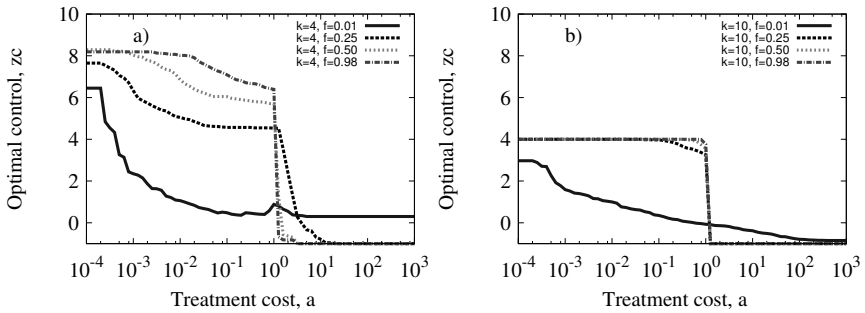

Fig. 2. No clustering: Control size, z_c, as a function of treatment cost, a, for different probabilities of spreading disease, f: $f = 0.01$ (solid/red lines), $f = 0.25$ (dashed/navy lines), $f = 0.5$ (dotted/blue lines), and $f = 0.98$ (dash-dotted/grey lines). Networks with degree $k = 4$ (left graph), with degree $k = 10$ (right graph). Other parameters: $q = 0.5$, and $v = 0.1$.

for the definition of z). The transition occurs at $a = 1$, except for small f (below invasion threshold), when it is best to treat the nearest neighbours ($z_c = 1$) for most values of a. For intermediate values of probability of disease spread, f, Local Strategy (LS), when treatment is applied to the neighbourhood of a detected individual, appears for all values of costs a smaller than 1. However, the radius of control, z_c, associated with LS is relatively high, $5 \leq z_c \leq 7$, and increases with increasing f. LS largely disappears for $f \simeq 1$, Fig. 2, as well as for the networks with degree $k = 10$.

In the case with clustering, we can identify three distinct control options, Fig. 3, the Global Strategy (GS), the Local Strategy (LS) and the Null Strategy (NS). However, regions of applicability for each scenario depend on the network properties and on whether the disease is invading or not.

Figure 3 (a) illustrates the situation when probability of disease spread, f, is very low ($f = 0.01$) and therefore the disease is not transmitted beyond the initial focus (*cf.* Fig. 1). All networks present the same behaviour. When the cost of treatment is very low ($a \leq 0.005$), GS is the cost-effective option but with increasing costs, a, z_c decreases gradually and reaches $z_c = -1$ that corresponds to NS. The exception are networks with low k, for which $z_c = 0$ (treating only the detected individual) is optimal for high a. Figures 3 (b), (c), (d) show the results with increasing probability of disease spread ($f = 0.25$ in (b), $f = 0.5$ in (c) and $f = 0.98$ in (d)). Three different strategies can still be found, similarly to the random network case.

Networks with 4-cliques and node degree $k = 4$ (thick black/red lines in Fig. 3) are characterized by the longest mean path length. Therefore, the optimal control, z_c, reaches the highest values when GS is the most cost-effective scenario. Moreover, the plateaux that corresponds to LS is the widest for networks with $c = k = 4$ and $p = 0.75$, Fig. 3. However, the plateaux is getting narrower with increasing probability of disease spread, f. Networks with $k = 10$ and with either $c = 4$ or $c = 10$ show results almost identical to random networks with $p = 0$. Increase of p to 0.75 extends the plateaux in this case as well, although the effect is small.

Number of cliques in networks affects the change between LS and NS. As the number of cliques in the population increases, the shift between LS and NS becomes sharper and moves towards lower treatment costs (approaching $a = 1$). The higher node degree, the smaller the difference between choice of control strategy for different number of cliques.

Finally, the network B with $c = 4$ and $k = 10$ largely follows the case of network C with $c = k = 10$ regardless of proportion of nodes in cliques, p, showing that the main effect is due to the change in the number of links per node, k, not the size of a clique, c. The apparent decrease in z_c in the region corresponding to GS (small values of a) is due to changes in the connectivity of the network. For $k = 10$, a single control event with $z = 5$

already reaches most of the nodes on the network, whereas for $k = 4$ it is necessary to extend z to $z = 8$ to achieve the same effect. Note that we keep the same f even though k increases, so the overall effect is of making the disease spread more rapidly.

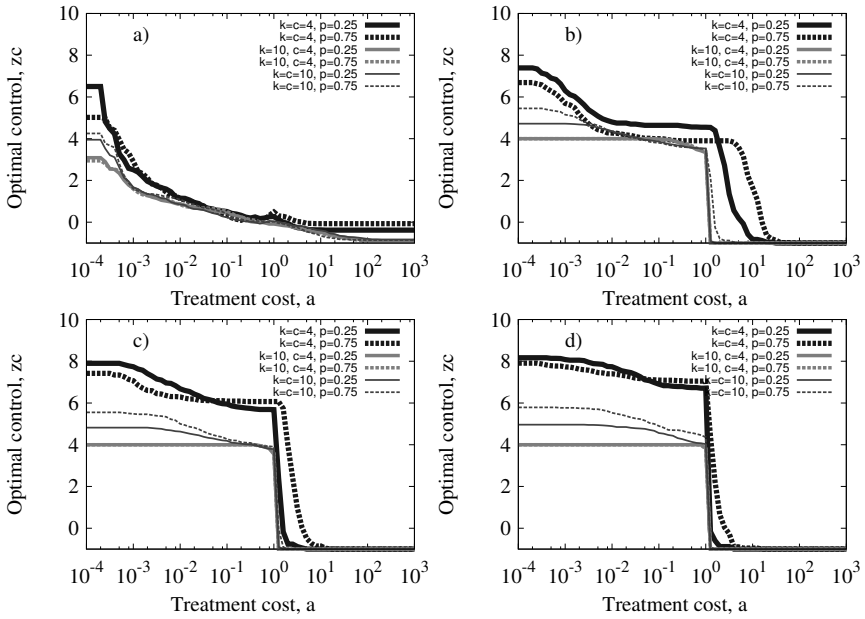

Fig. 3. With clustering: Control size, z_c, as a function of treatment cost, a, for different graph topology: networks A with node degree $k = 4$ and the size of cliques $c = 4$ (thick black/red lines), networks B with $k = 10$ and $c = 4$ (grey/blue lines) and networks C with $k = 10$ and $c = 10$ (thin/grey lines). All solid lines correspond to 75% ($p = 0.75$) of nodes in cliques, whereas dashed lines to 25% ($p = 0.25$). Probability of spreading disease, f, changes from $f = 0.01$ in (a), $f = 0.25$ in (b), $f = 0.50$ in (c) to $f = 0.98$ in (d). Other parameters: $q = 0.5$, and $v = 0.1$.

3.2. Effect of changing time until detection, $1/q$

The other important factor influencing the choice of the control strategy is the detection time, $1/q$. We first examine the effect of changing $1/q$ on random networks without clustering and then determine the effects of clustering.

We start with small values of $q = 0.01$ and, therefore, long times until detection, $1/q$. The longer it takes to examine the symptoms, the further the disease can spread without being noticed. This results in only two pos-

sibilities in the choice of the optimal control strategy: GS is chosen if costs $a < 1$ and NS (with $z_c = -1$) if $a \geq 1$, Fig. 4. The same sharp transition occurs when the network is clustered, Fig. 5 (a), although increasing proportion of individuals in cliques, p, shifts the values of control, z_c, in GS upwards. LS is not an optimal choice in that case (Fig. 5 (a)). The disease is transmitted without being detected and when the symptoms finally occur, pathogen already has reached the whole population.

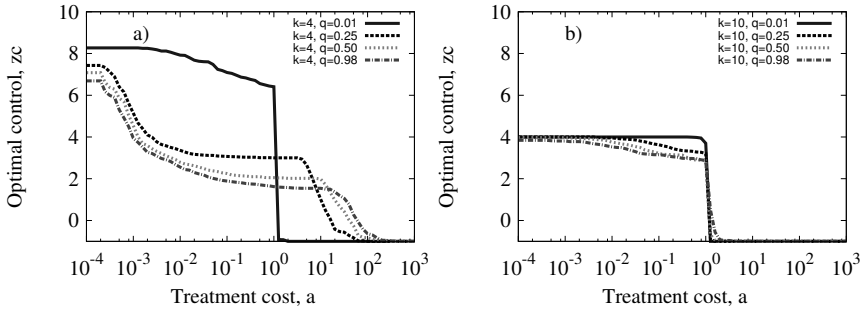

Fig. 4. No clustering: Control size, z_c, as a function of treatment cost, a, for different probabilities of occurring the symptoms, q: $q = 0.01$ (solid/red lines), $q = 0.25$ (dashed/navy lines), $q = 0.50$ (dotted/blue lines), and $q = 0.98$ (dash-dotted/grey lines). Networks with degree $k = 4$ (left graph), with degree $k = 10$ (right graph). Other parameters: $f = 0.1$, and $v = 0.1$.

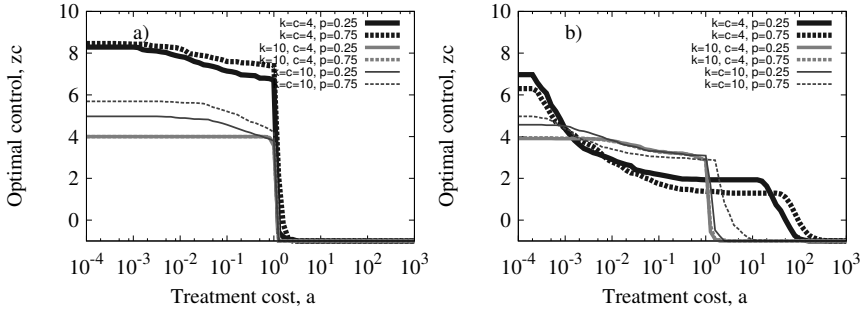

Fig. 5. With clustering: Control size, z_c, as a function of treatment cost, a, for different graph topology: networks A with node degree $k = 4$ and the size of cliques $c = 4$ (thick/red lines), networks B with $k = 10$ and $c = 4$ (grey/blue lines) and networks C with $k = 10$ and $c = 10$ (thin/grey lines). All solid lines correspond to 75% of nodes in cliques ($p = 0.75$), whereas dashed lines to 25% ($p = 0.25$). Probability of examine the symptoms $q = 0.01$ in (a) and $q = 0.50$ in (b). Other parameters: $f = 0.1$, and $v = 0.1$.

With decreasing detection time, $1/q$, the LS starts appearing with the associated z_c also decreasing, see Fig. 4 and compare with Fig. 5 (b). Interestingly, the region in which LS is optimal expands significantly as $1/q$ decreases. For fast detection times, LS can be applied even if the treatment is about 100 times more expensive than disease cases, $a \simeq 100$. The proportion of nodes in cliques, p, affects not only the region in which LS is valid, but also the value of z_c at the plateaux, Fig. 5, although the latter effect is relatively small. The larger p, the more cost-effective LS is, as the transition from LS and NS occurs at higher values of a for $p = 0.75$ than for $p = 0.25$. Also, increase in p results in small decrease in z_c at the plateaux. The biggest effect on the transition is, however, due to changes in k, for both non-clustered, Fig. 4, and clustered networks, Fig. 5.

As before, in the region where GS is valid, smaller values of z_c correspond to treating the whole population for $k = 10$ than for $k = 4$, see Fig. 5 and compare with Fig. 4. The results for $c = 4$ and $k = 10$ again follow the case with $c = k = 10$, so the main effect is associated with changing k. Interestingly, the effect of increasing k is opposite for LS, as z_c increases in this case, see Fig. 5 (b). This is due to the disease spreading much quicker for $k = 10$ than for $k = 4$, with the same f. This must be countered by increasing the size of the control neighbourhood.

3.3. Effect of changing time until treatment, $1/v$

Finally, we look at the efficiency of treatment, v. The balance between this parameter and probability of removal, r, determines the proportion of detected individuals that either are removed spontaneously, or are treated in control events. Thus, $1/v$ can be interpreted as time from detection to treatment, with the caveat that some individuals might become removed (**R**) (recover and become immune, or die) while waiting for treatment. Similarly to the case of detection rate, q, there is a big difference between low and high values of recovery, v, both for the non-clustered, Fig. 6, and clustered networks, Fig. 7.

When recovery rate, v, is small and the time until treatment, $1/v$, is long, the situation presents similar behaviour to the case of small probability of showing the symptoms,q. As long as the symptomatic individuals remain infectious, they continue to spread the disease while waiting for treatment. As a result, broadly speaking, there is only a choice between GS for $a < 1$ and NS for $a \geq 1$. However, there is some gradual change in control size, z_c, for GS and the transition at $a = 1$ is not as sharp as before, see Fig. 6.

Interestingly, although the fact that for long times till treatment, $1/v$, clustering introduces some evidence of a plateaux associated with LS, the values of control size, z_c is rather high ($z_c \simeq 6$). The plateaux is also

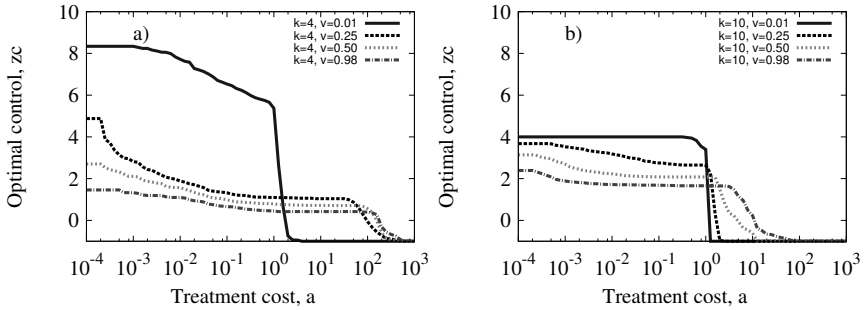

Fig. 6. No clustering: Control size, z_c, as a function of treatment cost, a, for different probabilities of recovery, v: $v = 0.01$ (solid/red lines), $v = 0.25$ (dashed/navy lines), $v = 0.50$ (dotted/blue lines), and $v = 0.98$ (dash-dotted/grey lines). Networks with node degree $k = 4$ (left graph) and with node degree $k = 10$ (right graph). Other parameters: $f = 0.1$, and $q = 0.5$.

extended towards treatment costs $a > 1$ when proportion of individuals in cliques $p = 0.75$ as compared to $p = 0.25$, Fig. 7. There is no consistent effect of clustering on control size, z_c, in the region of GS, Fig. 7. In addition, increase in degree of nodes, k, decreases the value of z_c for GS and shifts the transition from GS to NS towards costs $a = 1$. The reason of that behaviour is the infection that spreads easier in the networks with degree $k = 10$ than for networks with $k = 4$ (and for the same transmission rate, f).

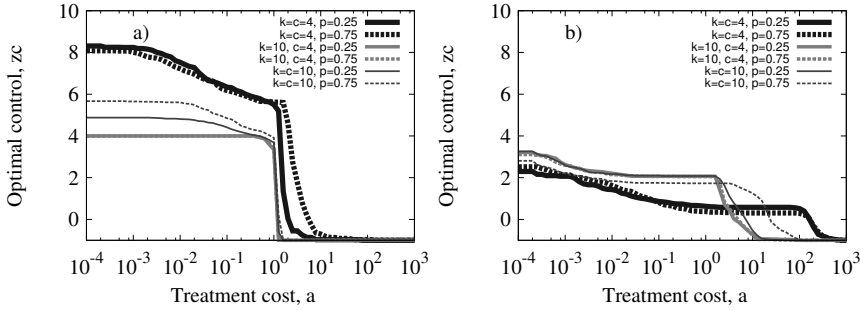

Fig. 7. With clustering: Control size, z_c, as a function of treatment cost, a, for different graph topology: networks A with node degree $k = 4$ and the size of cliques $c = 4$ (thick/red lines), networks B with $k = 10$ and $c = 4$ (grey/blue lines) and networks C with $k = 10$ and $c = 10$ (thin/grey lines). All solid lines correspond to 75% of nodes in cliques ($p = 0.75$), whereas dashed lines to 25% ($p = 0.25$). Probability of recovery $v = 0.01$ in (a) and $v = 0.50$ in (b). Other parameters: $f = 0.1$, and $q = 0.5$.

When treatment control can be applied without any delay (large v, small $1/v$), control size, z_c, is significantly lower than before. There is practically no evidence of GS as the optimal option and the plateaux associated with LS extends towards very small values of treatment costs, a. Thus, if we can act quickly, it is optimal to treat population locally even if the treatment cost, a, is very low and there is a temptation to treat indiscriminately (as in GS). We do not assume that some additional cost is associated with detection. Increase in clustering (from $p = 0$ to $p = 0.25$ and $p = 0.75$) shifts the extend of the plateaux towards higher values of costs, a, although the effect is small for small node degree, k, and the size go cliques, c.

The effect of changing k is similar as for the probability of showing the symptoms, q, both for GS (decrease in control size z_c as k increases) and for LS (increase in z_c), Fig. 7.

4. Discussion

Faced with an outbreak of a novel disease, the authorities need to decide on the approach to controlling its spread. One possibility might be to refrain from any preventive action and concentrate on palliative treatment of infected cases, effectively letting the epidemic to unfold itself (Null strategy). Alternatively, they can attempt to treat the whole population as quickly as possible (Global strategy). Finally, there is a possibility of a gradual responsive approach, whereby new cases are identified and then contact tracking is used to preventively treat individuals who might have links with the pre- and symptomatic individual (Local strategy). The extent of this "ring" control needs then to be determined by taking into account both epidemiological and economic factors.

In our previous work, we studied the dynamics of the disease spreading on regular, small-world and random networks. Although they capture some aspects of the structure of real networks of contacts between people, animals or plants, they do not include the effect of clustering noted in many real-life applications [47, 48]. This paper fills in this gap and takes an important step towards application of the modelling framework to realistic systems.

We have shown here and elsewhere [31] that the broad strategy choice (NS, GS or LS) is primarily determined by the relative cost of palliative and preventive treatments. In this paper, we are particularly interested in finding conditions under which the local strategy (LS) is optimal for as wide range of treatment costs, a, as possible. If the prevention is expensive ($a \gg 1$), the choice favours the NS. The GS becomes optimal for very low cost of vaccination ($a \ll 1$). However, the LS emerges for $a \simeq 1$ for disease agents with certain properties. Higher the probability of disease spread, f, decreases the range of optimality of LS and, at the same time, increases optimal control size, z_c, so that LS eventually merges with GS. Rise in

either detection q or treatment rate v (corresponding to the decline in the time till detection, $1/q$, and the time till treatment, $1/v$) expands the range in which LS is optimal, mainly towards high values of treatment costs, a. Thus, boost efficiency of detection and reaction of public health systems makes the LS more attractive, even if the actual treatment and prevention remain very expensive. The reason is that we are able to catch the outbreak early and stop it from expansion. Interestingly, higher treatment rate, v, also reduces the range of optimality for GS for very low treatment.

Network topology and its effect on the choice of the optimal control strategy form the key element addressed in this paper. Our analysis shows that the average path length, L appears to be the decisive factor — the larger L, the larger the interval for which LS is optimal. However, this is at the cost of growing control size, z_c. The degree-dependent clustering coefficient, c_k is the other crucial parameter. The large value of c_k leads to a small expansion of the range of LS applicability, particularly for $a > 1$. The relative insensitivity of the results to clustering is an important result for public health measures. We are not very likely to know the exact properties of the real network, therefore the knowledge of details of LS predicted by the mathematical models is significant, even if they do not exactly represent the real levels of clustering.

Altogether, in this paper, we studied the effect of topological and epidemiological factors on the choice of the optimal control strategy for epidemics spreading on clustered random networks. We particularly addressed the applicability of the local strategy (LS) in which individuals are treated in a neighbourhood of a detected case. The work can be extended in a number of directions. The network can be made more realistic, using real-world data collected for example by usage of mobile phones. The epidemiological model can also be extended to include different levels of mixing and changes in the network due to disease appearance. The current economic model is also very simple; there are many levels of costs that can be incorporated, including detection and contact tracing, hospitalisation, and delivery of vaccines.

We are very grateful to Professor James Gleeson (University of Limerick) for helpful discussions and his support during K.O.'s visit to Limerick. This work was partly supported by a project which operated within the Foundation for Polish Science International Ph.D. Projects Programme co-financed by the European Regional Development Fund covering, under the agreement No. MPD/2009/6, the Jagiellonian University International Ph.D. Studies in Physics of Complex Systems and was co-financed by the European Development Fund in the framework of the Innovative Economy Programme and within the framework of Exploring the Physics of Small Devices (EPSD) of the European Science Foundation (ESF).

REFERENCES

[1] R.M. Anderson, R.M. May, B. Anderson, *Infectious Diseases of Humans: Dynamics and Control*, vol. 28, Wiley Online Library, 1992.

[2] C. Dye, N. Gay, *Science* **300**, 1884 (2003).

[3] B. Hu *et al.*, *Science China Earth Sciences* **56**, 1380 (2012).

[4] M. Keeling, *Proc. R. Soc.* **B272**, 1195 (2005).

[5] D.T. Haydon, R.R. Kao, R.P. Kitching, *Nat. Rev. Microbiol.* **2**, 675 (2004).

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

[7] T.R. Gottwald *et al.*, *Phytopathology* **91**, 30 (2001).

[8] M.J. Keeling, K.T. Eames, *J. R. Soc. Interface* **2**, 295 (2005).

[9] M.E. Newman, *SIAM Rev.* **45**, 167 (2003).

[10] C. Moore, M.E. Newman, *Phys. Rev.* **E61**, 5678 (2000).

[11] R. Pastor-Satorras, A. Vespignani, *Phys. Rev. Lett.* **86**, 3200 (2001).

[12] S. Dorogovtsev, J. Mendes, *Evolution of Networks: From Biological Nets to the Internet and WWW*, Clarendon Press, Oxford 2003.

[13] M.E. Newman, D.J. Watts, S.H. Strogatz, *Proc. Natl. Acad. Sci. USA* **99**, 2566 (2002).

[14] A.-L. Barabâsi *et al. Physica A* **311**, 590 (2002).

[15] G. Caldarelli, A. Capocci, P. De Los Rios, M.A. Muñoz, *Phys. Rev. Lett.* **89**, 258702 (2002).

[16] D.B. West *et al.*, *Introduction to Graph Theory*, vol. 2, Prentice Hall Englewood Cliffs, 2001.

[17] M. Newman, *Networks: an Introduction*, Oxford University Press, 2009.

[18] J.P. Gleeson, *Phys. Rev.* **E80**, 036107 (2009).

[19] G. Ódor, *Phys. Rev.* **E88**, 032109 (2013).

[20] C.A. Gilligan, F. van den Bosch, *Annu. Rev. Phytopathol.* **46**, 385 (2008).

[21] M.E.J. Newman, C.R. Ferrario, *PLoS One* **8**, e71321 (2013).

[22] G. Fournié *et al.*, *Proc. Natl. Acad. Sci. USA* **110**, 9177 (2013).

[23] J. Ma, P. van den Driessche, F.H. Willeboordse, *J. Theor. Biol.* **325**, 12 (2013).

[24] M.D. Shirley, S.P. Rushton, *Ecol. Complex.* **2**, 287 (2005).

[25] E.E. Rees, B.A. Pond, R.R. Tinline, D. Bélanger, *J. Appl. Ecol.* **50**, 881 (2013).

[26] B. Lin, Z. Sun, X. Fu, G. Zhu, *Int. J. Biomath.* **06**, 1350025 (2013).

[27] S. Barrett, *J. Eur. Econ. Assoc.* **1**, 591 (2003).

[28] E. Campbell, M. Salathé, *Sci. Rep.* **3**, 1905 (2013).

[29] C. Buono, F. Vazquez, P.A. Macri, L.A. Braunstein, *Phys. Rev.* **E88**, 022813 (2013).

[30] G.A. Forster, C.A. Gilligan, *Proc. Natl. Acad. Sci.* **104**, 4984 (2007).
[31] A. Kleczkowski, K. Oleś, E. Gudowska-Nowak, C.A. Gilligan, *J. R. Soc. Interface* **9**, 158 (2012).
[32] K. Oleś, E. Gudowska-Nowak, A. Kleczkowski, *PLoS One* **7**, e36026 (2012).
[33] M. Gersovitz, J.S. Hammer, *Econ. J.* **114**, 1 (2004).
[34] B. Dybiec, A. Kleczkowski, C.A. Gilligan, *Phys. Rev.* **E70**, 066145 (2004).
[35] K. Oleś, E. Gudowska-Nowak, A. Kleczkowski, *PLoS One* **8**, e63813 (2013).
[36] B. Dybiec, A. Kleczkowski, C.A. Gilligan, *J. R. Soc. Interface* **6**, 941 (2009).
[37] M.L.N. Mbah, C.A. Gilligan, *Math. Med. Biol.* (2013), DOI:10.1093/imammb/dqt012.
[38] B. Dybiec, *Physica A* **387**, 4863 (2008).
[39] G. Smith, C. Cheeseman, D. Wilkinson, R. Clifton-Hadley, *J. Appl. Ecol.* **38**, 520 (2001).
[40] G. Smith, C. Cheeseman, R. Clifton-Hadley, D. Wilkinson, *J. Appl. Ecol.* **38**, 509 (2001).
[41] R. Woodroffe, S.D. Frost, R.S. Clifton-Hadley, *J. Appl. Ecol.* **36**, 494 (1999).
[42] F. Behlau *et al.*, *Crop Prot.* **27**, 807 (2008).
[43] J. Medlock, A.P. Galvani, *Science* **325**, 1705 (2009).
[44] H.-F. Zhang *et al.*, *Phys. Rev.* **E88**, 012813 (2013).
[45] D.J. Watts, S.H. Strogatz, *Nature* **393**, 440 (1998).
[46] M.E. Newman, D.J. Watts, *Phys. Rev.* **E60**, 7332 (1999).
[47] L. Danon *et al.*, *Proc. R. Soc.* **B280**, 20131037 (2013).
[48] E. Volz, *Phys. Rev.* **E70**, 056115 (2004).
[49] J.P. Gleeson, S. Melnik, A. Hackett, *Phys. Rev.* **E81**, 066114 (2010).
[50] M.A. Serrano, M. Boguñá, *Phys. Rev. Lett.* **97**, 088701 (2006).
[51] M.I. Meltzer *et al.*, *Emer. Infect. Dis.* **5**, 659 (1999).
[52] D. Weycker *et al.*, *Vaccine* **23**, 1284 (2005).
[53] J. Bilcke, P. Beutels, *PharmacoEconomics* **27**, 281 (2009).
[54] J. Luyten, P. Beutels, *PharmacoEconomics* **27**, 379 (2009).