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

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 is 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 the control 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.

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 ($\mathbf{S}$), except of a fixed small number of infected pre-symptomatic ($\mathbf{I}$) 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 ($\mathbf{I}$) or the detected individual ($\mathbf{D}$), the disease is passed to a susceptible individual ($\mathbf{S}$) that becomes infectious but pre-symptomatic individual ($\mathbf{I}$). Subsequently, the infected individual displays symptoms and the transition to a symptomatic state ($\mathbf{D}$) 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.
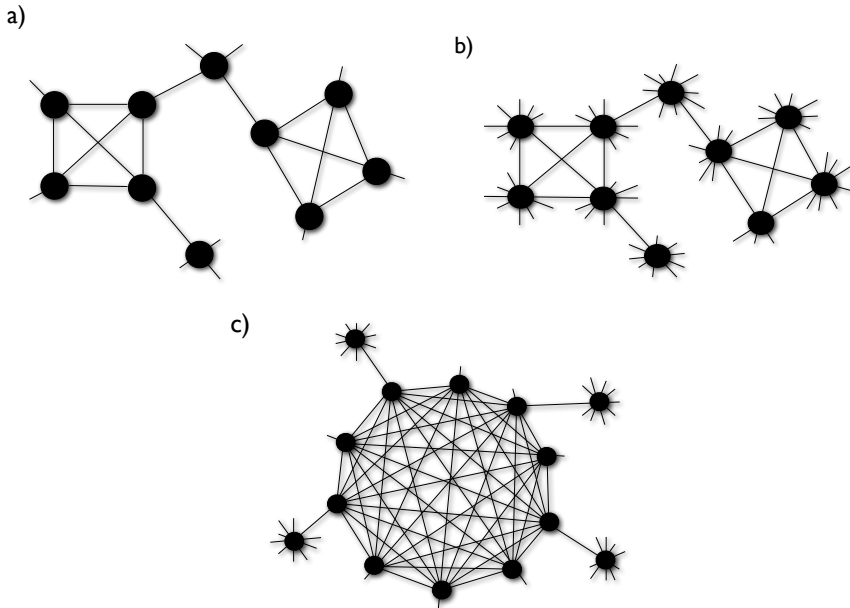

Fig. 1. Clustered random networks with a 4-cliques and single nodes of degree 4 — network A (a), 4-cliques and single nodes of degree 10 — network B (b), 10-clique and single nodes of degree 10 — network C (c).

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 $\langle k \rangle = \sum_k k P_k$. 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 $\binom{c-1}{2}$ triangles, so that its local clustering coefficient [18] is expressed by a fraction $\binom{c-1}{2} / \binom{k}{2}$. 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 mean 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 always $c \leq k$.

$k = \langle k \rangle$	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, $z_{inf} = 1$ (*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
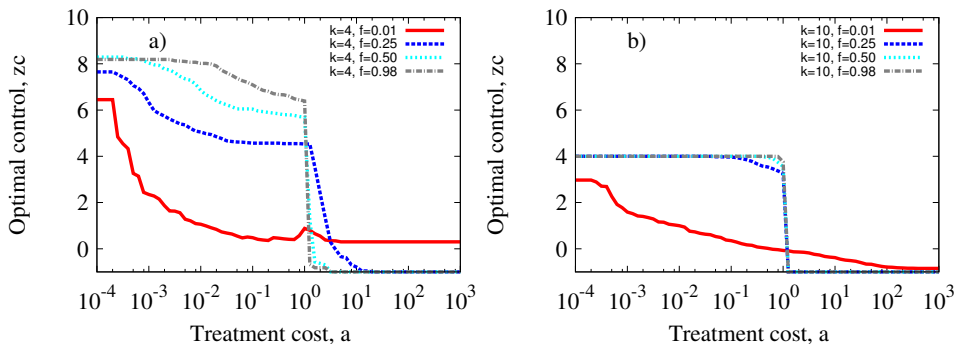

Fig. 2. No clustering: Control size, z_c, as a function of treatment cost, a, for different probabilities of spreading disease, f: $f = 0.01$ (solid/red lines), $f = 0.25$ (dashed/navy lines), $f = 0.5$ (dotted/blue lines), and $f = 0.98$ (dash-dotted/grey lines). Networks with degree $k = 4$ (left graph),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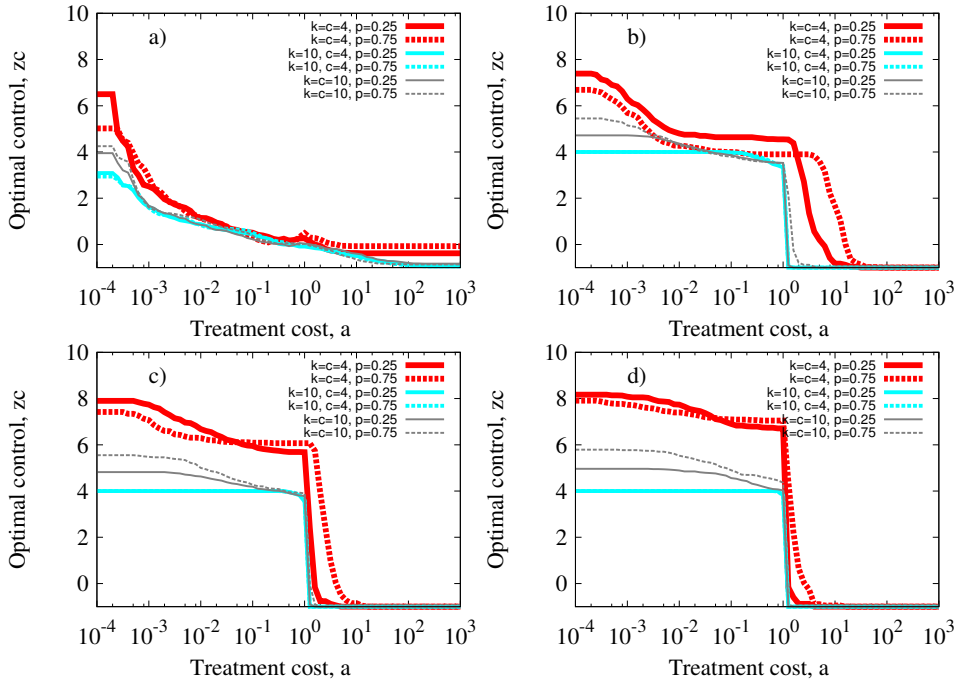

Fig. 3. With clustering: Control size, z_c, as a function of treatment cost, a, for different graph topology: networks A with node degree $k = 4$ and the size of cliques $c = 4$ (thick black/red lines), networks B with $k = 10$ and $c = 4$ (grey/blue lines) and networks C with $k = 10$ and $c = 10$ (thin/grey 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.
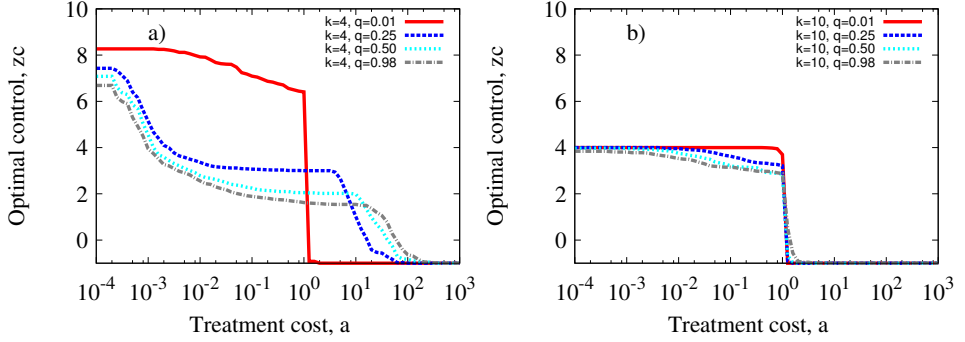

Fig. 4. No clustering: Control size, z_c, as a function of treatment cost, a, for different probabilities of occurring the symptoms, q: $q = 0.01$ (solid/red lines), $q = 0.25$ (dashed/navy lines), $q = 0.50$ (dotted/blue lines), and $q = 0.98$ (dash-dotted/grey lines). Networks with degree $k = 4$ (left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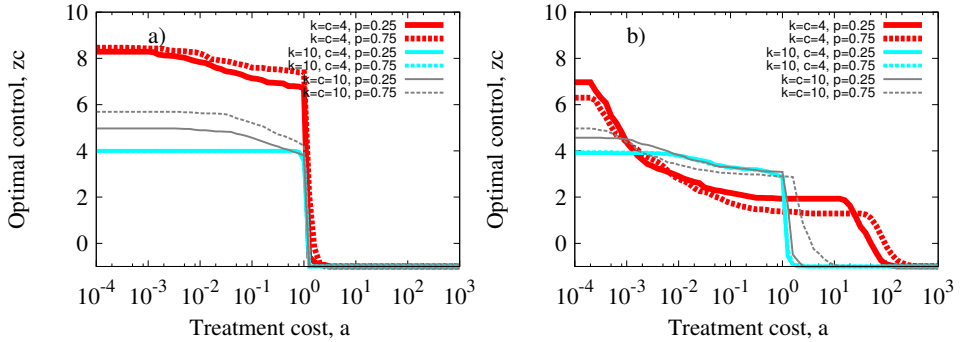

Fig. 5. With clustering: Control size, z_c, as a function of treatment cost, a, for different graph topology: networks A with node degree $k = 4$ and the size of cliques $c = 4$ (thick/red lines), networks B with $k = 10$ and $c = 4$ (grey/blue lines) and networks C with $k = 10$ and $c = 10$ (thin/grey 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 ($\mathbf{R}$) (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
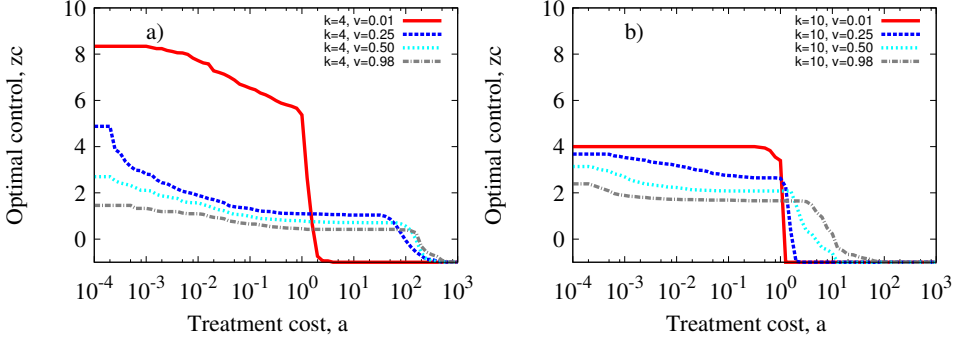
Fig. 6. No clustering: Control size, z_c, as a function of treatment cost, a, for different probabilities of recovery, v: $v = 0.01$ (solid/red lines), $v = 0.25$ (dashed/navy lines), $v = 0.50$ (dotted/blue lines), and $v = 0.98$ (dash-dotted/grey lines). Networks with node degree $k = 4$ (left graph) 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).
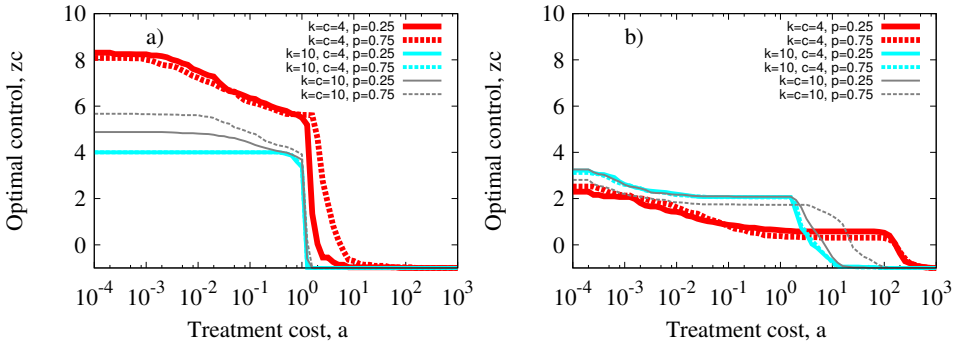

Fig. 7. With clustering: Control size, z_c, as a function of treatment cost, a, for different graph topology: networks A with node degree $k = 4$ and the size of cliques $c = 4$ (thick/red lines), networks B with $k = 10$ and $c = 4$ (grey/blue lines) and networks C with $k = 10$ and $c = 10$ (thin/grey 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.

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).