Journal of Theoretical Biology 288 (2011) 21–28

Contents lists available at ScienceDirect

Journal of Theoretical Biology

journal homepage: www.elsevier.com/locate/yjtbi

Predicting epidemic thresholds on complex networks: Limitations of mean-field approaches

Or Givan, Nehemia Schwartz, Assaf Cygelberg, Lewi Stone*

Biomathematics Unit, Faculty of Life Sciences, Tel Aviv University, Israel

ARTICLE INFO

Article history:
Received 1 September 2010
Received in revised form
30 June 2011
Accepted 22 July 2011
Available online 7 August 2011

Keywords:
SIS model
Markov chain
Poisson process
Eigenvalue

ABSTRACT

Over the last decade considerable research effort has been invested in an attempt to understand the dynamics of viruses as they spread through complex networks, be they the networks in human population, computers or otherwise. The efforts have contributed to an understanding of epidemic behavior in random networks, but were generally unable to accommodate specific nonrandom features of the network's actual topology. Recently, though still in the context of the mean field theory, Chakrabarti et al. (2008) proposed a model that intended to take into account the graph's specific topology and solve a longstanding problem regarding epidemic thresholds in both random and nonrandom networks. Here we review previous theoretical work dealing with this problem (usually based on mean field approximations) and show with several relevant and concrete counter examples that results to date breakdown for nonrandom topologies.

1. Introduction

Under what conditions will a virus or other infectious agent spread in a complex population network? This question has vexed epidemiologists, mathematicians and computer scientists alike for many decades (Anderson and May, 1991; May and Lloyd, 2001; Pastor-Satorras and Vespignani, 2001; Madar et al., 2004; Aparacio and Pascual, 2007; Berchenko et al., 2009). An early result arising from epidemic modeling is based on the so-called reproductive number R_0, the number of secondary infections a typical infected individual is able to generate (Anderson and May, 1991). If a typical infected individual is able to infect on average more than one other member of the population then $R_0>1$. In that case the virus is able to reproduce itself and trigger an epidemic in the population, allowing it to persist in time for an extensive period. In contrast, if the reproductive number is below unity then $R_0<1$, and the disease will rapidly die out in the population and an infection free equilibrium will be reached. This threshold result assumes that the population is homogeneous and randomly mixing, whereby an infected individual is equally likely to come into contact and infect any susceptible present, an assumption that has many limitations.

This result has been generalized to heterogeneous populations in which some individuals have more contacts than others. Historically, most notable are the studies of Dietz (1980) and May and Anderson (1988, 1984). For heterogeneous networks, the relative fraction of nodes having different degrees is referred to as the degree distribution. It is just the probability P_k that a randomly chosen node in the network has degree k, i.e, $P_k=prob(\deg(\text{node } i)=k)$. The mean degree of the network is $\langle k\rangle=\sum kP_k$ and variance of the degree distribution is $var(k)=\sigma^2=\langle k^2\rangle-\langle k\rangle^2$. Consider a large population and let d_k be the number of individuals in the population that have k contacts, with $\sum_k d_k=L$. Then (d_k/L) estimates the degree-distribution P_k.

The degree of heterogeneity in the population's contact structure may be gauged by the Coefficient of Variation $CV=(\sigma/\langle k\rangle)$. Equivalently one can write $CV^2=(\langle k^2\rangle/\langle k\rangle^2)-1$. The population is assumed to be randomly mixed subjected to the constraint so that the degree-distribution is always preserved. For such a heterogeneous population, it has been shown that (Anderson and May, 1991; Dietz, 1980; May and Anderson, 1988; May and Anderson, 1984):

$$R_0=\overline{R}(1+CV^2) \tag{1}$$

where $\overline{R}$ is the reproductive number for the equivalent homogeneous population where all individuals have $\langle k\rangle$ contacts and thus $CV=0$.

Eq. (1) will be referred to as the Dietz–May formula since both authors (May and Anderson, 1988; May and Anderson, 1984) were responsible for developing the formulation and applying it in practice. As before $R_0>1$ implies that an epidemic will ensue, while $R_0<1$ implies that the virus rapidly dies out and an infection free equilibrium is attained. These concepts have proved useful for studying contact networks with power-law distributions that

* Corresponding author.
E-mail address: lewi@post.tau.ac.il (L. Stone).

0022-5193/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jtbi.2011.07.015

might typify computer networks and in some cases be appropriate for studying the transmission of sexually distributed diseases (Pastor-Satorras and Vespignani, 2001; Lloyd and May, 2001). Measuring R_0 is often quite complicated, especially in complex networks as Aparacio and Pascual (2007) have discussed. Newman, 2002 and Cohen et al. (2002) have shown how percolation ideas and generating function methods can be used to provide exact solutions of epidemic models on simple networks and on bi-partite graphs. Their key epidemic threshold result is nevertheless the same as Eq. (1) obtained by the different methods.

Considerable work has been invested in exploring these issues in the biology, physics and mathematics literature. Concepts taken from the percolation theory continue to play a major role in current epidemic network research (Madar et al., 2004; Cohen et al., 2002; Kenah and Robins, 2007; Parshani et al., 2010). Moreover, there are many challenging open problems (see the interesting inaugural article of Durrett, 2010).

In recent years, there has been considerable interest in understanding the way in which the detailed network structure of the population, or its "topology," might affect the persistence threshold. That is, does the exact network structure, not just its degree distribution, give extra information from which it is possible to learn more about the spread of an epidemic? Since many real networks are non-random and sometimes highly clustered, the motivation to explore beyond random models is quite justified.

Chakrabarti, Wang, Faloutsos (Chakrabarti et al., 2008) introduced a new model, referred to here as the CWF model, which intended to identify exactly how a population's network structure controls the epidemic threshold. A very general epidemic threshold condition for any arbitrary network was derived. This condition is based on the network's topology as a mean field approximation will be elaborated shortly.

In this paper we show that many of the previous studies contribute to our understanding of epidemic thresholds for random networks, however for nonrandom network topologies (even regular graphs) accurate predictions of the epidemic threshold are hard to come by. We explore the mean field approximation formulated by CWF and show that its predictions often break down for nonrandom networks. This is because mean-field approximations fail to take into account the correlations in the state of indirect neighbors. Moreover, by mapping one model to another, we are able to retrieve known theoretical literature results (based on percolation theory) that contradict the CWF general threshold condition.

2. The CWF model

CWF (Chakrabarti et al., 2008) assume that the population is divided into two classes: individuals that are Susceptible ($\boldsymbol{S}$) and individuals that are Infected ($\boldsymbol{I}$). The model has the classical *SIS* structure whereby susceptible individuals may become infected upon contact with an infected individual. After contracting the disease an individual recovers after some fixed time period and becomes susceptible once again, thereby closing the *SIS* loop.

As each individual can be in one of the two states, for a complex network of N individuals, there are 2^N possible different states the population may be found in. It is appropriate to formulate the model in terms of a Markov chain, but this requires information specifying the probabilities between each of the possible states. In this formulation states correspond to particular configurations of the population network, with the configuration at each time step dependent on the former time step only. However, it is not a simple matter to determine the probabilities of the $2^N \times 2^N$ transition matrix, which in any case is impractical to work with even when N is modestly large, let alone of the order of millions of individuals as is appropriate for large cities. As such, CWF developed a method to approximate the Markov chain model.

In more detail they consider a population network, and define an individual's neighbors as all members of the population he or she can directly contact and transmit the disease. Set β as the probability that an infected individual/node will infect a susceptible neighboring node in the network, and let the probability that node-i is infected at time t be given by $p_{i,t}$. Over one time-step, the probability that node-i will not receive any infections from its neighbors is, according to CWF, given by

$$\zeta_{i,t} \simeq \prod_{j \in M_i} [(1-\beta)p_{j,t-1} + (1-p_{j,t-1})] = \prod_{j \in M_i} [1-\beta p_{j,t-1}] \tag{2}$$

where M_i is the set of all neighbors of node-i. Note Eq. (2) is exact only when it is assumed that the nodes $p_{j,t-1}$ are independent of each other. This "independence assumption" is of great importance and will be dealt with in detail in what follows. Thus, according to Chakrabarti et al. (2008) the probability that node-i is healthy at time t is given by

$$1-p_{i,t} = (1-p_{i,t-1})\zeta_{i,t} + \delta p_{i,t-1}\zeta_{i,t} \tag{3}$$

where δ is the probability that an infected node will recover at time-step t. Note that since recovery is geometrically distributed, the mean infection time is $1/\delta$. This last equation states that node-i is healthy at time t if it did not receive infections from its neighbors at t and either node-i was uninfected at time step $t-1$, or was infected at $t-1$ but was cured at t (Chakrabarti et al., 2008). (This last term in Eq. (3), which appears in the CWF model (Chakrabarti et al., 2008), is problematical as we explain in the discussion. It can however be dropped without affecting the results of the following stability analysis.) Combining Eqs. (2) and (3) yields the CWF model

$$p_{i,t} = 1-[1+(\delta-1)p_{i,t-1}] \prod_{j \in M_i} (1-\beta p_{j,t-1}) \tag{4}$$

It is clear that the model has an infection free equilibrium in which $p_i^* = 0\ \forall i$. (Here, the star notation indicates a state of equilibrium.) We now proceed to examine this equilibrium's local stability. Using vector notation, close to the equilibrium, Eq. (4) may be approximated as

$$\vec{p}_{\,t} = ((1-\delta)\boldsymbol{I} + \beta \boldsymbol{A}) \cdot \vec{p}_{\,t-1} \tag{5}$$

where, I is the identity matrix and A is the adjacency matrix of binary entries 1,0 representing the connectivity between the nodes. Thus, the infection free equilibrium ($p_i^* = 0\ \forall i$) is locally stable only if

$$(1-\delta) + \beta\rho < 1 \tag{6}$$

where ρ is the spectral radius of the matrix A. This is because the Perron–Frobenius theorem ensures that if A is a nonnegative, irreducible matrix then one of its eigenvalues is real, positive and greater than or equals to (in absolute value) all other eigenvalues (Horn and Johnson, 1985). This eigenvalue is the spectral radius ρ.

In terms of the reproductive number, the infection free equilibrium is locally stable if

$$R_0 = \frac{\beta}{\delta}\rho < 1 \tag{9}$$

The reproductive number R_0 has a simple interpretation. Returning to Eq. (6) we see that if $\vec{p}_{\,0}$ is an eigenvector corresponding to eigenvalue ρ of $\boldsymbol{A}$, the expected number of newly infected individuals in the next generation $\vec{p}_{\,1}$ is given by $\beta\rho$, while the expected number of recovered individuals is δ. Since the mean infectivity time is $1/\delta$, then $(\beta/\delta)\rho$ should be interpreted as the total number of new infections generated in a single time step multiplied by the actual infectivity period of the disease.

Hence R_0 is simply the mean number of secondary infections over the infectious period of the disease.

This conforms closely with the conventional view of R_0 as the number of secondary cases that one infected case can produce when placed in a wholly susceptible population. If it can infect more than one individual on average ($R_0>1$) an epidemic will ensue otherwise the infection will rapidly die out as the infection free equilibrium is reached. In what follows, (9) will be referred as the CWF threshold criterion, since stability of the infection free equilibrium depends on whether R_0 is greater or less than unity. In this way R_0, as given by Eq. (9), may be used as a reference frame for testing the CWF threshold

It should be pointed out that the above analysis concerns the underlying deterministic mean-field model presented by CWF, and this raises two issues. First, for the full stochastic model, in which the mean-field is supposed to mimic, one has to take into account the stochastic effects. In particular, if $R_0>1$ then demographic stochasticity at the initiation of an epidemic when infectives are in small numbers, can prevent the epidemic from triggering. This is the "stochastic epidemic theorem" (Renshaw, 1991): even though $R_0>1$ there is a finite probability that the epidemic will not trigger. However, if $R_0<1$ a major epidemic cannot occur.

3. Known epidemic thresholds for random networks

We first consider random networks making use of the results from Furedi and Komlos (1981). The latter authors studied random, symmetric, $N\times N$ matrices in which the elements a_{ij} are identically distributed having the same mean μ and variance σ^2. For such matrices the largest eigenvalue may be approximated by

$$\rho=\frac{\sum_{i,j}a_{i,j}}{N}+\frac{\sigma^2}{\mu}+O\left(\frac{1}{\sqrt{N}}\right) \tag{10}$$

Consider then Erdos Renyi networks, which comprise N nodes with a probability p, of having an edge between any pair of nodes. Thus, $\langle a_{ij}\rangle=\mu=p$ and $var(a_{ij})=\sigma^2=p(1-p)$. Therefore Eq. (10) may be rewritten as

$$\rho=\frac{\sum_{i,j}a_{i,j}}{N}+\frac{\sigma^2}{\mu}+O\left(\frac{1}{\sqrt{N}}\right)\approx Np+\frac{p(1-p)}{p}=(N-1)p+1=\langle k\rangle+1-p$$

for large N.

Hence, for an Erdos Renyi network, the CFW threshold is based on $R_0=(\beta/\delta)(\langle k\rangle+1-p)$. This coincides with the work of Dietz and May who, as we saw, argue that

$$R_0=\overline{R}(1+CV^2)=\frac{\beta Np}{\delta}\left(1+\frac{Np(1-p)}{N^2p^2}\right)=\frac{\beta}{\delta}((N-1)p+1)=\frac{\beta}{\delta}(\langle k\rangle+1-p)$$

It is of interest to examine *regular* random graphs in which each node has the same fixed number of edges k, but the edges are connected randomly between nodes. A simple calculation shows that the spectral radius of the adjacency matrix associated with any regular graph random or otherwise, must be $\rho=k$ (Restrepo et al., 2007). Thus the CWF threshold for a regular random network occurs at the point where $R_0=(\beta/\delta)k$ is unity. This threshold is in agreement with Dietz–May formula (taking $CV=0$) and deduced also by Kephart and White (1991).

Results are also available for the more general case of random networks having arbitrary degree distribution d_k. Chung et al. (2003) have shown that the spectral radius of the adjacency matrix associated with these networks is given by

$$\rho=\frac{\langle k^2\rangle}{\langle k\rangle}=\langle k\rangle(1+CV^2)$$

Thus the threshold condition for local stability as given by Eq. (9) becomes

$$\overline{R}(1+CV^2)<1$$

which is the Dietz–May formula given in Eq. (1).

4. Simulations of random networks:

We tested the above theoretical results by numerically simulating the spread of epidemics on Erdos Renyi networks ($N=50{,}000$) and Regular Random graphs ($N=100{,}000$). For each network studied, 1% of the nodes were randomly chosen and initially infected. Simulation then proceeded in steps of unit time increments. During each time step, an infected node was able to infect each of its neighbors with probability β. In addition, every infected node recovered with probability δ. In the case of $\delta=1$, infected nodes recovered in exactly one time-step. An infection attempt on an already infected node had no effect; however if a node recovers, it can be infected by its neighbors within the same time step (as simulated by CWF in Chakrabarti et al., 2008 and will be further discussed in the discussion).

Simulations were run for 50,000 time steps and were repeated 100 times with different initial conditions, for different values of $R_0=(\beta/\delta)\rho$.

Fig. 1 plots the proportion of infected nodes (i.e., the number of nodes infected divided by the total population N) at equilibrium as a function of R_0. One sees the presence of an epidemic threshold at $R_0=0.99$, while the CWF prediction is $R_0=1$ (see Eq. (9) above). The figure makes clear that the CWF threshold formula holds for both random networks and regular random networks, although the result has been known for decades in this context as given by the Dietz–May formula. We thus understand that the true importance of the CWF threshold formula concerns nonrandom graphs as treated in detail below.

5. Nonrandom graphs

5.1. One-dimensional chain

Consider regular graphs in which each node has exactly two neighbors. This forms a topology often referred to as a "one-dimensional chain" whereby each node-i is connected to node $i-1$ on the left and node $i+1$ on the right, for $i=1\ldots N$ (Fig. 2a).

For the particular case of a one-dimensional chain we show that it is possible to theoretically determine the threshold via the percolation theory. To achieve this we first have to show that the propagation of a virus in an infinite one-dimensional chain, where

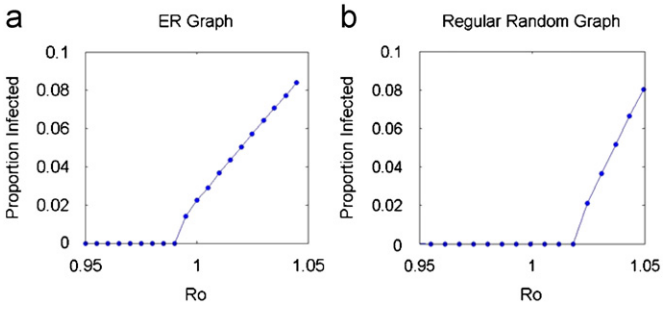

Fig. 1. (a) Random ER graph with average connectivity degree $\langle d\rangle=4$. (b) Regular random graph with fixed connectivity degree $\langle d\rangle=6$. The proportion of infected nodes at equilibrium is plotted for various values of $R_0=(\beta/\delta)\rho$ where $\delta=1$ is fixed and R_0 is determined by β. The CWF threshold prediction for the topology of the graph is simply $R_0=1$ in each figure. The simulated thresholds are in good agreement with the predictions.

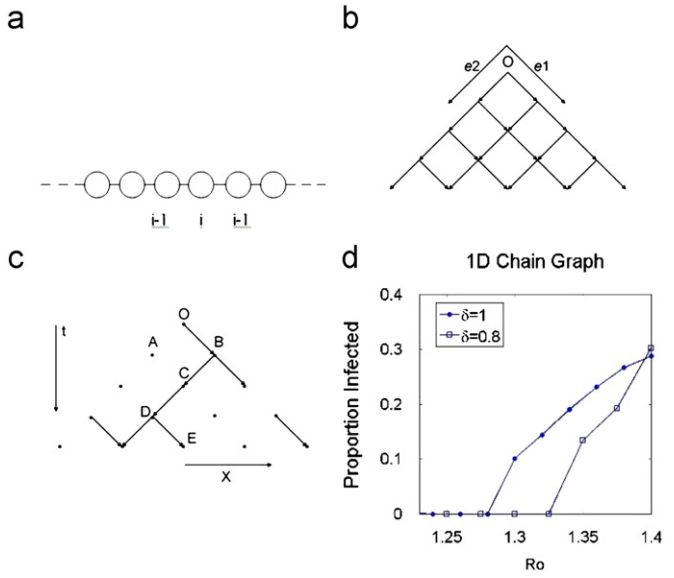

Fig. 2. (a) One-dimensional chain. (b) 2D directed lattice. (c) Percolation on directed lattice. (d) The proportion of infected nodes after 50,000 time steps plotted versus $R_0=(\beta/\delta)\rho$. Simulations were checked for both $\delta=1$ and $\delta=0.8$ and the value of R_0 is determined by β. Graph size is $N=15{,}000$. The CWF threshold prediction for the topology of the graph is $R_0=1$. The simulated threshold is $R_0\approx 1.29$ for $\delta=1$, as predicted in text.

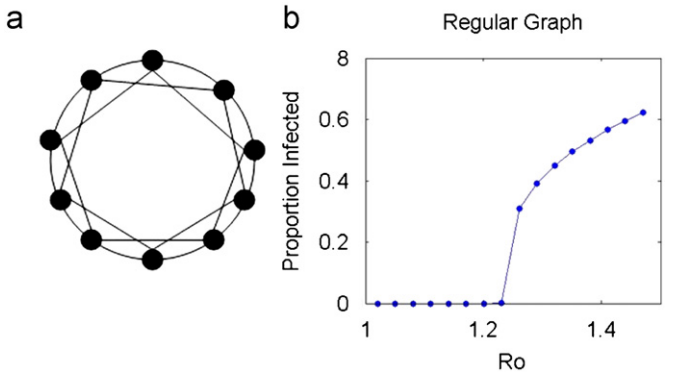

Fig. 3. (a) Regular graph $N=8$, $d=4$. The nodes are marked by black circles. (b) The proportion of infected nodes is plotted for various values of $R_0=(\beta/\delta)\rho$ where $\delta=1$ is fixed and R_0 is determined by β. The CWF threshold prediction for the topology of the graph is $R_0=1$. The simulated threshold is $R_0=1.23$.

the probability to recover in a time step is $\delta=1$, is analogous to directed bond percolation in an infinite 2D directed square lattice (Domany and Kinzel, 1981; Durrett, 1984). A directed square lattice is similar to a square lattice but differs in the fact that edges (bonds) always point in the positive direction of the axes as shown in Fig. 2b.

Our simulations for the 1D chain epidemic model assume $\delta=1$ and any infected node will recover after one time step. If we begin simulation by infecting the node O at the origin, then in the next time step the virus can exist only at the origin's neighbors. There is no possibility for the virus to exist at node O in the next time step. Observing Fig. 2b and c, one can define a time axis by $t=(e_1+e_2/\sqrt{2})$ where e_1 and e_2 are unit vectors pointing in the positive directions on the axes of the lattice. Moreover, a horizontal line can be defined by the coordinates $m_1e_1+m_2e_2$ where $m_1+m_2=M$, M being the line index. Each horizontal line will differ from its neighbors by integer units of t, i.e., assigned to a new time step. Notice that at $t=0$ only the origin exists while at $t=1$ only its neighbors exist and the coordinate of the origin node is vacant. Thus, one can deduce that the structure of the directed 2D lattice is adequate to describe the virus migration in a 1D chain for $\delta=1$. We now show that the behavior of the epidemic on a 1D chain with $\delta=1$ and a given β are analogous to a specific bond percolation on a 2D directed lattice.

In general, in bond percolation on a graph, all the bonds in the graph are processed in the following way. A bond will be removed with probability p or will be kept with probability $1-p$. Fig. 2c shows a directed square lattice after bond percolation. In the directed bond percolation model, a typical site O on a square lattice is chosen to be the origin of percolation, from which percolation can extend in a directed manner. If the base of the directed pyramid space can be reached from the origin following the bonds in a directed manner, percolation is said to occur.

In Fig. 2c the bond connecting origin O and node A was removed while the bond connecting the origin with node B was retained. This procedure is analogous to an infected origin that in the next time step will infect one neighbor and not the other. Indeed, when we set p to have the value of β the analogy is complete and the question whether a percolation exists is analogous to the question whether the epidemic is still alive. At a critical threshold p_c, long-range connectivity first appears and referred to as the percolation threshold as demonstrated in Fig. 2c. Here the base of the directed pyramid space can be reached from the origin following directed bonds, indicating that a chain of infection has percolated through the whole network.

The known threshold for 2D directed bond percolation is $p_c=0.6447$ (Essam et al., 1986; Jensen, 1996). Moreover, the spectral radius of the adjacency matrix of a one-dimensional chain is $\rho=2$ (since it is a regular graph where every node has 2 neighbors), Therefore we conclude that the epidemic threshold is $\beta_c=0.6447$ and $R_0=2(\beta/\delta)=2\times 0.6447\approx 1.29$. This known result from the literature is confirmed by the simulation results in Fig. 2d.

Since the epidemic threshold for the CWF model is $R_0=2(\beta/\delta)=1$, Fig. 2d shows that the simulation results are in contradiction with the CWF prediction with some 29–34% deviation. Hence we show both theoretically ($\delta=1$) and via simulations ($\delta=0.8$ and 1) that the mean field based CWF threshold is not appropriate for the 1-D chain. To show that the problem exists for other values of δ, and not just for the special case of $\delta=1$, Fig. 2d gives simulation results for $\delta=0.8$ that also contradict the CWF prediction.

5.2. *Regular graphs($d>2$):*

Examine now *regular graphs,* where it is assumed that each node in the graph has the same degree k ($k>2$). Here we consider regular graphs where each node is connected to its nearest neighbor and should thus be considered highly clustered regular lattices rather than random networks. We simulated the spread of a disease for regular graphs having degree $k=6$, and $N=15{,}000$ nodes for periods of 50,000 time steps. As shown in Fig. 3b the infection free equilibrium threshold was found at $R_0=1.23$ and deviated from the CWF predicted threshold ($R_0=1$) by some 23%. Hence again, the CWF prediction does not match the results obtained for simple *regular graphs*.

6. Star graphs

Another informative, yet simple example, is the star graph. It is defined as a central hub that connects to its N neighbors, but the neighbors are not connected to one another (see Fig. 4a). The hub is infected. Neglecting fluctuations and the discreteness of the system, the probability that the hub infects a neighbor, which

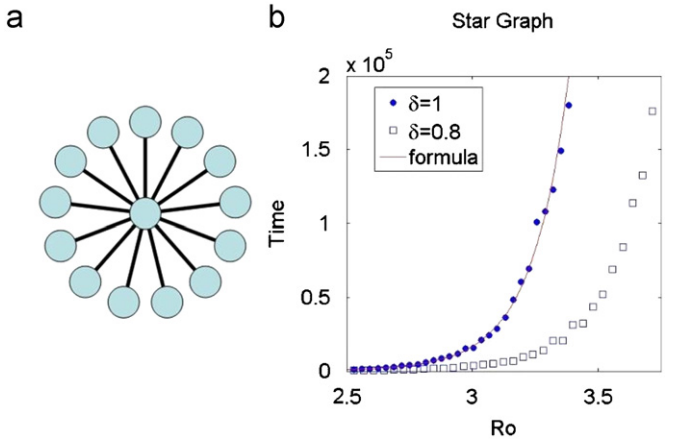

Fig. 4. (a) Star graph, $N=14$ neighbors and one hub. (b) The expected time steps a disease survives in a star graph according to Eq. (12). 100 simulations were executed for each $R_0=(\beta/\delta)\rho$, with the value of R_0 determined by β. Simulations were checked for both $\delta=1$ and $\delta=0.8$.

subsequently reinfects the hub, is β^2. Thus the mean number of secondary infections for an infected hub and a fixed $\delta=1$ is $R_0=N\beta_c^2$. The threshold based on R_0 is therefore calculated as $R_0=N\beta_c^2=1$, since the spectral radius ρ of the star graph is $\sqrt{N}$. The CWF threshold is thus $R_0=\sqrt{N}\beta=1$, and is identical to the one calculated before. However, we will show that for fixed $\delta=1$, not only is this threshold inaccurate, it does not exist.

We proceed by making an exact derivation of the disease average life time. Suppose the central hub is infected at time step $t=0$. The probability that the center node will not become infected by its neighbor node-i at time step $t=2$ is $(1-\beta^2)$. Thus the probability that the hub will not become infected from any of its neighbors at time step $t=2$ is $(1-\beta^2)^N$ and the probability that it will be infected at time step $t=2$ is $(1-(1-\beta^2)^N)$. Therefore the probability L_{2j} that the disease will survive to the $2j$th time step is

$$L_{2j}=(1-(1-\beta^2)^N)^j \tag{11}$$

Appendix A shows that the probability S_{2j+1} that the disease will survive to the $(2j+1)$th time steps but will not survive to the $2(j+1)$th time step is

$$S_{2k+1}=L_{2k}(1-\beta)^N[(1+\beta)^N-1]$$

and therefore the expected time steps a disease will survive in a star graph is (shown in Appendix A)

$$\langle j\rangle=\frac{2}{(1-\beta^2)^N}-\frac{1}{(1+\beta)^N}-1 \tag{12}$$

Our simulations of virus spread in $N=1000$ nodes star graph for different β values (represented in Fig. 4b), agree with Eq. (12). Eq. (12) shows that the star graph lacks an epidemic threshold for the special case of $\delta=1$ and $\beta<1$, since there are no phase transitions in the Equation. Thus, it appears that epidemic dynamics for the star graph lacks a threshold for $\beta<1$ and $\delta=1$.

In the absence of a more general analysis, it is difficult to conjecture about the star threshold, or its absence, for values of δ other than $\delta=1$. Instead, Fig. 4b provides a simulation for the case $\delta=0.8$ and it can be discerned that the scaling is indeed similar to the known threshold free case of $\delta=1$, notably without signs of a phase transition.

6.1. The independence assumption

The so-called independence assumption is a critical assumption used in deriving the CWF model. It assumes that the probabilities of the ith node.

$p_{j,t-1}$ ($j=1..N$ *neighbors of node* i) in Eq. (2) are independent of each other. The accurate way to formulate the disease dynamics is by examining all possible states. If the system has N nodes, each with a possible value of 1 (for infected) or 0 (for healthy), then the system has 2^N possible states. The probability of the system to be in state-k at time t is given by $P_{k,t}$. In general, $P_{k,t}$ has its dynamics in time as a function of β and δ.

The exact derivation of the probability that a node-i will not receive infections from its neighbors in the next time step should take into account the probability to be in that state and therefore Eq. (2) is more exactly written as

$$\zeta_{i,t}=\sum_k^{2^N}\left(\prod_{j\in M_i}(1-\beta n_{j,k})\right)P_{k,t-1} \tag{13}$$

where $P_{k,t-1}$ is the probability of the system to be at state-k at time step $t-1$, and M_i is the set of all neighbors of node-i. $n_{j,k}$ is the value (one or zero) of node-j at state-k.

As shown in Appendix B, CWF's model (Eq. (2)) is based on the mean field approximation

$$\left\langle\prod_{j\in M_i}n_j\right\rangle_t\approx\prod_{j\in M_i}\langle n_j\rangle_t \tag{14}$$

The brackets represent an averaging over all the 2^N possible states of the graph at time step t. In fact by application of this last approximation, Eq. (13) reduces directly to Eq. (2).

However, correlations between the neighbors may be significant for the propagation of infections through certain networks such as nonrandom regular graphs. In order to better understand the impact of correlations we examine two arbitrary nodes, say the first and second, and evaluate the term $\langle n_1n_2\rangle_t$ which is the product of the node values at the fixed time t averaged over many runs.

If there are no correlations we would expect $\langle n_1n_2\rangle_t=\langle n_1\rangle_t\langle n_2\rangle_t$. As an example we show in Fig. 5a and b the difference between the products $\langle n_1n_2\rangle_t$ and $\langle n_1\rangle_t\langle n_2\rangle_t$ in a 5000 nodes regular nonrandom graph (as depicted in Fig. 3a) and a regular random graph with connectivity degree $k=6$, $\beta=0.21$ and $\delta=1.5000$ simulations was made, which represent a sample of the 2^N possible states.

While for the regular random graph we find only 1% difference on average between the two sides of Eq. (14), the regular nonrandom graph yields a difference of 56%. Thus the independence assumption holds for random but not nonrandom regular graphs.

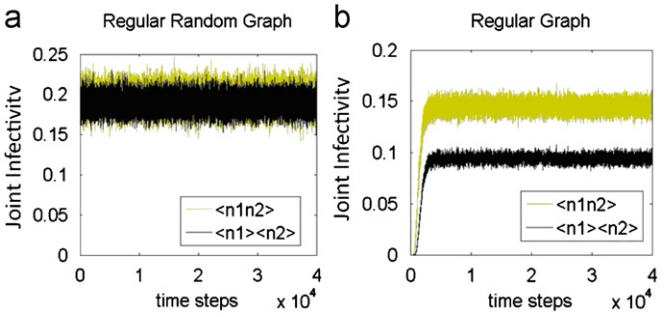

Fig. 5. Joint infectivity products $\langle n_1n_2\rangle$ and $\langle n_1\rangle\langle n_2\rangle$ calculated for nodes 1 and 2 are plotted versus time for (a) regular random graph and (b) regular nonrandom graph. The differences were measured after the system equilibrates.

7. Discussion

The dynamics of viruses as they spread through complex networks is a multidisciplinary research field that is currently receiving great attention (Pastor-Satorras and Vespignani, 2001; Madar et al., 2004; Berchenko et al., 2009; Lloyd and May, 2001). A general methodology that is often adopted to address the problem is to consider a simple parallel procedure that follows all newly infected nodes generated at each time step. This is the framework also adopted by Chakrabarti et al. (2008) when they proposed a general model and an epidemic threshold prediction applicable for complex network topologies. Unlike previous mean field models, CWF argue that their model gives correct threshold predictions for any arbitrary topology. In this article we have examined this claim in detail and presented various counter examples. Our examples range from modified random graphs to regular nonrandom graphs. In the case of the one-dimensional chain, it was possible to map the problem to an already existing and well-studied phenomenon: that of directed percolation in 2D. Known results from that field corroborate our simulations showing that the true threshold differs significantly from the CWF prediction. For the case of the star graph topology it was possible to show analytically that a threshold does not exist for fixed $\delta=1$. After a close investigation of the CWF model, a comparison with the exact Markov chain model made it possible to pinpoint exactly the mean-field approximation used by Chakrabarti et al. (2008) and deduce when this approximation is valid and when it fails. This relates closely to our section discussing the "independence assumption".

The analysis we report here is specifically relevant for the discrete time approach used in many network epidemic models including the work of CWF, but it is nevertheless worthwhile clarifying the underlying differences with continuous time approaches, which have received interest in recent years (Ganesh et al., 2005; Van Mieghem et al., 2009; Peyrard et al., 2008; Peyrard and Fran, 2005). For the discrete case, when modeling the spread of an epidemic through a network we may begin by making hypotheses about certain attributes of the nodes dynamics within a finite time interval. One can then attempt to write down the difference equations that realistically represent the dynamics of the system using a mean field approximation and probabilities. To achieve this, each node requires its own specific difference equation with time-step $\Delta t=1$ built in a manner that allows for the probability of multiple events to occur during that time interval. It may be easier to obtain analytic results by transforming the difference equations into more approachable differential equations. However, this needs careful treatment. Merely exchanging dt for Δt and rewriting the difference equation can be problematical. Transforming a difference equation into a differential one is a nontrivial problem. The analytic solution of a differential equation is not always representative of the "equivalent" difference equation.

Another possibility is to stay with the same features of the system but to let $\Delta t\to 0$. In this case several considerations need to be taken into account.

First, we do not deal with probabilities any more but with rates, i.e., single events have probability to happen within dt proportional to dt. One cannot just exchange the probabilities for rates of the same values.

Second, one should assess the logic of the model and equations as $\Delta t\to 0$. This requires making sure that multiple events cannot occur within infinitesimal dt so the difference equation is essentially linear in dt. In a correctly posed model, two events should not occur within the same infinitesimal dt otherwise we cannot appeal to the Poisson assumption implicit in the epidemic modeling approach.

Third, when simulating the epidemic dynamics on graph realizations, since $\Delta t\to 0$, the dynamics of an epidemic on a network is a Poisson process. Correct simulation of the system dynamics should make use of the Gillespie algorithm (Gillespie, 1977) that mimics the Poisson process. It is well known that the Gillespie algorithm may exhibit different results than an algorithm, which computes the changes of all of the nodes in parallel at each time step.

With these points in mind, let us return to the CWF model. In Eq. (3), which represents the CWF model, multiple events happen in the same dt, i.e., a node can recover and get infected once again within the same dt (last term RHS). Of course this is plausible only for a finite Δt but should not occur if $\Delta t\to 0$ since the underlying Poisson formulation only allows for one event per time step.

Similarly Eq. (2), which is also a key ingredient of the CWF model, explicitly defines $\zeta_{i,t}$ as the probability that a node-i will not receive any infection in the next time step. As can be seen in Eq. (2), the probability for this event is expressed as multiplication of probabilities of events, namely, that node-i does not get infected from all of its k-neighbors. Thus the formulation implicitly takes into account the likely possibility that node-i could be infected by more than one of its neighbors in the same time-unit Δt. This is precisely the reason one goes to the trouble of forming this product of probabilities of no event occurring. Hence Eq. (2) allows for the fact that multiple events may occur within the same time step, which can only have meaning when the time step is a finite Δt. Such a situation should not happen if $\Delta t\to 0$.

CWF claim that their analysis is valid when $\Delta t\to 0$ in Eq. (3), implying that it also holds for the continuous time case. Yet their difference equation takes into account the probability of multiple events within the same time interval. This prevents the development of a differential equation representing the system as $\Delta t\to 0$. Moreover, CWF simulate a parallel algorithm that follows the fate of all of the infected nodes at each iteration rather than an algorithm that mimics a Poisson process (*e.g.* the Gillespie algorithm, Gillespie, 1977) as should be applied when $\Delta t\to 0$.

Some work concerning the continuous time case has already appeared in the literature (Ganesh et al., 2005; Van Mieghem et al., 2009; Peyrard et al., 2008; Peyrard and Franc, 2005). However, the results of these modeling approaches are not directly equivalent to those found in the parallel discrete time procedure discussed here. For example, if β and δ are not small then $\beta\Delta t, \delta\Delta t\ll 1$ does not hold for $\Delta t\approx 1$, as required in a Poisson formulation, and the analogy between the rate equations and probability equations breaks down.

Finally we note that in Appendix C we explore the implications of the CWF scheme when used to make predictions about appropriate vaccinations and containment strategies in the event of an epidemic. In such cases, the CWF scheme yields misleading predictions.

Acknowledgements

We are grateful for support from the European Union FP7, the Israel Science Foundation and the Israel Ministry of Health.

Appendix A: star network

The probability the disease will survive to the $2j$th time step is

$$L_{2j}=P(\textit{time steps}=2j)=(1-(1-\beta^2)^N)^j \tag{A.1}$$

The probability the disease will survive to the $2j$th time steps but will not survive to the $(2j+1)$th time step

$$S_{2j}=L_{2j}(1-\beta)^N \tag{A.2}$$

and the probability the disease will survive to the $(2j+1)$th time steps but will not survive to the $2(j+1)$th time step is

$$S_{2j+1} = L_{2j}(\beta^N(1-\beta)^N + N\beta^{N-1}(1-\beta)(1-\beta)^{N-1} + \cdots + \binom{N}{1}\beta^1(1-\beta)^{N-1}(1-\beta)^1) = L_{2j}(1-\beta)^N\sum_{k=1}^{N}\binom{N}{k}\beta^k 1^{N-k}$$

$$= L_{2j}(1-\beta)^N[(1+\beta)^N - 1] \quad (A.3)$$

and hence the expected time steps the disease will survive are

$$\langle k\rangle = \sum_{k=0}^{\infty} 2kS_{2k} + \sum_{k=0}^{\infty}(2k+1)S_{2k+1} \quad (A.4)$$

By assigning the expressions of S_{2k+1} and S_{2k}, (A.4) turn to

$$\langle k\rangle = (1-\beta)^N\left[\sum_{k=0}^{\infty}(2kL_{2k}) + [(1+\beta)^N - 1]\sum_{k=0}^{\infty}((2k+1)L_{2k})\right]$$

$$= (1-\beta)^N\left[2(1+\beta)^N\sum_{k=0}^{\infty}kL_{2k} + ((1+\beta)^N - 1)\sum_{k=0}^{\infty}L_{2k}\right] \quad (A.5)$$

Since $\sum_{k=0}^{\infty}L_{2k}$ is a geometric series

$$\sum_{k=0}^{\infty}L_{2k} = \sum_{k=0}^{\infty}(1-(1-\beta^2)^N)^k = \frac{1}{1-1+(1-\beta^2)^N} = \frac{1}{(1-\beta^2)^N}$$

and

$$\sum_{k=0}^{\infty}kL_{2k} = \sum_{k=0}^{\infty}k(1-(1-\beta^2)^N)^k = \frac{1}{\ln\left[1-(1-\beta^2)^N\right]}\frac{\partial}{\partial\alpha}\left[\sum_{k=0}^{\infty}\left(1-(1-\beta^2)^N\right)^{\alpha k}\right]_{\alpha=1}$$

$$= \frac{1}{\ln\left[1-(1-\beta^2)^N\right]}\frac{\partial}{\partial\alpha}\left[\frac{1}{1-\left[1-(1-\beta^2)^N\right]^{\alpha}}\right]_{\alpha=1}$$

$$= \frac{\left(1-\left(1-\beta^2\right)^N\right)^{\alpha}}{\left[1-\left(1-\left(1-\beta^2\right)^N\right)^{\alpha}\right]^2}_{\alpha=1} = \frac{1-\left(1-\beta^2\right)^N}{\left(1-\beta^2\right)^{2N}}$$

Now, (A.5) can be written as

$$\langle k\rangle = (1-\beta)^N\left[2(1+\beta)^N\frac{(1-(1-\beta^2)^N)}{(1-\beta^2)^{2N}} + ((1+\beta)^N - 1)\frac{1}{(1-\beta^2)^N}\right]$$

$$\langle j\rangle = \frac{2}{(1-\beta^2)^N} - \frac{1}{(1+\beta)^N} - 1 \quad (A.6)$$

Appendix B: the independence assumption

The exact derivation of the probability that a node-i will not receive infections from its neighbors in the next time step is

$$\zeta_{i,t} = \sum_{k}^{2^N}\left(\prod_{j\in M_i}(1-\beta n_{j,k})\right)P_{k,t-1} \quad (B.1)$$

Eq. (B.1) can be revised into

$$\zeta_{i,t} = \sum_{k}\left(1-\beta\sum_{j\in M_i}n_{j,k} + \beta^2\sum_{j,q\in M, j\neq q}n_{j,k}n_{q,k} - \cdots + (-1)^M\beta^M\prod_{j\in M_i}n_{j,k}\right)P_{k,t-1}$$

$$= 1-\beta\sum_{j\in M_i}\langle n_j\rangle_{t-1} + \beta^2\sum_{j,q\in M, j\neq q}\langle n_j n_q\rangle_{t-1} - \cdots + (-1)^M\beta^M\left\langle\prod_{j\in M_i}n_j\right\rangle_{t-1} \quad (B.2)$$

where $M = |M_i|$ is the size of the neighbors set.

Using the approximation

$$\left\langle\prod_{j\in M_i}n_j\right\rangle_t \approx \prod_{j\in M_i}\langle n_j\rangle_t \quad (B.3)$$

where the product can be over a subset of the neighbors or all of them.

Eq. (B.2) turns into

$$\zeta_{i,t} \approx 1-\beta\sum_{j\in M_i}\langle n_j\rangle_{t-1} + \beta^2\sum_{j,q\in M, j\neq q}\langle n_j\rangle_{t-1}\langle n_q\rangle_{t-1} - \cdots + (-1)^M\beta^M\prod_{j\in M_i}\langle n_j\rangle_{t-1} \quad (B.4)$$

The average value of $\langle n_j\rangle_{t-1}$ is the probability of node-j to be infected at time $t-1$, which is equivalent to the CWF$p_{j,t-1}$ and therefore $\zeta_{i,t}$ is approximated to

$$\zeta_{i,t} \approx \prod_{j\in M_i}(1-\beta p_{j,t-1}) \quad (2)$$

Thus we understand that the mean field approximation suggested by Chakrabarti et al. was to approximate each averaged product into a product of averages (as seen in Eq. (B.3)).

Appendix C: vaccination strategy

One of the conclusions Chakrabarti et al. (2008) draw from the CWF model is that the most efficient way to immunize a network, is to vaccinate the nodes (*i.e.* subtract the nodes and their links from the graph) that will cause the most significant decrease in the spectral radius ρ of the adjacency matrix A. It is interesting to examine this proposition closer using their example as shown in Fig. 6.

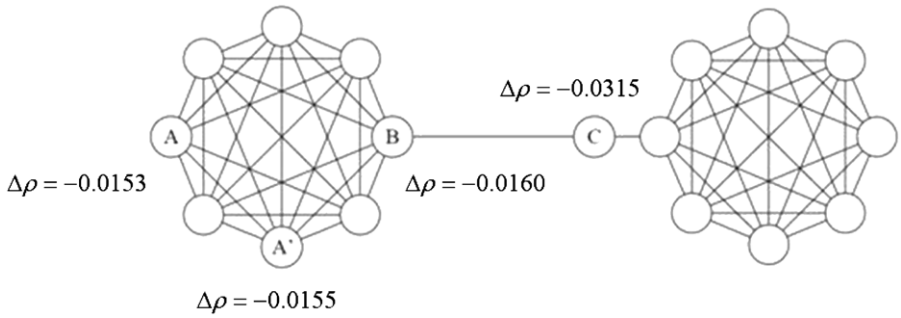

Fig. 6. The "bar-bell" graph discussed by Chakrabarti et al. (2008). Vaccinating any one of the nodes A, A′, B and C results in the change of the spectral radius ρ by $\Delta\rho$ demarcated. Since vaccination of node C is associated with largest $\Delta\rho = -0.0315$, the CWF method would suggest this to be the most effective strategy when only one node can be vaccinated.

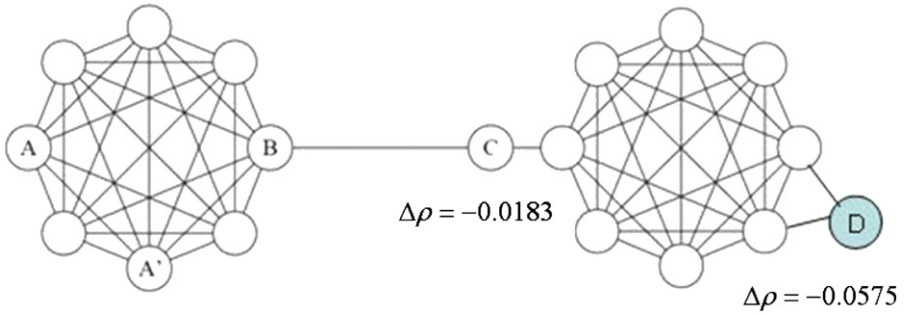

Fig. 7. Modification of the "bar-bell" graph. Node D is added to the right cluster. The effect of vaccination on ρ is noted next to nodes C and D.

In the case of a budget limited to the cost of the vaccination of only one node, arguments have been made in the past (Pastor-Satorras and Vespignani, 2002; Cohen et al., 2003) for vaccinating the node with the highest connectivity (several nodes are appropriate for this strategy in this example among them node A′). However, according to the CFW model the most efficient strategy would be to vaccinate node C since it achieves the maximum decrease of ρ ($\Delta\rho = -0.0315$).

Nevertheless there is a need to treat this conclusion with caution as the following example shows. Suppose a small change is made in the graph presented in Fig. 6, by adding a single node D to one of the clusters (Fig. 7).

This minor modification should, on the face of things, not change the fact that vaccinating node C is the most efficient strategy. But, at the same time, vaccinating the marginal node D, results in the enhanced reduction of ρ ($\Delta\rho = -0.0575$) in comparison to vaccinating node C ($\Delta\rho = -0.0183$). It is highly unlikely that this small network perturbation should change the vaccination policy to this degree, Thus, in our opinion, it is still an open question as to whether vaccinating nodes that cause the maximum reduction of ρ is in fact the most efficient one.

References

Albert, R., Barabási, A.L., 2002. Statistical mechanics of complex networks. Rev. Mod. Phys. 74, 47–97. doi:10.1103/RevModPhys.74.47.

Anderson, R.M., May, R.M., 1991. Infectious Diseases of Humans: Dynamics and Control. Oxford University Press, Oxford.

Aparacio, J.P., Pascual, M., 2007. Building epidemiological models from R_0: an implicit treatment of transmission in networks. Proc. R. Soc. B 274, 505–512. doi:10.1098/rspb.2006.0057.

Berchenko, Y., Artzy-Randrup, Y., Teicher, M., Stone, L., 2009. Emergence and size of the giant component in clustered random graphs with a given degree distribution. Phys. Rev. Lett. 102, 138701.

Chakrabarti, D., Wang, Y., Wang, C., Leskovec, J., Faloutsos, C., 2008. Epidemic thresholds in real networks. ACM Trans. Inf. Syst. Secur. 10 (4), 1–26.

Chung, F., Lu, L., Vu, V., 2003. Spectra of random graphs with given expected degrees. Proc. Natl. Acad. Sci. 100 (11), 6313–6318.

Cohen, R., ben-Avraham, D., Havlin, S., 2002. Percolation critical exponents in scale-free networks. Phys. Rev. E 66, 036113.

Cohen, R., Havlin, S., ben-Avraham, D., 2003. Efficient immunization strategies for computer networks and populations. Phys. Rev. Lett. 91, 247901.

Dietz, K., 1980. Models for vector-borne parasitic diseases. In: Barigozzi, C. (Ed.), Lecture Notes in Biomathematics. Proceedings of the Vito Volterra Symposium on Mathematical Models in Biology, vol. 39. Springer, Heidelberg, New York, pp. 264–277.

Domany, E., Kinzel, W., 1981. Directed percolation in 2 dimensions: numerical analysis and exact solution. Phys. Rev. Lett. 47, 5–8.

Durrett, R., 1984. Oriented percolation in two dimensions. Ann. Probab. 12, 999–1040.

Durrett, R., 2010. Some features of the spread of epidemics and information on a random graph. Proc. Natl. Acad. Sci. 107 (10), 4491–4498.

Essam, J., De'Bell, K., Adler, J., Bhatti, F.M., 1986. Analysis of extended series for bond percolation on the directed square lattice. Phys. Rev. B 33, 1982–1986.

Furedi, Z., Komlos, J., 1981. The eigenvalues of random symmetric matrices. Combinatorica 1 (3), 233–241.

Ganesh, A., Massoulie, L., Towsley, D., 2005. The effect of network topology on the spread of epidemics. In: Proceedings of the IEEE INFOCOM.

Gillespie, D.T., 1977. Exact stochastic simulation of coupled chemical reactions. J. Phys. Chem. 81 (25), 2340–2361.

Horn, R., Johnson, C., 1985. Matrix Analysis. Cambridge University Press, Cambridge.

Jensen, I., 1996. Temporally disordered bond percolation on the directed square lattice. Phys. Rev. Lett. 77, 4988.

Kenah, E., Robins, J.M., 2007. Second look at the spread of epidemics on networks. Phys. Rev. E 76, 036113.

Kephart, J.O., White, S.R., 1991. Directed-graph epidemiological models of computer viruses. In: Proceedings of the 1991 IEEE Computer Society Symposium on Research in Security and Privacy. pp. 343–359.

Lloyd, A.L., May, R.M., 2001. How viruses spread among computers and people. Science 292, 1316–1317.

Madar, N., Kalisky, T., Cohen, R., ben-Avraham, D., Havlin, S., 2004. Immunization and epidemic dynamics in complex networks. Eur. Phys. J. 38, 269–276B 38, 269–276.

May, R.M., Anderson, R.M., 1984. Spatial heterogeneity and the design of immunization programs. Math. Biosci. 72, 83–111.

May, R.M., Anderson, R.M., 1988. The transmission dynamics of human immunodeficiency virus (HIV) [and Discussion]. Philos. Trans. R. Soc. London B 321, 565–607. doi:10.1098/rstb.1988.0108.

May, R.M., Lloyd, A.L., 2001. Infection dynamics on scale-free networks. Phys. Rev. E 64, 066112.

Newman, M.E.J., 2002. Spread of epidemic disease on networks. Phys. Rev. E 66, 016128.

Parshani, R., Carmi, S., Havlin, S., 2010. Epidemic threshold for the susceptible–infectious–susceptible model on random networks. Phys. Rev. Lett. 104, 258701.

Pastor-Satorras, R., Vespignani, A., 2001. Epidemic spreading in scale free networks. Phys. Rev. Lett. 86, 3200–3203.

Pastor-Satorras, R., Vespignani, A., 2002. Immunization of complex networks. Phys. Rev. E 65, 036104.

Peyrard, N., Franc, A., 2005. Cluster variation approximations for a contact process living on a graph. Physica A: Stat. Mech. Appl. 358, 575–592.

Peyrard, N., Dieckmann, U., Franc, A., 2008. Long-range correlations improve understanding of the influence of network structure on contact dynamics. Theor. Popul. Biol. 73, 383–394.

Renshaw, E., 1991. Modelling Biological Populations in Time and Space. Cambridge University Press, Cambridge.

Restrepo, J.G., Ott, E., Hunt, B.R., 2007. Approximating the largest eigenvalue of network adjacency matrices. Phys. Rev. E 76, 056119.

Van Mieghem, P., Omic, J.S., Kooij, R.E., 2009. Virus spread in networks. IEEE/ACM Trans. Networking 17 (1), 1–14.