(This is a sample cover image for this issue. The actual cover is not yet available at this time.)

Journal of Theoretical Biology 315 (2012) 110–118

Contents lists available at SciVerse ScienceDirect

Journal of Theoretical Biology

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

Modelling the spread of diseases in clustered networks

Chai Molina [1], Lewi Stone *

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

HIGHLIGHTS

- We study epidemics spreading through clustered population networks.
- The reproductive number R_0 of epidemics is formulated on clustered networks.
- The formulation holds for networks of arbitrary degree distribution.
- The effects of clustering are often small and coupled with other network properties.
- The formulation can also take into account degree-dependent transmissibility.

ARTICLE INFO

Article history:
Received 31 January 2012
Received in revised form
25 August 2012
Accepted 28 August 2012
Available online 7 September 2012

Keywords:
Reproductive number
Transmissibility
Epidemic dynamics

ABSTRACT

It is now well appreciated that population structure can have a major impact on disease dynamics, outbreak sizes and epidemic thresholds. Indeed, on some networks, epidemics occur only for sufficiently high transmissibility, whereas in others (e.g. scale-free networks), no such threshold effect exists. While the effects of variability in connectivity are relatively well known, the effects of clustering in the population on disease dynamics are still debated. We develop a simple and intuitive model for calculating the reproductive number R_0 on clustered networks with arbitrary degree distribution. The model clearly shows that in general, clustering impedes epidemic spread; however, its effects are usually small and/or coupled with other topological properties of the network. The model is generalized to take into account degree-dependent transmissibility (e.g., relevant for disease vectors). The model is also used to easily rederive a known result concerning the formation of the giant component.

1. Introduction

Considerable significance has been given to the role of population structure in controlling the spread of diseases and to the impact of contact networks which shape the way infections pass through populations (Lloyd and May, 2001; Newman, 2003; Keeling and Eames, 2005). One of the most striking features of the interactions between individuals in real-world networks is the prevalence of triangle relationships. In social networks, these triangles reflect the high probability that two of a person's friends will also be friends. This probability is referred to as clustering, as discussed more formally below. Despite determined attempts, it is still unclear how, and to what extent, clustering affects epidemic spread (Britton et al., 2008; Miller, 2009; Newman, 2009; Volz et al., 2011); several highly influential studies sometimes report contradictory results. For instance Keeling (2005), using a correlation equation formulation, argued that clustering significantly reduces the basic reproductive number R_0 of a transmissible disease. However, Newman (2009) studied clustering in an extension of the configuration model and concluded, using a generating-function formalism, that higher clustering results in an increased R_0 and lowered epidemic threshold. Evidently, the effects of clustering are not yet clearly understood, despite persistent attempts at applying a variety of heavy mathematical machinery to the problem.

The most common mathematical tools used to model disease spread on networks are the generating function approach (Newman, 2003), pair approximations (Keeling and Eames, 2005), percolation theory (Cohen et al., 2002; Serrano and Boguñá, 2006) and spectral graph theory approaches (Diekmann and Heesterbeek, 2000). All of these methods have been highly successful in the study of epidemics on networks which contain little or no loops and are thus, in particular, unclustered. However, much less is known in the presence of clustering (Miller, 2009; Keeling, 2005; Newman, 2003; Trapman, 2007). In this study, we aim to find simple conditions for determining when a

* Corresponding author. Tel.: +972 3 6409806; fax: +972 3 6409403.
E-mail addresses: chai.molina@gmail.com (C. Molina), lewi@post.tau.ac.il (L. Stone).
[1] Present Address: Department of Mathematics and Statistics, McMaster University, Hamilton, Ontario, Canada.

0022-5193/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.jtbi.2012.08.036

transmissible disease invading a clustered population will result in an epidemic.

2. Network model

The population's contact structure may be visualized in terms of a network or graph. Each of the N nodes in the network represents an individual. An edge is placed between any pair of individuals who come into contact with one another. The resulting network topology is summarized by the symmetric adjacency matrix A whose elements $a_{ij}=1$ if there is a contact between the ith and jth individual, while $a_{ij}=0$ in the absence of such a contact. The degree of node i is the number of connections in which it is involved, $d_i=\deg(i)=\sum_{j=1}^{N} a_{ij}$. Here we assume that the graph is undirected, $a_{ij}=a_{ji}$.

The probability that a randomly chosen node x will have degree k is defined as $p_k=prob(\deg(x)=k)$, and the p_k are referred to as the network's degree distribution. The mean and the variance of the degree distribution are then $\langle k\rangle=\sum_{k=0}^{\infty} kp_k$ and $\mathrm{var}(k)=\langle k^2\rangle-\langle k\rangle^2$, respectively, where $\langle k^2\rangle=\sum_{k=0}^{\infty} k^2p_k$.

Note that the degree distribution of a node that is reached by a randomly chosen edge is different from p_k (Newman, 2003). For a graph with n_l edges, a particular node of degree k will be arrived at through a randomly chosen edge with probability $k/(2n_l)$. Thus, the probability of reaching a node of degree k via a random edge must be "boosted" by a factor of k, making it $\propto kp_k$ (see Fig. 1a).

The excess degree of a node that is reached by a random edge is defined as the number of neighbors of the node, excluding the neighbor at the other end of the edge by which it was arrived at, as shown in Fig. 1 (Feld, 1991; Molloy and Reed, 1995). For many diseases, a node that becomes infected often cannot transmit the disease back to the node by which it was infected. Thus only the excess degree nodes can potentially be infected. In light of the previous paragraph, the excess degree distribution is

$$q_k=\frac{(k+1)p_{k+1}}{\langle k\rangle}$$

(see Newman, 2003; Andersson, 1997; Cohen et al., 2000) and the mean excess degree is thus

$$\tilde{k}=\sum_{k=0}^{\infty} kq_k=\frac{1}{\langle k\rangle}\sum_{k=0}^{\infty} k(k+1)p_{k+1}$$
$$=\frac{1}{\langle k\rangle}\left(\sum_{k=0}^{\infty}(k+1)^2p_{k+1}-\sum_{k=0}^{\infty}(k+1)p_{k+1}\right)=\frac{\langle k^2\rangle-\langle k\rangle}{\langle k\rangle}. \quad (2.1)$$

The clustering coefficient (CC) is often defined as the proportion C of paths of length 2 (V-shapes) which have an additional edge connecting the two extreme nodes (see Fig. 1b). Put more simply, this quantifies the probability that two neighbors of a node are themselves neighbors. In the language of social networks, C is the probability that two of an individual's friends are also friends. In this sense, it is a measure of the prevalence of local connections. The clustering coefficient can easily be calculated from the adjacency matrix using the formula (Keeling and Eames, 2005)

$$C=\frac{tr\left(A^3\right)}{\sum_{i\in V} d_i(d_i-1)}. \quad (2.2)$$

It is not hard to show that the number of nodes in triangles, with multiplicity, is $\langle n_\Delta\rangle=\sum_{k=0}^{\infty} p_k\binom{k}{2}C$. Thus $\langle n_\Delta\rangle$ is three times the total number of triangles in the network. This probabilistic interpretation is relevant only when triangles are homogeneously distributed across the network. The probability of a connection between two nodes possessing a common neighbor may be correlated with the degree of this common neighbor, a situation which is referred to as degree-dependent clustering.

Lastly, in some networks the degrees of two neighbors are not independent random variables, but correlated (either negatively or positively). The Pearson correlation between the degrees of neighbors is called the assortativity of the network (Newman 2002). High assortativity has also been shown to have a marked effect on epidemic dynamics (e.g. Miller, 2009).

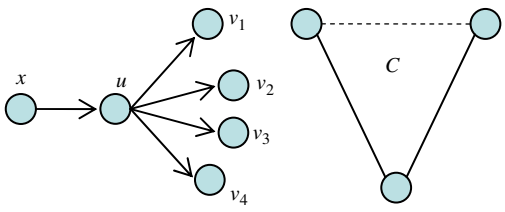

Fig. 1. (a) Left hand panel. Node u has degree 5 and excess degree 4. It is possible to arrive at u through any one of the 5 edges in which it participates. Here, we have arrived at u through the edge $x-u$. (b) Right hand panel. The clustering coefficient C is defined as the probability that two of a node's neighbors are connected.

2.1. The SIR epidemic model on a network:

We use the discrete-time SIR model formulation (Britton et al., 2008; Miller, 2009) to model a disease spreading along the edges of a graph. At each time step individuals may be classified as belonging to one of three states: susceptible, infective or recovered. If an individual is infective, each of his/her susceptible contacts has a probability T of becoming infected at each time step for the duration of infectiousness (it is assumed that the transmissibility is homogenous across the population; degree dependent transmission will be discussed later). Infective individuals recover after a deterministic time period of $\tau=1$.

A typical simulation begins by infecting a single randomly chosen node in an otherwise fully susceptible population. In the next time step, these initial infectives recover, but are given the opportunity to infect their neighbors with a probability T. Thus several new infectives might be generated in parallel in a single time step. This process of infection and recovery each time step continues, and for large enough values of T, can result in an epidemic whereby a large number of nodes in the network are at some point infected.

Note that for trees (graphs which contain no loops), it is trivial to generalize results for $\tau>1$, as only the transmission probability must be changed to reflect the multiple chances of infection: if the infection time is $\tau=n$, then the probability of an infected node not infecting its neighbor is $(1-T)^n$. It is thus possible to transform this scenario to a new model in which a single time-step corresponds to an infection time of n-time units and the equivalent transmission probability is $1-(1-T)^n$ rather than T. However, for clustered graphs it is not yet clear how to treat the case of $\tau>1$. Preliminary (unpublished) results show that there is some loss of generality in assuming $\tau=1$, which is a common practice in the literature (e.g. Miller, 2009).

2.2. Previous results for unclustered, uncorrelated networks

The criterion for whether or not an epidemic spreads through a network is often framed in terms of the basic reproductive number R_0. In the context of unstructured, random mixing models, R_0 gives the mean number of secondary infections caused by a typical single infective individual in an otherwise susceptible population. If $R_0>1$, the infection will, on average, be able to

reproduce itself and increase multiplicatively. If $R_0 < 1$ the epidemic will eventually die out.

When population structure is taken into account, an adjusted definition is needed. Now R_0 should not be calculated from the index case (the first infective), because the distribution of neighbors available for the index case (p_k) to infect is different from that for a typical infective (Miller, 2009; Aparacio and Pascual, 2007). In the absence of degree correlations and loops, the number of neighbors a typical infected node (other than the index case) will have available for infection is distributed according to q_k.

For unclustered random networks, such as those produced from the configuration model (CM; see Appendix I), it is well known that

$$R_0(T) = \frac{\langle k^2 \rangle - \langle k \rangle}{\langle k \rangle} T = \tilde{k} T \tag{2.3}$$

(see Newman, 2003; Andersson, 1997; Anderson and May, 1991; Cohen et al., 2000). This is simply the mean excess degree $\tilde{k}$ multiplied by the transmission probability T. The epidemic threshold, i.e. the critical transmission probability above which epidemics begin to occur, is found by setting $R_0 = 1$. Thus the critical transmission probability is

$$T_c = \frac{\langle k \rangle}{\langle k^2 \rangle - \langle k \rangle}, \tag{2.4}$$

which is the inverse of the mean excess degree.

Thus, it is interesting that for scale-free networks with $\gamma \in (2,3]$ (into which many real world—including social-networks—fall; see Albert and Barabási, 2002), $\langle k^2 \rangle$ diverges and so there is no critical transmission probability T_c. An epidemic outbreak can occur for any T (Pastor-Satorras and Vespignani, 2001; May and Lloyd, 2001). Random vaccination or health education campaigns, which act to effectively lower transmissibility, are thus not effective at achieving either herd immunity or the eradication of the disease in scale-free contact networks (Newman, 2003).

2.3. Previous results for calculating R_0 for clustered networks

Correcting for loops in describing epidemic dynamics requires taking into account multiple pathways resulting from these loops. Intuitively, increasing the loop length by one reduces the mean influence of these loops by a factor of T, since such a correction must factor in the probability of the disease passing along an additional edge. Thus, because transmission probabilities are less than unity, the corrections for loops decay exponentially with loop length. When relaxing the assumption that the network is a tree (i.e. completely devoid of loops), triangles are the first loops to correct for, because they are the smallest type of loop possible. Moreover when dealing with random networks, loops of large length rarely occur. Thus, as a first approximation, it is reasonable to just consider triangles and neglect higher order loops.

Because the set of nodes reached by a disease and the edges used by it are a sub-graph of the original network, the existence of a giant connected component (GCC) in the underlying network on which the epidemic spreads is a prerequisite for the development of an epidemic. Thus, examining the effect of clustering on the threshold for the formation of the GCC is a natural precursor for the study of its effects on epidemics. Berchenko et al. (2009) showed that clustering impedes the formation of the GCC and increases its threshold. In the analogy of epidemics (which occurs when $T = 1$) this implies that increasing clustering reduces R_0 and can prevent epidemic outbreaks.

However, isolating the effects of clustering on epidemic dynamics has proved difficult. To illustrate some of the confusion surrounding this issue, we cite a few illustrative cases. Newman (2003) examined a type of network model for which he observed a decrease in the epidemic threshold with increased clustering. However, Kiss and Green (2008) have shown that in this model, when clustering was changed, the degree distribution changed along with it. They further show that it was not clustering, but the increase in the variance of the degree distribution that was responsible for the lower epidemic threshold. Next, Newman (2009) studied a specific type of clustered random networks generalizing the CM. In these networks, triangles are isolated and do not share edges, in contrast for example to the configuration seen in Fig. 2. This leads to an almost tree-like graph, with triangles inserted randomly in the tree. Due to this "almost-loopless" structure, Newman was able to use a generating-function approach to model disease spread and conclude that clustering decreases the epidemic threshold. However, Miller (Miller, 2009) analyzed the same model and showed that it was not clustering, but in fact assortativity, built into the networks generated by their model, which was responsible for lowering the epidemic threshold. In fact, clustering serves to lower R_0 and increase the epidemic threshold.

Britton et al. (2008) also studied a particular class of random clustered graphs similar to Newman (2003) and likewise showed that if clustering is increased while keeping the mean degree constant, the epidemic threshold decreases. However, in their model, when clustering is increased, other factors of the degree distribution (such as the variance) are affected. Thus, as clustering is not varied in isolation, it is hard to attribute their observations to the effects of clustering alone.

Finally, Miller's (Miller, 2009) excellent paper is arguably the most successful treatment of epidemic dynamics on clustered networks to date. It describes a general formalism which takes both triangles and larger sub-structures into account, and finds

$$R_{0,1} = \frac{\langle k^2 - k \rangle}{\langle k \rangle} T - 2\frac{\langle n_\Delta \rangle}{\langle k \rangle} T^2 - \frac{\langle n_\square \rangle}{\langle k \rangle} T^3 + O\left(\frac{T^4}{\langle k \rangle}\right), \tag{2.5}$$

where $\langle n_\Delta \rangle$ and $\langle n_\square \rangle$ denote the average number of triangles and squares each node may be found in, respectively. To avoid unnecessary technicalities, we do not explain the precise definition of $R_{0,1}$ above, but it is shown in Miller (2009) that this is a good approximation of R_0. It is clear from this formula that in general, clustering serves to decrease R_0.

However, our approach, described in Section 3, has a number of advantages. The first is that it is readily extendable to degree dependent transmissibility and other scenarios (such as the extension of the CM developed by Newman (2009) and Miller

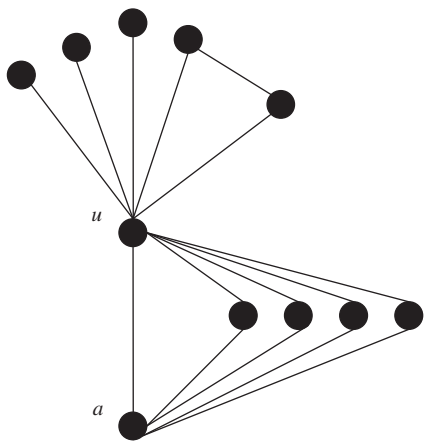

Fig. 2. If u was reached by edge $a-u$, its free excess degree is the number of its neighbors excluding a and any of a's neighbors, that is, 5.

(2009), mentioned earlier, in which triangles do not share edges). Second, and no less important, it allows an intuitive biological interpretation of the first two elements of the series obtained by Miller (2009), (Eq. (2.5)). Furthermore, the other coefficients in that expansion are in any case hard to estimate *a-priori* for different types of random networks and experimentally for large empirical ones.

3. Results

3.1. Constant transmission probability

Our goal is to estimate R_0, the reproductive number, in contact networks, taking into account triangles. In the following, it is assumed that clustering is not degree dependent and that degrees are uncorrelated. The necessity of these assumptions are touched on later.

Consider an arbitrary node u with $\deg(u)=k+1$ which has been infected in the previous iteration by an infecting neighboring node a, as in Fig. 3. We now calculate $R_0(T,k)$, the expected number of neighbors u will infect in the next time step.

Neighbors of u are either neighbors of a as well, or not. There are, then, k paths of length 2 starting from a, of which u is the middle node (e.g., $a-u-v$). Recall that C is the probability of the end-points of such a path being connected in a triangle (e.g., $a-u-w-a$). The mean number of neighbors shared by a and u in triangular formations is kC, and the number of remaining neighbors of u is $(1-C)k$. In fact, node a had an opportunity to infect the shared neighbors (such as w in Fig. 3) in the previous time step (in which u was also infected), and the mean number of successful infections of shared neighbors by a is kcT. Then, the number of nodes which remain available for infection by u is $k-kCT=k(1-CT)$. The mean number of successful infections by u is thus

$$R_0(T,k)=kT(1-CT). \tag{3.1}$$

Since u was reached via an edge connected to a random node, its excess degree distribution is given by q_k. Thus, the mean reproductive number is

$$\begin{aligned} R_0(T)=\langle R_0(T,k)\rangle &= \sum_{k=0}^{\infty} q_k\left(kT-kCT^2\right) \\ &= (1-CT)T\sum_{k=0}^{\infty} q_k k=(1-CT)T\frac{\left\langle k^2-k\right\rangle}{\langle k\rangle} \\ &= (1-CT)T\left(\left(1+CV^2\right)\langle k\rangle-1\right) \end{aligned} \tag{3.2}$$

where $CV=\sigma/\mu$ is the coefficient of variation of the degree distribution. (One should note, however, that if degree correlations are present, using $q_k=((k+1)/\langle k\rangle)p_k$ for the excess degree distribution of u may not be appropriate.)

Recalling that $R_0^*(T)=\tilde{k}T$ is the reproductive number for unclustered networks with an identical degree distribution, we obtain the simple correction

$$R_0(T)=(1-CT)R_0^*(T). \tag{3.3}$$

The epidemic threshold is obtained by equating $R_0=1$, to get

$$T_c=\frac{1-\sqrt{1-4(C/\tilde{k})}}{2C}. \tag{3.4}$$

From Eq. (3.3), it may already be seen that clustering has very little effect on the epidemic threshold, as the product CT, which determines the size of the correction to R_0 (in proportion to the unclustered case), is usually much smaller than 1 (both C and T being smaller than 1). However, the general direction of the correction is clear: clustering reduces R_0 and so increases the epidemic threshold. This echoes the results of Miller (2009) and Keeling (2005).

It should be noted that clustering may also cause triangles to exist between neighbors of u. The calculation above is exact when these triangles are between those neighbors of u which are not neighbors of a (green edge in Fig. 3). However, this treatment neglects the effect of squares (or higher-order loops) in the graph, which may be created by two (or more) triangles sharing an edge. For instance, denote one of u's neighbors by v. If a has a neighbor, say w, which is also connected to v (forming a triangle), then a may infect both u and w at the same time, which may in turn lead to concurrent infection of v by both u and w. This scenario is depicted in Fig. 3. Since we have assumed that there are no squares in the network which are not created by triangles, this can occur only when clustering causes two triangles to share an edge (as in Fig. 3).

We have formulated a corrected R_0 which takes this possibility into account and have simulated the process in order to assess the significance of this effect. Results show that it is negligible (theoretically, this is a term of order T^4). It should be remarked that Miller (2009) neglects such concurrent infections as well.

Interestingly, our prediction for constant transmission probability is equivalent to the first two elements in the expansion given by Miller (2009) and given in Eq. (2.5) above. The similarity can be seen through the following rearrangements, Since

$$\sum_{k=0}^{\infty} q_k k=\frac{\left\langle k^2-k\right\rangle}{\langle k\rangle}=\frac{1}{\langle k\rangle}\sum_{k=0}^{\infty} p_k\left(k^2-k\right)=\frac{2}{\langle k\rangle}\sum_{k=0}^{\infty} p_k\binom{k}{2}$$

and $\langle n_\Delta\rangle=\sum_{k=0}^{\infty} p_k\binom{k}{2}C$,

it follows that

$$R_0=T\sum_{k=0}^{\infty} q_k k-CT^2\sum_{k=0}^{\infty} q_k k=T\frac{\left\langle k^2-k\right\rangle}{\langle k\rangle}-T^2\frac{2\langle n_\Delta\rangle}{\langle k\rangle}. \tag{3.5}$$

Eq. (4.2) also has a simple intuitive interpretation that may be found by rewriting it as $R_0(T)=\left(\tilde{k}-C\tilde{k}T\right)T$. The first term in the

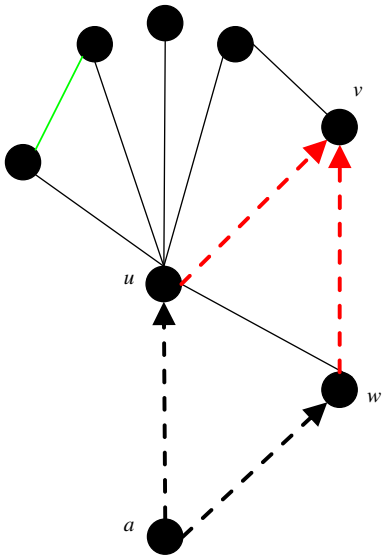

Fig. 3. w is a neighbor of both u and a, and is also connected to v, a neighbor of u. If w is infected at the same time as u (black dashed arrows), then it is possible that both u and w attempt to infect v (red arrows). In that case, it is impossible to attribute an infection of v to either w or u alone. Arrows represent directions of infection. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

brackets gives $R_0^*(T)$, the prediction for unclustered networks. The second term is the correction due to clustering. To return to the logic behind our derivation, in order to determine the number of neighbors a typical node can infect (u in Figs. 2 and 3), one starts with the mean excess degree, but subtracts those nodes which are neighbors of its predecessor (a in Figs. 2 and 3), and were infected by it. The fraction of neighbors of u shared with a is C. Thus, when u attempts to transmit the disease onwards, the number of u's neighbors previously infected by a is $C\tilde{k}T$, which is the correction term subtracted in the parentheses.

3.2. Extension to degree dependent transmission probability

As stated, this method is easily adjusted to account for various real-world scenarios, for example, when the transmissibility is degree dependent (Olinky and Stone, 2004). However, we cannot simply plug this dependency into the equation for the reproductive ratio, because the transmission probabilities and the degrees involved refer to different nodes. Instead, going back to the process by which the formula for R_0 (Eqs. (4.1) and (4.2)) was developed, we write

$$R_0(T,C)=\sum_{k_1,k_2=0}^{\infty} k_2T(k_2+1)(1-CT(k_1+1))q(k_1,k_2), \tag{3.6}$$

where $q(k_1,k_2)$ is the joint excess degree distribution of nodes connected by an edge; k_1 and k_2 refer to nodes a and u, respectively, and $T(k)$ is the degree dependent transmission probability. Thus, $q(k_1,k_2)$ is the probability that a's degree is k_1 and u's degree is k_2. Here, we see that any degree assortativity present in the network will be reflected in R_0, due to the dependence on $q(k_1,k_2)$.

However, when assortativity is negligible, it is possible to take $q(k_1,k_2)=q_{k_1}q_{k_2}$, which gives the somewhat simpler expression

$$R_0(T,C)=\left(1-C\sum_{k_1=0}^{\infty} q_{k_1}T(k_1+1)\right)\sum_{k_2=1}^{\infty} q_{k_2}T(k_2+1)k_2. \tag{3.7}$$

This allows us to compare R_0 for degree-dependent and constant transmission probabilities. For example, take $T(k)=\alpha/k$ for $k\geq 1$ (and $T(0)=0$), which may be realistic for macro-parasite infections and worms transmitted over peer-to-peer computer networks (Olinky and Stone, 2004). In such a case, where the amount of pathogen an infected host can spread is limited, the probability of infecting a susceptible neighbor decreases with the number of neighbors available to receive the pathogen. Having the transmission probability scale inversely with the number of neighbors in this context is natural.

Note that in order for $T(k)\leq 1$ to hold, $\alpha\leq \min\{k|p_k>0\}=k_{\min}$ (α must be no larger than the minimal degree in the network). It is easy to see that the mean number of infected neighbors in this case is always smaller than $\alpha(1-(1/k))<k_{\min}$. Thus, if $k_{\min}=1$, the model is never in the epidemic regime. This is, however, something of an artifact of the particular formula for T. Consider, for instance, a disease with transmission probability $T(k)=2/k$ spreading on a network with minimal degree $k_{\min}=3$. Clearly, adding one node with degree 1 to this network invalidates the entire transmission probability scheme (because for this added node, $T(1)=2>1$, meaning that $T(k)$ is no longer a transmission probability) which makes this setup highly pathological.

Eqs. (3.7) and (3.2) allow us to compare R_0 in this transmission probability scheme with the constant probability model (with $T=T_0$). We choose to fix α so that the mean transmission probability of nodes reached by a random edge is equal to the fixed transmission probability, T_0:

$$T_0=\sum_{k=0}^{\infty} q_kT(k+1)=\alpha\sum_{k=0}^{\infty} q_k\frac{1}{k+1} \quad \text{or} \quad \alpha=\frac{T_0}{\sum_{k=0}^{\infty} q_k(1/(k+1))}.$$

Note that we use q_k here because we are assuming that infected nodes are reached from following random edges (see explanation of excess degree distribution above).

As all the infections but the first case originate via edges from nodes which were themselves reached by other edges, using q_k as above is more appropriate than equating T_0 with the transmission probability averaged over nodes chosen at random i.e., $\sum_{k=0}^{\infty} p_kT(k)$.

Plugging this back into R_0 gives $R_0^{\text{fix}}(T,C)=(1-CT_0)T_0\sum_{k=1}^{\infty} q_kk$ for the fixed transmission probability, and $R_0^{\text{var}}(T_0,C)=(1-CT_0)T_0\left(\sum_{k_2=1}^{\infty} q_{k_2}(k_2/(k_2+1))/\sum_{k_1=1}^{\infty} q_{k_1}/(k_1+1)\right)$ for the degree dependent (variable) transmission probability. Their ratio is then

$$Q=\frac{R_0^{\text{var}}}{R_0^{\text{fix}}}=\frac{\sum_{k_2=1}^{\infty} q_{k_2}(k_2/(k_2+1))}{\sum_{k_1=0}^{\infty}(q_{k_1}/(k_1+1))\sum_{k_2=1}^{\infty} q_{k_2}k_2}. \tag{3.8}$$

Using $q_k=(k+1)p_{k+1}/\langle k\rangle$ gives

$$Q=\frac{\langle k\rangle\sum_{k_2=1}^{\infty} p_{k_2+1}k_2}{\sum_{k_1=1}^{\infty} p_{k+1}\sum_{k_2=1}^{\infty} p_{k_2+1}(k_2+1)k_2}. \tag{3.9}$$

Note that

$$\sum_{k_2=1}^{\infty} p_{k_2+1}k_2=\sum_{k_2=1}^{\infty} p_{k_2+1}(k_2+1)-\sum_{k_2=1}^{\infty} p_{k_2+1}=\sum_{k=2}^{\infty} p_kk-\sum_{k=2}^{\infty} p_k$$
$$=\langle k\rangle-p_1-(1-p_0-p_1)=\langle k\rangle+p_0-1$$
$$\sum_{k_2=1}^{\infty} p_{k_2+1}(k_2+1)k_2=\sum_{k=2}^{\infty} p_kk^2-\sum_{k=2}^{\infty} p_kk=\left\langle k^2\right\rangle-\left(p_1+\sum_{k=2}^{\infty} p_kk\right)$$
$$=\left\langle k^2\right\rangle-\langle k\rangle\sum_{k_1=0}^{\infty} p_{k+1}=1-p_0.$$

Hence Eq. (3.9) simplifies to $Q=\langle k\rangle((\langle k\rangle+p_0-1)/\left(\left\langle k^2\right\rangle-\langle k\rangle\right)(1-p_0))$. However, for all intents and purposes, nodes with degree 0 do not participate in the network or the epidemic dynamics, so we assume $p_0=0$. Thus

$$Q=\frac{\langle k\rangle^2-\langle k\rangle}{\left\langle k^2\right\rangle-\langle k\rangle}. \tag{3.10}$$

But, $\text{var}(k)=\left\langle k^2\right\rangle-\langle k\rangle^2\geq 0$, $Q=\left(\left(\langle k\rangle^2-\langle k\rangle\right)/\left(\left\langle k^2\right\rangle-\langle k\rangle\right)\right)\leq 1$ and equality is achieved only when $\text{var}(k)=0$, i.e. for regular graphs. It can be concluded that R_0 is greater for a constant transmission probability than for $T(k)=\alpha/k$, where the mean transmission probability (with respect to q_k) is equal to T_0.

This is not so surprising, as this type of degree dependence lowers transmission probability for high-degree nodes and raises it for low-degree nodes. Thus, the high-degree nodes' ability to act as hubs is reduced for the degree-dependent case. The importance of hubs for epidemic dynamics is already well known, so effects of reducing their disease-spreading abilities while trying to make up for it in the low-degree nodes were not unforeseeable. Nonetheless, it is interesting that this intuition can be backed up using the formulation presented here.

Moreover, recall that we used $\alpha=T_0/\sum_{k=0}^{\infty} q_k(1/(k+1))$ to equate T_0 with the mean transmission probability of a node reached by a random edge. However, had we equated $T_0=\sum_{k=0}^{\infty} p_kT(k)$ (so as to equate T_0 with the mean transmission probability of a randomly chosen node), we would have used $\tilde{\alpha}=T_0/\sum_{k=0}^{\infty} p_k(1/k)$ instead of α. From Jensen's inequality (see e.g. Ross (2002)), one can show that $\tilde{\alpha}<\alpha$. So using α rather than $\tilde{\alpha}$ leads to a higher overall mean transmission probability, for a randomly chosen node. That this does not compensate for the

discounting of the transmissibility of high-degree nodes is not immediately clear without this calculation.

4. Discussion

4.1. Constant transmission probability

We have tested the approximation for R_0 (Eq. (4.3)) on synthetic and real-world networks. Synthetic unclustered networks were generated by using the CM for randomized regular and Erdős-Rényi (ER) random networks, and the BA algorithm (Barabási and Albert, 1999) for scale-free networks. Their clustering was then modified using a Monte Carlo approach as described in Appendix I. Details on the simulation procedure and measurement of R_0 are found in Appendix II.

The potential for degree correlations (assortativity) to significantly alter R_0 from the uncorrelated prediction is known (e.g. Miller (2009)). However, there is no satisfactory method for creating clustered, uncorrelated networks with prescribed degree distributions. The method described in Appendix II results in a fairly low level of assortativity, and its effects can barely be felt. For example, the mean assortativity of the clustered ER networks used was $r=0.05\pm 6.3475\times 10^{-4}$.

An effective way to factor degree correlations out of the simulations is to use regular graphs, in which all the nodes have equal degrees. We generated 40 randomized regular graphs with mean $N=300$ and mean clustering $C=0.34$ and measured the average R_0 for this set of graphs. As seen in Fig. 4, the formula for R_0 (Eq. (4.2)) fits the simulations for different values of transmission probabilities T. The prediction is also in good agreement with the simulations of 40 clustered BA scale-free networks (Fig. 4).

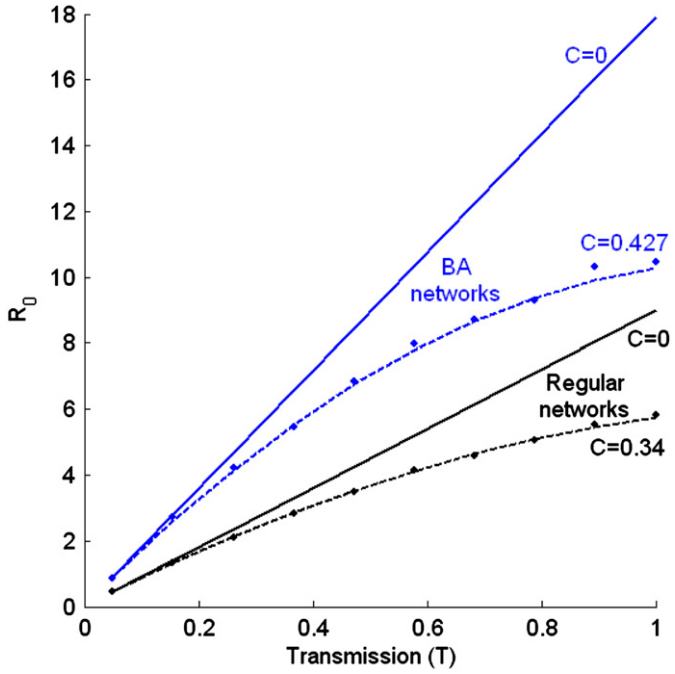

Fig. 4. Mean R_0 vs. transmission probability T, for clustered BA (blue) and clustered regular networks (black). Solid lines represent the classical prediction for $C=0$. Dashed lines represent theoretical predictions using Eq. (3.2). Solid circles (●) represent averaged numerical results obtained from simulations on clustered graphs as described in Appendices I and II. Parameters: $N=300$ and $\langle k\rangle=10$. $\langle C\rangle=0.427$ for the BA networks and $\langle C\rangle=0.34$ for the regular networks. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

We have also tested the prediction on an empirical social network from an American high school (Salathé et al., 2010). This social network data represents close proximity interactions (crucial to the transmission of many diseases) amongst school students and was collected using a wireless sensor network during 1 day in January 2010. The network could well portray the type of population contact structure relevant for disease dynamics, and interestingly it is characterized by unusually high clustering with $C=0.5$. Here, too, there is good agreement between theory and simulations (see Fig. 5). The high reproductive numbers reached in these simulations are noteworthy, and arise from the large excess degree of $\tilde{k}=334.74$. (We note that the mean degree of the graph is also large with $\langle k\rangle=299.8$).

4.2. Degree dependent transmission probability

Simulations have also been used to test the theoretical results for the degree dependent transmission scheme, $T(k)=\alpha/k$. Because of the explicit dependence of the degree-dependent transmission model on the joint excess degree distribution $q(k_1,k_2)$ (Eq. (3.6)), we can test the effects of moderate degree correlations for this transmission model. We have thus calculated the predicted R_0 neglecting degree correlations (using Eq. (3.7)), and devised a semi-empirical scheme to incorporate their effects. Starting out with $R_0(T,C)=\sum_{k_1,k_2=0}^{\infty}k_2T(k_2+1)(1-CT(k_1+1))q(k_1,k_2)$ (Eq. (3.6)), we partition the formula and substitute $T(k)=\alpha/k$ to obtain

$$\begin{aligned}R_0(T,C)&=\sum_{k_1,k_2=0}^{\infty}k_2T(k_2+1)q(k_1,k_2)\\&\quad-\sum_{k_1,k_2=0}^{\infty}CT(k_1+1)k_2T(k_2+1)q(k_1,k_2)\\&=\alpha\sum_{k_2=0}^{\infty}\frac{k_2}{k_2+1}q_{k_2}-\alpha^2C\sum_{k_1,k_2=0}^{\infty}\frac{1}{k_1+1}\frac{k_2}{k_2+1}q(k_1,k_2)\end{aligned}\qquad(4.1)$$

This separates the parameter governing the disease transmission (α) and the clustering C from factors governed by the degree distribution of the graph. It is now possible to simply empirically sample the joint excess degree probability distribution of the graphs to calculate R_0.

The discrepancy between the two modes of prediction was minute (on the order of 10^{-3}%), and thus only the prediction taking into account degree correlations (using Eq. (4.1)) was

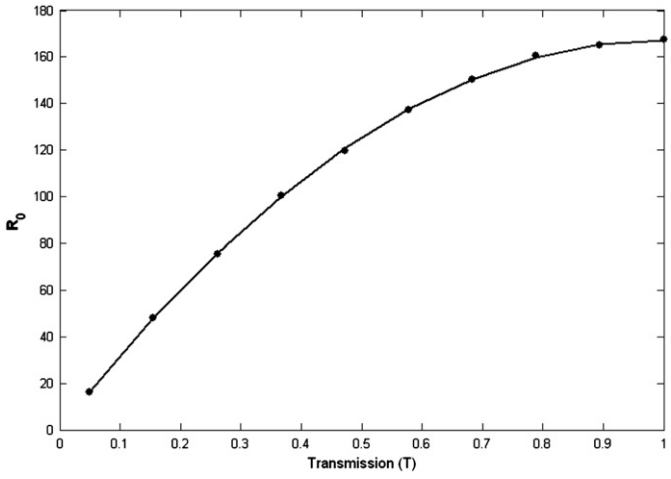

Fig. 5. Mean R_0 vs. transmission probability for an empirical contact network (Salathé et al., 2010) from an American high-school. Solid circles (●) represent simulation results and solid line represents theoretical prediction; $N=788$, $\langle k\rangle=299.8$, $C=0.5$ and $r=0.05$.

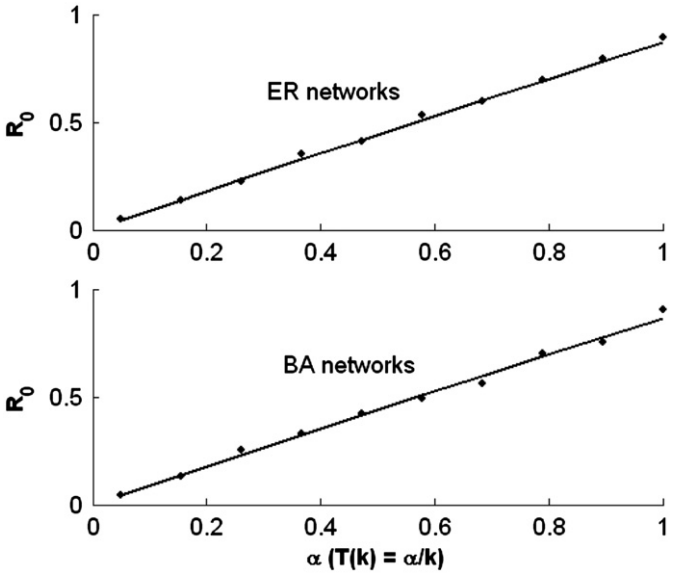

Fig. 6. Mean R_0 vs. α for the degree dependent transmission probability scheme $T(k)=\alpha/k$ on ER networks (upper panel) and BA networks (lower panel). Solid lines represent theoretical predictions of Eq. (4.1). Solid circles (♦) represent averaged numerical results obtained from simulations on clustered graphs as described in Appendices I and II. Parameters: $N=300$ and $\langle k\rangle=10$. $\langle C\rangle=0.427$ for the BA networks and $\langle C\rangle=0.341$ for the ER networks.

plotted (Fig. 6). In fact, in the 40 BA networks tested, the mean assortativity was $r=0.238\pm0.002$, and still, the prediction neglecting assortativity was almost identical to the semi-empirical one, and both were accurate. Since the difference between the two modes of prediction was negligible, in this particular transmission model, degree correlations appear to have little effect. Also, it is interesting to note the similarity in the plots of R_0 for the different networks in the degree-dependent transmission scheme. It seems that the effects of using this particular transmission scheme almost completely obscure the effects of network topology.

It is tempting to try to account for degree-dependent clustering using this scheme. However, simply plugging $C(k)$ into the equations fails to take into account the possibility that degree correlations are caused by the unequal distribution of triangles between the different degree nodes. In networks with degree dependent clustering, such as those used by Newman (2009) and Miller (2009), the node classes with higher clustering will necessarily be connected more often amongst themselves than with low-clustering nodes As stated, assortativity has an effect on the epidemic dynamics, typically increasing R_0 and decreasing the epidemic threshold. This increase in R_0 may be large enough to completely negate the decrease in R_0 caused by clustering.

4.3. Epidemic threshold

Using Eqs. (3.2) and (3.4) it is possible to explore the dependence of R_0 and T_c on clustering. From Eq. (3.4), it can be seen that if $\tilde{k}<4$, as clustering is increased, the value $\sqrt{1-4(C/\tilde{k})}$ and thus T_c can become imaginary. This transition reflects the fact that at this point, $R_0<1$ for any T. Given $\tilde{k}$ and C it is possible to calculate the maximum reproductive number, $R_0^{\max}$.

This maximum is attained at $T_{\max}=1/2C$ when $C>0.5$ and is equal to $R_0^{\max}=\tilde{k}/4C$. Of course, when $C<0.5$ the maximum is attained at $T_{\max}=1$ and is $R_0^{\max}=(1-C)\tilde{k}$. This maximum is not necessarily attained at $T=1$, since when both clustering and transmissibility are high, there is a significant reduction in the neighbors available for infection due to their already having been infected earlier (see Figs. 2 and 3).

It follows that for a given degree distribution, increasing clustering decreases $R_0^{\max}$. Furthermore, as we just saw, when $1<\tilde{k}<4$ (or $2<\left(\langle k^2\rangle/\langle k\rangle\right)<5$), even though for some values of C, for some transmission probabilities T, $R_0(T,C)>1$, one can choose C so that $R_0<1$ for any T. Thus, given a degree distribution for which $2<\left(\langle k^2\rangle/\langle k\rangle\right)<5$, modifying the clustering coefficient can in principle tip the scales so that $R_0<1$, regardless of T. However, when $\tilde{k}>4$, we are in the more complicated regime where for any value of C, there is a range of T for which $R_0>1$.

Moreover, Eqs. (3.3) and (3.4) show that clustering cannot "manufacture" an epidemic threshold where none exists due to an infinite 2nd moment of the degree distribution, as is the case with scale-free networks.

4.4. Relation to the giant component

It is well-known that for unclustered random graphs with a specified degree distribution, the condition for the existence of a giant component (GC) is

$$\frac{\langle k^2\rangle-\langle k\rangle}{\langle k\rangle}>1 \tag{4.2}$$

(see e.g. Newman, 2003). In their analysis of the emergence of the giant component in clustered random graphs, Berchenko et al. (2009) show that this condition scales as $1/(1-C)$; that is, a GC will form if

$$\frac{\langle k^2\rangle-\langle k\rangle}{\langle k\rangle}>\frac{1}{1-C}. \tag{4.3}$$

The latter is clearly a significant correction to the classical prediction (Eq. (4.2)) that neglects clustering. In contrast, the effect of clustering on the epidemic threshold appears to be very small indeed as shown in Eq. (3.3). Thus, an interesting question arises: what is the reason for the difference in the apparent magnitude of the effect of clustering on the GC formation and the epidemic threshold? This difference seems surprising at first, particularly because epidemics are often viewed as giant components in a graph created by a percolation process.

The answer, however, is also almost manifest from the viewpoint taken in this paper: the stochastic process of the spread of the disease in a network is equivalent to the exploration of this network's giant component if the transmissibility is equal to 1. The process is then completely deterministic and the resulting giant component will be identical to the giant component of the underlying network. In this case, R_0 is reduced to $R_0(T=1)=(1-C)(\langle k^2-k\rangle/\langle k\rangle)$, and the threshold condition $R_0>1$ gives $(\langle k^2-k\rangle/\langle k\rangle)>(1/(1-C))$, as obtained by Berchenko et al., (2009). However, for small transmission probability, the effects of clustering on the epidemic dynamics are limited. In effect, the epidemic does not "see" the triangles, because passing through them is highly unlikely.

5. Conclusions

In conclusion we have thus arrived at a simple and intuitive formula for R_0, showing that in order to account for clustering, one need only multiply the R_0^*, calculated assuming $C=0$, by the factor $(1-CT)$, that is $R_0(T)=(1-CT)R_0^*(T)$. This factor takes into account the fact that when a disease spreads on a clustered contact network, a node's predecessor in the epidemic process

can infect nodes which are themselves neighbors, effectively "stealing" susceptible nodes from them, as shown in Figs. 2 and 3. Thus, clustering leads to a reduced R_0, in comparison with a disease spreading on a network with a similar degree distribution but having $C=0$.

This simple model is flexible enough to allow for different infection schemes, in particular, degree dependent transmission probability. When transmissibility is inversely proportional to the degree, a lower R_0 and thus a higher epidemic threshold are obtained, in comparison with a fixed-transmission probability model (with the same mean transmission probability). In this sense, heterogeneity in transmission can be detrimental to disease spread, when transmission probability is inversely correlated with degree. The simplicity of this model suggests that it can be extended and used to study other cases such as degree-dependent clustering.

Our approximation for R_0 makes clear that clustering often fails to have a major impact on the epidemic threshold, because in general the term $CT \ll 1$. Nevertheless, increasing clustering will lead to an increase in this threshold, however small. Due to the small magnitude of the effect of clustering on R_0 and clustering's correlation with other topological properties such as assortativity, its effects may often be difficult to measure in simulations or experiments.

Acknowledgments

We are grateful for support from the European FP7 grant Epiwork, the Israel Science Foundation and the Israel Ministry of Health. We thank Or Givan and Rami Yaari for their discussions and technical assistance.

Appendix I. Generating clustered networks with prescribed degree distributions—the swap method

In order to create clustered networks, we generate unclustered graphs using the frequently used configuration model (CM) described in Newman (2003) and Britton et al. (2006) or the Barabási–Albert algorithm (Barabási and Albert, 1999) and increase their clustering using a Monte Carlo approach. The configuration model supplies a method of generating random graphs with a prescribed degree distribution. The clustering coefficient of CM graphs scales as $1/N$ (Newman, 2003). Unclustered ER and regular graphs were generated using the CM. Scale free graphs were generated using the Barabási–Albert preferential attachment algorithm (Barabási and Albert, 1999). The resulting networks have exponent $\gamma \to 3$.

One possible method for increasing clustering in these graphs without changing the degree distribution is called the Swap Method, described in full in Berchenko et al. (2009) and Artzy-Randrup and Stone (2005). Here, the adjacency matrix is searched for checkerboard patterns. These are sub-matrices of the form

$$C_1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \text{ or } C_2 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$

where the rows and columns are not necessarily adjacent. Then, a switch from C_1 to C_2 (or 2 to 1) is performed if it increases the clustering in the graph. In every step the algorithm searches for a single random checkerboard pattern and implements a switch after checking if the clustering coefficient C (Eq. (2.2) in text) increases from such a change. (Some switches are accepted randomly, in order to avoid becoming stranded on local "clustering peaks".) The process is halted when the desired level of clustering, C, is achieved.

It is important to note that switching (replacing C_1 by C_2) does not change row or column sums of the adjacency matrix and thus preserves the degree distribution. The local clustering in networks generated by applying this algorithm on ER and other network types exhibited minor degree dependence (Stone, 2011).

An improvement on this method is described in Bansal et al. (2009), and can be achieved by choosing the initial pair of links to be switched such that at least one triangle is created. This is done by choosing a random node with 2 or more edges. Then, one follows two random edges emanating from the initial node, and arrives at two new nodes. Following a random edge emanating from each one of these neighbors gives us two edges, a–b and c–d, where a and c are both neighbors of the initial node chosen. If these two edges satisfy the condition above (from the simple switching algorithm), at least one triangle is created in the switch.

It should, however, be noted that configuration model networks have a specified level of clustering predicted for a given degree distribution (Newman, 2003). By increasing C, we exit the domain of the configuration model, thereby invalidating some of its assumptions. The configuration model essentially assumes very little: the specified degree distribution and a random pairing of "stubs". Since the swap method does not alter the degree distribution, it follows that we depart from the random stub-pairing assumption and create graphs with some degree correlations (which may or may not be reflected in their Pearson correlation coefficient; see also Catanzaro et al. (2005)).

Appendix II. simulations and measurement of R_0

As explained in the text, a simulation begins by selecting a random node in the population to become infective, i_1, and one of its neighbors, i_2, to be the node for which we make our calculation of R_0. Next, i_1 is allowed to infect some of its neighbors i.e., each of its neighbors has a probability T of being infected in this one time step. If i_2 was not infected, we return to the initial state where only i_1 is infective and retry. Once i_2 is infected (as are perhaps some of the other neighbors of i_1), the simulation is allowed to proceed an additional time step so that i_1 becomes recovered ($\tau=1$) and all of its neighbors which have been infected may infect some of their neighbors. The number of neighbors of i_2 which are infected in this time step (i.e., excluding those infected by i_1) constitutes R_0 for the current simulation. For each graph, we ran 40 of these simulations. For each graph type (BA, random, regular) 40 different graphs were generated. Simulation results were averaged over all runs for each graph type.

References

Andersson, H., 1997. Epidemics in a population with social structures. Math. BioSci. 140, 79–84.
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, http://dx.doi.org/10.1098/rspb.2006.0057.
Anderson, R.M., May, R.M., 1991. Infectious Diseases of Humans: Dynamics and Control. Oxford University Press, New York.
Albert, R., Barabási, A.L., 2002. Statistical mechanics of complex networks. Rev. Mod. Phys. 74, 47–97.
Artzy-Randrup, Y., Stone, L., 2005. Generating uniformly distributed random networks: the ADD method. Phys. Rev. E 72, 05670.
Britton, T., Deijfen, M., Lageras, A.N., Lindholm, M., 2008. Epidemics on random graphs with tunable clustering. J. Appl. Probab. 45, 743–756.
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.
Barabási, A.L., Albert, R., 1999. Emergence of scaling in random networks. Science 286 (5439), 509–512, http://dx.doi.org/10.1126/science.286.5439.509.
Britton, T., Deijfen, M., Martin-Löf, A., 2006. Generating simple random graphs with prescribed degree distribution. J. Stat. Phys. 124, 1377–1397.
Bansal, S., Khandelwal, S., Meyers, L.A., 2009. Exploring biological network structure with clustered random networks. BMC Bioinformat. 10, 405.

Cohen, R., ben-Avraham, D., Havlin, S., 2002. Percolation critical exponents in scale-free networks. Phys. Rev. E 66, 036113.
Cohen, R., Erez, K., ben-Avraham, D., Havlin, S., 2000. "Resilience of the Internet to random breakdowns". Phys. Rev. Lett. 85, 4626.
Catanzaro, M., Boguñá, M., Pastor-Satorras, R., 2005. Generation of uncorrelated random scale-free networks. Phys. Rev. E 71, 027103.
Diekmann, O., Heesterbeek, J.A.P., 2000. Mathematical Epidemiology of Infectious Diseases: Model Building, Analysis and Interpretation. John Wiley & Sons, Chichester.
Feld, S., 1991. Why your friends have more friends than you do. Am. J. Sociol. 96, 1464–1477.
Keeling, M.J., Eames, K.T.D., 2005. Networks and epidemic models. J. R. Soc. Interface 2 (4), 295–307.
Keeling, M.J., 2005. Implications of network structure for epidemic dynamics. Theor. Popul. Biol. 67, 1–8.
Kiss, I.Z., Green, D.M., 2008. Comment on "properties of highly clustered networks". Phys. Rev. E 78 (4), ISSN 1539-3755.
Lloyd, A.L., May, R.M., 2001. How viruses spread among computers and people. Science 292, 1316–1317.
Miller, J.C., 2009. Spread of infectious disease through clustered populations. J. R. Soc. Interface 6, 1121.
Molloy, M., Reed, B., 1995. A critical point for random graphs with a given degree sequence. Random Struct. Algor. 6, 161–179.
Miller, J.C., 2009. Percolation and epidemics in random clustered networks. Phys. Rev. E 80, 020901(R).
May, R.M., Lloyd, A.L., 2001. Infection dynamics on scale-free networks. Phys. Rev. E 64, 066112.
Newman, M.E.J., 2003. Mixing patterns in networks. Phys. Rev. E 67, 026126.
Newman, M.E.J., 2009. Random graphs with clustering. Phys. Rev. Lett. 103, 58701.
Newman, M.E.J., 2003. Random graphs as models of networks. In: Bornholdt, S., Schuster, H.G. (Eds.), Handbook of Graphs and Networks. Wiley-VCH, Berlin, pp. 35–68.
Newman, M.E.J., 2003. Ego-centered networks and the ripple effect. Social Networks 25, 83–95.
Newman, M.E.J., 2002. Assortative mixing in networks. Phys. Rev. Lett. 89, 208701.
Newman, M.E.J., 2003. Properties of highly nclustered networks. Phys. Rev. E 68, 026121.
Olinky, R., Stone, L., 2004. Unexpected epidemic thresholds in scale-free networks. Phys. Rev. E 70, 30902(R).
Pastor-Satorras, R., Vespignani, A., 2001. Epidemic spreading in scale-free networks. Phys. Rev. Lett. 86, 3200–3203.
Ross, S.M., 2002. A First Course in Probability, sixth ed. Prentice-Hall, Englewood-Cliffs, NJ.
Serrano, M.A., Boguñá, M., 2006. Percolation and epidemic thresholds in clustered networks. Phys. Rev. Lett. 97, 088701.
Salathé, M., Kazandjieva, M., Lee, J.W., Levis, P., Feldman, M.W., Jones, J.H., 2010. A high-resolution human contact network for infectious disease transmission. Proc. Natl. Acad. Sci. 107, 22020–22025, http://dx.doi.org/10.1073/pnas.1009094108.
L. Stone, 2011. private communication.
Trapman, P., 2007. On analytical approaches to epidemics on networks. Theor. Popul. Bio. 71, 160–173.
Volz, E.M., Miller, J.C., Galvani, A.P., Meyers, L.A., 2011. Effects of heterogeneous and clustered contact patterns on infectious disease dynamics. PLoS Comput. Biol. 7, e1002042.