Exact Solutions of Epidemic Models on Networks

M. E. J. Newman

SFI WORKING PAPER: 2001-12-073

SFI Working Papers contain accounts of scientific work of the author(s) and do not necessarily represent the views of the Santa Fe Institute. We accept papers intended for publication in peer-reviewed journals or proceedings volumes, but not papers that have already appeared in print. Except for papers by our external faculty, papers must be based on work done at SFI, inspired by an invited visit to or collaboration at SFI, or funded by an SFI grant.

©NOTICE: This working paper is included by permission of the contributing author(s) as a means to ensure timely distribution of the scholarly and technical work on a non-commercial basis. Copyright and all rights therein are maintained by the author(s). It is understood that all persons copying this information will adhere to the terms and constraints invoked by each author’s copyright. These works may be reposted only with the explicit permission of the copyright holder.

www.santafe.edu
Exact solutions of epidemic models on networks

M. E. J. Newman

Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501, U.S.A.

(Dated: November 28, 2001)

Abstract

The study of social networks, and in particular the spread of disease on networks, has attracted considerable recent attention in the physical and mathematical literature. In this paper, we show that a large class of standard epidemiological models, the so-called susceptible/infective/recovered models, and many of their generalizations, can be solved exactly on a wide variety of networks. Solutions are possible for cases with heterogeneous or correlated probabilities of transmission, cases in which a portion of the individuals in the network are vaccinated against the disease, either uniformly at random or in some correlated fashion, and cases in which the network has complex structure of various kinds, such as separation of the population in groups according to personal, geographic, or social characteristics. We confirm the correctness of our exact solutions by comparison with computer simulations of epidemics propagating on the corresponding networks.
I. INTRODUCTION

Networks of various kinds have been the subject of much recent research. Social networks, technological networks, and biological networks have all been examined and modeled in some detail. Most of this work has focussed on the structural properties of the networks in question—patterns of connection between people or computers or species. Structure, however, while important, is in most cases only a prerequisite to answering the question of real interest: what is the behavior of networked systems? Knowing the topological structure of the Internet, for example, is a first step towards understanding the way Internet traffic flows. Knowing the structure of the World-Wide Web may help us to find information better or to understand the nature of Web communities.

One area in which substantial progress has been made in understanding system behavior is the study of the spread of disease. Recent simulation studies and approximate analytical treatments suggest that network structure can play a crucial role in defining the shape of a disease epidemic. In this paper, we show that the most fundamental standard model of disease propagation, the SIR model, and a large set of its generalized forms, are exactly solvable on a broad class of networks, including networks with complex community structure of various kinds. Our solutions provide exact criteria for when an epidemic will occur, how many people will be affected, and how the network structure or the transmission properties of the disease could be modified in order to prevent an epidemic.

II. SIR MODELS

The most fundamental mathematical model of the spread of disease, first formulated (though never published) by Lowell Reed and Wade Hampton Frost in the 1920s, is the susceptible/infective/recovered or SIR model. In this model a population is divided into three classes according to their status in relation to the disease of interest: susceptible (S), meaning they are free of the disease but can catch it, infective (I), meaning they have the disease and can pass it on to others, and recovered (R), meaning they have recovered from the disease and can no longer pass it on. There is a fixed probability per unit time that an infective individual will pass the disease to a susceptible individual with whom they have contact, rendering that individual infective.
Individuals who contract the disease remain infective for a certain time period before recovering and losing their infectivity.\textsuperscript{38}

The one other thing we need, in order to turn this process into a complete model of disease spread is the pattern of contacts between individuals. In the standard treatments, and indeed in almost all of mathematical epidemiology, researchers use the so-called “fully mixed” approximation, in which it is assumed that every individual has equal chance of contact with every other.\textsuperscript{18–20} This is clearly an unrealistic assumption, but it has proved popular because it allows one to write differential equations for the time evolution of the disease, which can be solved or numerically integrated to determine the course of an epidemic. More realistic versions of the model have also been studied in which populations are divided into groups, according to age or other characteristics. These models are still fully mixed within each group, however, again making them quite unlike the real world. In the real world, the pattern of contacts between individuals is far from fully mixed, forming a social network with well-defined structure. Only by incorporating this structure into our model of disease spreading can we hope to understand fully how epidemics develop and progress.

Consider then a network of initially susceptible individuals represented by the vertices of a graph. The edges of the graph represent connections between individuals by which disease can be transmitted. These connections might represent, for example, periodic physical proximity—two people working together in the same building perhaps, or living in the same house.

One of the most important results to come out of recent work on networks is the finding that the degree distributions of many networks are highly right-skewed. (Recall that the “degree” of a vertex is the number of other vertices to which it is connected.) In other words, most vertices have only a low degree, but there are a small number whose degree is very high.\textsuperscript{3,5,8} It is known that the presence of this small number of highly connected vertices can have a disproportionate effect on certain properties of the network.\textsuperscript{10,21–23} Recent work suggests that the same may true for disease propagation on networks,\textsuperscript{17,24} and so it will be important that we incorporate non-trivial degree distributions in our models.

By contrast, it has been shown\textsuperscript{25} that the standard fully mixed version of the SIR model is precisely equivalent to disease propagation on an Erdős–Rényi random graph, in which each vertex is connected with equal probability $p$ to each other, and the degree distribution is as a result Poissonian. The Poisson distribution has no significant skewness and is very different from the degree
distributions measured in most real-world networks, as has been much emphasized elsewhere.\textsuperscript{2,3,5}

In this paper we will also abandon two other unrealistic assumptions of the usual SIR model, the assumptions that all edges in our network represent equal probability of disease transmission, and that all individuals who catch the disease remain infective for the same amount of time. We will allow the probability per unit time $r_{ij}$ of transmission from an infective individual $i$ to a susceptible individual $j$ who are connected by an edge to be drawn from any arbitrary distribution $P(r)$. We will also allow the time $\tau$ for which individuals remain infective to be drawn from any arbitrary distribution $P(\tau)$. These generalizations increase enormously the range (and realism) of models to which our solutions are applicable.

III. EXACT SOLUTION FOR A SIMPLE CASE

The crucial observation that makes our solutions possible is that all epidemic processes in closed populations in which the host eventually either acquires immunity or dies are equivalent to a bond percolation process on the corresponding network of individuals and contacts. This correspondence appears first to have been pointed out by Grassberger for the case of the simple SIR model with fixed probabilities of infection and times of infectiveness.\textsuperscript{26} More recently, it has been observed numerically that the correspondence extends also to the case of variable probabilities and times.\textsuperscript{27} In fact, it is straightforward to show that the above generalized SIR process on a network corresponds to bond percolation on the same network with uniform bond occupation probability

$$T = 1 - \int_{0}^{\infty} dr d\tau P(r)P(\tau) e^{-r\tau}. \tag{1}$$

The quantity $T$, which we call the transmissibility of the disease, represents the average total probability that a susceptible individual will catch the disease from an infective contact. Using data from an outbreak of walking pneumonia in a hospital, for example, we have estimated that $T = 0.23(2)$ for transmission of that disease between patients in that setting.\textsuperscript{28}

In this paper we solve the bond percolation problem with bond probability $T$ on a variety of networks using a generating function technique similar to that introduced by Moore and Newman.\textsuperscript{29} We illustrate the method first for a simple class of unipartite graphs studied previously by a number of authors,\textsuperscript{30–32} in which the degree distribution is specified, but the graph is in other respects random.
Suppose that the probability of a randomly chosen vertex in our graph having degree \( k \) is \( p_k \). Following Newman et al.,\(^3\) we define two generating functions:

\[
G_0(x) = \sum_{k=0}^{\infty} p_k x^k, \quad G_1(x) = \frac{1}{z} G_0'(x),
\]

where \( z = G_0'(1) \) is the mean vertex degree in the network. These two functions generate respectively the probability distributions of the degrees of randomly chosen vertices, and vertices at the ends of randomly chosen edges. Not all edges leading from a vertex will be occupied however (i.e., result in transmission of the disease). The distribution of the number \( m \) of occupied edges around a randomly chosen vertex is generated by

\[
G_0(x; T) = \sum_{m=0}^{\infty} \sum_{k=m}^{\infty} p_k \binom{k}{m} T^m (1 - T)^{k-m} x^m
\]

\[
= \sum_{k=0}^{\infty} p_k \sum_{m=0}^{k} \binom{k}{m} (xT)^m (1 - T)^{k-m} = \sum_{k=0}^{\infty} p_k (1 - T + xT)^k
\]

\[
= G_0(1 + (x-1)T). \tag{3}
\]

And similarly the number around the vertex at the end of a randomly chosen edge is generated by \( G_1(x; T) = G_1(1 + (x-1)T) \). Now the generating function \( H_1(x; T) \) for the total number of people infected as a result of a single transmission along an edge in the network must satisfy a self-consistency condition of the form\(^29,32\)

\[
H_1(x; T) = xG_1(H_1(x; T); T). \tag{4}
\]

And the distribution of the number of people affected by an outbreak starting with a single disease carrier is generated by

\[
H_0(x; T) = xG_0(H_1(x; T); T). \tag{5}
\]

The average size \( \langle s \rangle \) of a disease outbreak is now given by the derivative of \( H_0 \) with respect to \( x \):

\[
\langle s \rangle = H_0'(1; T) = 1 + G_0'(1; T)H_1'(1; T) = 1 + \frac{G_0'(1; T)}{1 - G_1'(1; T)} = 1 + \frac{T G_0'(1)}{1 - T G_1'(1)}, \tag{6}
\]

where we have made use of Eq. (4) and the fact that all generating functions are 1 at \( x = 1 \) if the distributions that they generate are properly normalized. Eq. (6) diverges when \( T \) is equal to the critical value \( T_c = 1/G_1'(1) \), and this point marks the onset of epidemic behavior. For
transmissibilities below this epidemic threshold, \( T < T_c \), all outbreaks are finite in size, no matter how large the network, and the probability of any given individual being affected by an outbreak is zero in the limit of large graph size. For \( T > T_c \) there is always a finite chance of infection. The fraction of the population that is infected in an epidemic outbreak can be derived by observing that above \( T_c \), Eq. (5) generates the size distribution of outbreaks excluding epidemics,\(^{32}\) and hence the size \( S \) of the epidemic is given by the solution of

\[
S = 1 - G_0(u; T), \quad u = G_1(u; T).
\]

Unfortunately, it is not usually possible to find a closed form solution to this last equation, but it can be solved numerically by iteration from a suitable starting value of \( u \).

Note that it is not the case, even above \( T_c \), that all outbreaks give rise to epidemics of the disease. There are still finite outbreaks even in the epidemic regime, and the probability of an outbreak reaching epidemic proportions is also given by \( S \). While this appears very natural, it stands nonetheless in stark contrast to the standard fully mixed models, for which all outbreaks give rise to epidemics above the epidemic transition point.

IV. EXAMPLE APPLICATION

As an example of this first simple epidemic model, consider SIR disease outbreaks taking place on networks having a degree distribution with the truncated power-law form

\[
p_k = \begin{cases} 
0 & \text{for } k = 0 \\
C k^{-\alpha} e^{-k/\kappa} & \text{for } k \geq 1.
\end{cases}
\]

where \( \alpha \) and \( \kappa \) are constants, and \( C \) is set by the requirement that the distribution be normalized. This distribution, which we have studied elsewhere also,\(^{23,32}\) is interesting for a number of reasons: (1) it is seen in a number of networks in the real world,\(^{5,6}\) (2) it includes both pure power-law\(^ {8,9,33}\) and pure exponential\(^ {5}\) degree distributions as special cases, (3) it is normalizable and has all moments finite for all finite values of \( \kappa \) (by contrast with the pure power-law distribution, for instance). The distributions \( P(r) \) and \( P(\tau) \) of the probabilities of disease transmission and infectiveness times, we choose for simplicity to be uniform over the intervals \( 0 \leq r < r_{\text{max}} \) and \( 0 \leq \tau < \tau_{\text{max}} \) respectively, which makes Eq. (1) straightforward to evaluate.
Substituting (8) into the definitions of our generating functions, we then find that our disease has an epidemic transition at

\[ T_c = \frac{\text{Li}_{\alpha-1}(e^{-1/\kappa})}{\text{Li}_{\alpha-2}(e^{-1/\kappa}) - \text{Li}_{\alpha-1}(e^{-1/\kappa})}, \]  

where \( \text{Li}_n(x) \) is the \( n \)th polylogarithm of \( x \). Below this transition no epidemics are possible, only small outbreaks having average size

\[ \langle s \rangle = 1 + \frac{T[\text{Li}_{\alpha-1}(e^{-1/\kappa})]^2}{\text{Li}_{\alpha}(e^{-1/\kappa})[(T+1)\text{Li}_{\alpha-1}(e^{-1/\kappa}) - T\text{Li}_{\alpha-2}(e^{-1/\kappa})]}, \]  

while above it, epidemics occur with size and probability \( S \), whose value we can extract by numerical iteration of Eq. (7).

In Fig. 1 we compare the predictions of this solution against explicit simulations of epidemics spreading on networks with heterogeneous transmission rates \( r \) and infectiveness times \( \tau \). As the figure shows, agreement between analytic and numerical results is good.

To emphasize the difference between our results and those for the equivalent fully mixed model, we compare the position of the epidemic threshold in the two cases. In the case \( \alpha = 2, \kappa = 10 \) (the middle curve in each frame of Fig. 1), our analytic solution predicts that the epidemic threshold occurs at \( T_c = 0.329 \). The simulations agree well with this prediction, giving \( T_c = 0.32(2) \). By contrast, a fully mixed SIR model in which each infective individual transmits the disease to the same average number of others as in our network, gives a very different prediction of \( T_c = 0.558 \).

V. EXTENSIONS OF THE MODEL

Although the model above is already more realistic than the standard epidemic models in a number of ways (network structure, heterogeneous transmission, heterogeneous infectiveness times), there are many ways it can be further improved. For instance, with real diseases the transmission rates \( r \) or the infectiveness times \( \tau \) might not be iid random variables as we have assumed; they might be correlated. As an example of how this can be incorporated into the model, consider the case where the distribution of transmission rates \( r \) depends on the degree \( k \) of the vertex representing the infective individual. (One could imagine for example that people who have many contacts tend also to have more fleeting contacts, so that \( r \) would go down on average with increasing \( k \).) Then the transmissibility also becomes a function of \( k \) according to
FIG. 1: Epidemic size (top) and average outbreak size (bottom) for the SIR model on networks with degree distributions of the form Eq. (8) as a function of transmissibility. Solid lines are the exact solutions, Eqs. (6) and (7), for $\alpha = 2$ and (left to right in each panel) $\kappa = 20$, 10, and 5. Each of the points is an average result for 10000 simulations on graphs of 100000 vertices each with distributions of $r$ and $\tau$ as described in the text.

$T_k = 1 - \int dr \int d\tau P_k(r)P(\tau) e^{-rt}$ and the generating functions become a function of the complete set $\{T_k\}$. Alternatively, the distribution of $r$ might depend on the degree of the individual being infected, which gives us an similar set $\{U_k\}$ of transmissibilities. Or $r$ might depend on both degrees. The correct generalization of the generating functions is:

$$G_0(x; \{T_k\}, \{U_k\}) = \sum_k p_k (1 + (x-1)T_k)^k,$$

$$G_1(x; \{T_k\}, \{U_k\}) = \frac{\sum_k kp_k[1 + ((1 + (x-1)T_k)^{k-1} - 1)U_k]}{\sum_k kp_k}.$$  

The cases in which transmission depends only on one degree or the other can be derived from these expressions by setting either $T_k = 1$ or $U_k = 1$ for all $k$.

Once we have the generating functions, then the calculation proceeds as before, with mean
outbreak size below the epidemic transition being given by Eq. (6) and epidemic size above it by Eq. (7). The epidemic transition occurs as before at $G_1(1; \{T_k\}, \{U_k\}) = 1$.

Another area of current interest is models incorporating vaccination of individuals.\textsuperscript{15,34} This can be included in our models too—vaccination can be thought of as site percolation, in which occupied sites are capable of transmitting the disease and unoccupied ones are vaccinated. As we have shown elsewhere in the context of robustness models of the Internet,\textsuperscript{23} site percolation can be solved exactly on networks of the type considered here, both in the case of uniform independent vaccination probability (i.e., random vaccination of a population) or in the case of vaccination which is correlated with properties of individuals such as their degree (so that vaccination can be targeted at the so-called core group of the disease-carrying network—those with the highest degrees). It is straightforward to combine vaccination with any of the other features above, to provide complete solutions for disease spread in models with heterogeneous and correlated transmission or infectivness times and heterogeneous and correlated vaccination patterns.\textsuperscript{35}

\section{VI. MODELS WITH MORE COMPLEX NETWORK STRUCTURE}

The other main way in which we can make our models more realistic, while still retaining exact solvability, is to incorporate more realistic social structure into our networks. In the previous sections we have modeled a very simple network in which all individuals are of a single type and have statistically uniform properties. In the real world however individuals differ according to a variety of social and geographical parameters.

As an example of how network structure can be included in the model, consider the network by which a sexually transmitted disease is communicated, which is also the network of sexual partnerships between individuals. In a recent study of 2810 respondents Liljeros et al.\textsuperscript{7} found the degree distributions shown in Fig. 2 for number of sexual partners of men and women over the course of a year. From the figure it appears that both distributions are power-law in form with exponents that fall in the range $3.1$ to $3.3$. Of course, not all contacts between individuals carry the same probability of transmission of disease, and some may be inadequate for transmission altogether, but this makes no difference—Eq. (1) allows for such variation and our methods will still work.

If we assume that the disease of interest is transmitted primarily by contacts between men and
FIG. 2: Measured degree distributions for the network of sexual contacts of men and women, plotted on logarithmic scales. Note that the distribution shown in cumulative (a so-called rank/frequency plot). After Liljeros et al. Inset: the bipartite structure of the network in which there are two types of nodes (men and women) and links connecting nodes of different types, with the specified degree distribution for each type.

If women (true only for some diseases), then to a good approximation the network of contacts has the bipartite form shown in the inset of Fig. 2. We define two pairs of generating functions for males and females:

\[
F_0(x) = \sum_j p_j x^j, \quad F_1(x) = \frac{1}{\mu} \sum_j j p_j x^j, \quad (13)
\]

\[
G_0(x) = \sum_k q_k x^k, \quad G_1(x) = \frac{1}{\nu} \sum_k k q_k x^k, \quad (14)
\]

where \(p_j\) and \(q_k\) are the two degree distributions and \(\mu\) and \(\nu\) are their means. We can then develop expressions similar to Eqs. (6) and (7) for an epidemic on this new network. For instance, the epidemic transition takes place at the point where \(T_m f T_f m = 1/[F_1(1)G_1'(1)]\) where \(T_m f\) and \(T_f m\) are the transmissibilities for male-to-female and female-to-male infection respectively.

One important result which follows immediately is that if the degree distributions are truly
power-law in form, then there exists an epidemic transition only for a small range of values of the exponent \( \alpha \) of the power law. Let us assume, as appears to be the case, that the exponents are roughly equal for men and women: \( \alpha_m = \alpha_f = \alpha \). Then if \( \alpha \leq 3 \), we find that \( T_{mf}T_{fm} = 0 \), which is only possible if at least one of the transmissibilities \( T_{mf} \) and \( T_{fm} \) is zero. As long as both are positive, we will always be in the epidemic regime, and this would clearly be bad news. No amount of precautionary measures to reduce the probability of transmission would ever eradicate the disease. (Similar results have been seen in other types of models also.\textsuperscript{17,24}) Conversely, if \( \alpha > \alpha_c \), where \( \alpha_c = 3.4788 \ldots \) is the solution of \( \zeta(\alpha - 2) = 2\zeta(\alpha - 1) \), we find that \( T_{mf}T_{fm} = 1 \), which is only possible if both \( T_{mf} \) and \( T_{fm} \) are 1. When either is less than 1 no epidemic will ever occur, which would be good news.\textsuperscript{39} Only in the small intermediate region \( 3 < \alpha < 3.4788 \ldots \) does the model possess an epidemic transition. Interestingly, the real-world network of Fig. 2 appears to fall precisely in this region, with \( \alpha \approx 3.2 \). If true, this would be both good and bad news. On the bad side, it means that epidemics can occur. But on the good side, it means that that it is in theory possible to prevent an epidemic by reducing the probability of transmission, which is precisely what most health education campaigns attempt to do. The predicted critical value of the transmissibility is \( \zeta(\alpha - 1)/[\zeta(\alpha - 2) - \zeta(\alpha - 1)] \), which gives \( T_c = 0.3635\ldots \) for \( \alpha = 3.2 \). Epidemic behavior would cease were it possible to arrange that \( T_{mf}T_{fm} < T_c^2 \).

Some caveats are in order here. The error bars on the values of the exponent \( \alpha \) are quite large (about \( \pm 0.3 \)). Thus, assuming that the conclusion of a power-law degree distribution is correct in the first place, it is still possible that \( \alpha < 3 \), putting us in the regime where there is always epidemic behavior and no reduction in transmission probability will ever prevent spread of the disease on a large scale. On the other hand, it may also be that the distribution is not a perfect power law. Although Fig. 2 does appear to have a power-law tail, it seems moderately likely that the tail is cut off at some point. If this is the case, then there will always be an epidemic transition at finite \( T \), regardless of the value of \( \alpha \). Furthermore, if it were possible to reduce the number of partners that the most active members of the network have, so that the cutoff moves lower, then the epidemic threshold rises, making it easier to eradicate the disease. Interestingly, the fraction of individuals in the network whose degree need change in order to make a significant difference is quite small. At \( \alpha = 3 \), for instance, a change of cutoff from \( \kappa = \infty \) to \( \kappa = 100 \) affects only 1.3\% of the population, but increases the epidemic threshold from \( T_c = 0 \) to \( T_c = 0.52 \). In other words, targeting preventative efforts at changing the behavior of the most active members of the network
may be a much more promising way of preventing the spread of disease than targeting everyone. (This suggestion is certainly not new, but our models provide a quantitative basis for assessing its efficacy.)

Another application of the techniques presented here is described in Ref. 28. In that paper we model in detail the spread of walking pneumonia (Mycoplasma pneumoniae) in a closed setting (a hospital) for which network data are available from observation of an actual outbreak. In this example, our exact solutions agree well both with simulations and with data from the outbreak studied. Furthermore, examination of the analytic solution allows us to make specific suggestions about possible new control strategies for M. pneumoniae infections in settings of this type.

VII. CONCLUSIONS

In this paper we have shown how generating function techniques can be used to provide exact solutions for SIR-type models on a variety of networks incorporating social structure of various kinds. Our solutions are not restricted to the standard simplistic model in which parameters such as infectiveness times and transmission rates are homogeneous over the entire network, but cover cases of heterogeneous and correlated parameters. It is also possible to solve models in which a portion of the population is vaccinated and so cannot catch the disease, including models where vaccination patterns are not independent but are correlated with other properties of the network. We have given examples of applications of our methods to a simple unipartite network (although still with a population having heterogeneous properties) and to a bipartite network model of a sexually transmitted disease.

Applications of the techniques described here are also possible for networks specific to many other settings, and hold promise for the better understanding of the role that the structure of contact networks plays in the spread of disease.

Acknowledgments

The author thanks Lauren Ancel, László Barabási, and Duncan Callaway for useful comments. This work was supported in part by National Science Foundation grant number DMS–0109086.


31 M. Molloy and B. Reed, A critical point for random graphs with a given degree sequence. *Random


37 In common parlance, the word “infectious” is more often used, but in the epidemiological literature “infective” is the accepted term.

38 Diseases in which the victims may die rather than recover are covered by the same model. As far as the model is concerned recovery and death remove an individual from the infective pool with equal certainty. For this reason the R stage of the SIR is sometimes called “removed” rather than “recovered.”

39 Indeed for $\alpha > \alpha_c$ the network has no giant component, so clearly no epidemic will ever be possible.