CHAPTER 1

Network Models In Epidemiology: An Overview

Alun L. Lloyd
Biomathematics Graduate Program and Department of Mathematics, North Carolina State University, Raleigh, NC 27695

Steve Valeika
Department of Epidemiology, University of North Carolina, School of Public Health, Chapel Hill, NC 27599

1. Introduction

Mathematical modeling has provided many significant insights concerning the epidemiology of infectious diseases. The most notable of these include threshold conditions (involving the so-called 'basic reproductive number') that describe when invasion and persistence of an infection is possible[1,2,3].

The development of much of this theory has revolved around the use of extremely simple models, such as deterministic compartmental models. Typically, the population of interest is subdivided into a small number of compartments based on infection status (e.g. susceptible, infectious or recovered) and the flows between these compartments are described by a low dimensional set of ordinary differential equations. The derivation of these equations typically involves a number of simplifying assumptions, an important example of which is the 'mass-action' description of transmission, which will be discussed in detail below.

The simplicity of these models facilitates the use of analytic techniques to gain general understanding, but at the cost of oversimplifying the biology of real-world disease processes. The weaknesses of simple models have long been clear, particularly when model behavior has been compared to epidemiologic data, and this has lead to the development of increasingly complex models that do attempt to account for more details of the un-

derlying biology[1]. Much of this complexity can be incorporated within the population-level framework provided by compartmental models.

Individual-level models offer a fundamentally different way of describing biological populations. In this approach, every individual in the population is accounted for as a separate entity. The complexity of such models makes analysis difficult, and numerical simulation computationally intensive. Furthermore, these models must include some description of the interactions between the individuals that make up the population. Unless a large number of simplifying assumptions are made, specifying these interactions is a major task whenever there are more than a handful of individuals to be considered.

Network models (also known as graph models) provide a natural way of describing a population and their interactions. Nodes (vertices) of the graph represent individuals and edges (links) depict interactions between individuals that could potentially lead to transmission of infection. It is interesting to note that similar network representations can be used in a number of contexts, such as transportation networks, communication networks (including the internet and World Wide Web) and social networks (including friendship, movie actor and scientific collaboration networks) [4,5,6].

In this chapter we shall discuss the development and use of network models. While network models have long been discussed in the theoretical epidemiology literature, they have recently received a large amount of attention amongst the statistical physics community. This has been fueled by the desire to better understand the structure of social and large-scale technological networks, and the increases in computational power that have made the simulation of reasonably-sized network models a feasible proposition. A main aim of this review is to bridge the epidemiologic and statistical physics approaches to network models for infectious diseases, highlighting the important contributions made by both research communities.

This chapter is organized as follows. We shall first discuss some of the epidemiologic settings in which network models are employed. Our attention will then turn to ways in which networks are described, including measures that attempt to capture important properties of graphs. An important part of this discussion will include the feasibility of employing such methods to describe real-world networks, particularly when only incomplete information is available. We shall then describe some classes of networks that have received particular attention. Finally, we discuss the impact of network structure on the spread of infection and some of the ways in which control measures must account for this structure.

2. Network Model Settings

2.1. *Network Models as a Research Tool*

Since many epidemiologic systems can be most naturally described in terms of individual-level events and processes, network models have proved to be an valuable research tool to explore the relationship between individual-based and population-level models.

The formulation of population-level models typically involves making a large number of simplifying assumptions. Perhaps the most important of these is the mass-action assumption, in which the rate at which new infections occur is taken to be proportional both to the density of infective individuals and to the density of susceptible individuals [1,2,3]. This assumption has its roots in the theory of chemical kinetics, in which it is used to describe reaction kinetics in 'well-mixed' settings such as a vigorously stirred vessel.

Few epidemiologic settings could be described as being well-mixed: interactions within a population typically have some structure, for instance reflecting the social and spatial structure of a community. As examples, workplaces, schools and family homes provide settings in which particular groups of individuals spend considerable time in relatively close contact. Such settings are important sites for the transmission of many infections: a given individual is much more likely to acquire infection from such sources than from a person randomly chosen from the population at large. Further heterogeneities in transmission arise because individuals differ in other ways, such as their susceptibility to infection, their level of infectiousness once they become infected, and the number of people with whom they interact.

Network models have long been used to investigate the impact of spatial structure on the transmission of infection. Particular attention has been paid to the degree to which localized transmission of infection tends to slow the spread of an infection in a population[7]. This is in marked contrast with the mass-action setting, in which the presence of infection is immediately felt by every individual in the population, which can allow for rapid spread of infection.

2.2. *Epidemiologic Settings*

The epidemiologic settings in which network descriptions have the longest history of use involve sexually transmitted infections (STIs), such as gonorrhea or the human immunodeficiency virus (HIV) [8,9,10,11,12,13,14]. Here

there are natural, well-defined, network structures (sexual partnership networks) which have long been exploited by public health bodies in their attempts to track and control outbreaks of STIs. Network models have more recently been employed to describe the spread of a wider range of infections such as measles, SARS or foot and mouth disease (FMD)[15,16,17,18]. Increased interest in bioterrorism has also spurred much research, with the spread of smallpox coming under particular scrutiny[19,20].

The network structure appropriate for a given setting not only depends on the structure of the population, but on the infection itself. Within the same population, the network would be quite different for infections spread by sexual contact or by more casual contact. Even in the latter case, considerable differences would arise between infections that require prolonged close contact in order for transmission to occur and ones for which a brief encounter would be sufficient.

The contrast between networks describing sexual partnerships and more general social contact networks is particularly pronounced. It is instructive to look at some of these differences as they highlight many important aspects of network structure. The number of sexual partnerships is dwarfed by the number of social contacts in a population. An STI has far fewer chances to spread than an infection such as the common cold. Furthermore, since most individuals are monogamous (i.e. have only one sexual partner over a given time period), a large part of a sexual network consists of isolated pairs of individuals. Sexual networks often exhibit a high variance in the number of partners that different individuals have over a given time period[1,21,22]. Most individuals have just one partner, while a few individuals (such as sex workers) have a large number of partners.

In many cases, epidemiologic networks can be described by undirected graphs. Although transmission of infection is a directional event (from an infectious individual to a susceptible), the probability of transmission along an edge would often be the same if the placement of the two individuals (susceptible and infective) were reversed. Sexual transmission networks provide an example where this might not be the case, since the male to female transmission probability can differ from the female to male probability. In this setting a directional network may be more appropriate, with two directed edges between the partners having unequal transmission probabilities[23].

Transmission networks are dynamic structures: individuals' groups of contacts change over time. This is perhaps most pronounced in the case of sexual partnership networks. Partnership dynamics (the break up of existing partnerships and the formation of new partnerships) plays a major

role in the spread of infection through the large part of the network that consists of isolated pairs[8,9,14]. Considering a monogamous pair of individuals, infection can be readily transmitted between an infected individual and their susceptible partner, but further transmissions can only occur if the pair breaks up and the individuals find new susceptible partners.

The changing pattern of social contacts can have a major impact on transmission in more general settings. The classic example is provided by childhood infections, such as measles[24]. Schools are important sites for for the transmission of such infections: the congregation of children leads to much higher transmission rates during school terms than vacations. (This seasonal variation in transmission leads to large seasonal variations in disease incidence: the resulting multi-annual oscillations have been widely studied in the literature.)

The importance of the dynamic aspect of network structure depends on the timescale over which disease dynamics are of interest. For rapidly spreading infections, it is often assumed that a static network description will suffice. This leads to a considerable simplification, for both numerical simulation and mathematical analysis of transmission. As a consequence, much of the recent work has focused on static network settings.

2.3. *Epidemiologic Questions*

A large number of epidemiologically important questions can be addressed using modeling approaches. For a newly introduced infection, one may ask whether an epidemic can occur (i.e. whether the infection can invade the population), the timescale on which the ensuing outbreak will occur and the impact of the epidemic on the population (as measured, for instance, by the fraction of the population that will become infected). Questions of endemnicity and persistence of infection (whether there will just be a single outbreak, or whether the infection will be maintained within the population in the long-term) are also of interest.

An important observation is that it is often much easier to model newly introduced infections because the initial state of the system is simpler: the population is entirely susceptible. In general, though, one needs to have some idea of the susceptibility of the population. This question has been of particular interest in the context of smallpox: many people have been previously vaccinated against the disease and so the impact of any reintroduction of the infection would depend on the degree to which those individuals remain immune[19]. There are epidemiologic techniques, such as

seroprevalence surveys, that can be used to assess the susceptibility of a given community to a given infection.

From a public health viewpoint, the main questions to be addressed by modelers concern the impact of control measures: whether it is possible to prevent disease invasion, to eradicate an existing infection or the degree to which the spread of an infection can be slowed or contained. In a network setting, the key issue is understanding how the structure of the network affects transmission of the infection and whether network structure can be exploited to aid control measures targeted at the infection.

3. Describing Networks: Network Metrics

3.1. *Motivation*

A variety of network metrics are employed to describe the structure of a network. These have their origins in the mathematical theory of graphs, although some have been developed within the context of quite specific applications, such as social network theory or the exploration of large-scale technological networks. Many of these metrics describe properties that have a direct impact on transmission dynamics: we shall return to this point in a later section.

From a modeler's standpoint, such metrics can be used to ensure that their model network captures the required properties of the real-world network of interest. It is usually straightforward to calculate these metrics if the complete structure of the network is known. Unfortunately, this situation is rare in epidemiologic settings. Instead, the values must be estimated based on some sample of the network.

It is relatively straightforward to obtain information about the individuals that make up a population, either from census data or by sampling individuals. Standard statistical sampling theory can be deployed in the latter case. From a network viewpoint, however, knowledge about the composition of the population tells us little or nothing about the **structure** of the network: information about the edges of the network— describing how individuals are connected—is crucial. Thus, many network metrics require a sample of the edges of the network. Methodologies for sampling edges of networks are comparatively poorly developed, although the increasing use of network approaches is stimulating research in this area[10,11,13,25].

Our discussion of network metrics will include mention of what type of information is required in order to calculate or estimate their values. In some cases, sufficient information can be gained from just the sampled in-

dividuals. In other cases, we need to know not only about the individuals in the sample, but also about their neighbors. (In a practical setting, collection of this data clearly this involves considerable extra work.) We refer to both of these types of metrics as being local measures as they only require local information about the network. In sharp contrast, global measures require knowledge of either all or a major part of a network. Estimation of their values may be problematic in many settings.

It should also be pointed out that many (but not all) of the following metrics were developed in the context of static unweighted networks. Some of the notions carry over to more general situations. A simple way of achieving this in a dynamic network setting is to consider the edges to represent connections that existed at some point during a given time period. Many of the social networks that are commonly discussed (the actor network or scientific collaboration network) describe such 'time integrated' networks[23].

3.2. *Metrics*

A network is **connected** if it is possible to travel between any pair of individuals by moving along edges of the network. An epidemiologic interpretation of connectedness is that a single individual can transmit infection to any other individual in the population, typically via a number of intermediates. Clearly, connectedness can only be determined from global knowledge of the network. (Notice that the entire structure of the network need not be known in order to ascertain connectedness: this property can be demonstrated by finding any set of edges—a spanning set—that connects all individuals. It is much easier to show that a network is not connected: this can be achieved by finding an isolated set of individuals.)

The **degree** or **connectivity** of a node, often written as k, is equal to the number of neighbors that an individual has on the graph (that is, the number of people to whom our individual is directly connected). Since different individuals may have different numbers of neighbors, we talk about the **degree distribution**, often written as $P(k)$ or p_k, of the network. From this distribution, the average degree, written as $\bar{k}$ or $\langle k \rangle$, can be calculated as $\Sigma k p_k$. The variance of the degree distribution is given by $\sigma^2 = \Sigma(k-\bar{k})^2 p_k$. This variance equals zero if every individual has the same number of neighbors, in which case we say the network is homogeneous. Otherwise, the network is said to be heterogeneous. All of these quantities are local measures: they can be calculated once we know the connectivities of a number of individuals.

Several metrics attempt to describe the 'size' of the network. The **distance** between two nodes is the length of the shortest path that connects them. The **diameter** of a graph is the largest of these values when all pairs of nodes are examined. The **average path length** can be calculated and provides some idea of the typical number of steps between individuals on the network[4]. Clearly, one needs to have global knowledge of the network in order to calculate these quantities.

Connections between individuals are often described in terms of the mixing pattern of the network[26,27,28]. Mixing is usually described with respect to one or more relevant attributes (such as spatial location or an individual's age) and can be summarized by the **mixing matrix**. If the values that can be taken by the attribute(s) are labeled by the subscript i, then the entries of the mixing matrix, $p_{i\,j}$, depict the probabilities that a given contact of an individual of type i is with an individual of type j. In order to describe mixing patterns, the relevant attributes of both an individual and those to whom they are connected must be known.

Assortative mixing describes situations in which individuals are more likely to interact with other individuals who are similar to themselves in some respect[27,28]. **Disassortative mixing** describes the opposite situation, in which individuals tend to interact with dissimilar individuals. **Proportionate mixing** (also known as random mixing) occurs when interactions have no particular preference.

Mixing patterns have commonly been described in terms of the connectivities of individuals (Fig. 1). In this setting, assortative mixing means that highly connected individuals tend to interact with other highly connected individuals and that poorly connected individuals tend to interact with other poorly connected individuals. The opposite holds for disassortative mixing.

In order to define proportionate mixing, we imagine the process of constructing a network with a given connectivity distribution p_k. An individual of connectivity k will make k connections in the network. Listing all the connections to be made gives us a set, C, which we call the "connection pool". Since each edge of the network involves a connection between two individuals, the set C has twice as many elements as there are edges in the network. If there are N individuals in the population, then the Np_k individuals of type k contribute kNp_k connections to C. Consequently, we have that C has ΣkNp_k elements.

Proportionate mixing assumes that connections are made at random from the connection pool. Consequently, the fraction of connections that

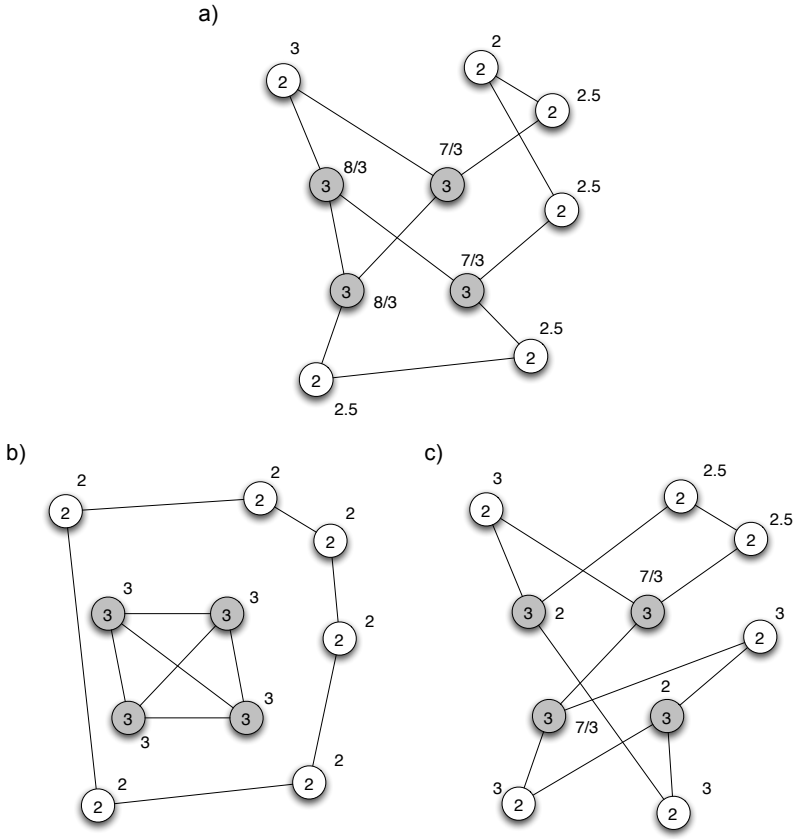

Fig. 1. Networks exhibiting (a) proportionate, (b) assortative and (c) disassortative mixing patterns. In each case, the network consists of ten individuals, six of whom have two neigbors and four of whom have three neighbors. In each of the three cases, the average connectivity of individuals is 2.4. The connection pool, C, contains $6\times2+4\times3=24$ elements, half of which arise from individuals of connectivity two and half of which arise from individuals of connectivity three. In the proportionate mixing case, therefore, half of the contacts of individuals of type two are with individuals of type two and half of their contacts are with type three individuals. The same is true for the contacts of individuals of type three. In the assortative network, greater fractions of connections are with individuals of their own type: the example network illustrates the extreme case where all contacts are amongst individuals of the same type, sometimes known as restricted mixing. In the disassortative network, more contacts are with individuals of the other type. The numbers next to the nodes depict the average connectivity of the neighbors of the particular node. Averaging these numbers illustrates the 'your friends have more friends than you do' phenomenon: the average connectivities of neighbors of individuals are given by (a) 2.5, (b) 2.4 and (c) 2.566... . The number in the proportionate mixing case is as predicted by the mean/variance formula discussed in the text, the number is lower in the assortative case and higher in the disassortative case.

are made to individuals of type k' is given by $k'Np_{k'}/\Sigma jNp_j$, regardless of the connectivity of the first individual. Notice that connections are not made at random from the population of individuals (which has connectivity distribution p_k), but rather from the connection pool (which has distribution $kp_k/\Sigma jp_j$).

An interesting consequence of proportionate mixing is that the average connectivity of the neighbors of individuals exceeds the average connectivity of individuals in the population. The former quantity can be shown to equal $\langle k\rangle + \mathrm{Var}(k)/\langle k\rangle$, which is clearly greater than $\langle k\rangle$ if the network is heterogeneous[29] (see Fig. 1).

Connectivity-based mixing patterns have commonly been used within the STI setting. Here, connectivity equates to the number of sexual partners (or, more likely, to the total number of partners over some period of time). Assortative mixing means that highly sexually active individuals tend to pair up with other highly active individuals and that individuals with few partners tend to be involved with similarly poorly connected individuals.

Another important property of networks is the degree to which they exhibit **local clustering**, also known as **cliquishness**, **mutuality** or **transitivity**[4,24,29]. Considering pairs of connected individuals, we consider how many of their neighbors are common to both of them. The existence of common neighbors leads to the appearance of triangles in the graph (i.e. paths from A to B to C and back to A, where A, B and C are vertices). One measure of local clustering is provided by ϕ, defined to equal the fraction of all triples on the graph (i.e. paths A to B to C) that form triangles[4,24,29]. Notice that this definition of clustering only looks at triples on the graph: more generally, we could ask if neighbors of connected pairs are "close" in a broader sense (e.g. whether they have distance less than or equal to some number m). A situation that would give rise to a locally clustered graph is one in which there is a strong preference for interactions to be spatially localized. We remark that clustering is a clearly a local property, although, in order to calculate ϕ, one needs to sample individuals, and ask about their neighbors and their neighbors' neighbors.

Betweenness and **centrality** attempt to quantify the importance of different individuals in terms of the population-level properties of the network[13,30]. More precisely, they provide information about the numbers of paths between pairs of nodes that pass through a given node. Clearly, these properties are global properties of the network.

Betweenness (also called betweenness centrality by some authors) measures the fraction of shortest paths in a connected component that contain

the node of interest. Let $b(j,k)$ represent all of the shortest paths between nodes j and k, and $b_i(j,k)$ represent the number of those paths that pass through node i. The betweenness of node i is then given by summing the fractions $g_i(j,k) = b_i(j,k)/b(j,k)$ over all pairs of nodes in the network[13,30].

Another measure of centrality, **information centrality**, is similar to betweenness but investigates all paths between nodes that include some other node, not just the shortest paths. The various paths are weighted according to the inverse of their lengths, thus assigning greater importance to the shorter paths which are likely to be more significant in the spread of infection[13].

Although consideration of static networks has dominated the literature to this point, several settings demand the use of dynamic networks. Most notably, sexual partnership networks change as partnerships are formed and break up. They have another notable property in that most individuals tend to be monogamous, so a large fraction of the partnership network consists of isolated nodes (singletons) who are not involved in a partnership and isolated connected pairs of nodes. Any further connections between nodes involve individuals who are involved in several simultaneous partnerships. Various measures attempt to capture this **concurrency** of partnerships [8].

3.3. *Canonical Network Types*

Given the extreme flexibility of the network approach it is often convenient to focus attention on a small set of canonical network models (Fig. 2). These are typically chosen on grounds of mathematical convenience (certain types of networks may lend themselves to the use of analytic techniques) or because they capture some particular important aspect of a more general class of networks.

The Erdös-Renyi random graph[31] is perhaps the best studied canonical network. Pairs of nodes in an N node network are independently connected at random, with per-pair connection probability p. This leads to a binomially distributed connectivity distribution, with mean $(N-1)p$. If N is sufficiently large, this distribution can be well approximated by a Poisson distribution with mean Np. This connectivity distribution is fairly closely centered about its mean: most individuals have a similar number of neighbors.

The connectedness of the graph depends on the value of Np: if this quantity is small then the graph consists of a large number of disconnected components, but when Np is large most sites are found to form a connected

component of the graph. This component is known as the 'giant component' of the graph. A celebrated theorem[31] makes this statement more precise, stating that (for large N) the random graph has a (single) giant component if and only if $\Phi = Np$ is greater than one. This component then contains a proportion z of the population, where z is the greatest root of the equation

$$z = 1 - \exp(-\Phi z). \tag{1}$$

The random nature of connections means that such graphs have little local structure, so exhibit low levels of clustering[4]. On the other hand, path lengths in random networks are relatively short. No individual is especially important in terms of the global structure of the network: since there are no preferred individuals, measures of betweenness and centrality tend to be low.

In marked contrast, connections in lattice models tend to be highly localized. Individuals are assumed to be situated on a regular lattice and are only connected to some local neighborhood. As an example, the lattice might be a rectangular lattice with individuals connected to their four nearest neighbors (up, down, left and right: the von Neumann neighborhood) or their eight nearest neighbors (up, down, left, right and diagonally: the Moore neighborhood). In order to avoid having to give special treatment to sites on the edges of the lattice, periodic boundaries conditions are sometimes imposed.

All individuals on a regular lattice (ignoring potential edge effects) have the same number of neighbors. Path lengths in lattices tend to be relatively long: one typically has to pass through a large number of intermediates in order to travel between any pair of nodes. In a one dimensional lattice, path lengths scale linearly with the network size N. Since connections are localized, lattices exhibit high values of the clustering coefficient[4]. As in the case of random graphs, there are no preferred nodes in the network so betweenness and centrality are low.

These first two canonical network types dominated the literature until Watts and Strogatz introduced "small world" networks in a paper[4] that has played a major role in stimulating interest in network modeling. Starting from a lattice model, a small world network can be generated by rewiring existing edges within the network. Each edge is examined in turn and is rewired with probability ψ: if it is to be rewired, then one of its ends is left in place and the other is reconnected to a randomly chosen node. (In an alternative formulation, connections are added between randomly chosen pairs of nodes with some probability[32].) This leads to a network that is,

in some sense, intermediate between the regular lattice and the random graph. If ψ equals zero, we have a regular lattice and if ψ equals one (all edges are rewired) then we have a random graph. When $0 < \psi << 1$, the majority of the connections are local in nature but there are a small number of long-range connections.

The surprising result of Watts and Strogatz is that it only takes a relatively small number of these long-range links to give the small world network many of the properties of the random graph. In particular, path lengths in the network rapidly decrease as ψ increases. In the small world regime, the network exhibits short path lengths (like the random graph) while still being highly locally clustered (like the lattice)[4].

The connectivity distribution of the small world network remains fairly tightly centered around its mean. This is in marked contrast to the final canonical network type that we shall consider, the scale free network. Studies of real-world technological networks (and indeed social and epidemiological networks) highlighted that many exhibit high levels of heterogeneity in their connectivity distribution. Barabási and Albert [33] proposed a mechanism by which such networks could arise: network growth with preferential attachment of edges. Starting with some initial number of nodes, additional nodes are added one by one. At each step, the new node makes m connections to existing nodes in the network. These connections are made at random, but the probability that the connection is made to a given existing node is taken to be proportional to the connectivity of that node. Thus new edges are more likely to be made to nodes that are already well connected and so "the rich get richer".

This process leads to a highly heterogeneous connectivity distribution: most individuals have few connections while a small number of individuals have a large number of connections. For the Barabási and Albert scale free network, the connectivity distribution can be shown to follow a power law, with $p_k \sim k^{-3}$. An important observation is that this distribution has infinite variance.

The highly heterogeneous nature of scale-free networks echoes an observation that has often been made by epidemiologists and sociologists in the sexual partnership setting: most individuals have few sexual partners, while a small number of individuals have a large number of partners[1]. It has been claimed that scale-free networks provide a good model for sexual partnership networks[21], although not all authors agree with this viewpoint[22].

Many other recipes for generating networks of various types have been described in the literature that has followed the work of Watts and Strogatz

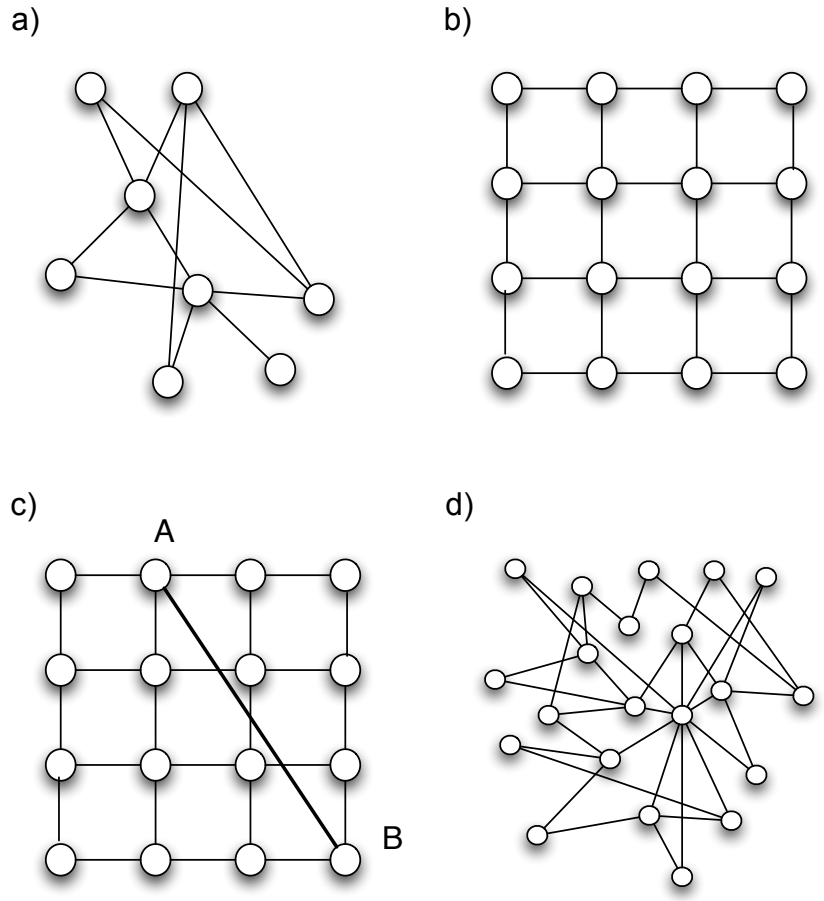

Fig. 2. Canonical network types: (a) random graph, (b) regular lattice, (c) small world network, (d) scale free network. For the small world network, notice how the addition of a single long-range link between nodes A and B of the lattice leads to the distance between the top left and bottom right nodes shrinking from six to two. Notice the heterogeneity of the scale free network: most nodes have two or three neighbors, while the most highly connected node has ten neighbors.

and Barabási and Albert. For instance scale free networks whose distributions have exponents other than minus three and ones that exhibit clustering have been produced.[34,35] For clarity, in what follows we shall reserve the term "scale free network" to mean the original Barabási and Albert formulation.

4. Epidemics on Networks

4.1. *Epidemic Processes*

In order to simulate an epidemic on a network structure, we first need to describe the natural history of the infection. The simplest descriptions are in the spirit of the compartmental models discussed at the start of this chapter. Individuals are assumed to be susceptible (S), infectious (I) or recovered (R). The SIR process assumes that susceptible individuals become infectious immediately upon infection, recover after some time, at which they acquire permanent immunity. The SIRS process assumes that immunity is not life-long, and so individuals return to the susceptible class after some time. The SIS process assumes that individuals return to the susceptible class immediately upon recovery: this corresponds to the SIRS model with a vanishingly short duration of immunity.

As an example, the following set of equations— the SIR model—is commonly used to describe the spread of a non-fatal infection in a well-mixed homogeneous closed population with no demography[1,2,3]

$$\frac{dS}{dt} = -\beta c \frac{SI}{N} \tag{2}$$

$$\frac{dI}{dt} = \beta c \frac{SI}{N} - \gamma I \tag{3}$$

$$\frac{dR}{dt} = \gamma I. \tag{4}$$

Here, the quantities S, I and R denote the numbers of susceptible, infectious and recovered individuals. The total population size, N, is constant. Births and deaths are assumed to be unimportant in this form of the model: such an assumption is appropriate if the timescale on which the epidemic plays out is short compared to the demographic timescale.

In the well-mixed setting, the transmission process is described by the the mass-action term, $\beta cSI/N$. Here, the parameter c depicts the rate at which any single individual makes contacts and the parameter β is the probability that infection would be transmitted during any one such contact. The simplest description of recovery assumes that infectious individuals recover at a constant rate, γ. We remark that this description of recovery implies that the duration of infectiousness is exponentially distributed with average $1/\gamma$. (This distribution is somewhat unrealistic biologically.)

The corresponding network model can be formulated in an analogous way. The simplest description of infection assumes that there is a constant rate (i.e. probability per unit time) at which an infective can infect a given

susceptible with whom they interact, and that this rate is identical for each edge in the network. Writing this rate as β, and noting that the interpretation of this parameter is slightly different in the network setting, the probability of transmission along a given edge over a short period of time, dt, is equal to $\beta\ dt$. Taking the recovery rate to be constant, as above, implies that an infectious individual has probability $\gamma\ dt$ of recovering over the time interval of length dt.

More general descriptions of infection and recovery are possible, such as allowing for a delay—known as the exposed period—between acquisition of infection and the start of infectiousness, or the inclusion of non-exponential distributions of infectiousness.

If the system is to be studied over a long time period, it may be necessary to include some description of the demographics (births and deaths) of the population. Deaths can be simulated by removing nodes from the network, births by adding nodes to the network.

4.2. *Basic Behavior of Epidemic Systems*

The rate at which new infections arise in the population (the incidence of infection) depends both on the number of infectious individuals (the prevalence of infection) and on the number of susceptibles. In most cases there is a threshold phenomenon related to the introduction of infection: an epidemic can only take off if the agent is sufficiently infectious that the rate at which new infections appear is greater than the rate at which infected individuals recover.

This threshold can be described by the basic reproductive number (R_0) of the system, which gives the average number of secondary infections that a single infective gives rise to in an otherwise entirely susceptible population over the course of their infectious period.

It is straightforward to derive expressions for R_0 in non-network settings. For instance, consider the early stages of an epidemic in the well-mixed SIR model described above. During this time, almost everyone will be susceptible ($S \approx N$), so the rate at which new infections occur is βcI. Each infective is giving rise to new infections at rate βc. Since infection lasts for an average of $1/\gamma$ time units, the average number of secondary infections is

$$R_0 = \frac{\beta c}{\gamma}. \tag{5}$$

4.2.1. *Dynamics in the Longer Term*

An important difference between the SIR and SIS (or SIRS) model is that the susceptible population is not replenished. In the SIR model, the progress of the epidemic continually reduces the susceptible population. Eventually, this depletion reduces the rate at which new infections can arise: SIR epidemics are self-limited and the infection eventually goes extinct.

This self-limitation typically occurs as the number of susceptibles passes below some threshold value: consequently, some fraction of the population will typically escape infection. In such settings, the severity of the epidemic is measured by the so-called size of the epidemic: the fraction (or number) of individuals who ever experience infection over the entire course of the outbreak.

In the SIS and SIRS settings, replenishment of the susceptible population means that it is possible for the infection to become permanently established in the population. In the simplest settings, the typical outcome is that the system approaches an equilibrium—the endemic equilibrium—at which there is a positive prevalence of infection. Endemic infections are possible in the SIR framework if demography is accounted for, since births provide another means by which the susceptible pool can be replenished.

5. The Impact of Network Structure on Epidemic Dynamics

Calculation of R_0 is more involved in the network setting and typically requires simplifying assumptions to be made. As an example, the presence of loops in the network is usually ignored. This enables analysis to be undertaken, albeit at the cost of neglecting some aspects of network structure—such as cliques— that may impact upon the spread of infection.

For a static network, each individual has a fixed set of contacts and so an important quantity [36,37] is the probability of transmission from an infective node to a susceptible node along a given edge over the entire duration of their infection. Newman calls this the "transmissibility" of infection[23] and represents its value by T. In the infection setting described earlier, in which infection is transmitted at rate β along a given edge and the duration of infectiousness is exponentially distributed with mean $1/\gamma$, it is easy to show that $T = \beta/(\beta + \gamma)$.

For a homogeneous network, in which every individual has k neighbors, the basic reproductive number equals

$$R_0 = T\,(k-1)\,. \tag{6}$$

Notice that the average number of secondary infections is proportional to the average number of neighbors minus one[36]. The minus one accounts for the fact that every infectious individual, except for the initial infective, must have acquired infection from one of their neighbors.

5.1. *Impact of Heterogeneity*

Heterogeneous networks must be treated with some care. In the case of proportionate (random) mixing, it is possible to show[2,23] that the basic reproductive number is given by the following formula

$$R_0 = T\left(\langle k\rangle - 1 + \frac{\mathrm{Var}(k)}{\langle k\rangle}\right). \tag{7}$$

This expression contains an extra term, involving the variance of the connectivity distribution, that leads to the value of R_0 being inflated in heterogeneous settings. This result was not unexpected, since similar "mean and variance" formulae for R_0 had earlier appeared in a wide number of epidemiological settings[1,2]. The attentive reader will notice the similarity between this result and the formula for the average connectivity of individuals' neighbors under proportionate mixing.

It should be noted that the value of R_0 no longer simply reflects the arithmetic mean of the numbers of secondary infections: in heterogeneous settings, one must adopt a more appropriate notion of the word "average" in the verbal definition of the basic reproductive number.

The appearance of the variance in formula (7) has a surprising impact on the spread of infection in scale free networks [38,39]. The basic reproductive number is infinite whenever the transmissibility is non-zero: infection can spread on a scale free network whenever there is some possibility of transmission. This result reflects the infinite variance of the connectivity distribution of the scale free network. It should be noted that this result only applies in the limit as the number of nodes becomes infinite: for a finite network, the variance will be large but can only be finite. Any real world scale free network can only have a finite number of nodes and so there would be an epidemic threshold, albeit for a much smaller transmissibility than would be the case in the corresponding homogeneous network (by which we mean a network with the same value of $\langle k\rangle$).

The impact of heterogeneity has long been recognized in the setting of sexually transmitted infections. Epidemiologists had realized that certain sections of the community, for instance highly sexually active individuals such as sex workers, were at much greater risk of infection than the gen-

eral population. Such "core groups" are responsible for a large fraction of the cases and transmission events[1,40]. The prevalence of infection is high within the core group, but low in the general population. In many cases the infection could not spread or persist without the core group: the heterogeneity in the population leads to the basic reproductive number being greater than one. This effect is often given as an explanation of why many infections are able persist at low levels in a population.

Heterogeneity in proportionate mixing settings, therefore, promotes the spread of infection compared to the corresponding homogeneous setting. Comparing two settings with the same value of R_0, heterogeneity leads to less severe outbreaks or lower prevalences of infection at endemic equilibrium, because infection tends to be concentrated amongst the highly connected individuals.

5.2. *Impact of Other Network Properties*

Local spatial structure and cliques slow the spread of infection. If the typical path length in the network is long then the infection must typically pass through many intermediates in order to cross the population. The presence of cliques results in many wasted transmission possibilities[24]: many fewer secondary infections will result if two infective individuals share a number of neighbors compared to the situation if they had no shared neighbors.

Regular lattices exhibit both long path lengths and high degrees of clustering and so lead to a slow spread of infection. The spread is, however, rapidly increased with the addition of the small number of long-range connections of the small world network. As the fraction of long-range links is increased, the speed of spread approaches that of the random graph, for which cliques are rare and path lengths are short.

Detailed exploration of such effects is far from straightforward, since they involve features of the network that are typically ignored in order to allow the use of analytic approaches. Much insight, however, has been provided by the use of approximate methods, such as the pair approximation approach[41,12,14,24,7]. In the well-mixed model (Eqs. 2-4), one only needs to know the numbers of susceptible, infectious and recovered individuals in order to describe transmission. In the network setting, transmission probabilities can be written in terms of the configuration of **pairs** of individuals: it is not enough to know how many S and I there are, one also needs to know how many susceptibles are connected to those infectives. The pair approach involves constructing differential equations that depict how the

numbers of the different types of pairs (such as S-I pairs) change over time. The difficulty with this approach is that the equations for pairs involve the numbers of triples. Typically, an approximation—a pair approximation—is employed to relate the numbers of triples to the numbers of pairs, leading to a closed set of equations.

Using the pair approximation approach, Keeling considered the impact of cliques in terms of the quantity ϕ, as defined earlier. Cliquishness was shown to reduce the value of the basic reproductive number and the severity of epidemics, with the largest impact occurring when individuals had only a small number of neighbors[7].

We remark that the impact of the core group effect discussed above can be modulated by the mixing pattern of the population. If mixing is assortative, then individuals within the core group will preferentially interact with each other, potentially giving rise to a cliquish network. With proportionate or disassortative mixing, there will be fewer interactions within the core group and lower degrees of cliquishness. An interesting observation is that the inclusion of clique structure within scale-free networks can lead to the reappearance of threshold behavior[35].

In the case of a sexual partnership network, concurrency plays a major role in the speed of spread. If all individuals were monogamous, then an infective individual could only infect a single other individual over the course of their partnership. Further transmissions could only occur with the break up of that partnership and the formation of new partnerships. Thus the spread is slowed by the time taken to break and form partnerships. Partnership concurrency enables the infection to spread from pair to pair without having to wait in this way. That concurrency can aid the spread of infection has been confirmed using both numerical[8,9] and pair approximation approaches[12,14].

6. Control of Infection

Many measures can be deployed in an attempt to control the spread of infection, such as isolation, quarantine and drug treatments. In this section, we shall focus on the use of vaccination. We may consider the effect of a perfect vaccine as preventing vaccinated nodes from acquiring and transmitting infection, essentially removing them from the network. In reality, vaccines are not perfect: not everyone gains protection against the infection, and the protection gained may only be partial.

For well-mixed models of the form (2-4), there is a critical vaccination

fraction, p_c, given by

$$p_c = 1 - 1/R_0 \tag{8}$$

such that vaccination of this fraction (or greater) of the population will guarantee eradication of the infection if it already exists, or prevent the infection from causing an outbreak in a naive population. This result makes the intuitive point that it is more difficult to eradicate a highly infectious disease than a less infectious one.

Given its impact upon the spread of infection, it is hardly surprising that network structure can have a major impact upon control of infection. Considerable attention has been directed towards the effects of heterogeneity. Anderson and May showed that uniform vaccination, in which individuals are vaccinated without regard to the heterogeneity, is always less effective than targeted vaccination and that the optimal vaccination strategy involves vaccinating those at highest risk[1]. In the case of sexually transmitted infections, this means that control measures should be directed towards the core group rather than the general population. This makes sense, particularly if the core group is responsible for the maintenance of the infection, and forms the basis of many public health policies.

Anderson and May's results were recently rediscovered in the context of vaccination of scale-free networks[42]. It was found that uniform vaccination was a completely ineffective approach since a randomly chosen individual in a scale-free network is likely to have a small number of neighbors. Removal of such individuals does little to affect the structure of the network. In contrast, removal of highly connected individuals, by targeting vaccinations, has a major impact and quickly leads to a situation in which the infection cannot spread.

One issue with targeted vaccination is that it requires the identification of individuals that are highly connected (or have some other high risk factor). This requires more effort than a simple uniform vaccination strategy. One intriguing approach[43] makes use of the fact, discussed above, that in most instances, randomly chosen neighbors of individuals have a higher connectivity than do randomly chosen individuals. A control strategy based on vaccinating randomly chosen neighbors of randomly chosen individuals can be shown to be more effective than uniform vaccination [43]. Of course, the potential benefit of this approach should be weighed up against the added complexity of its implementation.

Control measures can utilize local spatial structure, particularly during the early stages of an epidemic. If transmission is mainly local in nature,

effort can be concentrated in and around any foci of infection[16,17,20]. As an example, ring vaccination targets the area surrounding a geographically localized outbreak, much in the same way as fire-fighters might use fire breaks to contain a forest fire. Such approaches formed the cornerstone of control efforts during the 2001 outbreak of foot and mouth disease in the British livestock population [16,17]. Local control strategies become more difficult to employ as the infection becomes more widely disseminated in a given region.

The small world effect has a major impact on the use of local control strategies unless one can guarantee that long-range transmission events cannot occur. This was possible in the foot and mouth case as one of the earliest reactions of the UK authorities was to impose a ban on the movement of animals between farms. In a human setting, long-range travel such as transcontinental and intercontinental flights have reduced the entire planet to a small world, and so reliance on local control measures would appear to be unwise unless accompanied by stringent controls on travel.

7. Discussion

The highly detailed nature of network models is a double-edged sword: while they are more likely to provide a realistic framework within which the spread of infection can be studied, their complexity makes analysis difficult and very detailed population data is required in order to generate realistic networks. The deployment of network models in practical settings has been limited by this severe data requirement. Many of the instances in which network approaches have been successfully employed involve populations whose movements can be closely tracked (such as the livestock in the UK FMD outbreak).

Important work remains to be done to ascertain what data is needed in order to sufficiently characterize a network for epidemiologic modeling. As discussed above, many network properties can be deduced from knowledge of a sample of individuals, or from a sample of individuals and their contacts. Other properties are more global in nature. Even if local data is sufficient, much work remains to be done to determine the most appropriate sampling schemes and the sample sizes required for accurate characterization of networks.

Although network approaches have long been employed by epidemiological modelers, it is only with the recent increases in computing power that their simulation has become feasible for all but the most modest sizes

of networks. Input from statistical physicists, particularly with their study of large-scale technological networks, has caused a resurgence of interest in network approaches and led to many advances in our understanding. Despite this, much work remains to be done to turn theoretical studies into a practical tool that can routinely be employed by epidemiologists.

8. Acknowledgements

This work was supported by the University of North Carolina Center for AIDS Research (CFAR) by a CFAR Developmental Award. CFAR is funded by the National Institutes for Health.

References

1. R. M. Anderson and R. M. May. *Infectious Diseases of Humans: Dynamics and Control*. Oxford University Press, Oxford, 1991.
2. O. Diekmann and J. A. P. Heesterbeek. *Mathematical Epidemiology of Infectious Diseases*. John Wiley & Son, Chichester, 2000.
3. H. W. Hethcote. The mathematics of infectious diseases. *SIAM Rev.*, 42:599–653, 2000.
4. D. J. Watts and S. H. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393:440–2, 1998.
5. R. Albert, H. Jeong, and A.-L. Barabási. Diameter of the World Wide Web. *Nature*, 401:130–1, 1999.
6. M. E. J. Newman, S. H. Strogatz, and D. J. Watts. Random graphs with arbitrary degree distribution and their applications. *Phys. Rev. E*, 64:026118, 2001.
7. M. J. Keeling. The effects of local spatial structure on epidemiological invasions. *Proc. R. Soc. Lond. B.*, 266:859–67, 1999.
8. M. Kretzschmar and M. Morris. Measures of concurrency in networks and the spread of infectious disease. *Math. Biosci.*, 133:165–95, 1996.
9. M. Morris and M. Kretzschmar. Concurrent partnerships and the spread of HIV. *AIDS*, 11:641–8, 1997.
10. A. C. Ghani and G. P. Garnett. Measuring sexual partner networks for transmission of sexually transmitted diseases. *J. R. Stat. Soc. A*, 161:227–38, 1998.
11. A. C. Ghani, C. A. Donnelly, and G. P. Garnett. Sampling biases and missing data in explorations of sexual partner networks for the spread of sexually transmitted diseases. *Stat. Med.*, 17:2079–97, 1998.
12. N. M. Ferguson and G. P. Garnett. More realistic models of sexually transmitted disease transmission dynamics -sexual partnership networks, pair models, and moment closure. *Sex. Trans. Dis.*, 27:600–9, 2000.
13. A. C. Ghani and G. P. Garnett. Risks of acquiring and transmitting sexually transmitted diseases in sexual partner networks. *Sex. Trans. Dis.*, 27:579–87, 2000.

14. C. Bauch and D. A. Rand. A moment closure model for sexually transmitted disease transmission through a concurrent partnership network. *Proc. R. Soc. Lond. B.*, 267:2019–27, 2000.
15. C. J. Rhodes and R. M. Anderson. Power laws governing epidemics in isolated populations. *Nature*, 381:600–2, 1996.
16. M. J. Keeling, M. E. J. Woolhouse, D. J. Shaw, L. Matthews, M. Chase-Topping, D. T. Haydon, S. J. Cornell, J. Kappey, J. Wilesmith, and B. T. Grenfell. Dynamics of the 2001 UK foot and mouth epidemic: Stochastic dispersal in a heterogeneous landscape. *Science*, 294:813–7, 2001.
17. N. M. Ferguson, C. A. Donnelly, and R. M. Anderson. Transmission intensity and impact of control policies on the foot and mouth epidemic in Great Britain. *Nature*, 413:542–8, 2001.
18. L. A. Meyers, B. Pourbohloul, M. E. Newman, D. M. Skowronski, and R. C. Brunham. Network theory and SARS: predicting outbreak diversity. *J. Theor. Biol.*, 232:71–81, 2005.
19. M. E. Halloran, I. M. Longini Jr., A. Nizam, and Y. Yang. Containing bioterrorist smallpox. *Science*, 298:1428–32, 2002.
20. N. M. Ferguson, M. J. Keeling, W. J. Edmunds, R. Gani, B. T. Grenfell, R. M. Anderson, and S. Leach. Planning for smallpox outbreaks. *Nature*, 425:681–5, 2003.
21. F. Liljeros, C. R. Edling, L. A. Amaral, H. E. Stanley, and Y. Aberg. The web of human sexual contacts. *Nature*, 411:907–8, 2001.
22. J. H. Jones and M. S. Handcock. Social networks: Sexual contacts and epidemic thresholds. *Nature*, 423:605–6, 2003.
23. M. E. J. Newman. Spread of epidemic diseases on networks. *Phys. Rev. E*, 66:016128, 2002.
24. M. J. Keeling, D. A. Rand, and A. J. Morris. Correlation models for childhood epidemics. *Proc. R. Soc. Lond. B*, 264:1149–56, 1997.
25. M. Morris, editor. *Network Epidemiology: A Handbook for Survey Design and Data Collection*. Oxford University Press, 2004.
26. H. W. Hethcote and J. W. Van Ark. Epidemiological models for heterogeneous populations: proportionate mixing, parameter estimation, and immunization programs. *Math. Biosci.*, 84:85–118, 1987.
27. M. E. J. Newman. Assortative mixing in networks. *Phys. Rev. Lett.*, 89:208701, 2002.
28. M. E. J. Newman. Mixing patterns in networks. *Phys. Rev. E*, 67:026126, 2003.
29. M. E. J. Newman. Ego-centered networks and the ripple effect. *Social Networks*, 25:83–95, 2003.
30. K. I. Goh, E. Oh, B. Kahng, and D. Kim. Betweenness centrality correlation in social networks. *Phys. Rev. E*, 67:017101, 2003.
31. B. Bollobás. *Random Graphs*. Academic Press, 1985.
32. M. E. J. Newman, C. Moore, and D. J. Watts. Mean-field solution of the small-world network model. *Phys. Rev. Lett.*, 84:3201–4, 2000.
33. A.-L. Barabasi and R. Albert. Emergence of scaling in random networks. *Science*, 286:509–12, 1999.

34. R. Albert and A.-L. Barabasi. Topology of evolving networks: Local events and universality. *Phys. Rev. Lett.*, 85:5234–7, 2000.
35. V. M. Eguiluz and K. Klemm. Epidemic threshold in structured scale-free networks. *Phys. Rev. Lett.*, 89:108701, 2002.
36. O. Diekmann, M. C. M. De Jong, and J. A. J. Metz. A deterministic epidemic model taking account of repeated contacts between the same individuals. *J. Appl. Prob.*, 35:448–62, 1998.
37. M. J. Keeling and B. T. Grenfell. Individual-based perspectives on R(0). *J. Theor. Biol.*, 203:51–61, 2000.
38. R. Pastor-Satorras and A. Vespignani. Epidemic spreading in scale-free networks. *Phys. Rev. Lett.*, 86:3200–3, 2001.
39. R. M. May and A. L. Lloyd. Infection dynamics on scale-free networks. *Phys. Rev. E*, 64:066112, 2001.
40. J. A. Yorke, H. W. Hethcote, and A. Nold. Dynamics and control of the transmission of gonorrhea. *Sex. Trans. Dis.*, 5:51–6, 1978.
41. M. J. Keeling. Correlation equations for endemic diseases: externally imposed and internally generated heterogeneity. *Proc. R. Soc. Lond. B*, 266:953–60, 1999.
42. R. Pastor-Satorras and A. Vespignani. Immunization of complex networks. *Phys. Rev. E*, 65:036104, 2002.
43. R. Cohen, S. Havlin, and D. Ben-Avraham. Efficient immunization strategies for computer networks and populations. *Phys. Rev. Lett.*, 91:247901, 2003.