

Epidemic thresholds for infections in uncertain networks

L. Zager¹

G. Verghese

*Department of Electrical Engineering and Computer Science,
Massachusetts Institute of Technology*

Non-technical abstract: Population structure can have a significant impact on disease propagation, an impact that is often revealed by threshold tests for whether or not a disease will become an epidemic. In this paper, we begin by summarizing some recent work that attempts to clarify the relationships among different threshold tests and then apply the tools of spectral graph theory to the task of estimating these thresholds in populations whose interaction patterns are unknown.

Technical abstract: Over the last ten years, the field of mathematical epidemiology has piqued the interest of complex-systems researchers, resulting in a tremendous volume of work exploring the effects of population structure on disease propagation. Much of this research focuses on computing epidemic threshold tests, and in practice several different tests are often used interchangeably. We summarize recent literature that attempts to clarify the relationships among different threshold criteria, systematize the incorporation of population structure into a general infection framework, and discuss conditions under which interaction topology and infection characteristics can be decoupled in the computation of the basic reproductive ratio, R_0 . We then present methods for making predictions about disease spread when only partial information about the routes of transmission is available. These methods include approximation techniques and bounds obtained via spectral graph theory, and are applied to several data sets.

Number of pages: 16 (text), 4 (references), 2 (tables), 1 (figure captions), 3 (figures)

Keywords: epidemic, network, basic reproductive ratio, local asymptotic stability, spectral radius approximation, spectral graph theory, uncertain network

¹Corresponding author.

Mailing address: Rm 10-082, 77 Massachusetts Ave., Cambridge, MA 02139

Tel: 617-429-9888

Fax: 617-258-6774

E-mail: lzager@mit.edu

1 Thresholds in disease models

A major objective of mathematical epidemiology is to serve public health interests by modeling the essential characteristics of disease transmission. Until the acceptance of the germ theory of infection in the late nineteenth century, disease modeling was limited to *a posteriori* statistical analysis of outbreaks. Once the biological mechanisms of infection propagation were understood, dynamic modeling became the approach of choice. The central modeling paradigm of mathematical epidemiology is the *compartmental model*, which tracks individuals as they transition between different disease compartments, classified by the compartment's ability to acquire and transmit infection. The general model will be described in Section 2, but an orienting example is shown in Figure 1, which depicts a simple model for a susceptible-infected-susceptible (SIS) infection transmitted between two subpopulations. The state variable x_i counts the number of individuals in the i th compartment. Within each compartment, the individuals are assumed to be homogeneous with respect to disease status and behavior; different compartments might represent different disease states, susceptibilities or behaviors. This paper will explore the case in which potentially infectious contacts are constrained to occur over a contact network that links the subpopulations.

Researchers are often interested in using disease models to determine whether or not a disease will become an “epidemic”. In public health, this term is often loosely defined as “any marked upward fluctuation in disease incidence or prevalence” [1]. Often, researchers refer to a disease progression as an epidemic if a small initial infective population can grow in size, while others associate an epidemic with the establishment of an endemic presence, i.e., a sustained positive level of infection.

In stochastic models, one is often interested in the time scales over which the disease is likely to be present. For example, Ganesh et al. identify sufficient conditions for the expected time to extinction of an SIS infection to be of order $\log(n)$ (fast die-out, no epidemic) on a network of n nodes, or of order $\exp(n^\alpha)$, $\alpha > 0$ (slow die-out, or effectively endemic) [2]. For additional interpretations of “epidemic” in stochastic models, see the work of Nasell [3], Newman [4], and Miller [5].

Although the spread of infection is ideally modeled stochastically, as an individual-to-individual phenomenon, stochastic models can quickly become analytically intractable. Indeed, many results for these models are derived in the large-population limit, at which point the stochastic behavior is well-approximated by a corresponding deterministic model (for a more precise statement, see the work of Kurtz [6] [7] and Jacquez and Simon [8]). For the remainder of this paper, we will explicitly focus on deterministic models, but the results of Sections 4-5 are useful in both deterministic and stochastic settings.

In general, each of the various notions of epidemic behavior in both stochastic and deterministic models is associated with a function X of the model parameters, along with a threshold value c (which can be chosen as 1 without loss of generality) such that a disease will be an epidemic if and only if $X > c$. Table 1 presents a summary of the appearance of different notions of epidemic in deterministic models in the recent

literature. Even within the collection of deterministic models, definitions of epidemics and, correspondingly, the associated threshold tests, vary a great deal. Two of the most common epidemic definitions respectively track

- the *generation-to-generation* growth in the number of infected individuals;
- the *temporal* growth in the number of infected individuals.

Mathematically, these definitions respectively correspond to the following threshold tests:

- the basic reproductive ratio, R_0 , exceeds the threshold 1, where R_0 is canonically defined as “the expected number of secondary cases produced by a typical infected individual during its entire period of infectiousness in a completely susceptible population” [9] and is a measure of the *asymptotic per-generation growth factor* of an infection;
- a *disease-free equilibrium* (DFE) of the model is locally unstable, as determined by a threshold test on the eigenvalues of a linearized model describing the time-evolution of small initial deviations from this equilibrium.

The existence of multiple threshold criteria can create confusion when different criteria are erroneously assumed to be the same or equivalent, e.g., using the existence of an endemic equilibrium to conclude² that $R_0 > 1$. This issue has been raised by several authors, among them Heffernan et al. [14] and van den Driessche and Watmough [15]. To clarify this issue in the case of the two threshold criteria above, Section 3 of this paper will address the interpretation of the criteria using the formalism of van den Driessche and Watmough [15], establishing that these two tests ($R_0 > 1$ and local instability) are different but equivalent.

Within this mathematical framework, we then turn our attention in Section 4 to the class of models in which a population is divided into subpopulations with heterogeneously-structured interaction patterns. Under a commonly invoked set of simplifying assumptions, we demonstrate that the threshold R_0 can be factored into two terms, thereby decoupling the *biology* of infection from the *interaction patterns* of the subpopulations. Our central contribution in the paper is the introduction in Section 5 of several techniques for approximating and bounding the factor related to interaction patterns, in the face of uncertainty regarding these interaction patterns. We illustrate our ideas by application to several data sets.

2 A general compartmental model

To establish the framework for a general discussion, this paper focuses on a discrete-time model, adapted from the continuous-time framework presented by van den Driessche and Watmough in [15]. Although models based on ordinary differential equations have been the principal methodology in mathematical epidemiology, discrete-time formulations (in the form of difference equations) are a natural modeling paradigm for many applications and are more readily applicable to data sampled periodically.

²In general, the results of these threshold tests do not provide any information about the existence of endemic equilibria, or *vice versa*. In particular, many disease models exhibit *backwards bifurcations* in their equilibria structure, for example, models of recurrent immuno-suppressive infections [10], models with acquired immunity [11] and models in which interaction patterns change in response to infection [12]. For a deeper investigation of this topic, see [13].

Define a population (or state) vector $x = (x_1, \dots, x_n)$ that measures the number of individuals in each of n disease compartments, and let the first m of these compartments correspond to infected conditions (e.g., two different infected compartments might represent latent and symptomatic stages of an illness). Next, let

- $\mathcal{F}_i(x)$ represent the rate of appearance of new infections in compartment i ,
- $\mathcal{V}_i^+(x)$ represent the rate of movement of individuals into compartment i by means *other* than infection,
- $\mathcal{V}_i^-(x)$ represent the rate of removal of individuals from compartment i by any means.

Note that the distinction between terms included in $\mathcal{F}_i(x)$ and $\mathcal{V}_i^+(x)$ is not mathematical, but biological; this distinction impacts the computation of R_0 , as we'll see in the example at the end of Section 3. For the discrete-time model, the rates are measured as *changes per time step*. We can formulate a difference equation model of this process as follows:

$$x_i \leftarrow h_i(x) = x_i + \mathcal{F}_i(x) + \mathcal{V}_i^+(x) - \mathcal{V}_i^-(x) = x_i + \mathcal{F}_i(x) - \mathcal{V}_i(x) \quad (1)$$

The left arrow \leftarrow in Eq. 1 denotes a time-step update. The only restrictions placed on the form of the functions \mathcal{F}_i , \mathcal{V}_i^+ and \mathcal{V}_i^- are given by the following assumptions, suitably adapted for the discrete-time case from those given in [15]:

- (A1)** If $x \geq 0$, then $\mathcal{F}_i, \mathcal{V}_i^+, \mathcal{V}_i^- \geq 0$ for $i = 1, \dots, n$; all flows between compartments are nonnegative.
- (A2)** $\mathcal{V}_i^-(x) \leq x_i$; no more individuals can leave a compartment than currently occupy it.
- (A3)** $\mathcal{F}_i = 0$ for $i > m$; no new infections can arise in non-infected compartments.
- (A4)** If $x_i = 0$ for $i = 1, \dots, m$, then $\mathcal{F}_i(x) = 0$ and $\mathcal{V}_i^+(x) = 0$ for $i = 1, \dots, m$; when there are no infectives currently in the population, then no new infectives can arise, nor will there be any transitions into infected compartments, so the disease-free state is an invariant manifold in the dynamic model.

Assumptions (A1)-(A4) impose biologically reasonable restrictions on the behavior of any physically-based disease model, but put no limits on the functional forms that \mathcal{F}_i and \mathcal{V}_i can take. Additionally, we will take the entries of x to be real rather than integer; this approximation is routinely made in the literature and is appropriate for large population sizes.

A population vector \bar{x} will be called a *disease-free equilibrium (DFE)* if

- the first m components of \bar{x} are zero (corresponding to the absence of infected individuals);
- \bar{x} is an equilibrium of Eq. 1, i.e., $\bar{x} = h(\bar{x})$;
- all of the eigenvalues of the Jacobian matrix of the function $-\mathcal{V}$ at the equilibrium \bar{x} , denoted by $J = -D\mathcal{V}(\bar{x})$, have modulus less than one, ensuring the disease-free population dynamics (represented by the $\mathcal{V}_i(x)$) is locally stable within the disease-free invariant manifold, i.e., the equilibrium is stable to small perturbations that displace the state within this invariant manifold.

This paper will treat the case in which the population X can be broken into subpopulations X_1, \dots, X_n that interact in some constrained or patterned fashion. More specifically, we focus on the special case of heterogeneity in which all individuals have identical *biology*, i.e., each subpopulation moves through the same disease *stages* in the same manner. Note that a significant degree of heterogeneity in disease dynamics can still be incorporated within the assumption of identical biology. For example, one could allow different fractions of new infectives to exhibit different infection periods (see the example at the end of Section 3); identical biology only requires that these same dynamics apply to every subpopulation in the network.

Example: Two subpopulations with asymmetric interactions

The nature of our model is illustrated in Figure 1, which depicts two subpopulations undergoing simple susceptible-infected-susceptible (SIS) dynamics (appropriate for a non-lethal infection that can be repeatedly acquired). The dashed arrow from A to B indicates that disease can be transmitted from infected individuals in subpopulation A to susceptible individuals in subpopulation B (but not *vice versa* in this example). The state vector x will require one element for each disease stage within each subpopulation: $x = (x_1, x_2, x_3, x_4) = (I_1, I_2, S_1, S_2)$. Note that the dimension of x is the product of the number of disease stages and the number of subpopulations. Our infection model might take the particular form below:

$$\left\{ \begin{array}{l} x_1 \leftarrow x_1 + \underbrace{\beta x_1 x_3}_{\mathcal{F}_1} - \underbrace{(\gamma x_1 + dx_1)}_{\nu_1^-} \\ x_2 \leftarrow x_2 + \underbrace{\beta x_2 x_4 + \beta x_1 x_4}_{\mathcal{F}_2} - \underbrace{(\gamma x_2 + dx_2)}_{\nu_2^-} \\ x_3 \leftarrow x_3 + \underbrace{b + \gamma x_1}_{\nu_3^+} - \underbrace{(\beta x_1 x_3 + dx_3)}_{\nu_3^-} \\ x_4 \leftarrow x_4 + \underbrace{b + \gamma x_2}_{\nu_4^+} - \underbrace{(\beta x_2 x_4 + \beta x_1 x_4 + dx_4)}_{\nu_4^-} \end{array} \right. \quad (2)$$

Here b is the birthrate; β controls the rate of infection; $0 < \gamma < 1$ and $0 < d < 1$ respectively represent the fractions of the corresponding compartment populations that recover or die at each time step. The only potential DFE for this model is given by $\bar{x} = (\bar{x}_1, \bar{x}_2, \bar{x}_3, \bar{x}_4) = (0, 0, b/d, b/d)$, so each subpopulation size is $N = b/d$ at equilibrium. Note that the Jacobian matrix that governs small perturbations away from \bar{x} within the disease-free invariant manifold is given by

$$J = \begin{bmatrix} 1 - d & 0 \\ 0 & 1 - d \end{bmatrix}$$

and indeed has all eigenvalues of modulus less than one, thus satisfying the definition of a DFE.

3 R_0 and the stability of the DFE

The canonical methodology for determining R_0 for any type of deterministic infection dynamics utilizes the *next-generation operator* as defined by Diekmann et al. [9]. When the population is partitioned among only a finite number of static compartments, the next-generation operator can be written simply as a matrix, K , whose ij th element is the average number of direct infections of individuals of type i from an initial infective of type j . It is important to observe that R_0 is only defined in [9] for deterministic models; the “expected numbers” of individuals that comprise the entries of K are *population averages*, the value of that entry weighted by the fraction of the population corresponding to that value. Diekmann et al. propose that the appropriate measure for R_0 is the spectral radius, $\rho(K)$, of the matrix K . In this context, R_0 corresponds to the asymptotic per generation growth factor of the epidemic, assuming that new infections are replaced with fresh susceptibles. For nonnegative matrices like K , the spectral radius is also the largest, or dominant, eigenvalue.

Using the model of Eq. 1, the following result holds; the proof of this statement closely follows the procedure in [15] and can be found in [13].

Theorem 3.1. *Let \bar{x} be a DFE and define the $m \times m$ matrices $F = \{f_{ij}\}$ and $V = \{v_{ij}\}$ as:*

$$f_{ij} = \left. \frac{d\mathcal{F}_i}{dx_j} \right|_{\bar{x}}, \quad v_{ij} = \left. \frac{d\mathcal{V}_i}{dx_j} \right|_{\bar{x}} \quad \text{for infected compartments } i, j = 1, \dots, m$$

The next-generation matrix is given by $K = FV^{-1}$, so $R_0 = \rho(FV^{-1})$. In discrete-time models, the DFE \bar{x} is locally asymptotically stable if and only if the spectral radius of the Jacobian, $\rho(I + F - V)$, is less than 1, which in turn occurs if and only if R_0 is less than 1.

Note that the initial perturbations for which local stability is tested in Theorem 3.1 are no longer constrained to lie within the disease-free manifold. The result of Theorem 3.1 has intuitive appeal. Stability of the DFE invokes an approximation in *time*: if we replaced the system by its linearization around the DFE, an unstable DFE implies that the number of infected individuals will initially grow. More precisely, the size of the infected population cannot be kept arbitrarily small for all time, no matter how small the initial level of infection. Given a system described by the general model, which predicts the population $x[n]$ at *time* n , we can imagine constructing a related system $g[k]$ that counts the number of infected individuals in each new *generation* k of the disease. The condition on R_0 is exactly the condition for the local stability of the system $g[k]$: $R_0 > 1$ implies that the number of infected individuals per generation will initially grow. It should not be surprising, then, that the conditions for the stability of the DFE in time and by generation are identical in the parameter space of the model: the two are simply measures in different units of progression.

We stress, however, that the expressions for R_0 and the spectral radius of the Jacobian (in terms of model parameters) are not, in general, the same. This distinction is demonstrated explicitly by the following

example.

Example: Arbitrarily-distributed infectious period

In many discrete-time SIS compartmental models, the proportion of infected individuals who transition back into the susceptible state per unit time is a constant, δ . This implies a *geometric* infectious period distribution over the population, with mean $1/\delta$ (analogous to the exponential distribution in continuous-time models). As Wearing et al. have pointed out, the assumption of an exponentially-distributed infectious period can lead to erroneous results in prediction [16]. They (and others) have proposed a gamma distribution for the infectious period, as this has a tuning parameter that ‘interpolates’ between an exponential distribution and a fixed infectious period. Here, we present a model with an *arbitrarily*-distributed infectious period.

Let the infectious period be given by the discrete random variable X , which takes its values on the positive integers with $P(X = i) = q_i$. The range of values of X need not be finite, as long as X has a well-defined mean $\bar{X} = \sum_{i=1}^{\infty} i q_i$, but for ease of presentation we’ll assume that X can only take values from 1 to M . An individual, once infected, remains infected for exactly j time steps (which we shall refer to as being infected with duration j) with probability q_j . At the end of the j time steps, the individual is susceptible once again.

In order to incorporate this phenomenon into a deterministic disease model, we’ll interpret q_j as the *proportion* of infected individuals with an infectious period of exactly j time steps. Let I_{jk} denote the number of individuals who are infected with duration j and in the k th time step of their infection ($k \leq j$). Let S denote the number of susceptible individuals. We assume the same stable population dynamics as in the example of Figure 1, with b and $0 < d < 1$ representing births and deaths, respectively; β controls the rate of infection; and $N = b/d$, the population size at the unique DFE. A set of difference equations that describes this system is as follows:

$$\left\{ \begin{array}{l} S \leftarrow b + (1-d) \left[\underbrace{S}_{\text{susceptibles}} - \underbrace{S \frac{\beta}{N} \sum_{j=1}^M \sum_{k=1}^j I_{jk}}_{\text{new infections}} + \underbrace{\sum_{j=1}^M I_{jj}}_{\text{recovered infectives}} \right] \\ I_{j1} \leftarrow (1-d) \underbrace{q_j S \frac{\beta}{N} \sum_{j=1}^M \sum_{k=1}^j I_{jk}}_{\text{fraction of infectives with duration } j} \\ I_{jk} \leftarrow (1-d) \underbrace{I_{j(k-1)}}_{\text{transitions of infectives}} \quad \text{for } 1 < k \leq j \end{array} \right. \quad (3)$$

Observe that the only new infections are those that arise in the $j1$ compartments for $j = 1, \dots, M$, while the flow through the rest of the infected compartments simply represents transitions of already infected

individuals. The details of the calculation can be found in [13], but it is not difficult to show that

$$R_0 = \frac{\beta(1-d)}{d} \sum_{i=1}^M q_i (1 - (1-d)^i).$$

If the death rate is small (i.e., $d \ll 1$), then

$$R_0 \approx \beta \sum_{i=1}^M q_i i = \beta \bar{X}$$

where \bar{X} is the mean of the infectious period distribution.

By Theorem 3.1, the condition that $R_0 < 1$ is equivalent to the condition that the disease-free equilibrium is locally asymptotically stable. For the simple case of $M = 2$, where the probability of being infected with duration 1 is given by p and the probability of being infected with duration 2 is $1-p$, we can plot the spectral radius of the Jacobian J and that of the next-generation operator K as functions of p ; the result is given in Figure 2. Note that $R_0 = \rho(K) < 1$ if and only if $\rho(J) < 1$, even though the two are different functions of p . It is clear that the results of using either statistic for a threshold test are equivalent, but the threshold tests themselves are not identical. This is epidemiologically important: using the spectral radius of the Jacobian will either under- or over-estimate the basic reproductive ratio R_0 of the infection.

4 A structured population

We now apply the preceding framework to a population consisting of subpopulations or groups or “nodes” of individuals interacting across the edges of a network according to a deterministic infection model. This approach has been taken by a number of researchers; some recent examples include [17], [18], [19], [20] and [21]. We make the assumption of identical biology, as described in Section 2. How do the interaction patterns among groups, i.e., how does the *network topology*, influence the epidemic threshold R_0 ?

To begin, let us consider the simple structured population given by Figure 1 and modeled by Eq. 2. It is natural to describe the structure of the interactions by an *adjacency matrix* A , which has an entry of ‘1’ in the ij^{th} position if infected individuals in subpopulation i can infect susceptible individuals in subpopulation j , and an entry of ‘0’ otherwise; for the example in Figure 1,

$$A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}.$$

For the system of Eq. 2, we find that

$$F = \beta N \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} = \beta N A^\top \text{ and } V = (\delta + d) \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix},$$

where F and V are defined as in Theorem 3.1 and the superscript \top denotes matrix transposition. This implies that

$$R_0 = \rho(FV^{-1}) = \rho\left(\frac{\beta}{\delta + d} N A^\top\right) = \frac{\beta}{\delta + d} N \rho(A^\top). \quad (4)$$

It is not difficult to extend this example to more than two subpopulations with different interaction patterns; the general forms of F and V will remain the same.

For a more complex model with more disease stages, the assumption of identical biology allows us to generalize the factoring of R_0 in Eq. 4 using the *Kronecker product*, denoted by \otimes . If $C = \{c_{ij}\}$ is a $c_1 \times c_2$ matrix, and $D = \{d_{ij}\}$ is a $d_1 \times d_2$ matrix, then $C \otimes D$ is the $c_1 d_1 \times c_2 d_2$ matrix defined by

$$C \otimes D = \begin{bmatrix} c_{11}D & c_{12}D & \cdots & c_{1c_2}D \\ \vdots & \vdots & \ddots & \vdots \\ c_{c_1 1}D & c_{c_1 2}D & \cdots & c_{c_1 c_2}D \end{bmatrix}.$$

For models in which the ‘‘identical biology’’ assumption holds, the matrix F can be expressed as $F = F_h \otimes A^\top$, where:

- F_h is a square $m \times m$ matrix, where m is the number of infected stages, and the ij^{th} entry of F_h is the Jacobian at the DFE of the rate of new infections arising in infection stage i from individuals in infection stage j ;
- A is a *weighted adjacency matrix* whose pq^{th} entry is a scaling factor between subpopulations p and q which allows the rate of infection to vary from its nominal value in F_h due to factors like population size and interaction strength. When all pairs of interacting subpopulations have the same interaction strength (as in the example of Figure 1), A can be written as a 0–1 matrix. More generally, A will be a nonnegative matrix.

The Kronecker product $F_h \otimes A^\top$, in effect, repeats the matrix A^\top at each element of F_h . This operation restricts individuals in infection stage j to creating new infections in stage i only in those subpopulations that interact along the edges (i.e., the non-zero entries) of A . In the context of our subpopulations with identical biology, this corresponds to each subpopulation having its own set of the same disease stages through which individuals can progress; we simply repeat these stages for each subpopulation.

If we make the additional assumption that the movement between infected disease stages after initial infection is *not* a function of the state of neighboring subpopulations, then V can be factored as $V_h \otimes I$, where:

- V_h is the square $m \times m$ matrix whose ij^{th} entry represents the Jacobian around the DFE of the net rate of transitions out of infection stage i arising from individuals in infection stage j ;
- I is the $n \times n$ identity matrix.

This assumption is standard to most infection models; after infection, the progression through the remaining disease stages is an individual phenomenon and is not affected by social contacts.

The properties of Kronecker products of matrices have been well-studied; for example, for matrices A , B , C , and D of compatible dimensions, $(A \otimes B)(C \otimes D) = (AC) \otimes (BD)$. Additionally, if C and D are square matrices, $\rho(C \otimes D) = \rho(C)\rho(D)$. These properties allow the next-generation matrix K to be expressed as

$$K = (F_h \otimes A^\top)(V_h \otimes I)^{-1} = F_h V_h^{-1} \otimes A^\top \quad (5)$$

and

$$R_0 = \rho(K) = \rho(F_h V_h^{-1} \otimes A^\top) = R_h \rho(A^\top) = R_h \rho(A). \quad (6)$$

where $R_h = \rho(F_h V_h^{-1})$. The expression for R_0 in Eq. 6 has *decoupled* the biology of the infection (the progression through disease stages, summarized by R_h) and the impact of the subpopulation interaction topology (summarized by $\rho(A^\top)$). This decoupling allows us to focus separately on biological dynamics and interaction pattern issues in estimating R_0 , by separately considering the disease-specific R_h and the interaction-specific $\rho(A)$. The remainder of this paper will take advantage of this decoupling for this class of systems, and will focus on the bounding and approximation of $\rho(A)$, leaving the task of estimating R_h for a particular disease to the biological research community.

5 Uncertainty in interaction patterns

It is rare that complete information about the underlying contact network is known; a researcher might only have estimates of some of the following statistics:

- total number of nodes and edges, network girth or diameter;
- maximum or minimum degree, average degree and degree variance, degree distribution, or degree correlations;
- a collection of subgraphs (obtained, perhaps, by some network sampling method);
- parameters related to the growth mechanism underlying the creation and evolution of the network.

It is useful to interpret the unknown network structure as a random adjacency matrix \mathbf{A} , drawn from some distribution over the set of all adjacency matrices that satisfy the observed statistics. How might the stochasticity of network connections be incorporated into an otherwise deterministic disease model? For a

difference equation, this implies that the state vector is now a random vector \mathbf{x} that evolves according to the update

$$\mathbf{x} \leftarrow h(\mathbf{x}, \psi, \mathbf{A}) \quad (7)$$

where ψ is a known vector of parameters. This is a stochastic difference equation, in which every realization of the random matrix \mathbf{A} yields a different trajectory $\mathbf{x}[n]$. In general, determining the properties of the ensemble of possible trajectories $\mathbf{x}[n]$ is difficult. It is common to replace the random matrix \mathbf{A} by its expected value $E[\mathbf{A}]$, yielding a deterministic difference equation; in general, however,

$$E[\mathbf{x}[n+1]] \neq h(E[\mathbf{x}[n]], \psi, E[\mathbf{A}]).$$

Similarly, if \mathbf{A} is random, then so is the next-generation matrix \mathbf{K} , and therefore so is $\mathbf{R}_0 = \rho(\mathbf{K})$. In general, the next-generation matrix $\mathbf{K} = \mathbf{F}\mathbf{V}^{-1}$ is a nonlinear function of \mathbf{A} , as both \mathbf{F} and \mathbf{V} are functions of \mathbf{A} . However, in the case of the specially-structured populations described in the previous section, Eq. 5 shows that \mathbf{K} is a *linear* function of \mathbf{A} and therefore more amenable to analysis; in this case, $\mathbf{R}_0 = R_h \rho(\mathbf{A})$.

How does replacing the unknown adjacency matrix with its expected value, $E[\mathbf{A}]$, of R_0 ? If the network connections are undirected (i.e., a connection from subpopulation i to subpopulation j implies the existence of a connection of the same strength from j to i), then the adjacency matrix is real and symmetric, and it is not difficult to show that for these matrices,

$$\rho(E[\mathbf{A}]) \leq E[\rho(\mathbf{A})].$$

Thus, using $E[\mathbf{A}]$ leads one to underestimate the mean of the distribution of $\rho(\mathbf{A})$, which may be problematic; it means that the epidemic potential of the infection could be underestimated. For directed, weighted networks, the preceding inequality between $\rho(E[\mathbf{A}])$ and $E[\rho(\mathbf{A})]$ is no longer assured.

Additionally, there are several relevant parameters one could use to describe the distribution of the spectral radius (e.g., its mode, its maximum, an upper bound on its support); certainly $E[\rho(\mathbf{A})]$ is not necessarily the unique and best summary of the distribution. This is especially revealing when $\rho(E[\mathbf{A}])$ and $E[\rho(\mathbf{A})]$ diverge from each other as the number of subpopulations n grows; Chung et al. have identified conditions under which this divergence occurs, and present an example of a family of undirected random graphs for which this happens [22].

Given a value of R_h , if it is possible to upper bound the spectral radii of all possible realizations of \mathbf{A} by a constant c , then we use cR_h as an upper bound on \mathbf{R}_0 ; if this bound is less than one, we can conclude the local stability of the disease-free equilibrium, even in the face of uncertainty. Lower bounds on the spectral radii can similarly produce a condition for guaranteed local instability. Determining such bounds using the structural properties of the network is one of the tasks of *spectral graph theory*. In Tables 2 and 3,

a selection of upper and lower bounds is listed for the spectral radii of graphs that are simple (no self-loops or multiple edges) and connected. Tables 2 and 3 provide guaranteed bounds on the value of $\rho(\mathbf{A})$. To obtain an *approximation*, on the other hand, one can simply augment the known properties with additional assumptions that pin down the network structure.

We now present some examples illuminating the application of the ideas we have described.

Example 1: Imposing structure

Our first example explores the impact on $\rho(E[\mathbf{A}])$ of assuming various levels of structure. Suppose that only the total numbers of nodes n and edges e in the network are known. If we assume that the structure of the network is completely homogeneous, then the expected adjacency matrix will be the $n \times n$ matrix

$$E[\mathbf{A}] = \begin{bmatrix} \frac{2e}{n^2} & \cdots & \frac{2e}{n^2} \\ \vdots & \ddots & \vdots \\ \frac{2e}{n^2} & \cdots & \frac{2e}{n^2} \end{bmatrix}.$$

The largest (and only nonzero) eigenvalue of this rank-one matrix is $2e/n$. This is the average degree of a node, which we'll denote as $\langle k \rangle$.

Suppose we assume instead that the network is realized from some general degree distribution, with N_k nodes of degree k , and that the degrees of nodes are uncorrelated (i.e., the probability that nodes of degrees k_1 and k_2 are connected is proportional to $k_1 k_2$). The expected adjacency matrix will be the rank-one matrix given by the following outer product, where each vector has N_k entries of k for every k :

$$E[\mathbf{A}] = \frac{1}{\sum_k k N_k} \begin{bmatrix} 1 \\ \vdots \\ 1 \\ 2 \\ \vdots \\ 2 \\ \vdots \\ M \\ \vdots \\ M \end{bmatrix} \begin{bmatrix} 1 & \cdots & 1 & 2 & \cdots & 2 & \cdots & M & \cdots & M \end{bmatrix}. \quad (8)$$

Note that we require $\sum_k N_k = n$ and $\sum_k k N_k = 2e$. The largest (and only nonzero) eigenvalue of this matrix

is

$$\frac{\sum_k k^2 N_k}{\sum_k k N_k} = \frac{\langle k^2 \rangle}{\langle k \rangle}, \quad (9)$$

where $\langle \cdot \rangle$ indicates the average value. Comparing $\frac{\langle k^2 \rangle}{\langle k \rangle}$ to 1 is the threshold test derived by Anderson and May [23], then rederived by Pastor-Satorras and Vespignani [24]. Observe that $\frac{\langle k^2 \rangle}{\langle k \rangle} \geq \langle k \rangle$, which illustrates a more general trend: adding heterogeneity to the interaction patterns within a population increases the value of R_0 .

Thus, by supplementing known structural information with additional assumptions on interaction patterns, we can obtain an approximation of $\rho(\mathbf{A})$. The following subsections consider two further examples of bounding and approximating the largest eigenvalue of random graphs. These examples are not meant to provide definitive conclusions about the particular networks under study, but to simply illustrate the application of new tools to this task.

Example 2: Preferential attachment

Consider a network generated by a simple preferential attachment mechanism, slightly modified from the one described by Barabási and Albert [25]. A network is seeded with two nodes that have one edge between them; at each subsequent time step, a new node is added that connects to one existing node, with the probability of connection to any existing node being proportional to the existing node's degree. The procedure is terminated once the network reaches n nodes, which yields a simple, undirected network with $n - 1$ edges on n nodes, i.e., a tree. We can upper bound the maximum degree of any node in the network by $n - 1$ and can certify that the minimum degree is 1. It is known that as $n \rightarrow \infty$, the degree distribution of a preferential attachment graph follows a *power law*, in which the probability that a node has degree k is proportional to k^{-3} [25]. Figure 3 compares some theoretical upper bounds and approximations with simulation results. Curves (d) and (e) are the approximations described in Example 1; as observed in Eq. 9, making assumptions like these that reduce the heterogeneity of the network causes us to underestimate an infection's spreading potential.

Example 3: Egocentric network data

In 1997 and 1998, the U.S. National Institute on Drug Abuse sponsored a study of both drug-using and non-drug-using individuals in a low-income section of Houston, TX; this study was undertaken by Affiliated Systems Corporation and is described in [26], [27], and [28]. As part of the survey, each of 884 participants was asked to name a set of individuals within his or her social network, then asked to assess whether or not these named individuals knew *each other*. Although participants were able to name up to 18 members of their social network (in three groups of six), most of the participants named fewer than six and 237 of the participants listed none. Therefore, we restricted our attention to the first six individuals identified by each

participant and the connections among them. Participants who listed no contacts become isolated vertices in the social network, and consequently add a zero row and column to the matrix \mathbf{A} , which does not alter $\rho(\mathbf{A})$. Consequently, we will ignore these individuals and focus on the network formed by those with at least one contact. For every pair of individuals A and B named by a participant, the participant was asked whether A knew B and whether B knew A ; the 7 records whose relationships were asymmetric were removed from the set. A brief breakdown of the data set is provided in Table 4.

This kind of procedure reveals local subgraphs of the larger social network of this community, but does not connect them. This limited data, nevertheless, allows one to make several interesting approximations using several different network properties.

1. Assuming that the participants in the network were drawn from the population without regard to their number of social contacts, we can construct a histogram of the number of contacts listed by each participant as an approximation of the vertex degree distribution of the network.
2. Counting the number of edges between contacts listed by participant i is a measure of the local clustering C_i , defined as

$$C_i = \frac{2|e_{jk}|_i}{k_i(k_i - 1)}$$

where k_i is the degree of participant i and $|e_{jk}|_i$ is the number of edges between neighbors of participant i . Note that C_i is only defined if participant i listed more than one contact; let V' denote the set of such participant vertices. Following [29], we'll define the *average clustering coefficient* to be

$$C = \frac{1}{|V'|} \sum_{i \in V'} C_i.$$

The average clustering coefficient of the Houston data is calculated to be $C = 0.312$. The mean degree of participants who listed at least one contact is $d = 2.925$. Can this information be used to estimate a value of $\rho(A)$ for the network from which this data was drawn? For several of the bounds listed in Tables 2 and 3, a population size must be assumed; for these bounds, we fix $n = 1000$ individuals with at least one contact (the sensitivity of these bounds to the choice of n can be determined directly from the definition). From the degree distribution, then, it is possible to approximate several parameters (defined in Tables 2 and 3):

$$\delta = 1, \quad \Delta = 6, \quad e = \frac{dn}{2} = 1463,$$

which can be used to estimate the bounds presented in Tables 2 and 3. Observe that these results will no longer be actual bounds on the support of $\rho(\mathbf{A})$, because we have made structural assumptions in order to estimate the parameters δ , Δ and e . We can make additional structural assumptions to obtain other approximations of $\rho(\mathbf{A})$; if we assume that the degrees of adjacent nodes are uncorrelated, for example, then

the expression for $\rho(\mathbf{A})$ provided by Eq. 9 is an approximation. We summarize the results of these various computations, as well as others described below, in Figure 5.

Another approach to approximating $\rho(\mathbf{A})$ begins with the given degree distribution and average clustering coefficient, and asks what types of networks are possible? If it were possible to generate a set of networks with the observed degree distribution and clustering coefficient, a histogram could be constructed of the spectral radii of the adjacency matrices to get a sense for where $\rho(\mathbf{A})$ might fall. A procedure for doing precisely this is given by an algorithm developed by Volz in [30]. We used this algorithm to generate 100 networks with degree distributions and clustering coefficients close to those observed in the Houston data, obtained the associated adjacency matrices A_i , and recorded the mean and standard deviation of the values of $\rho(A_i)$.

Another technique for inferring global structure from local statistics chooses the parameters of a family of random graphs such that the observed graph is maximally likely; we can then use this “tuned” family to generate additional graphs that may have the same structural features. Here, we use the exponential random graph family of probability distributions (also called the ERGM or p^* family), which assumes that the probability of a given graph is an exponential function of a linear combination of relevant graph statistics; see [31] and [32]. Mathematically, this requires that the probability of a graph, denoted by a , takes the following form:

$$P(a) = \frac{1}{\kappa} \exp \left(\sum_k \theta_k z_k(a) \right) \quad (10)$$

where $z_k(a)$ is a particular graph statistic, $\theta_k \in \Re$ is a constant coefficient, and κ is a normalizing constant to ensure that $P(\cdot)$ is a valid probability distribution. In general, the statistics $z_i(a)$ can be any functions of the information that one has about the network, including both structural properties (like the strength and directionality of edges) and node identity properties (such as the gender or age of the individual represented by the node). We apply the exponential random graph structure to the Houston data to generate two different approximations, which differ in their choice of network statistics:

- ERGM-A - $z_k(a)$ comprise the number of edges and the number of triangles;
- ERGM-B - $z_k(a)$ comprise the number of edges, number of triangles, and degree distribution of the observed data.

To determine the optimal θ_k associated with each of these statistics within each of these models and then to generate draws from the resulting distribution, we use the *statnet* package for the R programming language [33]. This freely-available package utilizes Markov Chain Monte Carlo simulation techniques to produce pseudo-maximum-likelihood estimates of the θ_k ; more details can be found in a recent special volume of the *Journal of Statistical Software* [34].

For the Volz and ERGM approximations, histograms of the resulting values of $\rho(\mathbf{A})$ are depicted in Figure 4. The means of these respective histograms, along with the bounds and approximations described

earlier, are summarized in Figure 5. This figure suggests that it is likely the spectral radius of the unknown adjacency matrix is much closer to the approximate lower bounds than the upper bounds.

6 Observations and conclusions

Because of its prominence in the domain of mathematical epidemiology, the definition and use of R_0 naturally inspires both confusion and controversy. For example, the value of R_0 calculated for a vector-borne disease represents the average number of new infections in both vector and host, per generation. For public health decisions regarding diseases that alternate between two populations, one is often only interested in the behavior of the epidemic in just one of the subpopulations, so the even or odd powers of the next-generation matrix are more relevant. Roberts and Heesterbeek [35] suggest the use of an alternate statistic, also derived from the next-generation operator, which may be more appropriate when control measures can only be applied to a subpopulation of the affected individuals (for example, host treatment options for host-vector diseases).

Our observations in this note have been limited to discussion of the stability of various fixed points of the infection model, but any public health practitioner can point to many examples of disease behavior occurring periodically. Many disease models exhibit oscillatory behavior, either at a natural frequency or in response to periodic forcing. For example, Wearing and Rohani [36] observe that both seasonal variation and heterogeneity in infectivity seem to be required to explain the observed oscillations in the prevalence of dengue in Thailand. Additionally, we have only addressed the relationship between a threshold on R_0 and the stability of a disease-free equilibrium; what might happen if the population under study is in a stable disease-free limit cycle when an infection is introduced? In discrete time, analysis of T -periodic behavior requires examining the stability of fixed points of the state-evolution map $h(\cdot)$ of Eq. 7, composed with itself T times; in [37], Franke and Yakubu analyze a discrete-time SIS model in a periodic environment.

After looking at threshold conditions on the spectra of networks, a natural next step is to explore the spatial patterns of epidemic outbreaks. One might intuitively begin with the *eigenvector* of K associated with the eigenvalue R_0 . In most applications, this dominant eigenvector is typically interpreted as a centrality measure (see, for example, [38] and [39]); if disease is introduced into a population proportionally across subpopulations according to this eigenvector, then its initial spread will be maximally fast. What other kinds of information about spatial spread are embedded in the eigenvectors of K , or related matrices like the Laplacian of the graph? What kinds of inferences about spatial spread can be made for networks whose structure is not completely known?

Acknowledgments

The raw data from the Houston study was generously provided by Dr. Isaac Montoya of Affiliated Systems Corporation. We'd like to thank Richard Larson for helpful discussions, and the anonymous reviewer for improving the clarity and scope of this work. Support for this work was provided by the Martin Family Foundation for Environmental Sustainability and the Women's Technology Program at MIT.

References

- [1] D.J.P. Barker. *Practical epidemiology*. Churchill Livingstone, 1973.
- [2] A. Ganesh, L. Massoulié, and D. Towsley. The effect of network topology on the spread of epidemics. In *Proceedings of IEEE INFOCOM 2005*, volume 2, pages 1455–1466, 2005.
- [3] I. Nasell. The threshold concept in stochastic epidemic and endemic models. In D. Mollison, editor, *Epidemic models: their structure and relation to data*, pages 71–83. Cambridge University Press, 1995.
- [4] M.E.J. Newman. The spread of epidemic disease on networks. *Physical Review E*, 66:016128, 2002.
- [5] J.C. Miller. Epidemic size and probability in populations with heterogeneous infectivity and susceptibility. *Physical Review E*, 76:010101, 2007.
- [6] T. Kurtz. Limit theorems for sequences of jump Markov processes approximating ordinary differential processes. *Journal of Applied Probability*, 8(2):344–356, 1971.
- [7] T. Kurtz. The relationship between stochastic and deterministic models for chemical reactions. *Journal of Chemical Physics*, 57(7):2976–2978, 1972.
- [8] J. A. Jacquez and C. P. Simon. The stochastic SI model with recruitment and deaths - 1. comparison with the closed SIS model. *Mathematical Biosciences*, 117(1-2):77–125, 1993.
- [9] O. Diekmann, J. A. P. Heesterbeek, and J.A.J. Metz. On the definition and the computation of the basic reproduction ratio R_0 in models for infectious diseases in heterogeneous populations. *Journal of Mathematical Biology*, 28:365–382, 1990.
- [10] M. Safan, H. Heesterbeek, and K. Dietz. The minimum effort required to eradicate infections in models with backward bifurcation. *Journal of Mathematical Biology*, 53(4):703–718, 2006.
- [11] T. Reluga, J. Medlock, and A. Perelson. Backward bifurcations and multiple equilibria in epidemic models with structured immunity. *Journal of Theoretical Biology*, 252(1):155–165, 2008.
- [12] T. Gross, C. J. D. D’Lima, and B. Blasius. Epidemic dynamics on an adaptive network. *Physical Review Letters*, 96(20), 2006.
- [13] L. Zager. *Thresholds in infection processes on networks with structural uncertainty*. Ph.D. Thesis, EECS Dept., Massachusetts Institute of Technology, 2008.
- [14] J. M. Heffernan, R. J. Smith, and L. M. Wahl. Perspectives on the basic reproductive ratio. *Journal of the Royal Society Interface*, 2(4):281–293, 2005.
- [15] P. van den Driessche and J. Watmough. Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission. *Mathematical Biosciences*, 180:29–48, 2002.
- [16] H. J. Wearing, P. Rohani, and M. J. Keeling. Appropriate models for the management of infectious diseases. *PLoS Medicine*, 2(7):621–627, 2005.
- [17] N. G. Becker and K. Dietz. The effect of household distribution on transmission and control of highly infectious-diseases. *Mathematical Biosciences*, 127(2):207–219, 1995.
- [18] M. Boguna and R. Pastor-Satorras. Epidemic spreading in correlated complex networks. *Physical Review E*, 66:047104, 2002.
- [19] A. N. Hill and I. M. Longini. The critical vaccination fraction for heterogeneous epidemic models. *Mathematical Biosciences*, 181(1):85–106, 2003.

- [20] I. Z. Kiss, D. M. Green, and R. R. Kao. Infectious disease control using contact tracing in random and scale-free networks. *Journal of the Royal Society Interface*, 3(6):55–62, 2006.
- [21] N. Masuda and N. Konno. Multi-state epidemic processes on complex networks. *Journal of Theoretical Biology*, 243(1):64–75, 2006.
- [22] F. Chung, L. Lu, and V. Vu. Spectra of random graphs with given expected degrees. *Proceedings of the National Academy of Sciences of the United States of America*, 100(11):6313–6318, 2003.
- [23] R. Anderson and R. M. May. *Infectious diseases of humans: dynamics and control*. Oxford University Press, 1991.
- [24] R. Pastor-Satorras and A. Vespignani. Epidemic dynamics in finite size scale-free networks. *Physical Review E*, 65:035108(R), 2002.
- [25] A. Barabasi and R. Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, 1999.
- [26] D. C. Bell and R.A. Trevino. Modeling HIV risk. *Journal of Acquired Immune Deficiency Syndromes*, 22:280–287, 1999.
- [27] D. C. Bell, I. D. Montoya, and J. S. Atkinson. Partner concordance in reports of joint risk behaviors. *Journal of Acquired Immune Deficiency Syndromes*, 25:173–181, 2000.
- [28] D. C. Bell, I. D. Montoya, J. S. Atkinson, and S. J. Yang. Social networks and forecasting the spread of HIV infection. *Journal of Acquired Immune Deficiency Syndromes*, 31(2):218–229, 2002.
- [29] T. Schank and D. Wagner. Approximating clustering coefficient and transitivity. *Journal of Graph Algorithms and Applications*, 9(2):265–275, 2005.
- [30] E. Volz. *Random networks with tunable degree distribution and clustering*. M.S. Thesis, Dept. of Sociology, Cornell University, 2004.
- [31] J. Park and M. E. J. Newman. Solutions of the two-star model of a network. *Physical Review E*, 70:066146, 2004.
- [32] G. Robins, P. Pattison, Y. Kalish, and D. Lusher. An introduction to exponential random graph (p^*) models for social networks. *Social Networks*, 29:173–191, 2007.
- [33] M. Handcock, D.R. Hunter, C. Butts, S. M. Goodreau, and M. Morris. statnet: Software tools for the Statistical Modeling of Network Data. <http://statnetproject.org/>, 2003.
- [34] M. Handcock, D.R. Hunter, C. Butts, S. M. Goodreau, and M. Morris. ergm: A Package to Fit, Simulate and Diagnose Exponential-Family Models for Networks. *Journal of Statistical Software*, 24(4), 2008.
- [35] M. G. Roberts and J. A. P. Heesterbeek. A new method for estimating the effort required to control an infectious disease. *Proceedings of the Royal Society of London Series B-Biological Sciences*, 270(1522):1359–1364, 2003.
- [36] H. J. Wearing and P. Rohani. Ecological and immunological determinants of dengue epidemics. *Proceedings of the National Academy of Sciences of the United States of America*, 103(31):11802–11807, 2006.
- [37] J. E. Franke and A. A. Yakubu. Discrete-time SIS epidemic model in a seasonal environment. *SIAM Journal on Applied Mathematics*, 66(5):1563–1587, 2006.
- [38] G.S. Canright and K. Engo-Monsen. Spreading on networks: a topographic view. *Complexus*, 3:131–146, 2006.

- [39] A. S. Klovdahl, E. A. Graviss, A. Yaganehdoost, M. W. Ross, A. Wanger, G. J. Adams, and J. M. Musser. Networks and tuberculosis: an undetected community outbreak involving public places. *Social Science & Medicine*, 52(5):681–694, 2001.
- [40] G. R. Fulford, M. G. Roberts, and J. A. P. Heesterbeek. The metapopulation dynamics of an infectious disease: Tuberculosis in possums. *Theoretical Population Biology*, 61(1):15–29, 2002.
- [41] C. Fraser, S. Riley, R. Anderson, and N. M. Ferguson. Factors that make an infectious disease outbreak controllable. *Proceedings of the National Academy of Sciences of the United States of America*, 101(16):6146–6151, 2004.
- [42] Y. Wang, D. Chakrabarti, C. Wang, and C. Faloutsos. Epidemic spreading in real networks: an eigenvalue viewpoint. In *SRDS 2003*, pages 25–34, Florence, Italy, 2003.
- [43] J. M. Hyman and J. Li. An intuitive formulation for the reproductive number for the spread of diseases in heterogeneous populations. *Mathematical Biosciences*, 167(1):65–86, 2000.
- [44] M. E. Alexander and S. M. Moghadas. Bifurcation analysis of an SIRS epidemic model with generalized incidence. *SIAM Journal on Applied Mathematics*, 65(5):1794–1816, 2005.
- [45] I. Z. Kiss, D. M. Green, and R. R. Kao. The effect of contact heterogeneity and multiple routes of transmission on final epidemic size. *Mathematical Biosciences*, 203(1):124–136, 2006.
- [46] M. J. Keeling. The effects of local spatial structure on epidemiological invasions. *Proceedings of the Royal Society of London Series B-Biological Sciences*, 266(1421):859, 1999.
- [47] J. M. Hyman and J. Li. Differential susceptibility and infectivity epidemic models. *Mathematical Biosciences and Engineering*, 3(1):89–100, 2006.
- [48] R. Pastor-Satorras and A. Vespignani. Epidemic dynamics and endemic states in complex networks. *Physical Review E*, 63:066117, 2001.
- [49] K. B. Blyuss and Y. N. Kyrlychko. On a basic model of a two-disease epidemic. *Applied Mathematics and Computation*, 160(1):177–187, 2005.
- [50] J. A. Hyman and J. Li. The reproductive number for an HIV model with differential infectivity and staged progression. *Linear Algebra and Its Applications*, 398:101–116, 2005.
- [51] M. Salmani and P. van den Driessche. A model for disease transmission in a patchy environment. *Discrete and Continuous Dynamical Systems-Series B*, 6(1):185–202, 2006.
- [52] J. Arino and P. van den Driessche. The basic reproduction number in a multi-city compartmental epidemic model. In *Positive Systems, Proceedings*, volume 294 of *Lecture Notes in Control and Information Sciences*, pages 135–142. Springer Berlin/Heidelberg, 2003.
- [53] F. Juhász. On the spectrum of a random graph. In L. Lovász and V.T. Sós, editors, *Algebraic Methods in Graph Theory*, volume I, pages 313–316. North-Holland Publishing Company, 1981.
- [54] R.P. Stanley. A bound on the spectral radius of graphs with e edges. *Linear Algebra and Its Applications*, 67:267–269, 1987.
- [55] Y. Hong, J. L. Shu, and K. Fang. A sharp upper bound of the spectral radius of graphs. *Journal of Combinatorial Theory, Series B*, 81:177–183, 2001.
- [56] K. C. Das and P. Kumar. Some new bounds on the spectral radius of graphs. *Discrete Mathematics*, 281(1-3):149–161, 2004.

- [57] O. Favaron, M. Maheo, and J.-F. Sacle. Some eigenvalue properties in graphs (conjectures of Graffiti - II). *Discrete Mathematics*, 111:197–220, 1993.
- [58] X. Zhang. Eigenvectors and eigenvalues of non-regular graphs. *Linear Algebra and Its Applications*, 409:79–86, 2005.
- [59] J. L. Shu and Y. R. Wu. Sharp upper bounds on the spectral radius of graphs. *Linear Algebra and Its Applications*, 377:241–248, 2004.
- [60] V. Nikiforov. Bounds on graph eigenvalues I. *Linear Algebra and Its Applications*, 420(2-3):667–671, 2007.
- [61] M. Lu, H. Q. Liu, and F. Tian. A new upper bound for the spectral radius of graphs with girth at least 5. *Linear Algebra and Its Applications*, 414(2-3):512–516, 2006.
- [62] L. Collatz and U. Sinogowitz. Spektren endlicher grafen. *Abh. Math. Sem. Univ. Hamburg*, 21:63–77, 1957.

Tables

Approach	Reference
Approach	Reference
<i>next-generation operator</i>	[9], [17], [40], [41]
<i>local stability of disease-free equilibrium</i>	[18], [19], [42], [43], [44], [45], [46], [47]
<i>existence of an endemic equilibrium</i>	[23], [48], [21]
<i>multiple criteria</i>	[49], [50], [51], [52]

Table 1: Approaches taken to computing an epidemic threshold in a sampling of the literature.

Table 2: Upper bounds on $\rho(\mathbf{A})$ for simple, connected graphs. Structural properties are number of nodes (n), number of edges (e), maximum degree (Δ), minimum degree (δ), girth (G), diameter (D), degree of node i (d_i), and average degree of the neighbors of node i (m_i).

structural information	upper bound on $\rho(\mathbf{A})$	reference
$\{e\}$, self-loops allowed	$\sqrt{2e}$	[53]
$\{e\}$	$\frac{-1+\sqrt{1+8e}}{2}$	[54]
$\{n, \delta, e\}$	$\frac{(\delta-1)+\sqrt{(\delta+1)^2+4(2e-\delta n)}}{2}$	[55]
$\{m_i\}$	$\max\{\sqrt{m_i m_j} (i, j) \in E\}$	[56]
$\{d_i, m_i\}$	$\max\{\sqrt{d_i m_j} (i, j) \in E\}$	[57]
$\{n, e, \delta, \Delta\}$	$\sqrt{2e - (n-1)\delta + (\delta-1)\Delta}$	[56]
$\{n, D, \Delta, \delta\}$	$\Delta - \frac{\Delta+\delta-2\sqrt{\Delta\delta}}{Dn\Delta}$	[58]
$\{d_i\}$	$\min_{1 \leq i \leq n} \frac{d_i - 1 + \sqrt{(d_i+1)^2 + 4(i-1)(d_1-d_i)}}{2}$	[59]
$G \geq 5, \{n, \Delta\}$	$\min(\Delta, \sqrt{n-1})$	[60]
$G \geq 5, \{n, \Delta\}$	$\frac{-1+\sqrt{4n+4\Delta-3}}{2}$	[61]

Table 3: Lower bounds on $\rho(\mathbf{A})$. Structural properties are number of nodes (n), number of edges (e), maximum degree (Δ), minimum degree (δ), node degrees listed in descending order ($d_i \geq d_j$ for $i < j$), and average degree of the neighbors of node i (m_i).

structural information	lower bound on $\rho(\mathbf{A})$	reference
$\{n\}$, connected	$2 \cos \frac{\pi}{n+1}$	[62]
$\{\Delta\}$, simple	$\sqrt{\Delta}$	[57]
$\{n, e\}$, no multiple edges	$\frac{2e}{n}$	[53]
$\{n, d_i\}$, simple	$\sqrt{\frac{1}{n} \sum_i d_i^2}$	[57]
$\{n, d_i\}$, simple	$\frac{1}{e} \sum_{(i,j) \in E} \sqrt{d_i d_j}$	[57]

Table 4: A breakdown of the participant reports of social contacts in the Houston data set.

884 records	{	237 empty 7 asymmetric 640 symmetric & non-empty	{	166 listed ≤ 1 contact 444 listed ≥ 2 contacts	{	152 listed 2 contacts 112 listed 3 contacts 72 listed 4 contacts 50 listed 5 contacts 88 listed 6 contacts
-------------	---	--	---	---	---	--

Figure Captions

Figure 1. An example compartmental model for a susceptible-infected-susceptible (SIS) disease (see Eq. 2). Here, A and B represent two different subpopulations of individuals. The S and the I represent susceptible and infected compartments within each subpopulation. The dashed arrow from A to B indicates that disease can be transmitted from infected individuals in subpopulation A to susceptible individuals in subpopulation B . Births occur into the susceptible compartments, and deaths remove individuals from both compartments.

Figure 2. A comparison of the value of R_0 and the spectral radius of the Jacobian J for the discrete-time SIS model of Eq. 3 ($\beta = 2/3$ and $d = 0$).

Figure 3. (a) The mean (\pm std. dev.) of the spectral radius of the adjacency matrix of a simple preferential attachment model on n nodes ($n - 1$ edges), taken over 100 trials; (b) an upper bound on $\rho(\mathbf{A})$ obtained using the number of edges [54] (coincides with [61]); (c) an upper bound obtained using the number of nodes, edges, minimum degree and maximum degree [56] (coincides with [60]); (d) approximation assuming a degree distribution $\sim k^{-3}$, corresponding to preferential attachment, without degree correlations; (e) approximation assuming a homogeneous network on n nodes with $n - 1$ edges distributed identically.

Figure 4. Histograms of the values of $\rho(\mathbf{A})$ observed over 100 graphs drawn from each of the three simulation methods.

Figure 5. Bounds, approximations and simulation results for $\rho(\mathbf{A})$ based on the Houston data degree distribution and clustering statistics. Upper bounds are indicated by the convex curves, lower bounds by the concave curves, approximations by \times and simulation results by a horizontal line.

Figures

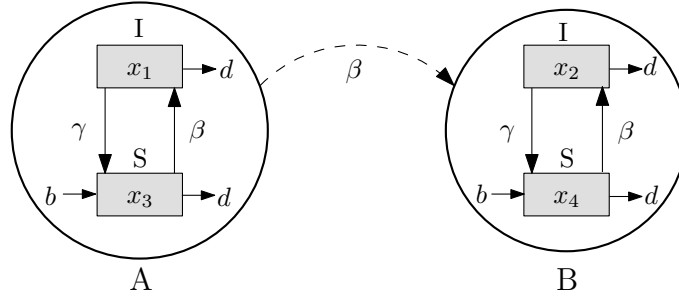


Figure 1: An example compartmental model for a susceptible-infected-susceptible (SIS) disease (see Eq. 2). Here, A and B represent two different subpopulations of individuals. The S and the I represent susceptible and infected compartments within each subpopulation. The dashed arrow from A to B indicates that disease can be transmitted from infected individuals in subpopulation A to susceptible individuals in subpopulation B . Births occur into the susceptible compartments, and deaths remove individuals from both compartments.

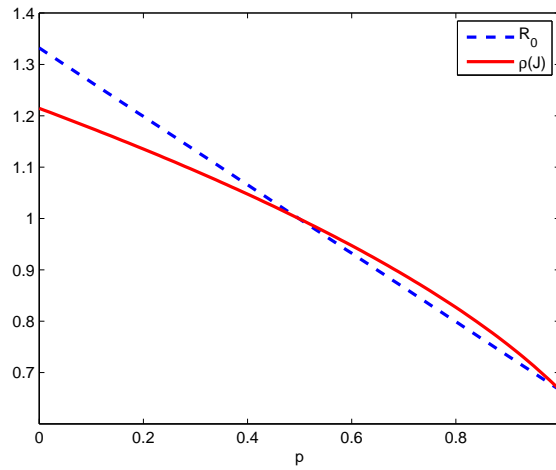


Figure 2: A comparison of the value of R_0 and the spectral radius of the Jacobian J for the discrete-time SIS model of Eq. 3 ($\beta = 2/3$ and $d = 0$).

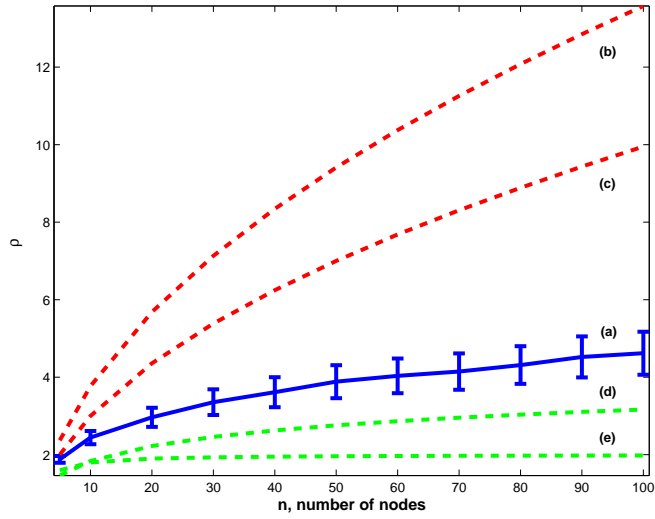


Figure 3: (a) The mean (\pm std. dev.) of the spectral radius of the adjacency matrix of a simple preferential attachment model on n nodes ($n - 1$ edges), taken over 100 trials; (b) an upper bound on $\rho(\mathbf{A})$ obtained using the number of edges [54] (coincides with [61]); (c) an upper bound obtained using the number of nodes, edges, minimum degree and maximum degree [56] (coincides with [60]); (d) approximation assuming a degree distribution $\sim k^{-3}$, corresponding to preferential attachment, without degree correlations; (e) approximation assuming a homogeneous network on n nodes with $n - 1$ edges distributed identically.

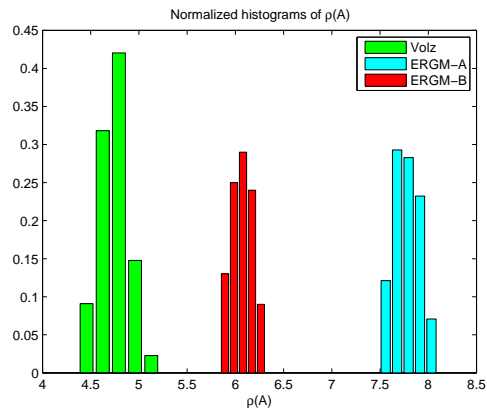


Figure 4: Histograms of the values of $\rho(\mathbf{A})$ observed over 100 graphs drawn from each of the three simulation methods.

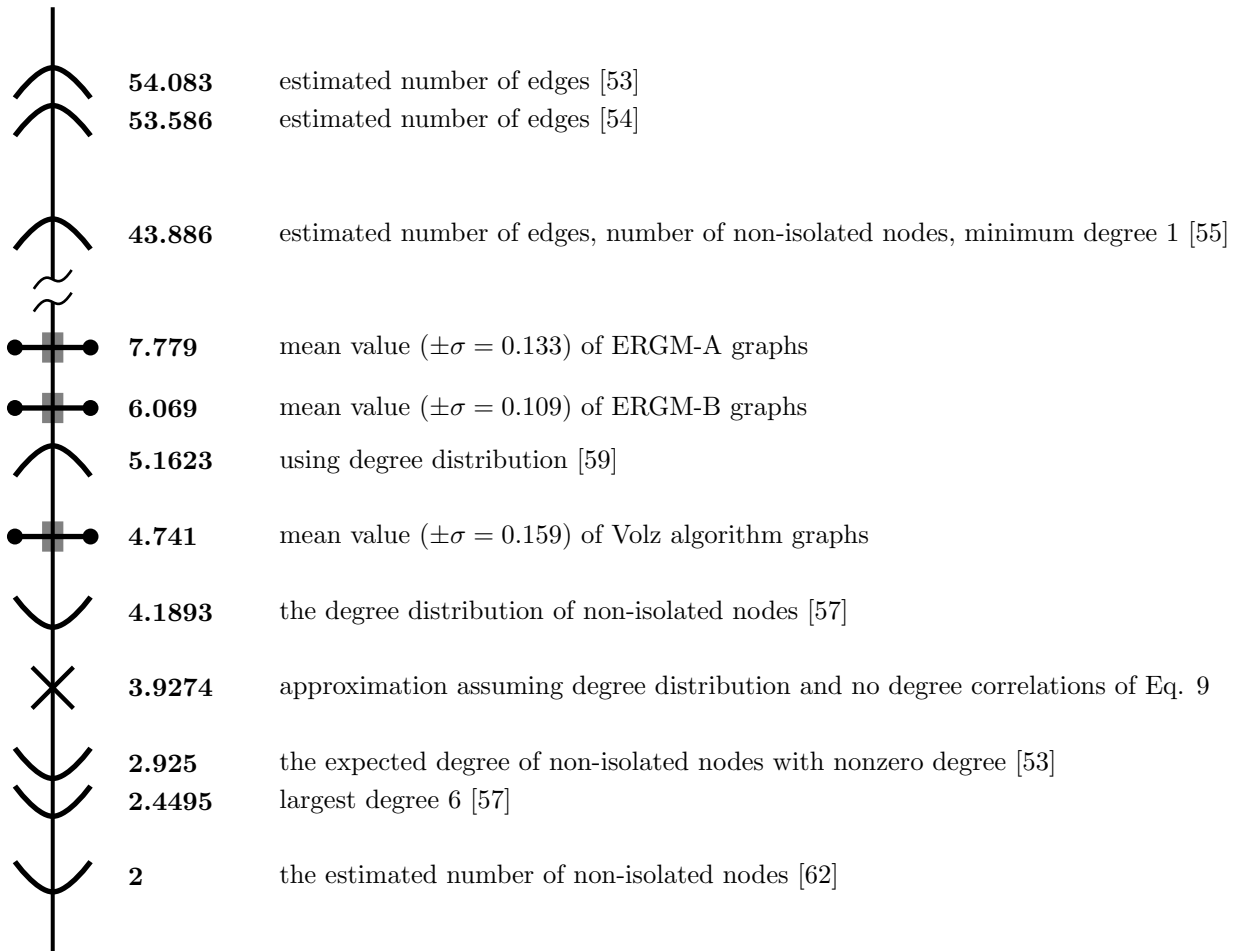


Figure 5: Bounds, approximations and simulation results for $\rho(\mathbf{A})$ based on the Houston data degree distribution and clustering statistics. Upper bounds are indicated by the convex curves, lower bounds by the concave curves, approximations by \times and simulation results by a horizontal line.