Infection processes on networks with structural uncertainty

by

Laura A. Zager

B.A. Mathematics and B.S. Engineering, Swarthmore College (2003)
S.M., Electrical Engineering and Computer Science, MIT (2005)

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2008

Author____
Department of Electrical Engineering and Computer Science
August 22, 2008

Certified by____
George Verghese
Professor of Electrical Engineering and Computer Science
Thesis Supervisor

Accepted by____
Terry P. Orlando
Professor of Electrical Engineering and Computer Science
Chair, Department Committee on Graduate Students

Infection processes on networks with structural uncertainty

by

Laura A. Zager

Submitted to the Department of Electrical Engineering and Computer Science
on August 22, 2008, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

Over the last ten years, the interest in network phenomena and the potential for a global pandemic have produced a tremendous volume of research exploring the consequences of human interaction patterns for disease propagation. The research often focuses on a single question: will an emerging infection become an epidemic? This thesis clarifies the relationships among different epidemic threshold criteria in deterministic disease models, and discusses the role and meaning of the basic reproductive ratio, R_0. We quantify the incorporation of population structure into this general framework, and identify conditions under which interaction topology and infection characteristics can be decoupled in the computation of threshold functions, which generalizes many existing results in the literature. This decoupling allows us to focus on the impact of network topology via the spectral radius of the adjacency matrix of the network.

It is rare, however, that one has complete information about every potential disease-transmitting interaction; this uncertainty in the network structure is often ignored in deterministic models. Neglecting this uncertainty can lead to an underestimate of R_0, an unacceptable outcome for public health planning. Is it possible to make guarantees and approximations regarding disease spread when only partial information about the routes of transmission is known? We present methods for making predictions about disease spread over uncertain networks, including approximation techniques and bounding results obtained via spectral graph theory, and illustrate these results on several data sets. We also approach this problem by using simulation and analytical work to characterize the spectral radii that arise from members of the exponential random graph family, commonly used to model empirical networks in quantitative sociology. Finally, we explore several issues in the spatiotemporal patterns of epidemic propagation through a network, focusing on the behavior of the contact process and the influence model.

Thesis Supervisor: George Verghese
Title: Professor of Electrical Engineering and Computer Science

Acknowledgements

My advisor, George Verghese, has been a paragon of patience and flexibility as I've followed my research interests over the past five years, as well as an endless source of creativity and insight and fixer of my "which's" and "that's"; you have my infinite admiration and gratitude. Many thanks are also due to my committee members, Richard Larson and Tamara Awerbuch-Friedlander, both of whom have introduced me to incredibly stimulating projects, ideas and people.

Alan Oppenheim was my very first mentor here at MIT; the care he puts into student advising is remarkable, and it's fantastic to see his philosophy taking root across the department. Richard Levins of the Harvard School of Public Health encouraged my interest in mathematical ecology and welcomed me warmly into his and Dr. Awerbuch-Friedlander's group of colleagues. Cynthia Skier and the rest of the Women's Technology Program staff for 2007 and 2008 have made my last two summers at MIT an absolute joy.

This thesis is dedicated to my family and friends. To Holly Waisanen-Hatipoglu, Alejandro Dominguez-Garcia, Victor Preciado, Bill Richoux, Erin Aylward and Tushar Parlikar: beacons of encouragement and humor and the greatest assets of the Institute. To Mary Carradine, Zac Marconi, John and Tanya and Elliott Brehm, Meg McDermott, Ryan Placko, Nate Young, and James and Philitsa and Eliana Hanson: all amazing, all the time. To Aimee Schultz, Alyssa Bonnoit, Andy Zuppann, Abram Falk, Mark Romanowsky and Christine James: I can't wait to see where you'll head next.

To Mom and Dad: the two of you have been with me for every second of this process, and there aren't any words for how deep and important that support has been. I'm totally overwhelmed with love for you. Oh, thank you. To Lisa: the person I most hope to be like when I grow up. To Gramma: the totally awesome glue that holds together the people that are the most precious to me. To Mike and Cheri and Larry and Robin and Karen and Melissa and Cousin Zach: there's nothing I look forward to more than coming home to see you. And eat pie.

To Kieran and Penny: the most incredible things on two and four legs, respectively.

But really, to Kieran.

This material is based upon work supported in part by a National Science Foundation Graduate Research Fellowship. Any opinions, findings, conclusions or recommendations expressed in this publication are those of the author and do not necessarily reflect the views of the National Science Foundation. Additional support for this work was provided by the Martin Family Fellowship for Environmental Sustainability, the MIT Women's Technology Program, and a graduate Instructor appointment in MIT's EECS Department.

Contents

Chapter 1
Mathematical epidemiology and networks

THIS chapter will motivate the work of this thesis by considering the history of mathematical modeling of disease, its central modeling approaches, and the critical importance of network methods to public health in a globalized world. We'll conclude by outlining the contributions of this thesis to this body of research.

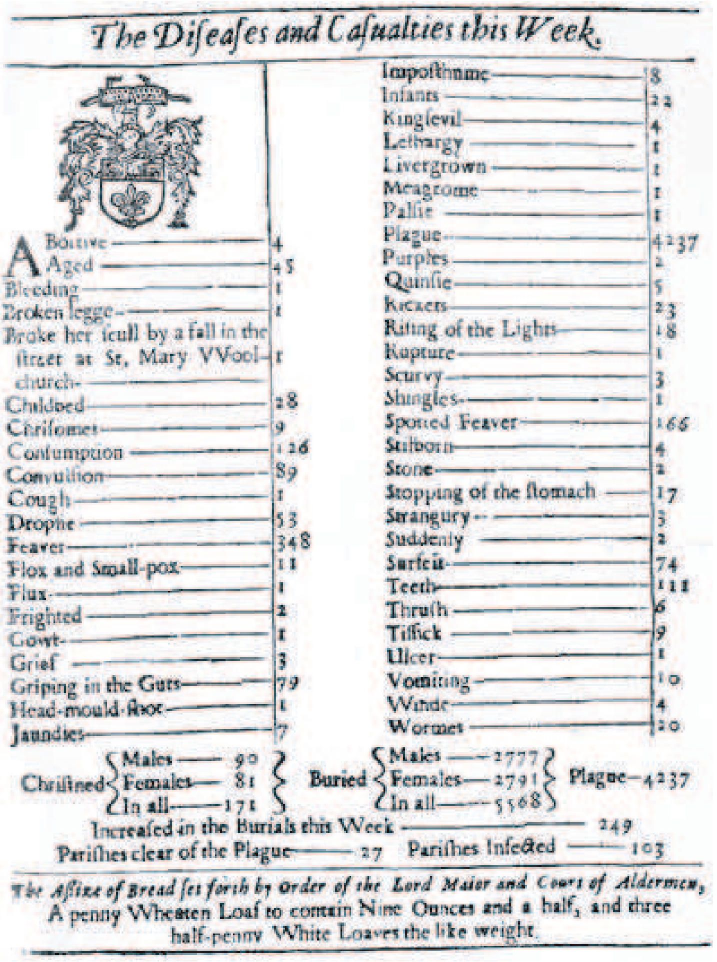
1.1 Early history

In the Western world, the first attempts to quantify and predict the extent of disease outbreaks were restricted to *a posteriori* statistical analyses of the demographics of infections. One of the earliest of these Western statistical demographers was John Graunt (1620-1674), a London merchant who published his analysis of the city's mortality records in 1665. A page from Graunt's *Natural and Political Observations made upon the Bills of Mortality* is depicted in Figure 1.1. The extent to which plague devastated the city of London in that year is evident; of the 5568 recorded deaths, 4237 were attributed to plague. Another milestone in Western disease demography was William Farr's 1840 report as the Registrar-General of England and Wales, in which Farr fitted parameterized curves to outbreaks of smallpox and rinderpest in cattle [1].

An interesting exception to the absence of dynamic models for disease transmission before the twentieth century was Daniel Bernoulli's differential equations model of smallpox dynamics, published in 1766 as *Essai d'une nouvelle analyse de la mortalité causée par la petite vérole et des avantages de l'inoculation pour la prévenir.* Bernoulli (1700-1782) was interested in quantifying the benefits of *variolation*, the inoculation of healthy individuals with small amounts of the smallpox virus in order to confer immunity from the disease. Bernoulli's model assumed that infection of healthy individuals occurred at a constant rate independent of the number of infected individuals in the population, and thus did not utilize an explicit model of disease transmission. For a more complete historical account of Bernoulli's work and its contributions, see [3].

1.2 Compartmental models

Importantly, Bernoulli's model was the first *compartmental model* in mathematical epidemiology, those in which individuals are classified by their *disease state.* A short list of common disease states is given in Table 1.1. For example, most individuals repeatedly contract the common cold, never achieving a state of permanent immunity. This type of illness is modeled as an *SIS* infectious process,



The Diſeaſes and Caſualties this Week.

Abortive	4
Aged	45
Bleeding	1
Broken legge	1
Broke her ſcull by a fall in the ſtreet at St. Mary VVoolchurch	1
Childbed	28
Chriſomes	9
Conſumption	126
Convulſion	89
Cough	1
Dropſie	53
Feaver	348
Flox and Small-pox	11
Flux	1
Frighted	2
Gowt	1
Grief	3
Griping in the Guts	79
Head-mould-ſhot	1
Jaundies	7
Impoſthume	8
Infants	22
Kingſevil	4
Lethargy	1
Livergrown	1
Meagrome	1
Palſie	1
Plague	4237
Purples	2
Quinſie	5
Rickets	23
Riſing of the Lights	18
Rupture	1
Scurvy	3
Shingles	1
Spotted Feaver	166
Stilborn	4
Stone	2
Stopping of the ſtomach	17
Strangury	3
Suddenly	2
Surfeit	74
Teeth	111
Thruſh	6
Tiſſick	9
Ulcer	1
Vomiting	10
Winde	4
Wormes	20

Chriſtned	Males	90	Buried	Males	2777	Plague—4237
	Females	81		Females	2791	
	In all	171		In all	5568	

Increaſed in the Burials this Week — 249

Pariſhes clear of the Plague — 27 Pariſhes Infected — 103

The Aſſize of Bread ſet forth by Order of the Lord Maior and Court of Aldermen,
A penny Wheaten Loaf to contain Nine Ounces and a half, and three
half-penny White Loaves the like weight.

Figure 1.1. Bills of Mortality for London, August 15-22, 1665, taken from John Graunt's *Natural and Political Observations made upon the Bills of Mortality.* For a modern interpretation of the disease names, see [2].

since individuals can transition from being susceptible to infected and then back to susceptible again. For infections which confer immunity, like the chicken pox, an *SIR* model is more appropriate, in which infected individuals transition to the removed state and remain there. The simple SIS and SIR models are the most common in the literature, but any number and combination of states is possible. Modeling the *dynamics* of transitions between compartments, however, was stymied by the absence of a coherent understanding of how infection was acquired and transmitted; it wasn't until the acceptance of the germ theory of disease in the nineteenth century that mathematical epidemiology was able to grow. Scientists like Lous Pasteur illuminated the mechanisms of the underlying biology, which enabled researchers to postulate mathematical models for the dynamics of transitions between compartments.

Table 1.1. Four of the most common compartments in infectious disease modeling.

compartment	description
susceptible	able to contract an infection
latent	contracted the disease but are not yet able to infect anyone else
infectious	able to infect others, symptomatic or not
removed	no longer able to transmit infection to anyone else (possibly via immunity or death)

1.3 Deterministic and stochastic models

Let us begin our discussion of disease models with a simple example: a non-lethal SIS infection moving through a well-mixed population. The dynamics of transitions between compartments is illustrated in Figure 1.2 and is described by the following variables and parameters:

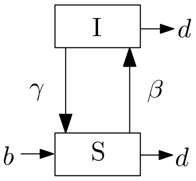


Figure 1.2. An illustration of a susceptible-infected-susceptible (SIS) disease model.

S - the number of susceptible individuals in the population

I - the number of infected individuals

β - the infectious transmission rate of interactions between susceptible and infected individuals

γ - the recovery rate of infected individuals

b - the birth rate (entirely to the susceptible class)

d - the death rate (equal for both susceptible and infected individuals)

N - b/d

How might we translate these ideas into a mathematical model? Most work in this area involves the construction and analysis of deterministic differential equations to describe infection spread. Indeed, the foundations of modern epidemic theory are often traced to the deterministic model of disease propagation formulated by W.O. Kermack and A.G. McKendrick in 1927 [4].[1] This approach

[1] Interestingly, a more general model was proposed and thoroughly explored by Ronald Ross, a British army physician, in a series of papers published 1916-1917 [5] [6] [7].

assumes that the populations under study undergo changes in state *continuously* in time and are large enough that the aggregate behavior of the population behaves *deterministically* (i.e., individual random effects can be ignored). Let us propose the following set of dynamical equations for the state of the total population, $x(t) = (I(t), S(t))$:

$$\begin{cases} \frac{dI}{dt} &= \beta S \frac{I}{S+I} - \gamma I - dI \\ \frac{dS}{dt} &= -\beta S \frac{I}{S+I} + \gamma I + b - dS. \end{cases} \tag{1.1}$$

The most interesting feature of this model is that new infections arise at rate $\beta S \frac{I}{S+I}$; the fraction of each susceptible's interactions that are with an infected individual is given by $\frac{I}{S+I}$, while β measures the rate of new infections per each infected-susceptible contact.

While this differential equation has plausible features, it is lacking in realism. Observe that this model is not constrained to yield integer values of S and I, and thus assumes an infinitesimally-divisible population. Because disease transmission is fundamentally an individual-individual phenomenon, it is much more naturally modeled stochastically, rather than as a deterministic process defined on the aggregate population. Let us construct another continuous-time model that is the stochastic counterpart to System 1.1 by defining a Markov process for $x(t)$. We assume that only one transition is possible in a small unit of time, Δt, and can write the probabilities of state transitions (infection, recovery, birth and death) as follows:

$$\begin{cases} P\{x(t+\Delta t) = (i+1, s-1) | x(t) = (i,s)\} &= \beta s \frac{i}{s+i-1} \Delta t + o(\Delta t) \\ P\{x(t+\Delta t) = (i-1, s+1) | x(t) = (i,s)\} &= \gamma i \Delta t + o(\Delta t) \\ P\{x(t+\Delta t) = (i, s+1) | x(t) = (i,s)\} &= b \Delta t + o(\Delta t) \\ P\{x(t+\Delta t) = (i, s-1) | x(t) = (i,s)\} &= ds \Delta t + o(\Delta t) \\ P\{x(t+\Delta t) = (i-1, s) | x(t) = (i,s)\} &= di \Delta t + o(\Delta t) \\ P\{x(t+\Delta t) = (s+k, i+j) | x(t) = (i,s)\} &= o(\Delta t) \text{ for all other transitions.} \end{cases} \tag{1.2}$$

Taking the limit as $\Delta t \to 0$, we obtain a system of differential equations for $p_{is}(t)$, the probability that $x(t) = (i,s)$:

$$\begin{aligned} \frac{dp_{is}(t)}{dt} &= \beta \left[(s+1) \frac{(i-1)}{s+i-1} p_{i-1,s+1} - s \frac{i}{s+i-1} p_{is} \right] \\ &\quad + b \left[p_{i,s-1} - p_{is} \right] + d \left[(i+1) p_{i+1,s} - i p_{is} + (s+1) p_{i,s+1} - s p_{is} \right]. \end{aligned} \tag{1.3}$$

The infinite system of differential equations represented by Eqn. 1.3 is clearly a great deal more complicated than its deterministic counterpart. Is the extra realism embedded in System 1.2 worth the additional complexity? Since the two models aim to mimic the same phenomena, we might hypothesize a relationship between them. In fact, in the limit of large population size, the *expected value* of the random state vector $x(t)$ in the stochastic model approaches the value of $x(t)$

in the deterministic model. This result is due to Kurtz [8], [9]. In Theorem 1.3.1, we present the restatement of Kurtz' result by Jacquez and Simon [10].

Theorem 1.3.1. Kurtz approximation theorem. *Let $X_N(t)$ be a one-parameter family of continuous-time Markov processes defined on the m-dimensional integer lattice $\mathbb{Z}^m$. Suppose that there is a continuous function $f : \mathbb{R}^m \times \mathbb{Z}^m \to \mathbb{R}$ that satisfies*

$$P\left\{X_N(t+\Delta t) = x+k | X_N(t) = x\right\} = Nf\left(\frac{x}{N}, k\right)\Delta t + o(\Delta t)$$

for all positive integers N and all positions x and increments k in $\mathbb{Z}^m$. Define $F : \mathbb{R}^m \to \mathbb{R}^m$ by

$$F(x) = \sum_k kf(x,k) = E(\Delta X_N | X_N = x),$$

the expected change in X_N from x. Suppose that there exists an open set S in $\mathbb{R}^m$ and a constant M such that

1. $|F(x) - F(y)| \leq M|x-y|$ *for all* $x, y \in S$*;*

2. $\sup_{x\in S} \sum_k |k| f(x,k) < \infty$*;*

3. $\lim_{d\to\infty} \sup_{x\in S} \sum_{|k|>d} |k| f(x,k) = 0$.

Let $Z(t; x_0)$ be the solution of the (deterministic) initial value problem

$$\frac{dZ}{dt} = F(Z), \qquad Z(0) = x_0.$$

Suppose that $Z(t;x_0) \in S$ for all $t \leq T$, and that $\lim_{N\to\infty}[X_N(0)/N] = x_0$ for the original family of Markov processes. Then, for every $\epsilon > 0$,

$$\lim_{N\to\infty} P\left\{\sup_{t\leq T}\left|\frac{1}{N}X_N(t) - Z(t;x_0)\right| > \epsilon\right\} = 0.$$

Do the models of Systems 1.1 and 1.2 meet the criteria of Theorem 1.3.1? Taking the population size to infinity in System 1.2 is equivalent to taking the birth rate b to infinity, so b will serve as the parameter for our family of models. We can satisfy the conditions on $f(\cdot,\cdot)$ by defining

$$\begin{aligned} f\left((i,s),(1,-1)\right) &= \beta s\frac{i}{s+i-1} \\ f\left((i,s),(-1,1)\right) &= \gamma i \\ f\left((i,s),(0,1)\right) &= 1 \\ f\left((i,s),(0,-1)\right) &= ds \\ f\left((i,s),(-1,0)\right) &= di. \end{aligned}$$

The existence of the open set S is guaranteed by the boundedness of the right-hand side of Systems 1.1 and 1.2 over any open interval in the positive quadrant. Thus, in the limit of large population size, the expected value of the state vector $x = (I, S)$ will be well-approximated by

$$\frac{dx}{dt} = \begin{bmatrix} \frac{dI}{dt} \\ \frac{dS}{dt} \end{bmatrix} = F(x) = \begin{bmatrix} \beta S \frac{I}{S+I} - \gamma I - dI \\ -\beta S \frac{I}{S+I} + \gamma I + b - dS \end{bmatrix},$$

precisely the deterministic model of System 1.1.

Practically, for what values of N is this approximation valid? The answer depends upon the behavior and time-scales of interest, but models similar to System 1.2 have agreed satisfactorily with the deterministic predictions for fewer than 100 individuals in a population; for some examples of such comparisons in the literature, see [11], [12] and [13].

Although models based on differential equations have been the principal methodology in mathematical epidemiology, deterministic *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. Kurtz' theorem compares the behavior of deterministic and stochastic models in continuous-time; do such results hold for discrete-time models as well? The answer is yes, but the proof is omitted: see [14] and [15] for a discussion, and [16] for simulation results. In Chapter 2, we will present a general framework for infection dynamics which can be implemented in continuous-time or discrete-time, and throughout this thesis, we will use examples of both kinds of models. For more examples of deterministic and stochastic modeling approaches, Appendix A contains a quick survey of some of the recent literature.

1.4 Model predictions and thresholds

The objective of mathematical epidemiology is to serve public health interests by modeling the essential characteristics of disease transmission. When preparing for a potential epidemic, public health officials face a number of questions:

- ▷ What kinds of policies might inhibit disease transmission?
- ▷ Can a vaccination campaign prevent an epidemic? What kind of vaccination strategy should be employed?
- ▷ Is quarantining necessary, or more more mild social regulations be just as effective?
- ▷ Are there preventative measures that will make an epidemic unlikely?

Thus, when formulating a mathematical model for disease transmission, one often has two sets of issues in mind: the testable hypotheses generated by the model, and the opportunities for active control of the model dynamics. Some common questions to ask and answer are:

- ▷ Will the disease become an epidemic?

- ▷ What percentage of the population will be affected? What percentage will die? What types of individuals are most at risk?

- ▷ How long will the disease persist?

- ▷ What parameters of disease transmission have the greatest impact on the epidemic outcome?

- ▷ Can we estimate the variability in the predictions of the model?

Of all of these predictive questions, "will the disease become an epidemic?" has been the focus of most of the work in mathematical epidemiology to date. Naturally, the answer to this question depends upon what one means by 'epidemic', which is often loosely defined as "any upward fluctuation in disease incidence or prevalence" [17]. This is a context-specific definition, necessarily designed to serve the public health officer and not the mathematician. 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 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 $X > c$. For example, inspection of System 1.1 reveals that there exists an endemic equilibrium

$$(I^e, S^e) = \left(\left(1 - \frac{\gamma + d}{\beta}\right) N, \frac{\gamma + d}{\beta} N\right).$$

if and only if $\frac{\beta}{\gamma+d} > 1$. This equilibrium is also a global attractor in this parameter regime, so all initial conditions will eventually reach this value. One could say, then, that the value of $\frac{\beta}{\gamma+d}$ serves as a *threshold* for this model: whether it is greater than or less than one determines whether or not the disease will establish an endemic presence.

In contrast to the deterministic model, the stochastic model of System 1.2 has a single absorbing state, $(I, S) = (0, N)$, independent of the values of the parameters. This observation is one of the key differences between stochastic and deterministic disease models. This is not to say, however, that the parameters of System 1.2 have no bearing on the stochastic dynamics. In particular, Jacquez and Simon find that if $\frac{\beta}{\gamma+d} > \frac{N}{N-1}$, the mean number of infected individuals will *always* increase from its initial value. If $\frac{\beta}{\gamma+d} < \frac{N}{N-1}$, the mean will decrease monotonically to zero. This observation suggests that the value of $\frac{\beta}{\gamma+d}$ serves as a threshold for the stochastic model as well. We will return to a discussion of thresholds in Chapter 2.

1.5 Mathematical models and network science

Tractable models like Systems 1.1 and 1.2 provide a tremendously valuable foundation, but typically rely on the assumption that each individual is equally likely to come into contact with every other individual in the population. Tools developed in the 1960s and 1970s, like computer simulations and percolation theory, however, opened up the range of testable hypotheses and, for the first time, allowed the modeling of an essential feature of human epidemics: the inherent locality of person-to-person disease transmission. These tools have enabled researchers to make more sophisticated predictions about key questions: which population density patterns encourage the spread of disease, and which inhibit epidemic formation? How quickly will a disease progress through a population? Will certain spatial patterns of vaccination halt or slow disease spread?

One fascinating area of research combines traditional epidemiologic models with the mathematics of network theory and dynamic systems to study the human-environment relationships that enable the emergence and spread of infectious disease. Gretchen Daily and Paul Ehrlich of the Center for Conservation Biology at Stanford have referred to these relationships as the 'epidemiological environment', a term that includes the biology of pathogenic parasites, the physical environment of parasite development and the human social patterns through which disease propagates [18].

Regarding the global re-emergence of malaria, Pim Martens and Lisbeth Hall of Maastricht University have noted that "as people move, they can increase their risk for acquiring the disease through the ways in which they change the environment and through the technology they introduce" [19]. The difficulty of modeling these kinds of interactions has led to public health policies that are often, in retrospect, short-sighted. The emergence of Lyme disease in the 1970s, for example, can be traced to the reforestation of the eastern U.S. after farmers relocated to the Midwest [20]. One 2005 assessment of the progress of the UN Millennium Project task force on environmental sustainability addressed the public health lessons of Lyme disease, among others, and noted with regret that "responses to the disease are still focused on individual treatment rather than better land use and wildlife management policies that might stem the spread of Lyme and possibly other new pathogens" [21].

Another element in the modeling of human epidemics concerns interactions between individuals that enable the spread and persistence of infection. Given the variety of ways in which modern individuals interact, the interactions which are critical to disease transmission can occur on vastly different scales. In one large-scale example, a recent article in *The Lancet* discussed the potential of the hajj, the yearly pilgrimage of over 2.5 million Muslims to Mecca in the 12th month of the Islamic calendar, as a potential epidemiological ground zero for many communicable diseases [22]. On a smaller scale, social epidemiologists focus on quantifying the impact of social networks on individual and local population health. Human-human contact is not simply a vehicle for disease transmission; one study found that the more diverse an individual's social network (defined by the number of different types of social interactions), the greater his or her resistance to the common cold [23]. The ability of social networks to both inhibit and promote the spread of infectious disease is just one of many interesting modeling challenges.

Not surprisingly, the interest in network science that arose in the 1990s has found exciting applications in mathematical epidemiology. Over the past ten years, interdisciplinary collaborations have revealed that the network structures seen by biologists, physicists, sociologists and ecologists often have the same interesting and peculiar features, as do the dynamic processes that take place within these networks. The parallels between the transmission of SARS, the propagation of a computer virus, and the techniques of viral marketing have generated an enormous amount of interest in the fundamental theory. Additionally, infectious diseases propagate over networks of many scales: from continent to continent via the air transportation network, from neighborhood to neighborhood via subway lines and bus routes, and from person to person via social contacts. The volume of work in the last few years on the dynamics of infection processes on networks is vast, but there is consensus on a single principle: the topology of a network can have critical consequences for the spread of infection.

1.6 Graph and matrix theory preliminaries

If we are to include these interaction patterns in our mathematical models, it is most natural to represent such networks as a *graph.* A graph $\mathcal{A} = G(\mathcal{N}_A, \mathcal{E}_A)$ consists of a set of *nodes*, $\mathcal{N}_A$, and a set of *edges*, $\mathcal{E}_A \subseteq \mathcal{N} \times \mathcal{N}$. A graph may be *directed* or *undirected*; if $\mathcal{A}$ is undirected, then $(u, v) \in \mathcal{E}_A$ implies that $(v, u) \in \mathcal{E}_A$ for $u, v \in \mathcal{N}_A$. In Chapter 5, we will also use $\mathcal{A}_{uv}$ to denote $(u, v) \in \mathcal{E}_A$. Two pictorial examples of graphs are given in Figures 1.3 and 1.4.

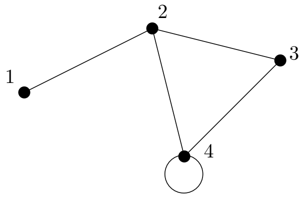


Figure 1.3. An undirected graph, $\mathcal{A}_u$.

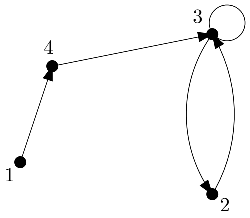


Figure 1.4. A directed graph, $\mathcal{A}_d$.

Every node i in a graph has an *out-degree* and an *in-degree*, the former being the number of edges that begin at i and the latter the number of edges that terminate at i. Note that in an undirected graph, the out-degree and in-degree of a node are equal. A *simple* graph is one which has no self-loops (an edge from a node to itself) or multiple edges. An undirected, *connected* graph is one which has a path between any two nodes i and j. A *tree* is a connected graph without any *cycles*, i.e.

paths from a vertex back to itself in which no edge is repeated. For an undirected graph, we'll also define the notion of a *clique* S, which is a set of vertices such that every vertex in S is connected to every other vertex in S. One measure of a graph's global structure is the *degree distribution*, a histogram of the degrees of each of its nodes; for a directed graph, there exist distributions of both in- and out-degrees.

We can also define *subgraphs* of a given graph by specifying subsets of $\mathcal{N}_A$ and $\mathcal{E}_A$. Two types of subgraphs will be useful to us:

- *node-induced subgraphs* - for a subset $\mathcal{N}_S \subseteq \mathcal{N}_A$, an edge in $\mathcal{E}_A$ is contained in $\mathcal{E}_S$ if and only if that edge connects two nodes in $\mathcal{N}_S$;

- *edge-induced subgraphs* - for a subset $\mathcal{E}_S \subseteq \mathcal{E}_A$, a node in $\mathcal{N}_A$ is contained in $\mathcal{N}_S$ if and only if that node appears at the end of one of the edges in $\mathcal{E}_S$.

In a graph representing a social network, connections between individuals are rarely best described as '1's and '0's; it is useful to be able to distinguish between a strong friendship and a weak acquaintanceship. This naturally leads to a notion of *weighted edges*, in which each edge i of a graph has a weight w_i (often restricted to the interval $[0, 1]$).

It is also possible to associates weights or attributes to the *nodes* in a graph: one could imagine a scenario in which the weight of a node indicated that node's relative importance to the graph, perhaps in modeling the hierarchy in a corporation. A higher-level approach to weighting nodes and edges in graphs is to label each node and/or edge with a vector of *attributes.* Attributes could be drawn from any class of descriptors; these could simply be weights as discussed previously, or they could be text strings or even functions.

A convenient way to represent a graph is a *node-node adjacency matrix* (also referred to simply as an *adjacency matrix*). If the cardinality of $\mathcal{N}_A$ is n_A, then the adjacency matrix A of this graph is an $n_A \times n_A$ matrix in which entry $[A]_{ij}$ is equal to 1 if and only if $(i, j) \in \mathcal{E}_A$, and is equal to 0 otherwise. Observe that the adjacency matrix of an undirected graph will always be symmetric. The adjacency matrices of the graphs in Figures 1.3 and 1.4 are given below.

Table 1.2. Adjacency matrix of $\mathcal{A}_u$.

$$A_u = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 \end{bmatrix}$$

Table 1.3. Adjacency matrix of $\mathcal{A}_d$.

$$A_d = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

We can extend the node-node adjacency matrix to graphs with weighted edges by allowing the ijth entry of A to take the value of the weight of edge (i, j). An weighted, nonnegative adjacency matrix is also a useful way of representing the volume of flow between nodes, e.g. the number of passengers traveling between two airports. Observe that the out- and in-degrees of each node can be found by simply summing over the rows and columns of the node-node adjacency matrix, respectively.

The identification of a graph with an adjacency matrix allows us to apply the tools of linear algebra to the study of graph properties. In particular, we will be interested in the ways that network structures influence the set of eigenvalues and eigenvectors of the adjacency matrix, a field known as *spectral graph theory.* Because an adjacency matrix is nonnegative, we will find the following fundamental results of Perron and Frobenius useful, where $\rho(A)$ is the *spectral radius* of A, the magnitude of the eigenvalue of A with largest magnitude [24].

- ▷ If $\lambda(A)$ is an eigenvalue of a real matrix A, then $\lambda(A)$ is also an eigenvalue of $A^\top$.
- ▷ If A is an $n \times n$ matrix with nonnegative entries, then $\rho(A)$ is an eigenvalue of A and there is a nonnegative vector $x \geq 0$, $x \neq 0$, such that $Ax = \rho(A)x$.
- ▷ An $n \times n$ nonnegative matrix A is irreducible if and only if $(I + A)^{n-1}$ has all positive entries.
- ▷ If A is $n \times n$, irreducible, nonnegative matrix, then $\rho(A)$ is positive and is an algebraically simple eigenvalue of A.

In Chapter 3, we will make extensive use of the *Kronecker product* of two matrices C and D, which we'll denote by $C \otimes D$. If $C = \{c_{ij}\}$ is an $c_1 \times c_2$ matrix, and $D = \{d_{ij}\}$ is an $d_1 \times d_2$ matrix, then $C \otimes D$ is the $c_1 d_1 \times c_2 d_2$ matrix defined by

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

The Kronecker product $C \otimes D$ simply repeats the matrix D at each element of C. One useful property of the Kronecker product is that for matrices A, B, C, and D of compatible dimensions, $(A \otimes B)(C \otimes D) = AC \otimes BD$. Additionally, $\rho(C \otimes D) = \rho(C)\rho(D)$.

Chapter 2

Epidemic thresholds and a general deterministic framework

CHAPER 1 introduced the idea of compartmental disease models; Appendix A contains a sample of the enormous volume of literature on different kinds of compartmental models, both deterministic and stochastic, in discrete time and continuous time, and for populations of varying heterogeneity. We also introduced the idea of an *epidemic*: a particular set of criteria by which an infection outbreak is assessed to be especially severe and noteworthy.

In stochastic models, one is often interested in the time scales over which the disease is likely to be present; Sections 1.4, 4.1 and A touch on different notions of "epidemic" in stochastic models. 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) [25].

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, as discussed in Section 1.3. This chapter (and Chapter 3) will explicitly focus on deterministic models, but the results of Chapters 4 and 5 are useful in both deterministic and stochastic settings.

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;
- whether the disease will establish a sustained presence in a community.

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" [26] 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 model exhibits a stable and/or attracting *endemic equilibrium*, one in which there is a positive level of infection.

Table 2.1 presents a summary of the appearance of these different notions of epidemic in deterministic models in the recent literature.

Table 2.1. Approaches taken to computing an epidemic threshold in a sample of the literature.

Approach	Reference
next-generation operator	Diekmann & Heesterbeek, 1990 [26], Becker & Dietz, 1995 [27], Fulford et al., 2002 [28]; Fraser et al., 2004 [29]
local stability of disease-free equlibrium	Boguna & Pastor-Satorras, 2002 [30]; Hill & Longini, 2003 [31]; Wang et al., 2003 [32]; Hyman & Li, 2000 [33]; Alexander & Moghadas, 2005 [34]; Kiss et al., 2006 [35]; Keeling, 1999 [36]; Hyman & Li, 2006 [37]
existence of an endemic equilibrium	Anderson & May, 1991 [38]; Pastor-Satorras & Vespignani, 2001 [39]; Masuda & Konno, 2006 [40]
multiple criteria	Blyuss & Kyrychko, 2005 [41]; Hyman & Li, 2005 [42]; Salmani & van den Driessche, 2006 [43]; Arino & van den Driessche 2003 [44]

The existence of multiple threshold criteria can create confusion when different criteria are erroneously assumed to be equivalent, e.g., using the existence of an endemic equilibrium to conclude that $R_0 > 1$. This issue has been raised by several authors, among them Heffernan et al. [45] and van den Driessche and Watmough [46]. To give a sense of where this confusion can arise, consider the simple deterministic SIS infection model presented in Chapter 1:

$$\begin{aligned}\frac{dS}{dt} &= -\beta S\frac{I}{S+I} + \gamma I + b - dS \\ \frac{dI}{dt} &= \beta S\frac{I}{S+I} - \gamma I - dI.\end{aligned}$$

In Chapter 1, we observed that an endemic equilibrium exists if and only if $\beta/(\gamma + d) > 1$. The model has a second, *disease-free*, equilibrium,

$$(I^{df}, S^{df}) = (0, N),$$

which always exists but which is locally asymptotically stable if and only if $\beta/(\gamma + d) < 1$ (Section 2.2 gives an extended discussion of this stability condition). Thus, the existence of an endemic equilibrium and the stability of the disease-free equilibrium coincide as threshold tests in the parameter space of the model. Furthermore, a single infective individual in an otherwise susceptible population can infect β individuals/time unit and remains infectious for $1/(\gamma+d)$ time units, so the value of R_0 associated with this model is $\beta/(\gamma + d)$ (a detailed discussion of the calculation of R_0 is deferred to Section 2.3). For this simple model, we see that all three threshold criteria are identical, occurring as $\beta/(\gamma + d)$ changes relative to 1. However, this is rarely the case in more detailed disease models, which often exhibit multiple equilibria with complex stability requirements. Here, we assemble many results and case studies from the literature to address the following question: what do these threshold criteria mean, and when do they yield the same predictions for disease behavior?

2.1 A general compartmental framework

An excellent general framework for infection modeling in structured populations was put forth by van den Driessche and Watmough in [46]. Here, we will extend their continuous-time results to discrete-time models, and will adopt their notation throughout this thesis. 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.

Define a population (or state) vector $x = (x_1, \ldots, x_n)$ 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). Any heterogeneity in the population or in the disease stages that one would like to model should be mapped to a different compartment. 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 of Section 2.5. For the discrete-time model, the rates are measured as *changes per time step*. We can formulate differential and difference equation models, respectively, of this process:

$$\dot{x}_i = h_i(x) = \mathcal{F}_i(x) + \mathcal{V}_i^+(x) - \mathcal{V}_i^-(x) = \mathcal{F}_i(x) - \mathcal{V}_i(x) \tag{2.1}$$

$$x_i \leftarrow h_i(x) = x_i + \mathcal{F}_i(x) + \mathcal{V}_i^+(x) - \mathcal{V}_i^-(x) = x_i + \mathcal{F}_i(x) - \mathcal{V}_i(x) \tag{2.2}$$

The left arrow ← in Eq. 2.2 denotes a time-step update. Although the use of some of the same symbols for the continuous- and discrete-time formulations might seem confusing, it is useful for highlighting the relationships between the two models, and the appropriate meaning should be clear from the context. 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 [46]:

(A1) If $x \geq 0$, then $\mathcal{F}_i, \mathcal{V}_i^+, \mathcal{V}_i^- \geq 0$ for $i = 1, \ldots, n$; all flows between compartments are nonnegative.

(A2) For continuous-time models, if $x_i = 0$, then $\mathcal{V}_i^-(x) = 0$, while for discrete-time models, $\mathcal{V}_i^-(x) \leq x_i$; 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 $i = 1, \ldots, m$, then $\mathcal{F}_i(x) = 0$ and $\mathcal{V}_i^+(x) = 0$ for $i = 1, \ldots, m$; 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 $\overline{x}$ will be called a *disease-free equilibrium (DFE)* if

▷ the first m components of $\overline{x}$ are zero (corresponding to the absence of infected individuals);

▷ $\overline{x}$ is an equilibrium of Eq. 2.2, i.e., $\overline{x} = h(\overline{x})$;

▷ all of the eigenvalues of the Jacobian matrix of the function $-\mathcal{V}$ at the equilibrium $\overline{x}$, denoted by $J = -D\mathcal{V}(\overline{x})$, have modulus less than one (in discrete-time), or have real part less than zero (in continuous-time), 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.

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. We will see, however, that strong conclusions can still be drawn within this very general framework.

To illustrate the utility of this approach, we can write the deterministic SIS model of Chapter 1 in this notation. Here, $x = (x_1, x_2) = (I, S)$, so $m = 1$. Then

$$\begin{aligned} \tfrac{dx_1}{dt} &= \underbrace{\beta x_2 \frac{x_1}{x_1 + x_2}}_{\mathcal{F}_1} - \underbrace{(\gamma x_1 + d x_1)}_{\mathcal{V}_1^-} \\ \tfrac{dx_2}{dt} &= \underbrace{b + \gamma x_1}_{\mathcal{V}_2^+} - \underbrace{\beta x_2 \frac{x_1}{x_1 + x_2} + d x_2}_{\mathcal{V}_2^-} . \end{aligned} \tag{2.3}$$

2.2 Local asymptotic stability of the DFE

In dynamic systems theory, the condition of local asymptotic stability dictates whether a small perturbation away from an equilibrium will grow or if the system will return to the equilibrium point. This has a natural implication for the dynamics of an emerging infection; here, a “perturbation” amounts to introducing a small number of infective individuals into a disease-free population. Mathematically, the criterion for local asymptotic stability of an equilibrium is a condition on the eigenvalues of the Jacobian matrix of the system, evaluated at the equilibrium. In continuous-time systems, if the real part of each of the eigenvalues is negative, then the equilibrium is stable. In discrete-time systems, all eigenvalues must have modulus less than one for local stability. How does this criterion translate to conditions on the general model of Eqs. 2.1 and 2.2?

We will consider the discrete-time case in detail; the continous-time results follow analogously. If $\overline{x}$ is a DFE, then the Jacobian matrix of the discrete-time system around the DFE takes the following form:

$$I + D\mathcal{F}(\overline{x}) - D\mathcal{V}(\overline{x}) = I + \begin{bmatrix} F & 0 \\ 0 & 0 \end{bmatrix} - \begin{bmatrix} V & 0 \\ J_3 & J_4 \end{bmatrix} \tag{2.4}$$

The partitions of $D\mathcal{F}(\overline{x})$ and $D\mathcal{V}(\overline{x})$ are a consequence of assumptions (A2)-(A4). Additionally, the matrix F is nonnegative; this follows from assumptions (A1) and (A4), and the argument can be found in the proof of Lemma 1 in [46]. Local asymptotic stability of the DFE requires that all eigenvalues of the linearization given in Eq. 2.4 fall within the unit circle. Given the partitioning of the linearization, the eigenvalues of the linearized system are the union of the eigenvalues of the matrices $I + F - V$ and $I - J_4$. The set of eigenvalues of $D\mathcal{V}(\overline{x})$ is the union of the set of eigenvalues of V and J_4; by the definition of a DFE, these are all assumed to be within the unit circle. Thus, the eigenvalues of $I - J_4$ are contained within the unit circle, so the condition for local asymptotic stability of the DFE is $\rho(I + F - V) < 1$.

In continuous time, the requirement for the stability of a DFE is that $\rho(F - V) < 0$, where F and V are defined identically as above. In the deterministic SIS model of Eq. 2.3, the DFE is given by $\overline{x} = (0, N)$; we can readily compute that F and V are the following $m \times m = 1 \times 1$ matrices

$$F = \beta, \quad V = \gamma + d,$$

so the condition for the local asymptotic stability of the DFE is

$$\beta - (\gamma + d) < 0 \quad \Leftrightarrow \quad \frac{\beta}{\gamma + d} < 1.$$

Section 2.5 will present an example of this threshold test for a discrete-time model.

2.3 R_0: its calculation and interpretation

Historically, the parameter that has received the most attention as the determining function for an epidemic threshold is R_0, the basic reproductive ratio, which 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" [26]. The history of the adoption of the parameter R_0 over the course of the 20th century is a complex and interesting story that weaves together developments in epidemiology and population ecology; the term as it is now understood was introduced by George MacDonald in 1952, rediscovered and used by Klaus Dietz in the 1970s, then canonized by Anderson and May in the early 1980s [47]. May et al. provide a heuristic description of the elements of R_0 and its relevance to epidemiology in [48]. Intuitively, if R_0 is greater than 1, then it is likely that the number of infected individuals in a population will increase after the introduction of an initial infective, and unlikely otherwise. Estimates for the R_0 value of some common diseases are given in Table 2.2.

Table 2.2. Ranges of R_0 for some well-known diseases, assuming homogeneous mixing with standard incidence [49].

AIDS	$2 - 5$
smallpox	$3 - 5$
measles	$16 - 18$
malaria	> 100

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. [26]. The next-generation operator is defined by the structure of the population (i.e., its relevant types or distinct subpopulations), the steady-state distribution of individuals in the disease-free equilibrium, and the number of infected individuals of each type produced by an infected individual of each type. The operator takes in a density that represents the likelihood of the initially infectious individual being of each type, and outputs the expected number of direct secondary infections caused by this individual over the course of the individual's lifetime within each of the different types. 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 ijth 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 [26] 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.

How does the spectral radius arise in this context? Let us begin with an initial distribution of individuals in infected compartments $1, \ldots, m$ defined by a vector ψ (with sum of entries denoted $\|\psi\|$). Note that R_0 is computed under the assumption that infected individuals operate in a completely susceptible population, i.e. there exists a never-ending supply of fresh susceptibles to take the place of those infected by the initial class. If the population of susceptible individuals is not depleted between generations, the next generation will produce $\|K\psi\|$ new infections, the second generation $\|K^2\psi\|$, and so on. Define $\|K\|$ as the maximum value of $\|K\psi\|$ for all ψ with $\|\psi\| = 1$; this is a definition of a *matrix norm.* The per generation growth rate, then, is $\|K^n\|^{1/n}$, a geometric mean. If we take the limit as $n \to \infty$, $\|K^n\|^{1/n} = \rho(K)$, the largest eigenvalue of K.[1]

How does this mathematical definition of R_0 correlate with the "word" definition given previously? Let us explore this question through a series of examples.

First, consider a host-vector disease. The dominant mode of transmission of malaria, for example, is back and forth between human and mosquito; in order for a human infection to cause another human infection, the disease must first pass through a mosquito. For this disease, the two sub-populations are human and mosquito (with only one infected compartment each, so the next generation matrix K will be 2×2), and it will only have off-diagonal entries since there can be no direct infections of a human by a human or a mosquito by a mosquito. If we use R_{HM} to denote the average number of secondary infections in the mosquito population caused by a single infective human in a completely susceptible mosquito population over the course of the human's lifetime, and R_{MH} to denote the analogous quantity caused by a single infective mosquito in the human population, then K is given by

$$K = \begin{bmatrix} 0 & R_{HM} \\ R_{MH} & 0 \end{bmatrix}.$$

[1] This result is called *Gelfand's formula*, and is true for any matrix norm.

What is the value of R_0 for this disease? Without referring to next-generation matrix arguments, a reasonable answer to this question is $\sqrt{R_{HM}R_{MH}}$. Certainly, our answer should depend both on R_{HM} and R_{MH}: the disease can't spread if both of those quantities is less than one. But it could spread, for example, if $R_{HM} > 1$ and $R_{MH} < 1$, depending on whether $R_{HM}R_{MH} > 1$. This is the key condition, but $R_{HM}R_{MH}$ itself is a "two-step" measure, and does not have the right units to be the number of secondary infections. If one would like a measure of the per generation growth factor, the geometric mean $\sqrt{R_{HM}R_{MH}}$ provides the right result (which is > 1 if and only if $R_{HM}R_{MH} > 1$). The largest eigenvalue of K is indeed $\sqrt{R_{HM}R_{MH}}$.

Now, consider an example of the type formulated by Larson in [50], in which an infected individual is equally likely to cause either 2 or 6 secondary infections. How do we interpret the phrase "equally likely" in the context of the deterministic models that we've been discussing? A naive approach might be to eliminate the randomness by assuming that all infected individuals cause the mean number of infections, 4; then $R_0 = 4$. Instead, what if you assume that half of the population infects 6 others and the other half infects 2? Call the former group N_6, and the latter group N_2. If an individual of N_6 has no preference for interacting with individuals of N_6 or N_2, then the number of new infections caused by an individual in N_i within the population N_j is given by $ij/(i+j)$ (for more discussion of this calculation, see Section 3.3.1); thus, the next-generation matrix K is

$$K = \begin{bmatrix} 36/8 & 12/8 \\ 12/8 & 4/8 \end{bmatrix},$$

where $\rho(K) = 5$. This is larger than the "homogeneous mixing assumption" - see Chapter 3 for more on this phenomenon. Taking this one step further, what if individuals in N_i only interacted with other individuals in N_i? Then

$$K = \begin{bmatrix} 6 & 0 \\ 0 & 4 \end{bmatrix}$$

and $\rho(K) = 6$.

What is happening in these examples? First, we see that the use of the phrases "average number of secondary infections" and "typical infectious individual" can be misleading; in the first and second examples, different mathematical interpretations of these phrases lead to different numerical outcomes for R_0. Additionally, from our last example, we see that even though half the population only infects 4 other individuals, the value of R_0 given by the mathematical definition is 6! In light of these observations, consider an alternative 'word' definition of the basic reproductive ratio: R_0 is the *asymptotic per generation growth factor* of the infection. Here, a *generation* refers to the time elapsed between the initial infection of an individual and that individual's removal from the infected class. If there are multiple types of individuals who may be simultaneously infected, generations may begin asynchronously, but can be interpreted as "waves" of infection. To determine whether a disease will grow or die, we want a measure of the maximum number of infections that could be produced in each generation; the infection will die out if this number if less than one.

For our general compartmental model, how do we compute the elements of K? An element K_{ij} in the next-generation matrix should represent the total number of secondary (direct) infections in compartment j caused by an infective introduced in compartment i over its entire infectious period. Note that this definition does not include a counting of the infections that occur at the tertiary stage and beyond. To find K_{ij}, we allow each initially infected individual to cause new infections according to $\mathcal{F}$ for every time step in an infected compartment, but do not allow the newly infected individuals to influence the dynamics. If only a small number of infectives are introduced into a stable disease-free population $\overline{x}$, then the inter-compartmental movement is well-described by the linearized system

$$x \leftarrow (I - D\mathcal{V}(\overline{x}))(x - \overline{x}). \tag{2.5}$$

Since we are interested in the dynamics when $\overline{x}$ is perturbed by the introduction of a few individuals into the infected compartments $i = 1, \dots, m$, we only need to follow the dynamics of the first m elements of the vector x. Denote these first m elements at time n by the vector $\psi(n)$; we require that $\psi(n)$ satisfy Eq. 2.5, and thus by the partitioning of Eq. 2.4

$$\psi(n+1) = (I - V)\psi(n). \tag{2.6}$$

The unique solution to Eq. 2.6 is given by $\psi(n) = (I - V)^n\psi(0)$, which counts the number of individuals in compartment i at time n for $i = 1, \dots m$, given an initial distribution. Every individual in this compartment is capable of transmitting infection, and in a mostly susceptible population with a small number of infectives, the number of infections caused by $\psi(n)$ is well-approximated by $F\psi(n)$. Observe that this approximation assumes that the susceptible population is not depleted as new infections occur, a key assumption in the computation of R_0. The total number of secondary infections in each compartment caused by the initial infective population is then given by

$$\sum_{i=0}^{\infty} F\psi(i) = \sum_{i=0}^{\infty} F(I - V)^i\psi(0) = FV^{-1}\psi(0).$$

We have already observed that the definition of a DFE requires that all eigenvalues of $I - D\mathcal{V}(\overline{x})$ have modulus less than one, which is equivalent to all eigenvalues of $D\mathcal{V}(\overline{x})$ being contained within the unit circle centered at 1. Thus, all eigenvalues of $D\mathcal{V}(\overline{x})$ have positive real parts, and V is invertible. The next-generation matrix, then, is $K = FV^{-1}$, a product of nonnegative matrices, and R_0 is defined to be its spectral radius, $\rho(FV^{-1})$. Since FV^{-1} is nonnegative, R_0 is an eigenvalue of K.

As presented in [46], an analogous argument holds for the determination of K in continuous-time, with an identical result: $K = FV^{-1}$. For the deterministic continuous-time SIS model represented by Eq. 2.3, we computed F and V in Section 2.2: K is a 1×1 matrix, which means that

$$R_0 = \rho(K) = K = \frac{\beta}{\gamma + d}.$$

2.4 Equivalence of threshold on R_0 and DFE stability

Thus far, this chapter has established two results on the behavior of the general compartmental model in discrete-time:

- ▷ The DFE is locally asymptotically stable if and only if $\rho(I + F - V) < 1$.
- ▷ R_0 is given by $\rho(FV^{-1})$.

What kind of relationship should we expect between the criteria for stability of the DFE and $R_0 < 1$? 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 themnumber 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. This idea is depicted in Figure 2.1. 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. The following theorem (for the discrete-time case, built on the continuous-time result in [46]) establishes this relationship.

Theorem 2.4.1. *Let $\overline{x}$ be a DFE and define the $m \times m$ matrices $F = \{f_{ij}\}$ and $V = \{v_{ij}\}$ as:*

$$f_{ij} = \frac{d\mathcal{F}_i}{dx_j}\bigg\|_{\overline{x}}, \quad v_{ij} = \frac{d\mathcal{V}_i}{dx_j}\bigg\|_{\overline{x}} \quad \text{for infected compartments } i, j = 1, \ldots, m$$

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

Proof. Note that the initial perturbations for which local stability is tested in Theorem 2.4.1 are no longer constrained to lie within the disease-free manifold. This proof follows that presented in [46].

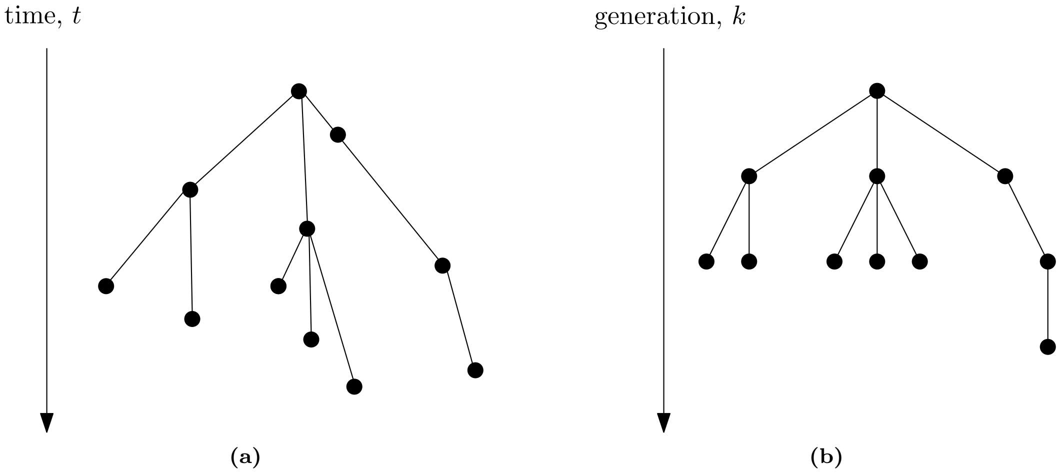


Figure 2.1. Schematic of new infections v. time (a) and generation (b).

First, we observe that V is an *M-matrix*; it is non-singular and has nonpositive off-diagonal entries (see the proof of Lemma 1 in [46]).

The matrix $I + F - V$ is nonnegative; to see this, expand V into its two components $V^- - V^+$ and observe that assumption (A2) implies that the diagonal entries of V^- are in the interval $[0, 1]$. Since V is an M-matrix, its off-diagonal entries are nonpositive. Combining these observations with the nonnegativity of F demonstrates the nonnegativity of $I + F - V$.

Now, we make an intermediate observation: if A is a nonnegative matrix, then $I - A$ is an M-matrix if and only if $\rho(A) < 1$. To see this, note that if $I - A$ is an M-matrix, then all of its eigenvalues are in the right-half plane, which in turn implies that all of the eigenvalues of A have real part less than one. Since A is nonnegative, $\rho(A)$ must be one of the eigenvalues of A, and thus $\rho(A) < 1$. Conversely, assume that $\rho(A) < 1$. Then all of the eigenvalues of $I - A$ are in the right-half plane. Since A is nonnegative, $I - A$ has nonpositive off-diagonal entries. Thus, $I - A$ is an M-matrix.

We complete the proof by demonstrating that $\rho(I + F - V) < 1$ if and only if $\rho(FV^{-1}) < 1$. Since FV^{-1} is nonnegative, $\rho(FV^{-1}) < 1$ if and only if $I - FV^{-1}$ is an M-matrix. Since V is an M-matrix, by Lemma 5 of [46], $I - FV^{-1}$ is an M-matrix if and only if $V - F$ is an M-matrix. We've observed that $I + F - V$ is nonnegative; by our intermediate result, $I - (I + F - V) = V - F$ is an M-matrix if and only if $\rho(I + F - V) < 1$. □

For the continous-time result, see [46], Lemma 1 and Theorem 2.

Theorem 2.4.1 establishes the equivalence of the thresholds obtained by the next generation matrix and local stability analysis. 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 analogous to the observation that for $a > 0$, $f(a) = a^2 > 1$ if and only if $g(a) = a > 1$, but $f(a) \neq g(a)$. We will see this explicitly in the discrete-time SIS model presented in Section 2.5. In particular, the correct computation of R_0 requires distinguishing between *new* infections and other types of transfers into infected compartments (as represented by the $\mathcal{F}$ and $\mathcal{V}^+$ functions, respectively). The following section presents a discrete-time example which illustrates this distinction.

2.5 Example: a discrete-time SIS model with 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 [51]. 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. In this section, we develop a model with an arbitrarily-distributed infectious period, a very general formalism.

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 $\overline{X} = \sum_{i=1}^{\infty} iq_i$, 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 $I_{jk}[t]$ denote the number of individuals at time t who are infected with duration j and in the kth time step of their infection ($k \leq j$). Let $S[t]$ denote the number of susceptible individuals at time t. We'll assume the simplest stable population dynamics: at each time step, a fixed number b of new susceptibles is born and a fraction d of individuals in all compartments die. These dynamics have the unique DFE at a total of $N = b/d$ individuals. We'll also define a transmission parameter β which measures the proportion of interactions between susceptible and infected individuals which result in new infections. A set of difference equations that describes this system is as follows:

$$\begin{cases} S[t+1] &= b+(1-d)\left[\underbrace{S[t]}_{\text{susceptibles}} - \underbrace{S[t]\frac{\beta}{N}\sum_{j=1}^{M}\sum_{k=1}^{j} I_{jk}[t]}_{\text{new infections}} + \underbrace{\sum_{j=1}^{M} I_{jj}[t]}_{\text{recovered infectives}}\right] \\ I_{j1}[t+1] &= (1-d) \underbrace{q_j S[t]\frac{\beta}{N}\sum_{j=1}^{M}\sum_{k=1}^{j} I_{jk}[t]}_{\text{fraction of infectives with duration } j} \\ I_{jk}[t+1] &= (1-d) \underbrace{I_{j(k-1)}[t]}_{\text{transitions of infectives}} \qquad \text{for } 1 < k \le j \end{cases} \tag{2.7}$$

Observe that the only new infections are those that arise in the $j1$ compartments for $j = 1, \ldots, M$, while flow through the rest of the infected compartments represents transitions of already infected individuals. Ordering the compartment populations in the vector x, defined as

$$x = \begin{bmatrix} I_{11} & I_{21} & I_{22} & I_{31} & \cdots & I_{MM} & S \end{bmatrix}^\top, \tag{2.8}$$

we readily determine that F and V as defined in Theorem 2.4.1 are the $M(M+1)/2 \times M(M+1)/2$ matrices

$$F = \beta(1-d)\begin{bmatrix} q_1 & q_1 & \cdots & q_1 \\ q_2 & q_2 & \cdots & q_2 \\ 0 & 0 & \cdots & 0 \\ q_3 & q_3 & \cdots & q_3 \\ 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ q_4 & q_4 & \cdots & q_4 \\ \vdots & \vdots & \cdots & \vdots \end{bmatrix} = \beta(1-d)\begin{bmatrix} q_1 \\ q_2 \\ 0 \\ q_3 \\ 0 \\ 0 \\ q_4 \\ \vdots \end{bmatrix}\begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & \cdots \end{bmatrix}$$

$$V = \begin{bmatrix} V_1 & & & \\ & V_2 & & \\ & & V_3 & \\ & & & \ddots \end{bmatrix}$$

where V is a block diagonal matrix with M blocks and the V_i are $i \times i$ matrices with entries of 1 on the diagonal and $-(1-d)$ on the first subdiagonal. To compute R_0, we seek the largest eigenvalue of FV^{-1}. Since V is block diagonal with blocks V_i, its inverse will be block diagonal with blocks V_i^{-1}; the V_i^{-1} are lower triangular matrices with entries of 1 on the diagonal, $(1-d)$ on the first subdiagonal, $(1-d)^2$ on the second subdiagonal, and so on. Since F is a rank-one matrix, the product FV^{-1} is also rank-one:

$$\begin{aligned} FV^{-1} &= \beta(1-d)\begin{bmatrix} q_1 \\ q_2 \\ 0 \\ q_3 \\ 0 \\ 0 \\ q_4 \\ \vdots \end{bmatrix}\begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & \cdots \end{bmatrix}\begin{bmatrix} V_1^{-1} & & & & \\ & V_2^{-1} & & & \\ & & V_3^{-1} & & \\ & & & \ddots & \\ & & & & V_M^{-1} \end{bmatrix} \\ &= \beta(1-d)\begin{bmatrix} q_1 \\ q_2 \\ 0 \\ q_3 \\ 0 \\ 0 \\ q_4 \\ \vdots \end{bmatrix}\begin{bmatrix} \tilde{d}_0 & \tilde{d}_1 & \tilde{d}_0 & \tilde{d}_2 & \tilde{d}_1 & \tilde{d}_0 & \tilde{d}_3 & \tilde{d}_2 & \tilde{d}_1 & \tilde{d}_0 \cdots \end{bmatrix} \end{aligned}$$

where $\tilde{d}_i = \sum_{j=0}^{i-1}(1-d)^j$. The largest eigenvalue of this rank-one matrix is the inner product of the two component vectors; thus,

$$R_0 = \rho(FV^{-1}) = \beta(1-d)\sum_{i=1}^{M} q_i \sum_{j=0}^{i-1}(1-d)^j = \frac{\beta(1-d)}{d}\sum_{i=1}^{M} q_i(1-(1-d)^i).$$

Observe that if the death rate is slow (i.e. $0 < d << 1$), then

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

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

By Theorem 2.4.1, 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, versus the largest eigenvalue of the next-generation operator; the result is given in Figure 2.2. Note that $R_0 < 1$ 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 largest eigenvalue of the Jacobian will either under or overestimate the basic reproductive ratio of the infection.

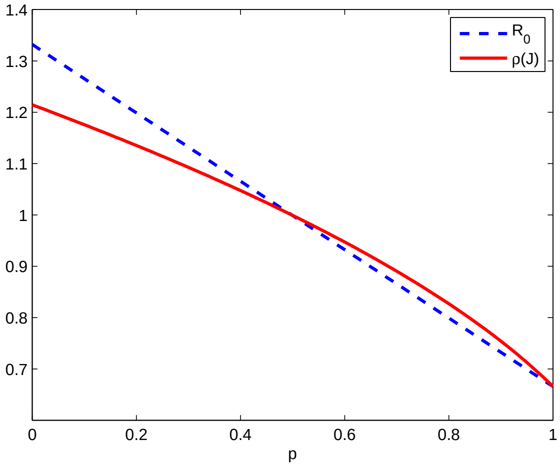


Figure 2.2. A comparison of the value of R_0 and the spectral radius of the Jacobian for the discrete-time SIS model discussed in Section 2.5 ($\beta = 2/3$ and $d = 0$).

Additionally, it is not difficult to demonstrate that this system has an endemic equilibrium, which exists as long as $R_0 > 1$. The number of susceptible individuals at this equilibrium is given by

$$S^* = \frac{N}{R_0}. \tag{2.9}$$

For $R_0 < 1$, then, the system will asymptotically converge to the disease-free equilibrium; for $R_0 > 1$, the system will converge to the endemic equilibrium. This behavior is depicted in Figure 2.3.

2.6 Equivalence with the existence of an endemic equilibrium

Is the condition $R_0 > 1$ equivalent to the existence of an endemic equilibrium in *every* infection model? Or are the conditions on local asymptotic stability and endemicity equivalent? Let us define a variable $I = \sum_{i=1}^{m} x_i$, which counts the total number of occupants in all infected compartments. We can begin to address this question by performing a *bifurcation analysis* of I, which determines the equilibrium values of I as a function of the parameters of the system. Section 2.4 demonstrated that the disease-free equilibrium $I = 0$ exists for all values of R_0 and changes from locally stable to unstable as R_0 increases past one, but in general, these conditions do not provide any information about the existence of endemic equilibria.

First, consider the bifurcation diagram in a region close to the DFE and to the point $R_0 = 1$.

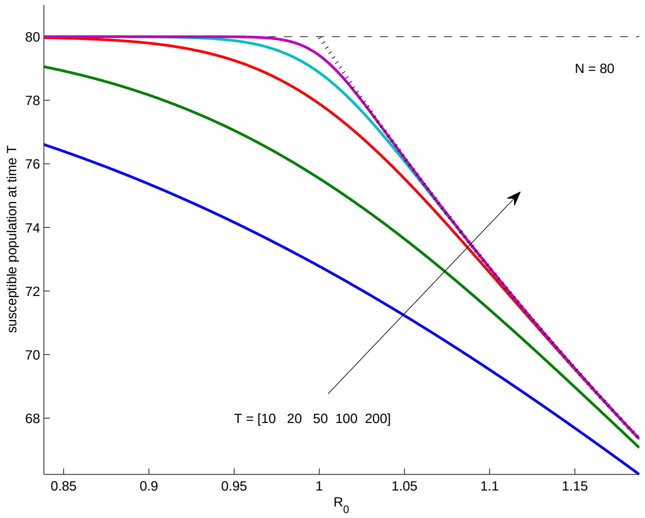


Figure 2.3. The behavior of System 2.7 for a initial introduction of 18 infectives into a population of $N = 80$ individuals with infectious period distribution $q = [1/3, 1/3, 1/3]$. The endemic equilibrium of Eq. 2.9 is represented by the dotted black line.

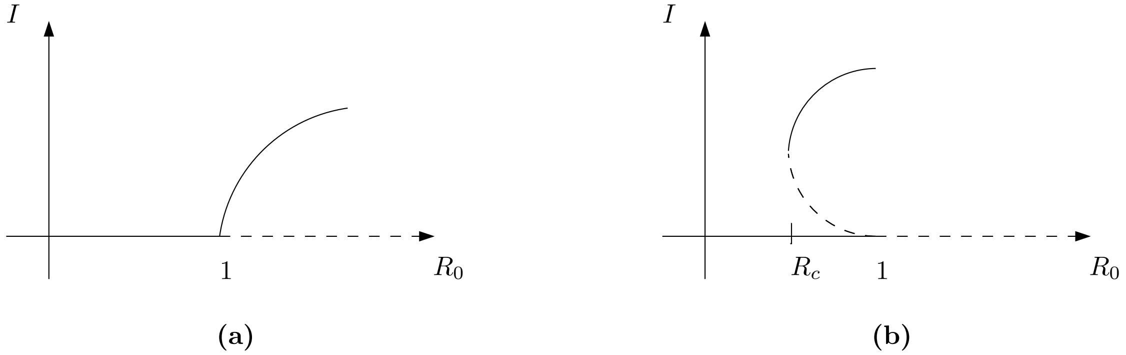


Figure 2.4. A forward bifurcation (a) and a backward bifurcation (b).

Figure 2.4 depicts two possible types of local behavior. In Figure 2.4(a), a stable endemic equilibrium is ‘born’ as R_0 increases past one, and is referred to as a *forward bifurcation*. In Figure 2.4(b), a *backward bifurcation* occurs. Here, a stable endemic equilibrium exists for a parameter range that overlaps with the interval $R_0 < 1$, with the dividing line of the basins of attraction of both stable equilibria demarcated by an unstable endemic equilibrium. In this case, the number of initial infectives introduced into the population is important; if that number is large, then the infection will reach the stable endemic equilibrium instead of dying out, for ranges of $R_c < R_0 < 1$. Additionally, if R_0 is initially greater than 1 when an infection spreads within a population, then the public health measures necessary to eliminate the disease from the population must push $R_0 < R_c$, requiring more effort than simply decreasing R_0 to below one.

In general, it can be difficult to obtain analytical solutions for the number and stability type of endemic equilibria, but there exists a class of methods from the theory of nonlinear dynamics that allows one to explore the behavior of equilibrium solutions in a small region around a bifurcation point; these methods utilize *center manifold theory* and the concept of *normal forms*. Recall that the DFE becomes locally unstable when one or more of its eigenvalues crosses a threshold curve: the imaginary axis in continuous-time systems, or the unit circle in discrete-time systems. This occurs when $R_0 = 1$. At that point, a certain number of the eigenvalues are on the threshold, with the remaining eigenvalues strictly within the stable region. Center manifold theory allows us to restrict our attention to the dynamics on a submanifold of the state-space that corresponds to the eigenvalues on the threshold. Putting the system into a normal form via a change of coordinates then classifies the type of local bifurcation. Indeed, in their development of the general framework that we've explored in this chapter, van den Driessche and Watmough obtain results on the existence of endemic equilibria via center manifold theory, but their results are fairly restrictive and difficult to interpret [46]. For a special case, Alexander and Moghadas perform a detailed local bifurcation analysis of a SIRS model using these methods in [34]. For an excellent discussion of the application of center manifold theory and normal forms to disease models, as well as many examples, see the work of Kribs-Zaleta [52]. For a more general exposition of the theory, see the work of Wiggins [53].

If the system is known to exhibit a unique forward bifurcation, then one *can* conclude the equivalence of $R_0 > 1$ and the existence of an endemic equilibrium. Backwards bifurcations are also common: some examples include an SIS model with imperfect vaccination [54], models of recurrent immuno-suppressive infections [55], malaria [56], and diseases that prompt a change in interaction patterns [57].

The global behavior of disease models is certainly not limited to the forwards and backwards bifurcations depicted in Figure 2.4; more complex dynamics are possible. Even when a center manifold analysis reveals a forward bifurcation at $R_0 = 1$, it is still possible for the system to exhibit a stable endemic equilibrium when $R_0 < 1$; Kribs-Zaleta describes an STD model that exhibits multiple endemic equilibria for $R_0 < 1$ and $R_0 > 1$ [52]. There do exist models that permit an analytical global analysis using other tools of dynamical systems theory. In [58], for example, Simon and Jacquez use Lyapunov functions to explore the stability of equilibria of an SI model.

In general, relating local stability of a fixed point to global stability is a nontrivial task, and there are few general results. For example, Reluga et al. provide sufficient conditions that preclude a backwards bifurcation in a continuous-time infection model with acquired immunity [59]. In [60], Castillo-Chavez et al. present a criterion in continuous time that guarantees the global asymptotic stability of the DFE when $R_0 < 1$, thereby precluding the existence of an endemic fixed point or limit cycle. We present their result in the following theorem.

Theorem 2.6.1. *Let ψ indicate the vector formed from the first m components of the vector x in system (2.1), corresponding to the infected compartments. If the DFE $\overline{x}$ is a globally asymptotically stable fixed point of System (2.1) when $\mathcal{F}$ is set to zero, and if the dynamics of ψ can be represented by*

$$\frac{d\psi}{dt} = (F - V)\psi - g(x)$$

where $V - F$ is an M-matrix and $g(x) \geq 0$ for all $x \geq 0$, then $\overline{x}$ is globally asymptotically stable.

2.7 Additional observations

As demonstrated in the host-vector example of Section 2.3, the value of R_0 calculated for a vector-bourne disease represents the average number of new infections in both vectors and their hosts, 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 is a more appropriate tool. Roberts and Heesterbeek [61] suggest the use of an alternate statistic, T_0, also derived from the next-generation operator, which may be more useful when control measures can only be applied to a single subpopulation.

Our attention in this chapter has been limited to discussion of the stability of equilibria of the infection model, but any epidemiologist can point to many examples of diseases occurring periodically. Many disease models exhibit oscillatory behavior, either at a natural frequency or in response to periodic forcing. For example, Wearing and Rohani observe that both seasonal variation and heterogeneity in infectivity are required to explain the observed oscillations in the prevalence of dengue in Thailand [62]. Additionally, we have only addressed the relationship between a threshold on R_0 and the stability of a disease-free population 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 map $h(x)$ composed with itself T times; in [63], Franke and Yakubu analyze such a discrete-time SIS model in a periodic environment.

Chapter 3

Network effects on thresholds

Chapter 2 presented a general framework capable of incorporating both demographic and topological heterogeneity in a population. This chapter, and the remainder of the thesis, will focus on the case in which the population X can be broken into subpopulations $X_1, \ldots, X_n$ that interact in some constrained fashion. More specifically, we assume that all individuals have identical *biology*, i.e., each subpopulation moves through the same disease *stages* in the same manner, but that the subpopulations differ in their interaction patterns. We'll begin by exploring some of the common types of models for the mixing of two subpopulations to gain some intuition for their impact on the computation of R_0, then look at the effects of interaction patterns on the general model of Chapter 2.

3.1 Population mixing and structure

The most common assumption underlying both deterministic and stochastic models of infection is the *homogeneous mixing* of individuals. As defined by Daley and Gani, "if the individuals in a population mix homogeneously, the rate of interaction between two different subsets of the population is proportional to the product of the numbers in each of the subsets concerned" [64]. The validity of this assumption is certainly context-dependent. For example, it may make sense to assume that the passengers in a subway car mix homogeneously with respect to an airborne influenza. For sexually-transmitted diseases, however, infections propagate along well-defined pathways from individual to individual within a social network and are thus poorly approximated by homogeneous mixing.

Mathematically, there are two common types of homogeneous mixing invoked in the literature. The first type of homogeneous mixing is referred to as *standard incidence*, and assumes that the rate of interaction between subpopulations of size A and B is $\alpha\frac{AB}{A+B}$; this type of mixing was invoked in the deterministic and stochastic SIS models first presented in Chapter 1. The second, *mass action incidence*, is rooted in the law of mass action, a principle from physical chemistry that describes the dynamics of well-mixed chemical reactions. Mass action incidence asserts that the rate of a reaction between two molecules which are present in quantities A and B is αAB for a constant α. As the size of one of the subpopulations under study grows large (e.g. $A \to \infty$), mass action incidence predicts a perpetually increasing rate of reaction, while the reaction rate under standard incidence remains bounded for fixed B.

How do these mixing assumptions impact the disease dynamics? To address this question, consider a single city with $N = b/d$ individuals who mix according to standard incidence, precisely the SIS model of System 1.1 in Chapter 1. Recall from our development in Section 2.3 that we can apply the formalism of the general model to obtain

$$R_0 = \frac{\beta}{\gamma + d},$$

with no dependence on population size N. Now, imagine that we take our N individuals and divide them between two cities with populations $N_1 = b_1/d$ and $N_2 = b_2/d$ (with $N_1 + N_2 = N$), and allow half of each city's population to travel to the other city continuously, regardless of their infection status. How might this change R_0? Both the numerator and denominator of the mixing term will decrease, so we might anticipate that the subdivision will have no effect on R_0. Our mathematical model now has 4 compartments, S_1, S_2, I_1 and I_2, corresponding to the susceptible and infected individuals in each city, and the dynamic model will have the following form (the equations for S_1 and S_2 are not shown):

$$\frac{dI_1}{dt} = \beta\frac{\frac{S_1}{2}\frac{I_1}{2}}{\frac{S_1+I_1}{2}} + \beta\frac{\frac{S_1}{2}\frac{I_2}{2}}{\frac{S_1+I_2}{2}} - \gamma I_1 - dI_1$$

$$\frac{dI_2}{dt} = \beta\frac{\frac{S_2}{2}\frac{I_1}{2}}{\frac{S_2+I_1}{2}} + \beta\frac{\frac{S_2}{2}\frac{I_2}{2}}{\frac{S_2+I_2}{2}} - \gamma I_2 - dI_2.$$

$$F = \begin{bmatrix} \frac{\beta}{2} & \frac{\beta}{2} \\ \frac{\beta}{2} & \frac{\beta}{2} \end{bmatrix}$$

$$V = \begin{bmatrix} \gamma & 0 \\ 0 & \gamma + d \end{bmatrix}$$

$$R_0 = \rho(K) = \rho(FV^{-1}) = \frac{\beta}{\gamma + d}.$$

Here, we see no difference in the value of R_0 computed for the subdivided population, because standard incidence has removed the impact of smaller population size. What if we repeat this example, but assume that the populations interact according to mass action incidence? With only one population of size N, our dynamic equations will take the following form:

$$\frac{dS}{dt} = -\beta SI + \gamma I + b - dS \tag{3.1}$$

$$\frac{dI}{dt} = \beta SI - \gamma I - dI. \tag{3.2}$$

It is not difficult to see that

$$R_0 = FV^{-1} = \frac{\beta N}{\gamma + d}, \tag{3.3}$$

which increases with the population size N. What will happen if we subdivide the population? We have effectively *reduced* the size of the population in which an individual mixes, so we'd expect the rate of new infections, and consequently R_0, to decrease. Applying the subdivision, we obtain the following model:

$$\begin{aligned}\frac{dI_1}{dt} &= \beta\frac{S_1}{2}\frac{I_1}{2} + \beta\frac{S_1}{2}\frac{I_2}{2} - \gamma I_1 - dI_1\\ \frac{dI_2}{dt} &= \beta\frac{S_2}{2}\frac{I_1}{2} + \beta\frac{S_2}{2}\frac{I_2}{2} - \gamma I_2 - dI_2.\end{aligned}$$

The F and V matrices for this system are given by

$$F = \begin{bmatrix}\beta\frac{N_1}{4} & \beta\frac{N_1}{4}\\ \beta\frac{N_2}{4} & \beta\frac{N_2}{4}\end{bmatrix}$$

$$V = \begin{bmatrix}\gamma+d & 0\\ 0 & \gamma+d\end{bmatrix},$$

which yields

$$K = FV^{-1} = \frac{\beta}{4(\gamma+d)}\begin{bmatrix}N_1 & N_1\\ N_2 & N_2\end{bmatrix},$$

which has largest eigenvalue

$$R_0 = \rho(K) = \frac{\beta}{4(\gamma+d)}(N_1+N_2) = \frac{\beta N}{4(\gamma+d)}.$$

This result confirms our intuition: by subdividing the population, we've decreased the total number of interactions, and thus slowed the potential growth of the epidemic.

Which is more correct? The type of interaction underlying one's model *must* depend on the infection under study. In mass action mixing, larger populations mean more infection opportunities; if the size of Boston doubles, then an infective can infect twice as many people as he could before the change. While this feature may be grossly incorrect for many kinds of interactions (e.g., sexually-transmitted diseases), such an assumption does have a place in certain kinds of infections. For example, consider a highly transmissible respiratory infection that only requires passing contact. There are roughly 14 times as many people in New York City as there are in Boston, and thus a tourist from Boston visiting New York may casually pass 14 times as many people in a given day as she would in Boston (e.g., on the subway or at large events). For a casually-transmissible infection, then, we'd like to see growth in the infection rate as population size increases. Indeed, any infection whose spread worsens in areas of high population density requires some measure of this effect.

Instead, we might want to model a situation in which the number of individuals that we interact with is limited to some maximum possible number; for a fixed number of infectives, as the size of the population increases to infinity, the rate of new infections will increase only to this limit. Standard incidence has this feature. Effectively, the parameter β is rescaled to β/N, where N is the total population. In our example, dividing one city into two smaller cities does not impact R_0 because the infection rate has been rescaled to decrease with the size of the mixing subpopulations. This subdivision simultaneously decreased the mixing pool and increased the infection rate so that R_0 remained the same. Similarly, connecting two cities that didn't previously interact would not change R_0 under standard incidence, because the mixing pool is increasing simultaneously with a decrease in the infection rate.

Choosing the right type of functional form for the mixing between subpopulations, then, has critical ramifications on the value of the basic reproductive ratio. In order for restricted interactions of subpopulations to impact the ability of a newly-introduced infection to spread, these interactions must change the *speed* with which the infection propagates from its nominal speed in a fully-mixed population. For this to appear in our predictions, we must choose a mathematical model with this property.[1] Throughout this thesis, we will often use mass action incidence as a proxy for any general form of mixing function which exhibits this behavior, but our general conclusions are not restricted by this specific form. The following section considers a second aspect of the choice of mixing function.

3.2 General incidence functions

Certainly, acceptable modeling simplifications are highly dependent on the nature of the infection under study, and there is a need for models which interpolate between extreme assumptions. Over the last several decades, researchers have considered many different mathematical forms for the rate of new infections as a function of the size of the populations within each compartment. The general model described in Chapter 2 puts no constraints on the precise form that the rate of new infections can take, beyond the biologically-required assumptions (A1)-(A5). However, there are functional forms that yield *degenerate* expressions for F: the all-zeros matrix, or a matrix whose entries are not all finite. For example, one family of models represents mixing with terms proportional to S^pI^q for some constants p and q [65].[2] If we replace SI in Eqs. 3.1-3.2 by S^pI^q, and compute the matrix F, we find that

$$F = q\beta S^p I^{q-1}\big|_{(S,I)=(N,0)} \,. \tag{3.4}$$

For $q < 1$, $F = \infty$, which leads to $R_0 = \infty$. For $q > 1$, $F = 0$ and $R_0 = 0$. These values of R_0 arise solely from the functional form chosen and have no dependence on the parameters of disease

[1] It is important to observe that although network topology does not change the value of R_0 under standard incidence, it certainly does impact the dynamics of the model! As we observed in Chapter 2, R_0 is a measure of an initial growth rate, and does not provide information about other phenomena of interest, including longer-term behavior or the spatial patterns of spread.

[2] These models have been invoked as natural generalizations of the bilinear form SI and justified as a way to incorporate population heterogeneities [66], but have not found wide application. In [65], the authors acknowledge that such functional forms would be difficult to distinguish from a bilinear form in empirical data.

transmission, which is both intuitively and mathematically problematic. Is it possible to classify the models which will yield non-degenerate expressions for R_0? Note that this definition of non-degeneracy for R_0 does not imply that R_0 itself cannot be zero, as long as that value does not arise from degeneracies in F or V^{-1}. The following definition characterizes such models.

Definition 3.2.1. *The general model described in Chapter 2 will yield a non-degenerate R_0 if and only if the matrix F has finite entries and is not identically zero.*

Remark 3.2.2. *In Chapter 2, we observed that V^{-1} is a nonnegative matrix. Its invertibility guarantees that the entries of V^{-1} are finite and the columns of V^{-1} form a linearly independent set. We also observed that F is a nonnegative matrix. Since V^{-1} is invertible, FV^{-1} is the zero matrix if and only if F itself is the zero matrix. Additionally, assume that $\rho(K) = \infty$. Since V^{-1} has finite entries, this occurs if and only if the F has at least one infinite entry.*

Throughout the remainder of this thesis, we will restrict our attention to models that produce non-degenerate R_0, i.e. those that satisfy the conditions of Definition 3.2.1.

3.3 Identical individuals interacting via a network

As stated at the start of this chapter, we'd like to focus on a special case of heterogeneity in which a population X can be broken into subpopulations $X_1, \ldots, X_n$ such that

- ▷ all individuals across subpopulations have identical *biology*, i.e. individuals in each subpopulation move through the same disease stages once infected,
- ▷ but differ in their *interaction patterns*, i.e. the level to which subpopulation X_j mixes with subpopulation X_i.

This idea is illustrated in Figure 3.1, 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. Assuming mass action incidence, our infection model might take the following form:

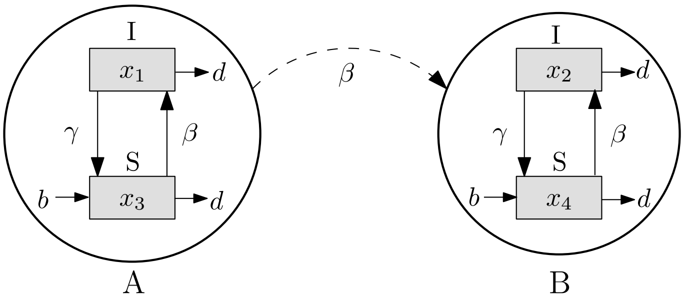


Figure 3.1. X_1 and X_2 represent two different subpopulations of individuals, S and the I represent susceptible and infected compartments within each subpopulation.

$$\begin{cases} x_1 & \leftarrow & x_1 + \underbrace{\beta x_1 x_3}_{\mathcal{F}_1} - \underbrace{(\gamma x_1 + dx_1)}_{\mathcal{V}_1^-} \\ x_2 & \leftarrow & x_2 + \underbrace{\beta x_2 x_4 + \beta x_1 x_4}_{\mathcal{F}_2} - \underbrace{(\gamma x_2 + dx_2)}_{\mathcal{V}_2^-} \\ x_3 & \leftarrow & x_3 + \underbrace{b + \gamma x_1}_{\mathcal{V}_3^+} - \underbrace{(\beta x_1 x_3 + dx_3)}_{\mathcal{V}_3^-} \\ x_4 & \leftarrow & x_4 + \underbrace{b + \gamma x_2}_{\mathcal{V}_4^+} - \underbrace{(\beta x_2 x_4 + \beta x_1 x_4 + dx_4)}_{\mathcal{V}_4^-} \end{cases} \tag{3.5}$$

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 $\overline{x} = (\overline{x}_1, \overline{x}_2, \overline{x}_3, \overline{x}_4) = (0, 0, b/d, b/d)$, so each subpopulation size is $N = b/d$ at equilibrium. Note that the Jacobian matrix that governs small perturbations away from $\overline{x}$ 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.

As discussed in Chapter 1, it is natural to describe the structure of these interactions by an *adjacency matrix* A, which has a '1' entry in the ij^{th} position if infected individuals in subpopulation i can infect susceptible individuals in subpopulation j; for the example in Figure 3.1,

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

How does the structure of the interaction between the subpopulations impact the matrices F and V? For System 3.5, we find that

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

which implies that

$$R_0 = \rho(FV^{-1}) = \rho\left(\frac{\beta}{\delta + d} N A^\top\right) = \frac{\beta}{\delta + d} N \rho(A^\top). \tag{3.6}$$

Observe that $\frac{\beta}{\delta+d}N$ is the value of R_0 obtained for a single population of size N in Eq. 3.3. Denote this value by R_h. When the subpopulations are connected according to the adjacency matrix A, the value of R_0 changes by a factor of $\rho(A^\top)$. It is not difficult to extend System 3.5 to more than two subpopulations with different interaction patterns; the general forms of F and V will remain the same. The expression for R_0 in Eq. 3.6 has *decoupled* the biology of the infection (the progression through disease stages, summarized by R_h) and the impact of the population topology (summarized by $\rho(A^\top)$).

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. 3.6 using the *Kronecker product* (discussed in Section 1.6). For any model 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.[3]

Recall that for matrices A, B, C, and D of compatible dimensions, $(A \otimes B)(C \otimes D) = (AC) \otimes (BD)$, and that $\rho(C \otimes D) = \rho(C)\rho(D)$. These properties allow us to write the next-generation matrix K as

$$K = (F_h \otimes A^\top)(V_h \otimes I)^{-1} = F_h V_h^{-1} \otimes A^\top \tag{3.7}$$

and

$$R_0 = \rho(K) = \rho(F_h V_h^{-1} \otimes A^\top) = R_h \rho(A^\top) = R_h \rho(A). \tag{3.8}$$

where $R_h = \rho(F_h V_h^{-1})$. The expression for R_0 in Eq. 3.8 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 $\rho(A)$. We will take advantage of the decoupling of biology and topology represented in Eqs. 3.7 and 3.3 throughout the remainder of this thesis. In light of this observation, many past and recent results regarding epidemics on complex networks can be seen as simple consequences. We conclude this chapter with a sampling of these results.

3.3.1 Examples from the literature

Anderson and May, 1991

Anderson and May explore a population model in which N_i individuals have i contacts with other individuals for $i = 1, \ldots, M$ [38]. The number of contacts of any individual is not correlated with the number of contacts of its neighbor, and thus the average number of contacts between an i-type individual and a j-type individual is given by

$$\frac{ij}{\sum_k kN_k}.$$

[3] Observe that these assumptions do *not* imply that there can be only one disease stage into which new infections can occur, only that new infections arise via interactions between subpopulations, while all other movements through disease stages do not.

Thus, A is the rank-one matrix given by the following vector product, which has N_i entries of i for every i:

$$A = \frac{1}{\sum_k kN_k} \begin{bmatrix} 1 \\ \vdots \\ 1 \\ 2 \\ \vdots \\ 2 \\ \vdots \\ M \\ \vdots \\ M \end{bmatrix} \begin{bmatrix} 1 & \cdots & 1 & 2 & \cdots & 2 & \cdots & M & \cdots & M \end{bmatrix} \tag{3.9}$$

It is not difficult to show that the largest eigenvalue of this matrix is given by:

$$\rho(A) = \frac{\sum_k k^2 N_k}{\sum_k kN_k} = \frac{\langle k^2 \rangle}{\langle k \rangle}$$

where $\langle \cdot \rangle$ indicates the average value. This observation corresponds to the threshold condition derived in [38].

Pastor-Satorras and Vespignani, 2001

In [39], Pastor-Satorras and Vespignani formulate mean-field equations for an SIS process on two different types of networks. The first type is referred to as *exponential networks*, which are characterized by a degree distribution that is sharply peaked at its average value $\langle k \rangle$ and decays exponentially fast on either side of $\langle k \rangle$. The authors mention the Watts-Strogatz "lattice rewiring" graph as an example of an exponential network. Mathematically, the peak of the degree distribution at $\langle k \rangle$ leads the authors to the approximation that all nodes have degree $\langle k \rangle$. For an N-node network, this yields the following $N \times N$ matrix A:

$$A = \begin{bmatrix} \frac{\langle k \rangle}{N} \\ \frac{\langle k \rangle}{N} \\ \vdots \\ \frac{\langle k \rangle}{N} \end{bmatrix} \begin{bmatrix} 1 & 1 & \cdots & 1 \end{bmatrix} \tag{3.10}$$

Clearly, the largest eigenvalue of this A is simply $\langle k \rangle$, which corresponds to the threshold condition determined in [39] for exponential networks.

The second type of network modeled in [39] are those with *power-law* degree distributions, with uncorrelated degrees. The authors cite the Barabasi-Albert (BA) model of preferential attachment as an example of this type of graph, but BA do models typically exhibit degree correlations (see Section 4.2.2 for a demonstration of this phenomenon). The power-law degree distribution is given by a probability density over the degree of a node, and is formalized in [39] as $P(k) = 2m^2k^{-3}$, where m is the minimum degree of a node in the network. This distribution is not clustered about its mean (which leads to the alternate term 'scale-free'), so the authors choose their dynamic variables to be the fraction of nodes of degree-k which are infected. One might observe that this construction is identical to that presented by Anderson and May in [38], and thus it should be no surprise that the threshold condition observed in [39] is dictated by the value of $\frac{\langle k^2 \rangle}{\langle k \rangle}$. For infinitely large power-law networks that have exponent ≥ 3, the quantity $\langle k^2 \rangle$ will not converge, and thus the authors conclude that there is no epidemic threshold for such graphs: any infection can propagate indefinitely.

Boguna and Pastor-Satorras, 2002

In [30], the authors address the possibility of degree correlations among the nodes in the network. Specifically, they focus on the case of *Markovian dependencies* between the degrees of adjacent nodes, in which the network structure is defined by the degree distribution $P(k)$ and the conditional distribution $P(k'|k)$, where the latter quantity denotes the probability that a neighbor of a degree-k node will have degree k'. They define a matrix C whose (k, k') entry is given by $kP(k'|k)$. This matrix C is the weighted adjacency matrix for an appropriately-defined network, so it is intuitive that the threshold condition observed by the authors is dictated by the largest eigenvalue of C.

Chapter 4

Approximating and bounding thresholds

THIS thesis focuses on the dynamic behavior of mathematical models of infection processes through structured populations. As in the previous chapters, we'll focus on the structure that arises via interaction constraints for a population of otherwise identical individuals. We conclude this chapter with a discussion of generalizations of R_0 that are appropriate for systems in which the interaction patterns are uncertain or changing with time.

4.1 Uncertainty in interaction patterns

There is no shortage of deterministic and stochastic models (in both continuous and discrete time) that have been proposed for various infections (biological, social and technological), and all of them have a common feature: the population structure is assumed to be completely known. Even in stochastic models, the population structure is rarely modeled as a random phenomenon. However, what if the population structure is unknown? It is rare that one has complete information about the connectivity of a network, especially when one is also required to estimate edge weights. Consider the following partial information scenarios.

1. A subgraph (or collection of subgraphs) is known. In social network analysis, for example, it is often the case that only local information is gathered (i.e., the neighborhood of individual network nodes), or that a subgraph of the complete population is mapped out via contact tracing (e.g., with tuberculosis diagnoses) or snowball sampling [67]. Similarly, web crawlers that attempt to map the structure of the WWW follow the outgoing links of an initial set of pages (and tend to exhibit biases in the structure they detect [68]). Certainly, different sampling methods produce different pictures of the population under study (and are often accompanied by errors), and all are necessarily incomplete.

2. A generation/evolution mechanism for the network can be hypothesized. The prevalence of power-law degree distributions in nature has inspired an enormous amount of interest in the network growth mechanisms that would generate such a distribution; among the most referenced explanations are Barabási and Albert's preferential attachment model [69] and Chung's duplication model [70]. A researcher might have knowledge of how the nodes connect with each other, which provides information about the resulting network structure.

3. Ranges for certain network parameters can be estimated. Common sense or empirical observations can bound the possible choices for some important network statistics. As a naive example, in a family tree, we can assume that no person has more than twenty children, which provides an upper bound on the maximum outdegree of a node.

All of these types of partial information suggest modeling the network as a probabilistic ensemble of all graphs that satisfy the known conditions. Perhaps this distribution arises out of some random network generation process, like preferential attachment, or perhaps it is a single snapshot of a network whose edges reflect the preferences of the individual nodes.

In general, let us consider a sample space, Ω, such that for each $\omega \in \Omega$, $\mathcal{A}(\omega)$ is one possible realization of the network. We will assume that a valid probability distribution $P(\cdot)$ is given over this sample space, which associates to each ω a probability that realization ω is seen. The graph $\mathcal{A}$ can be completely specified by the collection of random elements $\{N, \{\mathcal{N}_i\}, \mathcal{E}\}$ where

- ▷ N is the number of nodes in the graph,
- ▷ $\mathcal{N}_i$ is the set of attributes associated with node i (possibly vector-valued),
- ▷ $\mathcal{A}_{ij} \in \mathcal{E}$ indicates the strength or nature of a connection between nodes i and j (again, possibly vector-valued).

Our graph generation process, then, could be thought of a function associating each $\mathcal{A}(\omega)$ with a set $\{N(\omega), \{\mathcal{N}_i(\omega)\}, \mathcal{E}(\omega)\}$, where $\mathcal{A}(\omega)$ occurs with probability $P(\omega)$. It is rare, however, that this much generality is necessary or useful! There are many possible ways of defining an ensemble of random graphs and assigning probabilities to its realizations. In Chapter 5, we consider a particular family of random graph distributions known as the *exponential* or p^* random graphs, which are used extensively in social network analysis.

How do we incorporate this randomness into infection models? For stochastic models, we must include an initial step of choosing a particular network from the ensemble before applying the stochastic infection process. If the ensemble comprises networks whose edges are determined with some degree of independence, then we may be able to embed the randomness of the network structure within the process of infection spread; if edges are correlated, however, this extra step might destroy whatever tractability we began with. In deterministic models, which typically use differential or difference equations to model infection spread, incorporating random structure amounts to including random variables in our equations. For a difference equation, this implies that the state vector x, a random vector, evolves according to the update

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

where ψ is a known vector of parameters and $\mathbf{A}$ is the random *adjacency matrix* of the underlying network, $\mathcal{A}$. This is a stochastic differential 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 $\mathbf{x}(t)$ is difficult. Most deterministic models implicitly replace the random matrix $\mathbf{A}$ with its expected value $E[\mathbf{A}]$, yielding a deterministic differential equation; in general, however,

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

so even tracking $E[\mathbf{A}]$ is challenging. This seems like a desperate situation! However, instead of considering the full dynamics of these processes, what if we restrict our attention to computing the most widespread parameter in mathematical epidemiology, R_0? 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 chapter, Eq. 3.7 shows that $\mathbf{K}$ is a *linear* function of $\mathbf{A}$ and therefore more amenable to analysis; in this case,

$$\mathbf{R_0} = R_h \rho(\mathbf{A}).$$

The value of R_h is determined by the biology of infection; here, we'll assume that this is a known quantity. If we're interested in how $\mathbf{R_0}$ is distributed, then, we can go directly from information about the distribution of the spectral radius of $\mathbf{A}$ to a distribution on $\mathbf{R_0}$.

Note that we have only defined R_0 for deterministic dynamic models; how might the structure of the population impact the computation of stochastic thresholds? As discussed in Chapter 2, the definition of R_0 was given by Diekmann et al. as the expected number of secondary infections caused by a single infected individual in a completely susceptible population, but the "expectation" that this definition refers to is a *population average* over an infinitesimally divisible collection of individuals, not the *expectation* of a random variable [26]. One might naively guess that in stochastic models, the related quantity is the true "expectation" of an inherently random next-generation matrix $\mathbf{K}$, and indeed, $E[\mathbf{K}]$ often dictates threshold results. In [71], the authors present a continuous-time stochastic model in which individuals are of different *classes*, which define their infectious period and mixing patterns, and are also partitioned into *households*, with a higher frequency of contact within households (local) than across households (global). If the *type* of an infected household is the class to which its first infected individual belongs, the authors use a branching process approximation to determine that "a global epidemic occurs with non-zero probability if and only if" the spectral radius of a matrix M is greater than one, where "M_{ij} is the mean number of class-j global contacts that emanate from a typical type-i infected household" [71]. In order to obtain this result, however, the authors take the limit as the population size approaches infinity; this allows them to apply a branching process approximation to determine whether or not the infection will reach a finite fraction of the population before extinction. In most stochastic models, however, the population structure is assumed to be known. For example, Draief et al. describe a Reed-Frost model of infection over a deterministic network and find that $\rho(A)$ determines whether an outbreak will be "large" or "small" in a probabilistic sense [72]. Similarly, Ganesh et al. define a continuous-time Markov process for infection propagating through a deterministic network and find that $\rho(A)$ serves as a threshold for the expected duration of an epidemic [25].

Regardless of the modeling approach one uses, assuming that the network structure is drawn from some distribution of adjacency matrices gives you a distribution of the threshold parameter, rather than a single value. What if, to avoid this entire issue, we simply replaced the unknown population structure $\mathbf{A}$ with some kind of guess? A reasonable one might be the expected adjacency matrix of the entire ensemble, which we'll denote $E[\mathbf{A}]$. Is the value of $\rho(E[\mathbf{A}])$ an appropriate summary measure of the distribution of $\rho(\mathbf{A})$? In general, the answer is no. If we assume that contacts between individuals are symmetric, then the underlying network is undirected and the adjacency matrix of the network is symmetric. A nonnegative symmetric matrix is Hermitian, which has eigenvalues that are purely real, and we reference the following theorem from Horn and Johnson [24].

Theorem 4.1.1. *Let X and Y be $n \times n$ Hermitian matrices whose eigenvalues are given by $\lambda_i(X)$ and $\lambda_i(Y)$, respectively, and let the λ_i be arranged in increasing order from $i = 1, \ldots, n$. Then*

$$\lambda_k(X) + \lambda_1(Y) \leq \lambda_k(X+Y) \leq \lambda_k(X) + \lambda_n(Y).$$

A corollary quickly follows.

Corollary 4.1.2. *If $\mathbf{A}$ is the adjacency matrix of a random undirected graph on n nodes, then*

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

Proof. Since the set of all possible $n \times n$ adjacency matrices is bounded, $E[\mathbf{A}]$ exists and can be represented as

$$E[\mathbf{A}] = \sum_{i=1}^{2^n} p_i A_i$$

where A_i is a possible realization of $\mathbf{A}$ and p_i its associated probability. Since all of the A_i and $E[\mathbf{A}]$ are nonnegative and symmetric, they are Hermitian, and Theorem 4.1.1 implies that

$$\rho(E[\mathbf{A}]) = \rho\left(\sum_{i=1}^{2^n} p_i A_i\right) \leq \sum_{i=1}^{2^n} p_i \rho(A_i) = E[\rho(\mathbf{A})].$$

□

Thus, using $E[\mathbf{A}]$ can lead us to underestimate the mean of the distribution of eigenvalues, which is problematic; it means that we're underestimating the epidemic potential of the infection. For directed graphs, it is possible for either $\rho(E[\mathbf{A}])$ or $E[\rho(\mathbf{A})]$ to be the larger of the two. 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 [73].

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. Indeed, one may have only partial information about the structure via some of the following statistics and observations:

- ▷ total number of nodes and edges in the network;
- ▷ maximum or minimum degree, network girth (the length of the shortest cycle), or network diameter (the length of the longest path);
- ▷ average degree and variance, degree distribution, possibly accompanied by 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.

If we are purely interested in determinining whether or not $\mathbf{R_0} > 1$, partial information may allow us to make this assessment. For example, if we know that $\rho(\mathbf{A}) < 2$ and $R_{0,h} < 0.25$, then $R_{0,h}\rho(\mathbf{A}) < 0.5 < 1$ and an epidemic cannot occur. We can determine bounds on the spectral radius of the adjacency matrix using structural information via the tools of *spectral graph theory*, the focus of the following section.

4.2 Bounding and approximating $\rho(\mathbf{A})$

The literature of spectral graph theory is rich with bounds on the spectrum of adjacency matrices, given as functions of structural information. Determining such bounds using the structural properties of the network is one of the tasks of *spectral graph theory.* In Tables 4.1 and 4.2, 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 4.1 and 4.2 provide guaranteed bounds on the value of $\rho(\mathbf{A})$. We also present bounding results for the spectral radius of nonnegative matrices, which form the more general class of weighted adjacency matrices; these results are presented in Table 4.3. Finally, we summarize some results that bound the elements of the *eigenvector* associated with the largest eigenvalue in Table 4.4. Observe that most of these results are upper bounds; the literature on useful lower bounds for $\rho(\mathbf{A})$ is much more sparse. To see where the difficulty might arise, Theorem 4.1.1 allows us only to conclude that $\rho(E[\mathbf{A}]) \geq E[\lambda_{min}(\mathbf{A})]$, but λ_{min} is often negative. Since we know that $\rho(\mathbf{A}) > 0$, this bound does not provide any new information.[1]

Table 4.1. 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	$\sqrt{2e}$	[75]
$\{e\}$	$\frac{-1+\sqrt{1+8e}}{2}$	[76]
$\{n, \delta, e\}$	$\frac{(\delta-1)+\sqrt{(\delta+1)^2+4(2e-\delta n)}}{2}$	[77]
$\{m_i\}$	$\max\{\sqrt{m_i m_j} \mid (i,j) \in E\}$	[74]
$\{d_i, m_i\}$	$\max\{\sqrt{d_i m_j} \mid (i,j) \in E\}$	[78]
$\{n, e, \delta, \Delta\}$	$\sqrt{2e-(n-1)\delta+(\delta-1)\Delta}$	[74]
$\{n, D, \Delta, \delta\}$	$\Delta - \frac{\Delta+\delta-2\sqrt{\Delta\delta}}{Dn\Delta}$	[79]
$\{d_i\}$	$\min_{1 \leq i \leq n} \frac{d_i-1+\sqrt{(d_i+1)^2+4(i-1)(d_1-d_i)}}{2}$	[80]
$G \geq 5$, $\{n, \Delta\}$	$\min(\Delta, \sqrt{n-1})$	[81]
$G \geq 5$, $\{n, \Delta\}$	$\frac{-1+\sqrt{4n+4\Delta-3}}{2}$	[82]

Tables 4.1-4.3 presented bounds that can be rigorously established for the largest eigenvalue of the adjacency matrix of a graph for which only partial information is known. 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.

4.2.1 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

[1] See the survey of Das and Kumar [74] for several negative lower bounds.

Table 4.2. Lower bounds on $\rho(\mathbf{A})$. Structural properties are number of nodes (n), number of edges (e), maximum degree (Δ), minimum degree (δ), girth (G), diameter (D), 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	$2\cos\frac{\pi}{n+1}$	[83]
$\{\Delta\}$, simple	$\sqrt{\Delta}$	[78]
$\{n, e\}$, no multiple edges	$\frac{2e}{n}$	[75]
$\{n, d_i\}$, simple	$\sqrt{\frac{1}{n}\sum_i d_i^2}$	[78]
$\{n, d_i\}$, simple	$\frac{1}{e}\sum_{(i,j)\in E}\sqrt{d_i d_j}$	[78]

Table 4.3. Bounds on $\rho(\mathbf{A})$ for A a non-negative matrix. Matrix information includes the dimension (n), sum of the entries of the ith row (d_i), the minimum and maximum over these sums (δ and Δ, respectively), the minimum entry of $\mathbf{A}$ (b), the trace of $\mathbf{A}$ (t_1), the trace of $\mathbf{A}^2$ (t_2).

matrix information	bounds on $\rho(\mathbf{A})$	reference
$\{\delta, \Delta\}$	$\delta \leq \rho(A) \leq \Delta$	[84]
positive A, $\{\delta, \Delta, b\}$	$\delta + b(h-1) \leq \rho(A) \leq \Delta - b(1 - 1/g)$ $g = \frac{\Delta - 2b + \sqrt{\Delta^2 - 4b(\Delta-\delta)}}{2(\delta - b)}$ $h = \frac{-\delta + 2b + \sqrt{\delta^2 + 4b(\Delta-\delta)}}{2b}$	[85]
$\{n, t_1, t_2\}$	$\rho(A) \geq \frac{t_1}{n} + \sqrt{\frac{1}{n(n-1)}\left(t_2 - \frac{t_1^2}{n}\right)}$	[86]

Table 4.4. Bounds on the elements of the maximal eigenvector $\mathbf{v}$ of $\mathbf{A}$. Matrix information includes the dimension (n), sum of the entries of the ith row (d_i), the minimum and maximum over these sums (δ and Δ, respectively), the minimum diagonal entry of $\mathbf{A}$ (a_d).

matrix information	bounds on elements of $\mathbf{v}$	reference
$\delta, \Delta, \{A_{ij}\}$	$\sqrt{\frac{\Delta}{\delta}} \leq \max_{i,j}\frac{v_i}{v_j} \leq \max_{j,s,t}\frac{a_{sj}}{a_{tj}}$	[84]
δ, Δ, a_d, $A > 0$	$\max_{i,j}\frac{v_i}{v_j} = \sqrt{\frac{\Delta - a_d}{\delta - a_d}}$	[84]

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 known to have 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:

$$E[\mathbf{A}] = \frac{1}{\sum_k kN_k} \begin{bmatrix} 1 \\ \vdots \\ 1 \\ 2 \\ \vdots \\ 2 \\ \vdots \\ M \\ \vdots \\ M \end{bmatrix} \begin{bmatrix} 1 & \cdots & 1 & 2 & \cdots & 2 & \cdots & M & \cdots & M \end{bmatrix}. \tag{4.1}$$

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

$$\frac{\sum_k k^2 N_k}{\sum_k kN_k} = \frac{\langle k^2 \rangle}{\langle k \rangle}, \tag{4.2}$$

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 [38], then rederived by Pastor-Satorras and Vespignani [87]. 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.

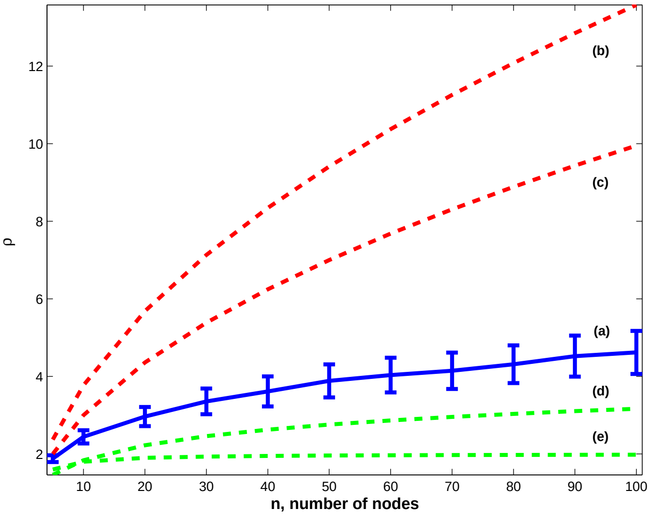


Figure 4.1. (a) The mean (± std. dev.) of the spectral radius of the adjacency matrix of a simple preferential attachment model on n nodes ($n-1$ edges), taken over 100 trials; (b) an upper bound on $\rho(\mathbf{A})$ obtained using the number of edges [76] (coincides with [82]); (c) an upper bound obtained using the number of nodes, edges, minimum degree and maximum degree [74] (coincides with [81]); (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.

4.2.2 Example 2: preferential attachment

Consider a network generated by a simple preferential attachment mechanism, slightly modified from the one described by Barabási and Albert [69]. 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 $n \to \infty$, 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} [69]. Figure 4.1 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. 4.2, making assumptions like these that reduce the heterogeneity of the network causes us to underestimate an infection's spreading potential.

4.2.3 Example 3: egocentric network data from Houston study

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 [88], [89], and [90]. As part of the survey, participants named up to 6 other individuals who were a part of their social network and assessed whether these individuals knew *each other*, which illuminates local subgraphs of the larger social network of this community. For a more complete description of this data source, refer to Appendix B.3.

From this data, we are able to measure three 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 *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 vertices. Following [91], we'll define the *average clustering coefficient* to be
$$C = \frac{1}{|V'|}\sum_{i\in V'} C_i.$$
3. The joint distribution of degree and clustering coefficient: see Appendix B.3 for a discussion of this property.

Figure 4.2 depicts the first network property, and we can compute the average clustering coefficient to be $C = 0.312$ over the participants in the study. The mean degree of participants who listed at least one contact is $d = 2.925$. Participants who listed no contacts become isolated nodes 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. Can this information be used to estimate a value of $\rho(A)$ for the network from which this data was drawn? First, a population size must be assumed, the choice of which will depend upon the population of interest; in order to illustrate these approximation techniques, we fix $n = 1000$ individuals with at least one contact. From the degree distribution, then, it is possible to estimate several parameters:

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

which we can use in the bounds presented in Section 4.2. Observe that these results will no longer be bounds on the support of $\rho(\mathbf{A})$, because we have made structural assumptions to guess the

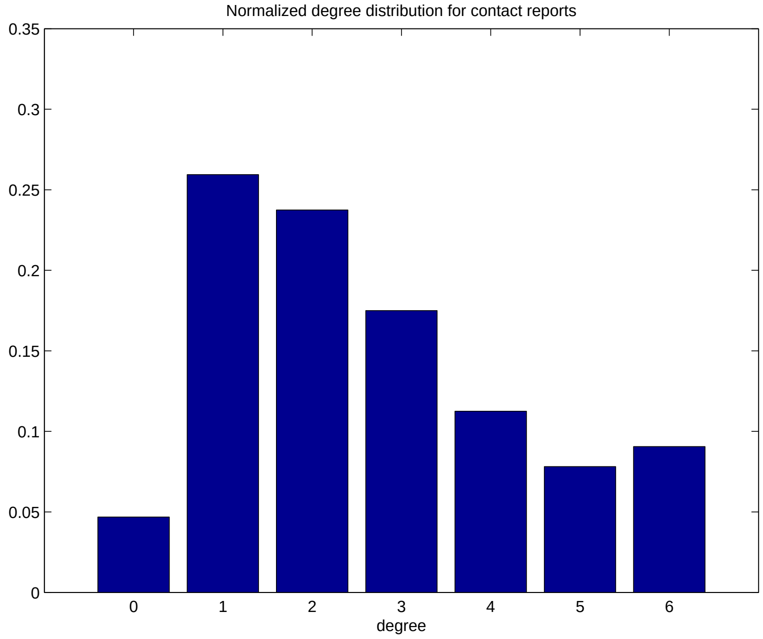


Figure 4.2. The degree distribution of the social network described by the Houston data set.

parameter. 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. 4.2 is an approximation. We summarize the results of these various computations, as well as others described below, in Figure 4.4.

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 [92]. 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.[2] Mathematically, this requires that the probability of a

[2] This family of random graphs will be the focus of Chapter 5; also see [93] and [94].

graph, denoted by a, takes the following form:

$$P(a) = \frac{1}{\kappa} \exp\left(\sum_k \theta_k z_k(a)\right) \tag{4.3}$$

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 [95]. 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* [96].

For the Volz and ERGM approximations, histograms of the resulting values of $\rho(\mathbf{A})$ are depicted in Figure 4.3. The means of these respective histograms, along with the bounds and approximations described earlier, are summarized in Figure 4.4. Using the degree distribution allows us to come quite close to the Volz algorithm simulations. 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; the reverse situation will be seen in the following section.

4.2.4 Example 4: airline traffic data

The Bureau of Transportation Statistics, an organization under the U.S. Department of Transportation, makes publicly available detailed data on domestic airline flights, among other modes of transportation. This section focuses on passenger flow between U.S. cities over the month of January 2007; for that month, we have an estimate of the number of passengers flying between 9986 directed pairs of U.S. cities. For more information regarding the collection and processing of this data, see Appendix B.1.

What is the appropriate 'adjacency matrix' to assemble from this data? If we are interested in the spread of a winter illness, like the flu or a common cold, then we might hypothesize that the rate at which such an infection spreads increases with the passenger volume and population of

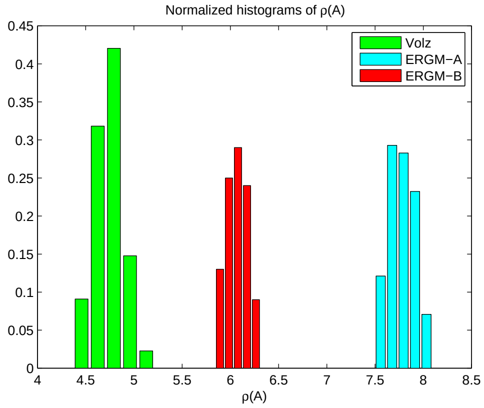
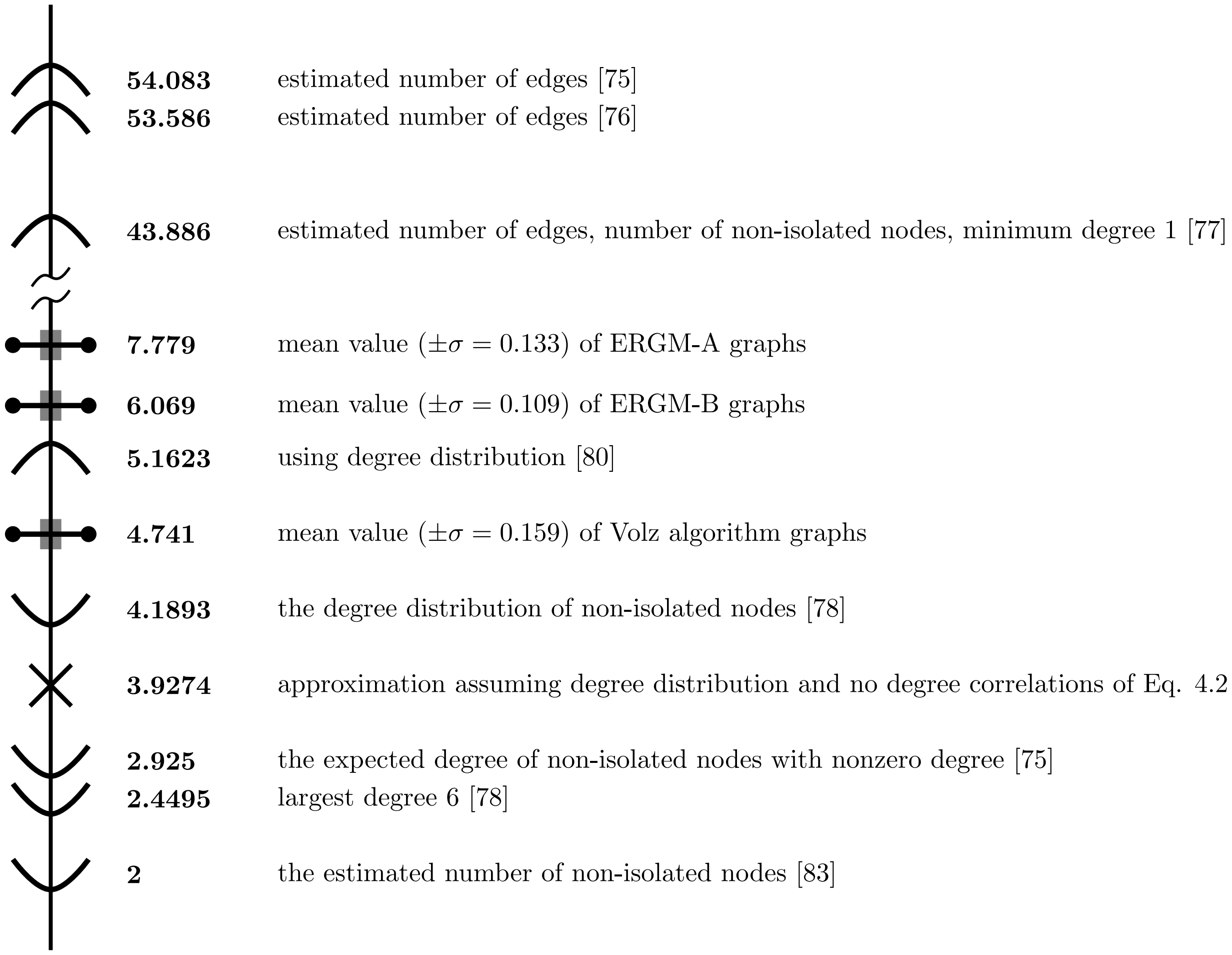


Figure 4.3. Histograms of the values of $\rho(\mathbf{A})$ observed over 100 graphs drawn from each of the three simulation methods.



54.083	estimated number of edges [75]
53.586	estimated number of edges [76]
43.886	estimated number of edges, number of non-isolated nodes, minimum degree 1 [77]
7.779	mean value ($\pm\sigma = 0.133$) of ERGM-A graphs
6.069	mean value ($\pm\sigma = 0.109$) of ERGM-B graphs
5.1623	using degree distribution [80]
4.741	mean value ($\pm\sigma = 0.159$) of Volz algorithm graphs
4.1893	the degree distribution of non-isolated nodes [78]
3.9274	approximation assuming degree distribution and no degree correlations of Eq. 4.2
2.925	the expected degree of non-isolated nodes with nonzero degree [75]
2.4495	largest degree 6 [78]
2	the estimated number of non-isolated nodes [83]

Figure 4.4. 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 × and simulation results by a horizontal line.

the destination cities. This kind of increase is appropriately modeled with mass action mixing (as described in Section 3.1). Thus, an expression for the rate of creation of newly-infected individuals in city j might take the following form:

$$\mathcal{F}_j = \sum_{i=1}^{n} \beta a_{ji} S_j \frac{I_i}{N_i}$$

where S_j is the number of susceptible individuals in city j, N_i is the population of city i, β is a biologically-determined infection parameter and a_{ji} is the number of passengers traveling from city i to city j. The appropriate value for a_{jj}, then, is the population N_j. If we assemble all of these a_{ji} into a matrix A such that $\{A\}_{ij} = a_{ij}$, and assemble all of the city populations into a diagonal matrix N, the matrix F in the computation of R_0 can be represented by

$$F = \beta N A^{\top} N^{-1}.$$

If we make the assumption of "identical biology" discussed in Section 3.3, the quantity of interest in determining R_0 will be $\rho(F)$. Since β is not known, we set it to 1 for the remainder of this analysis of $\rho(F)$; the choice of scale factor will not impact the qualitative results we seek here.

One of the features of disease transmission that we can investigate with this data set is how the inclusion of new routes of traffic changes the value of $\rho(F)$ from a nominal value. Our approach to answering this question is as follows: begin with the highest volume air traffic route in the U.S., and sequentially add additional routes in descending order of traffic until a desired number of cities have been included. If we then fill in the remaining traffic volumes between these cities, we've constructed a subgraph of the larger air transportation network that includes the highest volume routes for a given number of cities. If we adjust the number of cities that we consider, we can observe how $\rho(F)$, and thus R_0 increases. Analytical results of this procedure for subnetworks with five to fifty cities, along with upper and lower bounds, are given in Figures 4.5 and 4.6.

We see that adding new air routes does not dramatically change the value of $\rho(F)$, which appears to level out at roughly 822500 for more than twenty cities. The upper bounds of Figure 4.5 are much tighter than the lower bounds of Figure 4.6, and are certainly of the correct order of magnitude. The bound of [84] only requires knowledge of Δ, the largest row sum of F, which is simply the city with the most incoming traffic (scaled by origin population). This is certainly an easier quantity to estimate than the details of the full traffic pattern.

Importantly, this example also illustrates a critique of the reliance of epidemiologists on R_0. To determine whether this statistic is greater than or equal to 1 for a network of this size, one needs *extremely* precise estimates of the biological parameters in this system. If these parameters are determined experimentally, its possible for the estimated range for R_0 (determined by the error bounds on the experimental estimates and the error bounds on the network structure) to contain 1.

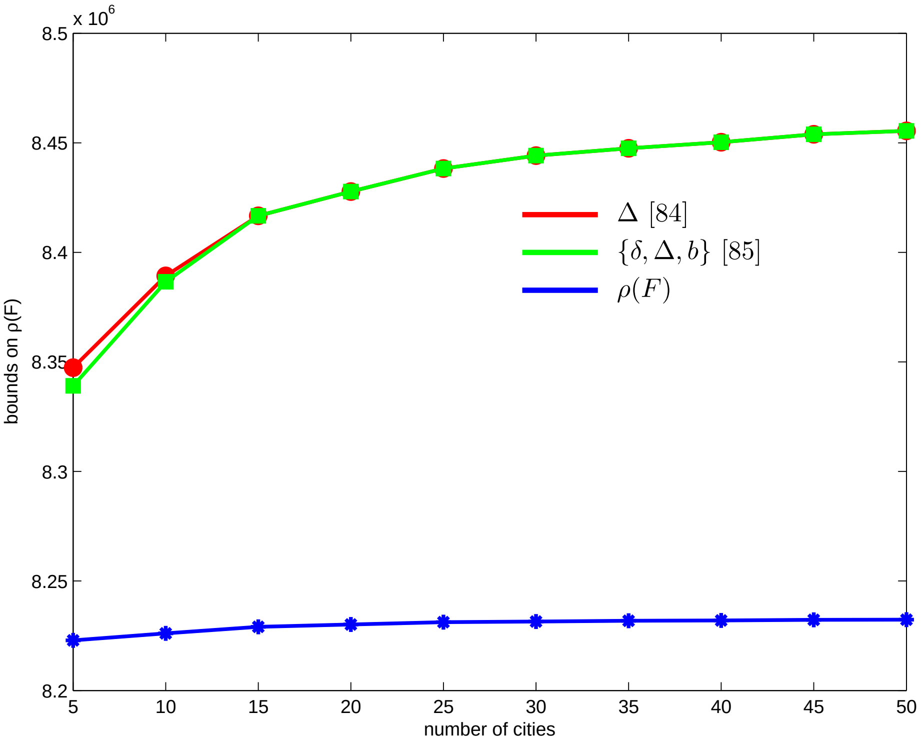


Figure 4.5. Upper bounds on $\rho(F)$ obtained via the results in Table 4.3. The matrix information required for each bound is indicated.

4.2.5 Example 5: Reality Mining proximity data

In 2004, the MIT Media Laboratory sponsored the Reality Mining Project, in which researchers distributed 100 Bluetooth-enabled Nokia 6600 smartphones to members of the MIT community. These phones contained software that, among other things, recorded all instances in which another Bluetooth-enabled device was detected within 5m, including the smartphones carried by other study participants. These proximity detections between study participants establish a time-dependent network of these users' (potential) physical interactions. Proximity data is very useful for predicting the spread of infections like the common cold, which can be transmitted by common handling of the same object (like a doorknob or public computer) or via inhalation of airborne droplets. To sample the interaction patterns of the study participants, we extracted a week's worth of data of this naturally time-varying social network (from a total of nine months of record-keeping). For more information on this data set and our processing techniques, see Appendix B.2.

A fundamentally time-varying network provides another source of uncertainty in modeling. The ability of an infection to become an epidemic might depend on not just where, but *when* the first infection arises. Seasonality has always been acknowledged as an important element in disease spread; as an example, there exists a strong correlation between November passenger volume on U.S. domestic flights and the severity of the annual flu season [97]. Periodicity like this occurs

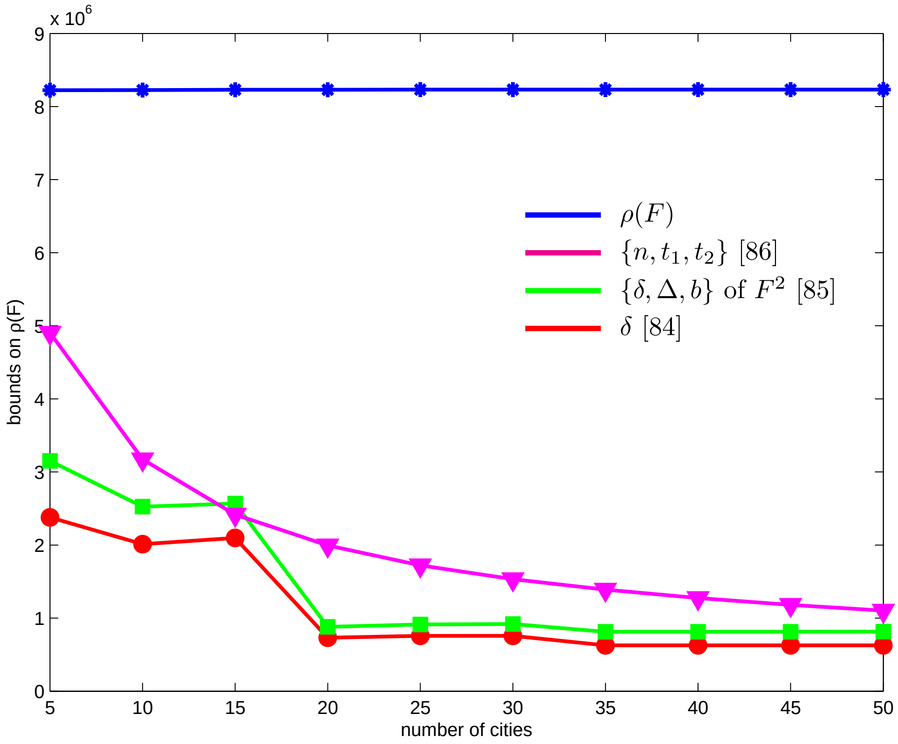


Figure 4.6. Lower bounds on $\rho(F)$ obtained via the results in Table 4.3. The matrix information required for each bound is indicated. For more than 10 cities, the matrix F has zero entries, and thus the lower bound of [85] could not be applied directly; instead, this bound was applied to the matrix F^2 to yield the bound on F.

naturally in the Reality Mining proximity data; if we aggregate the contacts over 12-hour periods as represented in Figures 4.7 and 4.8, we see distinct ‘night’ and ’day’ interaction patterns. Aggregating over smaller time periods reveals additional periodicities.

If one is to approximate a periodically-varying network with a static network for use in an epidemic model, what is the most appropriate time scale over which to aggregate the data? The answer certainly depends on the nature of the infection under study. Consider a simple SIR model, one in which an individual passes permanently into a recovered class after an infectious period. The duration of the infectious period will have some distribution, likely with a characteristic time-scale (e.g., the average infectious period duration). In this case, one should aggregate the network over at least the duration of the infectious period in order to obtain a conservative estimate that includes all possible transmission paths. To assess the effects of choosing a time-scale for aggregation, Figure 4.9 examines the number of participants who made at least one proximity detection with another participant from midnight on November 15, 2004 through the following week, as well as the value of $\rho(A)$ obtained for the network continually aggregated through the week. This figure illustrates that by the end of the day on Monday, individuals have already been in contact with the majority of distinct participants that they will interact with throughout the week. However, the new interactions that continue to accumulate push $\rho(A)$ from ~ 20 on Monday evening up to ~ 30 by Sunday evening.

If we examine these same statistics for the networks achieved by aggregating over 24- and 12-hour periods, a different perspective emerges; these results are depicted in Figures 4.10 and 4.11. The weekday values of $\rho(A)$ are consistently ~ 20, while the weeknight values are considerably smaller. Recall that the basic reproductive ratio is a measure of the initial growth rate of an infection, before saturation effects are seen. If the infectious period is on the order of a day or two, a reasonable proxy for the time-varying network is the snapshot taken over a single weekday. If the infectious period is longer, then the extra contacts made over the course of the week (which cause a 50% increase in $\rho(A)$) become relevant.

Rather than summarizing a naturally time-varying network by a single aggregate network through windowing, it is worthwhile to consider generalizations of R_0 that can accommodate changes in network structure over time. Although we don’t develop the connections here, Appendix C outlines several possible generalizations of the notion of spectral radius that might be interesting to explore in pursuit of a definition of R_0 more appropriate for time-varying networks.

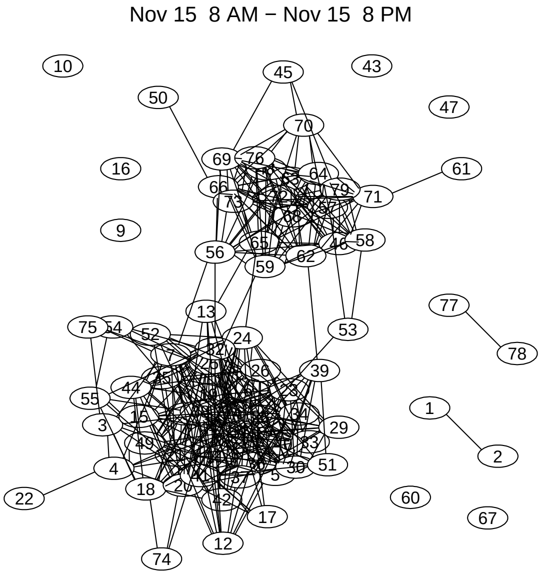

Nov 15 8 AM − Nov 15 8 PM

Figure 4.7. An example of the 'day' proximity network. The two clusters correspond to the two different groups included in the study: Media Lab affiliates and Sloan School of Business affiliates.

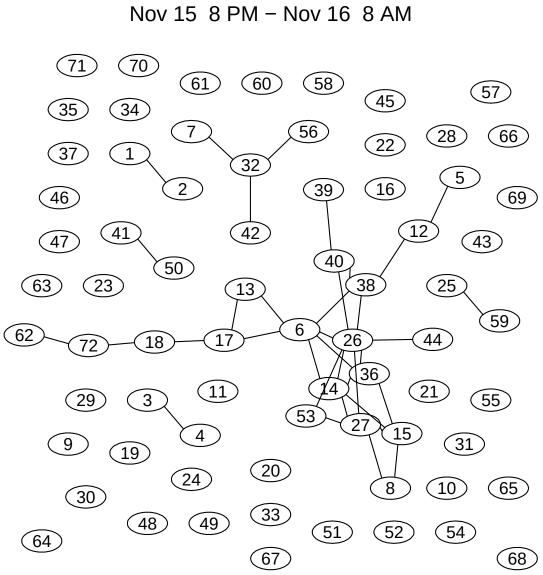

Nov 15 8 PM − Nov 16 8 AM

Figure 4.8. An example of the 'night' proximity network.

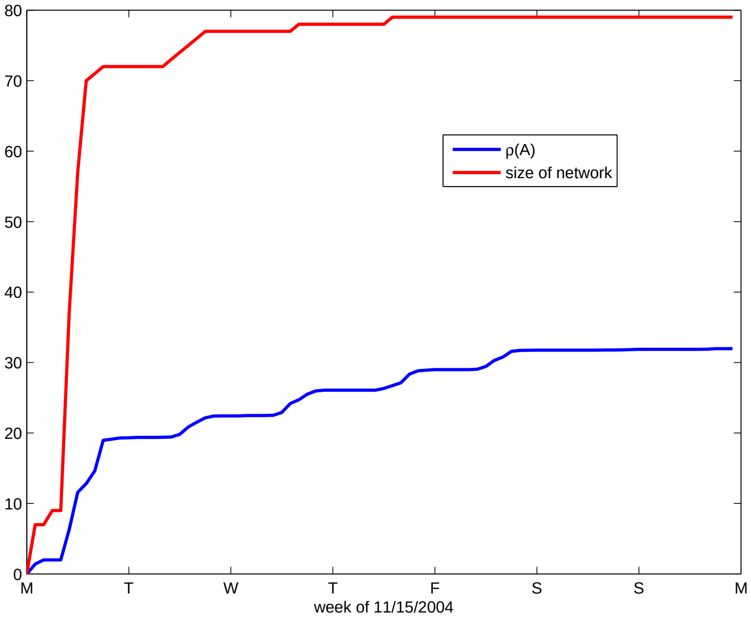


Figure 4.9. The number of individuals and $\rho(A)$ in the proximity network aggregated from midnight on Monday morning through the date indicated by the x-axis.

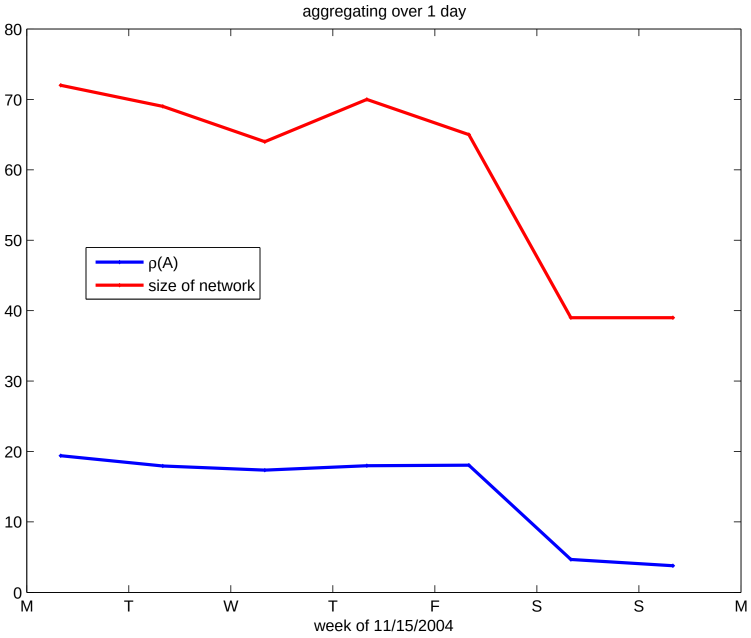


Figure 4.10. The number of individuals and $\rho(A)$ in proximity networks aggregated over 24 hour spans.

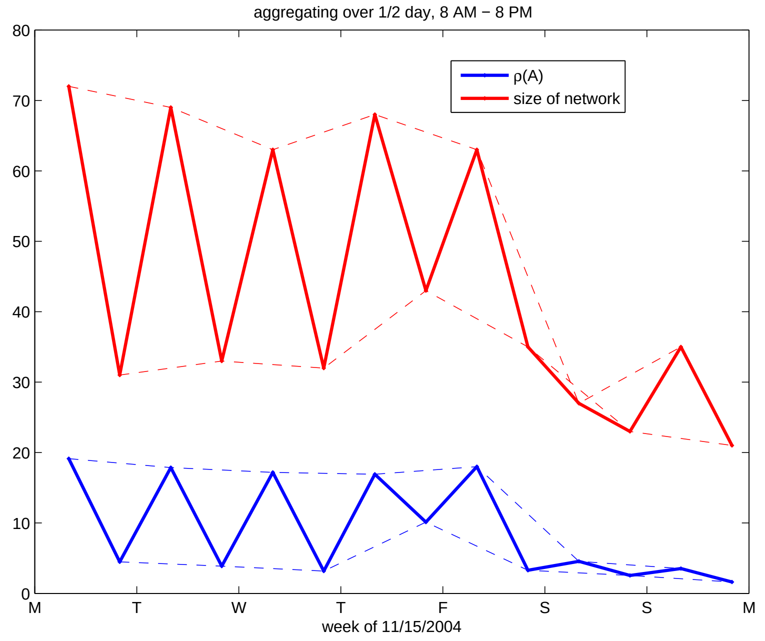


Figure 4.11. The number of individuals and $\rho(A)$ in proximity networks aggregated over 12 hour spans. The 'day' period runs from 8 AM to 8 PM, and the 'night' period runs from 8 PM to 8 AM. The dashed lines in the figure connect day periods to day periods, and night periods to night periods.

Chapter 5

Characterizing the spectral radii of exponential random graph models

THIS chapter begins with an introduction to *exponential random graph models*, a particular family of probability distributions over networks. We will focus on models that are defined by simple structural graph statistics and build intuition regarding the parameters that characterize these models. We'll also present preliminary results to characterize the spectral radii of these models as functions of the parameters, providing a link between static network characterization and the dynamic processes that occur on these networks.

5.1 Probability distributions over graphs

In Chapter 4, we began to think of an uncertain network as a realization of an underlying random ensemble of graphs, and suggested that there are many ways of defining the probabilities over a set of such graphs. Let's begin by considering one particular distribution with a single degree of freedom. The *Erdös-Rényi (ER) random graph* (also called a *Bernoulli random graph* in the sociology and statistics literature) begins with a fixed number n of nodes and considers only the existence of edges [94]. If $\mathcal{A}$ is an ER graph, each undirected edge exists (i.e. $\mathcal{A}_{ij} = 1$) with probability p, independently of the existence of all other edges, and no self-loops are allowed. Formally, these conditions can be written as

$$\Pr(\mathcal{A}_{ij}|\mathcal{A}_{kl} \text{ for all } k \neq i, l \neq j) = P(\mathcal{A}_{ij}),$$

where $P(\mathcal{A}_{ij} = 1) = p$ for $i \neq j$ and $P(\mathcal{A}_{ii} = 1) = 0$ for all i. The independence of edges is very attractive from an analysis standpoint, and much work has been done to characterize the structure of the ensemble of resulting graphs as a function of the parameter p. This model, however, is not especially useful for situations in which edges between nodes *do* have some kind of dependence.

A first step towards relaxing the independence assumption is given by the *Markov random graph*. Again, we begin with a fixed number n of nodes; however, now we assume a *conditional independence* between $\mathcal{A}_{ij}$ and all *non-adjacent* edges:

$$\Pr(\mathcal{A}_{ij}, \mathcal{A}_{kl}|\overline{\mathcal{A}}_{ij,kl}, \{i,j\} \cap \{k,l\} = \emptyset) = \Pr(\mathcal{A}_{ij}|\overline{\mathcal{A}}_{ij,kl})\Pr(\mathcal{A}_{kl}|\overline{\mathcal{A}}_{ij,kl})$$

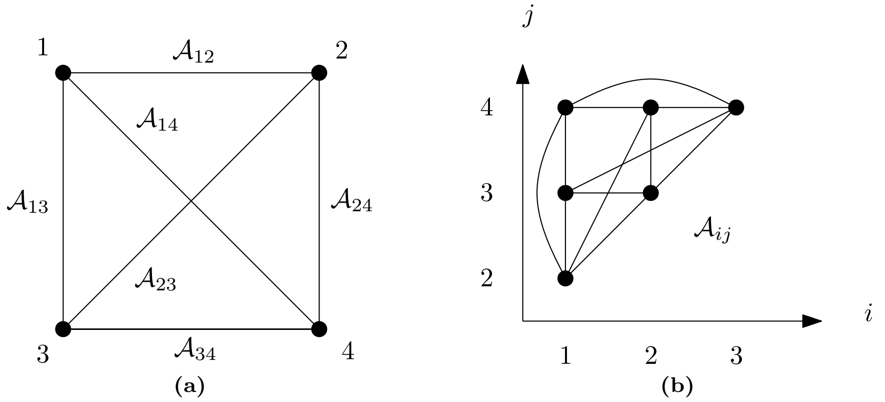


Figure 5.1. A Markov random graph (a) and its associated dependence graph when regarded as a Markov random field (b).

where $\emptyset$ denotes the null set and $\overline{A}_{ij,kl}$ denotes the set of all edges excluding $\mathcal{A}_{ij}$ and $\mathcal{A}_{kl}$ [98]. That is, the dependence of $\mathcal{A}_{ij}$ on the other edges of the graph is limited to those which are adjacent to either node i or node j.

Markov random graphs are a special case of a more general structure called a *Markov random field*. A Markov random field is a collection of variables $\mathcal{V} = \{\mathcal{V}_1, \ldots, \mathcal{V}_m\}$ that serve as the vertices of a *dependence graph* $\mathcal{D}$, which has an edge $\mathcal{D}_{ij}$ connecting $\mathcal{V}_i$ and $\mathcal{V}_j$ if and only if $\mathcal{V}_i$ and $\mathcal{V}_j$ are not conditionally independent, given the state of the rest of the vertices [99]. Translating the Markov random field structure to the special case of our Markov random graph $\mathcal{A}$, the nodes of the associated dependence graph are the edges $\mathcal{A}_{ij}$, with $\mathcal{A}_{ij}$ connected to $\mathcal{A}_{kl}$ if and only if $\{i,j\} \cap \{k,l\} \neq \emptyset$. This is illustrated in Figure 5.1 for a graph on four nodes.

5.2 The Hammersley-Clifford theorem

We introduced the Markov random graph as a more general structure than the Erdös-Rényi random graph, but how analytically tractable is this new structure? The key result underlying Markov random field computations is known as the Hammersley-Clifford theorem[1] and is presented in Theorem 5.2.1.

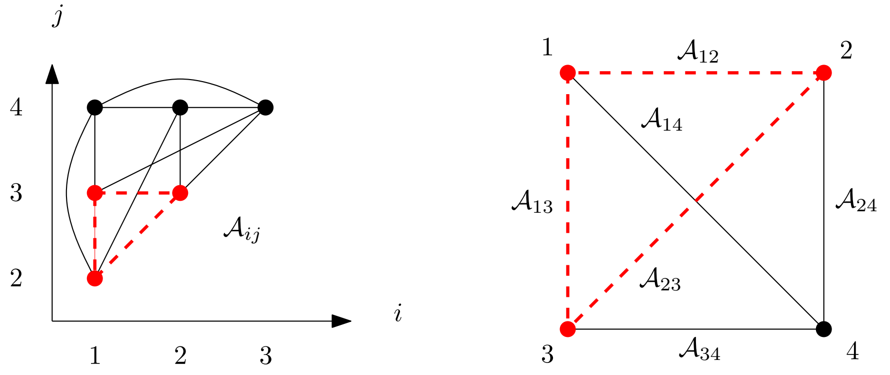
Theorem 5.2.1. *Let $\mathcal{V} = \{\mathcal{V}_1, \ldots, \mathcal{V}_m\}$ be a collection of discrete random variables such that*

- ▷ *for any collection of realized values $v = \{v_1, \ldots, v_m\}$ for which $Pr(\mathcal{V}_i = v_i) > 0$ for every i, $P(v) \equiv Pr(\mathcal{V} = v) = Pr(\mathcal{V}_1 = v_1, \ldots, \mathcal{V}_m = v_m) > 0$, and*
- ▷ *the all-zeros state is possible: $P(0) = Pr(\mathcal{V}_1 = 0, \ldots, \mathcal{V}_m = 0) > 0$.*

Define $Q(v) = \ln\{P(v)/P(0)\}$. Then

$$P(v) = \frac{\exp\{Q(v)\}}{\sum_v \exp\{Q(v)\}}$$

[1] This result was first stated (but not published) by Hammersley and Clifford in the early 1970s, but a more elegant proof was devised by Besag in 1974; the original authors preferred Besag's method, and never published the theorem themselves. Our statement of the theorem is summarized from Besag's work [99].



(a) **(b)**

Figure 5.2. A clique in the dependence graph of Figure 5.1 (a) and its associated triangle in the Markov random graph (b).

and $Q(v)$ can be uniquely expanded into

$$Q(v) = \sum_{1 \le i \le m} v_i G_i(v_i) + \sum \sum_{1 \le i < j \le m} v_i v_j G_{i,j}(v_i, v_j) + \cdots + v_1 v_2 \cdots v_m G_{1,2,\ldots,m}(v_1, v_2, \ldots, v_m)$$

where for any $1 \le i < j < \cdots < s \le m$, the function $G_{i,j,\ldots,s}$ can be arbitrarily chosen to be any nonzero value if and only if the variables $\mathcal{V}_i, \mathcal{V}_j, \ldots, \mathcal{V}_s$ form a clique in the dependence graph; otherwise, $G_{i,j,\ldots,s} = 0$.

Recall that a *clique* is defined as a group of vertices such that every vertex in the group is connected to every other vertex in this group, i.e., the vertices of the clique form a *complete* subgraph. The Hammersley-Clifford theorem states that the probability of any particular realization of $\mathcal{V}_1, \ldots, \mathcal{V}_m$ can be written as a function of the realized values of only the cliques in the dependence graph; the joint realization of an arbitrary collection of $\mathcal{V}_i$ is not necessary.

What is the consequence of this theorem for Markov random graphs? In this case, each random variable $\mathcal{A}_{ij}$ can only take the values 0 and 1, and we have imposed a "nearest-neighbor" dependence assumption. In [98], Frank and Strauss observed that each clique in the dependence graph associated with a Markov random graph corresponds to either a star or a triangle in the Markov random graph itself. Figure 5.2 illustrates this idea for the dependence clique $\{\mathcal{A}_{12}, \mathcal{A}_{13}, \mathcal{A}_{23}\}$, which corresponds to a triangle in the Markov random graph. This observation leads to the following result.

Theorem 5.2.2. *Any undirected Markov graph $\mathcal{A}$ on n nodes has probability*

$$Pr(\mathcal{A} = a) = \frac{1}{\kappa} \exp\left\{ \sum \sum_{k=1}^{n-1} \frac{1}{k!} \sigma_{u_0 u_1 \cdots u_k}(a) + \sum \tau_{uvw}(a) \right\}$$

where

- ▷ *$\sigma_{u_0 u_1 \cdots u_k}(a)$ is nonzero if and only if node u_0 is the center of a k-star connected to nodes $u_1, \ldots, u_k$ in graph a, and*
- ▷ *$\tau_{uvw}(a)$ is nonzero if and only if a triangle connects node u, v and w in a, and*
- ▷ *κ is a normalizing constant.*

If we impose an additional *homogeneity* requirement that any two *isomorphic* graphs should have the same probability, i.e., the labeling of the nodes does not affect the probability of the realization, then this result simplifies even further; now, we need only *count* the number of triangles and each type of star.

Theorem 5.2.3. *Any homogeneous undirected Markov graph $\mathcal{A}$ on n nodes has probability*

$$Pr(\mathcal{A} = a) = \frac{1}{\kappa} \exp \left\{ \sum_{k=1}^{n-1} \theta_k S_k(a) + \tau T_1(a) \right\}$$

where

- ▷ $S_k(a)$ *is the number of k-stars in a, i.e. the number of distinct combinations of a single node and k adjacent edges,*
- ▷ $T_1(a)$ *is the number of triangles in a, and*
- ▷ κ *is a normalizing constant.*

5.3 The exponential random graph family

The exponential random graph family of probability distributions (also called the p^* family or ERGMs) generalizes the analytical structure of Markov random graph probabilities by assuming that the probability of a given graph is an exponential function of a linear combination of relevant graph statistics. Mathematically, this requires that the probability of a graph takes the following form:

$$\Pr(\mathcal{A} = a) \equiv P(a) = \frac{1}{\kappa} \exp \left(\sum_k \theta_k z_k(a) \right) \tag{5.1}$$

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 (which nodes are connected to which nodes), the strength and directionality of these connections, and node identity properties (such as the gender or age of the individual represented by the node). Some of the most commonly used structural properties are:

- ▷ $D_k(a)$ - the number of nodes in a with degree k.
- ▷ $S_k(a)$ - the number of k-stars in a, i.e. the number of distinct combinations of a single node and k adjacent edges.
- ▷ $T_k(a)$ - the number of k-triangles in a, i.e. the number of k distinct triangles that share a common edge.

As pointed out by Anderson et al., this analytical form corresponds to an *autologistic regression model*, one in which the log odds of the probability of a particular network is a linear combination of functions of the variables in the network [100]. One attractive feature of this family of graphs is that they serve as the *entropy-maximizing* distribution given the expected values of the statistics z_k [101]. That is, if $E[z_k] = \mu_k$ for the ERGM family represented by Eq. 5.1, then all other distributions with these same statistics have a smaller entropy $H(\cdot)$, where the entropy of a distribution $P(\cdot)$ is defined as

$$H(P) = -\sum_a P(a) \ln P(a).$$

In a sense, then, the ERGM family over a given set of statistics is maximally general. These models are the foundation of much of quantitative sociology, where they are used to extract information about the processes relevant to the structure of empirically-observed networks. To do this, one begins by generating a list of all possible network statistics that might be relevant to the formation of the network (e.g., *homophily*, the tendency of nodes with similar attributes to connect) or are evident in its structure (e.g., many triangles in a social network). Next, one would like to compute the maximum-likelihood estimate (MLE) of the values of the coefficients θ_k given an observed network. However, evaluating the normalizing factor κ typically requires enumerating all of the possible graphs in the ensemble; this is a prohibitively large number for graphs much larger than thirty nodes. As a consequence, approximations to the ML estimator are often used, most often the maximum pseudolikelihood estimator developed by Strauss and Ikea (discussed in [102]). More recently, a family of approximate MLE methods based on Markov chain Monte Carlo (MCMC) techniques have been developed; see [103] and [94]. Estimates of the θ_k are returned along with confidence intervals, from which a sociologist can identify the most important statistics in the structure of the observed network, then draw conclusions or refine the model and repeat the process.

There are two issues that complicate the practical utility of the p^* family. The first is referred to as *model degeneracy*; for certain combinations of statistics and parameter ranges, the ensemble of graphs has the bulk of the probability density on a very small subset of the total set, often on only the fully-connected graph. An illustration of model degeneracy is given in Figure 5.3, reproduced from [102]. This figure considers a 7-node two-statistic ERGM family that uses both the S_1 and S_2 statistics, and plots the probability that an ERGM with the parameters θ_1 and θ_2 will produce the empty graph (a) and the complete graph (b). One might say that a parameter combination yields a degenerate distribution if the probability of producing the fully-connnected or the fully-disconnected graph is sufficiently close to one; Figure 5.3 demonstrates that degeneracy exists for a wide range of parameter combinations (θ_1, θ_2), and that the transition from a degenerate to a non-degenerate set of parameters is often very abrupt. This observation leads directly to the second problematic issue, *inferential degeneracy*, which occurs when the MLE or pseudo-MLE does not converge to finite values, or when the estimates have very wide confidence intervals corresponding to sensitive dependence on the network data. When we're looking for robust features of the network, non-convergence or sensitive dependence indicate that the model is poorly chosen; see [102] for a detailed treatment of these issues.

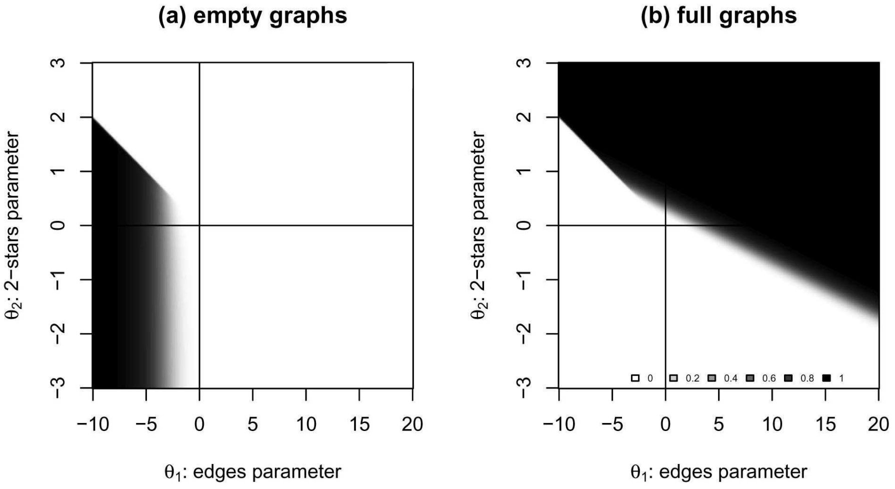


Figure 5.3. The probability of obtaining the empty graph (a) or the complete graph (b) in an ERGM on 7 nodes with S_1 and S_2 parameters. The degree of shading is proportional to the probability. This figure appears as Figure 2 in [102].

Even when degeneracies such as those represented by Figure 5.3 do not occur, the shape of the densities of ERGMs can often lead to undesirable behavior in simulations. For example, there exist parameter ranges for which the distribution is multi-model over different graph densities; an MCMC simulation algorithm might spend millions of iterations in one "regime" before transitioning to another qualitatively different "regime" (examples of this phenomenon are presented in [103]). Alternatively, there may exist a "potential barrier" in the landscape over which the MCMC algorithm operates, such that the algorithm will spend most of its time in one regime before permanently transitioning into another. Burda et al. have explored this phenomenon for ERGMs that use a triangle statistic, $T_1(\cdot)$ [104]. In these cases, the model does not represent a useful ensemble of graphs.

In spite of these complications, we can still derive many useful results from simulations of this family, as long as we remain tuned for the presence of some of these unattractive behaviors.

5.4 Spectra of ERGMs

Much of this thesis explores the ways in which the topology of a network influences the spread of infection among its nodes, and in particular, on the dynamic information conveyed by the largest eigenvalue of the adjacency matrix of the network. Since ERGMs are the dominant network modeling paradigm in the social sciences, considering the spectra of this family is a natural first step in exploring the dynamic processes that occur through these networks. This is an area that has yet to be explored by researchers in sociology, epidemiology or dynamic systems. In the remainder of this chapter, we'll focus on the effects of the inclusion of purely structural statistics and their coefficients on the spectral radii of the adjacency matrices of the resulting undirected networks.

Most analytical results for complex network spectra are achieved asymptotically as the number of nodes goes to infinity. For real social networks, which involve a finite (and often small) number of individuals, an asymptotic analysis is often inappropriate. Therefore, we begin our investigation empirically, by generating many realizations of ERGMs from a fixed distribution and recording the spectral radius $\rho(A)$ of their adjacency matrices A. Our goals for this preliminary analysis are modest: to characterize the mean and variance of $\rho(A)$ as a function of the number of nodes in the network and the coefficients θ_k for the graph statistics included in the model. To generate many draws from an ERGM distribution, we use the *statnet* package for the R programming environment. R is a language designed for statistical analysis, and is freely available via the R Project for Statistical Computing.[2] The *statnet* package was developed by Mark Handcock at the Center for Statistics and the Social Sciences at the University of Washington, and is an excellent tool for simulation and parameter estimation of ERGMs.[3] In particular, *statnet* provides a convenient interface for performing MCMC simulations of a model, using the Metropolis-Hastings update step. The package allows a user to specify many of the MCMC settings, such as the burn-in and sampling intervals, and returns an adjacency matrix that can be exported to a text file for analysis in any software package.

We would also like to make confidence estimates of the means and standard deviations that we observe. Throughout the remainder of this chapter, the error bars at each data point indicate the 95% confidence intervals. For calculations of the mean, these confidence intervals are obtained as an estimate of the mean of a distribution whose variance is unknown; with probability 0.95, the true mean of the distribution is contained within the interval $\bar{\rho} \pm b$ where $\bar{\rho}$ is the sample mean and

$$b = \frac{ts}{N}$$

where

- ▷ N is the number of sample points,
- ▷ t is the value of Student's t-distribution for $N-1$ degrees of freedom at 95% confidence, and
- ▷ s is the unbiased sample standard deviation

$$s = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(\rho_i - \bar{\rho})^2}.$$

[2] `http://www.r-project.org/`

[3] The authors maintain a useful web resource for *statnet* users: `http://csde.washington.edu/statnet/`.

A pragmatic way to construct confidence intervals on the standard deviation of an unknown distribution is to assume the underlying distribution is Gaussian, with unknown mean and variance. In this case, σ can be bounded with 95% confidence by

$$\sqrt{\frac{(N-1)s^2}{\chi_R^2}} < \sigma < \sqrt{\frac{(N-1)s^2}{\chi_L^2}}$$

where χ_R^2 is the value of the chi-squared distribution such that the area to its right is $(1-0.95)/2 = 0.025$ and χ_L^2 is the value of the chi-squared distribution such that the area to its left is $(1+0.95)/2 = 0.975$ [105].

5.5 The S_1 statistic

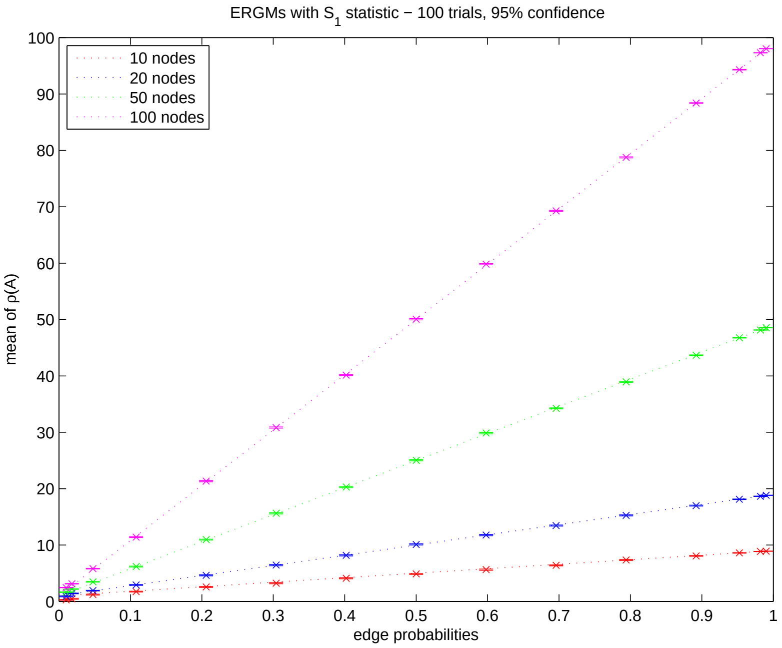
Let's return to the Erdös-Rényi random graph model discussed at the beginning of the chapter, in which each undirected edge exists with probability p, independently of the existence of all other edges. This is certainly a special case of a homogeneous Markov random graph; here, the associated dependence graph has no edges at all. The statistic $S_1(a)$ measures the number of 1-stars (i.e., edges) in graph a, and simple algebra allows us to readily see this case as an example of Theorem 5.2.3:

$$P(a) = p^{S_1(a)}(1-p)^{\left(\frac{n(n-1)}{2} - S_1(a)\right)} = (1-p)^{\frac{n(n-1)}{2}} e^{\ln\left(\frac{p}{1-p}\right)^S_1(a)} = \frac{1}{\kappa} e^{\theta_1 S_1(a)},$$

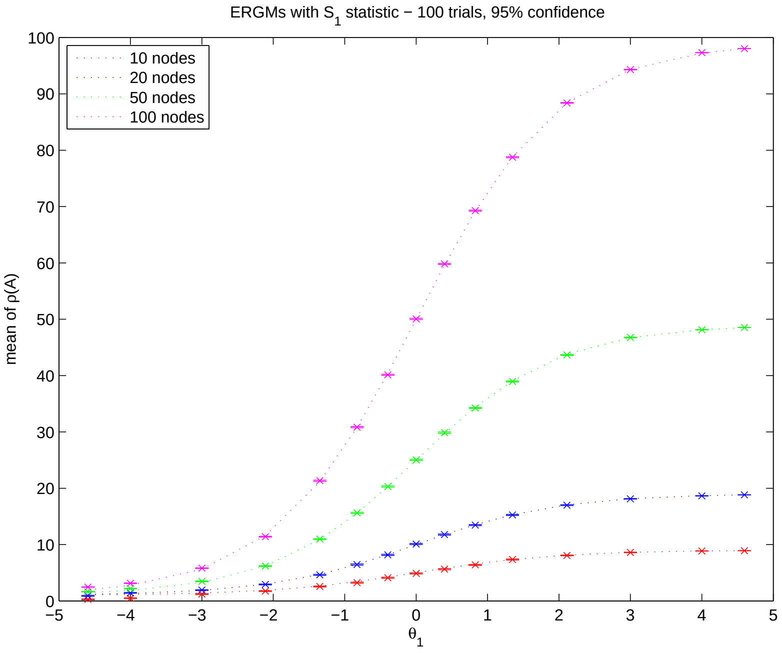
where $\kappa = (1-p)^{-\frac{n(n-1)}{2}}$ and $\theta_1 = \ln \frac{p}{1-p}$. As p varies from 0 to 1, θ_1 varies from $-\infty$ to ∞. For this case, asymptotic analytical results predict that the distribution of the largest eigenvalue will be Gaussian with mean $(n-1)p + (1-p)$ and variance $2p(1-p)$ [106]. These analytical results can be compared to simulation results to validate our approach before attempting more complex ERGMs.

Simulation results are presented in Figures 5.4 and 5.5. Figure 5.4(a) presents the average value $\bar{\rho}(A)$ obtained over 100 trials for the indicated values of n and p, while Figure 5.4(b) plots the same results versus the parameter θ_1. These results align well with the asymptotic prediction. Figure 5.5 presents the estimates of standard deviation obtained via the simulations. That these results (for small values of n) demonstrate the same behavior as the asymptotic predictions provides validation that the simulations are indeed constructing the family of random graphs that we desire.

Additionally, we can correlate our simulation results with the qualitative predictions of Figure 5.3. The ER graph corresponds to the slice of Figure 5.3 taken at $\theta_2 = 0$. The figure indicates two symmetric transitions: a decrease in the probability of the empty graph (in the figure, around $\theta_1 \approx -5$) and an increase in the probability of the complete graph (around $\theta_1 \approx 5$).

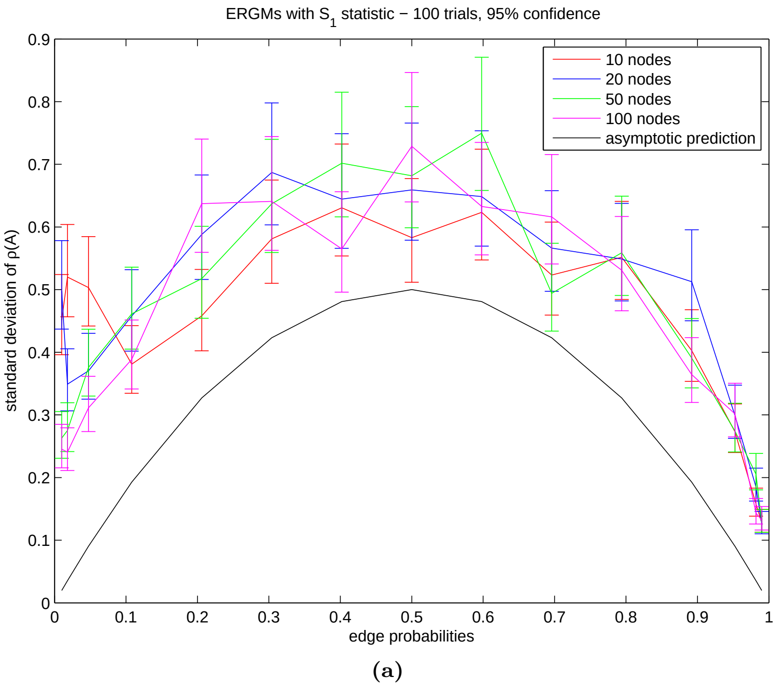


(a)

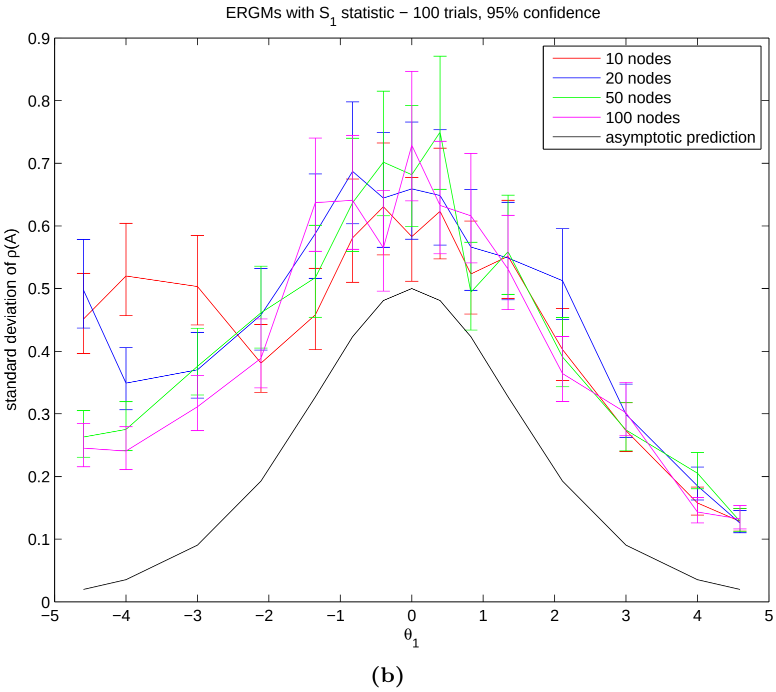


(b)

Figure 5.4. Simulation results using the *statnet* package of the mean of the largest eigenvalue of an ERGM with graph statistic S_1, plotted versus p in (a) and θ_1 in (b). The analytical prediction is given by the dotted lines, which appear to exactly interpolate the experimental data. 95% confidence intervals are indicated by the error bars; in this plot, they are indistinguishable from the data points.



(a)



(b)

Figure 5.5. Simulation results using the *statnet* package of the standard deviation of the largest eigenvalue of an ERGM with graph statistic S_1, plotted versus p in (a) and θ_1 in (b). The asymptotic analytical prediction is given by the solid black line. 95% confidence intervals are indicated by the error bars.

5.6 The S_2 statistic

The S_2 statistic counts the number of 2-stars in the graph, i.e., pairs of edges connected by a central node. In this section, we consider the family of ERGMs parameterized solely by the coefficient θ_2 associated with the S_2 statistic, i.e., the probability of a graph a is given by

$$P(a) \propto \exp\{\theta_2 S_2(a)\}.$$

Let's begin by considering some limiting cases, which we can use as a "sanity check" for our simulations. As $\theta_2 \to \infty$, graphs with large numbers of 2-stars are weighted increasingly heavily. Since adding additional edges can only increase this number, we'd expect the limiting distribution to place the bulk of the probability density on the complete graph, which implies that $\rho(A) \to (n-1)$. As $\theta_2 \to -\infty$, graphs with large numbers of 2-stars are increasingly penalized. In the limit, we'd expect that the only graphs with positive probability will be those with no 2-stars at all. What does a graph in this set look like? A graph with no 2-stars permits node degrees of 0 and 1. Such a graph is a collection of components of size 2 (i.e. connected pairs) and components of size 1 (i.e. isolated nodes). A procedure that is equally likely to generate any such graph is as follows. First, we divide the nodes into sets N_0 and N_1 by assigning each of the n nodes to one or the other with equal probability. The set N_1 will correspond to nodes with degree 1, while N_0 corresponds to singletons. Since each realization of this procedure is equally likely, the mean degree of a node in this assignment scheme is $1\frac{1}{2} + 0\frac{1}{2} = \frac{1}{2}$. Recalling that the mean degree is often a good first approximation to $\rho(A)$, we might hypothesize that as $\theta_2 \to -\infty$, $\rho(A) \to \frac{1}{2}$, independent of n.

Graphs based on the S_2 statistic alone are a special case of those studied analytically by Newman and Park in [93], who applied tools from statistical mechanics to the ERGM family parameterized by both 1- and 2-star statistics (with coefficients θ_1 and θ_2, respectively). Park and Newman present first and second order approximations to the mean degree and the mean squared-degree of the resulting ensemble of graphs as a function of the number of nodes and the parameters θ_1 and θ_2.[4] To begin, define the parameters J and B as

$$J = \frac{1}{2}(n-1)\theta_2,$$

$$B = \frac{1}{2}(\theta_1 - \theta_2)$$

and define ϕ_0 as the solution to

$$\phi_0 = \frac{1}{2}\left(\tanh[2J\phi_0 + B] + 1\right). \tag{5.2}$$

[4] In [93], the authors use θ_i where we've been using $-\theta_i$. The notation in this section is consistent with our usage.

For the remainder of this section, we will continue to use the notation of J, B and ϕ_0, but will additionally assume that $\theta_1 = 0$. With this assumption, Eq. 5.2 has a unique solution for all values of θ_2 except for those in a tiny range around $\theta_2 = 0$, which becomes increasingly small as n increases. In this narrow range, there exist two additional solutions to Eq. 5.2. Since our interest is in qualitative changes in the type of network produced by these ERGMs that are relatively robust to small parameter changes, we will not explore this small intermediate regime further. The behavior of the solution to Eq. 5.2 is depicted in Figure 5.6 for a graph on 20 nodes. As $\theta_2 \to \infty$, $\phi_0 \to 1$ very quickly as θ_2 increases from zero; as $\theta_2 \to -\infty$, $\phi_0 \to 1/(2n-2)$ (much more slowly than for $\theta_2 > 0$).

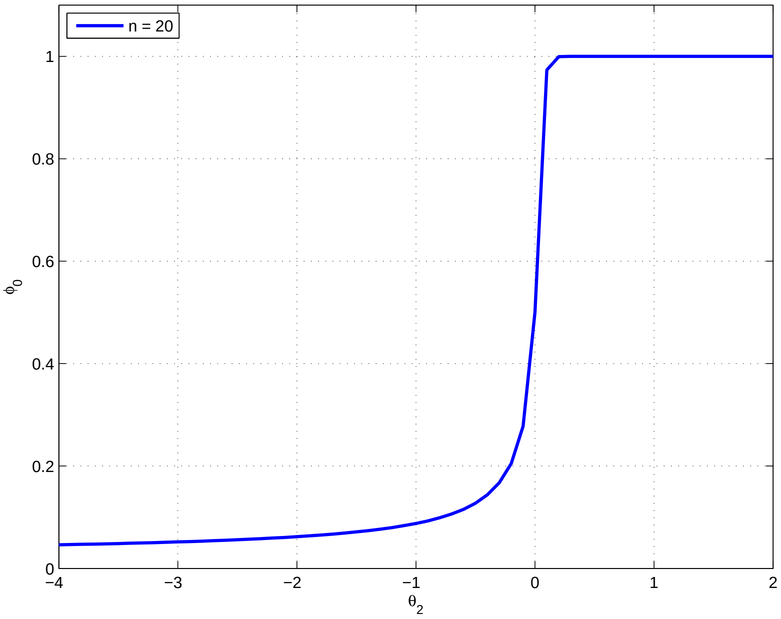


Figure 5.6. Solutions to Eq. 5.2 for n = 20.

From an analytical expression, Park and Newman develop two levels of approximation. First, a *mean-field* approximation is made, which assumes that all nodes are identical with degree equal to the mean degree $\langle k \rangle_1$ (the subscript notation denotes the level of approximation), which is given by

$$\langle k \rangle_1 = (n-1)\phi_0. \tag{5.3}$$

This assumption also implies that the expected value of the squared-degree $\langle k^2 \rangle_1$ is equal to $\langle k \rangle_1^2$. The second level of approximation allows fluctuations in degree about the mean, but assumes no degree correlations, i.e. an edge connected to a node of a given degree is equally likely to have its other end connected to a node of any other degree. Park and Newman then obtain the following results for $\langle k \rangle_2$ and $\langle k^2 \rangle_2$:

$$\langle k \rangle_2 = \langle k \rangle_1 + \frac{2J\phi_0(1-\phi_0)(1-2\phi_0)}{[1-4J\phi_0(1-\phi_0)][1-2J\phi_0(1-\phi_0)]} \tag{5.4}$$

$$\langle k^2 \rangle_2 = \langle k \rangle_1^2 + \frac{(n-1)\phi_0(1-\phi_0)(1-4J\phi_0^2)}{[1-4J\phi_0(1-\phi_0)][1-2J\phi_0(1-\phi_0)]}. \tag{5.5}$$

Because of the rapid transition to a fully-connected graph that occurs when $\theta_2 > 0$, we will focus on comparing the analytical predictions with simulation results for $\theta_2 < 0$. Figure 5.7 compares the mean degree approximations $\langle k \rangle_1$ and $\langle k \rangle_2$ with simulation results for graphs on 50 and 250 nodes, with the average degree taken over 100 trials for each value of θ_2. It is clear that the approximation $\langle k \rangle_2$ is closer to the experimental values than $\langle k \rangle_1$ and that both approximations are better for 250 v. 50 nodes.

Figure 5.8 depicts experimental results for $\langle k^2 \rangle$, taken over 100 trials at the specified values of θ_2 for graphs on 50 and 250 nodes. Again, we see that the analytical approximations hold very well for the larger graph.

Our primary interest, however, is in obtaining expressions for $\rho(A)$, not $\langle k \rangle$ or $\langle k^2 \rangle$; can we use this information to obtain an analytical approximation for $\rho(A)$ as a function of n and θ? Recall that $\langle k \rangle_1$ was obtained by making the mean-field assumption that all nodes are identical with degree $\langle k \rangle_1$. As addressed in Eq. 3.10 of Section 3.3.1, this assumption implies the first approximation

$$\rho_1(A) = \langle k \rangle_1. \tag{5.6}$$

Correspondingly, the second level approximation of Park and Newman is identical to the development of Eq. 3.9, which implies a second approximation for $\rho(A)$:

$$\rho_2(A) = \frac{\langle k^2 \rangle_2}{\langle k \rangle_2}. \tag{5.7}$$

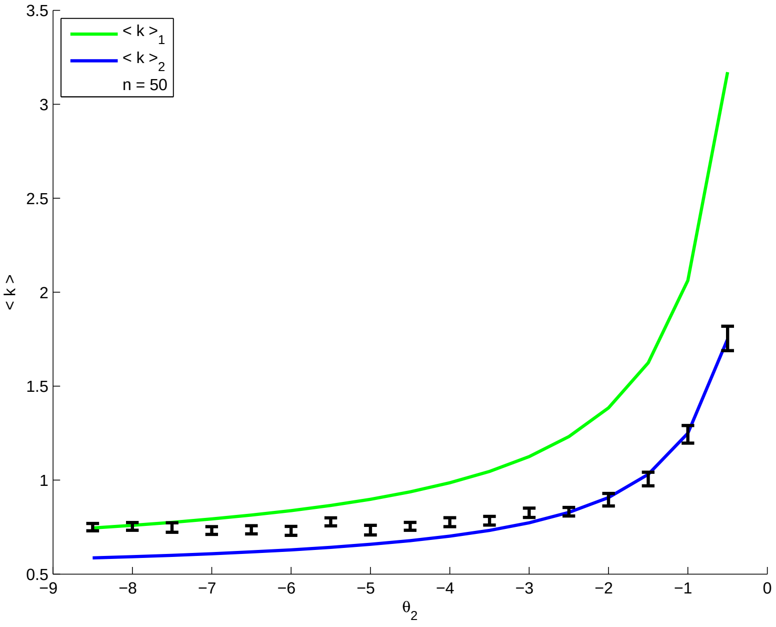
Observe that both of these approximations demonstrate the limiting behavior that we predicted at the start of this section:

▷ as $\theta_2 \to \infty$, $\phi_0 \to 1$, and thus $\rho_1(A) = \langle k \rangle_1 \to n-1$ and

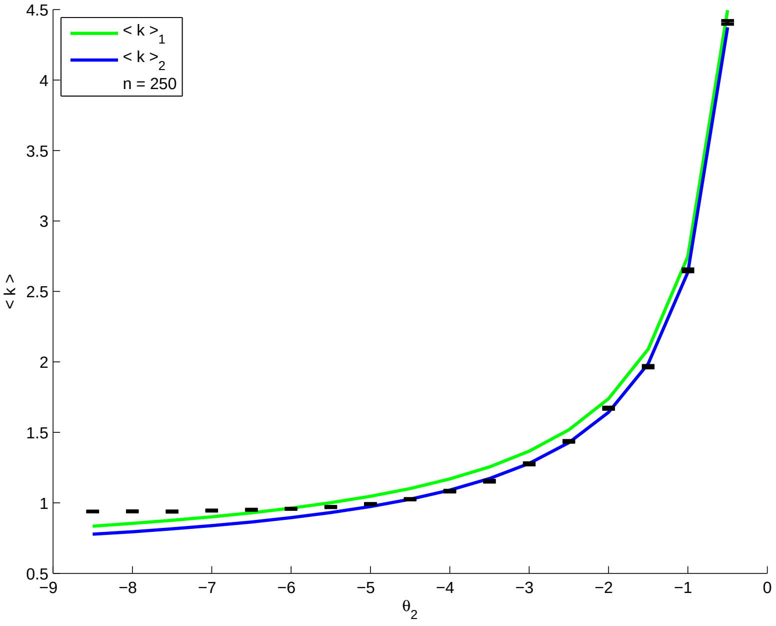
$$\rho_2(A) = \frac{\langle k^2 \rangle_2}{\langle k \rangle_2} \to \frac{\langle k \rangle_1^2}{\langle k \rangle_1} = \langle k \rangle_1 = n-1;$$

▷ as $\theta_2 \to -\infty$, $\phi_0 \to 1/(2n-2)$ and $J \to -\infty$ and thus $\rho_1(A) = \langle k \rangle_1 \to \frac{n-1}{2n-2} = \frac{1}{2}$

$$\rho_2(A) = \frac{\langle k^2 \rangle_2}{\langle k \rangle_2} \to \frac{\langle k \rangle_1^2}{\langle k \rangle_1} = \langle k \rangle_1 = \frac{1}{2}.$$

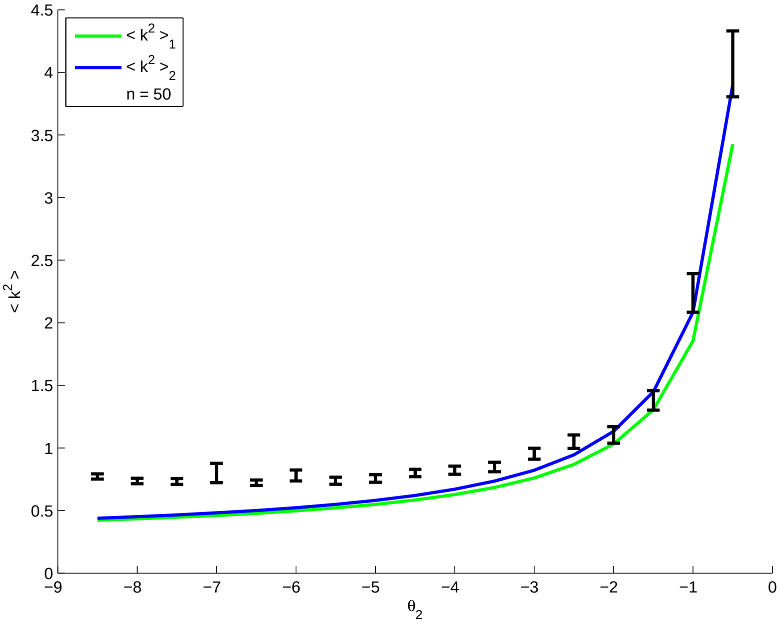


(a)

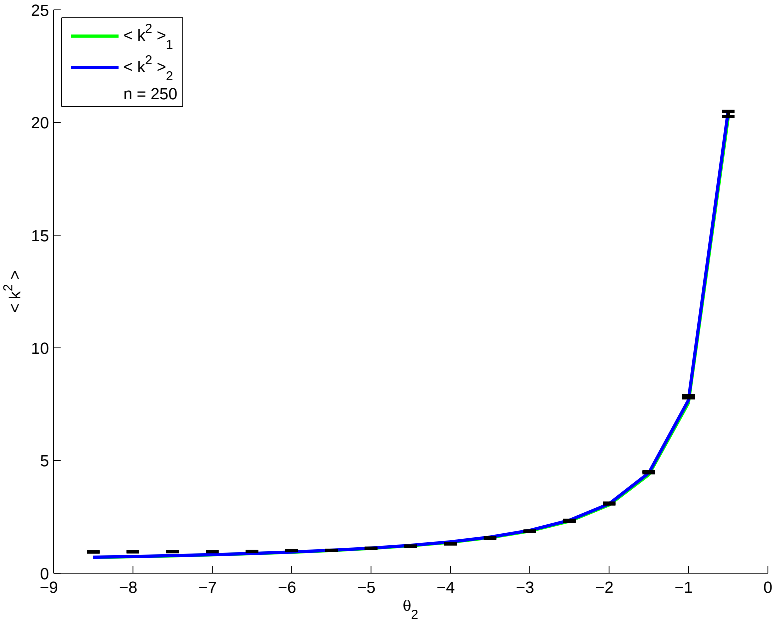


(b)

Figure 5.7. Comparing the analytical approximations of [93] and simulation results on the mean of the degree of nodes with given θ_2 in graphs of (a) 50 and (b) 250 nodes (averaged over 100 trials, with 95% confidence intervals indicated).



(a)



(b)

Figure 5.8. Comparing the analytical approximations of [93] and simulation results on the mean of the squared-degree of nodes with given θ_2 in graphs of (a) 50 and (b) 250 nodes (averaged over 100 trials, with 95% confidence intervals indicated).

Figures 5.9 and 5.10 present experimental results for the mean and standard deviation, respectively, of the largest eigenvalue of graphs realized from the ERGM family with the single parameter S_2. These results confirm the very narrow transition region for $\rho(A)$ as a function of θ_2. For positive values of θ_2, the resulting graphs are all fully-connected, with $\rho(A) = n - 1$. As θ_2 decreases through large negative values, $\rho(A)$ appears to approach a constant, non-zero value for each n. Returning to the degeneracy illustration of Figure 5.3 and examining its predictions for $\theta_1 = 0$, we expect to see one sharp transition in the experimental data corresponding to the sharp transition in the probability of a complete graph in Figure 5.3 (b) (corresponding to the vertical line $\theta_1 = 0$). Figure 5.10 indicates that the widest distribution of $\rho(A)$ occurs when $\theta_2 = 0$, which corresponds to the Erdös-Rényi random graph with edge probability 1/2. Figure 5.11 presents sample draws from the distribution at varying values of θ_2, and Figure 5.12 depicts the degree distributions of sample draws at various values of θ_2.

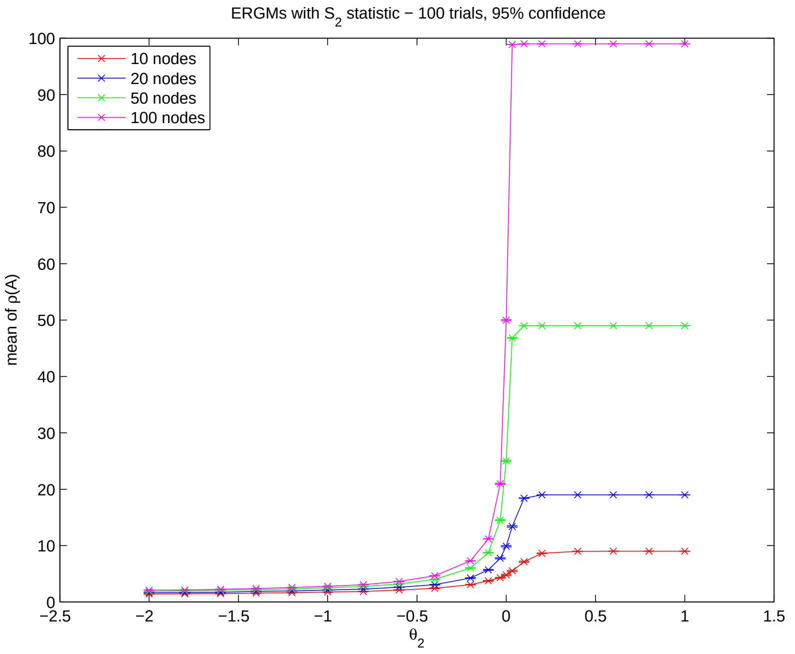


Figure 5.9. Simulation results using the *statnet* package of the mean of the largest eigenvalue of an ERGM with graph statistic S_2, plotted versus θ_2. 95% confidence intervals are indicated by the error bars; in this plot, they are indistinguishable from the data points.

Figure 5.13 compares the approximations ρ_1 and ρ_2 with the simulation results for $\theta_2 < 0$ on the 50- and 250-node graphs. Certainly, the trend is correct, but the approximations fail in precisely the way we should expect: additional heterogeneities beyond the second-order approximation increase the value of $\rho(A)$. The S_2 ERGM specification will result in degree correlations that are not included in the approximations, so they will necessarily underestimate $\rho(A)$.

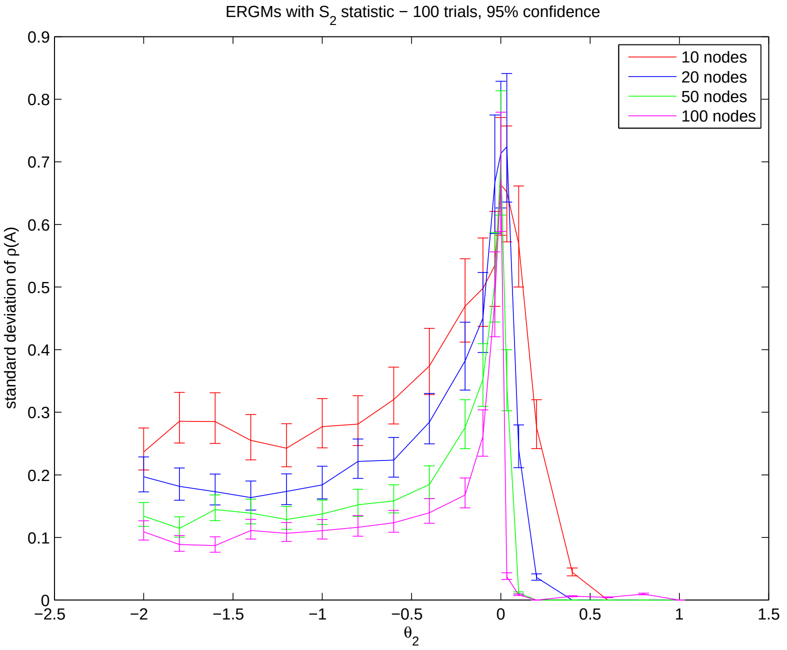


Figure 5.10. Simulation results using the *statnet* package of the standard deviation of the largest eigenvalue of an ERGM with graph statistic S_2, plotted versus θ_2. 95% confidence intervals are indicated by the error bars.

5.7 The T_1 statistic

A third statistic that often appears in ERGM specifications is T_1, the total number of triangles in the graph. Triangles are especially relevant to social network researchers, since they signal the presence of clustering in a network, which occurs when “friends of mine are friends with each other”. Again, let us begin by considering some limiting cases of the following distribution:

$$P(a) \propto \exp\{\tau T_1(a)\}.$$

As $\tau \to \infty$, graphs with large numbers of triangles are increasingly rewarded; as in the S_2 case, the density will center on the complete graph and $\rho(A) \to n-1$. As $\tau \to -\infty$, the only graphs with positive probability will be those without triangles. How can we characterize this set? A first guess is that this is the set of all bipartite graphs, but such a set also unnecessarily excludes graphs with *any* odd-length cycles, not just triangles. This set is, however, a good approximation to the one we desire, and in fact, in the limit of large n, the difference between these sets is a vanishingly small fraction of their size (a result demonstrated by Erdös, Kleitman and Rothschild in [107]). If we'd like a procedure that is equally likely to generate any bipartite graph, we can first divide the n nodes into two sets of size A and B by assigning node x to one or the other set with equal probability. Let n_A and n_B be the number of vertices in each of the sets A and B; these are each binomial random

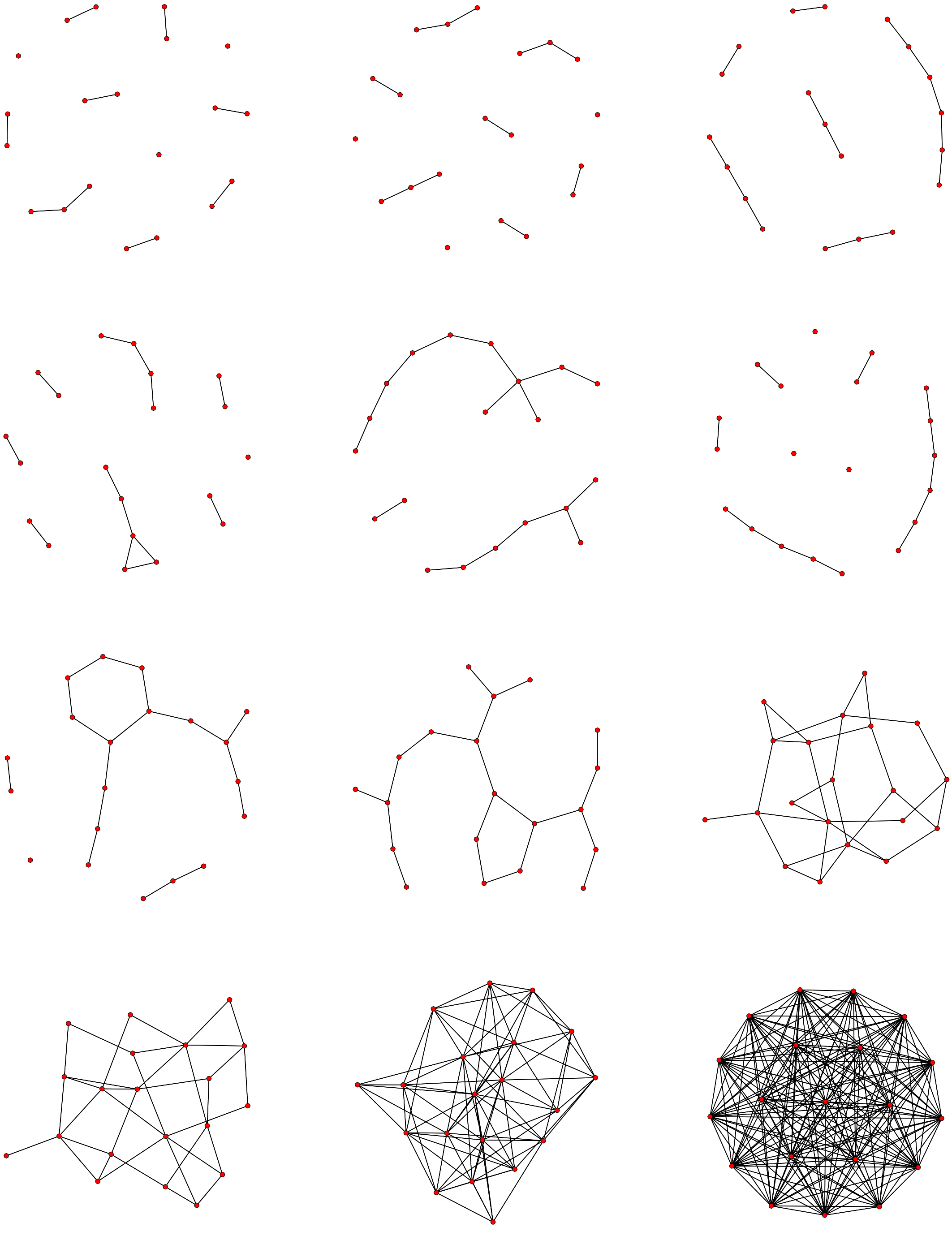


Figure 5.11. From left to right then top to bottom, samples of an ERGM using the S_2 statistic, with θ_2 increasing from -2 to 0.2 in increments of 0.2.

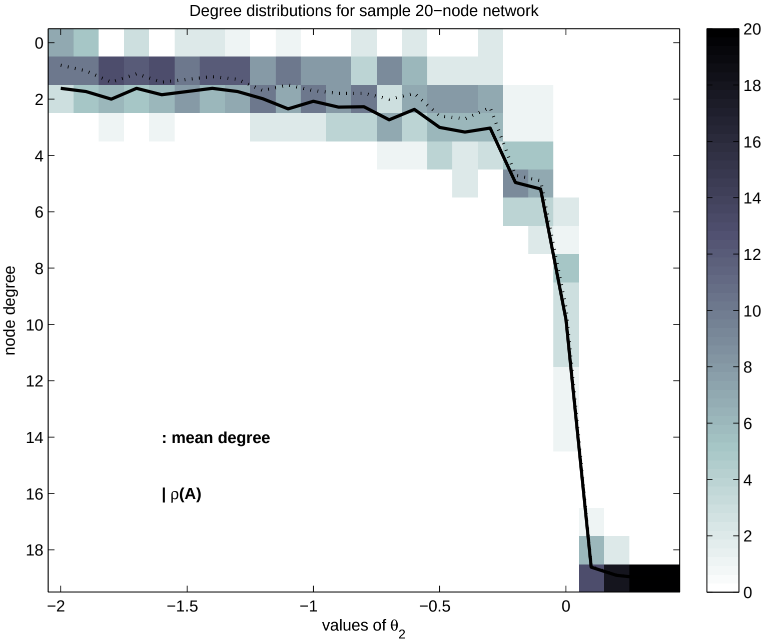
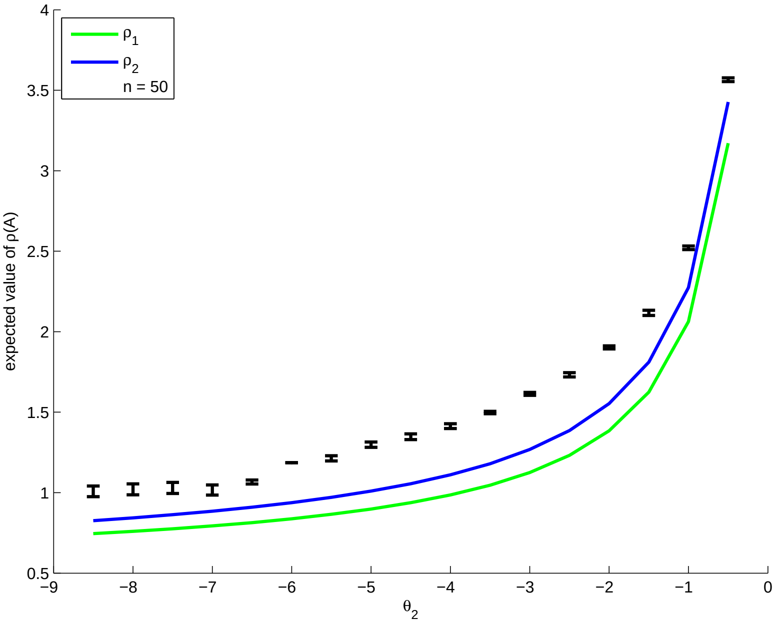


Figure 5.12. A vertical view of the degree distributions of a sample of the ERGM using the S_2 statistic with the corresponding θ_2. The corresponding $\rho(A)$ and the mean degree are also plotted.

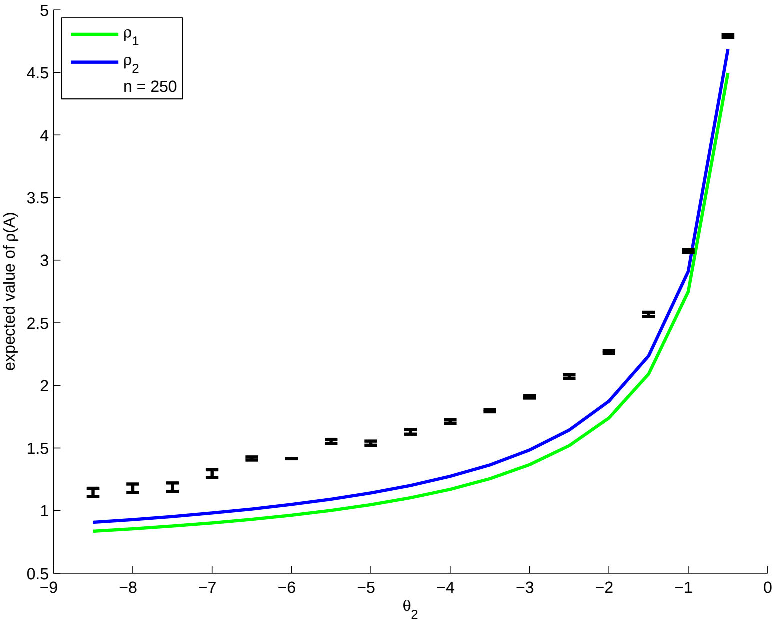
variables on n trials. For every pair of nodes $x \in A$ and $y \in B$, connect x to y with probability 1/2. Conditioned on the value of n_A, the expected degree of a node in set B is $n_A/2$, so the expected degree of a node in set B is $E[n_A/2] = n/4$. The argument for a node in set A is identical, so the expected degree of any node in the graph is $n/4$. Thus, we might anticipate that as $\tau \to -\infty$, $\rho(A) \to n/4$.

To confirm these analytical predictions and explore the behavior of this family of ERGMs in intermediate ranges of τ, we conducted *statnet* simulations in the typical fashion; these results are depicted in Figure 5.14 for each fixed value of n and τ, with a burn-in period of 500,000 iterations and a sampling interval of 50,000 iterations of the MCMC procedure. Our asymptotic predictions are born out (Figure 5.15 demonstrates this for $\tau \to -\infty$), but for larger numbers of nodes, an interesting phenomenon arose: there appears to be a dip in $\rho(A)$ in the interval $\tau \in [-1.5, 0]$ before reaching its asymptotic value as $\tau \to -\infty$ and its known value at $\tau = 0$.

Is this a genuine feature of the distribution or a simulation artifact? To investigate, we allowed the simulations to run for longer burn-in periods and collected a single data point at the end of the burn-in. The results are depicted in Figure 5.16, and seem to suggest that the MCMC algorithm remains in a quasi-stationary state for indefinitely long periods of time (whose duration increases as $\tau \to 0$ from below). This is one of the types of degeneracy discussed in Section 5.3 that often plague ERGM families, and may or may not represent the existence of two separate regions of high

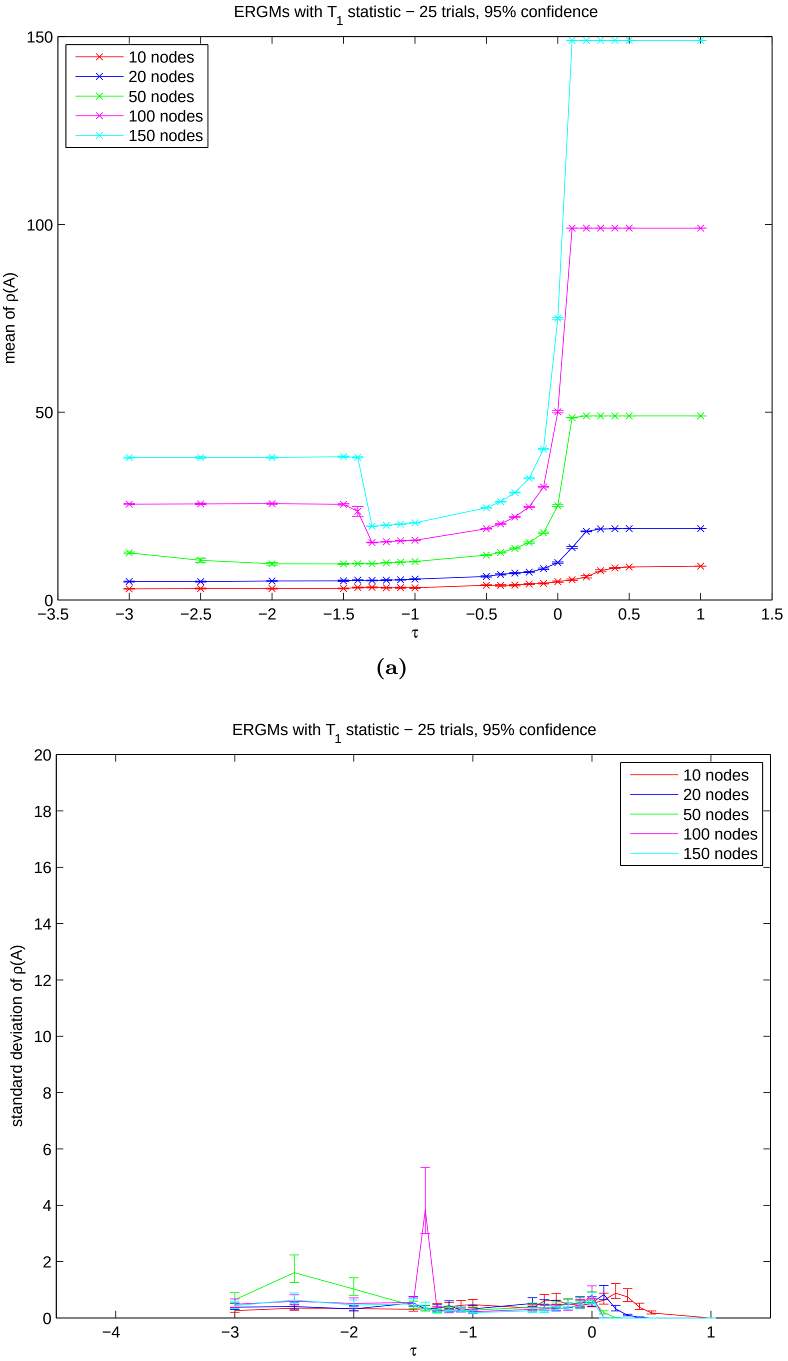


(a)



(b)

Figure 5.13. Comparing the analytical approximations ρ_1 and ρ_2 of $\rho(A)$, as described in Eqs. 5.6-5.7 for graphs of (a) 50 and (b) 250 nodes, with simulation results (95% confidence intervals on the mean are shown).



(a)

(b)

Figure 5.14. Mean (a) and standard deviation (b) of $\rho(A)$ for 25 trials of the ERGM family based on the T_1 statistic for varying values of n and τ.

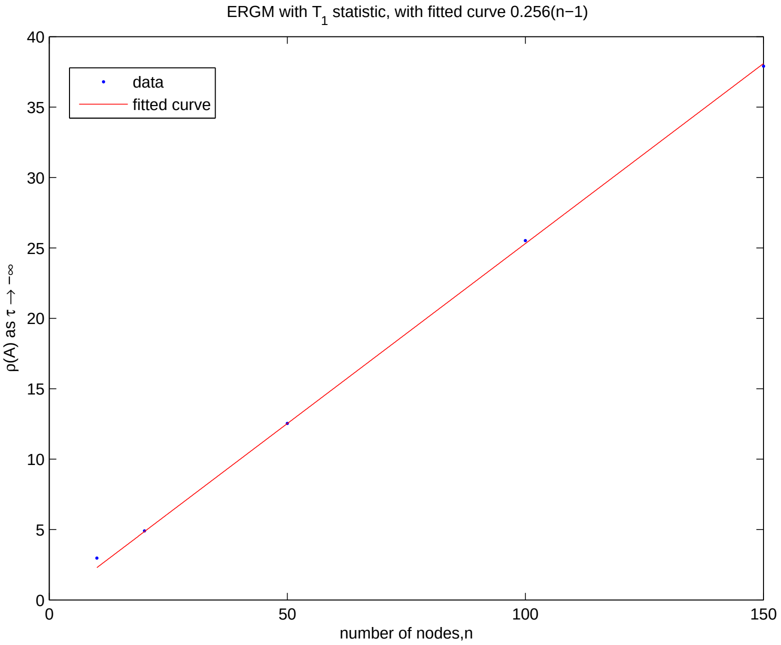


Figure 5.15. A detailed view of the behavior of $\rho(A)$ as $\tau \to -\infty$, confirming our asymptotic prediction of $n/4$.

probability density, i.e. a bimodal distribution. The structure of the graphs generated in these two regions are very different; Figure 5.17 depicts sample degree distributions for a 150 node graph with $\tau = -1.32$ at iterations 5×10^6 and 5.5×10^7.

5.8 The GWD statistic

A relatively new addition to the set of common ERGM statistics is the geometrically-weighted degree (GWD) statistic, defined on a graph a as

$$u(a; d_g) = e^{d_g} \sum_{i=1}^{n-1} \left[1 - \left(1 - d^{-d_g}\right)^i \right] D_i(a)$$

where $D_i(a)$ counts the number of nodes of degree i in graph a and d_g is a fixed parameter. The GWD statistic was described by Hunter in [108] as a more intuitive alternative to the alternating k-star statistic proposed by Snijders et al. in [109]. Both of these statistics involve measuring several structural properties of the graph (like degree distributions) and combining them via the fixed proportions set by the functional form, and both are suggested as statistics that have better convergence and degeneracy properties than standard statistics (like T_1).

This section will considering the family of ERGMs that depend on the GWD statistic. Unlike the single-parameter families that we've discussed so far, this family is parameterized by two values:

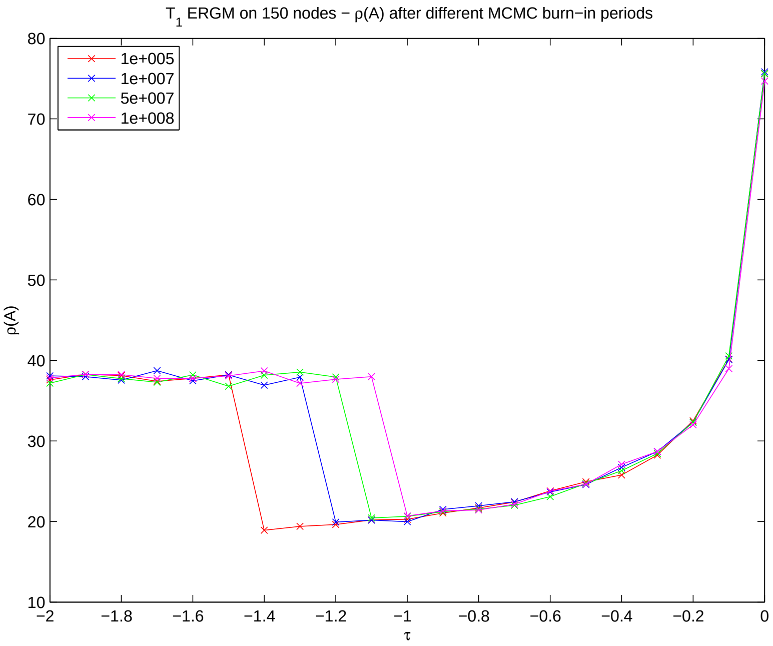


Figure 5.16. Taking a sample of $\rho(A)$ produced by the MCMC procedure at varying burn-in periods.

the coefficient θ_g and the weight parameter d_g, i.e.

$$P(a) \propto \exp\{\theta_g u(a; d_g)\}.$$

As in previous sections, let us consider some limiting cases of these parameters to get a sense of the characteristics of this family of graphs. First, fix θ_g. As $d_g \to \infty$, the term

$$e^{d_g}\left[1-\left(1-d^{-d_g}\right)^i\right] \to i,$$

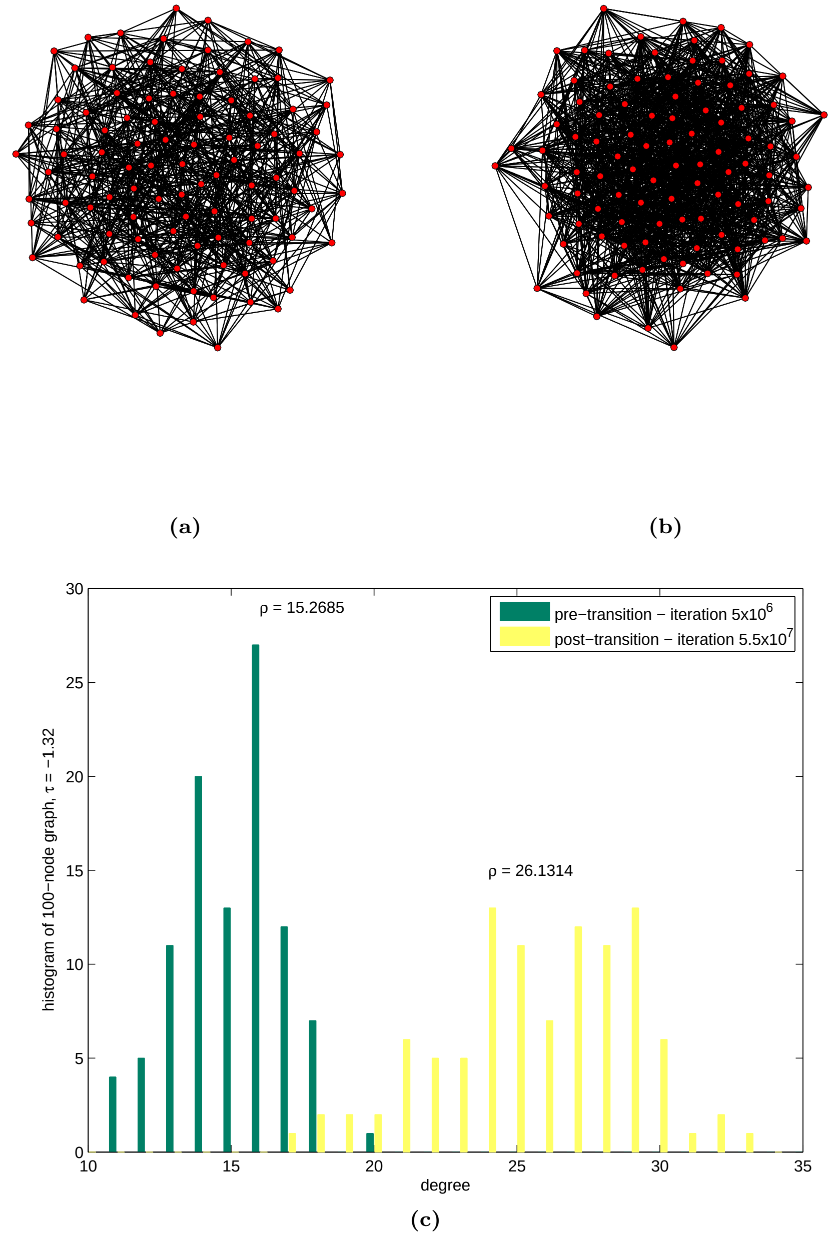
so that

$$\lim_{d_g \to \infty} u(a; d_g) \to \sum_{i=1}^{n-1} iD_i(a) = n\langle k \rangle = 2S_1(a),$$

where $\langle k \rangle$ is the mean degree of nodes in a and $S_1(a)$ is the number of edges. Then

$$P(a) \propto \exp\{2\theta_g S_1(a)\}$$

and the model reduces to the S_1 family described in Section 5.5 with parameter $\theta_1 = 2\theta_g$. In this family, as $\theta_g \to \infty$, we obtain the complete graph and $\rho(A) = n - 1$; as $\theta_g \to -\infty$, we obtain the empty graph and $\rho(A) = 0$.



(a)

(b)

(c)

Figure 5.17. Two graphs from the ERGM family parameterized by the T_1 statistic and generated by *statnet*'s `simulate` function at iterations (a) 5×10^6 and (b) 5.5×10^7, representing samples drawn before and after transition to the higher density state in the MCMC routine. The degree distributions of these graphs (100 nodes with $\tau = -1.32$) are compared in (c).

At $d_g = 0$, the GWD statistic reduces to

$$u(a;0) = \sum_{i=1}^{n-1} D_i(a) = n - s,$$

where s is the number of singleton (isolated) nodes in a. Thus, all graphs on n nodes with the same number of singletons have the same probability. What fraction of the graphs on n nodes have at least one singleton node? The total number of graphs on n nodes is given by

$$2^{\frac{n(n-1)}{2}},$$

while the number of graphs with at least one singleton is

$$n\left(2^{\frac{(n-1)(n-2)}{2}}\right).$$

The ratio of these two quantities is $2n/2^n$, a number that progresses rapidly to zero with increasing n. Thus, the bulk of the graphs with positive probability at $d_g = 0$ are those with no singletons, which each have the same probability

$$P(a) \propto \exp\{\theta_g n\}.$$

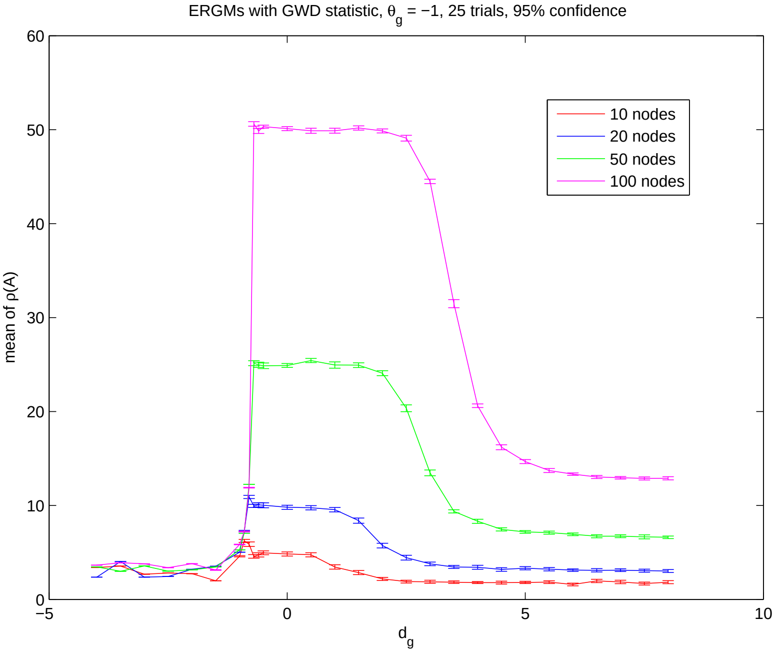
This distribution of graphs is very close to the Erdös-Rényi distribution on n nodes with edge probability 1/2, in the sense that all graphs in the ensemble have roughly the same probaiblity. The difference is in the ensembles, which for the Erdös-Rényi construction includes graphs with isolated vertices. Again, however, these graphs form a vanishingly small fraction of the total number of graphs, so one might reasonably approximate the GWD-induced distribution at $d_g = 0$ with an ER graph with edge probability 1/2, and expect that $\rho(A)$ will be well-approximated by $(n-1)/2$.

What might happen as $d_g \to -\infty$? The quantity $e^{d_g}\left[1 - \left(1 - d^{-d_g}\right)^i\right]$ is large and positive for i odd, and large and negative for i even. Observe that when $d_g = d_g^* = -\ln(2) = -0.6931$, the quantity $1 - \left(1 - d^{-d_g}\right)^i$ is zero when i is even, and thus

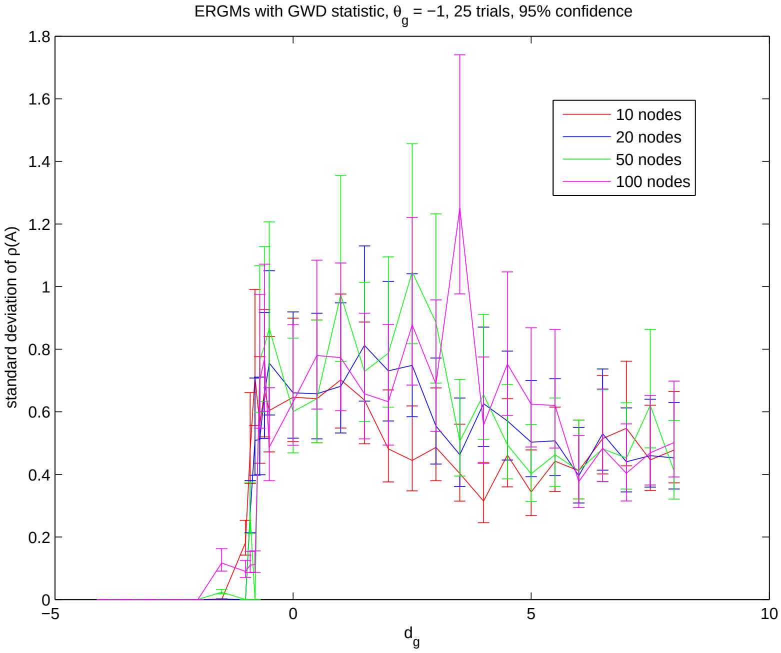
$$u(a; d_g^*) = \sum_{i\ odd} \frac{1}{2} D_i(a) = \frac{n_{odd}}{2}$$

where n_{odd} is the number of odd-degree nodes. For $d_g < d_g^*$, the preference will be for odd-degree nodes when θ_g is positive, and for even-degree nodes when θ_g is negative.

Simulation results are presented in Figures 5.18 and 5.19 for negative and positive values of θ_g, respectively. Observe that our asymptotic predictions are confirmed, and that we indeed see critical behavior at $d_g = d_g^*$.

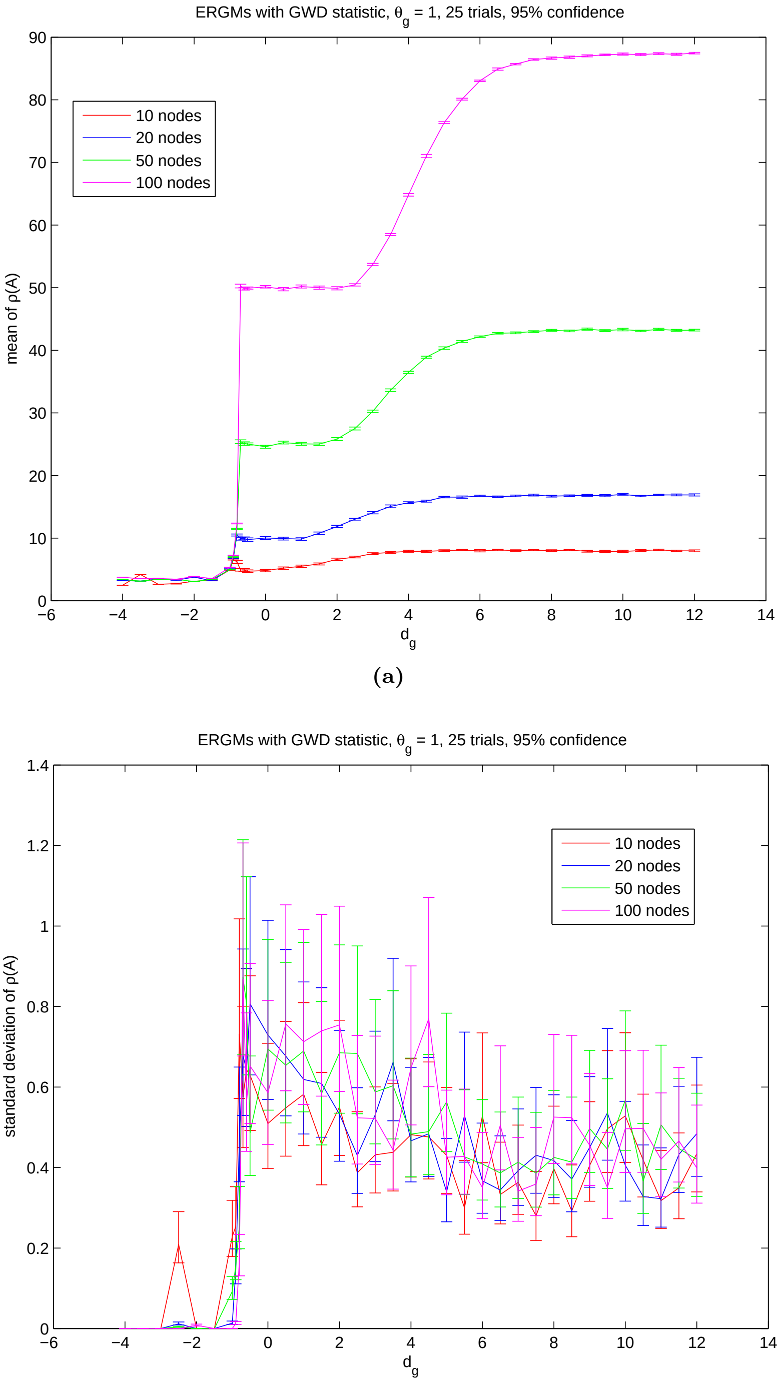


(a)



(b)

Figure 5.18. Simulation results using the *statnet* package of the mean (a) and standard deviation (b) of the largest eigenvalue of an ERGM with the geometrically weighted degree (GWD) statistic. 95% confidence intervals are indicated by the error bars.



(a)

(b)

Figure 5.19. Simulation results using the *statnet* package of the mean (a) and standard deviation (b) of the largest eigenvalue of an ERGM with the geometrically weighted degree (GWD) statistic. 95% confidence intervals are indicated by the error bars.

Figures 5.20 and 5.21 confirm the Erdös-Rényi behavior that we expected for $d_g = 0$ and $d_g = 10$, respectively.

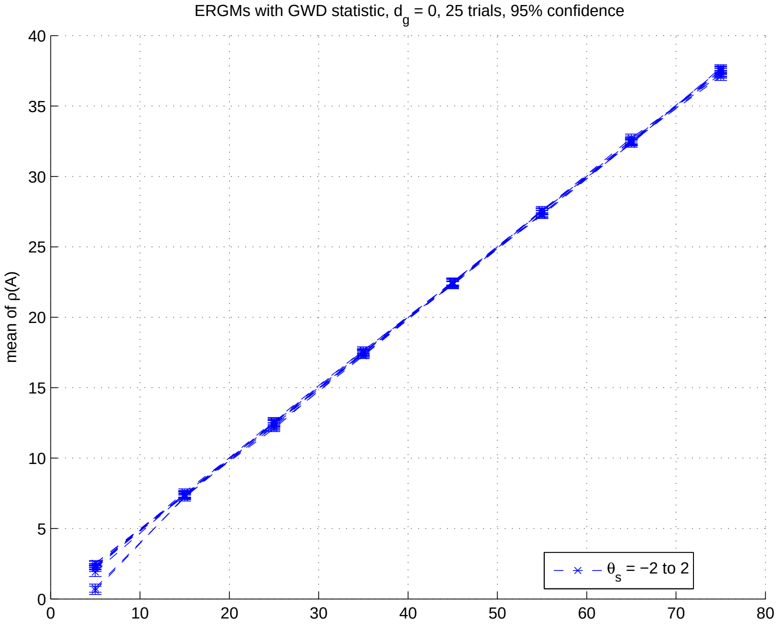


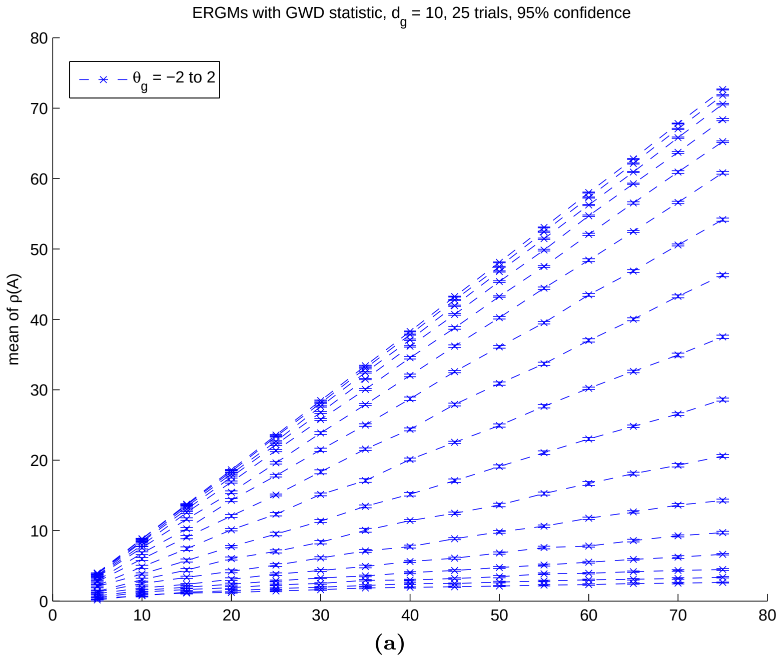
Figure 5.20. Mean values of $\rho(A)$ obtained when $d_g = 0$.

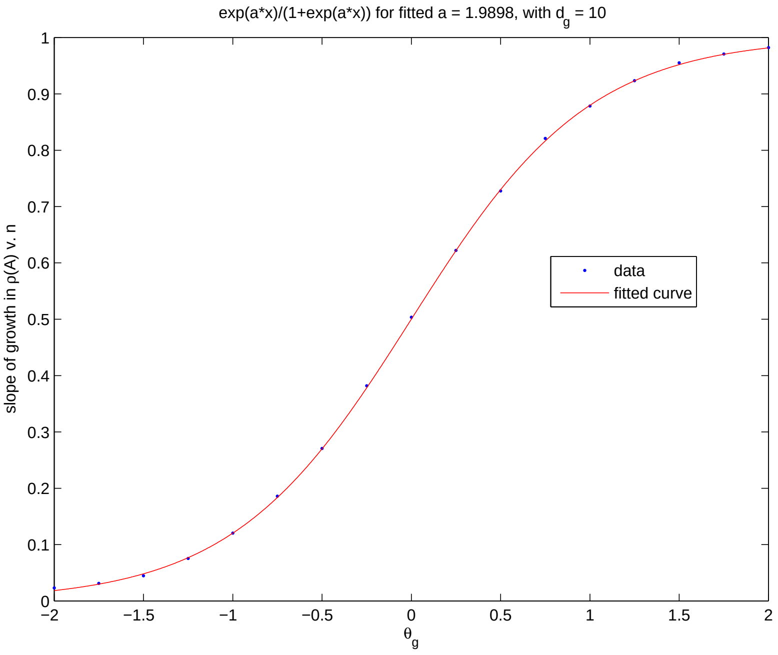
Additionally, Figures 5.22 and 5.23 depict sample degree distributions for random draws from the GWD-based ERGM family. In Figure 5.22, for which $\theta_g > 0$, $d_g \to -\infty$, the black bands at 1, 3 and 5 indicates the preference for odd-degree nodes occurring at $d_g = d_g^*$; similarly, Figure 5.23 shows the preference for even-degree nodes when $\theta_g < 0$ as $d_g \to -\infty$.

We've made many observations of the types of networks that result when using the GWD statistic in an ERGM model; how is the GWD statistic used and interpreted in the social networks community? Hunter et al. demonstrate some of its properties by considering how the probability of a particular graph changes when a single edge is added [108]. Suppose this edge connects two nodes of degree k and l, respectively. Then the ratio of the probabilities of the "after" and "before" graphs is given by

$$\frac{p_{after}}{p_{before}} = \exp\{\theta_g(\phi^k + \phi^l)\}$$

where $\phi = 1 - \exp\{-d_g\}$. As d_g increases from 0 to ∞, ϕ increases from 0 to 1. When $\theta_g > 0$, the terms ϕ^k and ϕ^l then can be interpreted as having an "anti-preferential attachment" effect; the increase in probability that arises from adding an edge decreases with the degree of the nodes to which the edge connects. For $\theta_g < 0$, the preference is for having *fewer* edges. Interestingly, the case of $d_g < 0$ is explicitly avoided in the literature, likely because its consequent even/odd favoring doesn't have a ready sociological interpretation.





(b)

Figure 5.21. Mean values of $\rho(A)$ obtained when $d_g = 10$ (a) and comparing the results with the predicted behavior (b).

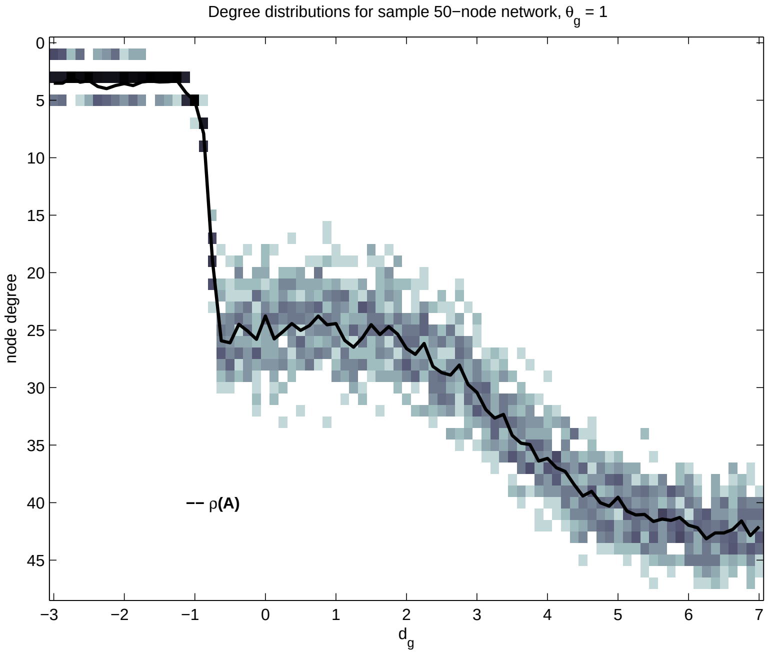


Figure 5.22. Degree distributions of sample GWD graphs at varying values of d_g, for $\theta_g = 1$; the frequency of a given degree is proportional to the intensity of the shading.

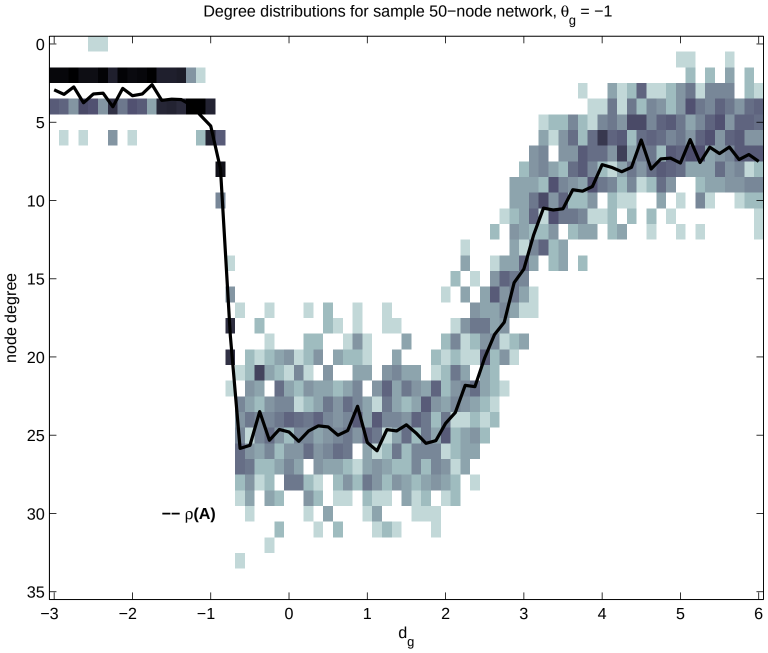


Figure 5.23. Degree distributions of sample GWD graphs at varying values of d_g, for $\theta_g = -1$; the frequency of a given degree is proportional to the intensity of the shading.

5.9 Relevance to public health

Our discussion of the exponential random graph family of probabilistic distributions over networks was motivated by a desire to link static descriptions of topology and the dynamic processes that occur on top of these topologies, and there are ready public health consequences for the kinds of observations we've made in this chapter. For example, consider the following (simplistic) example: in order to contain the spread of infection in a hospital, all patients are isolated from one another, and medical personnel are instructed to only interact with patients and not with each other. Since the individuals in the population can be partitioned into two sets (patients and medical personnel), each of which only interact with members of the other set, the resulting interaction network will be bipartite and contain no triangles. In Section 5.7, we observed that if a network's structure depends only upon minimizing the appearance of transitive relationships (i.e., choosing a very negative value of τ, the parameter associated with the triangle statistic T_1), then $\rho(A)$ can only be decreased to its minimum possible value of $n/4$, where n is the population size. If a further decrease is required to ensure that $R_0 < 1$ (i.e., if the biological factor R_h is larger than $n/4$), another social policy must be put in place or the value of R_h must be driven down by pharmaceutical means. Alternately, one can imagine a cost associated with changing each of the θ parameters that underlies the formation of any given network; knowing the functional dependence of $\rho(A)$ on the values of θ allows one to evaluate the cost of potential policies versus their benefit in reducing R_0.

Chapter 6

Spatiotemporal characteristics of outbreaks

AFTER looking at the relationship between network topology and epidemic thresholds, a natural next step is to explore the patterns of infection propagation through networks, on either side of the epidemic threshold. Classical problems in ecology, including infection spread, have been well-studied as diffusion phenomena in continuous time and space, while relevant results for populations that interact along a network structure have arisen in bursts over the last several decades and from very disparate academic communities. We begin this chapter by considering spatial results for deterministic and stochastic models in turn. Our focus then shifts to the contact process, a continuous-time stochastic model for the spread of SIS infections; we'll explore the behavior of this process on several different topologies to improve our understanding of the mechanisms underlying the extinction behavior. This chapter will conclude with a comparison between the contact process and an analogous influence model for infection spread.

6.1 Special topologies

Before we begin a discussion of spatial phenomena, we will highlight and define several important undirected network topologies that will arise as examples and special cases throughout this chapter.

- ▷ the *complete graph* on n vertices: every vertex is connected to every other vertex (but no self-connections are made).
- ▷ the *hypercube* on $n = 2^m$ vertices: each vertex corresponds to a binary string of length $m = \log_2(n)$; two nodes are adjacent when the Hamming distance between their string representations is 1, i.e., the strings differ in only one entry. This graph is also referred to as the $m - cube$, e.g. the 5-cube has 32 nodes. We will use $\mathbf{0}$ and $\mathbf{1}$ to denote the vertices corresponding to the strings $00 \cdots 0$ and $11 \cdots 1$, respectively.
- ▷ the *star* with n leaves: a graph with $n + 1$ vertices, in which each of the n edges joins to one of the n remaining vertices a common vertex (called the "center" or "hub").
- ▷ the *infinite d-dimensional lattice*, $\mathbb{Z}^d$: each vertex corresponds to a d-tuple $x = (x_1, x_2, \ldots, x_d)$ where $x_i \in \mathbb{Z}$; two vertices x and y are adjacent if their Euclidean distance is 1, i.e. the

entries differ in only one position, and only by ±1. The vertex corresponding to the d-tuple $(0, 0, \ldots, 0)$ will be called the "origin", and will be denoted by $\mathbf{0}$.

- ▷ the *torus*: a graph constructed by excising a square portion of $\mathbb{Z}^2$, then connecting "opposite sides" of the square. The torus can also be thought of as a "window" of $\mathbb{Z}^2$ with periodic boundary conditions.
- ▷ the *infinite homogeneous tree*, $\mathbb{T}_d$: for $d \geq 2$, a tree in which all vertices have degree $d + 1$, except for a single "root" vertex which has degree d. $\mathbb{T}_d$ is also called a *Bethe lattice*, denoted $\mathcal{B}_{d+1}$.

6.2 Deterministic models

This section will highlight several deterministic compartmental models whose results go beyond identifying a threshold to making predictions about the patterns of infection.

Rass and Radcliffe, 2003

In [110], Rass and Radcliffe present an integro-differential equation model of an SIR infection, which is general enough to allow an individual's infectivity (i.e., ability to cause new infections) to vary over the course of the infectious period. More importantly in the context of networks, their formulation includes multiple *types* of individuals who mix heterogeneously, along with the possibility of introducing infection *exogeneously* to a native population at a single point in time. This model assumes a *closed* population, one in which no births or deaths occur. Let $x_i(t)$ denote the proportion of type-i individuals who are susceptible at time t; then

$$\frac{dx_i}{dt}(t) = -x_i(t)\left(\sum_{j=1}^{n} \sigma_j \int_0^t I_j(t,\tau)\lambda_{ij}(\tau)d\tau + \sum_{k=1}^{m} \int_0^\infty \sigma\lambda^*_{ik}(t+\tau)\epsilon_k(\tau)d\tau\right)$$

$$I(t,\tau) = I(t-\tau, 0)$$

where

- ▷ σ_j is the number of type-j individuals and $\sigma = \sum_{j=1}^{n} \sigma_j$ is the total population size,
- ▷ $I_j(t,\tau)$ is the proportion (of σ) of type-j individuals who were infected in the time interval $(t - \tau - d\tau, t - \tau)$,
- ▷ $\lambda_{ij}(\tau)$ is the rate of infection of a type-i susceptible by a type-j infected who was infected τ time units ago (similarly for $\lambda^*_{ik}(\tau)$, with the infections caused by individuals from *exogeneous* type k),[1]

[1] The units of λ_{ij} are the number of contacts that transmit infectious material per unit time per infected individual.

▷ $\epsilon_k(\tau)$ is the proportion (of σ) of outside individuals of type k who are introduced into the population at time 0 and were infected in the time interval $(-\tau - d\tau, -\tau)$ (observe that $\epsilon_k(\tau)$ could be greater than 1).

Additionally, define

$$\epsilon_j = \int_0^\infty \epsilon_j(\tau)d\tau,$$

and similarly,

$$\gamma_{ij}(\tau) = \sigma_j\lambda_{ij}(\tau), \qquad \gamma_{ij} = \int_0^\infty \sigma_j\lambda_{ij}(\tau)d\tau$$

with $\gamma^*_{ij}(\tau)$ and γ^*_{ij} defined analogously. These functions $\gamma_{ij}(\tau)$ and $\gamma^*_{ij}(\tau)$ are required to be bounded with continuous, bounded derivatives. Define a matrix $\mathbf{\Gamma}$ that has as its ijth entry γ_{ij}. Rass and Radcliffe say that an "epidemic" has occurred if the asymptotic size of the infected population is nonzero (i.e., the infection is endemic), and present the following result on the conditions for such a situation and the final epidemic size.

Theorem 6.2.1. Theorem 2.3 of [110]. *Define the fraction of type-i individuals ultimately affected by the infection by*

$$v_i = 1 - \lim_{t\to\infty} x_i(t)$$

and let $\boldsymbol{v}$ denote the vector of these fractions. Denote the vector of ϵ_j by $\boldsymbol{\epsilon}$.

1. *If $\rho(\mathbf{\Gamma}) \leq 1$, $\boldsymbol{v} \to \mathbf{0}$ as $\boldsymbol{\epsilon} \to \mathbf{0}$.*
2. *If $\rho(\mathbf{\Gamma}) > 1$,*

   (i) when $\rho(\mathbf{\Gamma})$ is finite, then $\boldsymbol{v} \geq \boldsymbol{\eta}$ component-wise and $\boldsymbol{v} \to \boldsymbol{\eta}$ as $\epsilon \to 0$, where $\boldsymbol{\eta}$ is the unique positive solution to

   $$-\log(1-\boldsymbol{\eta}) = \mathbf{\Gamma}\boldsymbol{\eta}.$$

   (ii) when $\mathbf{\Gamma}$ has at least one infinite element in each row, $\boldsymbol{v} = \mathbf{1}$.

   (iii) when $\mathbf{\Gamma}$ can be partitioned into

   $$\mathbf{\Gamma} = \begin{bmatrix} \mathbf{\Gamma}_{11} & \mathbf{\Gamma}_{12} \\ \mathbf{\Gamma}_{21} & \mathbf{\Gamma}_{22} \end{bmatrix}$$

   where $\mathbf{\Gamma}_{11}$ and $\mathbf{\Gamma}_{12}$ are finite and $[\mathbf{\Gamma}_{21}\,\mathbf{\Gamma}_{22}]$ has at least one infinite element in each row, partition $\boldsymbol{v}$ and $\boldsymbol{a}$ similarly as $\boldsymbol{v}' = [\boldsymbol{v}_1'\,\boldsymbol{v}_2']$ and $\boldsymbol{a}' = [\boldsymbol{a}_1'\,\boldsymbol{a}_2']$; then $\boldsymbol{v}_2 = \mathbf{1}$ and $\boldsymbol{v}_1 \to \boldsymbol{\eta}$ as $\boldsymbol{\epsilon} \to \mathbf{0}$ where $\boldsymbol{\eta}$ is the unique solution to

   $$-\log(1-\boldsymbol{\eta}) = \mathbf{\Gamma}_{11}\boldsymbol{\eta} + \mathbf{\Gamma}_{12}\mathbf{1}.$$

How do these results compare to the general model discussed in Chapter 2? Recall that the definition of R_0 is a measure of the rate of *initial growth* of an infection, while the results of Theorem 6.2.1 refer to the final size of the epidemic. Additionally, the Rass/Radcliffe model has an uncountably infinite number of infected types, distinguished by their time of infection; our general model requires a countable number of infective compartments. However, at any time t, the only infected compartments whose membership is increasing due to new infections are the $I_j(t,0)$ for $j = 1,\ldots,n$, with the increase occuring at rate $-\frac{dx_j}{dt}$. Additionally, the only way for individuals to transfer into these compartments is via new infection. Thus, we can compute an $n \times n$ next-generation matrix $\boldsymbol{K}$ that counts the number of new infected individuals of each of the n types generated by all infectives of each of the other types. If we are interested in the spread of the epidemic when infection arises in the native population (i.e., no exogeneous infectives are introduced), the next-generation matrix is given by

$$\{\boldsymbol{K}\}_{ij} = \sigma_j \int_0^\infty \lambda_{ij}(\tau) d\tau = \gamma_{ij}$$

and thus $\boldsymbol{K} = \boldsymbol{\Gamma}$. Returning to Theorem 6.2.1, we see that the criteria for local asymptotic stability and the existence of an endemic equilibrium coincide for this class of models.

These results are derived for any type of heterogeneity, and can be interpreted as *spatial* results when applied to the case in which each "type" corresponds to a different node in a network (as was discussed in Chapter 3).

Barthélemy et al., 2004

In a second example, Barthélemy et al. work with an SI model operating on a network and write the following set of differential equations for the infection density at time t for nodes of degree k, assuming that the degrees of adjacent nodes are uncorrelated:

$$\frac{di_k}{dt}(t) = \lambda k[1 - i_k(t)]\theta(t)$$

where $\theta(t)$ is the density of infected neighbors [111]. They linearize this system, then obtain an expression for the time constant τ that governs the initial exponential growth of the total number of infected nodes; this τ is proportional to $\langle k \rangle / \langle k^2 \rangle$ where $\langle \cdot \rangle$ denotes the average value. Additionally, for a given initial condition and pair of degrees $k > k'$, there exists a time t^* such that the number of susceptible nodes of degree k is less than the number of susceptible nodes of degree k' for $t > t^*$; the authors interpret this statement as a prediction that the nodes with the highest degree will be the first infected.

To confirm their analytical results, Barthélemy et al. conduct simulations to confirm this "hierarchical" spread on Barabási-Albert (BA) preferential attachment graphs in which each new node connects to m existing nodes.[2] These results are presented in Figure 6.1, which depicts two statistics. The first is the average degree of newly infected nodes as a function of time. The second is a measure called the *inverse participation ratio*, $Y_2(t)$, which measures the heterogeneity of the degrees of infected nodes and is defined by

$$Y_2(t) = \sum_k \left(\frac{i_k(t)}{i(t)} \right)^2$$

[2] This is a natural generalization of the preferential attachment mechanism described in Section 4.2.2.

where $i(t) = \sum_k i_k(t)$. If the infection is concentrated on a single degree class, then Y_2 achieves its maximum value of 1; Y_2 decreases as infectives are spread more uniformly among the degree classes.

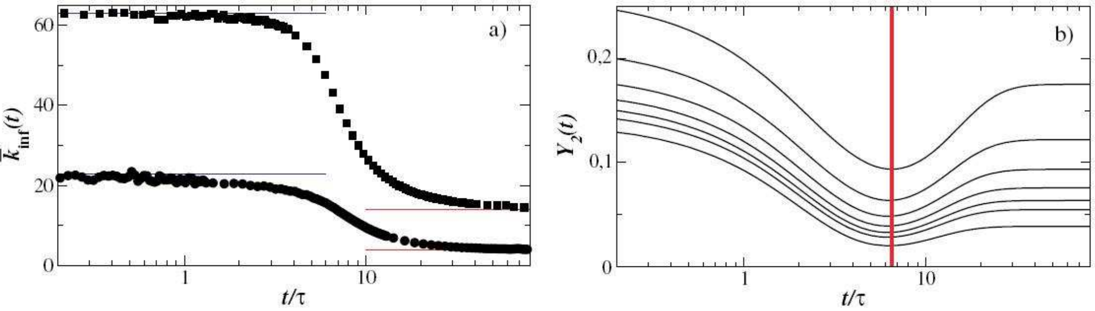


Figure 6.1. Figure 2 from [111]. The caption reads "(a) Time behavior of the average degree of the newly infected nodes for SI outbreaks in BA networks of size $N = 10^4$. Time is rescaled by τ. Reference lines are drawn at the asymptotic values $\langle k^2 \rangle / \langle k \rangle$ for $t << \tau$ and m for $t >> \tau$. The two curves are for $m = 4$ (bottom) and $m = 14$ (top). (b) Inverse participation ratio Y_2 versus time for BA network of size $N = 10^4$ with minimum degree $m = 4, 6, 8, 10, 12, 14$ and 20, from top to bottom. Time is rescaled with τ. The reference line indicates the minimum of Y_2 around $t/\tau \approx 6.5$."

In Figure 6.1(a), we certainly see evidence of a progression of infection from high- to low-degree nodes, but the patterns of outbreak are unclear and certainly depend on the particular structure of BA graphs. For other networks with a power-law degree distribution not generated by the preferential attachment mechanism, will the shape of this curve be different? This question is even more relevant in interpreting the results of Figure 6.1(b); the preferential attachment mechanism necessarily puts the low-degree nodes at the "fringes" of the network, whereas a network of individuals organized into communities, linked by long-distance connections, might not have the same strictly degree-hierarchical spread.

Canright and Engø-Munson, 2006

In a final example, Canright and Engø-Munson examine a discrete-time SI model in which susceptible nodes are infected by each of their infected neighbors independently at each time step with probability $p = 0.05$ [112]. They begin by presenting a heuristic argument for why the *centrality* of a node (as measured by the entries of the dominant eigenvector of the network's adjacency matrix) should be relevant to epidemic spread. They argue that a network can be uniquely decomposed into *regions* by considering the centrality scores as "heights" above the plane of the network, and grouping nodes by identifying the "peak" to which a steepest-ascent algorithm converges (when constrained to move along the edges of the graph). All nodes whose steepest paths converge to the same "peak" are identified as a region.

Additionally, they argue that each of these regions introduces its own S-shaped curve into a plot of the total number of infected nodes v. time; infection enters a region via a node of low centrality, at which point the rate of infection begins to accelerate until a "peak" node is reached. The infection then slows down as it spread through the remaining (lower centrality) nodes in the region. The authors acknowledge that a first approximation to a node's centrality is its degree; their hypothesis, then, refines the observations of Barthélemy et al. in [111].

To test their hypothesis, Canright and Engø-Munson performed simulations on several real networks (snapshots of the Gnutella network, a student social network, and two collaboration networks) and discussed "typical" results (number of infected nodes v. time and average centrality of newly infected nodes v. time). Figure 6.2 presents their results on a collaboration graph of the researchers at the Santa Fe Institute, a graph in which three regions were identified. The three lower curves in (a) represent the number of infected individuals in each of the three regions, and the circles represent the infection times of the "peak" node in each region; in (b), $\mu(EVC)$ is the mean eigenvector centrality of all infected nodes. The authors conclude by discussing several mathematical models whose predictions reduce to the centrality measure under enough simplifying approximations.

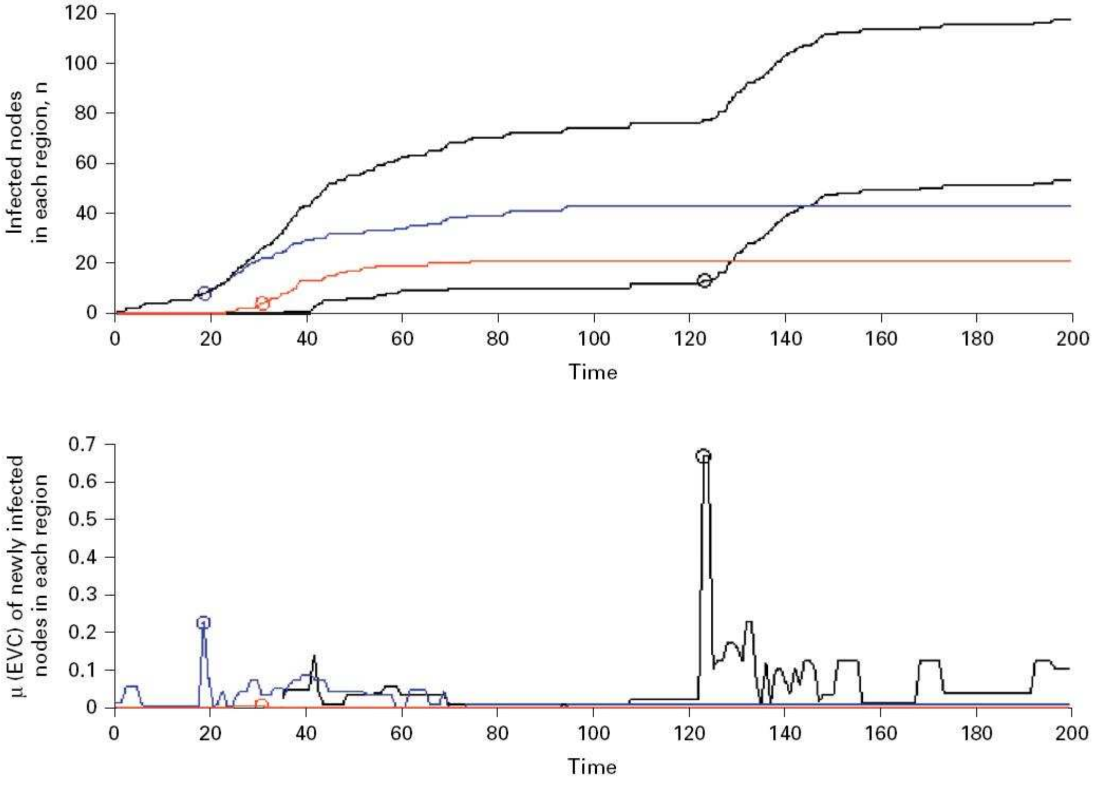


Figure 6.2. Figure 6 from [112].

This argument certainly has intuitive appeal. Returning to the Rass and Radcliffe model described at the beginning of this section, consider the case of $\rho(\mathbf{K}) = 1 + \epsilon$ for some small $\epsilon > 0$. Then the steady-state vector of affected individuals will be small, so one might reasonably invoke

Theorem 6.2.1 to approximate $\boldsymbol{v}$ as

$$\boldsymbol{\Gamma v} = -log(1-\boldsymbol{v}) \approx \boldsymbol{v},$$

which implies that $\boldsymbol{v}$ is close to the dominant eigenvector of $\boldsymbol{\Gamma} = \boldsymbol{K}$. This, however, is not the argument of Canright and Enø-Munson, and not all of the simulations in [112] demonstrate the straightforward relationship between eigenvector centrality and the S-curves as those depicted in Figure 6.2. Part of Canright and Engø-Munson's explanation for why the eigenvector centrality may not work well is that it implicitly allows a node to influence *itself* via closed walks from the node back to itself. If one seeks to quantify the ability of a node i to infect another node j by counting and weighting the numbers of routes between them, one would not want to include any routes from i to j that have an intermediate stop at j; in graph theory terms, one would rather count only *paths*, not all *walks*.

6.3 Stochastic models

We explore in this section some of the literature on stochastic spatial infection spread, addressing three of the major approaches which can be interpreted as SI, SIR and SIS models, respectively. Much of this work originated in the statistical physics community, and its main results concern asymptotic behavior in both time and the size of the systems under study. It is perhaps not surprising, then, that this work (though treating processes highly akin to infection spread) has yet to be fully integrated into the practical mathematical epidemiology toolbox. Some of the more recent results, however, have made analytical predictions on finite networks; we'll see several such examples. Throughout this section, let I_n denote the set of infected nodes at time n in discrete-time (respectively $I(t)$ for continuous-time). If A is the set of infected nodes at time 0, we shall denote the subsequent number of infectives at time n by I_n^A (respectively, $I^A(t)$). Additionally, denote the set of neighbors of node x_i by $\mathcal{N}(x_i)$.

6.3.1 SI

The first stochastic model we'll consider is the SI discrete-time Markov model known as *Richardson's model*, in which each infected neighbor of a susceptible node x_i successfully infects x_i with probability p, independently of all other neighbors [113]. Thus, state transitions occur according to

$$\Pr(x_i \in I_{n+1} | x_i \in I_n) = 1$$

$$\Pr(x_i \notin I_{n+1} | x_i \notin I_n) = (1-p)^{|\mathcal{N}(x_i) \cap I_n|},$$

where $|B|$ denotes the cardinality of set B. Given enough time, every site in the network will become infected with probability 1; is it possible to characterize the likely patterns of spread? The spatial evolution of Richardson's model on the d-dimensional lattice $\mathbb{Z}^d$ has been studied, and in particular, the following "shape theorem" has been established. Let $\mathbf{0}$ denote the node at the origin of the lattice.

Theorem 6.3.1. Theorem 1 of [113]. *There is a convex set D such that for any $\epsilon > 0$*

$$n(1-\epsilon)D \cap \mathbb{Z}^d \subset I_n^{\{\mathbf{0}\}} \subset n(1+\epsilon)D$$

for all n sufficiently large.

Theorem 6.3.1 demonstrates that the asymptotic growth rate of the radius of the infected cluster on the lattice is linear. Additionally, Durrett has demonstrated that for values of p above a certain threshold p_c, the convex set D has "flat edges" in the sense that the intersection of the boundary of $I_n^{\{0\}}$ with the line $\{x|x_1 + x_2 = n\}$ is a non-empty interval; see [114] for details.[3] Some variants of Richardson's model, as well as variants of the SIR and SIS models discussed in the following sections, have also been explored by the pattern recognition community; for an example, see the work of Thompson and Rosenfeld [117]. A continuous-time version of Richardson's model, in which directed edges are "activated" at exponentially-distributed times (with mean 1), has been studied by Fill et al. for the hypercube [118]. Starting with the single infected vertex $\mathbf{0}$, the following upper and lower bounds hold for the time until the entire hypercube is infected.

Theorem 6.3.2. Corollary 6.3 and Theorem 6.4 of [118]. *Consider the hypercube on $n = 2^m$ nodes, and let the infection time along any directed edge be independently realized from an exponential distribution with mean 1. Denote the vertex set by V_m. For any $\epsilon > 0$,*

$$Pr\left(I^{\{\mathbf{0}\}}\left(4\ln(4+2\sqrt{3})+6+\epsilon\right) = V_m\right) \to 1$$

as $m \to \infty$. Also, for any $\epsilon > 0$,

$$Pr\left(I^{\{\mathbf{0}\}}\left(\ln(2+\sqrt{5})/2+\ln 2-\epsilon\right) = V_m\right) \to 0$$

as $m \to \infty$.

6.3.2 SIR

Next, consider a simple continuous-time Markov process model for an SIR infection. The instantaneous rate of transition for a susceptible node x_i to the infected state is given by $\lambda|I(t) \cap \mathcal{N}(x_i)|$; once x_i is infected, it spends a random amount of time (realized independently per the distribution F) in the infected state before permanently transitioning to a recovered state, at which point it no longer participates in infection propagation.

[3] Richardson's model has also been used to study the spatial dynamics of competition between two exclusive species. In this case, a node is either empty or occupied by one of the two species, each of which has its own infection probability. For results of this model, see the work of Deijfen et al., [115] and [116].

What kinds of questions have typically been asked about this process? Most work has built upon the connection to a classical model in statistical physics, *bond percolation*, in which each edge in a network is independently "open" with probability p or "closed" with probability $1-p$. One is then interested in the characteristics of the subgraphs induced by the open edges, e.g., in the induced graph, how large is the component containing a given node? What is the probability of an open path existing between two given nodes, or a given node and set of nodes?

To take advantage of bond percolation theory in analyzing the SIR model, we can construct a unique mapping between the steady-state behavior of the infection process and the bond percolation formulation. Consider an equivalent characterization of the process: each infected node emits a "germ" at rate λ to each of its neighbors. In order for another node to be infected along a given edge, a germ must be transmitted along that edge before the infected node recovers (and ceases to emit germs). For any t, the probability that a germ is transmitted within t time units is simply $1-e^{-\lambda t}$. Then each edge in the network will successfully transmit infection (given the opportunity) independently with probability

$$p = \int_0^\infty (1-e^{-\lambda t})F(t)dt.$$

Thus, the steady-state behavior of the infection process can be analyzed by looking at the equivalent bond percolation model with probability p.

Most analytical results on bond percolation focus on graphs with an infinite number of nodes and with p close to a critical probability p_c; when $p > p_c$, there exists an infinitely large connected component of the graph induced by the "open" edges.[4] In the infection process, we might define p_c (equivalently, λ_c) as

$$p_c = \inf\left\{p \mid \Pr(|I^{\{\mathbf{0}\}}(\infty)| = \infty) > 0\right\}.$$

On the infinite 2-D lattice, $p_c = 1/2$, and for this case, Cox and Durrett used the bond percolation equivalence to develop the following shape theorem for the continuous-time model.

Theorem 6.3.3. Theorem 1 of [119]. *Assume that the second moment of the distribution of F is finite and that $\lambda > \lambda_c$ where λ_c is a critical rate derived from p_c. Let $I^{\{\mathbf{0}\}}(\infty)$ denote the set of sites that will ever become infected when only the origin is initially infected, $R_t^{\{\mathbf{0}\}}$ denote the set of recovered sites at time t, and $I_t^{\{\mathbf{0}\}}$ denote the set of infected sites at time t. Then there is a convex set D such that for any $\epsilon > 0$,*

$$Pr\left(I^{\{\mathbf{0}\}}(\infty) \cap t(1-\epsilon)D \subset R_t^{\{\mathbf{0}\}} \subset t(1+\epsilon)D\right) = 1$$

and

$$Pr\left(I_t^{\{\mathbf{0}\}} \subset t(1+\epsilon)D - t(1-\epsilon)D\right) = 1$$

for all sufficiently large t.

[4] The exact definition of p_c depends upon the phenomenon of interest, and is also defined differently when finite graphs are considered.

This theorem tells us that the diameter of the set of recovered individuals grows from the origin linearly in time, and that the "front" of active infection is a convex curve that follows the boundary of the convex set D (the set difference $t(1+\epsilon)D - t(1-\epsilon)D$). In another example, Braga et al. have studied bond percolation on the infinite homogeneous tree, $\mathbb{T}_d$; they find that $p_c = 1/d$ [120]. Fill and Pemantle present the following result of the probability of directed percolation on the hypercube.

Theorem 6.3.4. Theorem 3.2 of [118]. *Let each directed edge of the hypercube on $n = 2^m$ nodes be independently open with probability $p = c/m$. Then $Pr(\mathbf{0}$ is connected to $\mathbf{1}$ by an oriented open path $)$ converges to a limit as $m \to \infty$. The limit is 0 if $c < e$ and is $(1 - x(c))^2$ if $c \geq e$, where x(c) is the solution in $(0, 1)$ to $x = e^{c(x-1)}$.*

There are fewer results on finite deterministic graphs, like subsets of $\mathbb{Z}^d$. Sander et al. begin with a similar model in discrete-time; the infection probability between every pair of nodes is drawn independently and identically from an arbitrary distribution, and recovery occurs after a fixed interval [121]; this model can also be mapped to simple bond percolation, and the authors perform simulations to determine under what conditions an "epidemic" will occur, defined as reaching the edge of the 200 × 200 lattice with an initial infective at the origin. They also explore a different statistic; the length of the path that the infection took to each ultimately-infected node. The idea behind tracking this quantity is that it can be readily compared to the *phylogenetic distance* between two infectious microbes. Assuming that there is a correlation between the number of genetic mutations an infection has undergone and the number of hosts through which it has passed, one can extract information about the patterns of transmission from a biological analysis of active strains. Some simulation results from [121] are given in Figure 6.3, with the distribution of infection probabilities X given by

$$f_X(x) = \frac{1}{15x}, \qquad e^{-15} \leq x \leq 1.$$

Borgs et al. take an analytical approach to percolation on a finite window of the 2-D lattice $\mathbb{Z}^2$, and explore how the size of the largest component in the graph induced by the open edges within that window scales with the size of the window as a function of the edge probability [122]. Let W_N be the size of the largest connected component in the induced graph in a window of $\mathbb{Z}^2$ centered at the origin with side length N. Then with probability one,

$$W_N \asymp \begin{cases} \log N & p < p_c \\ N^{2-(1/\rho)} & p = p_c \\ N^2 & p > p_c \end{cases}$$

where $f(p) \asymp g(p)$ means that there exist positive constants c_1 and c_2 such that $c_1 g(N) \leq f(N) \leq c_2 g(N)$, and where ρ is a constant. Moreover, there exists a range of p around p_c such that within this range, all of the clusters scale as $N^{2-1/\rho}$; above this range, there is one dominant cluster.

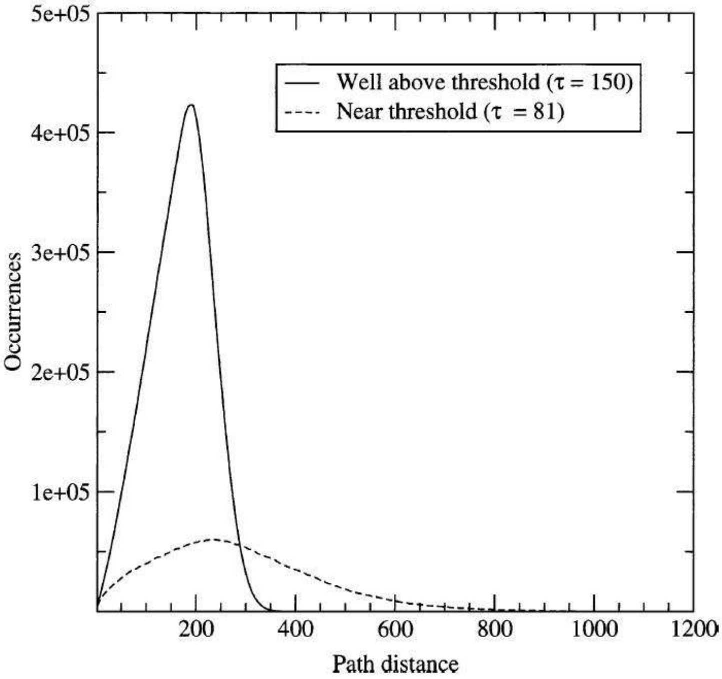


Figure 6.3. Figure 4 from [121]. The critical threshold was varied by adjusting τ, the fixed infectious period. The caption reads “The frequency of occurrence of path distances from a recovered site through its infectors back to the origin averaged over 1000 simulations on a 200×200 lattice. Note that the overall number of paths is larger well above threshold; near threshold there are many bottlenecks in the spread of the epidemic.”

Interestingly, there is a substantial body of literature on percolation on finite *random* graphs; here, tractability increases from the deterministic case when the probabilistic nature of the *existence* of an edge can be coupled into questions regarding the *openness* of that edge. For graphs whose node degrees are drawn independently from an arbitrary degree distribution, Callaway et al. develop expressions for the generating function associated with the distribution of cluster sizes under bond percolation [123], using the generating function methodology described by Newman, Strogatz and Watts in [124]. A quick sketch of this approach is useful. Assume that a graph is generated by choosing node degrees independently from an arbitrary distribution; once all nodes have been assigned degrees, one of the networks consistent with this degree distribution is chosen uniformly at random.[5] If p_k is the probability that a node has degree k, then one can define the moment-generating function of this distribution as

$$G_0(x) = \sum_k p_k x^k.$$

Now, let us consider a bond percolation model on this random graph. If an edge is open with probability T, then the moment-generating function $G_0(x;T)$ for the number of open edges attached to a vertex is given by

$$G_0(x;T) = G_0(1 + (x-1)T).$$

Similarly, one can build moment-generating functions for other quantities, like the distribution of cluster sizes. These functions must satisfy certain self-consistency properties, and numerical methods can be used to solve for quantities like the mean cluster size. Newman extends these results to more general infection processes in [126]. Kalisky and Cohen also use generating function methods to examine the form of the *survivability function*, $S(p,l)$, the probability that, under a bond percolation model with probability p and starting from a randomly-chosen node in a cluster, there exists at least one node at distance l in that same cluster [127]. Around the critical probability $p = p_c$, they find that this probability is exponential in l.

In [128], Ferrari et al. apply simulations and the generating function approach of Newman et al. to a discrete-time stochastic SIR model to explore how removal of nodes via immunity influences the susceptibility of a network to future epidemics. At each time step, a susceptible node x_i is infected with probability $1 - e^{-\lambda |I \cap \mathcal{N}(x_i)|}$, and an infected node recovers with probability γ; in [128], the authors use fixed values of $\lambda = 0.05$ and $\gamma = 0.1$.[6] Three types of networks are simulated: a Watts-Strogatz small-world network[7], a BA network (also called a "scale-free" network), and an

[5] This construction is often called the "configuration model"; see [125].

[6] The authors report varying λ from 0.01 to 0.05 without a qualitative difference in their results, but do not address the critical behavior that might arise from the interaction of β, γ and the network topology.

[7] A Watts-Strogatz network on n nodes begins with a ring graph on n nodes, with edges connecting each node to d of its nearest neighbors. Edges are then rewired uniformly at random with some probability. This yields a graph with high local clustering and short path lengths between any two randomly chosen nodes: see [129].

ER network (defined in Section 5.1), each of which had 1000 nodes and a mean degree of 10. For each type of network, an infection is seeded and allowed to spread, then the removed individuals are subtracted from the network (to form a *residual* network) and new degrees are calculated for each node. Comparing the original and new degrees allows the authors to compare the impact of immunity from two different possible mechanisms: as a consequence of previous infection and as a consequence of random vaccination programs. They define two measures, *frailty* ϕ and *interference* θ as follows:

$$\phi = \frac{\langle k \rangle - \langle k \rangle_r}{\langle k \rangle}$$

$$\theta = \frac{\langle k \rangle_r - \langle k_r \rangle_r}{\langle k \rangle}$$

where $\langle k \rangle$ is the mean original degree, $\langle k \rangle_r$ is the mean original degree of nodes which remain in the residual network, and $\langle k_r \rangle_r$ is the mean residual degree. Frailty, then, measures the preferential immunity that an epidemic might give to nodes of different degree, while interference measures how the removal of immunized nodes changes the distribution. An example of these results is given in Figure 6.4.

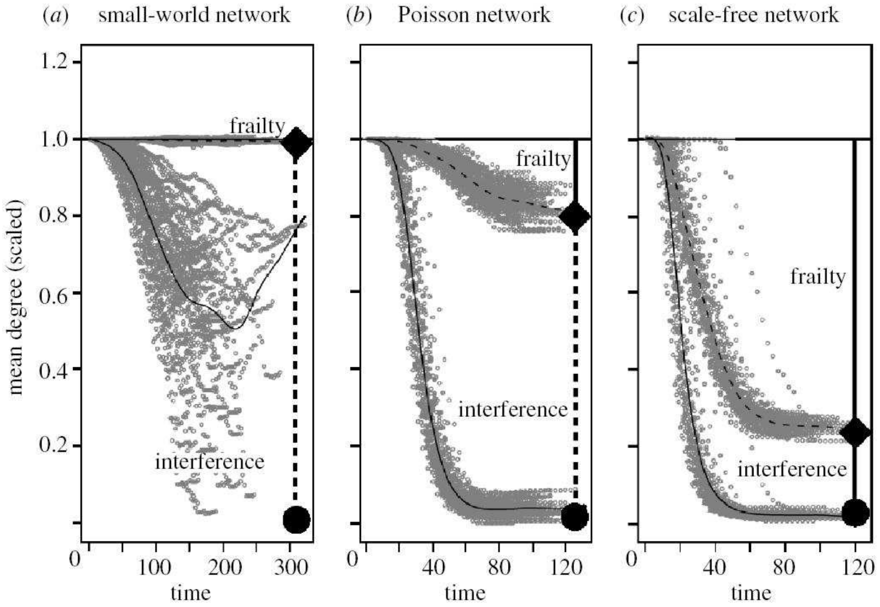


Figure 6.4. Figure 2 from [128]. The caption reads "Mean original degree and mean residual degree (scaled to $\langle k \rangle = 1$) of the active epidemic network (susceptible and infectious nodes) for 100 simulated network epidemics and analytical predictions for (a) small-world, (b) Poisson and (c) scale-free networks of 1000 nodes. Each epidemic was simulated on a separate network with $\beta = 0.05$. The dashed curve gives the mean original degree of nodes across all the networks and the solid curve gives the mean residual degree across time. Points indicate the simulated trajectories. The solid diamond indicates the predicted final mean original degree, $\langle k \rangle_r$, and the solid circle indicates the predicted mean residual degree, $\langle k_r \rangle_r$. The solid vertical bar indicates the predicted frailty and the dashed bar indicates the predicted interference."

In this figure, it is striking how the analytical predictions for the small-world network diverge from the simulation results. This, however, is to be expected; the generating function methodology allows one to assume an arbitrary degree distribution, but assumes that node connections are uncorrelated. This kind of assumption destroys the local clustering that is critical to the structure of a small-world network, and also explains why these approaches are inappropriate for similarly-clustered graphs like $\mathbb{Z}^2$.

To remedy this gap, Serrano and Boguna have developed a generating function methodology for dealing with random graphs with degree correlations and clustering (the tendency of the neighborhoods of two connected nodes to overlap) [130] [131]. To do this, the authors define a set of probability densities $g(s|k)$, which denotes the probability that a vertex can reach s other vertices given that it is connected to a vertex v of degree k and that it cannot visit either v or the neighborhood of v. By looking at solutions for the generating function of this distribution

$$\widehat{g}(z|k) = \sum_s z^s g(s|k)$$

given the constraints that it must satisfy, Serrano and Boguna find that the expected number of reachable nodes diverges for every k (i.e., a giant component forms) when the largest eigenvalue of the following matrix is greater than 1:

$$(k' - 1 - m_{kk'})P(k'|k)$$

where

- $\triangleright$ $m_{kk'}$ is the average number of triangles in which an edge connecting nodes of degree k and k' participates, and
- $\triangleright$ $P(k'|k)$ is the probability that an edge with one end at a degree k node has its other end at a degree k' node.

6.4 The contact process

A third class of probabilistic model that has received a great deal of attention is the continuous-time *contact process*, appropriate for SIS infections. In this model, a susceptible node x_i becomes infected at rate $\lambda|I_t \cap \mathcal{N}(x_i)|$ (just as in the previous section). However, once infected, a node returns to the susceptible state at rate 1 and can be infected again [113]. Like the SI and SIR models discussed in previous sections, most research on the contact process aims to identify important values of λ that separate different regimes of behavior for networks with an infinite number of nodes. In particular, two interesting thresholds on λ are often studied. The first is the *lower critical value* (also called the *global survival critical value*), λ_1, defined as the smallest value of λ such that the infection survives indefinitely with positive probability. The *upper critical value* λ_2 is the smallest value of λ such that an arbitrary node will be infected infinitely often with positive probability; when $\lambda > \lambda_2$, the infection is said to *survive strongly.* In [132], Liggett demonstrates that in $\mathbb{Z}^d$, $\lambda_1(d) = \lambda_2(d) = \lambda_c(d)$. The value of $\lambda_c(d)$ is bounded by

$$\frac{1}{2d-1} \leq \lambda_c(d) \leq \frac{2}{d},$$

and for some choices of d, more exact bounds are known [133]. Durrett has also proven a shape theorem for the contact process in $\mathbb{Z}^d$, analogous to those for the SI and SIR processes, which demonstrates that, among other results, the infection front grows linearly from an initial infective at the origin and that it is contained in a convex set: see [134].

On the infinite homogeneous tree $\mathbb{T}_d$, $\lambda_1(d)$ is strictly less than $\lambda_2(d)$; these values are bounded as follows [132]:

$$\frac{1}{d+1} \leq \lambda_1(d) \leq \frac{1}{d-1},$$

$$2-\sqrt{2} \leq \liminf_{d\to\infty} \sqrt{d}\lambda_2(d) \leq \limsup_{d\to\infty} \sqrt{d}\lambda_2(d) \leq 1.$$

6.4.1 Finite graphs

For finite graphs, the all-susceptible state is the unique absorbing state of the contact process, and will be reached eventually with probability 1. However, the results on finite and infinite graphs are not unrelated. For example, the same $\lambda_c(d)$ that determines different regimes of behavior in $\mathbb{Z}^d$ also has relevance for finite d-dimensional "windows" of $\mathbb{Z}^d$ in terms of how the *time to extinction*, a random variable denoted by τ_N, scales with the window side length N.[8] In particular, the process is called *subcritical* when $\lambda < \lambda_c(d)$, because in this range

$$\frac{\tau_N}{\log N} \to \frac{d}{\gamma_-(\lambda)} \tag{6.1}$$

in probability as $N \to \infty$, where $\gamma_-(\lambda)$ is a positive, decreasing function with $\gamma_-(0) = 1$ and the initial condition is the all-infected state.[9] By contrast, when the process is *supercritical*, $\lambda > \lambda_c(d)$,

$$\frac{\log(\tau_N)}{N^d} \to \gamma_+(\lambda)$$

[8] This is simply a generalization of the square windows of $\mathbb{Z}^2$ discussed earlier in the chapter.

[9] For the special case of $\lambda = 0$, in which no new infections arise, we can observe this result directly. Suppose that a fraction α of the N^d nodes in the window are initially infected. Then τ_N can be written as

$$\tau_N = X_{0\to1} + X_{1\to2} + \cdots + X_{\alpha N^d-1\to\alpha N^d}$$

where $X_{i\to j}$ is the random variable that represents the time elapsed between the ith recovered node and the jth recovered node. Because each of the nodes is recovering independently at rate 1, the random variable $X_{i\to j}$ will be exponentially distributed with parameter $\lambda_{i\to j} = \alpha N^d - i$ (the number of nodes that are still infected after i have recovered). Moreover, the memorylessness of the recovery process implies that the $X_{i\to j}$ form a mutually independent in probability as $N \to \infty$, where $\gamma_+(\lambda)$ is positive and decreasing [132].

For finite regular trees of depth h, Stacey has demonstrated the following results for τ_h^0, the time until extinction given a single inital infective at the root:

Theorem 6.4.1. Proposition 1.2, Theorem 1.3 and Theorem 1.4 of [137]. *Consider a finite regular tree with degree d whose depth from the root is h. Let $\lambda_1(d)$ and $\lambda_2(d)$ be the critical values for the infinite homogeneous tree $\mathbb{T}_d$. Then*

1. *when $\lambda < \lambda_1(d)$, there exists a γ such that for any h,*

$$Pr(\tau_h^0 > t) \le e^{-\gamma t}; \tag{6.2}$$

2. *when $\lambda_1(d) < \lambda < \lambda_2(d)$, there exists a function $r(h, d) > 0$ such that for $s < r(h, d)$,*

$$\liminf_{h \to \infty} Pr(\tau_h^0 > sh) > 0$$

and for $s > r(h, d)$

$$\lim_{h \to \infty} Pr(\tau_h^0 > sh) = 0;$$

3. *when $\lambda_2(d) < \lambda$, and for $\aleph < 1$, there exist $c, \epsilon > 0$, $\Upsilon > 1$ such that for any h*

$$Pr\left(\tau_h^0 \ge c\Upsilon^{(d\aleph)^h}\right) \ge \epsilon.$$

set. We can compute the expectation and variance of τ_N to be

$$\begin{aligned} E[\tau_N] &= E[X_{0\to1}] + E[X_{1\to2}] + \cdots + E[X_{\alpha N^d - 1 \to \alpha N^d}] \\ &= \frac{1}{\alpha N^d} + \frac{1}{\alpha N^d - 1} + \cdots \frac{1}{1} \\ &= \sum_{i=1}^{\alpha N^d} \frac{1}{i} \end{aligned}$$

$$\begin{aligned} \mathrm{var}(\tau_N) &= \mathrm{var}(X_{0\to1}) + \mathrm{var}(X_{1\to2}) + \cdots + \mathrm{var}(X_{\alpha N^d - 1 \to \alpha N^d}) \\ &= \frac{1}{(\alpha N^d)^2} + \frac{1}{(\alpha N^d - 1)^2} + \cdots \frac{1}{(1)^2} \\ &= \sum_{i=1}^{\alpha N^d} \frac{1}{i^2}. \end{aligned}$$

As $N \to \infty$, the variance summation converges to $\pi^2/6$ (a fact demonstrated by Euler in 1736 [135]). The expectation summation does not converge, but approaches a function of N,

$$E[\tau_N] = \log(\alpha N^d) + \gamma = d \log(N) + \log(\alpha) + \gamma,$$

where γ is the Euler-Mascheroni constant, roughly 0.577 [136]. If we consider the scaled random variable $\tau_N / \log(N)$, then as $N \to \infty$,

$$E\left[\frac{\tau_N}{\log(N)}\right] \to d$$

$$\mathrm{var}\left(\frac{\tau_N}{\log(N)}\right) \to 0$$

and the relationship of Eq. 6.1 holds.

One of the consequences of Eq. 6.2 is that the cumulative density function (CDF) of the extinction time τ_h^0 is lower bounded by the CDF of an exponential random variable with parameter γ:

$$\Pr(\tau_h^0 < t) = 1 - \Pr(\tau_h^0 > t) \geq 1 - e^{-\gamma t}.$$

Thus, $E[\tau_h^0]$ is upper bounded by $1/\gamma$, a fixed value independent of h and d. This implies that when $\lambda < \lambda_1(d)$, the expected extinction time from a single initial infective cannot grow without bound as $n \to \infty$; this is distinctly different from the time to extinction for finite windows of $\mathbb{Z}^d$, which grows as $\log(n)$. However, when $\lambda < \lambda_2(d)$, starting from the all-infected state leads to *linear* growth of the extinction time as $n \to \infty$; see [137] for the details of this result.

There is one topology that is amenable to direct analysis: the complete graph. Since every pair of nodes is joined by an edge, the infection rates for all nodes are the same, and are proportional to the number of infected nodes. Analytically, one can write:

$$\begin{cases} P(I_n(t+dt) = i+1 | I_n(t) = i) & = & \lambda i(n-i)dt + o(dt) \\ P(I_n(t+dt) = i-1 | I_n(t) = i) & = & idt + o(dt) \end{cases} \tag{6.3}$$

where $I_n(t)$ denotes the total number of infected individuals at time t. The model of System 6.3 is referred to the *stochastic logistic epidemic*, and has been studied by many researchers; in particular, Anderson and Djehiche present results on the asymptotic *distribution* of τ_n:

Theorem 6.4.2. From Theorem 1 of [138]. *The time to extinction τ_n has the following asymptotic properties:*

1. *If $n\lambda < 1$ and a nonzero fraction $\bar{a}$ of the population is initially infected, then the following convergence in distribution occurs:*

$$(1 - n\lambda(1-\bar{a}))\tau_n - \log n - \log \bar{a} - \log(1 - n\lambda(1-\bar{a})) \to W$$

where W has the extreme value distribution

$$P(W \leq w) = \exp\{-e^{-w}\}.$$

2. *If $n\lambda > 1$ and a nonzero fraction $\bar{a}$ of the population is initially infected, then $\tau_n / E[\tau_n] \to Z$ in distribution, where Z is exponentially distributed with parameter 1 and*

$$E[\tau_n] \asymp \sqrt{\frac{2\pi}{n}} \frac{n\lambda}{(n\lambda - 1)^2} e^{n\overline{V}}$$

where $\overline{V} = \log(n\lambda) - 1 + 1/(n\lambda)$.

The expected value of the random variable W in Thm. 6.4.2 is γ, the Euler-Mascheroni constant. Thus, when $n\lambda > 1$, $E[\tau_n]$ grows exponentially as $n \to \infty$; when $n\lambda < 1$, $E[\tau_n]$ has the following behavior, which is $O(\log n)$:

$$E[\tau_n] \to \frac{\gamma + \log n + \log \bar{a} + \log(1 - n\lambda(1-\bar{a}))}{(1 - n\lambda(1-\bar{a}))}.$$

6.5 Arbitrary topologies

In [25], Ganesh et al. look at the contact process on a finite graph with arbitrary topology (represented by the undirected adjacency matrix A) and explore the behavior of $E[\tau]$; in particular, they seek conditions under which $E[\tau]$ grows as the logarithm of n versus exponentially in n (as seen on either side of λ_c for windows of $\mathbb{Z}^d$). Stacey's results and others tell us that these two regimes are not collectively exhaustive, so we should not expect to find a single threshold dividing the regimes for an arbitrary topology. Ganesh et al. achieve the following sufficient conditions.

Theorem 6.5.1. Theorem 3.1 of [25]. *If $\lambda < 1/\rho(A)$, then*

$$E[\tau] \leq \frac{\log(n) + 1}{1 - \lambda\rho(A)}.$$

Theorem 6.5.2. Corollary 4.1 of [25]. *Define $\eta(G, m)$, the* generalized isoperimetric constant[10] *of the graph G, as*

$$\eta(G, m) = \inf_{S \subset \{1,\ldots,n\}, |S| \leq m} \frac{E(S, \overline{S})}{|S|}, 0 < m < \lfloor n/2 \rfloor,$$

where $E(S, \overline{S})$ counts the number of edges connecting vertices in set S to vertices in $\overline{S}$. Define

$$r(G, m) = \frac{1}{\lambda\eta(G, m)}.$$

For a sequence of graphs G_n indexed by n, suppose there exists an $a > 0$ and a sequence $m_n = \Theta(n^a)$ such that $r(G_n, m_n) < 1$ uniformly in n. Then $\log(E[\tau]) = \Omega(n^a)$.[11]

In short, Ganesh et al. observe fast die-off when $\lambda < 1/\rho(A)$ (i.e., $E[\tau] = O(\log(n))$) and slow die-off when $\lambda > 1/\eta(G, m)$ (i.e., $\log(E[\tau]) = \Omega(n^a)$). Observe that Theorem 6.5.2 gives a bound on the rate of growth of $E[\tau]$ as a function of n that is related to the rate of growth of λ as a function of n. In general, the two threshold values in these theorems do not coincide. However, Ganesh et al. consider several special topologies and improve the conditions from Theorems 6.5.1 and 6.5.2; for example, the gap between the two regimes is "closed" for the hypercube and the complete graph.

As an additional example, let us apply Theorems 6.5.1 and 6.5.2 to one particular topology: the torus on n nodes. All nodes in this graph have the same degree, 4, so $\rho(A) = 4$ and Theorem 6.5.1 tells us that the extinction time will grow as $O(\log n)$ for $\lambda < 1/4$. To determine a sufficient condition for slow die-off via Theorem 6.5.2, we must determine the generalized isoperimetric constant. For any subset S of nodes with cardinality $|S|$, the smallest value of $E(S, \overline{S})$ that can be obtained occurs when the $|S|$ nodes are arranged in the closest approximation to an $\sqrt{|S|} \times \sqrt{|S|}$ square. In this

[10] When $m = \lfloor n/2 \rfloor$, this quantity is known as the isoperimetric constant, the edge-isoperimetric constant, Cheeger's constant or the edge expansion of G.

[11] We will use the following order notation conventions throughout this chapter, where $g(n)$ is a positive function:

- ▷ $f(n) = O(g(n))$ if there exist $c, N > 0$ such that $f(n) \leq cg(n)$ for all $n > N$.
- ▷ $f(n) = \Theta(g(n))$ if there exist $c, d, N > 0$ such that $cg(n) \leq f(n) \leq dg(n)$ for all $n > N$.
- ▷ $f(n) = \Omega(g(n))$ if there exist $c, N > 0$ such that $f(n) \geq cg(n)$ for all $n > N$.

case, the number of edges between nodes in S and nodes in $\overline{S}$ is $4\sqrt{|S|}$, which yields

$$\frac{E(S,\overline{S})}{|S|} = \frac{4\sqrt{|S|}}{|S|} = \frac{4}{\sqrt{|S|}}.$$

This quantity is minimized when $|S|$ is chosen to be as large as possible; thus, for any m, $\eta(G,m) \approx 4/\sqrt{m}$. In order to apply Theorem 6.5.2, we must be able to construct a sequence m_n such that $r(m_n) < 1$ for all n. If $m_n = \Theta(n^a)$ for $a \in (0,1)$, then there exists a constant c and a positive integer N such that $m_n > cn^a$ for all $n > N$. This implies that for $n > N$,

$$1 > r(m_n) = \frac{1}{\lambda\eta(G,m_n)} = \frac{\sqrt{m_n}}{4\lambda} > \frac{\sqrt{cn^a}}{4\lambda} = \frac{\sqrt{c}}{4\lambda}n^{a/2}.$$

For this condition to be satisfied, λ must grow faster than $n^{a/2}$; if this occurs, then $\log(E[\tau])$ will be $\Omega(n^a)$.

6.5.1 Simulating the contact process

To begin to get a handle on this behavior, we began by verifying Theorems 6.5.1 and 6.5.2 empirically for four topologies: the star, the hypercube, the complete graph and the torus. The first three of these topologies are treated in detail in [25]. Table 6.1 summarizes the critical ranges of λ for these topologies on n nodes. Observe the agreement of the threshold for the complete graph (in the large n limit) with the results of Andersson and Djehiche in Theorem 6.4.2.

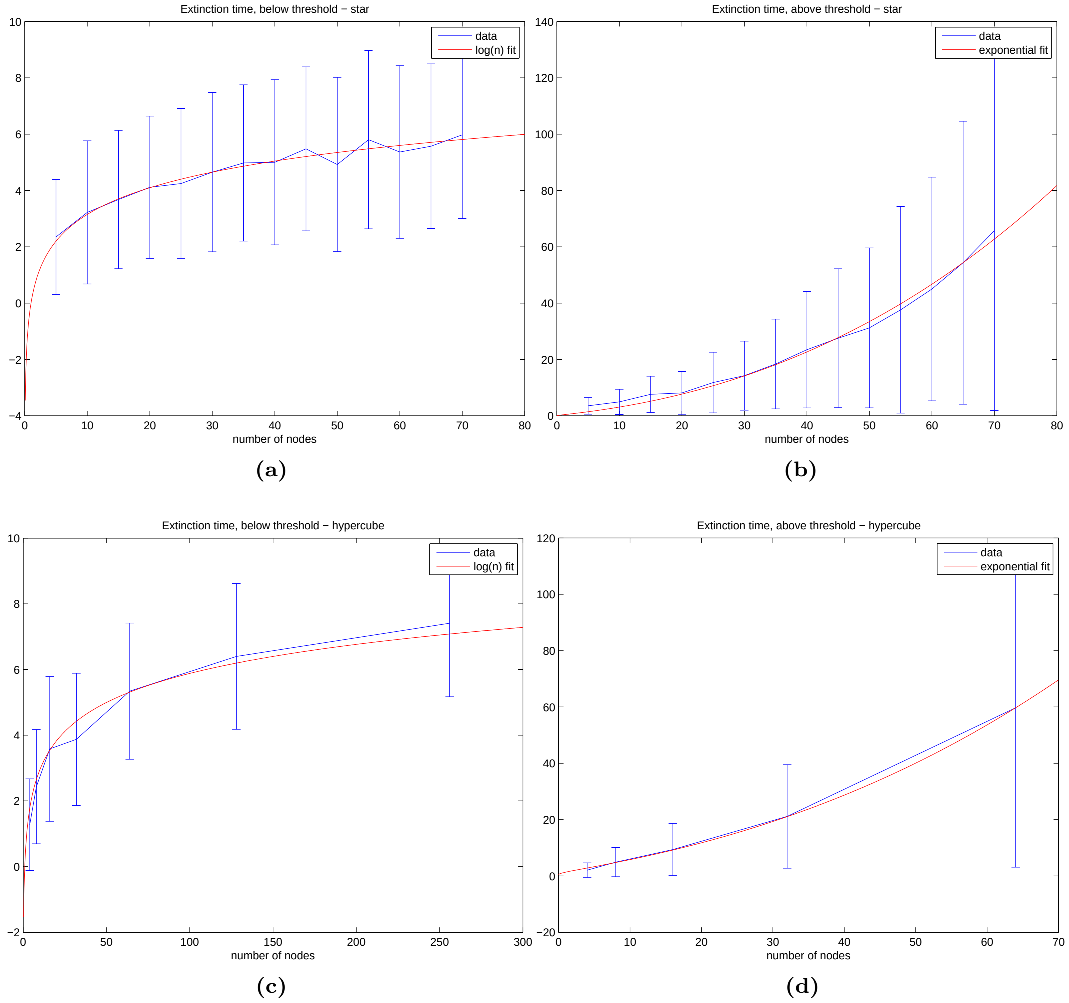
Table 6.1. Extinction regimes for the contact process on n-node graphs, [25].

topology	$E[\tau] = O(\log n)$	$\log(E[\tau]) = \Theta(n^a)$
star	$\lambda < \frac{C}{\sqrt{n}}, C > 0$	$\lambda > n^{a-1/2}, a \in (0,1)$
hypercube	$\lambda < \frac{1}{\log_2(n)}$	$\lambda > \frac{1}{(1-a)\log_2(n)}, a \in (0,1)$
complete graph	$\lambda < \frac{1}{n-1}$	$\lambda > \frac{1}{n-n^a}, a \in (0,1)$
torus	$\lambda < \frac{1}{4}$	$\lambda > \frac{1}{4}n^{a/2}, a \in (0,1)$

Figures 6.5 and 6.6 depict the time to extinction for these topologies for various values of n for both λ below and above the thresholds listed in Table 6.1. The values of λ used in each of these simulations is given in Table 6.2, as is the number of trials over which sample means and standard deviations were computed. The number of trials conducted above threshold is often smaller than the number conducted below threshold because of the simulation time required. Each node in each network was initially infected with probability 1/4; for the star, the center node was always initially infected.

Table 6.2. Values of λ used in simulations of Section 6.5.1. The number of trials for each simulation is given in parentheses.

topology	**below threshold**	**above threshold**
star	$\frac{1}{2}\frac{1}{\sqrt{n}}$ (500)	$n^{-1/4}$ (500)
hypercube	$\frac{1}{2}\frac{1}{\log_2(n)}$ (500)	$\frac{3}{2}\frac{1}{\log_2(n)}$ (300)
complete graph	$\frac{1}{2}\frac{1}{n-1}$ (500)	$\lambda > \frac{1}{0.7n}$ (300)
torus	$\lambda < \frac{1}{5}$ (500)	$\frac{1}{4}n^{1/8}$ (300)



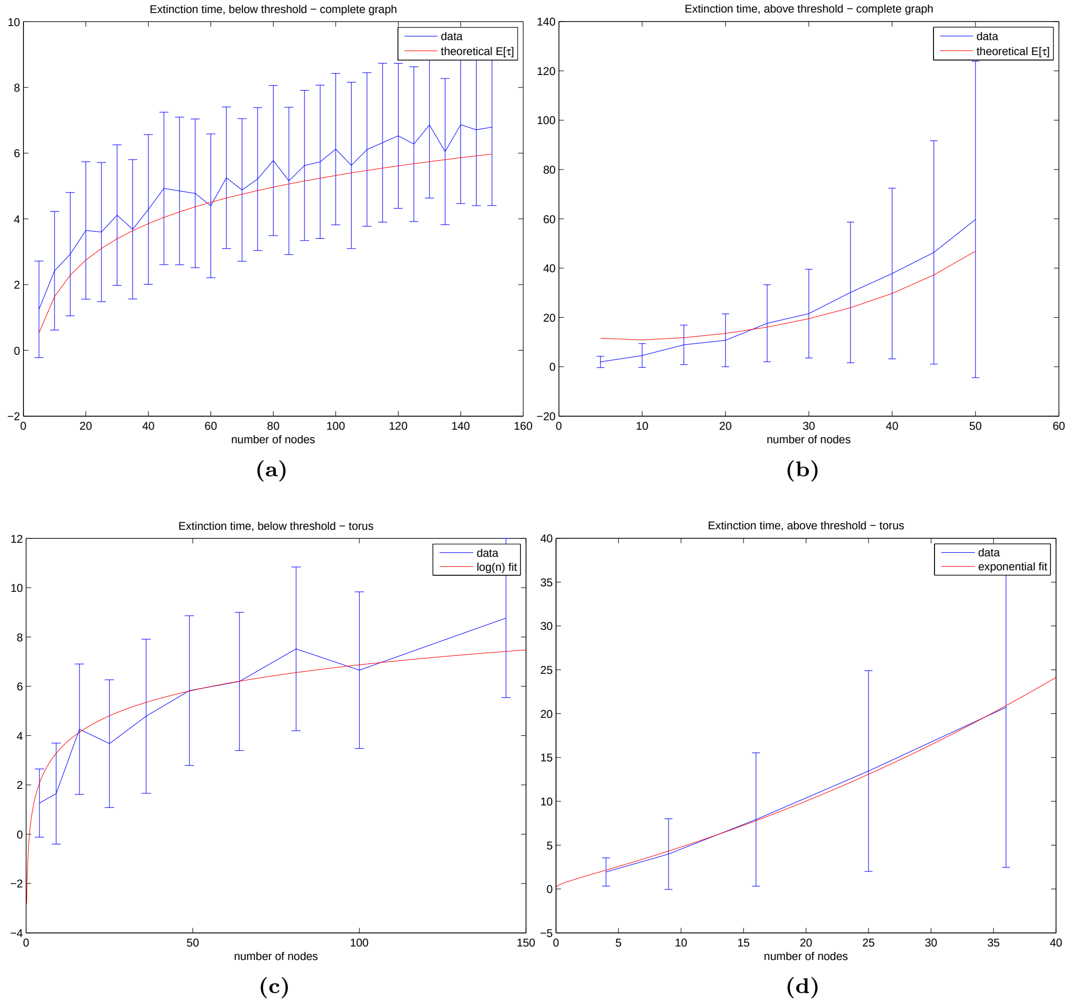
(a)

(b)

(c)

(d)

Figure 6.5. Mean and standard deviation of the time to extinction for the star and the hypercube. Each node in each network was initially infected with probability 1/4. The range of number of nodes considered is smaller for $\lambda > \lambda_c$ because of the amount of simulation time required.



(a) **(b)**

(c) **(d)**

Figure 6.6. Mean and standard deviation of the time to extinction for the complete graph and the torus. Each node in each network was initially infected with probability 1/4. The range of number of nodes considered is smaller for $\lambda > \lambda_c$ because of the amount of simulation time required.

Indeed, we see the behavior predicted by Theorems 6.5.1 and 6.5.2. The "$\log(n)$" curves result from a least-squares fit of the data to a function $c_1 \log(n)$. The "exponential" curves correspond to a least-squares fit to $c_1 \exp(c_2 x^a)$ where a represents the growth rate of λ as a function of n: see Table 6.2. For the complete graph, we've fitted the data to the curves predicted by Andersson and Djehiche in Theorem 6.4.2, but note that the growth rates predicted by the theorem are asymptotic for large n.

We can also look at the mean number of *transitions*, i.e., changes in state, as a function of n; these are depicted in Figures 6.7 and 6.8. It is interesting to note that in the regime below threshold, the number of transitions to extinction appears to grow linearly with n, with a slope not much less than 1. This suggests that nodes are not becoming infected repeatedly before extinction, an observation we'll make again in Section 6.5.3.

6.5.2 Extinction time distributions

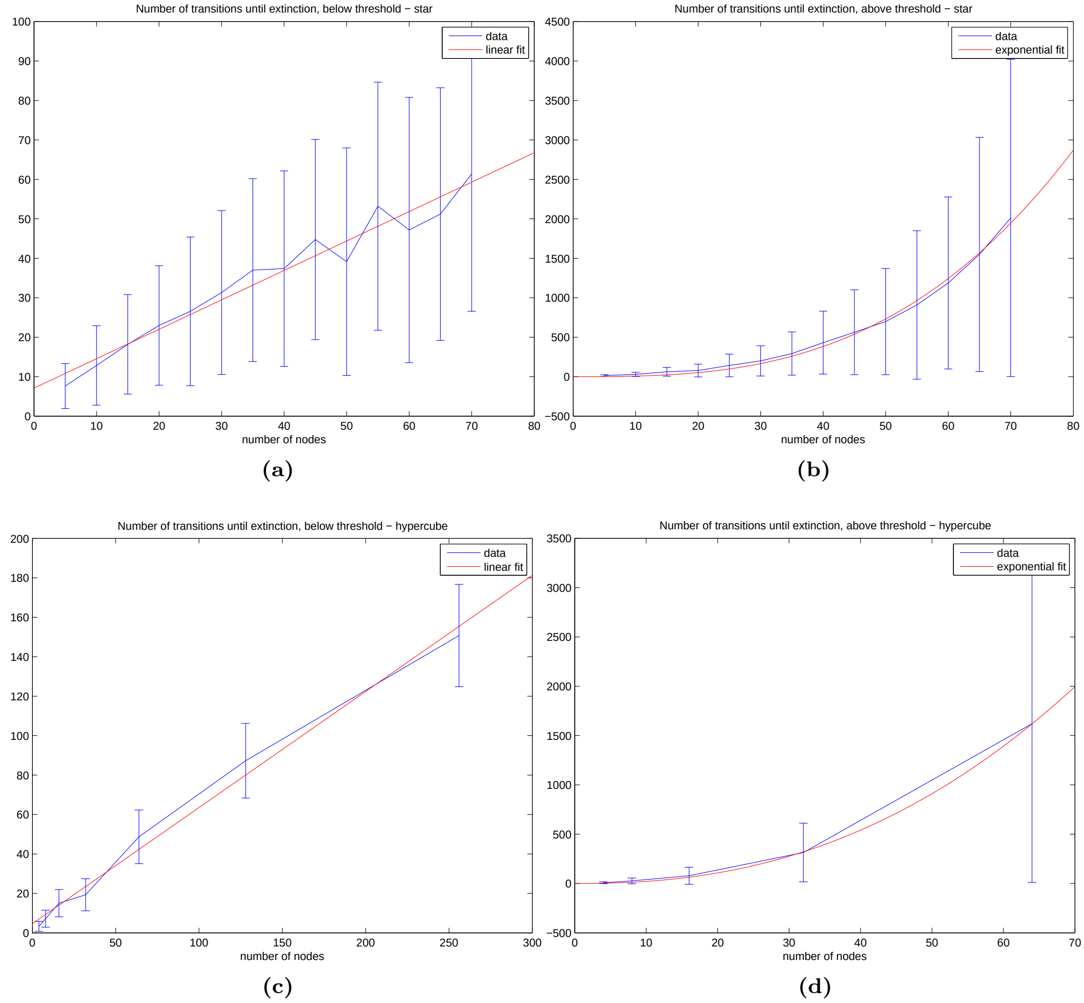
What change in distribution of the extinction time underlies this change in expected value? Figure 6.9 presents histograms of the time to extinction for the star topology, below and above the threshold. The shapes of these distributions are very evocative of those described in Thm. 6.4.2 (exponential and extreme-value shapes), even though the star and complete graph topologies are quite different!

To test this observation, each histogram was fitted to the appropriate distribution, using maximum-likelihood methods to estimate the parameters. The resulting fit was then evaluated using a χ^2 goodness-of-fit test with significance level $\alpha = 0.05$. The results are given in Figures 6.10 and 6.11, with the resulting p-values indicated, as well as whether or not $p > \alpha$.

Interestingly, the torus is the only topology that fails the goodness-of-fit tests, below and above threshold. What makes the torus different from the other topologies? Observe that the torus is the only topology whose maximum degree does not grow with the size of the graph. Additionally, the maximum possible path length on the torus with n nodes is $\sqrt{n}/2$; in the hypercube on n nodes (the only other topology in which this length increases as the number of nodes increases), the maximum path length is $\log_2(n)$, a much slower rate of growth. This is one possible explanation for the longer extinction times on the hypercube; the rate of spread is limited by local clustering and thus the peak of infection is delayed.

6.5.3 Analyzing cluster sizes

The histograms of extinction time in Section 6.5.2 give us some physical feeling for the two regimes of behavior, but focusing on extinction time alone provides a very narrow window into the underlying phenomena. Our interest in spatial behavior suggests that we extend our investigation to the patterns traced as an infection "cluster" progresses through the network. In epidemiology, a cluster is often loosely defined as a set of epidemiological events that are related to each other, typically a group



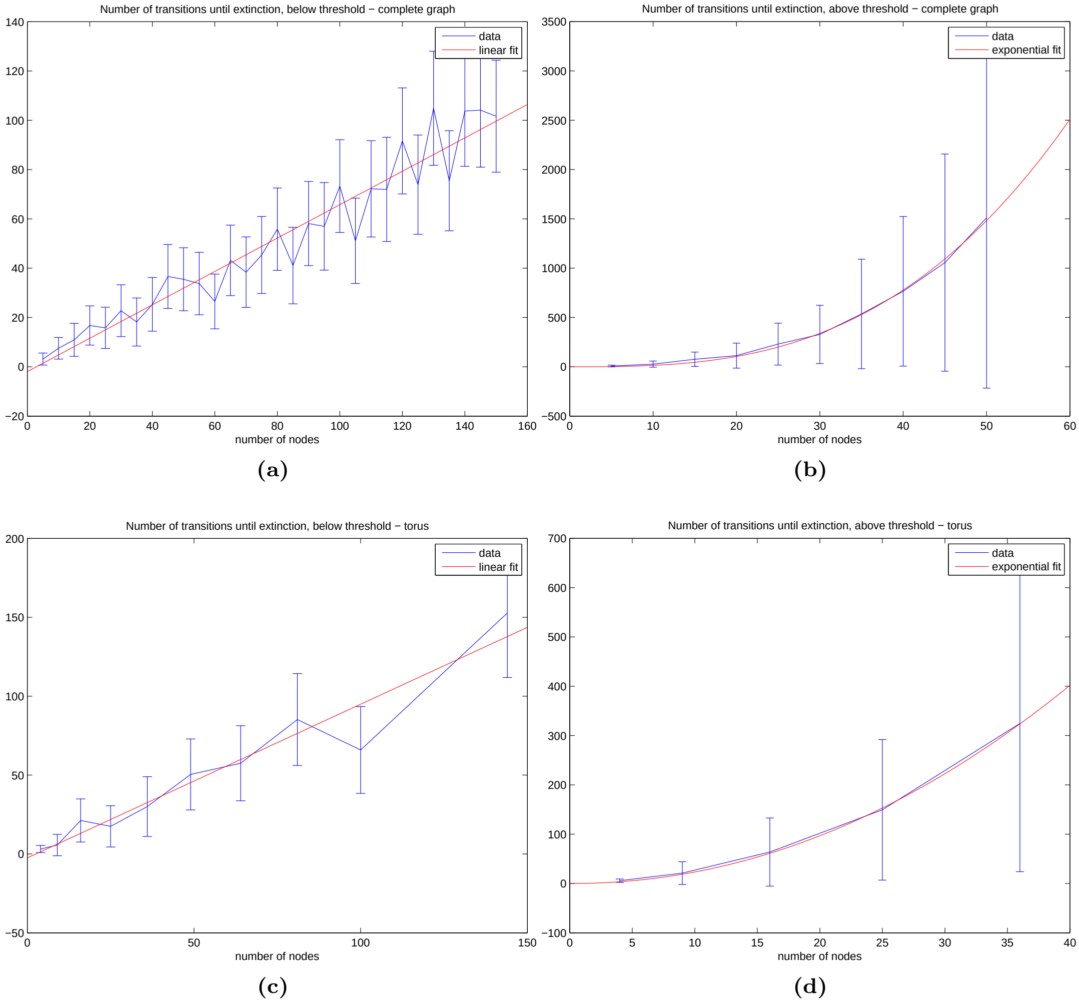
(a)

(b)

(c)

(d)

Figure 6.7. Mean and standard deviation of the number of transitions until extinction for the star and the hypercube. Each node in each network was initially infected with probability 1/4. The range of number of nodes considered is smaller when $\log(E[\tau]) = \Omega(n^a)$ because of the amount of simulation time required.



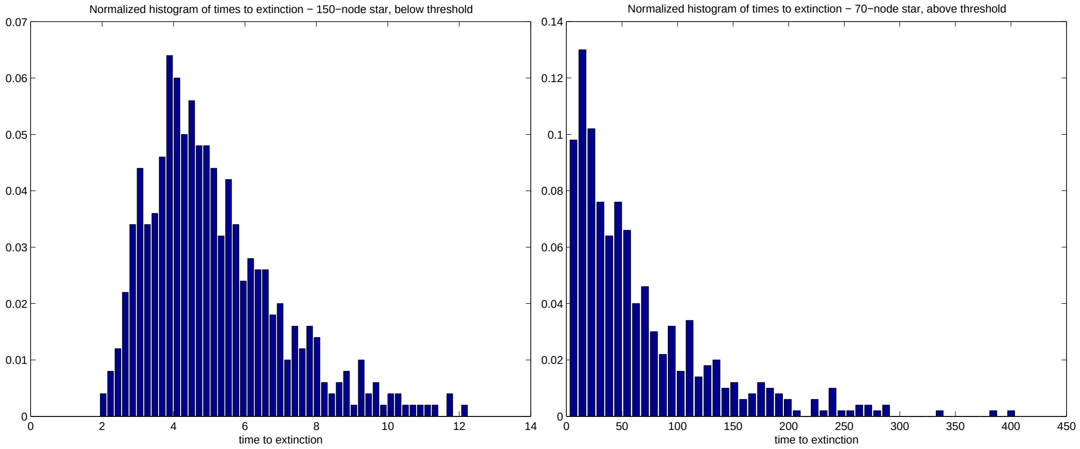
(a)

(b)

(c)

(d)

Figure 6.8. Mean and standard deviation of the number of transitions until extinction for the complete graph and the torus. Each node in each network was initially infected with probability 1/4. The range of number of nodes considered is smaller when $\log(E[\tau]) = \Omega(n^a)$ because of the amount of simulation time required.



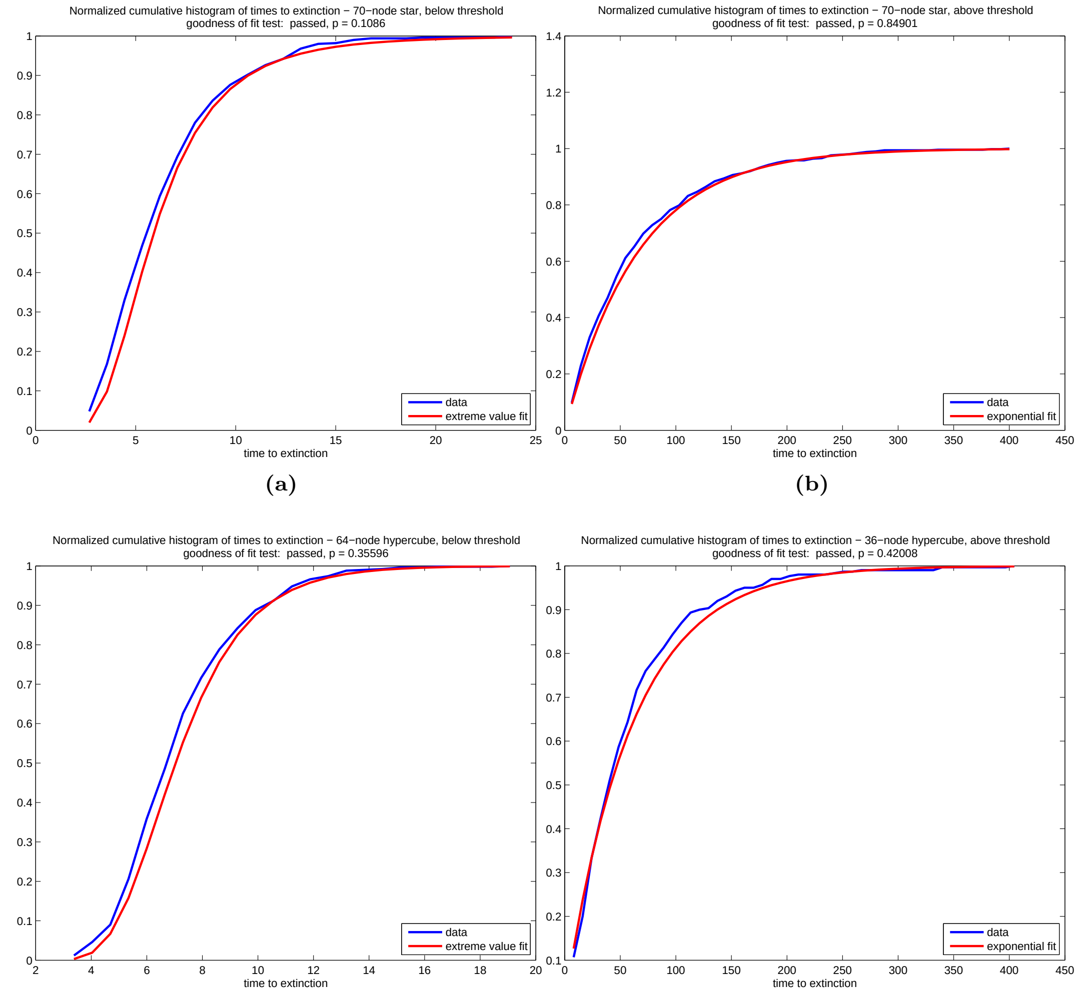
(a) **(b)**

Figure 6.9. Normalized histogram of times to extinction for the contact process on the star.

of new infections that can be traced to a single source. For our purposes, we'll define an infection cluster to be a connected set of simultaneously infected nodes.

To investigate the distribution of these clusters as the infection progresses, the following analysis was performed on the simulations discussed in the previous sections. At each instant of time, we measured the number and size of the connected components comprising the subgraph induced by only the infected nodes. This gives us a histogram at each moment of time of the size of the infected clusters; for each of these histograms, the mean number of clusters, the maximum cluster size, and the total number of infected individuals was recorded. To compare different trials within the same topology, the time scale of each trial was normalized so that extinction occured at time 1; we then computed the average and standard deviation of the recorded statistics across all of the trials. We performed this analysis for each of the four topologies, above and below the extinction time threshold, for the largest graphs simulated of each type. We have excluded the complete graph from the results on mean number of clusters and mean cluster size; since all nodes are connected to all others, these statistics are identical to the total number of infected nodes and 1, respectively. Clustering results on the star must be interpreted carefully; if the center node is infected, there is only one cluster, and otherwise, there are as many clusters as there are infected leaves.

Figure 6.12 depicts the mean number of clusters, while Figure 6.13 depicts the mean cluster size, both versus the normalized time. The results in these figures for the hypercube provide some intuition for the idea of a "quasi-stationary" state, a level of endemic infection that persists for an extended time before the infection is ultimately driven from the system. Below threshold, the system quickly fragments into many smaller infected clusters as nodes recover (the initial increase in the number of clusters), then these clusters disappear steadily. Above threshold, the initial infected mass coalesces into a stable pattern, which persists until a rapid transition to extinction. The torus



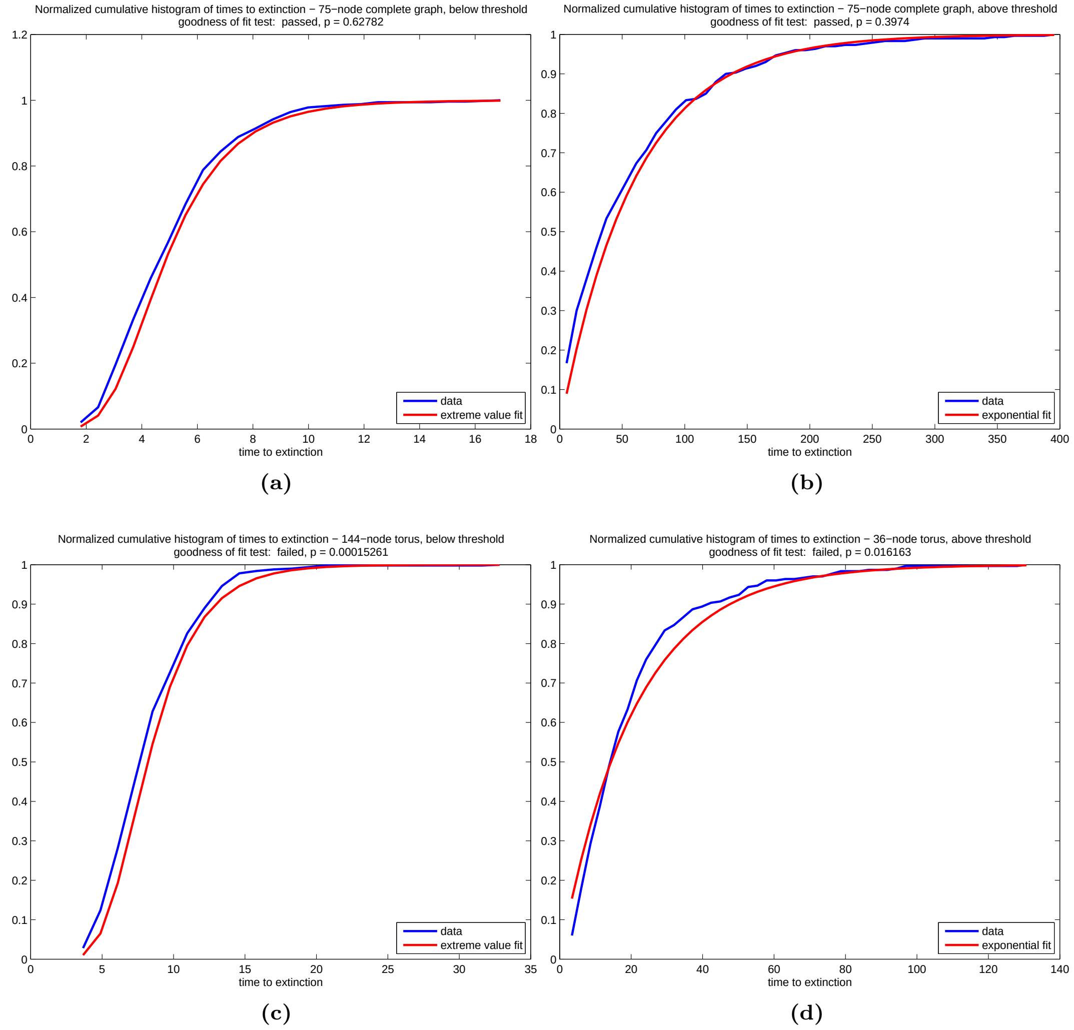
(a)

(b)

(c)

(d)

Figure 6.10. Normalized cumulative histogram of the time until extinction for the star and the hypercube.



(a) **(b)**

(c) **(d)**

Figure 6.11. Normalized cumulative histogram of the time until extinction for the complete graph and the torus.

displays a similar quasi-statationary state above threshold. With the star, the number of clusters increases as extinction approaches because infected leaves are all that remain.

Figures 6.14 and 6.15 depict the total number of infected nodes. These results suggest several possible analytical threshold tests that might have the same behavior as a test on the extinction time; for example, one might be able to test for the existence of an inflection point in the mean number of infectives.

6.6 The binary influence model

We shift our focus to a more general model that can be applied to infection processes. The *influence model*, a probabilistic framework proposed and analyzed by Asavathiratham et al. in [139], [140], which provides both a point of comparison for the contact process and is an interesting and tractable model in its own right. We'll introduce the influence model by way of a simple infection scenario through a weighted, directed network on n nodes.

Assume that each node can either be infected (status '1') or susceptible to infection (status '0'). At each time step, a node j chooses one of its *in-neighbors* i with probability c_{ij}, and copies the status of node i with probability p, and otherwise retains its current state. The sum of the weights of incoming edges to a single node is 1, i.e., $\sum_i c_{ij} = 1$ Note that self-loops in the network allow a node to be its own influencer, retaining its current state for another time step.

Let $s_i[k]$ denote the status of node i at time k, and assemble all of these statuses into a single status vector $s[k]$; we'll call the collection of statuses of the sites the *state* of the network. Define the matrix C such that $\{C\}_{ij} = c_{ij}$ (so C is column-stochastic). Using slightly different notation than [139], we can represent the conditional probability of the state at the next time step, given the current state, as

$$E[s[k+1]|s[k]] = pC^\top s[k]$$

which implies that

$$E[s[k+1]] = pC^\top E[s[k]]. \tag{6.4}$$

Since the entries of $s[k]$ are indicator random variables, $E[s[k+1]]$ yields the probability that each node is in status 1. Eq. 6.4 provides a simple linear update for the expected state of the network and allows us to connect the topology of the network and the dynamics of the influence process. Throughout these notes, we will consider matrices $C^\top$ that can be decomposed into the following form

$$C^\top = dI + (1-d)A,$$

for $d \in [0, 1]$ and where A is a row-stochastic matrix with zero diagonal entries. This decomposition allows us to isolate the effects of a universal *self-influence* parameter d from the effects of the rest

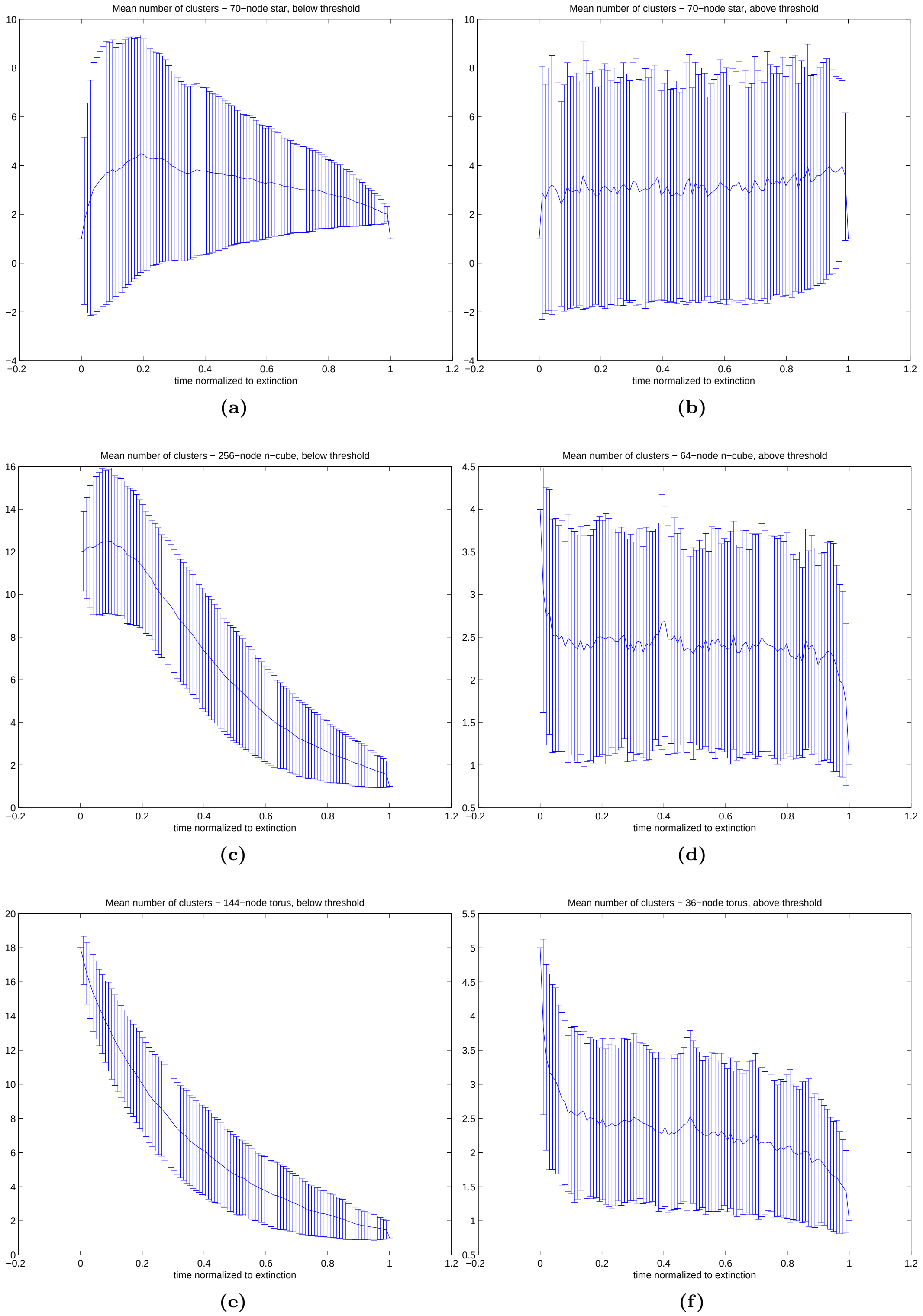
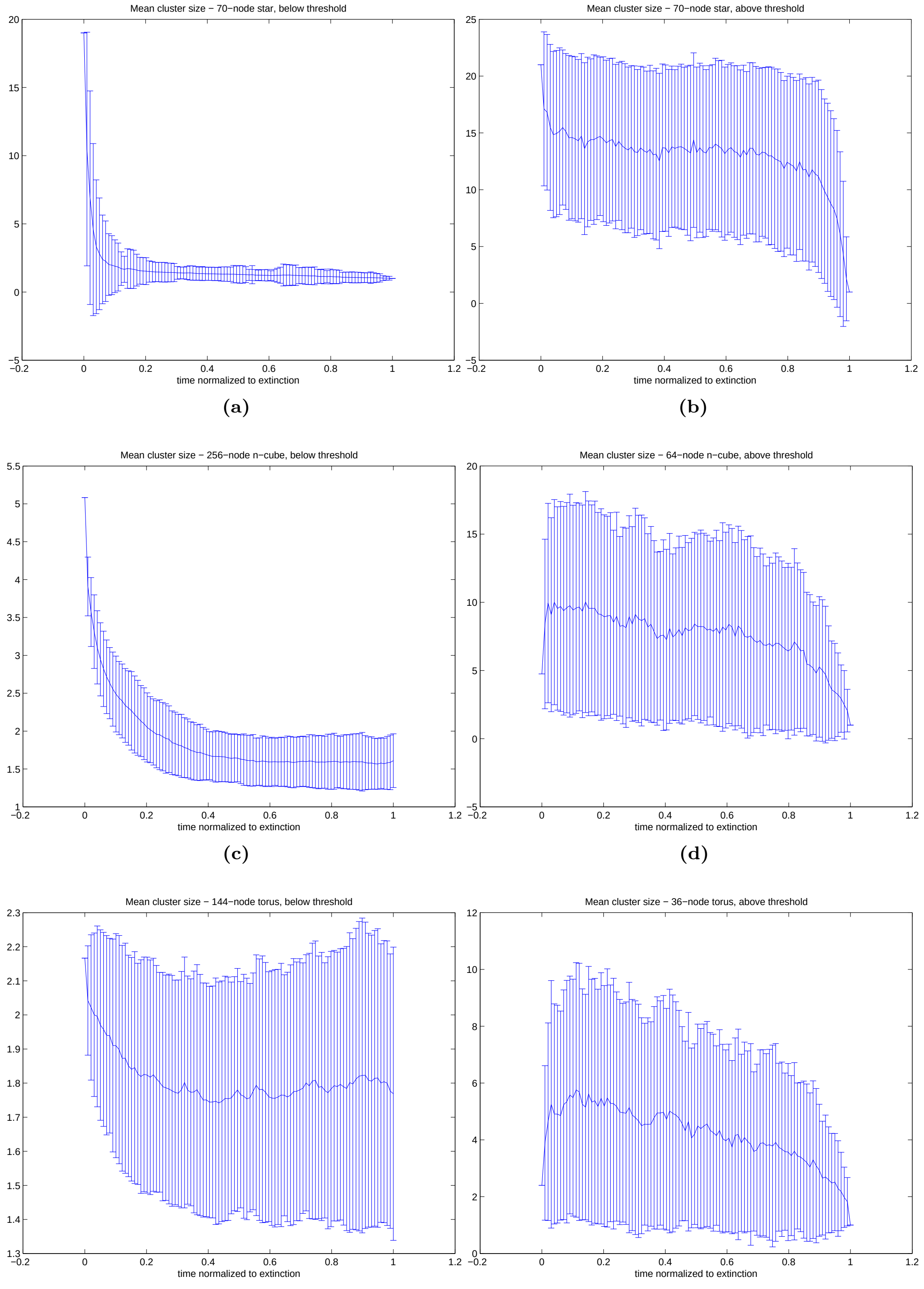


Figure 6.12. Mean and standard deviation of the number of infected clusters until extinction for the star, hypercube and torus.

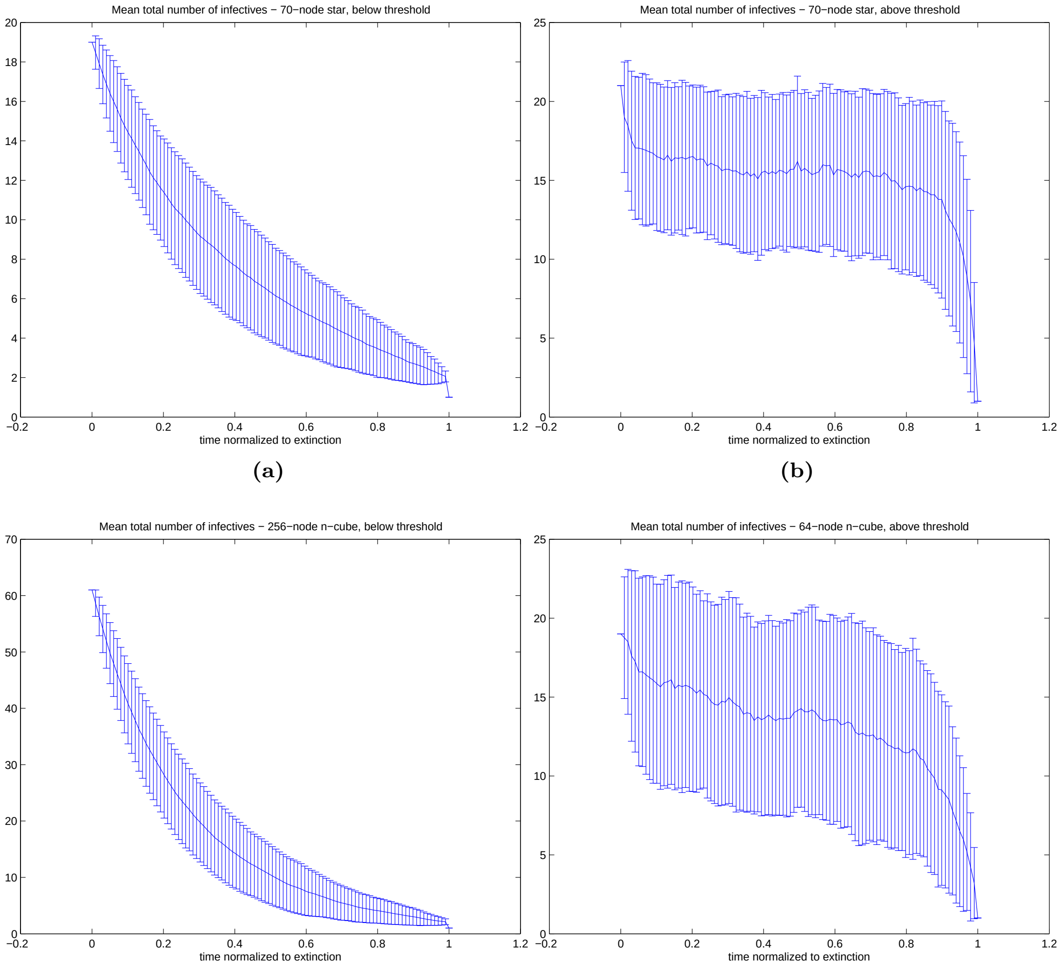


(a) **(b)**

(c) **(d)**

(e) **(f)**

Figure 6.13. Mean and standard deviation of the size of infected clusters until extinction for the star, hypercube and torus.



Mean total number of infectives − 70−node star, below threshold

Mean total number of infectives − 70−node star, above threshold

Mean total number of infectives − 256−node n−cube, below threshold

Mean total number of infectives − 64−node n−cube, above threshold

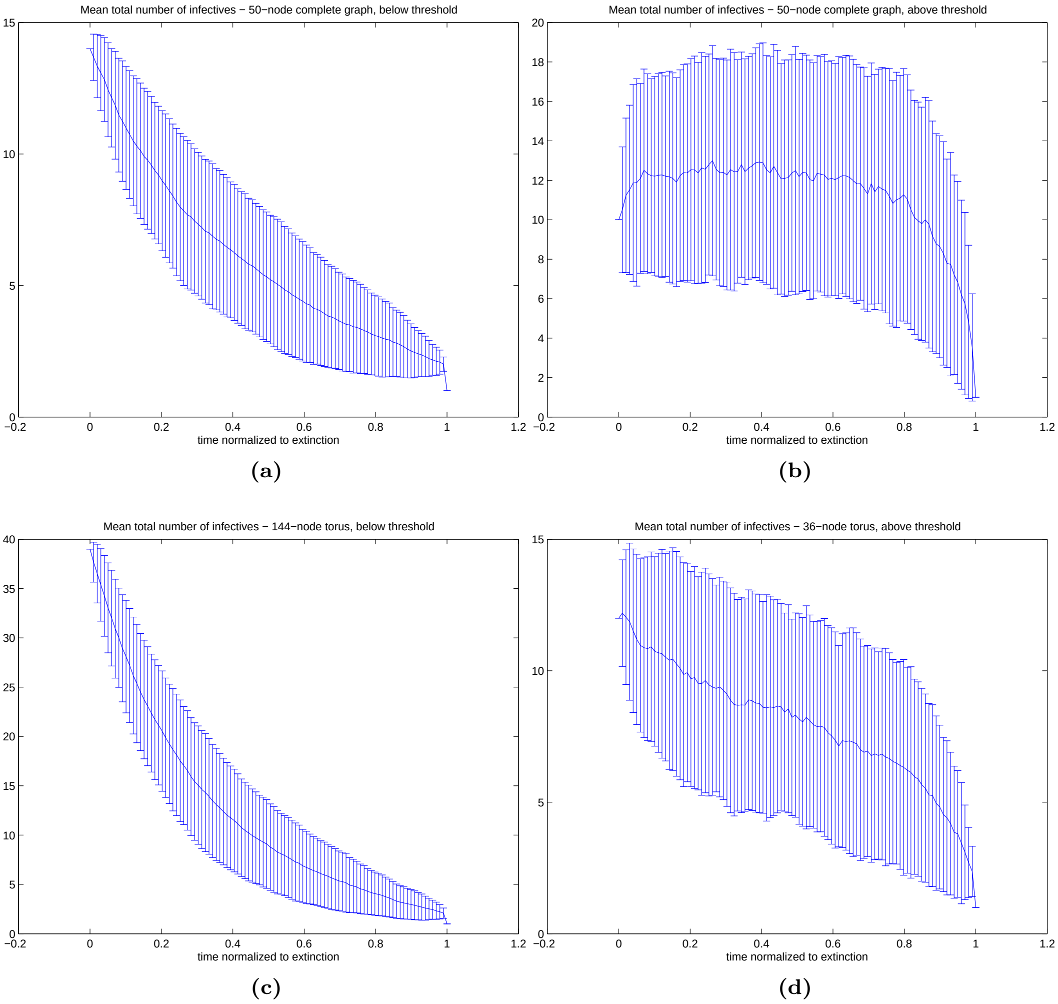
(a)

(b)

(c)

(d)

Figure 6.14. Mean and standard deviation of the total number of infectives for the star and the hypercube.



(a) **(b)** **(c)** **(d)**

Figure 6.15. Mean and standard deviation of the total number of infectives for the complete graph and the torus.

of the topology, represented by A. (Note that our notation differs slightly from that in [139], for better comparison to the contact process.)

Eq. 6.4 is a *linear* propagation rule for the expected value of the statuses of sites in the network, and is thus amenable to all of the tools of linear dynamical systems theory.

6.6.1 Comparing the influence model and the contact process

Given the analytical intractability of the contact process on an arbitrary network topology, it is tempting to use the influence model as an approximation to the contact process. What kinds of parameter regimes and conditions might make this a reasonable comparison? The influence model evolves in discrete time, and allows all nodes to change their state at each time step, while the contact process evolves continuously with negligible probability of two transitions occuring simultaneously. However, one can consider a discrete-time version of the contact process in which the system is sampled at the transition times; in this sampled contact process, one node changes state per time step.

Consider the binary influence model on an arbitrary topology, but assume that each node is equally likely to choose any of its neighbors. If I_j denotes the number of infected neighbors of node j, then the probability that j is infected at the next time step is given by pI_j/m_j, where m_j is the degree of node j.

Now, let us consider the contact process from the perspective of a fixed node j. This node registers arrivals from a Poisson process with rate λ from each of its infected neighbors, so a "germ" arrives at rate λI_j. At the same time, the "antidote" arrives at rate 1. The probability that the next arrival for node j is infectious material is simply

$$\frac{\lambda I_j}{\lambda I_j + 1}.$$

This is a rather loose connection, but one might be tempted to say that the influence model and the contact process may well mimic each other under the condition that these infection probabilities are equal:

$$\frac{\lambda I_j}{\lambda I_j + 1} = p\frac{I_j}{m_j}.$$

If $\lambda I_j \ll 1$, then this condition becomes

$$\lambda I_j \approx p\frac{I_j}{m_j} \Longrightarrow p \approx \lambda m_j. \tag{6.5}$$

If all nodes in the network have the same degree $k_j = m$, then $\rho(A) = m$ and this condition requires that

$$p < 1 \Longrightarrow \lambda < \frac{1}{m} = \frac{1}{\rho(A)},$$

the condition required to apply Theorem 6.5.1. This suggests that the influence model might have some relevance in describing the dynamics of the contact process in the *subcritical* regime. To explore this, we'll look at the relaxation behavior of the total number of infected individuals in the contact process and influence model, starting from the all-infected state.[12] Results are given for the hypercube in Figure 6.16 and the torus in Figure 6.17, for different values of p. In these figures, the time axes are normalized so that extinction occurs at time 1; for the discrete-time influence model, a linear interpolation of the number of infected individuals was performed on each trial before the results were averaged. For both the hypercube and the torus, the results are slightly counterintuitive, at least in light of the loose argument that led to the approximation of Eq. 6.5; in Figures 6.16 and 6.17, we see that the mean and standard deviation of the total number of infected individuals in the influence model matches most closely (and rather well) for intermediate values of p, $p \approx 0.5$, rather than small values.

6.6.2 Some future explorations of spatiotemporal patterns

The ideas discussed in this chapter are only an introduction to the array of interesting questions regarding spatial patterns of infection spread. In particular, the connection between the influence model and other probabilistic infection processes is worth careful consideration; leveraging the influence model's tractability might yield very valuable approximations of important quantities in the contact process and others. Here, we shall discuss one of these connections.

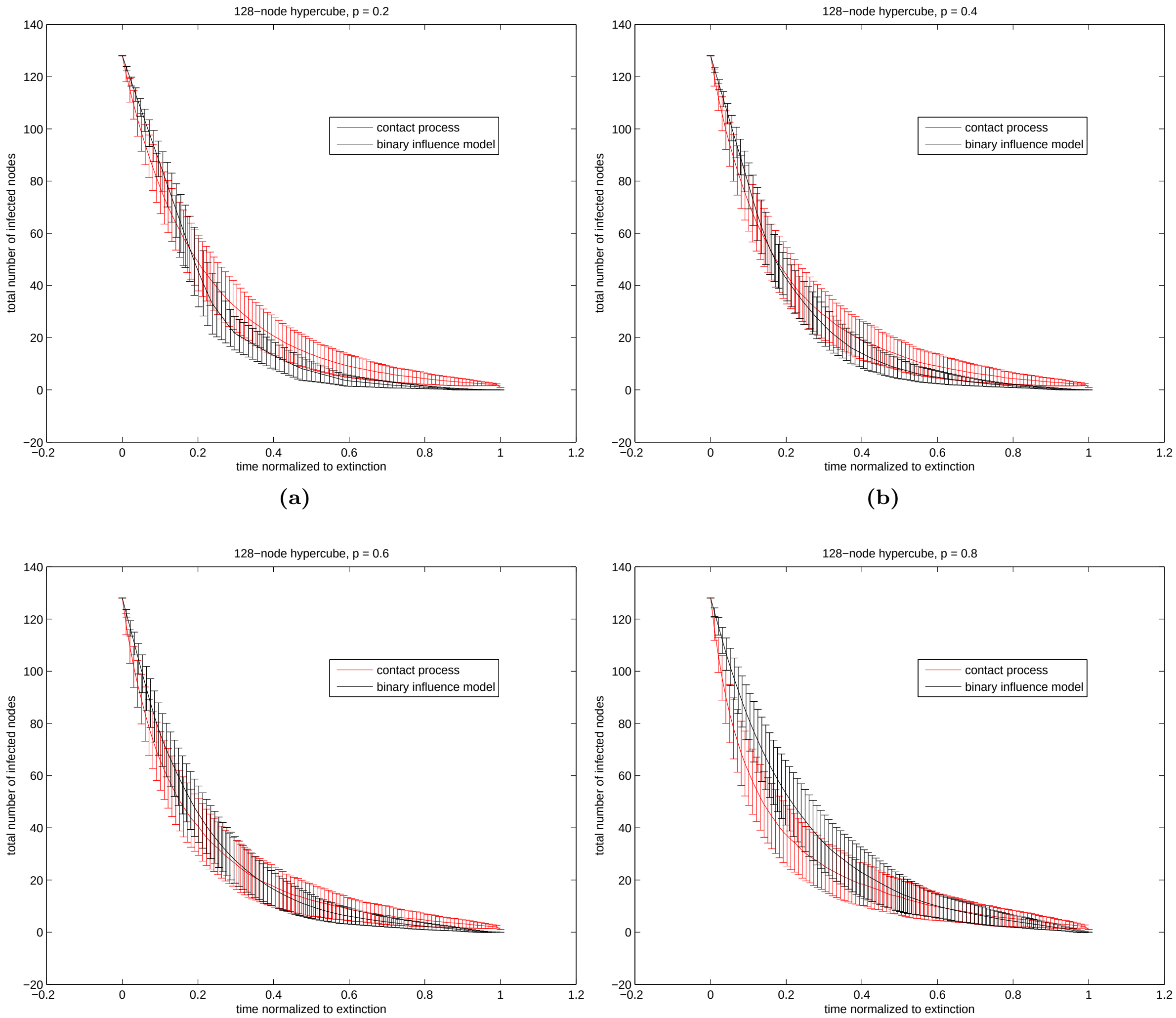
Given our interest in spatial patterns and our previous discussion of clustering, it would be useful to be able to track the node status *correlations* $E[s_i[k]s_j[k]]$. Assemble these products into a matrix $\{M_k\}_{ij} = s_i[k]s_j[k]$, and define $E[s_{k+1}|s_k] \equiv p_{k+1}$. Then, conditioned on the current state,

$$E[M_{k+1}|s_k] = (p_{k+1}\mathbf{1}^\top) \circ I + (p_{k+1}p_{k+1}^\top) \circ (\mathbf{1}_M - I) = (\mathbf{1}^\top D^\top s_k) \circ I + (D^\top s_k s_k^\top D) \circ (\mathbf{1}_M - I) \quad (6.6)$$

where $\mathbf{1}$ is a vector of all ones, $\mathbf{1}_M$ is the matrix of all ones whose dimensions are identical to those of M and $\circ$ denotes the Hadamard, or entrywise, matrix product. Observe that this decomposition into two terms arises from the distinction between the diagonal entries of $E[M_k]$, which are terms of the form $E[s_i^2[k]|s_k] = E[s_i[k]|s_k] \neq p_i^2[k]$. Observe that $(v\mathbf{1}^\top) \circ I = \text{diag}(v)$, where $\text{diag}(v)$ is the diagonal matrix whose entries are the elements of the vector v. Indeed, let us represent M_k as

$$M_k \equiv N_k + Q_k \equiv I \circ M_k + (\mathbf{1}_M - I) \circ M_k.$$

[12] This is simply one of many comparisons that could be made between the contact process and the influence model; other important comparisons include the existence of different extinction time regimes in the influence model and cluster distribution and persistence.



(a) **(b)** **(c)** **(d)**

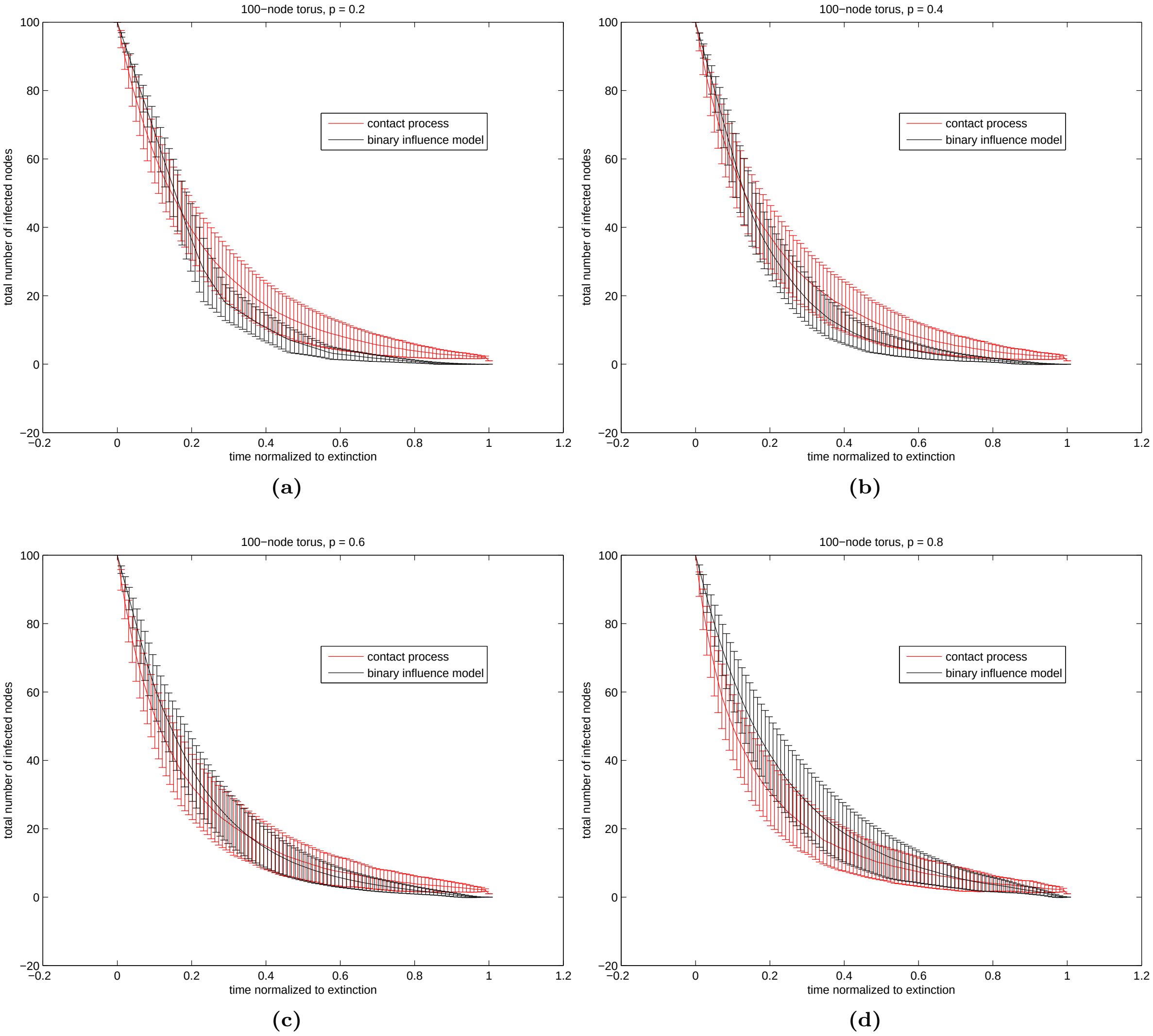
Figure 6.16. Mean and standard deviation of the number of infected nodes for the 128-node hypercube in the influence model and the contact process. The value of p for the influence model is fixed, and $\lambda = p/m$, where m is the degree of all nodes (here, $m = 7$).

If we take the expectation of Eq. 6.6, we obtain

$$E[M_{k+1}] = (\mathbf{1}^\top D^\top (I \circ E[M_k])\mathbf{1}) \circ I + (D^\top E[M_k]^\top D) \circ (\mathbf{1}_M - I).$$

This implies the following expressions for $E[N_{k+1}]$ and $E[Q_{k+1}]$

$$\begin{aligned} E[N_{k+1}] &= I \circ E[M_{k+1}] = I \circ (D^\top (I \circ E[M_k])\mathbf{1}\mathbf{1}^\top) = I \circ (D^\top E[N_k]\mathbf{1}\mathbf{1}^\top) \\ E[Q_{k+1}] &= (\mathbf{1} - I) \circ E[M_{k+1}] = (\mathbf{1}_M - I) \circ (D^\top (I \circ E[M_k] + (\mathbf{1}_M - I) \circ E[M_k])D) \qquad (6.7) \\ &= (\mathbf{1}_M - I) \circ (D^\top E[N_k + Q_k]D). \end{aligned}$$



(a) **(b)** **(c)** **(d)**

Figure 6.17. Mean number of infected nodes for the 100-node torus in the influence model and the contact process. The value of p for the influence model is fixed, and $\lambda = p/m$, where m is the degree of all nodes (here, $m = 4$).

If we apply the vec$(\cdot)$ operation to both sides of Eqs. 6.7, we obtain

$$\begin{aligned}
\text{vec}(E[N_{k+1}]) &= \text{vec}(I) \circ ((\mathbf{1}_M \otimes D)\text{vec}(E[N_k])) = \text{diag}(\text{vec}(I))(\mathbf{1}_M \otimes D)\text{vec}(E[N_k]) \\
\text{vec}(E[Q_{k+1}]) &= \text{vec}(\mathbf{1}_M - I) \circ ((D^\top \otimes D^\top)(\text{vec}(E[N_k]) + \text{vec}(E[Q_k]))) \\
&= \text{diag}(\text{vec}(\mathbf{1}_M - I))(D^\top \otimes D^\top)(\text{vec}(E[N_k]) + \text{vec}(E[Q_k]))
\end{aligned}$$

To simplify the notation, denote $I^v = \text{diag}(\text{vec}(I))$ and $\overline{I}^v = \text{diag}(\text{vec}(\mathbf{1}_M - I))$. Observe that $I^v + \overline{I}^v = I$ and that $I^v\overline{I}^v = 0$. Recalling that $\text{vec}(M_k) = \text{vec}(N_k) + \text{vec}(Q_k)$, and that

$\text{vec}(N_k) = I^v \text{vec}(M_k)$, we have

$$\text{vec}(E[M_{k+1}]) = [I^v(1_m \otimes D)I^v + \overline{I}^v(D^\top \otimes D^\top)]\text{vec}(E[M_k]) \equiv W\text{vec}(E[M_k]). \tag{6.8}$$

As a simple example, consider the model depicted in Figure 6.18, known as the "evil rain" scenario [139]. The evil rain node, denoted by 1, is permanently infected and has the ability to infect any node that it influences. Similarly, there is a "recovery" node, denoted by 0. The presence of these special nodes ensures that the system will not reach a consensus state.

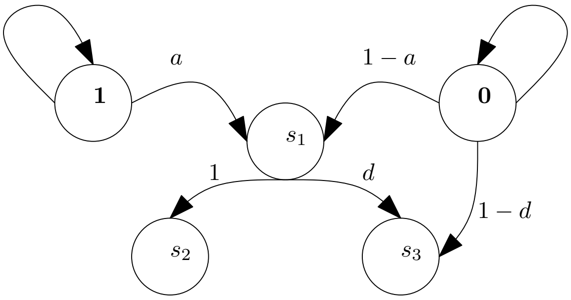


Figure 6.18. An "evil rain" influence model.

For this model,

$$D^\top = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ a & 1-a & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 1-d & d & 0 & 0 \end{bmatrix}.$$

In [139], Asavathiratham computes an analytical solution for the steady-state $E[s]$ for evil rain models, yielding

$$E[s] = \begin{bmatrix} 1 \\ 0 \\ a \\ a \\ da \end{bmatrix}$$

which appears on the diagonal in the steady-state solution for $E[M]$ given by Eq. 6.8:

$$E[M] = \begin{bmatrix} 1 & 0 & a & a & ad \\ 0 & 0 & 0 & 0 & 0 \\ a & 0 & a & a^2 & a^2d \\ a & 0 & a^2 & a & ad \\ ad & 0 & a^2d & ad & ad \end{bmatrix}.$$

Correlation results can reveal useful relationships between the node statuses; for example, nodes s_3 and s_1 are less correlated than nodes s_3 and s_2, even though s_1 is the only influencer of s_3.

Independently of its relationship with the contact process, the influence model has many potential applications in the public health setting. For example, tracking correlations might be very useful for food supply chain management and traceability. If a particular node (a transportation center, producer, restaurant, etc.) tests positive for some contaminant, which other nodes in the network are likely to be simultaneously contaminated? To answer this, we could apply a simple hypothesis testing framework. If we observe that site i has status 1 at time k (denoted by $s_i[k] = 1$), let H_0 denote the hypothesis that site j has status 0 ($s_j[k] = 0$), and H_1 denote the hypothesis that site j is in status 1 ($s_j[k] = 1$). The MAP rule tell us that we should choose H_0 if $E[s_i[k]]$ is greater than $2E[s_i[k]s_j[k]]$, and choose H_1 otherwise. Knowing the correlations between sites allows us to make this prediction.

Chapter 7
Conclusions

THIS thesis has explored several interesting issues in the mathematical modeling of infection processes; here, we shall summarize our work, highlight what we see as the key contributions, and suggest directions for future research.

7.1 Summary

Chapter 1 briefly surveyed the history of Western mathematical epidemiology, and introduced the *compartmental model* of disease transmission. As an example, we presented a related set of deterministic and stochastic models for the spread of a susceptible-infected-susceptible (SIS) disease, and noted the link between the two in the large-population limit. Chapter 1 also introduced the notion of a *threshold test* for whether a disease will become an epidemic, then considered the convergence of interest in epidemic prevention with research in network science to control the spread of disease through structured populations.

Chapter 2 established a general deterministic framework for infection modeling, based on the work of van den Driessche and Watmough [46], and explored the relationship between two of the most common threshold tests: whether the *basic reproductive ratio*, R_0 is greater than 1, and whether a *disease-free equilibrium* is locally asymptotically stable. We concluded that although these tests are equivalent in their epidemic predictions, they do not involve identical functions of the parameters of the model. Chapter 2 also compared these threshold tests with a third common test, the existence of an *endemic* equilibrium, using results from the literature to demonstrate different phenomena.

Chapter 3 began by considering how the choice of different mathematical functions to represent the *mixing* of two subpopulations can have a dramatic impact on the computation of R_0, then broadened the discussion to consider the central topic of the thesis: the computation of R_0 for populations that can be broken into smaller subgroups restricted to interact over the edges of a *network*. Under a common set of simplifying assumptions, we proved that R_0 can be expressed as the product of two factors: a "biology-based" factor R_h, and a "topology-based" factor $\rho(A)$, where A is the *adjacency matrix* representation of the network structure. The chapter concluded by demonstrating that many results in the literature (both canonical and more recent) are special cases of this decoupling of the biology of infection from the topology governing its spread.

Invoking the decoupling result of Chapter 3, Chapter 4 focused on the impact of topology via $\rho(A)$ and asked "What can be concluded about R_0 when the network structure is not completely known?" We showed that simply replacing an inherently random adjacency matrix $\mathbf{A}$ by its expected value $E[\mathbf{A}]$, can lead to *underestimating* the value of R_0, a problematic outcome in public health. For different types of partial information (e.g., a generation mechanism for the network, or a collection of network statistics), we presented two approaches for dealing with uncertainty: identifying *bounds* on $\rho(\mathbf{A})$ using spectral graph theory; and *approximating* $\rho(\mathbf{A})$ by making assumptions to fill in the missing information. These approaches were illustrated on several data sets, including preferential attachment graphs, the results of an egocentric social survey in a Houston community, the U.S. airline transportation network, and proximity detections from the Reality Mining project.

Continuing with the theme of a random $\mathbf{A}$, Chapter 5 considered a particular family of distributions that is widely used in quantitative sociology: the *exponential random graphs.* This chapter combined simulation and analytical work to explore the distribution of $\rho(\mathbf{A})$ for some of the most used members of this family, those that are parameterized by simple network statistics.

Finally, Chapter 6 moved beyond the focus on threshold tests for epidemics on networks to consider the *spatiotemporal patterns* of infection spread through these networks, on either side of the threshold. After surveying the literature on deterministic and stochastic models, we focused on two stochastic process: the contact process and the influence model. Through simulation and analytical approximations, we explored several spatial statistics, like number and size of infected clusters, which complement existing results to illuminate the underlying mechanisms of spread.

7.2 Contributions of thesis

Some highlights of our work are listed below:

- ▷ extends van den Driessche and Watmough's continuous-time framework [46] to discrete time models via difference equations;
- ▷ demonstrates the disconnect between the canonical "word" and "mathematical" definitions of R_0, and presents a better "word" definition (the asymptotic per generation growth rate);
- ▷ provides a clear interpretation of the relationship between R_0 and local asymptotic stability of the disease-free equilibrium;
- ▷ as an example, presents a new mathematical model for incorporating an arbitrarily-distributed infectious period;
- ▷ establishes the possibility of decoupling "biology" and "interaction patterns" in the computation of R_0 for a large class of models, thereby connecting several disparate results in the literature as special cases;

- ▷ clarifies the ways in which uncertainty has been *implicitly* embedded in deterministic models, and addresses the potential faults of these unstated assumptions;

- ▷ presents new tools for dealing with uncertainty in network structure via approximations and bounding, using spectral graph theory, and illustrates these techniques on four data sets;

- ▷ characterizes several members of the exponential random graph family by their spectral radius, via simulation and analytical work, thereby providing a first link between static network descriptions and the dynamic processes that unfold on the networks;

- ▷ investigates the spatiotemporal patterns of the contact process on several topologies, in the "slow die-off" and "fast die-off" regimes, via clustering statistics.

7.3 Directions for future work

Throughout the thesis, we have suggested additional areas of potential research and interesting open questions; we assemble these and additional thoughts here.

- ▷ Application to a case study. This thesis has focused on a very general class of infection models, but applying its results in a useful way to any particular disease could be of substantial public health benefit. Simply identifying the appropriate "network" on which an infection spreads is a non-trivial task, particularly when a population is partitioned into groups of varying contact levels, susceptibility and transmissibility.

- ▷ Generalizing R_0. In Appendix C, we suggest additional mathematical criteria that generalize the notion of "spectral radius"; in particular, these alternatives are useful when a population's interaction pattern varies over time. One does not have to look far for examples of this kind of phenomenon: the Reality Mining data set is a ready candidate. Exploring the utility of these generalizations to infection progression, both through analytical approximations and simulation work with time-varying data sets, would be an entirely new contribution to mathematical epidemiology.

- ▷ Identifying "better" summary statistics of the distribution of R_0. As a first step in characterizing the distribution of the $\rho(\mathbf{A})$ that arise from members of the exponential random graph family in Chapter 5, we focused on calculating the mean and variance. As we argued in Chapter 4, however, the mean of the distribution of $\rho(\mathbf{A})$ (and more generally, the mean of $\mathbf{R_0}$) might not be the most epidemiologically useful summary statistic. For example, in estimation problems, means naturally arise when one is interested in minimizing a *mean-square-error* criterion, but medians arise when the *minimum absolute error* is to be minimized. It would be interesting to investigate, through analytical and simulation work, the information conveyed by *other* summary statistics on the behavior of infection spread.

- ▷ Exploring the issue of degeneracy in ERGM models by looking at the distribution of $\rho(\mathbf{A})$. In Chapter 5, our characterization of $\rho(\mathbf{A})$ was limited to the study of means and standard deviations, but as we've just discussed, these are limited summary statistics. Exploring how the shape of the distribution of the spectral radius changes as the parameters change might shed light on the fundamental mechanisms behind the graph ensembles, particularly in parameter ranges around the onset of degenerate behavior.

- ▷ Relating classical stochastic models of infection and the analogous influence model. The tractability of the influence model, discussed in Chapter 6, suggests that it might be a useful approximation for the behavior of some of the canonical stochastic models of infection (e.g., the contact process) on finite graphs. For example, an "evil rain" model like the one described in Section 6.6.2 has a steady state with a non-zero level of infection that could be compared to the quasi-stationary state seen in the contact process in the "slow die-off" regime.

- ▷ Uncovering the limitations of R_0 through spatial analysis. As demonstrated in Chapter 2, a threshold test on R_0 only provides information about the growth factor of the infection in a completely susceptible population. This does not mean that two networks with the same $\rho(A)$ will exhibit identical infection trajectories, even in the initial phases of spread! Constructing a collection of non-isomorphic graphs with the same $\rho(A)$ and exploring the differences in their spatial patterns of infection would provide particularly useful information about the range of possible behaviors that can be exhibited by systems whose R_0 value is identical.

Appendix A

Approaches in mathematical modeling of infection processes

SECTIONS A.1 and A.2 will describe the most common approaches to the mathematical modeling of infection and will highlight some recent results in the field for both deterministic and stochastic models. Section A.3 will consider particular results for models that operate on an underlying contact network.

A.1 Deterministic approaches

The canonical text by Anderson and May presents a tremendous number of permutations of the deterministic differential equations model presented in Chapter 1, including extensions that incorporate the age structure and demographic details of a population [38]. Lloyd [141], Wearing et al. [51] and Keeling and Grenfell [142] (among others) address the assumption that recovery from infection is 'memoryless,' which leads to an exponentially-distributed infectious period. These authors explore the consequences of more realistic assumptions on the duration of the infectious period. Dodds and Watts present what they call a "threshold model," in which an individual receives a dose of infection from each contact and maintains a finite-length memory of such doses; if the sum of all doses received in the memory window exceed a threshold amount, that individual becomes infected [143]. Fraser et al. construct an infection model that allows the time variation of the infectiousness of an individual and that individual's likelihood to exhibit symptoms, measured since the onset of infection [29]. They use this model to explore the efficacy of contact tracing as a public health intervention. Significantly, they introduce a parameter θ that represents the proportion of secondary infections caused by an initial infective *before* the onset of symptoms. The larger this fraction θ, the more difficult it is for public health officials to identify and treat affected individuals before they've spread the disease.

For an extensive and thorough survey of deterministic epidemic models, see Hethcote [144].

A.2 Stochastic approaches

Most stochastic models of disease transmission fall into one of two categories. The first are those that approximate the spread of infection as a *branching process*, one in which new "branches" of a tree are created for each new infected individual (in a manner similar to Figure 2.1). One of the earliest stochastic models of disease spread was the *chain-binomial* model, first presented in a series of lectures by Lowell Reed and Wade Hampton Frost at Johns Hopkins University in 1928, which operates on a finite population in discrete time units representing generations of infection (see the discussion by Daley and Gani in [64]). At each time step, a susceptible individual has probability β of being infected by any given infective individual. These infections occur independently with each susceptible-infective pair. At the end of the time step, the current infectives are removed from the population and the new infectives emerge to infect a new generation of susceptibles. If there are S_k susceptibles and I_k infectives at time step k, then the number of infectives at time step $k+1$ is a binomial random variable over S_k trials with probability of success $1-(1-\beta)^{I_k}$. Some closed-form results can be obtained for this type of model, but its behavior when the number of infectives is small compared to the number of susceptibles is most often approximated by a discrete-time branching process; see Andersson and Britton for more detail regarding this approach [145].

The second category of stochastic model comprises continuous-time Markov processes, similar to the SIS example presented in Section 1.3. As suggested in Chapter 1, these models can exhibit behaviors that their deterministic counterparts cannot. In simulations, it is often seen that the number of infected individuals will fluctuate around the endemic equilibrium of its deterministic counterpart for a long period of time before the infection dies out. This behavior is referred to as the *quasi-stationary state* of the system, and is quantified by examining the behavior of the Markov process $X(t)$ conditioned on non-absorption in the all-susceptible state. This conditioned Markov process, $\widetilde{X}(t)$, cannot be solved in closed form, so several approximations are commonly used. The first is to modify the process $X(t)$ by eliminating the possibility of transitioning from one infected individual to zero infected individuals. As the population size $N \to \infty$, the equilibrium distribution of this new Markov process converges to the equilibrium distribution of the conditioned process $\widetilde{X}(t)$, and is sharply peaked at $I = (1-\gamma/\beta)N$, the deterministic endemic equilibrium [10]. Nåsell discusses another method of approximation: the modification of the process $X(t)$ to maintain one permanently infected individual [146] (much like the "evil rain" influence model discussed in Section 6.6.2). Nåsell then goes on to approximate this modified process by a normal distribution when R_0 is distinctly larger than 1 (centered at the deterministic endemic equilibrium) and a geometric distribution when R_0 is distinctly less than 1. Nåsell also identifies a transition region for R_0 between the two types of behaviors, which shrinks as $N \to \infty$. Srivastava presents simulation results of these dynamics for a model of virus systems [147].

In Chapter 1, we invoked Kurtz' theorem to approximate the expected value of the state variables by the deterministic predictions. However, if we choose to write down the differential equation for the mean of $X(t)$, we find that the nonlinear terms required by the standard incidence model (e.g., the product SI) introduce second moments into the expression. Similarly, any dynamic equation for second moments will involve third moments, and so on. Thus, it is impossible to construct a closed system of equations for any of the moments. One technique for dealing with this phenomenon is the use of *moment-closure* methods. These methods assume *a priori* a type of distribution for the moments, and use the moment properties of the distribution to specify higher-order moments in terms of lower-order moments, thus closing the system of equations. For example, if one assumes that S and I are distributed as a multivariate Gaussian, then the third-order central moments are zero, allowing one to close the equations at second-order [148]. Keeling et al. have proposed the use of *multiplicative moments* [149]: for example, the multiplicative second moment $\hat{V}$ of a random variable X is defined by

$$E[X^2] = E[X]^2 + E[(X - E[X])^2] \equiv E[X]^2\hat{V}.$$

Keeling et al. demonstrate that the assumption that all third-order and higher multiplicative moments are 1 is equivalent to assuming that the random variable follows a log-normal distribution, which has a non-negative support. Krishnarajah et al. suggest a beta-binomial distribution, which has both a non-negative support and an upper bound [150]. The authors also employ a mixed distribution with a non-zero probability mass at zero infectives to model the probability of extinction of the disease, plus a probability distribution on the positive integers to model the quasi-stationary state.

Isham's 2004 survey provides an excellent overview of the state of stochastic epidemic modeling [151]. For additional interpretations of "epidemic" in stochastic models, see the work of Newman [126] and Miller [152].

A.3 Network models

We concluded Chapter 3 by discussing some threshold results from the literature on network epidemiology and continued this discussion in Chapter 6. This section will highlight some of the other contemporary work in this field.

Barthelemy et al. make a significant contribution to the literature by building upon the scale-free network results of Pastor-Satorras and Vespignani discussed in Section 3.3.1 to explore the time scales over which an epidemic will occur [153]. Using an SI model, the authors find that the time scale of epidemic outbreak is inversely proportional to the skewness of the degree distribution (i.e., the magnitude of $\frac{<k^2>}{<k>}$). The authors also observe a characteristic pattern to disease spread in scale-free networks, from high-degree nodes to low-degree nodes after an initial transient period that depends on where the infection originates.

As discussed in Section A.2, differential equations for the moments of the distribution of the number of infected individuals in stochastic models require the knowledge of higher-order moments. In order to close a finite number of these equations, some approximation technique must be applied. When the population under study has a network structure, one can include information about the network topology by observing the 'moments' of the connection pattern: counting the number of connected pairs of nodes for each combination of disease states. Note that these are still deterministic models; the moments here are those that describe the degree distributions and correlations for different disease states. Let $[S_i]$ denote the number of susceptible individuals with i connections to other individuals (i.e., the number of nodes of degree i), and let $[S_i I_m]$ denote the the number of connected pairs of degree-i susceptible individuals with degree-m infected individuals. A common formulation of this problem is given by Roy and Pascual in [66]:

$$\frac{d[S_i]}{dt} = -\beta \sum_m [S_i I_m] + \gamma [I_m] \tag{A.1}$$

In [66], the authors approximate the $[S_i I_m]$ terms using powers of $[S_i]$ and $[I_m]$. Keeling et al. close their set of differential equations at third order with a single parameter that measures the proportion of complete triangles in the network [154]. Sharkey et al. present a more complicated moment-closure approximation that applies to directed networks [155].

Read and Keeling address a different phase of infection dynamics in [156]; how will the characteristics of an infection change if it is allowed to evolve as it progresses through a structured population? The authors perform simulations on several different types of networks and observe evolutionary trends in the disease parameters (infectivity and duration of infectious period).

In a final example, Watts et al. model the population as a set of subgroups that interact with each other stochastically via a 'metapopulation' structure [157]. The researchers use simulations to demonstrate that this kind of model can produce the kind of periodic re-emergence of disease that is seen in many populations, and can provide an intuitive way of structuring a population that is suitable for large-scale simulations.

Appendix B
Network data sets

TO validate an epidemiological model, one needs empirical data. While data collection in any field presents challenges, acquiring data on networks of actors compounds these challenges. Occasionally, network data is readily available from a governing agency or central data source; for example, one can assemble networks of corporate board members from tax disclosures. Social networks, however, are not as easily observed; to record relationship patterns in a community, one needs to either interview individuals about their behavior, or extensively monitor interactions. Many different approaches to studying social networks via individual surveys have been proposed and explored in the literature; for a comprehensive overview, see [67].

In order to explore the applications and limitations of the techniques discussed in this thesis, we have assembled network data sets representing several different kinds of data acquisition methods.

1. The volume of passenger flow through the U.S. airline transportation network in January 2007.
2. Proximity relationships of individuals participating in the MIT Media Lab Reality Mining study of 2004.
3. Self-reports of local social networks from a community in Houston.

The remainder of this appendix will describe how these data sets were acquired and processed.

B.1 U.S. airline transportation snapshot

The Bureau of Transportation Statistics, an organization under the U.S. Department of Transportation, is charged with the regular collection of data on many modes of transportation, including aviation, maritime, highway, public transit, rail and pedestrian/bike traffic. Much of this data has been made available to the public at `http://www.transtats.bts.gov`, and can be aggregated and exported for individual use.

In this study, we focus on domestic airline flights, as reported by both domestic and international carriers in the Air Carrier Statistics database via the T-100 reporting form. In particular, we examine the volume of passenger flow between U.S. cities over the month of January in 2007.[1] The raw data obtained from BTS consisted of 21633 records for the specified period (in the rows of a CSV file), each of which contained:

[1] The smallest time resolution available from this data set is 1 month.

▷ the origin airport code and city (name and number)

▷ the destination airport code and city (name and number)

▷ the number of passengers recorded from the origin to the destination airport aggregated over the specified period

Since multiple airlines service any given pair of cities, the data typically contained several entries for each airport pair, each reported by a different airline. The first step in processing this data was to combine the reports from different airlines to obtain total flow volumes for each airport pair, which was accomplished using the 'roll-up' command offered in the DigDB set of add-in tools for Microsoft Excel.[2] The roll-up command was applied a second time to combine multiple airports within a single city. Finally, all of the city pairs with zero recorded passenger traffic were removed from the data set, leaving 9986 directed city pairs with nonzero flow.

To accompany this data, we also assembled population statistics for many of the cities in the data set. We began by sorting the city pairs by descending traffic volume, and filled in population information until the top 100 cities had been identified. Most of these cities were listed in the U.S. Census Bureau's *Annual Estimates of the Population for Incorporated Places Over 100,000, Ranked by July 1 2006, Population: April 1, 2000 to July 1, 2006.* Matching the cities described in this document to the cities described in the air traffic data required some manual tuning: for example, the Census Bureau's entries for 'Minneapolis' and 'St. Paul' were combined to match the airline joint designation 'Minneapolis/St. Paul'. To fill in the remaining cities, we selected the most recent population statistics from each city's Wikipedia entry (typically from U.S. or state projections).

In a final step of data processing, we removed an interesting anomaly from the air traffic data: one of the highest volume routes in the U.S., which connects Kahului and Honolulu, HI. Because of the volume of tourists using this route to move between Hawaiian islands, its total flow over the month of January was several times the populations of both cities combined. Clearly, any disease model which incorporates mixing the between the passengers moving between cities and the city populations themselves would have to address this kind of mixing distinctly from a city like New York or Chicago, and thus we chose to remove these cities from the larger data set.

A sample of the data from the highest volume routes is given in Table B.1.

B.2 Reality Mining proximity data

The Reality Mining Project is the product of a collaboration between Nathan Eagle and Alex Pentland at the MIT Media Laboratory. In 2004, the researchers distributed 100 Nokia 6600 smartphones to members of the MIT community, each of which was able to detect and record:

[2] DigDB provides many useful extensions to Excel functionality: see `http://www.digdb.com`.

Table B.1. Top 5 U.S. cities with most total inter-city air traffic over January 2007. The ij^{th} entry corresponds to air traffic from city i to city j. The diagonal entries of the matrix are the city populations.

	New York, NY	Chicago, IL	Orlando, FL	Atlanta, GA	Ft. Lauderdale, FL
New York, NY	8214426	125833	80336	93778	111792
Chicago, IL	124729	3849378	70582	76853	40681
Orlando, FL	81825	74127	2833321	121363	20140
Atlanta, GA	96252	81528	114146	2144491	82491
Ft. Lauderdale, FL	114889	41065	21775	87334	1512986

▷ phone calls - start and end times, the other participant in the call, and the ID of the cell tower through which the call was routed;

▷ phone activity - on/off status and application usage;

▷ Bluetooth devices within 5-10 m - time of detection and a Bluetooth ID for the detected device.

These detections were recorded and sent to a central server over the nine-month course of the study. While any active Bluetooth device could be detected, we are particularly interested in detecting the proximity of the phones of other study participants; these proximity detections establish a time-dependent network of these users' (potential) physical interactions. Proximity data is very useful for predicting the spread of infections that can be transmitted by common handling of the same object (like a doorknob or public computer) or via inhalation of airborne droplets. For example, the infectious period for the common cold (a designation that comprises many particular viral infections, including rhinovirus, coronavirus and influenza) begins roughly one day prior to the onset of symptoms and continues for roughly 5 days after symptom onset (see the entry for "respiratory disease, acute viral" in [158]). During this roughly weeklong period, an infectious individual could infect anyone that he or she came into contact with. Thus, for this analysis, we extracted a week's worth of data on this naturally time-varying social network.

```
select s.starttime, s.person_oid, v.person_oid
from devicespan s inner join device v
on v.oid = s.device_oid
where v.person_oid > 0
and s.starttime > ‘2004-11-15 00:00:00’
and s.endtime < ‘2004-11-22 00:00:00’
into outfile ‘week1115.txt’
```

Figure B.1. A SQL query of the Reality Mining data set.

The data set is packaged as an SQL database, whose organizational structure is described in [159].[3] To extract information from the database, we installed MySQL Server 5.0 on a personal computer, loaded the database, and ran queries via a terminal window. A sample query is given in Figure B.1, which performs the following task:

1. compares the two database tables that
   - ▷ associate users with devices
   - ▷ associate Bluetooth detections with devices
2. matches the unique device ID numbers between the two, then
3. extracts all of those devices which
   - ▷ belong to a study participant *and*
   - ▷ represent interactions between November 15, 2004 and November 22, 2004.

This particular week of data was chosen to follow an initial period of difficulty with the memory storage on some of the phones in the study, which resulted in a loss of data from several users during the months of September and October. As Eagle and Pentland note, this proximity data is certainly not a perfect representation of the interaction patterns of study participants. Since the RF Bluetooth signals are able to penetrate walls, false proximity detections are likely recorded. Additionally, to conserve battery life, Bluetooth device scans were performed only once every five minutes, rather than continuously. Additionally, there are certainly interactions that went unrecorded when participants turned their phones off during certain activities, allowed the batteries to run down, or forgot to bring their phones with them [160].

B.3 Social contacts in a Houston community

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

[3] The complete data set is available to the public at `http://reality.media.mit.edu/`.

Table B.2. A summary 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

by Affiliated Systems Corporation and is described in [88], [89], and [90]. As part of the survey, participants were asked to name up to 18 other individuals who were a part of their social network, and describe the nature of their relationships. The participants were also asked to assess whether these individuals knew *each other*, which illuminates local subgraphs of the larger social network of this community. The researchers then attempted to link up these disconnected networks by matching named individuals who were mentioned by multiple participants, or who were participants themselves (the 'partner identification' step).

Dr. Isaac Montoya of Affiliated Systems Corp. provided us with SPSS files of the data collected during intake and followup interviews with participants. SPSS is a proprietary software package used primarily in the social sciences for aggregating and analyzing experimental results, and stores variables which includes a participant ID number as well as answers to all of the survey questions. To extract the relevant data, we used SPSS 16.0 to export the social network information to an Excel spreadsheet. Unfortunately, the accompanying documentation which describes the details of the data collection procedure are missing, as are the partner identifications that connect participants to each other. As a result, we are limited to analyzing the local networks of study participants, which allows us to test out disease prediction strategies that are limited to local network data information.

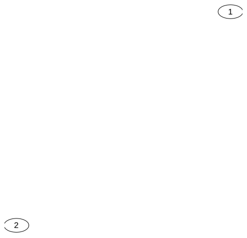
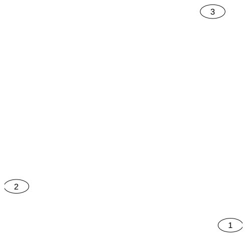
The data files identified 884 records of participants, six of which have the same identification number and which may represent the same individual or an error in number assignment. Although participants were asked to name up to 18 members of their social network (in three groups of six), most of the 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 between them. For any pair of named individuals A and B, participants were asked whether A know 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 B.2.

To aggregate the 640 non-trivial participant records, the following procedure was implemented in MATLAB:

1. Read in participant i's local network description.
2. Convert the description into an adjacency matrix A_i between named individuals.
3. Check the graph $G(A_i)$ for isomorphism against a list of graphs that have already been recorded, $\{G_1, \ldots, G_m\}$. If $G(A_i)$ is a previously unseen structure, add it to this list as G_{m+1} and set $count_{m+1} = 1$. If $G(A_i)$ is isomorphic to G_j, then increment $count_j$.

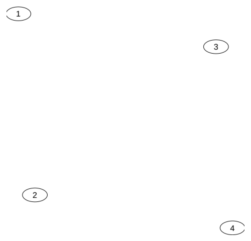
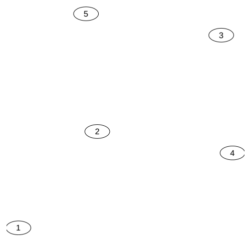
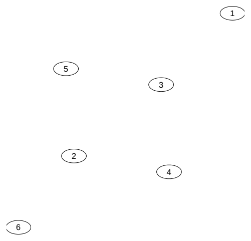
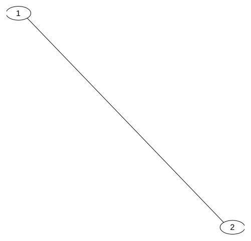
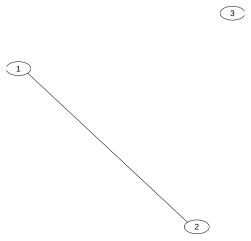
The results of this procedure are presented in Table B.3.

Table B.3. The list of local contact networks reported in the Houston data set.

local network structure	number of nodes	number of edges	frequency in data set
	1	0	166 (25.9%)
	2	0	91 (14.2%)
	3	0	54 (8.4%)

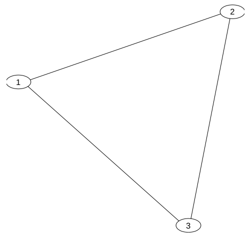
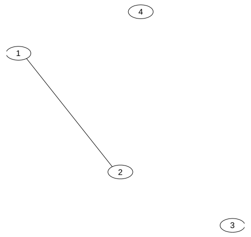
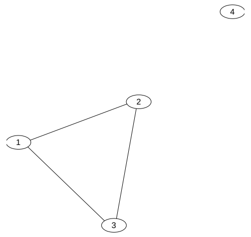
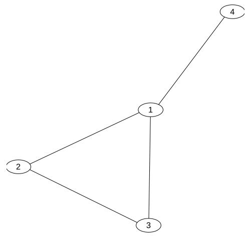
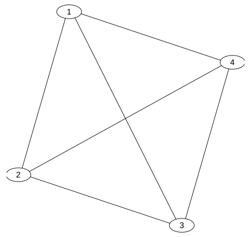
continued on next page

continued from previous page

local network structure	number of nodes	number of edges	frequency in data set
	4	0	34 (5.3%)
	5	0	26 (4.1%)
	6	0	30 (4.7%)
	2	1	61 (9.5%)
	3	1	24 (3.8%)

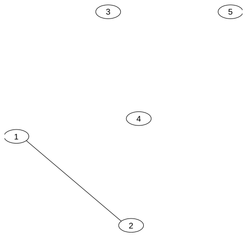
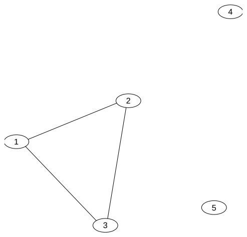
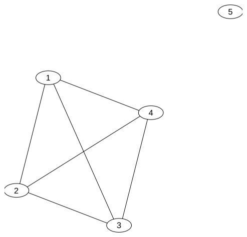
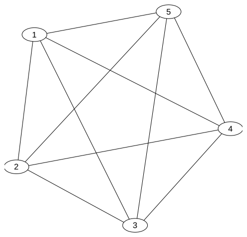
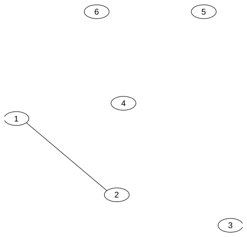
continued on next page

continued from previous page

local network structure	number of nodes	number of edges	frequency in data set
	3	3	34 (5.3%)
	4	1	20 (3.1%)
	4	3	8 (1.3%)
	4	4	1 (0.2%)
	4	6	9 (1.4%)

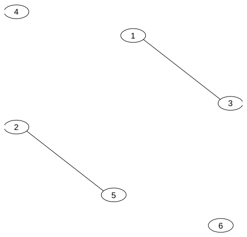
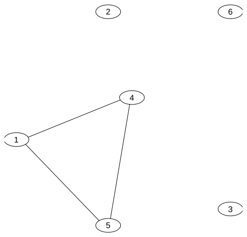
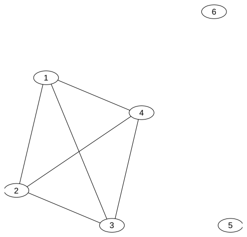
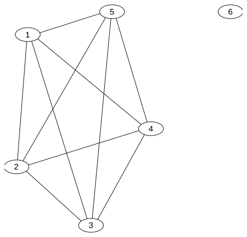
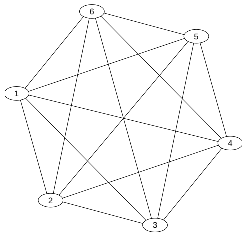
continued on next page

continued from previous page

local network structure	number of nodes	number of edges	frequency in data set
	5	1	7 (1.1%)
	5	3	9 (1.4%)
	5	6	5 (0.8%)
	5	10	3 (0.5%)
	6	1	17 (2.7%)

continued on next page

continued from previous page

local network structure	number of nodes	number of edges	frequency in data set
	6	2	1 (0.2%)
	6	3	21 (3.3%)
	6	6	8 (1.3%)
	6	10	3 (0.5%)
	6	15	8 (1.3%)

B.3.1 Generating clustered random graphs

The data from the Houston study discussed in the previous section provides local information regarding the structure of the social network in the community. Constructing a histogram of the number of contacts listed by each participant provides a measure of the *degree distribution* of the network, while counting the number of edges between contacts is a measure of *clustering* (the likelihood that two of a node's neighbors are connected to each other). There are several measures of clustering common in the literature: the first is the *clustering coefficient*, given by

$$C_\Delta = \frac{3N_\Delta}{N_T}$$

where N_Δ is the number of triangles in the graph and N_T is the number of *transitive triads* or 2-paths. If $C_\Delta = 1$, then every pair of nodes with a common neighbor is itself connected, while $C_\Delta = 0$ implies that there are no triangles in the graph at all. A related measure is 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 [91], we'll define the *average clustering coefficient* to be

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

Local information like that provided in the Houston data set allows us to compute $C = 0.3124$ over the participants in the study. These local measures are necessarily imperfect; participants were only able to list up to six contacts and had to make their best guesses about the relationships between them, but using them as approximations to the real network allows us to make some predictions.

In particular, given a degree distribution and an average clustering coefficient, what types of complete networks are possible? The generation of random graphs with specified properties has been a very active field of research over the last several years. One proposed method for producing a graph with a general degree distribution and clustering coefficient was developed by Volz in [92]. This algorithm begins by generating the desired number of nodes, and attaching to node i a number k_i of "stubs", where k_i is the desired degree of node i. Volz' algorithm then begins to connect stubs between nodes while maintaining the desired clustering coefficient, C_{input}. To generate a list of the desired degrees of the nodes in the network from a distribution, a Matlab script was written to repeatedly and independently sample node degrees from this distribution until the desired number of nodes had been generated: all of these degrees were compiled into a list, which could then be

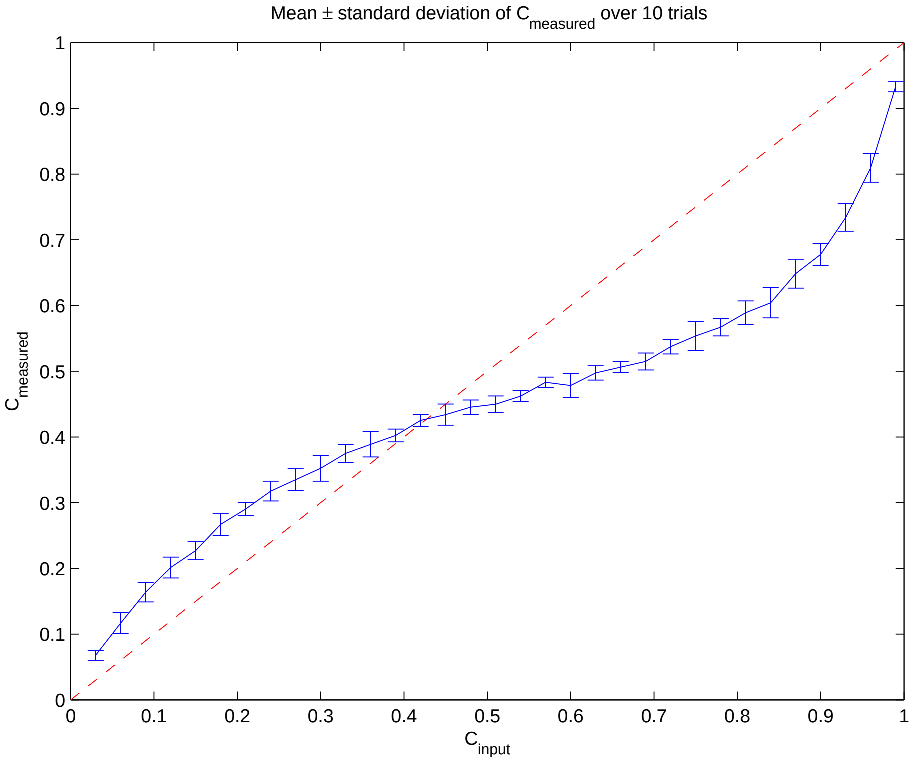


Figure B.2. Average clustering coefficient C v. C_{input} for the degree distribution of the Houston data set. Means and standard deviations were taken over 10 trials.

imported into a Java implementation of the Volz algorithm.[4]

Interestingly, Volz' algorithm appears to be able to generate networks of a specified degree distribution for any fixed $C \in [0, 1]$ (or C_{Δ}), but achieving this C might require specifying a different C_{input}. For example, the Houston data set has the degree distribution given by the blue bars in Figure B.3; when this degree distribution is provided to Volz' algorithm for a 1000 node network, the resulting C v. C_{input} is given in Figure B.2. This is a nonlinear, but bijective map, and is similar to the results for a Poisson random network that Volz documents in [92].

To obtain an average clustering coefficient of $C = 0.3124$ (the Houston data value), then, it appears that we should choose $C_{input} \sim 0.2375$. The red curves in Figure B.3 give the degree distributions of 50 random networks generated by `RandomClusteringNetwork.jar`, a Java executable. This program was run via the MS-DOS command line, and a sample command is given in Figure B.4. Some summary results for the remaining trials are presented in Table B.4.

[4] The author has made this executable available at `http://www.people.cornell.edu/pages/emv7/clustering/`.

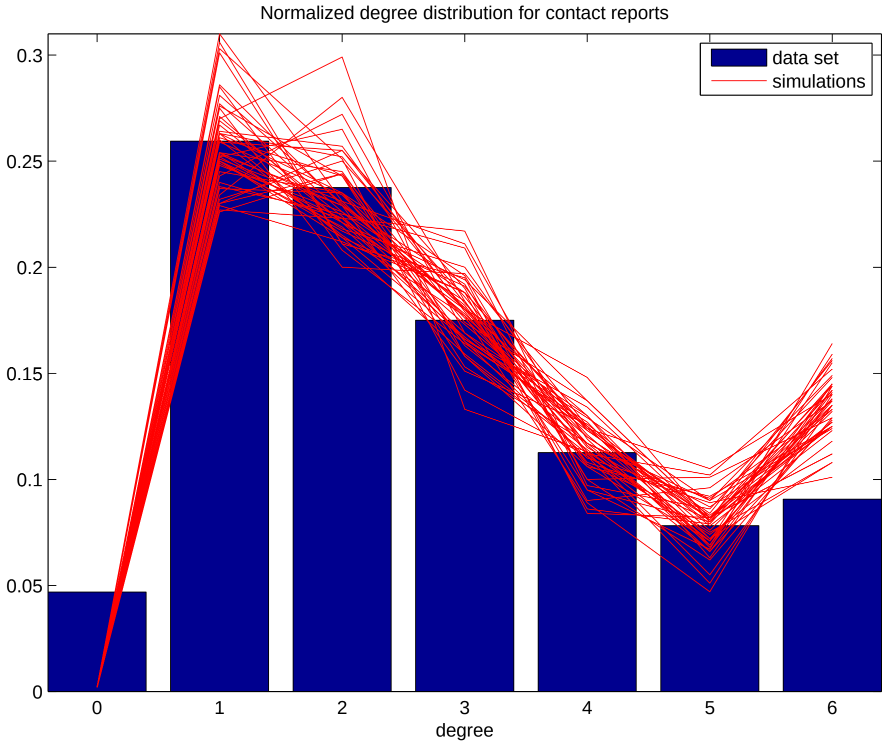


Figure B.3. Empirically-observed degree distribution in the Houston data set, and the resulting degree distributions of 50 networks generated via the Volz algorithm with $C_{input} = 0.2375$.

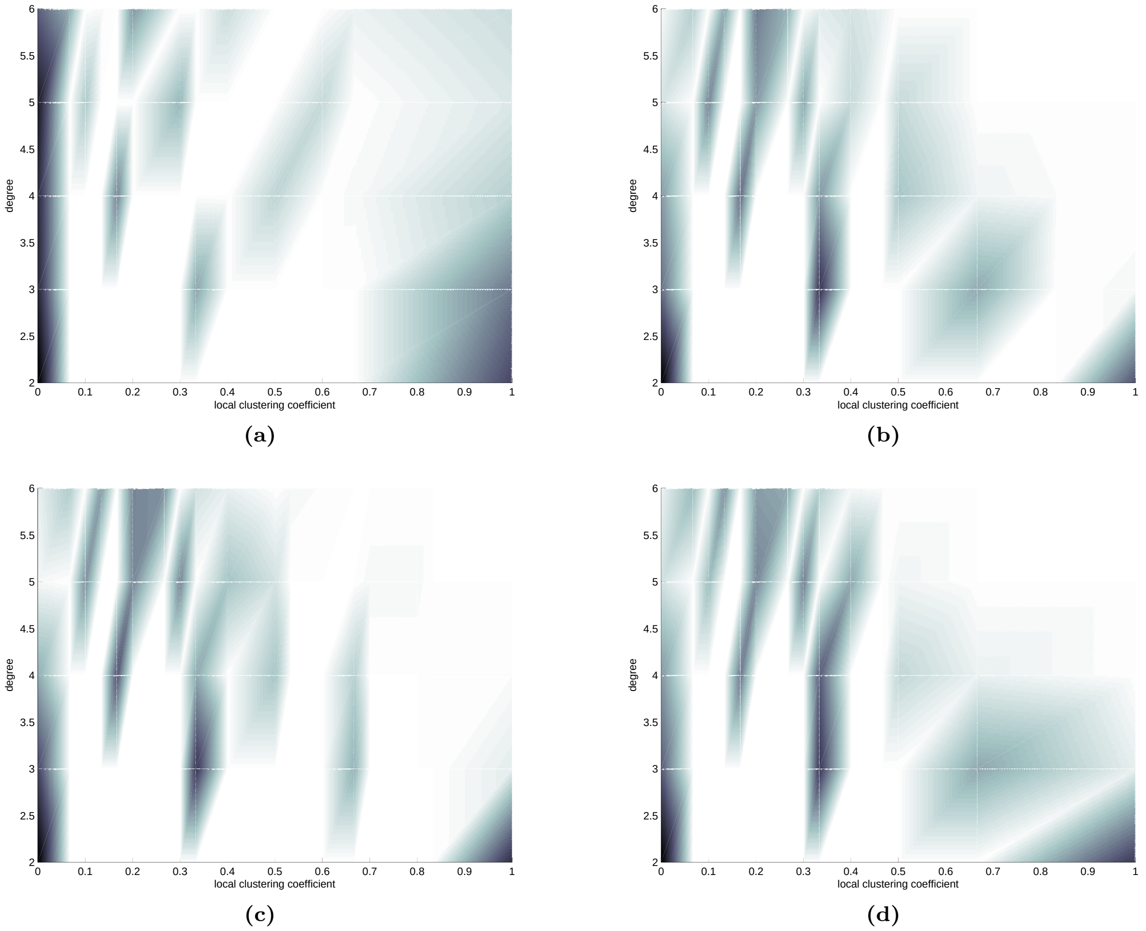
```
for /L %j in (1,1,50) do java -jar RandomClusteringNetwork.jar
     degrees%j.txt 1000 0.3 edgelist%j.txt
```

Figure B.4. MS-DOS command line execution of the `RandomClusteringNetwork.jar` executable. Here, `degrees%j.txt` is a tab-delimited file containing the desired degrees of each of the nodes in the network, 1000 is the number of nodes, 0.3 is the input clustering parameter C_{input} and `edgelist%j.txt` is a text file to which the edge list will be written. This code increments `%j` from 1 to 50 in increments (the middle argument) of 1.

Table B.4. Summary statistics for the 88 random graphs generated on 1000 nodes with the degree distribution and clustering observed in the Houston data set. 100 graphs were generated, but 12 of these resulted in nodes with degree greater than six, and were discarded.

statistic	**value**
max. C	0.395
min. C	0.292
$\overline{C} \pm \sigma_C$	0.338 ± 0.0245
max. number of edges, E	1545
min. E	137
$\overline{E} \pm \sigma_E$	1460 ± 39.964
max. $\rho(A)$	5.211
min.$\rho(A)$	4.373
$\overline{\rho(A)} \pm \sigma_{\rho(A)}$	4.741 ± 0.159
max. degree	6
min. degree	1

Our fundamental goal is to generate a distribution of random graphs from which the Houston data may have arisen. Are there any ways of validating whether or not the Volz algorithm produces such graphs? It would be useful if we had a third statistic, besides degree distribution and clustering coefficient, which we could compare between the Houston data and the Volz graphs; if these statistics matched, we'd be more confident that the Volz graphs were good approximations of the network from which the Houston data was derived. Indeed, we have such a statistic: the joint distribution of degree and local clustering coefficient. For each node of degree $i > 1$, the Houston data yields a histogram of the observed local clustering coefficient. We can compute such histograms for sample random graphs generated by the Volz distribution, and see if the two produce the same kind of behavior. A comparison of the Houston data and three sample Volz graphs is given in Figure B.5. The three Volz graphs have qualitatively similar distributions: a decreasing correlation between node degree and local clustering. The Houston data, however, seems to have the bulk of its probability density distributed over two separate regions: low and high local clustering for all degrees. In the context of how the Houston data was collected, this has an intuitive explanation; the contacts named by a participant are likely to have been randomly selected from that participant's social group (and thus not necessarily aware of each other) or they may all have been drawn from one group (where there are many interrelationships). Since the degree distribution decreases quickly after two contacts, participants may have not been especially motivated to give full accounts of their social relationships, which is a likely explanation for the bulk of the distribution in the region of low clustering.



(a) **(b)** **(c)** **(d)**

Figure B.5. Joint distribution of degree and clustering coefficient, in the original data set (a) and in three different simulations using the Volz algorithm (b)-(d). The (discrete) data has been interpolated to make trends easier to identify; black areas correspond to high probability regions, while white corresponds to low probability regions.

Appendix C

Possible generalizations of R_0

IN Section 2.3, we defined $R_0 = \rho(K)$, where K is the next-generation matrix, and observed that $\rho(K) = \lim_{n\to\infty} \|K^n\|^{1/n}$ for any matrix norm $\|\cdot\|$. We interpreted multiplication of the initial distribution of infectives ϕ by the matrix K as the creation of a new "generation" with infectives distributed as $K\phi$; thus, this model assumes that each new generation grows according to the same rules as the previous generation (i.e., by repeated left-multiplication with K). Continuing to explore the theme of uncertainty, it is likely that the population dynamics undergo some inherent *jitter*, such that K_i and K_{i+1} are slightly different as a result of stochastic phenomena. We can also imagine that K *changes* from generation to generation as individuals adjust their behavior seasonally, or in response to news of an impending epidemic. Is it possible to guarantee performance on threshold tests under these scenarios?

One might conjecture that as long as $\rho(K_i) < 1$ for every i, then $\lim_{s\to\infty} \rho\left(\prod_{i=0}^{s} K_i\right) < 1$ and there will be no epidemic. Unfortunately, this statement is not true in general. However, Hartfiel has demonstrated that if each $\rho(K_i) < 1$ *and* the matrices K_i do not change too quickly from one generation to the next, then the desired decay can be achieved [161].

Theorem C.0.1. From Theorem 12.1 of [161]. *Let ϕ_i denote the ith generation of new infections, and thus $\phi_{i+1} = K_i\phi_i$. If there exists a matrix norm $\|\cdot\|$ and $M_1, M_2 > 0$ such that $\|K_i\| \leq M_1$ and $\rho(K_i) \leq M_2 < 1$ for all $i \geq 0$, then there exists an $\epsilon > 0$ such that if*

$$\|K_{i+1} - K_i\| \leq \epsilon$$

for all i, then

$$\lim_{i\to\infty} \phi_i \to 0.$$

Next, consider a more general model of uncertainty. Let Σ denote a *bounded* set of matrices, i.e., there exists a constant M such that for some matrix norm $\|\cdot\|$ we have $\|A\| \leq M$ for all $A \in \Sigma$. What if the K_i are pulled at random from Σ; can we bound the R_0 of the result? What are the best- and worst-case scenarios? To explore this question, we'll begin by considering a possible generalization of the spectral radius. Define

$$\rho_k(\Sigma) = \sup\left\{\rho\left(\prod_{i=1}^{k} A_i\right) \mid A_i \in \Sigma\right\}$$

and

$$\rho(\Sigma) = \lim_{k\to\infty} \sup \left(\rho_k(\Sigma)^{1/k} \right).$$

The quantity $\rho(\Sigma)$ is called the *generalized spectral radius* of the set Σ; it represents a “worst-case” value of R_0 [162].

Recall that for a fixed K, $\rho(K)$ has an equivalent characterization: $\rho(K) = \lim_{n\to\infty} \|K^n\|^{1/n}$. Might this provide a second generalization of the spectral radius? Define the *joint spectral radius*, $\widehat{\rho}(\Sigma)$, by

$$\widehat{\rho}_k(\Sigma) = \sup \left\{ \left\| \prod_{i=1}^{k} A_i \right\| \mid A_i \in \Sigma \right\}$$

and

$$\widehat{\rho}(\Sigma) = \lim_{k\to\infty} \sup \left(\widehat{\rho}_k(\Sigma)^{1/k} \right).$$[1]

In [162], Berger and Wang demonstrate that these two generalizations are, in fact, equal.

There exist a number of bounds on this “worst-case” quantity; Hartfiel provides that

$$\widehat{\rho}(\Sigma) \leq \sup_{A\in\Sigma} \|A\|.$$

If Σ is a finite set, e.g. $\Sigma = \{A_1, \ldots, A_s\}$, Blondel and Nesterov obtain that

$$\frac{1}{s}\rho(A_1 + \ldots + A_s) \leq \widehat{\rho}(\Sigma) \leq \rho(A_1 + \ldots + A_s)$$

and obtain another characterization of the joint spectral radius:

$$\widehat{\rho}(\Sigma) = \lim_{i\to\infty} \rho(A_1^{\otimes i} + \ldots + A_s^{\otimes i})^{1/i},$$

where $A^{\otimes i}$ denotes the ith Kronecker power of A [163]. Computationally, computing the joint spectral radius of a set of matrices is difficult, but many approximation algorithms exist (although it has been shown that unless $P = NP$, no polynomial-time approximations exist [164]). Since we are only interested in assessing whether or not $R_0 < 1$, is our task any easier? In fact, it is unknown whether determining if $\widehat{\rho}(\Sigma) < 1$ is a decidable problem, and Blondel and Tsitsiklis have shown that assessing whether $\widehat{\rho}(\Sigma) \leq 1$ is undecidable [165].[2]

Finally, it is interesting to consider the “best-case” value of R_0 when the next-generation matrices K_i are chosen from a bounded set Σ; what is the smallest value of R_0 that can be obtained? Define

[1] In [161], Hartfiel proves that this quantity does not depend on the particular matrix norm chosen, so we omit it from the notation.

[2] A problem is *decidable* if there exists an algorithm to solve it that is guaranteed to halt in a finite number of steps for all possible inputs.

the *generalized spectral subradius* as

$$\rho_*(\Sigma) = \lim_{k\to\infty} \inf \left(\rho_k(\Sigma)^{1/k}\right)$$

and the *joint spectral subradius* as

$$\widehat{\rho_*}(\Sigma) = \lim_{k\to\infty} \inf \left(\widehat{\rho}_k(\Sigma)^{1/k}\right).$$

Czornik has demonstrated that $\rho_*(\Sigma) = \widehat{\rho_*}(\Sigma)$ [166].

Bibliography

[1] M. Whitehead. William Farr's legacy to the study of inequalities in health. *Bulletin of the World Health Organization*, 78(1), 2000.

[2] B. J. Becker. Plagues and people: infectious and epidemic disease in history, `http://eee.uci.edu/clients/bjbecker/PlaguesandPeople/index.html`, 2005.

[3] K. Dietz and J. A. P. Heesterbeek. Daniel Bernoulli's epidemiological model revisited. *Mathematical Biosciences*, 180(1-2):1–21, 2002.

[4] W.O. Kermack and A.G. McKendrick. A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society of London*, 115, 1927.

[5] R. Ross. An application of the theory of probabilities to the study of *a priori* pathometry - Part I. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, 92(638):204–230, 1916.

[6] R. Ross and H. P. Hudson. An application of the theory of probabilities to the study of *a priori* pathometry - Part II. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, 93(650):212–225, 1917.

[7] R. Ross and H. P. Hudson. An application of the theory of probabilities to the study of *a priori* pathometry - Part III. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, 93(650):225–240, 1917.

[8] T. Kurtz. Limit theorems for sequences of jump Markov processes approximating ordinary differential processes. *Journal of Applied Probability*, 8(2):344–356, 1971.

[9] T. Kurtz. The relationship between stochastic and deterministic models for chemical reactions. *Journal of Chemical Physics*, 57(7):2976–2978, 1972.

[10] 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.

[11] S. E. Chick, A. L. Adams, and J. S. Koopman. Analysis and simulation of a stochastic, discrete-individual model of STD transmission with partnership concurrency. *Mathematical Biosciences*, 166(1):45–68, 2000.

[12] R. Huerta and L. S. Tsimring. Contact tracing and epidemics control in social networks. *Physical Review E*, 66(5), 2002.

[13] L. J. S. Allen and P. van den Driessche. Stochastic epidemic models with a backward bifurcation. *Mathematical Biosciences and Engineering*, 3(3):445–458, 2006.

[14] F. Klebaner and O. Nerman. Autoregressive approximation in branching processes with a threshold. *Stochastic Processes and their Applications*, 51:1–7, 1994.

[15] J. Watkins. Consistency and fluctuation theorems for discrete time structured population models having demographic stochasticity. *Journal of Mathematical Biology*, 41:253–271, 2000.

[16] L. J. S. Allen and A. M. Burgin. Comparison of deterministic and stochastic SIS and SIR models in discrete time. *Mathematical Biosciences*, 163(1):1–33, 2000.

[17] D.J.P. Barker. *Practical epidemiology.* Churchill Livingstone, 1973.

[18] G. C. Daily and P. R. Ehrlich. Global change and human susceptibility to disease. *Annual Review of Energy and the Environment*, 21:125–144, 1996.

[19] P. Martens and L. Hall. Malaria on the move: human population movement and malaria transmission. *Emerging Infectious Diseases*, 6(2):103–109, 2000.

[20] A. G. Barbour and D. Fish. The biological and social phenomenon of Lyme disease. *Science*, 260(5114):1610–1616, 1993.

[21] D. J. Melnick, Y. K. Navarro, J. McNeely, G. Schmidt-Traub, and R. R. Sears. The Millennium Project: the positive health implications of improved environmental sustainability. *The Lancet*, 365(9460):723–725, 2005.

[22] Qanta A. Ahmed, Yaseen M. Arabi, and Ziad A. Memish. Health risks at the Hajj. *The Lancet*, 367(9515):1008–1015, 2006.

[23] S. Cohen, W. Doyle, D. Skoner, B. Rabin, and J. Gwaltney. Social ties and susceptibility to the common cold. *JAMA-Journal of the American Medical Association*, 277(24), 1997.

[24] R.A. Horn and C.R. Johnson. *Matrix Analysis.* Cambridge University Press, 1990.

[25] A. Ganesh, L. Massoulie, 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.

[26] 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.

[27] 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.

[28] 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.

[29] 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.

[30] M. Boguna and R. Pastor-Satorras. Epidemic spreading in correlated complex networks. *Physical Review E*, 66:047104, 2002.

[31] A. N. Hill and I. M. Longini. The critical vaccination fraction for heterogeneous epidemic models. *Mathematical Biosciences*, 181(1):85–106, 2003.

[32] 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.

[33] 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.

[34] 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.

[35] 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.

[36] 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.

[37] J. M. Hyman and J. Li. Differential susceptibility and infectivity epidemic models. *Mathematical Biosciences and Engineering*, 3(1):89–100, 2006.

[38] R. Anderson and R. M. May. *Infectious diseases of humans: dynamics and control*. Oxford University Press, 1991.

[39] R. Pastor-Satorras and A. Vespignani. Epidemic dynamics and endemic states in complex networks. *Physical Review E*, 63:066117, 2001.

[40] N. Masuda and N. Konno. Multi-state epidemic processes on complex networks. *Journal of Theoretical Biology*, 243(1):64–75, 2006.

[41] K. B. Blyuss and Y. N. Kyrychko. On a basic model of a two-disease epidemic. *Applied Mathematics and Computation*, 160(1):177–187, 2005.

[42] 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.

[43] 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.

[44] 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.

[45] 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.

[46] 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.

[47] J. A. P. Heesterbeek. A brief history of R_0 and a recipe for its calculation. *Acta Biotheoretica*, 50(3):189–204, 2002.

[48] Robert M. May, S. Gupta, and A. McLean. Infectious disease dynamics: what characterizes a successful invader? *Philosophical Transactions of the Royal Society of London, Series B*, 356:901–910, 2001.

[49] M. J. Keeling. The mathematics of diseases. *Plus Magazine*, (14), 2001.

[50] R. Larson. Revisiting R_0, the basic reproductive number for pandemic influenza. Technical Report ESD-WP-2008-10, 2008.

[51] H. J. Wearing, P. Rohani, and M. J. Keeling. Appropriate models for the management of infectious diseases. *PLoS Medicine*, 2(7):621–627, 2005.

[52] C.M. Kribs-Zaleta. Center manifolds and normal forms in epidemic models. In C. Castillo-Chavez, S. Blower, P. van den Driessche, D. Kirschner, and A. A. Yakubu, editors, *Mathematical approaches for emerging and remeerging infectious diseases: an introduction*, volume 125 of *The IMA Volumes in Mathematics and its Applications*, pages 269–286. Springer-Verlag, New York, 2002.

[53] S. Wiggins. *Introduction to applied nonlinear dynamical systems and chaos*, volume 2 of *Texts in Applied Mathematics*. Springer-Verlag, 1997.

[54] F. Brauer. Backward bifurcations in simple vaccination models. *Journal of Mathematical Analysis and Applications*, 298(2):418–431, 2004.

[55] 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.

[56] N. Chitnis, J. M. Cushing, and J. M. Hyman. Bifurcation analysis of a mathematical model for malaria transmission. *SIAM Journal on Applied Mathematics*, 67(1):24–45, 2006.

[57] T. Gross, C. J. D. D'Lima, and B. Blasius. Epidemic dynamics on an adaptive network. *Physical Review Letters*, 96(20), 2006.

[58] C. Simon and J. A. Jacquez. Reproduction numbers and the stability of equilibria of SI models for heterogeneous populations. *SIAM Journal on Applied Mathematics*, 52(2):541–576, 1992.

[59] 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.

[60] C. Castillo-Chavez, Z. Feng, and W. Huang. On the computation of R_0 and its role in global stability. In C. Castillo-Chavez, S. Blower, P. van den Driessche, D. Kirschner, and A. A. Yakubu, editors, *Mathematical approaches for emerging and remeerging infectious diseases: an introduction*, volume 125 of *The IMA Volumes in Mathematics and its Applications*, pages 229–250. Springer-Verlag, New York, 2002.

[61] 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.

[62] 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.

[63] 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.

[64] D.J. Daley and J. Gani. *Epidemic modelling: an introduction*, volume 14 of *Cambridge Studies in Mathematical Biology.* Cambridge University Press, Cambridge, 1999.

[65] H. W. Hethcote and P. van den Driessche. Some epidemiological models with nonlinear incidence. *Journal of Mathematical Biology*, 29:271–287, 1991.

[66] M. Roy and M. Pascual. On representing network heterogeneities in the incidence rate of simple epidemic models. *Ecological Complexity*, 3(1):80–90, 2006.

[67] M. Morris. *Network epidemiology: a handbook for survey design and data collection.* International Studies in Demography. Oxford University Press, 2004.

[68] L. Vaughan and M. Thelwall. Search engine coverage bias: evidence and possible causes. *Information Processing and Management*, 40(4):693–707, 2004.

[69] A. Barabasi and R. Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, 1999.

[70] F. Chung and L. Lu. *Complex graphs and networks*, volume 107 of *CBMS Regional Conference Series in Mathematics.* American Mathematical Society, 2006.

[71] F. Ball, T. Britton, and O. Lyne. Stochastic multitype epidemics in a community of households: estimation and form of optimal vaccination schemes. *Mathematical Biosciences*, 191(1):19–40, 2004.

[72] M. Draief. Epidemic processes on complex networks. *Physica A-Statistical Mechanics and Its Applications*, 363(1):120–131, 2006.

[73] 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.

[74] K. C. Das and P. Kumar. Some new bounds on the spectral radius of graphs. *Discrete Mathematics*, 281(1-3):149–161, 2004.

[75] 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.

[76] R.P. Stanley. A bound on the spectral radius of graphs with e edges. *Linear Algebra and Its Applications*, 67:267–269, 1987.

[77] 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.

[78] O. Favaron, M. Maheo, and J.-F. Sacle. Some eigenvalue properties in graphs (conjectures of Graffiti - II). *Discrete Mathematics*, 111:197–220, 1993.

[79] X. Zhang. Eigenvectors and eigenvalues of non-regular graphs. *Linear Algebra and Its Applications*, 409:79–86, 2005.

[80] 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.

[81] V. Nikiforov. Bounds on graph eigenvalues I. *Linear Algebra and Its Applications*, 420(2-3):667–671, 2007.

[82] 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.

[83] L. Collatz and U. Sinogowitz. Spektren endlicher grafen. *Abh. Math. Sem. Univ. Hamburg*, 21:63–77, 1957.

[84] H. Minc. *Non-negative matrices*. Wiley Interscience Series in Discrete Math and Optimization. John Wiley and Sons, 1988.

[85] A. Brauer. The theorems of Ledermann and Ostrowski on positive matrices. *Duke Mathematical Journal*, 24:265–274, 1957.

[86] J.K. Merikoski and A. Virtanen. The best possible lower bound for the Perron root using traces. *Linear Algebra and Its Applications*, 388:301–313, 2004.

[87] R. Pastor-Satorras and A. Vespignani. Epidemic dynamics in finite size scale-free networks. *Physical Review E*, 65:035108(R), 2002.

[88] D. C. Bell and R.A. Trevino. Modeling HIV risk. *Journal of Acquired Immune Deficiency Syndromes*, 22:280–287, 1999.

[89] 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.

[90] 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.

[91] T. Schank and D. Wagner. Approximating clustering coefficient and transitivity. *Journal of Graph Algorithms and Applications*, 9(2):265–275, 2005.

[92] E. Volz. *Random networks with tunable degree distribution and clustering*. M.S. Thesis, Dept. of Sociology, Cornell University, 2004.

[93] J. Park and M. E. J. Newman. Solutions of the two-star model of a network. *Physical Review E*, 70:066146, 2004.

[94] 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.

[95] 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.

[96] M. Handcock, D.R. Hunter, C. Butts, S. M. Goodreau, and M. Morris. ergm: A Package to Fit, Simulate and Diagnoze Exponential-Family Models for Networks. *Journal of Statistical Software*, 24(4), 2008.

[97] J.S. Brownstein, C.J. Wolfe, and K. D. Mandl. Empirical evidence for the effect of airline travel on inter-regional influenza spread in the united states. *PLoS Medicine*, 3(10), 2006.

[98] O. Frank and D. Strauss. Markov graphs. *Journal of the American Statistical Association*, 81(395):832–842, 1986.

[99] J. Besag. Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society Series B-Statistical Methodology*, 36(2):192–236, 1974.

[100] C. Anderson, S. Wasserman, and B. Crouch. A p* primer: logit models for social networks. *Social Networks*, 21:37–66, 1999.

[101] A. Willsky. Supplementary Notes No. 3, Notes for 6.972: Algorithms for Estimation and Inference, Massachusetts Institute of Technology, Fall 2006.

[102] M. Handcock. Assessing degeneracy in statistical models of social networks. Technical Report 39, University of Washington, 2003.

[103] T. Snijders. Markov chain Monte Carlo estimation of exponential random graph models. *Journal of Social Structure*, 3(2), 2002.

[104] Z. Burda, J. Jurkiewicz, and A. Krzywicki. Network transitivity and matrix models. *Physical Review E*, 69:026106, 2004.

[105] Allie Pearson. Confidence intervals for variance and standard deviation, `http://www2.selu.edu/Academics/Faculty/apearson/math241.htm`, 2007.

[106] Y. Ioannides. Random graphs and social networks: an economics perspective. Technical report, Tufts University, 2006.

[107] P. Erdos, D.J. Kleitman, and B.L. Rothschild. Asymptotic enumeration of K_n-free graphs. In *International Colloquium on Combinatorial Theory, Atti dei Convegni Lincei*, volume 2, pages 19–27, Rome, 1976.

[108] D.R. Hunter. Curved exponential family models for social networks. *Social Networks*, 29:216–230, 2007.

[109] T. Snijders, P. Pattison, G. Robins, and M. S. Handcock. New specifications for exponential random graph models. *Sociological Methodology*, 36(1):99–153, 2006.

[110] L. Rass and J. Radcliffe. *Spatial deterministic epidemics*, volume 102 of *Mathematical surveys and monographs*. American Mathematical Society, 2003.

[111] M. Barthelemy, A. Barrat, R. Pastor-Satorras, and A. Vespignani. Velocity and hierarchical spread of epidemic outbreaks in scale-free networks. *Physical Review Letters*, 92(17), 2004.

[112] G.S. Canright and K. Engo-Monsen. Spreading on networks: a topographic view. *Complexus*, 3:131–146, 2006.

[113] R. Durrett. Crabgrass, measles and gypsy moths: an introduction to modern probability. *Bulletin of the American Mathematical Society*, 18(2):117–143, 1988.

[114] R. Durrett and T. Liggett. The shape of the limit set in Richardson's growth model. *Annals of Probability*, 9(2):186–193, 1981.

[115] M. Deijfen and O. Haggstrom. Nonmonotonic coexistence regions for the two-type Richardson model on graphs. *Electronic Journal of Probability*, 11:331–344, 2006.

[116] M. Deijfen and O. Haggstrom. The pleasures and pains of studying the two-type Richardson model. Technical Report 2007:17, Stockholm University, 2007.

[117] S. Thompson and A. Rosenfeld. Isotropic growth on a grid. *Pattern Recognition*, 28(2):241–253, 1995.

[118] J. Fill and R. Pemantle. Percolation, first-passage percolation and covering times for Richardson's model on the n-cube. *Annals of Applied Probability*, 3(2):593–629, 1993.

[119] J. T. Cox and R. Durrett. Limit theorems for the spread of epidemics and forest fires. *Stochastic Processes and their Applications*, 30(2):171–191, 1988.

[120] G. A. Braga, R. Sanchis, and T. A. Schieber. Critical percolation on a Bethe lattice revisited. *SIAM Review*, 47(2):349–365, 2005.

[121] L. M. Sander, C. P. Warren, I. M. Sokolov, C. Simon, and J. Koopman. Percolation on heterogeneous networks as a model for epidemics. *Mathematical Biosciences*, 180:293–305, 2002.

[122] C. Borgs, J. T. Chayes, H. Kesten, and J. Spencer. The birth of the infinite cluster: Finite-size scaling in percolation. *Communications in Mathematical Physics*, 224(1):153–204, 2001.

[123] D. S. Callaway, M.E.J. Newman, S. H. Strogatz, and D.J. Watts. Network robustness and fragility: percolation on random graphs. *Physical Review Letters*, 85:5468–5471, 2000.

[124] M.E.J. Newman, S. H. Strogatz, and D.J. Watts. Random graphs with arbitrary degree distributions and their applications. *Physical Review E*, 64:026118, 2001.

[125] M. A. Serrano and M. Boguna. Weighted configuration model. In *CNET2004*, volume 776, pages 101–107. American Institute of Physics, 2004.

[126] M.E.J. Newman. The spread of epidemic disease on networks. *Physical Review E*, 66:016128, 2002.

[127] T. Kalisky and R. Cohen. Width of percolation transition in complex networks. *Physical Review E*, 73(3), 2006.

[128] M. J. Ferrari, S. Bansal, L. A. Meyers, and O. N. Bjornstad. Network frailty and the geometry of herd immunity. *Proceedings of the Royal Society B-Biological Sciences*, 273(1602):2743–2748, 2006.

[129] D.J. Watts and S. H. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393, 1998.

[130] M. A. Serrano and M. Boguna. Clustering in complex networks. I. General formalism. *Physical Review E*, 74:056114, 2006.

[131] M. A. Serrano and M. Boguna. Clustering in complex networks. II. Percolation properties. *Physical Review E*, 74(5):056115, 2006.

[132] T. Liggett. *Stochastic interacting systems: contact, voter and exclusion processes*. Springer-Verlag, 1999.

[133] R. Gouet, F.J. Lopez, and G. Sanz. Estimation of critical values in interacting particle systems. *Journal of Applied Probability*, 37(1):118–125, 2000.

[134] R. Durrett. The contact process, 1974-1989. In *Mathematics of Random Media*, volume 27 of *Lectures in Applied Mathematics*, pages 1–18. American Mathematical Society, 1991.

[135] C.E. Sandifer. Euler's solution of the Basel problem – the longer story. In R.E. Bradley, L.A. D'Antonio, and C.E. Sandifer, editors, *Euler at 300: An Appreciation*. Mathematical Association of America, 2007.

[136] R.P. Boas. Partial sums of infinite series and how they grow. *The American Mathematical Monthly*, 84(4):237–258, 1977.

[137] A. Stacey. The contact process on finite homogeneous trees. *Probability Theory and Related Fields*, 121:551–576, 2001.

[138] H. Andersson and B. Djehiche. A threshold limit theorem for the stochastic logistic epidemic. *Journal of Applied Probability*, 35(3):662–670, 1998.

[139] C. Asavathiratham. *The influence model: a tractable representation for the dynamics of networked Markov chains*. Ph.D., Massachusetts Institute of Technology, 2000.

[140] C. Asavathiratham, S. Roy, B. Lesieutre, and G. Verghese. The influence model. *IEEE Control Systems Magazine*, 21(6):52–64, 2001.

[141] A. L. Lloyd. Realistic distributions of infectious periods in epidemic models: changing patterns of persistence and dynamics. *Theoretical Population Biology*, 60:59–71, 2001.

[142] M. J. Keeling and B. T. Grenfell. Effect of variability in infection period on the persistence and spatial spread of infectious diseases. *Mathematical Biosciences*, 147(2):207–226, 1998.

[143] P.S. Dodds and D.J. Watts. A generalized model of social and biological contagion. *Journal of Theoretical Biology*, 232:587–604, 2005.

[144] H. W. Hethcote. The mathematics of infectious diseases. *SIAM Review*, 42(4):599–653, 2000.

[145] H. Andersson and T. Britton. *Stochastic epidemic models and their statistical analysis*. Springer, 2000.

[146] I. Nasell. On the quasi-stationary distribution of the stochastic logistic epidemic. *Mathematical Biosciences*, 156(1-2):21–40, 1999.

[147] R. Srivastava, L. You, J. Summers, and J. Yin. Stochastic vs. deterministic modeling of intracellular viral kinetics. *Journal of Theoretical Biology*, 218(3):309–321, 2002.

[148] A. L. Lloyd and V. A. A. Jansen. Spatiotemporal dynamics of epidemics: synchrony in metapopulation models. *Mathematical Biosciences*, 188:1–16, 2004.

[149] M. J. Keeling. Multiplicative moments and measures of persistence in ecology. *Journal of Theoretical Biology*, 205(2):269–281, 2000.

[150] I. Krishnarajah, A. Cook, G. Marion, and G. Gibson. Novel moment closure approximations in stochastic epidemics. *Bulletin of Mathematical Biology*, 67(4):855–873, 2005.

[151] V. Isham. Stochastic models for epidemics. Technical Report 263, University College London, 2004.

[152] J.C. Miller. Epidemic size and probability in populations with hetereogeneous infectivity and susceptibility. *Physical Review E*, 76:010101, 2007.

[153] M. Barthelemy, A. Barrat, R. Pastor-Satorras, and A. Vespignani. Dynamical patterns of epidemic outbreaks in complex heterogeneous networks. *Journal of Theoretical Biology*, 235(2):275–288, 2005.

[154] M. J. Keeling, D. A. Rand, and A. J. Morris. Correlation models for childhood epidemics. *Proceedings of the Royal Society of London Series B-Biological Sciences*, 264(1385):1149–1156, 1997.

[155] K. J. Sharkey, C. Fernandez, K. L. Morgan, E. Peeler, M. Thrush, J. F. Turnbull, and R. G. Bowers. Pair-level approximations to the spatio-temporal dynamics of epidemics on asymmetric contact networks. *Journal of Mathematical Biology*, 53(1):61–85, 2006.

[156] J. M. Read and M. J. Keeling. Disease evolution on networks: the role of contact structure. *Proceedings of the Royal Society of London Series B-Biological Sciences*, 270(1516):699–708, 2003.

[157] D.J. Watts, R. Muhamad, D. C. Medina, and P.S. Dodds. Multiscale, resurgent epidemics in a hierarchical metapopulation model. *Proceedings of the National Academy of Sciences of the United States of America*, 102(32):11157–11162, 2005.

[158] J. Chin. Control of communicable diseases manual. Technical Report 17th edition, American Public Health Association, 2000.

[159] M.J. Lambert. *Visualizing and analyzing human-centered data streams*. M.Eng., Massachusetts Institute of Technology, 2005.

[160] N. Eagle and A. Pentland. Reality mining: sensing complex social systems. *Personal and Ubiquitous Computing*, 10(4):255–268, 2006.

[161] D.J. Hartfiel. *Nonhomogeneous matrix products*. World Scientific Publishing, Singapore, 2002.

[162] M. Berger and Y. Wang. Bounded semigroups of matrices. *Linear Algebra and Its Applications*, 166:21–27, 1992.

[163] V.D. Blondel and Y. Nesterov. Computationally efficient approximations of the joint spectral radius. *SIAM Journal on Matrix Analysis and Applications*, 27(1):256–272, 2005.

[164] J.N. Tsitsiklis and V.D. Blondel. The Lyapunov exponent and joint spectral radius of pairs of matrices are hard - when not impossible - to compute and to approximate. *Mathematics of Control, Signals and Systems*, 10(1):31–40, 1997.

[165] V.D. Blondel and J.N. Tsitsiklis. The boundedness of all products of a pair of matrices is undecidable. *Systems and Control Letters*, 41(2):135–140, 2000.

[166] A. Czornik. On the generalized spectral subradius. *Linear Algebra and Its Applications*, 407:242–248, 2005.