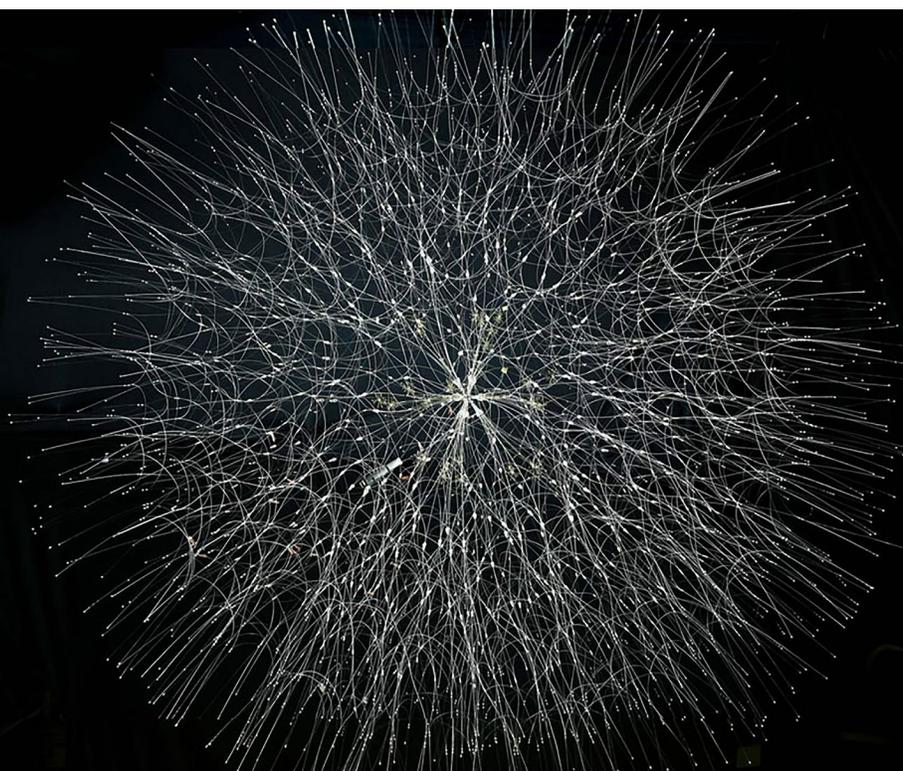


10

ALBERT-LÁSZLÓ BARABÁSI

NETWORK SCIENCE

SPREADING PHENOMENA



**ACKNOWLEDGEMENTS**

MÁRTON PÓSFAI  
NICOLE SAMAY  
ROBERTA SINATRA

SARAH MORRISON  
AMAL HUSSEINI  
PHILIPP HOEVEL

# INDEX

Introduction	1
Epidemic Modeling	2
Network Epidemics	3
Contact Networks	4
Beyond the Degree Distribution	5
Immunization	6
Epidemic Prediction	7
Summary	8
Homework	9
ADVANCED TOPICS 10.A	
Microscopic Models of Epidemic Processes	10
ADVANCED TOPICS 10.B	
Analytical Solution of the SI, SIS and SIR Models	11
ADVANCED TOPICS 10.C	
Targeted Immunization	12
ADVANCED TOPICS 10.D	
The SIR Model and Bond Percolation	13
Bibliography	14

**Figure 10.0 (cover image)**

**Bill Smith**

*An epidemiological model of the perfect infectious disease (evolved growth system) is an artwork by Bill Smith, an Illinois-based artist (2009, mixed media, 84x84x84 inches) (<http://www.widicus.org>).*



This work is licensed under a  
Creative Commons: CC BY-NC-SA 2.0.

PDF V26, 05.09.2014

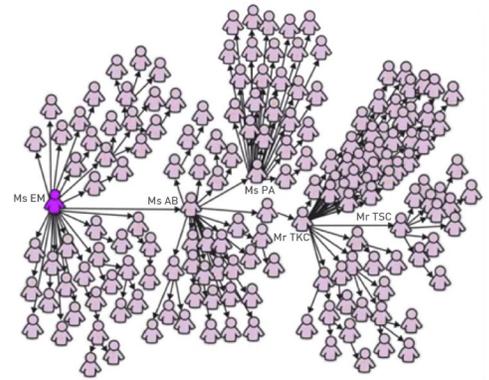
# INTRODUCTION

On the night of February 21, 2003 a physician from Guangdong Province in southern China checked into the Metropole Hotel in Hong Kong. He previously treated patients suffering from a disease that, lacking a clear diagnosis, was called *atypical pneumonia*. Next day, after leaving the hotel, he went to the local hospital, this time as a patient. He died there several days later of atypical pneumonia [1].

The physician did not leave the hotel without a trace: That night sixteen other guests of the Metropole Hotel and one visitor also contracted the disease that was eventually renamed Severe Acute Respiratory Syndrome, or SARS. These guests carried the SARS virus with them to Hanoi, Singapore, and Toronto, sparking outbreaks in each of those cities. Epidemiologists later traced close to half of the 8,100 documented cases of SARS back to the Metropole Hotel. With that the physician who brought the virus to Hong Kong become an example of a *super-spreader*, an individual who is responsible for a disproportionate number of infections during an epidemic.

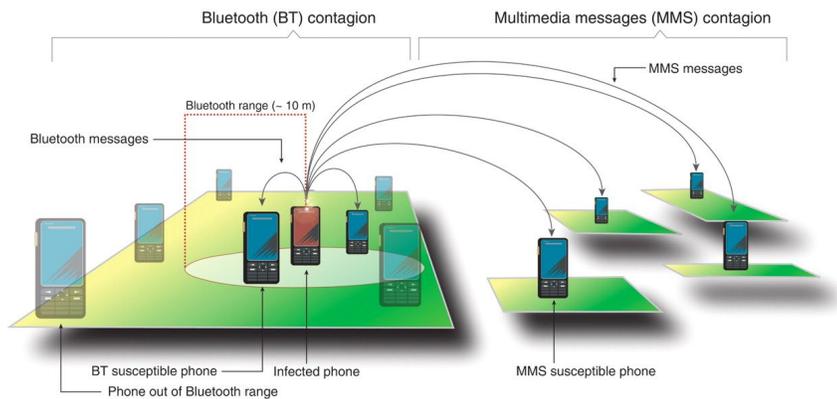
A network theorist will recognize super-spreaders as hubs, nodes with an exceptional number of links in the contact network on which a disease spreads. As hubs appear in many networks, super-spreaders have been documented in many infectious diseases, from smallpox to AIDS [2]. In this chapter we introduce a network based approach to epidemic phenomena that allows us to understand and predict the true impact of these hubs. The resulting framework, that we call *network epidemics*, offers an analytical and numerical platform to quantify and forecast the spread of infectious diseases.

Infectious diseases account for 43% of the global burden of disease, as captured by the number of years of lost healthy life. They are called *contagious*, as they are transmitted by contact with an ill person or with their secretions. Cures and vaccines are rarely sufficient to stop an infectious disease - it is equally important to understand how the pathogen responsible for the disease spreads in the population, which in turn determines the way we administer the available cures or vaccines.



**Figure 10.1**  
**Super-spreaders**

One-hundred-forty-four of the 206 SARS patients diagnosed in Singapore were traced to a chain of five individuals that included four *super-spreaders*. The most important of these was *Patient Zero*, the physician from Guangdong Province in China, who brought the disease to the Metropole Hotel. After [1].



**Figure 10.2**  
**Mobile Phone Viruses**

Smart phones, capable of sharing programs and data with each other, offer a fertile ground for virus writers. Indeed, since 2004 hundreds of smart phone viruses have been identified, reaching a state of sophistication in a few years that took computer viruses about two decades to achieve [3]. Mobile viruses are transmitted using two main communication mechanisms [4]:

#### Bluetooth (BT) Viruses

A BT virus infects all phones found within BT range from the infected phone, which is about 10-30 meters. As physical proximity is essential for a BT connection, the transmission of a BT virus is determined by the owner's location and the underlying mobility network, connecting locations by individuals who travel between them (SECTION 10.4). Hence BT viruses follow a spreading pattern similar to influenza.

#### Multimedia Messaging Services (MMS)

Viruses carried by MMS can infect all susceptible phones whose number is in the infected phone's phonebook. Hence MMS viruses spread on the social network, following a long-range spreading pattern that is independent of the infected phone's physical location. Consequently the spreading of MMS viruses is similar to the patterns characterizing computer viruses.

The diversity of phenomena regularly described as spreading processes on networks is staggering:

#### Biological

The spread of pathogens on their respective contact network is the main subject of this chapter. Examples include airborne diseases like influenza, SARS, or tuberculosis, transmitted when two individuals breathe the air in the same room; contagious diseases and parasites transmitted when people touch each other; the Ebola virus, transmitted via contact with a patient's bodily fluids, HIV and other sexually transmitted diseases passed on during sexual intercourse. Infectious diseases also include cancers carried by cancer-causing viruses, like HPV or EBV, or diseases carried by parasites like bedbugs or malaria.

#### Digital

A computer virus is a self-reproducing program that can transmit a copy of itself from computer to computer. Its spreading pattern has many similarities to the spread of pathogens. But digital viruses also have many unique features, determined by the technology behind the specific virus. As mobile phones morphed into hand-held computers, lately we also witnessed the appearance of mobile viruses and worms that infect smartphones (Figure 10.2).

#### Social

The role of the social and professional network in the spread and acceptance of innovations, knowledge, business practices, products, behavior, rumors and memes, is a much-studied problem in social sciences, marketing and economics [5, 6]. Online environments, like Twitter, offer unprecedented ability to track such phenomena. Consequently a staggering number of studies focus on social spreading, asking for example why can some messages reach millions of individuals, while others struggle to get noticed.

The examples discussed above involve diverse spreading agents, from biological to computer viruses, ideas and products; they spread on differ-

PHENOMENA	AGENT	NETWORK
Venereal Disease	Pathogens	Sexual Network
Rumor Spreading	Information, Memes	Communication Network
Diffusion of Innovations	Ideas, Knowledge	Communication Network
Computer Viruses	Malwares, Digital viruses	Internet
Mobile Phone Virus	Mobile Viruses	Social Network/Proximity Network
Bedbugs	Parasitic Insects	Hotel - Traveler Network
Malaria	Plasmodium	Mosquito - Human network

ent types of networks, from social to computer and professional networks; they are characterized by widely different time scales and follow different mechanisms of transmission (Table 10.1). Despite this diversity, as we show in this chapter, these spreading processes obey common patterns and can be described using the same network-based theoretical and modeling framework.

**Table 10.1**  
**Networks and Agents**

The spread of a pathogen, a meme or a computer virus is determined by the network on which the agent spreads and the transmission mechanism of the responsible agent. The table lists several much studied spreading phenomena, together with the nature of the particular spreading agent and the network on which the agent spreads.

# EPIDEMIC MODELING

Epidemiology has developed a robust analytical and numerical framework to model the spread of pathogens. This framework relies on two fundamental hypotheses:

## i. Compartmentalization

Epidemic models classify each individual based on the stage of the disease affecting them. The simplest classification assumes that an individual can be in one of three *states* or *compartments*:

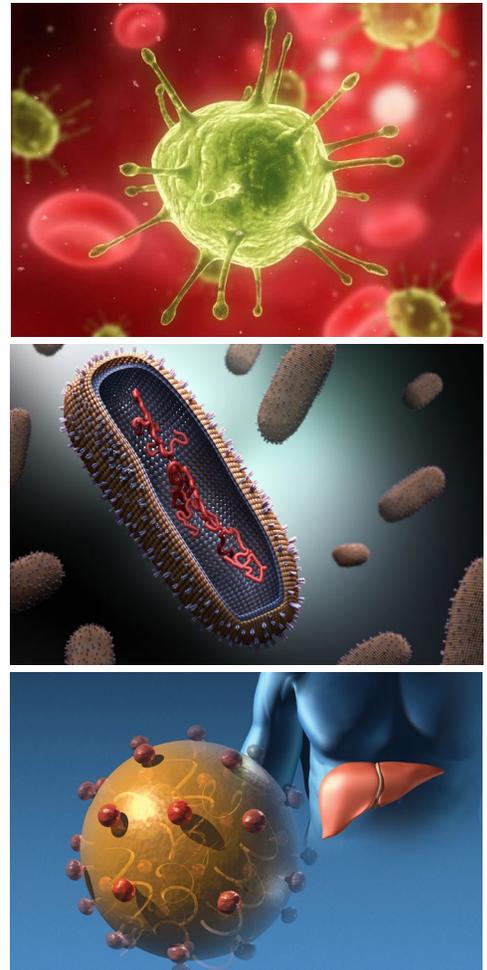
- *Susceptible (S)*: Healthy individuals who have not yet contacted the pathogen (Figure 10.3).
- *Infectious (I)*: Contagious individuals who have contacted the pathogen and hence can infect others.
- *Recovered (R)*: Individuals who have been infected before, but have recovered from the disease, hence are not infectious.

The modeling of some diseases requires additional states, like *immune* individuals, who cannot be infected, or *latent* individuals, who have been exposed to the disease, but are not yet contagious.

Individuals can move between compartments. For example, at the beginning of a new influenza outbreak everyone is in the susceptible state. Once an individual comes into contact with an infected person, she can become infected. Eventually she will recover and develop immunity, losing her susceptibility to the the particular strain of influenza.

## ii. Homogenous Mixing

The homogenous mixing hypothesis (also called *fully mixed* or *mass-action approximation*) assumes that each individual has the same chance of coming into contact with an infected individual. This hypothesis eliminates the need to know the precise contact network on which the disease spreads, replacing it with the assumption that



**Figure 10.3**  
**Pathogens**

A *pathogen*, a word rooted in the Greek words “suffering, passion” (*pathos*) and “producer of” (*genēs*), denotes an infectious agent or germ. A pathogen could be a disease-causing microorganism, like a virus, a bacterium, a prion, or a fungus. The figure shows several much-studied pathogens, like the HIV virus, responsible for AIDS, an influenza virus and the hepatitis C virus. After <http://www.livescience.com/18107-hiv-therapeutic-vaccines-promise.html> and [http://www.huffingtonpost.com/2014/01/13/deadly-viruses-beautiful-photos\\_n\\_4545309.html](http://www.huffingtonpost.com/2014/01/13/deadly-viruses-beautiful-photos_n_4545309.html)

anyone can infect anyone else.

In this section we introduce the epidemic modeling framework built on these two hypotheses. To be specific, we explore the dynamics of three frequently used epidemic models, the so-called SI, SIS and SIR models, that help us understand the basic building blocks of epidemic modeling.

### SUSCEPTIBLE-INFECTED (SI) MODEL

Consider a disease that spreads in a population of  $N$  individuals. Denote with  $S(t)$  the number of individuals who are susceptible (healthy) at time  $t$  and with  $I(t)$  the number individuals that have been already infected. At time  $t=0$  everyone is susceptible ( $S(0)=N$ ) and no one is infected ( $I(0)=0$ ). Let us assume that a typical individual has  $\langle k \rangle$  contacts and that the likelihood that the disease will be transmitted from an infected to a susceptible individual in a unit time is  $\beta$ . We ask the following: If a single individual becomes infected at time  $t=0$  (i.e.  $I(0)=1$ ), how many individuals will be infected at some later time  $t$ ?

Within the homogenous mixing hypothesis the probability that the infected person encounters a susceptible individual is  $S(t)/N$ . Therefore the infected person comes into contact with  $\langle k \rangle S(t)/N$  susceptible individuals in a unit time. Since  $I(t)$  infected individuals are transmitting the pathogen, each at rate  $\beta$ , the average number of new infections  $dI(t)$  during a timeframe  $dt$  is

$$\beta \langle k \rangle \frac{S(t)I(t)}{N} dt.$$

Consequently  $I(t)$  changes at the rate

$$\frac{dI(t)}{dt} = \beta \langle k \rangle \frac{S(t)I(t)}{N}. \quad (10.1)$$

Throughout this chapter we will use the variables

$$s(t) = S(t)/N, \quad i(t) = I(t)/N, \quad (10.2)$$

to capture the fraction of the susceptible and of the infected population at time  $t$ . For simplicity we also drop the  $(t)$  variable from  $i(t)$  and  $s(t)$ , re-writing (10.1) as (ADVANCED TOPICS 10.A)

$$\frac{di}{dt} = \beta \langle k \rangle si = \beta \langle k \rangle i(1-i), \quad (10.3)$$

where the product  $\beta \langle k \rangle$  is called the *transmission rate* or *transmissibility*. We solve (10.3) by writing

$$\frac{di}{i} + \frac{di}{(1-i)} = \beta \langle k \rangle dt.$$

Integrating both sides, we obtain

$$\ln i - \ln(1-i) + C = \beta \langle k \rangle t.$$

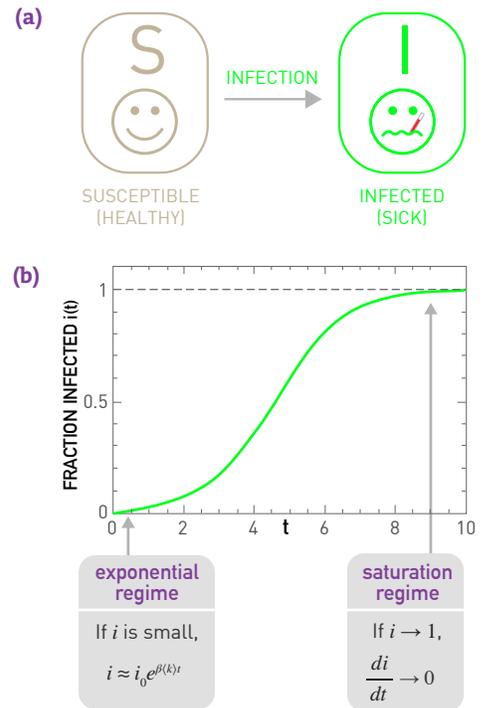


Figure 10.4  
The Susceptible-Infected (SI) Model

- (a) In the SI model an individual can be in one of two states: susceptible (healthy) or infected (sick). The model assumes that if a susceptible individual comes into contact with an infected individual, it becomes infected at rate  $\beta$ . The arrow indicates that once an individual becomes infected, it stays infected, hence it cannot recover.
- (b) Time evolution of the fraction of infected individuals, as predicted by (10.4). At early times the fraction of infected individuals grows exponentially. As eventually everyone becomes infected, at large times we have  $i(\infty)=1$ .

With the initial condition  $i_0 = i(t=0)$ , we get  $C = i_0 / (1 - i_0)$ , obtaining that the fraction of infected individuals increases in time as

$$i = \frac{i_0 e^{\beta\langle k \rangle t}}{1 - i_0 + i_0 e^{\beta\langle k \rangle t}}. \quad (10.4)$$

Equation (10.4) predicts that:

- At the beginning the fraction of infected individuals increases exponentially (Figure 10.4b). Indeed, early on an infected individual encounters only susceptible individuals, hence the pathogen can easily spread.

- The *characteristic time* required to reach an  $1/e$  fraction (about 36%) of all susceptible individuals is

$$\tau = \frac{1}{\beta\langle k \rangle}. \quad (10.5)$$

Hence  $\tau$  is the inverse of the speed with which the pathogen spreads through the population. Equation (10.5) predicts that increasing either the density of links  $\langle k \rangle$  or  $\beta$  enhances the speed of the pathogen and reduces the characteristic time.

- With time an infected individual encounters fewer and fewer susceptible individuals. Hence the growth of  $i$  slows for large  $t$  (Figure 10.4b). The epidemic ends when everyone has been infected, i.e. when  $i(t \rightarrow \infty) = 1$  and  $s(t \rightarrow \infty) = 0$ .

### SUSCEPTIBLE-INFECTED-SUSCEPTIBLE (SIS) MODEL

Most pathogens are eventually defeated by the immune system or by treatment. To capture this fact we need to allow the infected individuals to recover, ceasing to spread the disease. With that we arrive at the so-called *SIS model*, which has the same two states as the SI model, susceptible and infected. The difference is that now infected individuals recover at a fixed rate  $\mu$ , becoming susceptible again (Figure 10.5a). The equation describing the dynamics of this model is an extension of (10.3),

$$\frac{di}{dt} = \beta\langle k \rangle i(1-i) - \mu i, \quad (10.6)$$

where  $\mu$  is the *recovery rate* and the  $\mu i$  term captures the rate at which the population recovers from the disease. The solution of (10.6) provides the fraction of infected individuals in function of time (Figure 10.5b)

$$i = \left(1 - \frac{\mu}{\beta\langle k \rangle}\right) \frac{C e^{\beta\langle k \rangle t - \mu t}}{1 + C e^{\beta\langle k \rangle t - \mu t}}, \quad (10.7)$$

where the initial condition  $i_0 = i(t=0)$  gives  $C = i_0 / (1 - i_0 - \mu / \beta\langle k \rangle)$ .

While in the SI model eventually everyone becomes infected, (10.7) predicts that in the SIS model the epidemic has two possible outcomes:

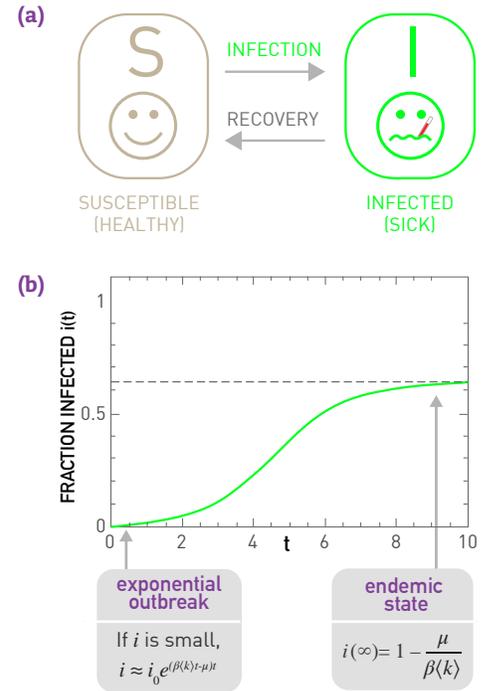


Figure 10.5  
The Susceptible-Infected-Susceptible (SIS) Model

- (a) The SIS model has the same states as the SI model: susceptible and infected. It differs from the SI model in that it allows recovery, i.e. infected individuals are cured, becoming susceptible again at rate  $\mu$ .
- (b) Time evolution of the fraction of infected individuals in the SIS model, as predicted by (10.7). As recovery is possible, at large  $t$  the system reaches an *endemic state*, in which the fraction of infected individuals is constant,  $i(\infty)$ , given by (10.8). Hence in the endemic state only a finite fraction of individuals are infected. Note that for high recovery rate  $\mu$  the number of infected individuals decreases exponentially and the disease dies out.

- **Endemic State** ( $\mu < \beta \langle k \rangle$ )

For low recovery rate the fraction of infected individuals,  $i$ , follows a logistic curve similar to the one observed for the SI model. Yet, not everyone is infected, but  $i$  reaches a constant  $i(\infty) < 1$  value (Figure 10.5b). This means that at any moment only a finite fraction of the population is infected. In this stationary or *endemic state* the number of newly infected individuals equals the number of individuals who recover from the disease, hence the infected fraction of the population does not change with time. We can calculate  $i(\infty)$  by setting  $di/dt = 0$  in (10.6), obtaining

$$i(\infty) = 1 - \frac{\mu}{\beta \langle k \rangle}. \quad (10.8)$$

- **Disease-free State** ( $\mu > \beta \langle k \rangle$ )

For a sufficiently high recovery rate the exponent in (10.7) is negative. Therefore,  $i$  decreases exponentially with time, indicating that an initial infection will die out exponentially. This is because in this state the number of individuals cured per unit time exceeds the number of newly infected individuals. Therefore with time the pathogen disappears from the population.

In other words, the SIS model predicts that some pathogens will persist in the population while others die out shortly. To understand what governs the difference between these two outcomes we write the characteristic time of a pathogen as

$$\tau = \frac{1}{\mu(R_0 - 1)}, \quad (10.9)$$

where

$$R_0 = \frac{\beta \langle k \rangle}{\mu} \quad (10.10)$$

is the *basic reproductive number*. It represents the average number of susceptible individuals infected by an infected individual during its infectious period in a fully susceptible population. In other words,  $R_0$  is the number of new infections each infected individual causes under ideal circumstances. The basic reproductive number is valuable for its predictive power:

- If  $R_0$  exceeds unity,  $\tau$  is positive, hence the epidemic is in the endemic state. Indeed, if each infected individual infects more than one healthy person, the pathogen is poised to spread and persist in the population. The higher is  $R_0$ , the faster is the spreading process.
- If  $R_0 < 1$  then  $\tau$  is negative and the epidemic dies out. Indeed, if each infected individual infects less than one additional person, the pathogen cannot persist in the population.

Consequently, the reproductive number is one of the first parameters

DISEASE	TRANSMISSION	$R_0$
Measles	Airborne	12-18
Pertussis	Airborne droplet	12-17
Diphtheria	Saliva	6-7
Smallpox	Social contact	5-7
Polio	Fecal-oral route	5-7
Rubella	Airborne droplet	5-7
Mumps	Airborne droplet	4-7
HIV/AIDS	Sexual contact	2-5
SARS	Airborne droplet	2-5
Influenza (1918 strain)	Airborne droplet	2-3

**Table 10.2**  
**The Basic Reproductive Number,  $R_0$**

The reproductive number (10.10) provides the number of individuals an infectious individual infects if all its contacts are susceptible. For  $R_0 < 1$  the pathogen naturally dies out, as the number of recovered individuals exceeds the number of new infections. If  $R_0 > 1$  the pathogen will spread and persist in the population. The higher is  $R_0$ , the faster is the spreading process. The table lists  $R_0$  for several well-known pathogens. After [7].

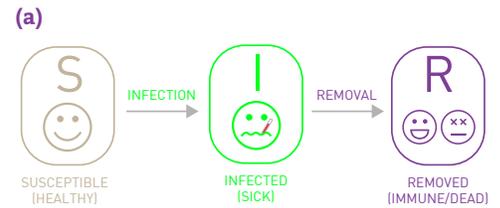
epidemiologists estimate for a new pathogen, gauging the severity of the problem they face. For several well-studied pathogens  $R_0$  is listed in Table 10.2. The high  $R_0$  of some of these pathogens underlies the dangers they pose: For example each individual infected with measles causes over a dozen subsequent infections.

### SUSCEPTIBLE-INFECTED-RECOVERED (SIR) MODEL

For many pathogens, like most strains of influenza, individuals develop immunity after they recover from the infection. Hence, instead of returning to the susceptible state, they are “removed” from the population. These recovered individuals do not count any longer from the perspective of the pathogen as they cannot be infected, nor can they infect others. The SIR model, whose properties are discussed in Figure 10.6, captures the dynamics of this process.

In summary, depending on the characteristics of a pathogen, we need different models to capture the dynamics of an epidemic outbreak. As shown in Figure 10.7, the predictions of the SI, SIS, and SIR models agree with each other in the early stages of an epidemic: When the number of infected individuals is small, the disease spreads freely and the number of infected individuals increases exponentially. The outcomes are different for large times: In the SI model everyone becomes infected; the SIS model either reaches an endemic state, in which a finite fraction of individuals are always infected, or the infection dies out; in the SIR model everyone recovers at the end. The reproductive number predicts the long-term fate of an epidemic: for  $R_0 < 1$  the pathogen persists in the population, while for  $R_0 > 1$  it dies out naturally.

The models discussed so far have ignored the fact that an individual comes into contact only with its network-based neighbors in the pertinent contact network. We assumed homogenous mixing instead, which means that an infected individual can infect any other individual. It also means that an infected individual typically infects only  $\langle k \rangle$  other individuals, ignoring variations in node degrees. To accurately predict the dynamics of an epidemic, we need to consider the precise role the contact network plays in epidemic phenomena.

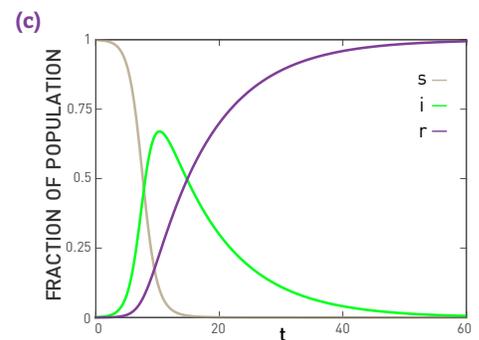


(b)

$$\frac{ds}{dt} = -\beta \langle k \rangle i [1 - r - i]$$

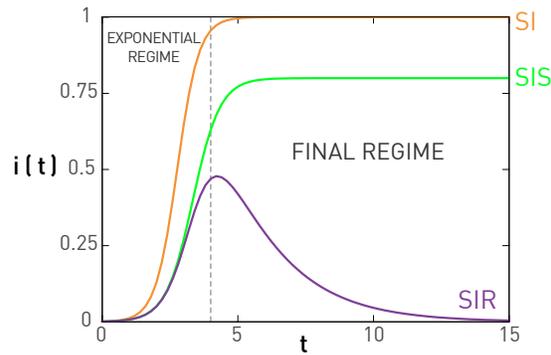
$$\frac{di}{dt} = -\mu i + \beta \langle k \rangle i [1 - r - i]$$

$$\frac{dr}{dt} = \mu i$$



**Figure 10.6**  
**The Susceptible-Infected-Recovered (SIR) Model**

- (a) In contrast with the SIS model, in the SIR model recovered individuals enter a *recovered* state, meaning that they develop immunity rather than becoming susceptible again. Flu, SARS and Plague are diseases with this property, hence we must use the SIR model to describe their spread.
- (b) The differential equations governing the time evolution of the fraction of individuals in the susceptible  $s$ , infected  $i$  and the removed  $r$  state.
- (c) The time dependent behavior of  $s$ ,  $i$  and  $r$  as predicted by the equations shown in (b). According to the model all individuals transition from a susceptible (healthy) state to the infected (sick) state and then to the recovered (immune) state.



**Figure 10.7**  
**Comparing the SI, SIS and SIR Models**

The plot shows growth of the fraction of infected individuals,  $i$ , in the SI, SIS and SIR models. Two different regimes stand out:

**Exponential Regime**

The models predict an exponential growth in the number of infected individuals during the early stages of the epidemic. For the same  $\beta$  the SI model predicts the fastest growth (smallest  $\tau$ , see (10.5)). For the SIS and SIR models the growth is slowed by recovery, resulting in a larger  $\tau$ , as predicted by (10.9). Note that for sufficiently high recovery rate  $\mu$  the SIS and the SIR models predict a disease-free state, when the number of infected individuals *decays* exponentially with time.

**Final Regime**

The three models predict different long-term outcomes: In the SI model everyone becomes infected,  $i(\infty)=1$ ; in the SIS model a finite fraction of individuals are infected  $i(\infty)<1$ ; in the SIR model all infected nodes recover, hence the number of infected individuals goes to zero  $i(\infty)=0$ .

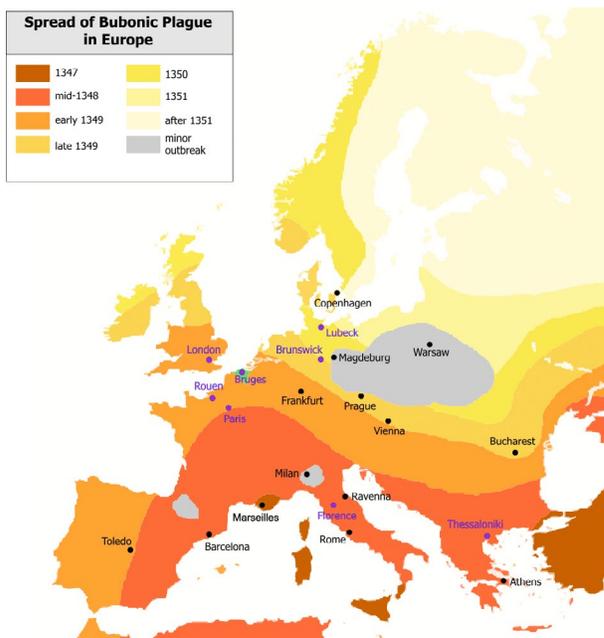
The table summarizes the main properties of each model.

	SI	SIS	SIR
<b>Exponential Regime:</b> Number of infected individuals grows exponentially	$i = \frac{i_0 e^{\beta(i_0)t}}{1 - i_0 + i_0 e^{\beta(i_0)t}}$	$i = \left(1 - \frac{\mu}{\beta\langle k \rangle}\right) \frac{C e^{(\beta\langle i \rangle - \mu)t}}{1 + C e^{(\beta\langle i \rangle - \mu)t}}$	No closed solution
<b>Final Regime:</b> Saturation at $t \rightarrow \infty$	$i(\infty) = 1$	$i(\infty) = 1 - \frac{\mu}{\beta\langle k \rangle}$	$i(\infty) = 0$
<b>Epidemic Threshold:</b> Disease does not always spread	No threshold	$R_0 = 1$	$R_0 = 1$

# NETWORK EPIDEMICS

The ease of air travel, allowing millions to cross continents on a daily basis, has dramatically accelerated the speed with which pathogens travel around the world. While in medieval times a virus took years to sweep a continent (Figure 10.8), today a new virus can reach several continents in a matter of days. There is an acute need, therefore, to understand and predict the precise patterns that pathogens follow as they spread around the globe.

The epidemic models discussed in the previous section do not incorporate the structure of the contact network that facilitates the spread of a pathogen. Instead they assume that any individual can come into contact with any other individual (homogenous mixing hypothesis) and that all individuals have comparable number of contacts,  $\langle k \rangle$ . Both assumptions are false: Individual can transmit a pathogen only to those they come into contact with, hence pathogens spread on a complex contact network. Furthermore, these contact networks are often scale-free, hence  $\langle k \rangle$  is not sufficient to characterize their topology.



**Figure 10.8**  
**The Great Plague**

The Black Death, one of the most devastating pandemics in human history, was an outbreak of bubonic plague caused by the bacterium *Yersinia pestis*. The figure shows the gradual advance of the disease throughout Europe, taking years to sweep the continent. It started in China and traveled along the Silk Road to reach Crimea around 1346. From there, probably carried by Oriental rat fleas on the black rats that were regular passengers on merchant ships, spread throughout the Mediterranean and Europe. Its slow spread reflected the slow travel speed of its era. The black death is estimated to have killed 30% to 60% of Europe's population [8]. The resulting devastation has caused a series of religious, social and economic upheavals, having a profound impact on the history of Europe.

After Roger Zenner, Wikipedia.

The failure of the basic hypotheses prompted a fundamental revision of the epidemic modeling framework. This change began with the work of Romualdo Pastor-Satorras and Alessandro Vespignani, who in 2001 extended the basic epidemic models to incorporate in a self-consistent fashion the topological characteristics of the underlying contact network [9]. In this section we introduce the formalism developed by them, familiarizing ourselves with *network epidemics*.

### SUSCEPTIBLE-INFECTED (SI) MODEL ON A NETWORK

If a pathogen spreads on a network, individuals with more links are more likely to be in contact with an infected individual, hence they are more likely to be infected. Therefore the mathematical formalism must consider the degree of each node as an implicit variable. This is achieved by the *degree block approximation*, that distinguishes nodes based on their degree and assumes that nodes with the same degree are statistically equivalent (Figure 10.9). Therefore we denote with

$$i_k = \frac{I_k}{N_k} \quad (10.11)$$

the fraction of nodes with degree  $k$  that are infected among all  $N_k$  degree- $k$  nodes in the network. The total fraction of infected nodes is the sum of all infected degree- $k$  nodes

$$i = \sum_k p_k i_k \quad (10.12)$$

Given the different node degrees, we write the SI model for each degree  $k$  separately:

$$\frac{di_k}{dt} = \beta(1 - i_k)k\Theta_k. \quad (10.13)$$

This equation has the same structure as (10.3): The infection rate is proportional to  $\beta$  and the fraction of degree- $k$  nodes that are not yet infected, which is  $(1 - i_k)$ . Yet, there are some key differences:

- The average degree  $\langle k \rangle$  in (10.3) is replaced with each node's actual degree  $k$ .
- The density function  $\Theta_k$  represents the fraction of infected neighbors of a susceptible node  $k$ . In the homogenous mixing assumption  $\Theta_k$  is simply the fraction of the infected nodes,  $i$ . In a network environment, however, the fraction of infected nodes in the vicinity of a node can depend on the node's degree  $k$  and time  $t$ .
- While (10.3) captures with a single equation the time dependent behavior of the whole system, (10.13) represents a system of  $k_{\max}$  coupled equations, one equation for each degree present in the network.

We start by exploring the early time behavior of  $i_k$ , a choice driven by both theoretical interest and practical considerations. Indeed, developing vaccines, cures, and other medical interventions for a new pathogen can take months to years. If we lack a cure, the only way to alter the course

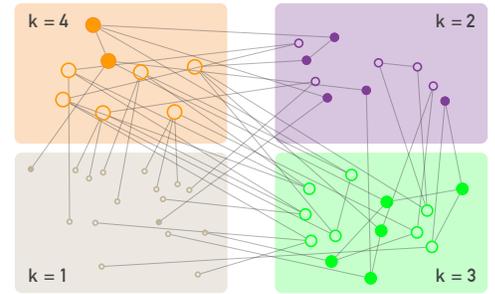


Figure 10.9  
Degree Block Approximation

The epidemic models discussed in SECTION 10.2 grouped each node into compartments based on their state, placing them into susceptible, infected, or recovered compartments. To account for the role of the network topology, the *degree block approximation* adds an additional set of compartments, placing all nodes that have the same degree into the same block. In other words, we assume that nodes with the same degree behave similarly. This allows us to write a separate rate equation for each degree, as we did in (10.13). The degree block approximation does not eliminate the compartments based on the state of an individual: Independent of its degree an individual can be susceptible to the disease (empty circles) or infected (full circles).

of an epidemic is to do so early, using quarantine, travel restrictions and transmission-slowing measures to halt its spread. To make the right decision about the nature, the timing and the magnitude of each intervention, we need an accurate estimate of the number of individuals infected in the early stages of the epidemic.

At the beginning of the epidemic  $i_k$  is small and the higher order term in (10.13)  $\beta i_k k \Theta_k$  can be neglected. Hence we can approximate (10.13) with

$$\frac{di_k}{dt} \approx \beta k \Theta_k. \quad (10.14)$$

As we show in ADVANCED TOPICS 10.B, for a network lacking degree correlations the  $\Theta_k$  function is independent of  $k$ , so using (10.40), (10.14) becomes

$$\frac{di_k}{dt} \approx \beta k i_0 \frac{\langle k \rangle - 1}{\langle k \rangle} e^{t/\tau^{\text{SI}}}, \quad (10.15)$$

where  $\tau^{\text{SI}}$  is the characteristic time for the spread of the pathogen

$$\tau^{\text{SI}} = \frac{\langle k \rangle}{\beta(\langle k^2 \rangle - \langle k \rangle)}. \quad (10.16)$$

Integrating (10.15) we obtain the fraction of infected nodes with degree  $k$

$$i_k = i_0 \left( 1 + \frac{k(\langle k \rangle - 1)}{\langle k^2 \rangle - \langle k \rangle} (e^{t/\tau^{\text{SI}}} - 1) \right). \quad (10.17)$$

Equation (10.17) makes several important predictions:

- The higher the degree of a node, the more likely that it becomes infected. Indeed, for any time  $t$  we can write (10.17) as  $i_k = g(t) + kf(t)$ , indicating that the group of nodes with higher degree has a higher fraction of infected nodes (Figure 10.10).
- According to (10.12) the total fraction of infected nodes grows with time as

$$i = \int_0^{k_{\max}} i_k p_k dk = i_0 \left( 1 + \frac{\langle k \rangle^2 - \langle k \rangle}{\langle k^2 \rangle - \langle k \rangle} (e^{t/\tau^{\text{SI}}} - 1) \right). \quad (10.18)$$

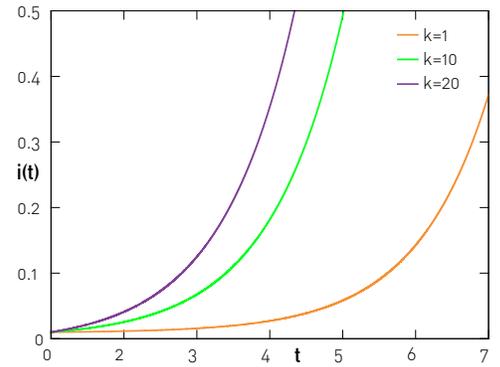
According to (10.16) the characteristic time  $\tau$  depends not only on  $\langle k \rangle$ , but also on the network's degree distribution through  $\langle k^2 \rangle$ . To fully understand the significance of the prediction (10.16), let us derive  $\tau^{\text{SI}}$  for different networks:

- **Random Network**

For a random network  $\langle k^2 \rangle = \langle k \rangle (\langle k \rangle + 1)$ , obtaining

$$\tau_{\text{ER}}^{\text{SI}} = \frac{1}{\beta \langle k \rangle}, \quad (10.19)$$

recovering the result (10.5) for homogenous networks.



**Figure 10.10**  
**Fraction of Infected Nodes in the SI Model**

Equation (10.17) predicts that the a pathogen spreads with different speed on nodes with different degrees. To be specific, we can write  $i_k = g(t) + kf(t)$ , indicating that at any time the fraction of high degree nodes that are infected is higher than the fraction of low degree nodes. The figure shows the fraction of infected nodes with degrees  $k=1, 10$  and  $100$  in an Erdős-Rényi network with average degree  $\langle k \rangle = 2$ . It shows that at  $t=3$  less than 3% of the  $k=1$  nodes are infected, in contrast with close to 20% of the  $k=10$  nodes and close to 30% of the  $k=20$  nodes. Consequently, at any time virtually all hubs are infected, but small-degree nodes tend to be disease free. Hence the disease is maintained in the hubs, which in turn broadcast the disease to the rest of the network.

- **Scale-free Network with  $\gamma \geq 3$**

If the contact network on which the disease spreads is scale-free with degree exponent  $\gamma \geq 3$ , both  $\langle k \rangle$  and  $\langle k^2 \rangle$  are finite. Consequently  $\tau^{\text{SI}}$  is also finite and the spreading dynamics is similar to the behavior predicted for a random network but with an altered  $\tau^{\text{SI}}$ .

- **Scale-free Networks with  $\gamma \leq 3$**

For  $\gamma < 3$  in the  $N \rightarrow \infty$  limit  $\langle k^2 \rangle \rightarrow \infty$  hence (10.16) predicts  $\tau^{\text{SI}} \rightarrow 0$ . In other words, *the spread of a pathogen on a scale-free network is instantaneous*. This is perhaps the most unexpected prediction of network epidemics.

The vanishing characteristic time reflects the important role hubs play in epidemic phenomena. Indeed, as illustrated in Figure 10.10, in a scale-free network the hubs are the first to be infected, as through the many links they have, they are very likely to be in contact with an infected node. Once a hub becomes infected, it “broadcasts” the disease to the rest of the network, turning into a super-spreader.

- **Inhomogenous Networks**

A network does not need to be strictly scale-free for the impact of the degree heterogeneity to be detectable. Indeed, (10.16) predicts that as long as  $\langle k^2 \rangle > \langle k \rangle (\langle k \rangle + 1)$ ,  $\tau^{\text{SI}}$  is reduced. Hence heterogenous network enhance the speed of any pathogen.

In the SI model with time the pathogen reaches all individuals. Consequently the degree heterogeneity affects only the characteristic time, which in turn determines the speed with which the pathogen sweeps through the population. To understand the full impact of the network topology, we need to explore the behavior of the SIS model on a network.

### SIS MODEL AND THE VANISHING EPIDEMIC THRESHOLD

The continuum equation describing the dynamics of the SIS model on a network is a straightforward extension of the SI model discussed in SECTION 10.2,

$$\frac{di_k}{dt} = \beta(1 - i_k)k\Theta_k(t) - \mu i_k. \quad (10.20)$$

The difference between (10.13) and (10.20) is the presence of the recovery term  $-\mu i_k$ . This changes the characteristic time of the epidemic to (ADVANCED TOPICS 10.B)

$$\tau^{\text{SIS}} = \frac{\langle k \rangle}{\beta \langle k^2 \rangle - \mu \langle k \rangle}. \quad (10.21)$$

For sufficiently large  $\mu$  the characteristic time is negative, hence  $i_k$  decays exponentially. The condition for the decay depends not only on the recovery rate and  $\langle k \rangle$ , but also on the network heterogeneity, through  $\langle k^2 \rangle$ . To predict when a pathogen persists in the population we define the *spreading rate*

$$\lambda = \frac{\beta}{\mu}, \quad (10.22)$$

which depends only on the biological characteristics of the pathogen, namely the transmission probability  $\beta$  and the recovery rate  $\mu$ . The higher is  $\lambda$ , the more likely that the disease will spread. Yet, the number of infected individuals does not increase gradually with  $\lambda$ . Rather, the pathogen can spread only if its spreading rate exceeds an *epidemic threshold*  $\lambda_c$ . Next we calculate  $\lambda_c$  for random and scale-free networks.

- **Random Network**

If a pathogen spreads on a random network, we can use  $\langle k^2 \rangle = \langle k \rangle (\langle k \rangle + 1)$  in (10.21), obtaining that the pathogen persists in the population if

$$\tau_{ER}^{SIS} = \frac{1}{\beta(\langle k \rangle + 1) - \mu} > 0. \quad (10.23)$$

Using (10.22) we obtain

$$\lambda > \frac{1}{\langle k \rangle + 1}, \quad (10.24)$$

obtaining the *epidemic threshold of a random network* as

$$\lambda_c = \frac{1}{\langle k \rangle + 1}. \quad (10.25)$$

As  $\langle k \rangle$  is always finite, a random network always has a nonzero epidemic threshold (Figure 10.11), with key consequences:

- If the spreading rate  $\lambda$  exceeds the epidemic threshold  $\lambda_c$ , the pathogen will spread until it reaches an endemic state, where a finite fraction  $i(\lambda)$  of the population is infected at any time.
- If  $\lambda < \lambda_c$ , the pathogen dies out, i.e.  $i(\lambda) = 0$ .

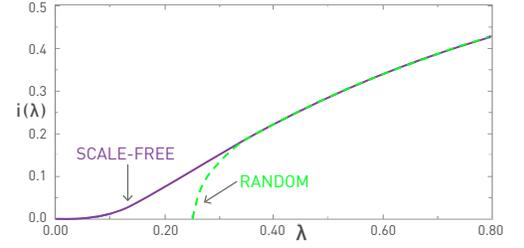
Hence the epidemic threshold allows us to decide if a pathogen can or cannot persist in a population. This transition from the absence to the presence of an epidemic outbreak by increasing the spreading rate  $\lambda$  is at the basis of most campaigns to stop a pathogen (SECTION 10.6).

- **Scale-free Network**

For a network with an arbitrary degree distribution we set  $\tau^{SIS} > 0$  in (10.21), obtaining the epidemic threshold as

$$\lambda_c = \frac{\langle k \rangle}{\langle k^2 \rangle}. \quad (10.26)$$

As for a scale-free network  $\langle k^2 \rangle$  diverges in the  $N \rightarrow \infty$  limit, for large networks the epidemic threshold is expected to vanish (Figures 10.11 and 10.12). This means that *even viruses that are hard to pass from individual to individual can spread successfully*, repre-



**Figure 10.11**  
**Epidemic Threshold**

The fraction of infected individuals  $i(\lambda) = i(t \rightarrow \infty)$  in the endemic state of the SIS model. The curves are for a random (green) and for a scale-free contact network (purple). The random network has a finite epidemic threshold  $\lambda_c$ , implying that a pathogen with a small spreading rate ( $\lambda < \lambda_c$ ) must die out, i.e.  $i(\lambda_c) = 0$ . If, however, the spreading rate of the pathogen exceeds  $\lambda_c$ , the pathogen becomes endemic and a finite fraction of the population is infected at any time. For a scale-free network we have  $\lambda_c = 0$ , hence even viruses with a very small spreading rate  $\lambda$  can persist in the population.

$2 < \gamma < 3$	$\lambda_c = 0$ $\Theta(\lambda) \sim (k_{\min} \lambda)^{(\gamma-2)(3-\gamma)}$ $i(\lambda) \sim \lambda^{1/(3-\gamma)}$
$\gamma = 3$	$\lambda_c = 0$ $\Theta(\lambda) \approx \frac{e^{-1/k_{\min} \lambda}}{\lambda k_{\min}} (1 - e^{-1/k_{\min} \lambda})^{-1}$ $i(\lambda) \sim 2e^{-1/k_{\min} \lambda}$
$3 < \gamma < 4$	$\lambda_c > 0$ $i(\lambda) \sim \left( \lambda - \frac{\gamma-3}{k_{\min}(\gamma-2)} \right)^{1/(\gamma-3)}$
$\gamma > 4$	$\lambda_c > 0$ $i(\lambda) \sim \lambda - \frac{\gamma-3}{k_{\min}(\gamma-2)}$

**Figure 10.12**  
**The Asymptotic Behavior of the SIS Model**

The fraction of individuals infected in the endemic state,  $i(\lambda) = i(t \rightarrow \infty)$ , depends on the structure of the underlying network and the disease parameters  $\beta$  and  $\mu$ . The figure summarizes the key properties of the epidemic threshold  $\lambda_c$ , the density function  $\Theta(\lambda)$  and  $i(\lambda)$  for a scale-free network with degree exponent  $\gamma$ . The results indicate that only for  $\gamma > 4$  does the epidemics on a scale-free network converge to the results of the traditional epidemic models. After [10].

MODEL	CONTINUUM EQUATION	$\tau$	$\lambda_c$
SI	$\frac{di_k}{dt} = \beta[1-i_k]k\theta_k$	$\frac{\langle k \rangle}{\beta(\langle k^2 \rangle - \langle k \rangle)}$	0
SIS	$\frac{di_k}{dt} = \beta[1-i_k]k\theta_k - \mu i_k$	$\frac{\langle k \rangle}{\beta\langle k^2 \rangle - \mu\langle k \rangle}$	$\frac{\langle k \rangle}{\langle k^2 \rangle}$
SIR	$\frac{di_k}{dt} = \beta s_k \theta_k - \mu i_k$ $s_k = 1 - i_k - r_k$	$\frac{\langle k \rangle}{\beta\langle k^2 \rangle - (\mu + \beta)\langle k \rangle}$	$\frac{1}{\frac{\langle k^2 \rangle}{\langle k \rangle} - 1}$

**Table 10.3**  
**Epidemic Models on Networks**

The table shows the rate equation for the three basic epidemic models (SI, SIS, SIR) on a network with arbitrary  $\langle k \rangle$  and  $\langle k^2 \rangle$ , together with the corresponding characteristic  $\tau$  and the epidemic threshold  $\lambda_c$ . For the SI model  $\lambda_c = 0$ , as in the absence of recovery ( $\mu=0$ ) a pathogen spreads until it reaches all susceptible individuals. The listed  $\tau$  and  $\lambda_c$  are derived in **ADVANCED TOPICS 10.B**.

sending the second fundamental prediction of network epidemics.

The vanishing epidemic threshold is a direct consequence of the hubs. Indeed, a pathogen that fails to infect other nodes before the infected individual recovers, will slowly disappear from the population (**ADVANCED TOPICS 10.A**). In a random network all nodes have comparable degree,  $k \approx \langle k \rangle$ , hence if the spreading rate is under the epidemic threshold, the pathogen has no avenues to spread. In a scale-free network, however, even if a pathogen is only weakly infectious, if it infects a hub, the hub can pass it on to a large number of other nodes, allowing it to persist in the population.

In summary, the results of this section show that accounting for the network topology greatly alters the predictive power of the epidemic models. We derived two fundamental results:

- In a large scale-free network  $\tau=0$ , which means that a virus can instantaneously reach most nodes.
- In a large scale-free network  $\lambda_c=0$ , which means that even viruses with small spreading rate can persist in the population.

Both results are the consequence of hubs' ability to broadcast a pathogen to a large number of other nodes.

Note that these results are not limited to scale-free networks. Rather **(10.16)** and **(10.26)** predict that both  $\tau$  and  $\lambda_c$  depend on  $\langle k^2 \rangle$ , hence the effects discussed above will impact any network with high degree heterogeneity. In other words, if  $\langle k^2 \rangle$  is larger than the random expectation  $\langle k \rangle \langle k+1 \rangle$ , we will observe an enhanced spreading process, resulting in a smaller  $\tau$  and  $\lambda_c$  than predicted by the traditional epidemic models. As this implies a faster spread of the pathogen than predicted by the traditional epidemic models, efforts to control an epidemic cannot ignore this difference.

The results of this section were based on the degree-block approximation, which treats the detailed time-dependent infection process in a mean-field

fashion. Note, however, that this approximation, while simplifies the presentation, is not necessary. The underlying stochastic problem can be treated in its full mathematical complexity [11-14]. Such calculations show that due to the fact that the hubs can be re-infected in the SIS model, the epidemic threshold vanishes even for  $\gamma > 3$ , in contrast with the finite threshold predicted by the mean-field approach (Figure 10.12). Hence hubs play an even more important role than our earlier calculations indicate.

# CONTACT NETWORKS

Network epidemics predicts that the speed with which a pathogen spreads depends on the degree distribution of the relevant contact network. Indeed, we found that  $\langle k^2 \rangle$  affects both the characteristic time  $\tau$  and the epidemic threshold  $\lambda_c$ . None of these findings are consequential if the network on which a pathogen spreads is random - in that case the predictions of network epidemics are indistinguishable from the predictions of the traditional epidemic models encountered in SECTION 10.2. In this section we inspect the structure of several contact networks encountered in epidemic phenomena, offering direct empirical evidence of the significance of the underlying degree heterogeneities.

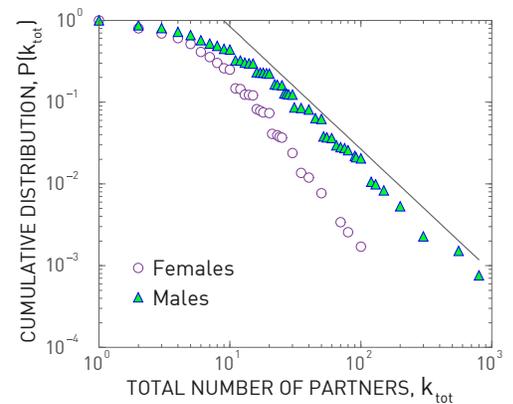
## SEXUALLY TRANSMITTED DISEASES

The HIV virus, the pathogen responsible for AIDS, spreads mainly through sexual intercourse. Consequently, the relevant contact network captures who had sexual relationship with whom. The structure of this sex web was first revealed by a study surveying the sexual habits of the Swedish population [15]. Through interviews and questionnaires, researchers collected information from 4,781 randomly chosen Swedes of ages 18 to 74. The participants were not asked to reveal the identity of their sexual partners, but only to estimate the number of sexual partners they had during their lifetime. Hence the researchers could reconstruct the degree distribution of the sexual network [16], finding that it is well approximated with a power law (Figure 10.13). This was the first empirical evidence of the relevance of scale-free networks to the spread of pathogens. The finding was confirmed by data collected in Britain, US and Africa [17].

The scale-free nature of the sexual network indicates that most individuals have relatively few sexual partners. A few individuals, however, had hundreds of sexual partners during their lifetime. Consequently the sexual network has a high  $\langle k^2 \rangle$ , which lowers both  $\tau$  and  $\lambda_c$ .

## AIRBORNE DISEASES

For airborne diseases, like influenza, SARS or H1N1, the contact network captures the set of individuals a person comes into physical proxim-



**Figure 10.13**  
**The Sex Web**

Cumulative distribution of the total number of sexual partners  $k$  since sexual initiation for individuals interviewed in the 1996 study on sexual patterns in Sweden [15]. For women a power law fit to the tail indicates  $\gamma=3.1\pm 0.3$  for  $k>20$ ; for men  $\gamma=2.6\pm 0.3$  in the range  $20<k<400$ . Note that for men the average number of partners is higher than for women. This difference may be rooted social bias, prompting males to exaggerate and females to suppress the number of sexual partners they report. After [16].

## BOX 10.1

### SEXUAL HUBS

Anecdotal evidence suggests that sexual hubs are real. Take for example Wilt Chamberlain, a Hall of Fame basketball player in the 1980s, who claimed having sex with a staggering number of 20,000 partners. “Yes, that’s correct, twenty thousand different ladies,” he wrote in his autobiography [18]. “At my age, that equals to having sex with 1.2 woman a day, every day, since I was fifteen years old.” Within the AIDS literature the story of Geetan Dugas, a flight attendant with approximately 250 homosexual partners, is well documented [19]. He is often called *patient zero*, whom, given his extensive travel, became a super-spreader of AIDS within the gay community. Hubs are observed even in high school romantic networks (Figure 10.14).

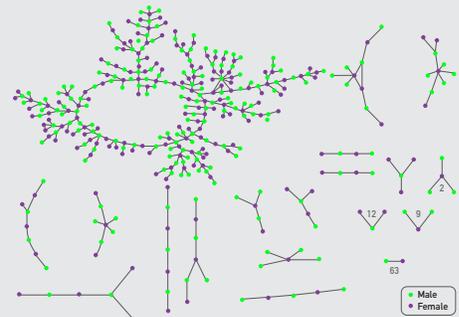


Figure 10.14  
Romantic Links in a High School

Romantic and sexual links between high school students in midwestern United States. Each circle represents a student and the links represent romantic relationships during six months preceding the interview. The numbers indicate the frequency of each subgraph: there are 63 couples isolated from the rest of the network. After [20].

ity. The structure of this contact network is explored at two levels. First, the global travel network allows us to predict the worldwide spread of a pathogen, representing the input of several large-scale epidemic prediction tools (SECTION 10.7). Second, digital badges probe the local properties of the contact network, i.e. the number of individuals a person directly interacts with.

### Global Travel Network

To predict the spread of pathogens, we must know how far infected individuals travel. Our understanding of individual travel patterns exploded with the use of mobile phones, that offer direct information about individual mobility [21-24]. In the context of epidemic phenomena, the most studied mobility data comes from air travel, the mode of transportation that determines the speed with which a pathogen moves around the globe. Consequently the *air transportation network*, that connects airports with direct flights, plays a key role in modeling and predicting the spread of pathogens [25-27]. As Figure 10.15 shows, this network is scale-free with degree exponent  $\gamma=1.8$ . This low value is possible because there are multiple flights between two airports, hence the network is not simple. A similar power law distribution is detected for the link weights, indicating that the number of passengers traveling between two airports is typically low, but between some airports the traffic can be extraordinary. As we discuss in SECTION 10.5, these heterogeneities play a key role in the spread of specific pathogens.

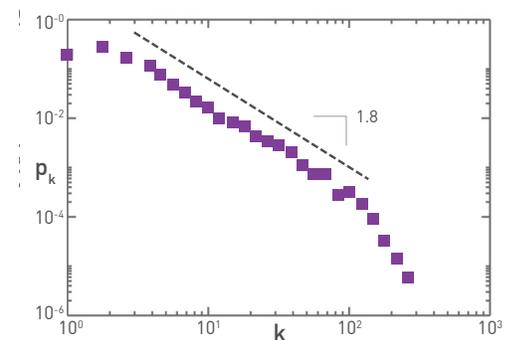


Figure 10.15  
Air Transportation Network

The degree distribution of the air transportation network is well approximated by a power-law with  $\gamma=1.8\pm 0.2$ . The map was built using the International Air Transport Association database that contains the world list of airport and the direct flights between them in 2002. The resulting network is a weighted graph containing the  $N=3,100$  largest airports as nodes that are connected by  $L=17,182$  direct flights as links, together accounting for 99% of the worldwide traffic. After [25].

### Local Contact Patterns

Many airborne diseases spread thanks to face-to-face interactions [28-31]. These interaction patterns can be monitored using Radio-Frequency Identification Devices (RFID) [29,31], mobile-phone based sociometric badges [32,33], and other wireless technologies [34].

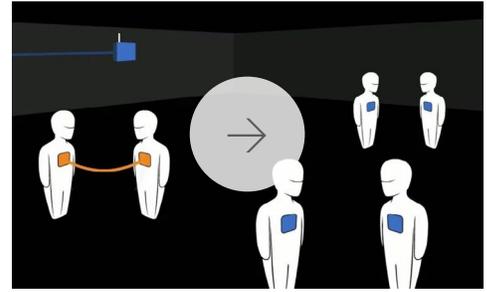
RFID are digital badges that detect the proximity of other individuals that wear a badge (Online Resource 10.1). They have been deployed in various environment, capturing for example the interactions between more than 14, 000 visitors of a Science Gallery over a three month period or between 100 participants of a three-day conference [29]. An RFID-mapped network shown in Figure 10.16 captures the interactions between high school students and their teachers during a two-day period. Several findings stand out:

- RFID tags detect interactions only with individuals that wear the same badge and face each other, limiting the number of detected contacts. Consequently the contact networks mapped out in these studies typically have an exponential degree distribution.
- The duration of each face-to-face interaction follows a power law distribution over several orders of magnitude. Therefore most contacts are brief, but there are a few lasting interactions, documenting bursty temporal pattern [35] with key consequences for the spread of pathogens (SECTION 10.5).
- The link weights, which capture the *cumulative time* two individuals have spent together, also follow a power law distribution. Therefore individuals spend most of their time with only a few others, again with important implications on spreading patterns (SECTION 10.5).
- For most airborne pathogens spatial proximity is sufficient for transmission. For example, standing next to an infected individual in the elevator may be sufficient to transmit SARS or H1N1, an interaction not recorded by a RFID tag.

In summary, RFID tags provide remarkably detailed temporal and spatial information about local contacts. To be useful these studies must be scaled up, using for example mobile phone based technologies [36].

### LOCATION NETWORKS

For many airborne pathogens the relevant contact network is the so-called *location network*, whose nodes are the locations that are connected by individuals that move regularly between them. Measurements combined with agent-based simulations indicate that the location network is fat tailed [37]: malls, airports, schools or supermarkets act as hubs, being linked to an exceptionally large number of smaller locations, like homes and offices. Therefore, once the pathogen infects a hub, the disease can rapidly reach many other locations.



#### Online Resource 10.1 Detecting Networks via RFIDs

A video introducing the RFID technology and their use in mapping social interactions.

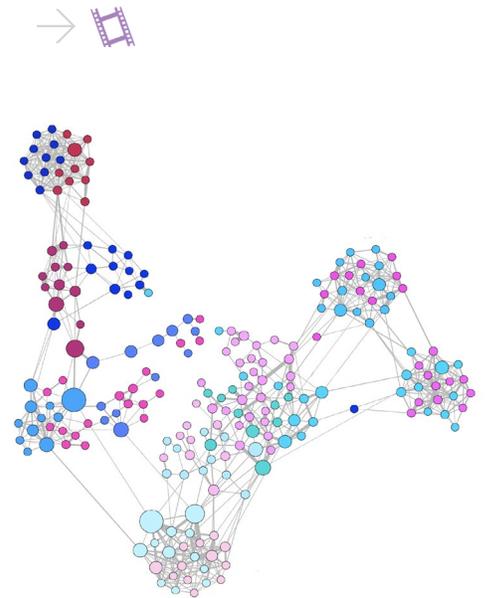


Figure 10.16  
Face-to-face Interactions

A face-to-face contact network mapped out using RFA tags, capturing interactions between 232 students and 10 teachers across 10 classes in a school [31]. The structure of the maps obtained by RFID tags depend on the context in which they are collected. For example the school network shown here reveals the presence of clear communities. In contrast, a study capturing the interactions between individuals that visited a museum reveal an almost linear network [29]. Finally, a network of attendees of a small conference is rather dense, as most participants interact with most others [29]. After [31].

## DIGITAL VIRUSES

The study of digital viruses, that infect computers and smart phones, represents an increasingly important application of epidemic phenomena. As we discuss next, the relevant contact networks are determined by the spreading mode of the respective digital pathogen.

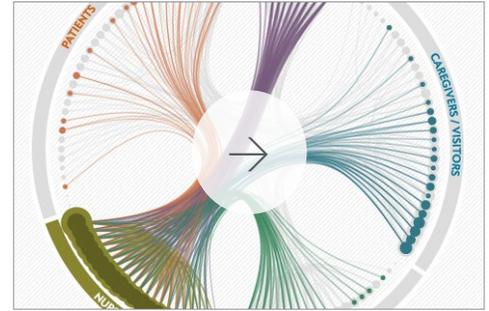
### Computer Viruses

Computer viruses display just as much diversity as biological viruses: depending on the nature of the virus and its spreading mechanism, the relevant contact network can differ dramatically. Many computer viruses spread as email attachments. Once a user opens the attachment, the virus infects the user's computer and mails a copy of itself to the email addresses found in the computer. Hence the pertinent contact network is the email network, which, as we discussed in Table 4.1, is scale-free [58]. Other computer viruses exploit various communication protocols, spreading on networks that reflect the Internet's pattern of interconnectedness, which is again scale-free (Table 4.1). Finally, some malware scan IP addresses, spreading on fully connected networks.

### Mobile Phone Viruses

Mobile phone viruses spread via MMS and Bluetooth (Figure 10.2). An MMS virus sends a copy of itself to all phone numbers found in the phone's contact list. Therefore MMS viruses exploit the social network behind mobile communications. As shown in Table 4.1, the mobile call network is scale-free with a high degree exponent. Mobile viruses can also spread via Bluetooth, passing a copy of themselves to all susceptible phones with a BT connection in their physical proximity. As discussed above, this co-location network is also highly heterogenous [4].

In summary, in the past decade technological advances allowed us to map out the structure of several networks that support the spread of biological or digital viruses, from sexual to proximity-based contact networks (see also ONLINE RESOURCE 10.2). Many of these, like the email network, the internet, or sexual networks, are scale-free. For others, like co-location networks, the degree distribution may not be fitted with a simple power law, yet show significant degree heterogeneity with high  $\langle k^2 \rangle$ . This means that the analytical results obtained in the previous section are of direct relevance to pathogens spreading on most networks. Consequently the underlying heterogenous contact networks allow even weakly virulent viruses to easily spread in the population.



### Online Resource 10.2 Hospital Outbreaks

Bacteria resistant to current antibiotics pose an important threat to global health. Such bacteria are particularly prevalent in hospitals and health care facilities. The Interactive Feature by Scientific American describes the tracking of bacterial outbreaks in hospitals.



# BEYOND THE DEGREE DISTRIBUTION

So far we have kept our models simple: We assumed that pathogens spread on an unweighted network uniquely defined by its degree distribution. Yet, real networks have a number of characteristics that are not captured by  $p_k$  alone, like degree correlations or community structure. Furthermore, the links are typically weighted and the interactions have a finite temporal duration. In this section we explore the impact of these properties on the spread of a pathogen.

## TEMPORAL NETWORKS

Most interactions that we perceive as social links are brief and infrequent. As a pathogen can be only transmitted when there is an actual contact, an accurate modeling framework must also consider the timing and the duration of each interaction. Ignoring the timing of the interactions can lead to misleading conclusions [39-41]. For example, the static network of Figure 10.17b was obtained by aggregating the individual interactions shown in Figure 10.17a. On the aggregated network the infection has the same chance of spreading from D to A as from A to D. Yet, by inspecting the timing of each interaction, we realize that while an infection starting from A can infect D, an infection that starts at D cannot reach A. Therefore, to accurately predict an epidemic process we must consider the fact that pathogens spread on *temporal networks*, a topic of increasing interest in network science [40-43]. By ignoring the temporality of these contact patterns, we typically overestimate the speed and the extent of an outbreak [42,43].

## BURSTY CONTACT PATTERNS

The theoretical approaches discussed in the SECTIONS 10.2 and 10.3 assume that the timing of the interactions between two connected nodes is random. This means that the interevent times between consecutive contacts follow an exponential distribution, resulting in a random but uniform sequence of events (Figure 10.18a-c). The measurements indicate otherwise: The interevent times in most social systems follow a power law distribution [35,44] (Fig. 10.18d-f). This means that the sequence of contacts

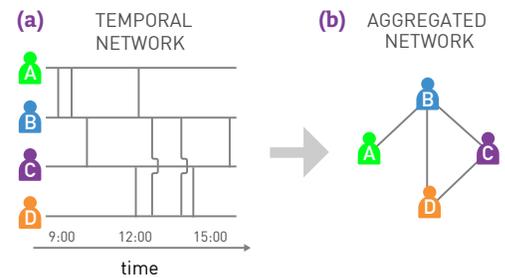


Figure 10.17  
Temporal Networks

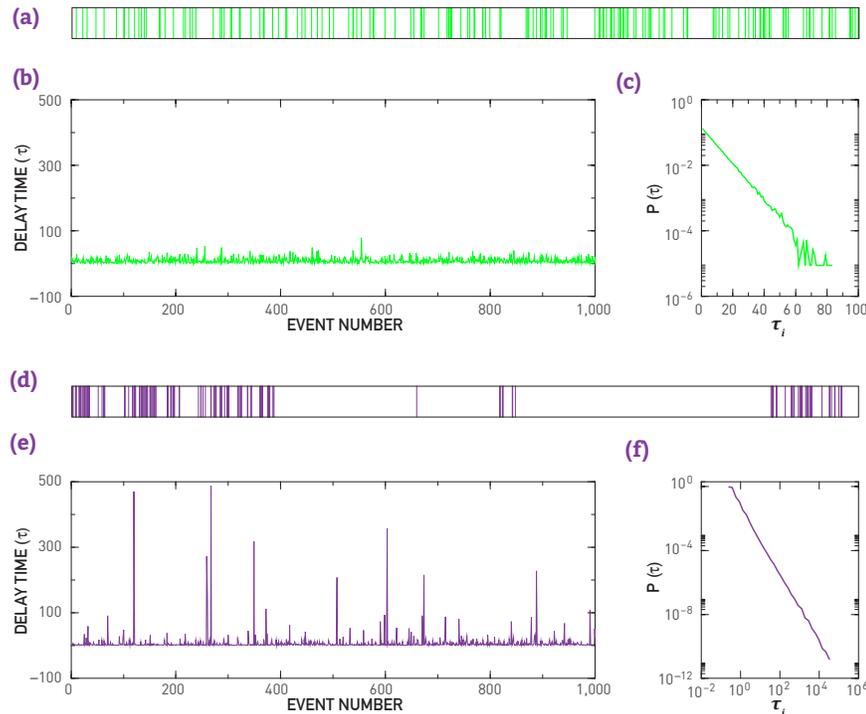
Most interactions in a network are not continuous, but have a finite duration. We must therefore view the underlying networks as *temporal networks*, an increasingly active research topic in network science.

### (a) Temporal Network

The timeline of the interactions between four individuals. Each vertical line marks the moment when two individuals come into contact with each other. If A is the first to be infected, the pathogen can spread from A to B and then to C, eventually reaching D. If, however, D is the first to be infected, the disease can reach C and B, but not A. This is because there is a temporal path from A to D.

### (b) Aggregated Network

The network obtained by merging the temporal interactions shown in (a). If we only have access to this aggregated representation, the pathogen can reach all individuals, independent of its starting point. After [40].



**Figure 10.18**  
**Bursty Interactions**

- (a) If the pattern of activity of an individual is random, the interevent times follow a Poisson process, which assumes that in any moment an event takes place with the same probability  $q$ . The horizontal axis denotes time and each vertical line corresponds to an event whose timing is chosen at random. The observed inter-event times are comparable to each other and very long delays are rare.
- (b) The absence of long delays is visible if we show the inter-event times  $\tau_i$  for 1,000 consecutive random events. The height of each vertical line corresponds to the gaps seen in (a).
- (c) The probability of finding exactly  $n$  events within a fixed time interval follows the Poisson distribution  $P(n,q)=e^{-qt}(qt)^n/n!$ , predicting that the inter-event time distribution follows  $P(\tau_i)\sim e^{-q\tau_i}$ , shown on a log-linear plot.
- (d) The succession of events for a temporal pattern whose interevent times follow a power-law distribution. While most events follow each other closely, forming bursts of activity, there are a few exceptionally long interevent times, corresponding to long gaps in the contact pattern. The time sequence is not as uniform as in (a), but has a bursty character.
- (e) The waiting time  $\tau_i$  of 1,000 consecutive events, where the mean event time is chosen to coincide with the mean event time of the Poisson process shown in (b). The large spikes correspond to exceptionally long delays.
- (f) The delay time distribution  $P(\tau_i)\sim\tau_i^{-2}$  for the bursty process shown in (d) and (e). After [35].

between two individuals is characterized by periods of frequent interactions, when multiple contacts follow each other within a relatively short time frame. Yet, the power law also implies that occasionally there are a very long time gaps between two contacts. Therefore the contact patterns have an uneven, “bursty” character in time (Figure 10.18d,e).

Bursty interactions are observed in a number of contact processes of relevance for epidemic phenomena, from email communications to call patterns and sexual contacts. Once present, burstiness alters the dynamics of the spreading process [43]. To be specific, power law interevent times increase the characteristic time  $\tau$ , consequently the number of infected individuals decays slower than predicted by a random contact pattern. For example, if the time between consecutive emails would follow a Poisson distribution, an email virus would decay following  $i(t)\sim\exp(-t/\tau)$  with a decay time of  $\tau\approx 1$  day. In the real data, however, the decay time is  $\tau\approx 21$  days, a much slower process, correctly predicted by the theory if we use power law interevent times [43].

### DEGREE CORRELATIONS

As discussed in CHAPTER 7, many social networks are assortative, implying that high degree nodes tend to connect to other high degree nodes. Do these degree correlations affect the spread of a pathogen? The calculations indicate that degree correlations leave key aspects of network epidemics in place, but they alter the speed with which a pathogen spreads in a network:

- Degree correlations alter the epidemic threshold  $\lambda_c$ : assortative correlations decrease  $\lambda_c$  and disassortative correlations increase it [45,46].

- Despite the changes in  $\lambda_c$ , for the SIS model the epidemic threshold vanishes for a scale-free network with diverging second moment, whether the network is assortative, neutral or disassortative [47]. Hence the fundamental results of SECTION 10.3 are not affected by degree correlations.
- Given that hubs are the first to be infected in a network, assortativity accelerates the spread of a pathogen. In contrast disassortativity slows the spreading process.
- Finally, in the SIR model assortative correlations were found to lower the prevalence but increase the average lifetime of an epidemic outbreak [48].

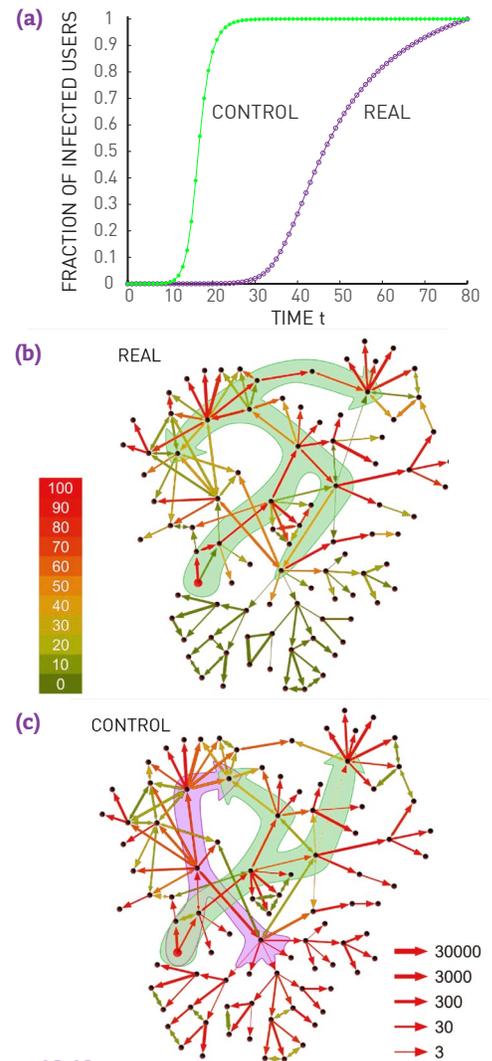
### LINK WEIGHTS AND COMMUNITIES

Throughout this chapter we assumed that all tie strengths are equal, focusing our attention on pathogens spreading on an unweighted network. In reality tie strengths vary considerably, a heterogeneity that plays an important role in spreading phenomena. Indeed, the more time an individual spends with an infected individual, the more likely that she too becomes infected.

In the same vein, previously we ignored the community structure of the network on which the pathogen spreads. Yet, the existence communities (CHAPTER 9) leads to repeated interactions between the nodes within the same community, altering the spreading dynamics.

The mobile phone network allows us to explore the role of tie strengths and communities on spreading phenomena [49]. Let us assume that at  $t=0$  we provide a randomly selected individual with some key information. At each time step this “infected” individual  $i$  passes the information to her contact  $j$  with probability  $p_{ij} \sim \beta w_{ij}$ , where  $\beta$  is the spreading probability and  $w_{ij}$  is the strength of the ties captured by the number of minutes  $i$  and  $j$  have spent with each other on the phone. Indeed, the more time two individuals talk, the higher is the chance that they will pass on the information. To understand the role of the link weights in the spreading process, we also consider the situation when the spreading takes place on a *control network*, that has the same wiring diagram but all tie strengths are set equal to  $w = \langle w_{ij} \rangle$ .

As Figure 10.19a illustrates, information travels significantly faster on the control network. The reduced speed observed in the real system indicates that the information is trapped within communities. Indeed, as we discussed in CHAPTER 9, strong ties tend to be within communities while weak ties are between them [50]. Therefore, once the information reaches a member of a community, it can rapidly reach all other members of the same community, given the strong ties between them. Yet, as the ties between the communities are weak, the information has difficulty escaping the community. Consequently the rapid invasion of the community is followed by long intervals during which the infection is trapped within a



**Figure 10.19**  
**Information Diffusion in Mobile Phone Networks**

The spread of information on a weighted mobile call graph, where the probability that a node passes information to one of its neighbors is proportional to the strength of the tie between them. The tie strength is the number of minutes two individuals talk on the phone.

- (a) The fraction of infected nodes in function of time. The blue circles capture the spread on the network with the real tie strengths; the green symbols represent the control case, when all tie strengths are equal.
- (b) Spreading in a small network neighborhood, following the real link weights. The information is released from the red node, the arrow weight indicating the tie strength. The simulation was repeated 1,000 times; the size of the arrowheads is proportional to the number of times the information was passed along the corresponding direction, and the color indicates the total number of transmissions along that link. The background contours highlight the difference in the direction the information follows in the real and the control simulations.
- (c) Same in (b), but we assume that each link has the same weight  $w = \langle w_{ij} \rangle$  (control). After [49].

community. When all link weights are equal (control), the bridges between communities are strengthened, and the trapping vanishes.

The difference between the real and the control spreading process is illustrated by Figure 10.20b,c, that shows the spreading pattern in a small neighborhood of the mobile call network. In the control simulation the information tends to follow the shortest path. When the link weights are taken into account, information flows along a longer backbone with strong ties. For example, the information rarely reaches the lower half of the network in Figure 10.20b, a region always reached in the control simulation shown in (c).

### COMPLEX CONTAGION

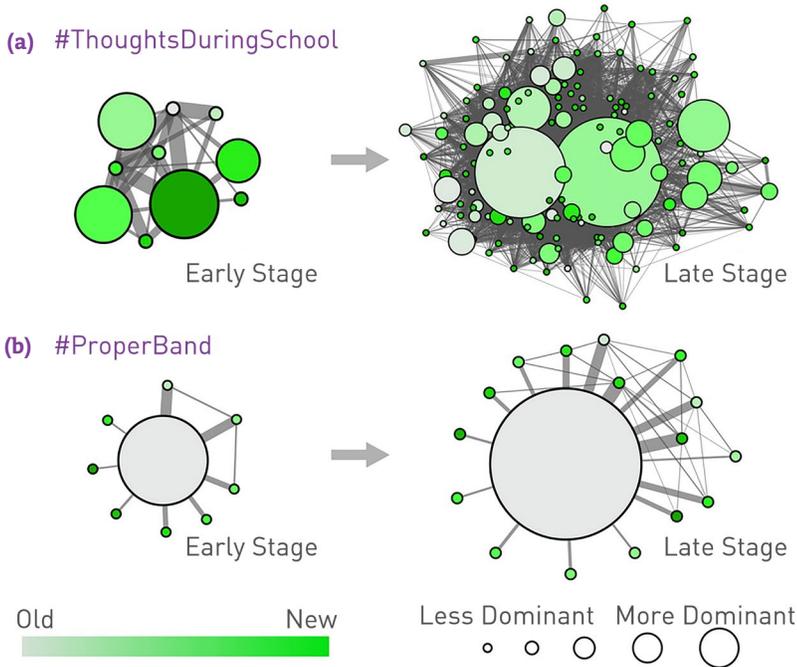
Communities have multiple consequences for spreading, from inducing global cascades [51,52] to altering the activity of individuals [53].

The diffusion of memes, representing ideas or behavior that spread from individual to individual, further highlights the important role of communities [54]. Meme diffusion has attracted considerable attention from marketing [5, 55] to network science [56,57], communications [58], and social media [59-61]. Pathogens and memes can follow different spreading patterns, prompting us to systematically distinguish simple from complex contagion [54,62,63].

*Simple contagion* is the process we explored so far: It is sufficient to come into contact with an infected individual to be infected. The spread of memes, products and behavior is often described by *complex contagion*, capturing the fact that most individuals do not adopt a new meme, product or behavioral pattern at the first contact. Rather, adoption requires reinforcement [64], i.e. repeated contact with several individuals who have already adopted. For example, the higher is the fraction of a person's friends that have a mobile phone, the more likely that she also buys one.

In simple contagion communities trap an information or a pathogen, slowing the spreading (Figure 10.19a). The effect is reversed in complex contagion: Because communities have redundant ties, they offer social reinforcement, exposing an individual to multiple examples of adoption. Hence communities can incubate a meme, a product or a behavioral pattern, enhancing its adoption.

The difference between simple and complex contagion is well captured by Twitter data. Tweets, or short messages, are often labeled with *hashtags*, which are keywords acting as memes. Twitter users can follow other users, receiving their messages; they can forward tweets to their own followers (*retweet*), or mention others in tweets. The measurements indicate that most hashtags are trapped in specific communities, a signature of complex contagion [54]. A high concentration of a meme within a certain community is evidence of reinforcement. In contrast, viral memes spread across communities, following a pattern similar to that encountered in bi-



**Figure 10.20**  
**Simple vs. Complex Contagion**

The community structure of the Twitter follower network. Each circle corresponds to a community and its size is proportional to the number of tweets produced by the respective community. The color of a community represents the time when the studied hashtag (meme) is first used in the community. Lighter colors denote the first communities to use a hashtag, darker colors denote the last community to adapt it.

**(a) Simple Contagion**

The evolution of the viral meme captured by the #ThoughtsDuringSchool hashtag from its early stage (30 tweets, left) to the late stage (200 tweets, right). The meme jumps easily between communities, infecting many of them, following a contagion pattern encountered in the case of biological pathogens.

**(b) Complex Contagion**

The evolution of a non-viral meme captured by the #ProperBand hashtag from the early stage (left) to the final stage (65 tweets, right). The tweet is trapped in a few of communities, having difficulty to escape them. This is a signature of reinforcement, an indication that the meme follows complex contagion. After [54].

ological pathogens. In general the more communities a meme reaches, the more viral it is (Figure 10.20).

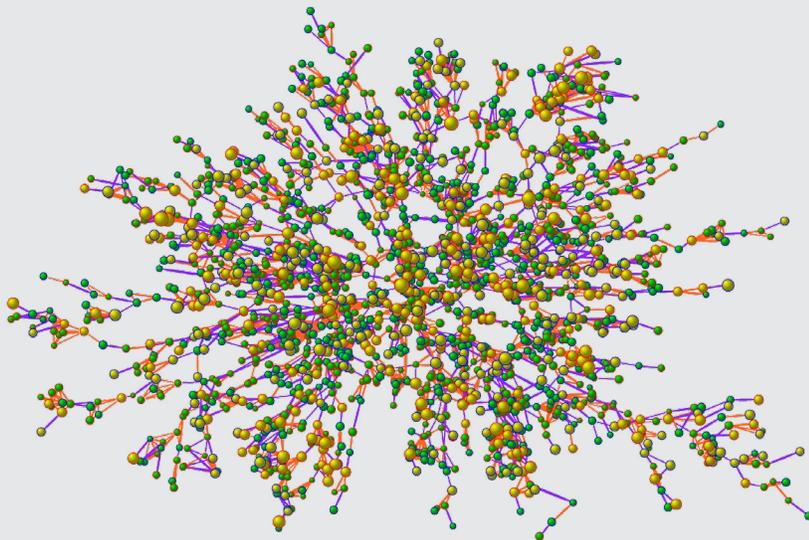
In summary, several network characteristics can affect the spread of a pathogen in a network, from degree correlations to link weights and the bursty nature of the contact pattern. As we discussed in this section, some network characteristics slow a pathogen, others aid their spread. These effects must therefore be accounted for if we wish to predict the spread of a real pathogen. While these patterns are of obvious relevance for infectious diseases, they also influence the spread of such non-infectious diseases as obesity (BOX 10.2).

## BOX 10.2

### DO OUR FRIENDS MAKE US FAT?

Infectious diseases, like influenza, SARS, or AIDS, spread through the transmission of a pathogen. But could the social network aid the spread of noninfectious diseases as well? Recent measurements indicate that it does, offering evidence that social networks can impact the spread of obesity, happiness, and behavioral patterns, like giving up smoking [65,66].

Obesity is diagnosed through an individual's body-mass index (BMI), which is determined by numerous factors, from genetics to diet and exercise. The measurements show that our friends also play an important role. The analysis of the social network of 5,209 men and women has found that if one of our friends is obese, the risk that we too gain weight in the next two to four years increases by 57% [65]. The risk triples if our best friend is overweight: In this case, our chances of weight gain jumps by 171% (Figure 10.21). For all practical purposes, obesity appears to be just as contagious as influenza or AIDS, despite the fact that there is no "obesity pathogen" that transmits it.



### Online Resource 10.3

#### Spreading in Social Networks

"If your friends are obese, your risk of obesity is 45 percent higher. ... If your friend's friends are obese, your risk of obesity is 25 percent higher. ... If your friend's friend's friend, someone you probably don't even know, is obese, your risk of obesity is 10 percent higher. It's only when you get to your friend's friend's friend's friends that there's no longer a relationship between that person's body size and your own body size."

Watch Nicholas Christakis explaining the spread of health patterns in social networks.



Figure 10.21

#### The Web of Obesity

The largest connected component of the social network capturing the friendship ties between 2,200 individuals enrolled in the Framingham Heart Study. Each node represents an individual; nodes with blue borders are men, those with red borders are women. The size of each node is proportional to the person's BMI, yellow nodes denoting obese individuals ( $BMI \geq 30$ ). Purple links are friendship or marital ties and orange links are family ties (e.g. siblings). Clusters of obese and non-obese individuals are visible in the network. The analysis indicates that these clusters cannot be attributed to homophily, i.e. the fact that individuals of similar body size may befriend with each other. They document instead a complex contagion process, capturing the "spread" of obesity along the links of the social network. After [65].

# IMMUNIZATION

Immunization strategies specify how vaccines, treatments or drugs are distributed in the population. Ideally, should a treatment or vaccine exist, it should be given to every infected individual or those at risk of contracting the pathogen. Yet, often cost considerations, the difficulty of reaching all individuals at risk, and real or perceived side effects of the treatment prohibit full coverage. Given these constraints, immunization strategies aim to minimize the threat of a pandemic by most effectively distributing the available vaccines or treatments.

Immunization strategies are guided by an important prediction of the traditional epidemic models: If a pathogen's spreading rate  $\lambda$  is reduced under its critical threshold  $\lambda_c$ , the virus naturally dies out (Figure 10.11). Yet, the epidemic threshold vanishes in scale-free networks, questioning the effectiveness of this strategy. Indeed, if the epidemic threshold vanishes, immunization strategies can not move  $\lambda$  under  $\lambda_c$ . In this section we discuss how to use our understanding of the network topology to design effective network-based immunization strategies that counter the impact of the vanishing epidemic threshold.

## RANDOM IMMUNIZATION

The main purpose of immunization is to protect the immunized individual from an infection. Equally important, however, is its secondary role: Immunization reduces the speed with which the pathogen spreads in a population. To illustrate this effect consider the situation when a randomly selected  $g$  fraction of individuals are immunized in a population [8].

Let us assume that the pathogen follows the SIS model (10.3). The immunized nodes are invisible to the pathogen, and only the remaining  $(1-g)$  fraction of the nodes can contact and spread the disease. Consequently, the effective degree of each susceptible node changes from  $\langle k \rangle$  to  $\langle k \rangle (1-g)$ , which decreases the spreading rate of the pathogen from  $\lambda = \beta/\mu$  to  $\lambda' = \lambda(1-g)$ . Next we explore the consequences of this reduction in both random and scale-free contact networks.

• **Random Networks**

If the pathogen spreads on a random network, for a sufficiently high  $g$  the spreading rate  $\lambda'$  could fall below the epidemic threshold (10.25). The immunization rate  $g_c$  necessary to achieve this is calculated by setting

$$\frac{(1-g_c)\beta}{\mu} = \frac{1}{\langle k \rangle + 1},$$

obtaining

$$g_c = 1 - \frac{\mu}{\beta} \frac{1}{\langle k \rangle + 1}. \tag{10.27}$$

Consequently, if vaccination increases the fraction of immunized individuals above  $g_c$ , it pushes the spreading rate under the epidemic threshold  $\lambda_c$ . In this case  $\tau$  becomes negative and the pathogen dies out naturally. This explains why health officials encourage a high fraction of the population take the influenza vaccine: The vaccine protects not only the individual, but also the rest of the population by decreasing the pathogen's spreading rate. Similarly, a condom not only protects the individual who uses it from contacting the HIV virus, but also decrease the rate at which AIDS spreads in the sexual network. Hence for random networks a sufficiently high immunization rate can eliminate the pathogen from the population.

• **Heterogenous Networks**

If the pathogen spreads on a network with high  $\langle k^2 \rangle$ , and random immunization changes  $\lambda$  to  $\lambda(1-g)$ , we can use (10.26) to determine the critical immunization  $g_c$

$$\frac{\beta}{\mu}(1-g_c) = \frac{\langle k \rangle}{\langle k^2 \rangle} \tag{10.28}$$

obtaining

$$g_c = 1 - \frac{\mu}{\beta} \frac{\langle k \rangle}{\langle k^2 \rangle}. \tag{10.29}$$

For a random network (10.29) reduces to (10.27). For a scale-free network with  $\gamma < 3$  we have  $\langle k^2 \rangle \rightarrow \infty$ , hence (10.29) predicts  $g_c \rightarrow 1$ . In other words if the contact network has a high  $\langle k^2 \rangle$ , we need to immunize virtually all nodes to stop the epidemic. This prediction is consistent with the finding that for many diseases we must immunize 80%-100% of the population to eradicate the pathogen. For example, measles requires 95% of the population to be immunized [8]; for digital viruses the strategies relying on random immunization call for close to 100% of the computers to install the appropriate antivirus software [67].

To illustrate the role degree heterogeneity plays in immunization let us consider a digital virus spreading on the email network. If we make the email network random and undirected, we

## BOX 10.3

### HOW TO HALT AN EPIDEMIC?

Health safety officials rely on several interventions to control or delay an epidemic outbreak. Some of the most common interventions include:

#### Transmission-Reducing Interventions

Face masks, gloves, and hand washing reduces the transmission rate of airborne or contact based pathogens. Similarly, condoms reduce the transmission rate of sexually transmitted pathogens.

#### Contact-Reducing Interventions

For diseases with severe health consequences officials can quarantine patients, close schools and limit access to frequently visited public spaces, like movie theaters and malls. These make the network sparser by reducing the number of contacts between individuals, hence decreasing the transmission rate.

#### Vaccinations

Vaccinations permanently remove the vaccinated nodes from the network, as they cannot be infected nor can they spread the disease. Vaccinations also reduce the spreading rate, enhancing the likelihood that the pathogen dies out.

have  $\langle k \rangle = 3.26$ . Using  $\lambda = 1$  in (10.27) we obtain  $g_c = 0.76$ . In other words, to eradicate the virus we need to convince 76% of computer users to update their antivirus software. Yet, the email network is scale-free with  $\langle k^2 \rangle = 1,271$  (undirected version), hence (10.27) does not apply. In this case (10.29) predicts  $g_c = 0.997$  for  $\lambda = 1$ , meaning that more than 99.7% of the users must install the software to halt the email virus. It is virtually impossible to achieve this level of compliance - many users simply ignore all warnings. This is the reason why email viruses linger for years and disappear only after the operating systems that supports them is phased out [67].

### VACCINATION STRATEGIES IN SCALE-FREE NETWORKS

The ineffectiveness of random immunization is rooted in the vanishing epidemic threshold. Consequently, to successfully eradicate a pathogen in heterogenous networks, we must find ways to increase the epidemic threshold. This requires us to reduce the variance,  $\langle k^2 \rangle$ , of the underlying contact network.

The hubs are responsible for the large variance of heterogenous networks. Therefore if we immunize the hubs, i.e. all nodes whose degree exceeds some preselected  $k'_{\max}$ , we decrease the variance and increase the epidemic threshold according to (10.26) [68,69]. Indeed, if nodes with degrees  $k > k'_{\max}$  are absent, the epidemic threshold changes to (ADVANCED TOPICS 10.C)

$$\lambda_c \approx \frac{\gamma - 2}{3 - \gamma} \frac{k_{\min}^{2-\gamma}}{(k'_{\max})^{\gamma-3}}. \quad (10.30)$$

Therefore, for  $\gamma < 3$ , the more hubs we cure (i.e. the smaller is  $k'_{\max}$ ), the larger will be the epidemic threshold (Figure 10.22). By immunizing a sufficient fraction of the hubs we can drop  $\lambda_c$  below  $\lambda = \beta/\mu$  that characterizes the pathogen. This procedure is equivalent with altering the underlying network: By immunizing the hubs, we are fragmenting the contact network, making more difficult for the pathogen to reach the nodes in other components (Figure 10.23).

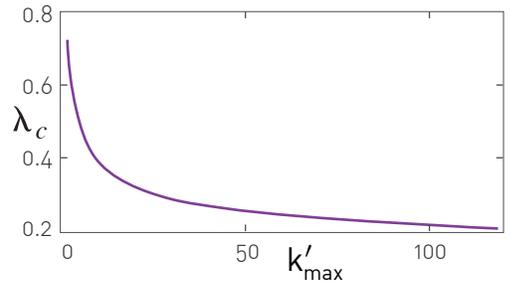


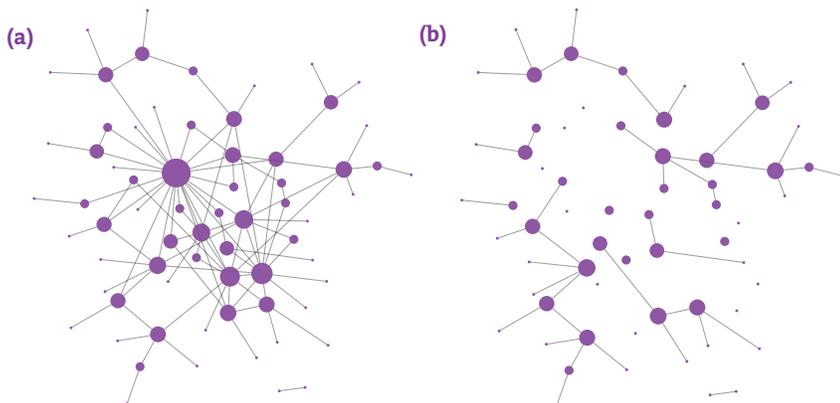
Figure 10.22  
Immunizing the Hubs

In heterogenous networks a virus can be eradicated by increasing the epidemic threshold through hub immunization. The figure shows the expected epidemic threshold if we immunize all nodes with degree larger than  $k'_{\max}$ . The more hubs are immunized (i.e. the smaller is  $k'_{\max}$ ), the larger is  $\lambda_c$ , increasing the chance that the disease dies out. Immunizing the hubs changes the network on which the disease spreads, making the hubs invisible to the pathogen (Figure 10.23).

Figure 10.23  
Robustness and Immunization

Scale-free networks show a remarkable resilience to random node and link failures (CHAPTER 8). At the same time, they are vulnerable to attacks: If we remove their most connected nodes, scale-free networks break apart. This phenomena has many similarities to the immunization problem: Random immunization is unable to eradicate a disease, but selective immunization, that targets the hubs, can restore a finite critical threshold, helping us eradicate the disease. The analogy is not accidental: The robustness and the immunization problem can be both linked to the diverging  $\langle k^2 \rangle$ . Indeed, the vanishing epidemic threshold is equivalent with the finding that the percolation threshold under random node removal problem converges to one (ADVANCED TOPICS 10.D). Similarly, the re-emergence of the epidemic threshold under hub immunization is equivalent with the small percolation threshold characterizing a scale-free network under attack. Therefore, the attack and targeted immunization problems represent two sides of the same coin.

To illustrate the equivalence between attacks and targeted immunization, consider the network shown in (a). An attack that removes its five largest hubs breaks the network into many isolated islands, as shown in (b). Targeted immunization plays the same role: By making the hubs immune to the disease, the network on which the pathogen spreads becomes the fragmented network in (b). As the immunized network is broken into small islands, the pathogen will be stuck in one of the small clusters, unable to infect the nodes in the other clusters.



Hub immunization represents a perspective change in immunization protocols: instead of trying to decrease the spreading rate using random immunization, we must alter the topology of the contact network, which in turn increases  $\lambda_c$  above the biologically determined  $\lambda = \beta/\mu$ .

The problem with a hub-based immunization strategy is that for most epidemic processes we lack a detailed map of the contact network. Indeed, we do not know the number of sexual partners each individual has in a population, nor can we accurately identify the super-spreaders during an influenza outbreak. In other words it is difficult to identify the hubs. Yet, we can still exploit the network topology to design more efficient immunization strategies. To do so, we rely on the friendship paradox, the fact that on average the neighbors of a node have higher degree than the node itself (BOX 7.1). Therefore, by immunizing the acquaintances of a randomly selected individual, we target the hubs without having to know precisely which individuals are hubs. The procedure consists of the following steps [70]:

- 1) Choose randomly a  $p$  fraction of nodes, like we do during random immunization. Call these nodes Group 0.
- 2) Select randomly a link for each node in Group 0. We call Group 1 the set of nodes to which these links connect to. For example, we ask each individual from Group 0 to nominate one of its acquaintance with whom he/she engaged in an activity that could have resulted in the transmission of the pathogen. In the case of HIV, ask them to name a sexual partner.
- 3) Immunize the Group 1 individuals.

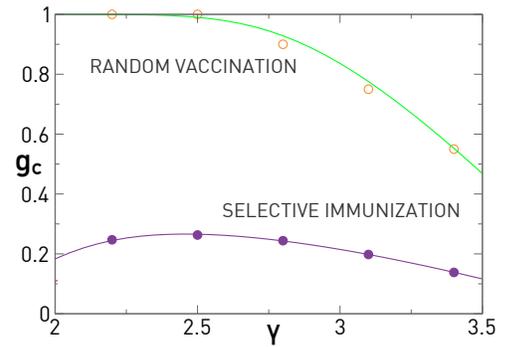
This strategy requires no information about the global structure of the network. Yet, according to (7.3) the probability that a node with  $k$  links belongs to Group 1 is proportional to  $kp_k$ . Consequently the Group 1 individuals have higher average degree than the Group 0 individuals. The implications of this bias are illustrated in Figure 10.24, which shows the critical threshold required to eradicate a pathogen for a scale-free network with degree exponent  $\gamma$ . The figure offers several key insights:

### 1) Random Immunization

The top curve shows  $g_c$  for random immunization. For heterogeneous networks (small  $\gamma$ ) we find that  $g_c \approx 1$ , indicating that we must immunize all nodes to eradicate the disease. As  $\gamma$  approaches 3 the network develops a finite epidemic threshold and  $g_c$  drops. Hence for large  $\gamma$ , immunizing a sufficiently high fraction of the population can eradicate the pathogen.

### 2) Selective Immunization

For the biased strategy  $g_c$  is systematically under 30%. Therefore by immunizing a randomly chosen neighbor of 30% of the nodes, we could eradicate the disease. The efficiency of this strategy depends



**Figure 10.24**  
Selective Immunization of Scale-free Networks.

The critical immunization threshold  $g_c$  in function of the degree exponent  $\gamma$  of the contact network on which the pathogen spreads following the SIS model. The curves correspond to two immunization strategies: *random immunization* (green) and *selective immunization* (purple), that immunizes a first neighbor of a randomly selected node. The continuous lines represent the analytical results while the symbols represent simulation data for  $N=10^6$  and  $m=1$ . As the population has a finite size, we have  $g_c < 1$  for random immunization even for  $\gamma < 3$ . Redrawn after [70].

only weakly on  $\gamma$ . Selective immunization is more efficient than random immunization even for high  $\gamma$ , when hubs are less prominent.

In summary, if we have the resources to immunize everyone at risk of contacting a pathogen, we should do that - this was the strategy of the eradication campaigns (BOX 10.4). When extensive immunization is not feasible, we need to employ various immunization strategies to maximize the impact of our resources. The effectiveness of each strategy depends on the structure of the contact network on which the pathogen spreads. In general, random immunization is inefficient for pathogens that spread on heterogeneous networks: For random immunization to succeed we need to reach and immunize close to 100% of the susceptible nodes, which is impossible in most circumstances. In contrast, strategies that immunize the hubs have high effectiveness. Selective immunization, that immunizes the neighbors of randomly selected nodes, can significantly enhance effectiveness, without requiring an accurate map of the contact network. This strategy is efficient for both random and heterogeneous networks.

## BOX 10.4

### CAN PATHOGENS BE ERADICATED?

At the end of the 1960s smallpox was still widespread in Africa and Asia. Before 1967 the smallpox eradication strategy relied on mass vaccination, a strategy that was ineffective in densely populated areas. Health officials eventually developed network-based protocols to stop the transmission: They set out to find and treat anyone who had been in contact with an infected individual. This strategy allowed smallpox to become the first disease to be officially eradicated (Figure 10.25).

*Eradication* is the complete elimination of a pathogen from the population. To select an infectious disease for eradication, health officials must make sure that the targeted pathogen does not have a non-human reservoir, so human vaccination can truly eradicate it. There is also need for an efficient and practical vaccine or drug to interrupt its transmission. So far eradication campaigns had mixed success: smallpox and rinderpest were successfully eradicated, but programs targeting hookworm, malaria, and yellow fever have failed.



**Figure 10.25**  
**Eradicating Smallpox**

Rahima Banu, the last smallpox infected patient in Bangladesh in 1976. After [71].

# EPIDEMIC PREDICTION

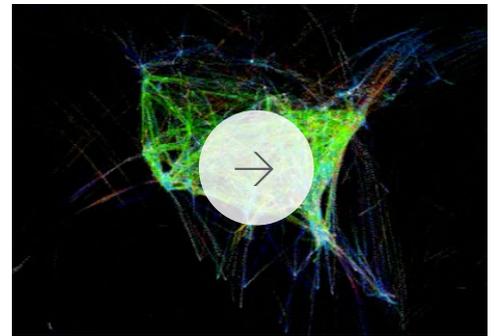
During much of its history humanity has been helpless when faced with a pandemic. Lacking drugs and vaccines, infectious diseases repeatedly swept through continents, decimating the world's population. The first vaccine was tested only in 1796 and the systematic development of vaccines and cures against new pathogens became possible only in the 1990s. Despite the spectacular medical advances, we have effective vaccines only against a small number of pathogens. Consequently transmission-reducing and quarantine-based measures remain the main tools of health professionals in combatting new pathogens. For the combination of vaccines, treatments and quarantine-based measures to be effective, we need to predict when and where the pathogen emerges next, allowing local health officials to best deploy their resources.

The real-time prediction of an epidemic outbreak is a very recent development. The ground was set by the development of the epidemic modeling framework in the 1980s [72] and by the 2003 SARS epidemic, which resulted in worldwide reporting guidelines about ongoing outbreaks. The subsequent systematic availability of data pertaining to a pandemic [1] offered real-time input to modeling efforts. The 2009 H1N1 outbreak was the first beneficiary of these developments, becoming the first pandemic whose spread was predicted in real time.

The emergence of any new pathogen raises several key questions:

- Where did the pathogen originate?
- Where do we expect new cases?
- When will the epidemic arrive at various densely populated areas?
- How many infections are to be expected?
- What can we do to slow its spread?
- How can we eradicate it?

Today these questions are addressed using powerful epidemic simula-



**Online Resource 10.4**  
**North American Flight Patterns**

Real time flights across North America, relying on data released by the Federal Aviation Administration. This global transportation network is responsible for the spread of pathogens across continents. Consequently flight schedules represent the input for epidemic forecasts. While this video, produced by Aaron Koblin, could easily be seen as a purely scientific illustration, it is also viewed as digital art by the art community. Indeed, the video is now in Media Art collection of the Museum of Modern Art (MoMA) in New York.



tors that consider as input demographic, mobility-related (Online Resource 10.4), and epidemiological data [73-75]. The algorithms behind these tools range from stochastic meta-population models [76-78] to agent-based computer simulations that capture the behavior and the interactions of millions of individuals [79]. In this section we summarize the capabilities of these tools, highlighting the role of network science in these developments.

### REAL-TIME FORECAST

Epidemic forecast aims to foresee the real time spread of a pathogen, predicting the number of infected individuals expected each week in each major city [79,80]. The first successful real time pandemic forecast based on network science relied on the Global Epidemic and Mobility (GLEAM) computational model [80] (Figure 10.26, Online Resource 10.5), a stochastic framework that uses as input high-resolution data on worldwide human demography and mobility. GLEAM employs a network-based computational model:

- GLEAM maps each geographic location into the nodes of a network.
- Transport between these nodes, representing the links, are provided by global transportation data, like airline schedules (Online Resource 10.4).
- GLEAM estimates the epidemic parameters, like the transmission rate or reproduction number, using a network-based approach: It relies on chronological data that captures the worldwide spread of the



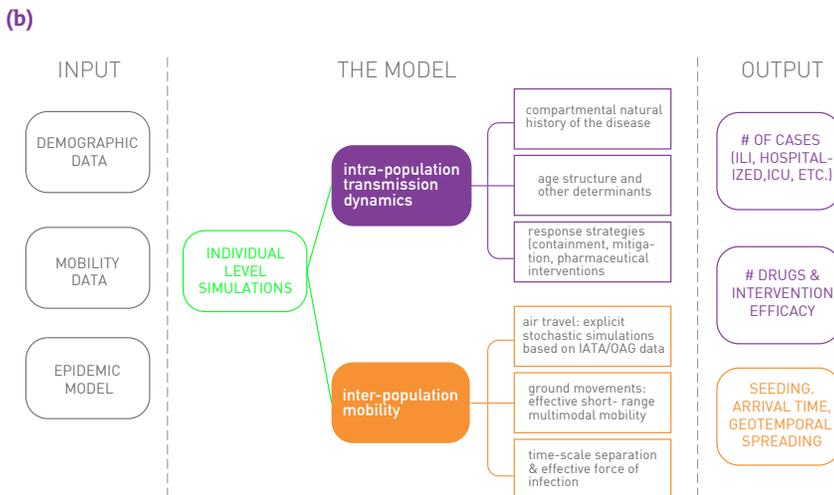
### Online Resource 10.5 GLEAM

A video describing the GLEAM software package for epidemic prediction.



Figure 10.26  
Modeling the 2009 H1N1 Pandemic.

(a) The spread of the H1N1 virus during the early stage of the 2009 outbreak. The arrows represent the arrival of the first infections in previously unaffected countries. The color code indicates the time of the virus' arrival.



(b) The flowchart of the Global Epidemic and Mobility (GLEAM) computational model, used to predict the real-time spread of pathogens like H1N1 or Ebola. The left column (Input) represents the input databases, capturing demographic, mobility and epidemiological information. The center column (model) describes the network-based dynamic processes that are modeled at each time step. The right column (Output) offers examples of quantities the model can predict. After [82].

pandemic, rather than medical reports [81].

GLEAM then implements the network-based epidemic framework described in SECTION 10.3, generating a large number of potential outcomes of the pathogen’s global progression for the coming months. For H1N1 the predictions were compared with data collected from surveillance and virologic sources in 48 countries during the full course of the pandemic [80], resulting in several key findings:

- **Peak Time**

Peak time corresponds to the week when most individuals are infected in a particular country. Predicting the peak time helps health officials decide the timing and the quantity of the vaccines or treatments they distribute. The peak time depends on the arrival time of the first infection and the demographic and the mobility characteristics of each country. The observed peak time fell within the prediction interval for 87% of the countries (Figure 10.27). In the remaining cases the difference between the real and the predicted peak was at most two weeks.

- **Early Peak**

GLEAM predicted that the H1N1 epidemic will peak out in November, rather than in January or February, the typical peak time of influenza-like viruses. This unexpected prediction turned out to be correct, confirming the model’s predictive power. The early peak time was a consequence of the fact that H1N1 originated in Mexico, rather than South Asia (where many flu viruses come from), hence it took the virus less time to arrive to the northern hemisphere.

- **The Impact of Vaccination**

Several countries implemented vaccination campaigns to accelerate the decline of the pandemic. The simulations indicated that these mass vaccination campaigns had only negligible impact on the course of the epidemic. The reason is that the timing of these campaigns was guided by the expectation of a January peak time, prompting the deployment of the vaccines after the November 2009 peak [83], too late to have a strong effect.

### ‘WHAT IF’ ANALYSIS

By incorporating the time and nature of each containment and mitigation procedure, simulations can estimate the efficiency of specific contingency plans [73-75,77,84]. Next we discuss the impact of two such interventions.

- **Travel Restrictions**

Given the important role air travel plays in the spread of a pathogen, faced with a dangerous pandemic, like an Ebola outbreak (Figure 10.28), the first instinct is to restrict travel. Yet, in a world where key resources travel by air, a travel ban leads to economic collapse. Therefore before resorting to a travel ban, we must make sure that

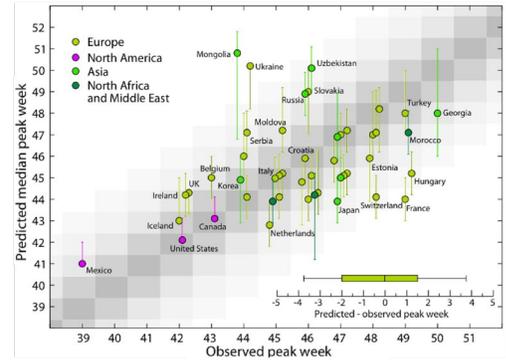


Figure 10.27  
Activity Peaks for H1N1

The predicted and the observed peak time for the H1N1 virus in several countries. The peak time corresponds to the week when most individuals are infected by the pathogen, and is measured in weeks after the beginning of the epidemic. The model predictions were obtained by analyzing 2,000 stochastic realizations of the outbreak, generating the error bars in the figure. After [82].

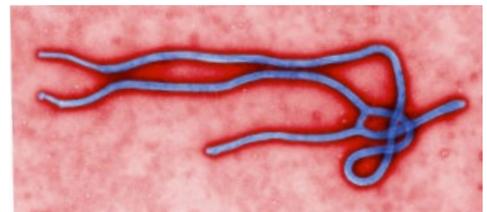


Figure 10.28  
The Deadliest Outbreak

With a fatality rate in the vicinity of 80%, the Ebola virus is one of the deadliest viruses known to humans. Its first known incidence was in 1976 in Zaire, killing 280 of the 312 infected individuals by hemorrhagic fever, a combination of high fever and bleeding disorder. The virus can be transmitted by contact with the blood or the secretion of an infected individual.

# BOX 10.5

## A NIGHT AT THE MOVIES

For a fictionalized but plausible depiction of a major pandemic, watch *Contagion*, the 2011 medical thriller directed by Steven Soderbergh, featuring Marion Cotillard, Bryan Cranston, Matt Damon, Laurence Fishburne, Jude Law, Gwyneth Paltrow, Kate Winslet, and Jennifer Ehle. The movie follows the desperate attempts of public health officials to stop a virus and the ensuing panic from sweeping the globe, hence addressing the impact of both biological and social contagion. The 1995 medical disaster film *Outbreak* directed by Wolfgang Petersen, starring Dustin Hoffman, Rene Russo and Morgan Freeman, focuses on a deadly Ebola-like virus that starts from a small village in Zaire and reaches the United States. Both movies illustrate the difficult choices civilian and military agencies must take to contain the spread of a deadly pathogen.



**Figure 10.30**  
**Outbreak: Fiction and Truth**

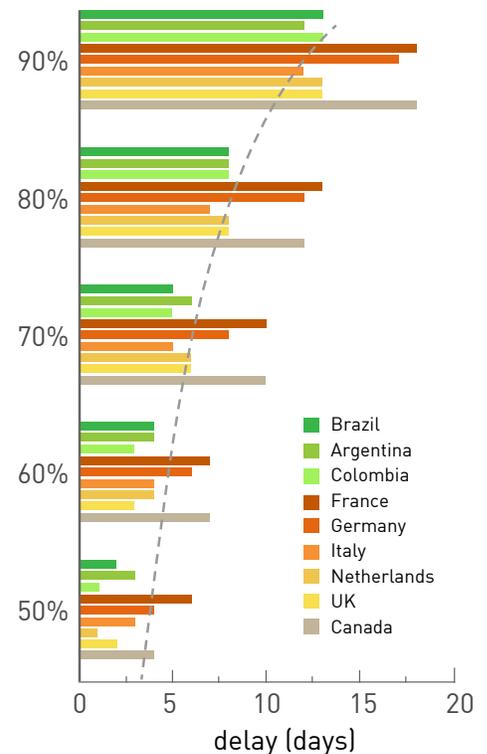
The theatrical release posters of two pandemic-related movies, *Contagion* and *Outbreak*.

travel restrictions have beneficial effects on the pandemic. For this we must realize that awareness of a viral outbreak results in self-imposed travel reductions. For example, there was a 40% decline in travel to and from Mexico in May 2009, during the H1N1 outbreak, as individuals canceled non-necessary business and leisure activities in the infected region. The modeling indicates [80,82] that this 40% reduction delayed the arrival of the first infection with less than 3 days in various countries around the world. Furthermore, even if travel dropped 90%, the peak time is delayed with less than 20 days (Figure 10.29).

Most important, travel restrictions do not decrease the number of infected individuals. They only delay the outbreak, offering local authorities more time to prepare for the pandemic. Hence travel restrictions are effective only if the delay caused by them increases local vaccination levels or helps the deployment of cures.

- **Antiviral Treatment**

During the 2009 H1N1 pandemic Canada, Germany, Hong Kong, Japan, the UK, and the USA distributed antiviral drugs to mitigate the impact of the disease [85]. This prompted modelers to ask what would have been the impact if all countries that had drug stockpiles would have distributed it to their population [86]. The simulations indicate that peak times would have been delayed with about 3 to 4 weeks, offering time to immunize a larger fraction of the population before the pandemic reached its peak.



**Figure 10.29**  
**The Impact of Travel Reduction**

The impact of travel reduction on the arrival time of the H1N1 virus from Mexico to various countries, compared with the reference scenario of no travel reduction. The percentages on the vertical axis show the degree of travel reduction implemented around the world. The largest delay is less than 20 days, observed for a 90% travel restriction. After [77].

## EFFECTIVE DISTANCE

Before cars and airplanes pathogens traveled on foot or at most with the speed of a horse. Hence a pandemic like the Black Death in Europe moved slowly from village to village (Figure 10.8), following a diffusive process described by simple reaction-diffusion models [87,88]. As the next infection always emerged in the geographic proximity of the previous infections, there was a strong correlation between the time of the outbreak and the physical distance from the origin of the outbreak.

Today, with airline travel, physical distance has lost its relevance for epidemic phenomena. A pathogen that emerges in Manhattan can just as easily travel to London than to Garrison, NY, a village an our drive from Manhattan. This prompts us to ask: Is there a better space to view the spread of an epidemic than the physical space? Such space does exist if we replace the conventional geographic distance with an effective distance derived from the mobility network [89]. The nodes of the mobility network are cities and the links represent the amount of travel between them. Each link is directed and weighted, characterized by a flux-fraction  $0 \leq p_{ij} \leq 1$ , that represents the fraction of travelers that leave node  $i$  and arrive at node  $j$ . The values of  $p_{ij}$  can be extracted from airline schedules, having  $p_{ij} > 0$  only if there is direct travel from  $i$  to  $j$ .

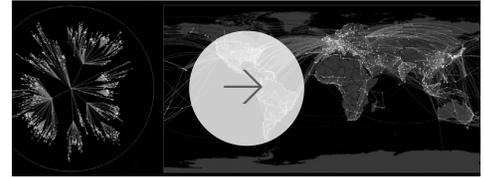
Given the multiple routes a person can take between any two cities, a pathogen can follow multiple paths on the mobility network. Yet, its spread is dominated by the most probable trajectories predicted by the mobility matrix  $p_{ij}$ . This allows us to define the *effective distance*  $d_{ij}$  between two connected locations  $i$  and  $j$ , as

$$d_{ij} = (1 - \ln p_{ij}) \geq 0. \quad (10.31)$$

If  $p_{ij}$  is small, implying that only a small fraction of individuals that leave from  $i$  travel to  $j$ , then the effective distance between  $i$  and  $j$  is large. Note that  $d_{ij} \neq d_{ji}$ : For a small village  $i$  located near a metropolis  $j$  we expect  $d_{ij}$  to be small, as most travelers from  $i$  go to  $j$ . Yet,  $d_{ji}$  is large as only a small fraction of travelers leaving the metropolis head to the small village. The logarithm in (10.31) accounts for the fact that effective distances are additive, whereas probabilities along multi-step paths are multiplicative.

As Figure 10.31 indicates (see also ONLINE RESOURCE 10.6), if we use (10.31) to represent the distance of each city from the source of an epidemic, the pathogen follows circular wave fronts. This is in contrast with the complex spreading pattern we observe if we view the pandemic in the geographical space. Furthermore, while the arrival time of H1N1 appears to be random if plotted in function of the physical distance, it correlates strongly with the effective distance (Figure 10.32). We can therefore use the effective distance to determine the speed of a pathogen (ONLINE RESOURCE 10.6).

A surprising but welcome aspect of epidemic forecast is that the predictions of different models are rather similar, despite the fact that they use different mobility data (airline schedules [25,26] or dollar bill movement



## Online Resource 10.6 The Speed of a Pandemic

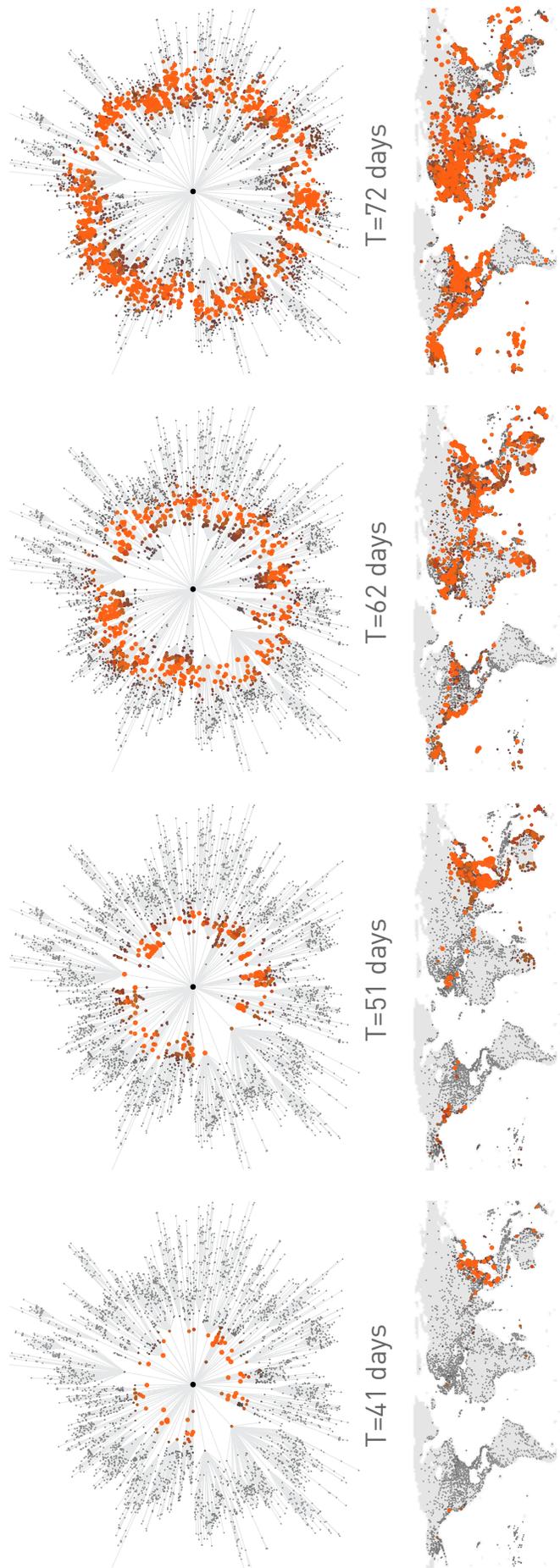
The spread of a pathogen, as predicted by GLEAM, from three initial outbreak locations. While the geographic spreading pattern is difficult to interpret, in the effective distance representation the pandemic follows a regular radial pattern (Figure 10.31).

The observed spreading patterns prompt us to ask: What is the speed of a typical pathogen as it spreads around the globe? The speed depends on three key parameters:

1. The basic reproduction number  $R_0$ , which is in the vicinity of 2 for influenza type viruses (Table 10.2).
2. The recovery rate, which is approximately 3 days for influenza.
3. The mobility rate, which represents the total fraction of the population that travels during a day. This parameter is in the range of 0.01-0.001.

Running GLEAM (Figure 10.26) with these parameters we can compute the correlation between the arrival time and the geographic distance to the source of the epidemic, obtaining a speed of about 250-300 km/day. Therefore an influenza virus moves through a continent with the speed of a sports car or of a smaller airplane [89].





**Figure 10.31**  
**Effective Distance**

The spread of a pandemic with an initial outbreak in Hong Kong. Regions with a large number of infections are shown as red nodes. Each panel compares the state of the system in the conventional geographic representation (bottom) with the effective distance representation (top). The complex spatial pattern observed in the geographic representation becomes a circular wave that moves outwards at constant speed in the effective distance representation (see also the [Online Resource 10.6](#)). After [89].

[24]) and different assumptions about the epidemic parameters (recovery rate, transmission rate, etc). The effective distance helps us understand why the various model predictions converge. Indeed, we can write the arrival time of a pathogen to location  $a$  as [89]

$$T_a = \frac{d_{\text{eff}}(P)}{V_{\text{eff}}(\beta, R_0, \gamma, \varepsilon)} \quad (10.32)$$

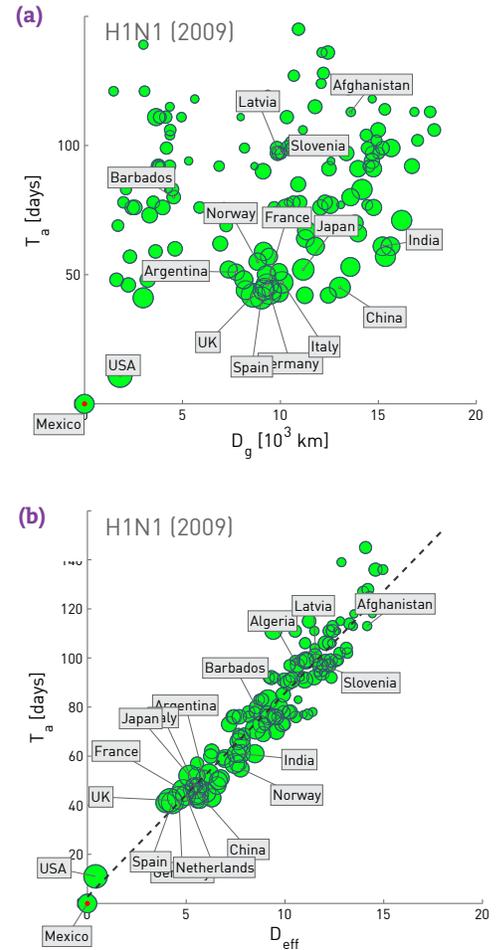
Therefore the arrival time is the ratio of the effective distance  $d_{\text{eff}}$  and an effective speed  $V_{\text{eff}}$ . The effective speed is determined only by the epidemiological parameters of the pathogen, whereas the effective distance  $d_{\text{eff}}$  depends only on the topology of the mobility network encoded by  $p_{ij}$ . When confronted with a new outbreak, the pathogen-specific epidemiological parameters are unknown in the beginning. However, (10.32) predicts that the *relative arrival times are independent of the epidemiological parameters*. For example, for an outbreak that starts at node  $i$ , the ratio of the arrival times to nodes  $j$  and  $l$  is

$$\frac{T_a(j/i)}{T_a(l/i)} = \frac{d_{\text{eff}}(j/i)}{d_{\text{eff}}(l/i)},$$

i.e. the ratio depends only on the effective distances. Therefore, the relative arrival times of the disease depend only on the topology of the mobility network. As the mobility patterns around the world are unique and model-independent, the predictions of different models converge, independent of the choice of the epidemiological parameters.

In summary, joint advances in data collection and network epidemics have offered the capability to predict the real-time spread of a pathogen. The developed models can help design response and mitigation scenarios, train health and emergency personnel, can be used to explore the impact of various interventions, from quarantine to travel restrictions, and to optimize the deployment of treatments and vaccines.

Interestingly, the recent success of epidemic forecast is not due to the improved understanding of the underlying biology of infectious pathogens. It can be attributed instead to the lucky situation that when it comes to the spreading of a pathogen, the epidemic parameters are of secondary importance. The most important factor is the structure of the mobility network. That, however, can be accurately estimated from travel schedules, allowing us to turn human mobility patterns into accurate predictions about the course of a pandemic.



**Figure 10.32**  
**Effective Distance and Arrival Time**

**(a) Geographic Distance**

Arrival times vs. geographic distance from its source (Mexico) for the 2009 H1N1 pandemic. Each circle represents one of the 140 affected countries and the symbol size indicates the total traffic in each country. Arrival times are the date of the first confirmed case in a given country after the beginning of the outbreak on March 17, 2009. In this representation the arrival time and the geographic distance are largely independent of each other ( $R_o=0.0394$ ).

**(b) Effective Distance**

Epidemic arrival time  $T_a$  vs. effective distance  $D_{\text{eff}}$  for H1N1, demonstrating the strong correlations between the effective distance (10.31) and the arrival time. After [89].

# BOX 10.6

## IDENTIFYING THE SOURCE OF A PANDEMIC

Identifying the source of an epidemic is an important component of epidemic control. The source could be the first individual in a contact network, or the city where the pathogen first emerged in the mobility network. The mathematical formulation of the problem [91] inspired a burst of research on the subject [92-99].

The difficulty in finding the source is rooted in the stochastic nature of the infection process: different initial conditions can lead to similar infection patterns at the observation time. The approach we take depends on the information we have about the epidemic:

- In the simplest case at a given moment  $t$  we know the nodes that have been infected and the network on which the pathogen spreads. The task is to find the source  $i$  [91] (Figure 10.33).
- If we also have the time of infection for each node, we can reconstruct the dynamics of the epidemic, significantly enhancing our ability to detect the source.
- The best strategy is to monitor the hubs, as they have the earliest and the most accurate information about a breakout. For example, for a pathogen spreading on a scale-free network, monitoring the state of 18% of the highest degree nodes can offer a 90% success rate in detecting the source. In contrast, to achieve the same level of accuracy we need to monitor 41% of the nodes if we select randomly the nodes we monitor [93].
- In the effective distance representation (Figure 10.31) the infection follows a circular pattern only if we use the right outbreak location. Otherwise the observed pattern is asymmetric. Therefore, we can detect the source by finding the location (node) from which the outbreak pattern shows the highest radial symmetry [89].



**Figure 10.33**  
**Epidemic Sources**

Finding the source of an epidemic is like finding the source of a water ripple. As pathogens do not spread in a uniform medium, the challenge is to identify the appropriate “ripples” in the mobility network.

# SUMMARY

Most networks facilitate transfer along their links: transfer of trust, knowledge, habits or information (social networks), electricity (power grid), money (financial networks), goods (trade networks). To understand these phenomena, we must understand how the network topology affects these dynamical processes. In this chapter we focused on the spread of pathogens along the links of the network, the area where our understanding of the interplay between dynamical phenomena and network topology is the most advanced. We showed that the network topology has a drastic impact on the dynamics of the spreading process, offering distinct predictions for spreading on random and on scale-free networks. This finding laid the ground for addressing a wider class of problems: the need to systematically understand the impact networks have on various dynamical processes [100], an increasingly active chapter of network science [101,102].

Modeling the spread of pathogens also represents an important practical application of network science. The advances in this area were rather spectacular, giving birth to accurate epidemic forecasts, something that was only a dream a decade earlier. Two advances made this possible. The first is the emergence of a robust theoretical framework to describe network-based epidemics. The second is access to accurate real time data on human travel and demographics, allowing us to reconstruct the mobility network that is responsible for the global spread of a pathogen. As we have seen in SECTION 10.7, the biological parameters and the network contributions to the accuracy of the observed predictive power are decoupled. Consequently, an accurate forecast requires primarily an accurate knowledge of the mobility network.

The analytical framework of network epidemics has offered a number of unexpected results, the most important being the vanishing characteristic spreading time and epidemic threshold in heterogeneous networks. As most contact networks encountered in epidemic processes have a broad degree distribution, these results are of immediate and of lasting theoretical and practical interest.

## BOX 10.7

### AT A GLANCE: NETWORK EPIDEMICS

**Infection Rate:**  $\beta$

**Recovery Rate:**  $\mu$

**Spreading Rate:**  $\lambda = \frac{\beta}{\mu}$

**Reproductive Number:**  $R_0 = \frac{\beta \langle k \rangle}{\mu}$

#### SI Model:

$$i(t) = \frac{i_0 e^{\beta \langle k \rangle t}}{1 - i_0 + i_0 e^{\beta \langle k \rangle t}}$$

#### SIS Model:

$$i(t) = \left(1 - \frac{\mu}{\beta \langle k \rangle}\right) \frac{C e^{(\beta \langle k \rangle - \mu)t}}{1 + C e^{(\beta \langle k \rangle - \mu)t}}$$

#### Characteristic time:

**SI:**  $\tau = \frac{\langle k \rangle}{\beta \langle k^2 \rangle - \langle k \rangle}$

**SIS:**  $\tau = \frac{\langle k \rangle}{\beta \langle k^2 \rangle - \mu \langle k \rangle}$

**SIR:**  $\tau = \frac{\langle k \rangle}{\beta \langle k^2 \rangle - (\mu + \beta) \langle k \rangle}$

#### Epidemic Threshold:

**SIS:**  $\lambda_c = \frac{\langle k \rangle}{\langle k^2 \rangle}$

**SIR:**  $\lambda_c = \frac{1}{\frac{\langle k^2 \rangle}{\langle k \rangle} - 1}$

#### Immunization Threshold (SIS):

$$g_c = 1 - \frac{\mu \langle k \rangle}{\beta \langle k^2 \rangle}$$

## BOX 10.8

### HISTORICAL NOTE: NETWORK EPIDEMICS

Epidemic phenomena became a central topic in network science after Romualdo Pastor-Satorras and Alessandro Vespignani introduced the continuum theory that can account for the properties of the underlying contact network. They also discovered the dependence of the epidemic threshold and characteristic time on the second moment of the degree distribution, a central result of network epidemics. Subsequently Vespignani and his research group have developed GLEAM, a computational framework that offers real-time predictions for the spread of a pathogen.



Figure 10.34

### Romualdo Pastor-Satorras and Alessandro Vespignani

Physicists by training, Pastor-Satorras was a postdoctoral associate with Vespignani at ICTP in Trieste when they discovered the impact of the scale-free property on the epidemic threshold. Subsequently both researchers had major contributions to network science, from the discovery of degree correlations (CHAPTER 7) to our understanding of weighted networks.

Equally important are the insights network epidemiology offers for immunization strategies. As we showed in SECTION 10.6, while random immunization can successfully eradicate a virus that spreads on a random network, this strategy is suboptimal in a scale-free network. As most contact networks are heterogenous, this is a rather depressing conclusion. Yet, we showed that selective immunization strategies can restore the epidemic threshold and suppress the prevalence of a pathogen. Selective immunization succeeds by systematically altering the topology of the network on which a pathogen spreads.

# HOMework

## 10.1. Epidemics on Networks

Calculate the characteristic time  $\tau$  and the epidemic threshold  $\lambda_c$  of the SI, SIS and SIR models for networks with

- (a) Exponential degree distribution.
- (b) Stretched exponential degree distribution.
- (c) Delta distribution (all nodes have the same degree).

Assume that the networks are uncorrelated and infinite. Refer to [Table 4.2](#) for the functional form of the distribution and the corresponding first and second moments.

## 10.2. Random Obesity in Social Networks

Consider a social network with degree distribution  $p_k$ , where 50% of the nodes are obese. Make the assumption that obese nodes are distributed randomly within the network.

- (a) If the network has degree correlation, encoded in the joint probability  $e_{kk'}$ , what is the probability  $P(\emptyset o)$  that a non-obese ( $\emptyset$ ) individual is friend with an obese individual ( $o$ )? And what is the probability  $P(oo)$  that two obese individuals are friends?
- (b) Assume that the network is uncorrelated. How many second neighbors of a degree- $k$  node are obese?

Calculate the same quantities of (a)-(b) if the percentage of obese increases to 70%.

## 10.3. Immunization

Choose four networks from [Table 4.1](#) (assume that directed networks behave like undirected and uncorrelated networks with  $p_k = p_{k_{in}}$ ) and consider an epidemic process spreading on them. Remember: not only pathogens, but also ideas or opinions can spread on a network! Determine for each network the critical fraction  $g_c$  necessary to stop the ep-

idemic if we randomly immunise a  $g$ -fraction of the nodes. How would the epidemic threshold  $\lambda_c$  change if all nodes with degree higher than 1,000 are immunized?

#### 10.4. Epidemic on Bipartite Networks

Consider a bipartite network, with two types of nodes, which we indicate as male (M) and female (F). On this network a pathogen can be transmitted only from the node of one set to the node of the other set. Assume that the rate of transmission from an M node to an F node,  $\beta_{M \rightarrow F}$ , is different from the rate of transmission from an F node to an M node,  $\beta_{F \rightarrow M}$ . Write the equations of the corresponding SI model, assuming the degree block approximation and that the network is uncorrelated.

# ADVANCED TOPICS 10.A

## MICROSCOPIC MODELS OF EPIDEMIC PROCESSES

In SECTIONS 10.2 and 10.3 we relied on the continuum approach to describe epidemic phenomena. In this section we show that the key results can be derived using microscopic models and probability-based reasoning. These arguments help us understand the origin of the continuum approach and improve our understanding of epidemic phenomena.

### DERIVING THE EPIDEMIC EQUATION

We start by deriving the continuum model (10.3) from the microscopic processes that describe the interactions between two individuals [101]. Consider a susceptible individual in contact with an infected individual, so that the susceptible individual becomes infected with probability  $\beta dt$  during the time interval  $dt$ . The probability that the susceptible individual is *not* infected in the  $dt$  interval is  $(1 - \beta dt)$ . If the susceptible individual  $i$  has degree  $k_i$ , each of its  $k_i$  links could in principle infect it. Therefore the probability that it avoids infection is  $(1 - \beta dt)^{k_i}$ . Finally the total probability that node  $i$  becomes infected in time  $dt$  is  $1 - (1 - \beta dt)^{k_i}$ , or one minus the total probability that it is not infected. Assuming  $\beta dt \ll 1$ , at the leading order the probability that a susceptible individual becomes infected is

$$1 - (1 - \beta dt)^{k_i} \approx \beta k_i dt. \quad (10.33)$$

In a random network all nodes have approximately  $\langle k \rangle$  neighbors. Replacing  $k_i$  with  $\langle k \rangle$  in (10.33) we obtain the first term of the continuum equation (10.3). If we do not replace  $k_i$  with  $\langle k \rangle$ , we obtain to the first term of (10.13), capturing the spread of a pathogen in a heterogenous network.

### EPIDEMIC THRESHOLD AND NETWORK TOPOLOGY

A key result of SECTION 10.3 connects the network topology to the epidemic threshold  $\lambda_c$ , a result derived using the continuum theory. We can arrive at the same result using a mechanistic argument that illustrates the connection between the epidemic threshold and the network topology.

Consider a pathogen that is transmitted with probability  $\beta$  in a unit time. Therefore in a unit time an infected node with degree  $k$  will infect

$\beta k$  neighbors. If each infected node recovers at rate  $\mu$ , then the characteristic time that a node stays infected is  $1/\mu$ . The pathogen can persist in the population only if during this  $1/\mu$  time interval the infected node infects at least one other node. Otherwise, the pathogen gradually dies out.

In other words, if  $\beta k/\mu < 1$ , then our degree- $k$  node recovers before it could infect other nodes. If we consider a random network, where most nodes have comparable degrees,  $k \sim \langle k \rangle$ , the condition  $\beta k/\mu = 1$  allows us to calculate the epidemic threshold. Using  $\lambda = \beta/\mu$  we obtain  $\lambda_c = 1/\langle k \rangle$ , which is the high- $k$  limit of the result (10.25) derived for random networks. It tells us that the ability of a pathogen to spread is determined by the interplay between the epidemiological characteristics of the pathogen ( $\beta$  and  $\mu$ ) and the network topology ( $\langle k \rangle$ ).

In a scale-free network nodes have widely different degrees. Therefore while the network's average degree may satisfy  $\beta \langle k \rangle / \mu < 1$ , suggesting that the virus will die out, for all nodes with  $k > \langle k \rangle$  we have  $\beta k / \mu > 1$ . If such a high degree node is infected, even if the spreading rate  $\lambda$  is under the threshold  $1/\langle k \rangle$ , the disease can spread, persisting in the hubs. This is the reason why the epidemic threshold vanishes in networks with high  $\langle k^2 \rangle$ .

# ADVANCED TOPICS 10.B

## ANALYTICAL SOLUTION OF THE SI, SIS AND SIR MODELS

In this section we solve the SI, SIS and SIR models on a network, deriving the results summarized in [Table 10.3](#), namely the characteristic spreading time  $\tau$  and the epidemic threshold  $\lambda_c$  for each model.

### THE DENSITY FUNCTION

The density function  $\Theta_k$  provides the fraction of infected nodes in the neighborhood of a susceptible node with degree  $k$ . As discussed in [SECTION 10.3](#), to calculate  $i_{k'}$ , we must first determine  $\Theta_k$ . If a network lacks degree correlations, the probability that a link points from a node with degree  $k$  to a node with degree  $k'$  is independent of  $k$ . Hence the probability that a randomly chosen link points to a node with degree  $k'$  is the excess degree [\(7.3\)](#),

$$\frac{k' p_{k'}}{\sum_k k p_k} = \frac{k' p_{k'}}{\langle k \rangle}.$$

At least one link of each infected node is connected to another infected node, the one that transmitted the infection. Therefore the number of links available for future transmission is  $(k'-1)$ , allowing us to write

$$\Theta_k = \frac{\sum_{k'} (k'-1) p_{k'} i_{k'}}{\langle k \rangle} = \Theta. \quad (10.34)$$

In other words, in the absence of degree correlations  $\Theta_k$  is independent of  $k$ . Differentiating [\(10.34\)](#) we obtain

$$\frac{d\Theta}{dt} = \sum_k \frac{(k-1) p_k}{\langle k \rangle} \frac{di_k}{dt}. \quad (10.35)$$

To make further progress, we need to consider the specific model the pathogen follows.

## SI MODEL

Using (10.13) and (10.35) we obtain

$$\frac{d\Theta}{dt} = \beta \sum_k \frac{(k^2 - k)p_k}{\langle k \rangle} [1 - i_k] \Theta. \quad (10.36)$$

To predict the early behavior of the epidemics, we consider the fact that for small  $t$  the fraction of infected individuals is much smaller than one. Therefore we can neglect the second order terms in (10.36), obtaining

$$\frac{d\Theta}{dt} = \beta \left( \frac{\langle k^2 \rangle}{\langle k \rangle} - 1 \right) \Theta. \quad (10.37)$$

This has the solution

$$\Theta(t) = C e^{t/\tau} \quad (10.38)$$

where

$$\tau = \frac{\langle k \rangle}{\beta(\langle k^2 \rangle - \langle k \rangle)}. \quad (10.39)$$

Using the initial condition

$$\Theta(t=0) = C = i_0 \frac{\langle k \rangle - 1}{\langle k \rangle},$$

which means that initially an  $i_0$  fraction of nodes are infected uniformly (hence  $i_k(t=0) = i_0$  for all  $k$ ), we obtain the time dependent  $\Theta$  as

$$\Theta(t) = i_0 \frac{\langle k \rangle - 1}{\langle k \rangle} e^{t/\tau}. \quad (10.40)$$

We insert this into (10.13) to arrive at (10.15).

## SIR MODEL

In the SIR model the density of infected nodes follows

$$\frac{di_k}{dt} = \beta(1 - i_k - r_k)k\Theta - \mu i_k, \quad (10.41)$$

where  $r_k$  is the fraction of recovered nodes with degree  $k$ . Keeping only the first order terms (which means that we ignore  $i_k$  and  $r_k$  in the parenthesis above, as for small  $t$  they are much smaller than one), we obtain

$$\frac{di_k}{dt} = \beta k \Theta - \mu i_k. \quad (10.42)$$

Multiplying this equation with  $(k-1)p_k/\langle k \rangle$  and summing over  $k$  we have

$$\frac{d\Theta}{dt} = \left( \beta \frac{\langle k^2 \rangle - \langle k \rangle}{\langle k \rangle} - \mu \right) \Theta. \quad (10.43)$$

The solution of (10.43) is

$$\Theta(t) = Ce^{t/\tau}, \tag{10.44}$$

where the characteristic time for the SIR model is

$$\tau = \frac{\langle k \rangle}{\beta \langle k^2 \rangle - \langle k \rangle (\beta + \mu)}. \tag{10.45}$$

A global outbreak is possible only if  $\tau > 0$ , i.e. when the number of infected nodes grows exponentially with time. This yields the condition for a global outbreak as

$$\lambda = \frac{\beta}{\mu} > \frac{\langle k \rangle}{\langle k^2 \rangle - \langle k \rangle}, \tag{10.46}$$

allowing us to write the epidemic threshold for the SIR model as (Table 10.3)

$$\lambda_c = \frac{1}{\frac{\langle k^2 \rangle}{\langle k \rangle} - 1}. \tag{10.47}$$

### SIS MODEL

In the SIS model the density of infected nodes is given by (10.18),

$$\frac{di_k}{dt} = \beta(1 - i_k)k\Theta - \mu i_k. \tag{10.48}$$

There is a small but important difference in the density function of the SIS model. For the SI and the SIR models, if a node is infected, then at least one of its neighbors must also be infected or recovered, hence at most  $(k-1)$  of its neighbors are susceptible, the origin of the  $(-1)$  term in the parenthesis of (10.34). However, in the SIS model the previously infected neighbor can become susceptible again, therefore all  $k$  links of a node can be available to spread the disease. Hence we modify the definition (10.34) to obtain

$$\Theta_k = \frac{\sum_{k'} k' p_{k'} i_{k'}}{\langle k \rangle} = \Theta. \tag{10.49}$$

Again keeping only the first order terms we obtain

$$\frac{di_k}{dt} = \beta k \Theta - \mu i_k. \tag{10.50}$$

Multiplying the equation with  $(k-1)p_k/\langle k \rangle$  and summing over  $k$  we have

$$\frac{d\Theta}{dt} = \left( \beta \frac{\langle k^2 \rangle}{\langle k \rangle} - \mu \right) \Theta. \tag{10.51}$$

This again has the solution

$$\Theta(t) = Ce^{t/\tau}, \tag{10.52}$$

where the characteristic time of the SIS model is

$$\tau = \frac{\langle k \rangle}{\beta \langle k^2 \rangle - \langle k \rangle \mu}. \quad (10.53)$$

A global outbreak is possible if  $\tau > 0$ , which yields the condition for a global outbreak as

$$\lambda = \frac{\beta}{\mu} > \frac{\langle k \rangle}{\langle k^2 \rangle}, \quad (10.54)$$

and the epidemic threshold for the SIS model as (Table 10.3)

$$\lambda_c = \frac{\langle k \rangle}{\langle k^2 \rangle}. \quad (10.55)$$

# ADVANCED TOPICS 10.C

## TARGETED IMMUNIZATION

In this section we derive the epidemic threshold for the SIS and SIR models on scale-free networks under hub immunization. We start with an uncorrelated network with power law degree distribution  $p_k = c \cdot k^{-\gamma}$  where  $c = (\gamma - 1) / k_{\min}^{\gamma+1}$  and  $k \geq k_{\min}$ . In SECTION 10.16 we obtained for the critical spreading rate,

$$\lambda_c = \frac{\langle k \rangle}{\langle k^2 \rangle} = \frac{1}{\kappa} \quad (\text{SIS model})$$

and

$$\lambda_c = \frac{1}{\frac{\langle k^2 \rangle}{\langle k \rangle} - 1} = \frac{1}{\kappa - 1} \quad (\text{SIR model}).$$

Under hub immunization we immunize all nodes whose degree is larger than  $k_0$ . From the perspective of the epidemic this is equivalent with removing the high degree nodes from the network. Therefore to calculate the new critical spreading rate, we need to determine the average degree  $\langle k' \rangle$  and the second moment  $\langle k'^2 \rangle$  after the hubs have been removed. This problem was addressed in the ADVANCED TOPICS 8.F, where we studied the robustness of a network under attack. We have seen that hub removal has two effects:

- 1) The maximum degree of the network changes to  $k_0$ .
- 2) The links connected to the removed hubs are also removed, as if we randomly remove an

$$\tilde{f} = \left( \frac{k_0}{k_{\min}} \right)^{-\gamma+2} \quad (10.56)$$

fraction of links.

The degree distribution of the resulting network is

$$p'_k = \sum_{k=k_{\min}}^{k_0} \binom{k}{k'} f^{-k-k'} (1-\tilde{f})^{k'} p_k.$$

According to (8.39) and (8.40) this yields

$$\begin{aligned}\langle k' \rangle &= (1 - \tilde{f})\langle k \rangle, \\ \langle k'^2 \rangle &= (1 - \tilde{f})^2 \langle k^2 \rangle + \tilde{f}(1 - \tilde{f})\langle k \rangle,\end{aligned}$$

where  $\langle k \rangle$  is the average and  $\langle k^2 \rangle$  is the second moment of the degree distribution before the link removal, but with maximum degree  $k_0$ . For the SIS model this means

$$\lambda'_c = \frac{(1 - \tilde{f})\langle k \rangle}{(1 - \tilde{f})^2 \langle k^2 \rangle + \tilde{f}(1 - \tilde{f})\langle k \rangle} = \frac{1}{(1 - \tilde{f})\kappa + \tilde{f}}, \quad (10.57)$$

where, according to equation (8.47), for  $2 > \gamma > 3$

$$\kappa = \frac{\gamma - 2}{3 - \gamma} k_0^{3-\gamma} k_{\min}^{\gamma-2}. \quad (10.58)$$

Combining (10.56), (10.57) and (10.58) we obtain

$$\lambda'_c = \left[ \frac{\gamma - 2}{3 - \gamma} k_0^{3-\gamma} k_{\min}^{\gamma-2} - \frac{\gamma - 2}{3 - \gamma} k_0^{5-2\gamma} k_{\min}^{2\gamma-4} + k_0^{2-\gamma} k_{\min}^{\gamma-2} \right]^{-1}. \quad (10.59)$$

For the SIR model a similar calculation yields

$$\lambda'_c = \left[ \frac{\gamma - 2}{3 - \gamma} k_0^{3-\gamma} k_{\min}^{\gamma-2} - \frac{\gamma - 2}{3 - \gamma} k_0^{5-2\gamma} k_{\min}^{2\gamma-4} + k_0^{2-\gamma} k_{\min}^{\gamma-2} - 1 \right]^{-1}. \quad (10.60)$$

For both the SIR and SIS models if  $k_0 \gg k_{\min}$  we have

$$\lambda'_c \approx \frac{3 - \gamma}{\gamma - 2} k_0^{\gamma-3} k_{\min}^{2-\gamma}. \quad (10.61)$$



ing its neighbor is  $p_b$ .

Since the network is randomly connected, as long as the epidemic has not spread yet, the average number of neighbors infected by the selected node is

$$\langle R_i \rangle = p_b \sum \frac{p_k k(k-1)}{\langle k \rangle}.$$

An endemic state can be reached only if  $\langle R_i \rangle > 1$ , obtaining the condition for the epidemic as [107,108]

$$\left( \frac{\langle k^2 \rangle}{\langle k \rangle} - 1 \right) > \frac{1}{p_b}. \quad (10.62)$$

Equation (10.62) agrees with the result (10.46) derived earlier from the dynamical models: Scale-free networks with  $\gamma \leq 3$  have a divergent second moment, hence such networks undergo a percolation transition even at  $p_b \rightarrow 0$ . That is, a virus can spread on this network regardless of how small is the infection probability  $\beta$  or how small is the recovery time  $\tau$ .

# BIBLIOGRAPHY

[1] D. Normile. The Metropole, Superspreaders and Other Mysteries. *Science*, 339:1272-1273, 2013.

[2] J.O. Lloyd-Smith, S.J. Schreiber, P.E. Kopp, and W.M. Getz. Super-spreading and the effect of individual variation on disease emergence. *Nature*, 438:355-359, 2005.

[3] M. Hypponen. Malware Goes Mobile. *Scientific American*, 295:70, 2006.

[4] P. Wang, M. Gonzalez, C. A. Hidalgo, and A.-L. Barabási. Understanding the spreading patterns of mobile phone viruses. *Science*, 324:1071-1076, 2009.

[5] E.M. Rogers. *Diffusion of Innovations*. Free Press, 2003.

[6] T.W. Valente. *Network models of the diffusion of innovations*. Hampton Press, Cresskill, NJ, 1995.

[7] History and Epidemiology of Global Smallpox Eradication From the training course titled "Smallpox: Disease, Prevention, and Intervention". The CDC and the World Health Organization. Slides 16-17.

[8] R.M. Anderson and R.M. May. *Infectious Diseases of Humans: Dynamics and Control*. Oxford University Press, Oxford, 1992.

[9] R. Pastor-Satorras and A. Vespignani. Epidemic spreading in scale-free networks. *Physical Review Letters*, 86:3200–3203, 2001.

[10] R. Pastor-Satorras and A. Vespignani. Epidemic dynamics and endemic states in complex networks. *Physical Review E*, 63:066117, 2001.

[11] Y. Wang, D. Chakrabarti, C. Wang, and C. Faloutsos. Epidemic spreading in real networks: an eigenvalue viewpoint. *Proceedings of 22nd International Symposium on Reliable Distributed Systems*, pg. 25-34, 2003.

[12] R. Durrett. Some features of the spread of epidemics and information on a random graph. *PNAS*, 107:4491-4498, 2010.

[13] S. Chatterjee and R. Durrett. Contact processes on random graphs with power law degree distributions have critical value 0. *Ann. Probab.*, 37: 2332-2356, 2009.

[14] C. Castellano, and R. Pastor-Satorras. Thresholds for epidemic spreading in networks. *Physical Review Letters*, 105:218701, 2010.

[15] B. Lewin. (ed.), *Sex i Sverige. Om sexuallivet i Sverige 1996 [Sex in Sweden. On the Sexual Life in Sweden 1996]*. National Institute of Public Health, Stockholm, 1998.

[16] F. Liljeros, C. R. Edling, L. A. N. Amaral, H. E. Stanley, and Y. Åberg. The web of human sexual contacts. *Nature*, 411:907-8, 2001.

[17] A. Schneeberger, C. H. Mercer, S. A. Gregson, N. M. Ferguson, C. A. Nyamukapa, R. M. Anderson, A. M. Johnson, and G. P. Garnett. Scale-free networks and sexually transmitted diseases: a description of observed patterns of sexual contacts in Britain and Zimbabwe. *Sexually Transmitted Diseases*, 31: 380-387, 2004.

[18] W. Chamberlain. *A View from Above*. Villard Books, New York, 1991.

[19] R. Shilts. *And the Band Played On*. St. Martin's Press, New York, 2000.

[20] P. S. Bearman, J. Moody, and K. Stovel. Chains of affection: the structure of adolescent romantic and sexual networks. *Am J Sociol.*, 110:44-91, 2004.

[21] M. C. González, C. A. Hidalgo, and A.-L. Barabási. Understanding individual human mobility patterns. *Nature*, 453:779-782, 2008.

[22] C. Song, Z. Qu, N. Blumm, and A.-L. Barabási. Limits of Predictability in Human Mobility. *Science*, 327:1018-1021, 2010.

[23] F. Simini, M. González, A. Maritan, and A.-L. Barabási. A universal model for mobility and migration patterns. *Nature*, 484:96-100, 2012.

[24] D. Brockmann, L. Hufnagel, and T. Geisel. The scaling laws of human travel. *Nature*, 439:462-465, 2006.

[25] V. Colizza, A. Barrat, M. Barthelemy, and A. Vespignani. The role of the airline transportation network in the prediction and predictability of global epidemics. *PNAS*, 103:2015, 2006.

[26] L. Hufnagel, D. Brockmann, and T. Geisel. Forecast and control of epidemics in a globalized world. *PNAS*, 101:15124, 2004.

[27] R. Guimerà, S. Mossa, A. Turtschi, and L. A. N. Amaral. The worldwide air transportation network: Anomalous centrality, community struc-

ture, and cities' global roles. *PNAS*, 102:7794, 2005.

[28] C. Cattuto, et al. Dynamics of Person-to-Person Interactions from Distributed RFID Sensor Networks. *PLoS ONE*, 5:e11596, 2010.

[29] L. Isella, C. Cattuto, W. Van den Broeck, J. Stehle, A. Barrat, and J.-F. Pinton. What's in a crowd? Analysis of face-to-face behavioral networks. *Journal of Theoretical Biology*, 271:166-180, 2011.

[30] K. Zhao, J. Stehle, G. Bianconi, and A. Barrat. Social network dynamics of face-to-face interactions. *Physical Review E*, 83:056109, 2011.

[31] J. Stehlé, N. Voirin, A. Barrat, C Cattuto, L. Isella, J-F. Pinton, M. Quagiotto, W. Van den Broeck, C. Régis, B. Lina, and P. Vanhems. High-resolution measurements of face-to-face contact patterns in a primary school. *PLoS ONE*, 6:e23176, 2011.

[32] B.N. Waber, D. Olguin, T. Kim, and A. Pentland. Understanding Organizational Behavior with Wearable Sensing Technology. Academy of Management Annual Meeting. Anaheim, CA. August, 2008.

[33] L. Wu, B.N. Waber, S. Aral, E. Brynjolfsson, and A. Pentland. Mining Face-to-Face Interaction Networks using Sociometric Badges: Predicting Productivity in an IT Configuration Task. In Proceedings of the International Conference on Information Systems. Paris, France. December 14-17 2008.

[34] M. Salathé, M. Kazandjievab, J.W. Leeb, P. Levisb, M.W. Feldmana, and J.H. Jones. A high-resolution human contact network for infectious disease transmission. *PNAS*, 107:22020–22025, 2010.

[35] A.-L. Barabási. The origin of bursts and heavy tails in human dynamics. *Nature*, 435:207-11, 2005.

[36] V. Sekara, and S. Lehmann. Application of network properties and signal strength to identify face-to-face links in an electronic dataset. Proceedings of CoRR, 2014.

[37] S. Eubank, H. Guclu, V.S.A. Kumar, M.V. Marathe, A. Srinivasan, Z. Toroczkai, and N. Wang. Modelling disease outbreaks in realistic urban social networks. *Nature*, 429:180-184, 2004.

[38] H. Ebel, L-I. Mielsch, and S. Bornholdt. Scale-free topology of e-mail networks. *Physical Review E*, 66:035103, 2002.

[39] M. Morris, and M. Kretzschmar. Concurrent partnerships and transmission dynamics in networks. *Social Networks*, 17:299-318, 1995.

[40] N. Masuda and P. Holme. Predicting and controlling infectious diseases epidemics using temporal networks. *F1000 Prime Rep.*, 5:6, 2013.

[41] P. Holme, and J. Saramäki. Temporal networks. *Physics Reports*,

[42] M. Karsai, M. Kivelä, R. K. Pan, K. Kaski, J. Kertész, A.-L. Barabási, and J. Saramäki. Small but slow world: how network topology and burstiness slow down spreading. *Physical Review E*, 83:025102(R), 2011.

[43] A. Vazquez, B. Rácz, A. Lukács, and A.-L. Barabási. Impact of non-Poissonian activity patterns on spreading processes. *Physical Review Letters*, 98:158702, 2007.

[44] A. Vázquez, J.G. Oliveira, Z. Dezsö, K.-I. Goh, I. Kondor, and A.-L. Barabási. Modeling bursts and heavy tails in human dynamics. *Physical Review E*, 73:036127, 2006.

[45] A.V. Goltsev, S.N. Dorogovtsev, and J.F.F. Mendes. Percolation on correlated networks. *Physical Review E*, 78:051105, 2008.

[46] P. Van Mieghem, H. Wang, X. Ge, S. Tang and F. A. Kuipers. Influence of assortativity and degree-preserving rewiring on the spectra of networks. *The European Physical Journal B*, 76:643, 2010.

[47] M. Boguná, R. Pastor-Satorras, and A. Vespignani. Absence of epidemic threshold in scale-free networks with degree correlations. *Physical Review Letters*, 90:028701, 2003.

[48] Y. Moreno, J. B. Gómez, and A.F. Pacheco. Epidemic incidence in correlated complex networks. *Physical Review E*, 68:035103, 2003.

[49] J.-P. Onnela, J. Saramaki, J. Hyvonen, G. Szabó, D. Lazer, K. Kaski, J. Kertész, and A.-L. Barabási. Structure and tie strengths in mobile communication networks. *PNAS*, 104:7332, 2007.

[50] M. S. Granovetter. The strength of weak ties. *American Journal of Sociology*, 78:1360–1379, 1973.

[51] A. Galstyan, and P. Cohen. Cascading dynamics in modular networks. *Physical Review E*, 75:036109, 2007.

[52] J. P. Gleeson. Cascades on correlated and modular random networks. *Physical Review E*, 77:046117, 2008.

[53] P. A. Grabowicz, J. J. Ramasco, E. Moro, J. M. Pujol, and V. M. Eguiluz. Social features of online networks: The strength of intermediary ties in online social media. *PLOS ONE*, 7:e29358, 2012.

[54] L. Weng, F. Menczer and Y.-Y. Ahn. Virality Prediction and Community Structure in Social Networks. *Scientific Reports*, 3:2522, 2013.

[55] S. Aral, and D. Walker. Creating social contagion through viral product design: A randomized trial of peer influence in networks. *Management Science*, 57:1623–1639, 2011.

[56] J. Leskovec, L. Adamic, and B. Huberman. The dynamics of viral

marketing. *ACM Trans. Web*, 1, 2007.

[57] L. Weng, A. Flammini, A. Vespignani, and F. Menczer. Competition among memes in a world with limited attention. *Scientific Reports*, 2:335, 2012.

[58] J. Berger, and K. L. Milkman. What makes online content viral? *Journal of Marketing Research*, 49:192–205, 2009.

[59] S. Jamali, and H. Rangwala. Digging digg: Comment mining, popularity prediction and social network analysis. *Proc. Intl. Conf. on Web Information Systems and Mining (WISM)*, 32–38, 2009.

[60] G. Szabó and, B. A. Huberman. Predicting the popularity of online content. *Communications of the ACM*, 53:80–88, 2010.

[61] B. Suh, L. Hong, P. Pirolli, and E. H. Chi. Want to be retweeted? Large scale analytics on factors impacting retweet in twitter network. *Proc. IEEE Intl. Conf. on Social Computing*, 177–184, 2010.

[62] D. Centola. The spread of behavior in an online social network experiment. *Science*, 329:1194–1197, 2010.

[63] L. Backstrom, D. Huttenlocher, J. Kleinberg, and X. Lan. Group formation in large social networks: membership, growth, and evolution. *Proc. ACM SIGKDD Intl. Conf. on Knowledge discovery and data mining*, 44–54, 2006.

[64] M. Granovetter. Threshold Models of Collective Behavior. *American Journal of Sociology*, 83:1420–1443, 1978.

[65] N.A. Christakis, and J.H. Fowler. The Spread of Obesity in a Large Social Network Over 32 Years. *New England Journal of Medicine*, 35:370–379, 2007.

[66] N. A. Christakis and J. H. Fowler. The collective dynamics of smoking in a large social network. *New England Journal of Medicine*, 358:2249–2258, 2008.

[67] R. Pastor-Satorras, and A. Vespignani. *Evolution and structure of the Internet: A statistical physics approach*. Cambridge University Press, Cambridge, 2007.

[68] Z. Dezső and A-L. Barabási. Halting viruses in scale-free networks. *Physical Review E*, 65:055103, 2002.

[69] R. Pastor-Satorras and A. Vespignani. Immunization of complex networks. *Physical Review E*, 65:036104, 2002.

[70] R. Cohen, S. Havlin, and D. ben-Avraham. Efficient Immunization Strategies for Computer Networks and Populations. *Physical Review Letters*, 91:247901, 2003.

[71] F. Fenner et al. *Smallpox and its Eradication*. WHO, Geneva, 1988.  
<http://www.who.int/features/2010/smallpox/en/>

[72] L. A. Rvachev, and I. M. Longini Jr. A mathematical model for the global spread of influenza. *Mathematical Biosciences*, 75:3-22, 1985.

[73] A. Flahault, E. Vergu, L. Coudeville, and R. Grais. Strategies for containing a global influenza pandemic. *Vaccine*, 24:6751-6755, 2006.

[74] I. M. Longini Jr, M. E. Halloran, A. Nizam, and Y. Yang. Containing pandemic influenza with antiviral agents. *Am. J. Epidemiol.*, 159:623-633, 2004.

[75] I.M. Longini Jr, A. Nizam, S. Xu, K. Ungchusak, W. Hanshaworakul, D. Cummings, and M. Halloran. Containing pandemic influenza at the source. *Science*, 309:1083-1087, 2005.

[76] V. Colizza, A. Barrat, M. Barthélemy, A.-J. Valleron, and A. Vespignani. Modeling the world-wide spread of pandemic influenza: baseline case and containment interventions. *PLoS Med*, 4:e13, 2007.

[77] T. D. Hollingsworth, N.M. Ferguson, and R.M. Anderson. Will travel restrictions control the International spread of pandemic influenza? *Nature Med.*, 12:497-499, 2006.

[78] C.T. Bauch, J.O. Lloyd-Smith, M.P. Coffee, and A.P. Galvani. Dynamically modeling SARS and other newly emerging respiratory illnesses: past, present, and future. *Epidemiology*, 16:791-801, 2005.

[79] I. M. Hall, R. Gani, H.E. Hughes, and S. Leach. Real-time epidemic forecasting for pandemic influenza. *Epidemiol Infect.*, 135:372-385, 2007.

[80] M. Tizzoni, P. Bajardi, C. Poletto, J. J. Ramasco, D. Balcan, B. Gonçalves, N. Perra, V. Colizza, and A. Vespignani. Real-time numerical forecast of global epidemic spreading: case study of 2009 A/H1N1pdm. *BMC Medicine*, 10:165, 2012.

[81] D. Balcan, H. Hu, B. Gonçalves, P. Bajardi, C. Poletto, J. J. Ramasco, D. Paolotti, N. Perra, M. Tizzoni, W. Van den Broeck, V. Colizza, and A. Vespignani. Seasonal transmission potential and activity peaks of the new influenza A/H1N1: a Monte Carlo likelihood analysis based on human mobility. *BMC Med.*, 7:45, 2009.

[82] P. Bajardi, et al. Human Mobility Networks, Travel Restrictions, and the Global Spread of 2009 H1N1 Pandemic. *PLoS ONE*, 6:e16591, 2011.

[83] P. Bajardi, C. Poletto, D. Balcan, H. Hu, B. Gonçalves, J. J. Ramasco, D. Paolotti, N. Perra, M. Tizzoni, W. Van den Broeck, V. Colizza, and A. Vespignani. Modeling vaccination campaigns and the Fall/Winter 2009 activity of the new A/H1N1 influenza in the Northern Hemisphere. *EHT Journal*, 2:e11, 2009.

[84] M.E. Halloran, N.M. Ferguson, S. Eubank, I.M. Longini, D.A.T. Cummings, B. Lewis, S. Xu, C. Fraser, A. Vullikanti, T.C. Germann, D. Wagener, R. Beckman, K. Kadau, C. Macken, D.S. Burke, and P. Cooley. Modeling targeted layered containment of an influenza pandemic in the United States. *PNAS*, 105:4639-44, 2008.

[85] G. M. Leung, A. Nicoll. Reflections on Pandemic (H1N1) 2009 and the international response. *PLoS Med*, 7:e1000346, 2010.

[86] A.C. Singer, et al. Meeting report: risk assessment of Tamiflu use under pandemic conditions. *Environ Health Perspect.*, 116:1563-1567, 2008.

[87] R. Fisher. The wave of advance of advantageous genes. *Ann. Eugen.*, 7:355-369, 1937.

[88] J. V. Noble. Geographic and temporal development of plagues. *Nature*, 250:726-729, 1974.

[89] D. Brockmann and D. Helbing. The Hidden Geometry of Complex, Network-Driven Contagion Phenomena. *Science*, 342:1337-1342, 2014.

[90] 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 Med*, 3:e40, 2006.

[91] D. Shah and T. Zaman, in SIGMETRICS'10, Proceedings of the ACM SIGMETRICS international conference on Measurement and modeling of computer systems, pp. 203-214, 2010.

[92] A. Y. Lokhov, M. Mezard, H. Ohta, L. Zdeborová. Inferring the origin of an epidemic with dynamic message-passing algorithm. *Phys. Rev E*, 90:012801, 2014.

[93] P. C. Pinto, P. Thiran, M. Vetterli. Locating the Source of Diffusion in Large-Scale Networks. *Physical Review Letters*, 109:068702, 2012.

[94] C. H. Comin and L. da Fontoura Costa. Identifying the starting point of a spreading process in complex networks. *Phys. Rev. E*, 84:056105, 2011.

[95] D. Shah and T. Zaman. Rumors in a Network: Who's the Culprit? *IEEE Trans. Inform. Theory*, 57:5163, 2011.

[96] K. Zhu and L. Ying. Information source detection in the SIR model: A sample path based approach. *Information Theory and Applications Workshop (ITA)*; 1-9, 2013.

[97] B. A. Prakash, J. Vreeken, and C. Faloutsos. Spotting culprits in epidemics: How many and which ones? *ICDM'12; Proceedings of the IEEE International Conference on Data Mining*, 11:20, 2012.

[98] V. Fioriti and M. Chinnici. Predicting the sources of an outbreak

with a spectral technique. *Applied Mathematical Sciences*, 8:6775-6782, 2012.

[99] W. Dong, W. Zhang and C.W. Tan. Rooting out the rumor culprit from suspects. *Proceedings of CoRR*, 2013.

[100] B. Barzel, and A.-L. Barabási. Universality in network dynamics. *Nature Physics*, 9:673, 2013.

[101] A. Barrat, M. Barthélemy and A. Vespignani. *Dynamical Processes on Complex Networks*. Cambridge University Press, 2012.

[102] S. N. Dorogovtsev, A.V. Goltsev, and J. F. F. Mendes. Critical phenomena in complex networks. *Reviews of Modern Physics* 80, 1275, 2008.

[103] R. Cohen and S. Havlin. *Complex Networks - Structure, Robustness and Function*. Cambridge University Press, 2010.

[104] P. Grassberger. On the critical behavior of the general epidemic process and dynamical percolation. *Mathematical Biosciences*, 63:157, 1983.

[105] M. E. J. Newman. The spread of epidemic disease on networks. *Physical Review E*, 66:016128, 2002.

[106] C. P. Warren, L. M. Sander, and I. M. Sokolov. Firewalls, disorder, and percolation in networks. *Mathematical Biosciences*, 180:293, 2002.

[107] R. Cohen, K. Erez, D. ben-Avraham, and S. Havlin. Resilience of the Internet to random breakdown. *Physical Review Letters*, 85:4626-4628, 2000.

[108] 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.