

Spreading Phenomena

SECTION 1

Introduction

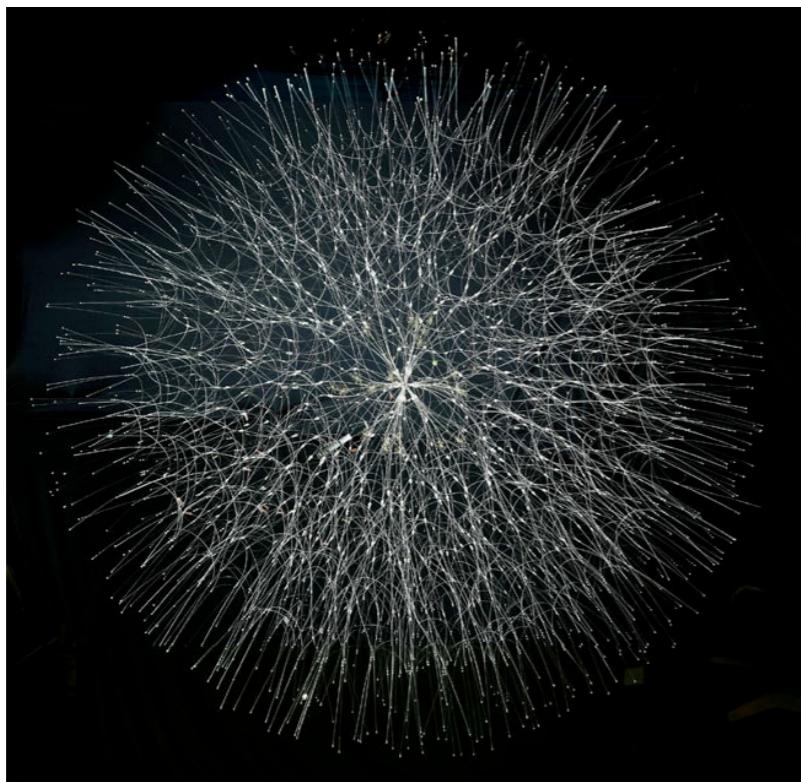


Figure 10.1 Bill Smith

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

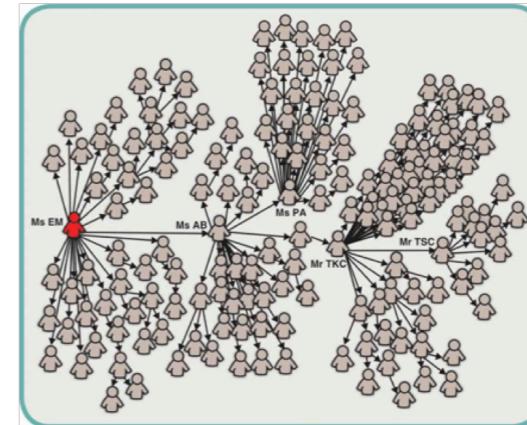


Figure 10.2 Super-spreaders.

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].

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 better name and clear diagnosis, was called “atypical pneumonia”. Next day, after leaving the hotel, he went to the local hospital, this time as a patient and died there several days later of atypical pneumonia [1].

Yet, the physician did not leave the hotel without a trace. Sixteen other guests of the Metropole Hotel and one visitor also contracted atypical pneumonia, that was later renamed Severe Acute Respiratory Syndrome, or SARS. These guests carried with them the SARS coronavirus 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 a textbook example of a *super-spreader*, an individual who is responsible for a disproportionate number of infections during an epidemic.

Those familiar with network theory 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 are not unique to SARS, but have been documented in many infectious diseases, from smallpox to AIDS [2]. The goal of this chapter is to develop a network based approach to epidemic phenomena that can help us understand and predict the true impact of these hubs. Such approaches, leading to a new formalism that we call network epidemics, will allow us to accurately 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. Typically cures and vaccines are not 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.

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

- **Biological:** The spread of pathogens (Figure 10.3) on their respective contact network is the main subject of this chapter. Examples include airborne diseases like influenza, SARS, or

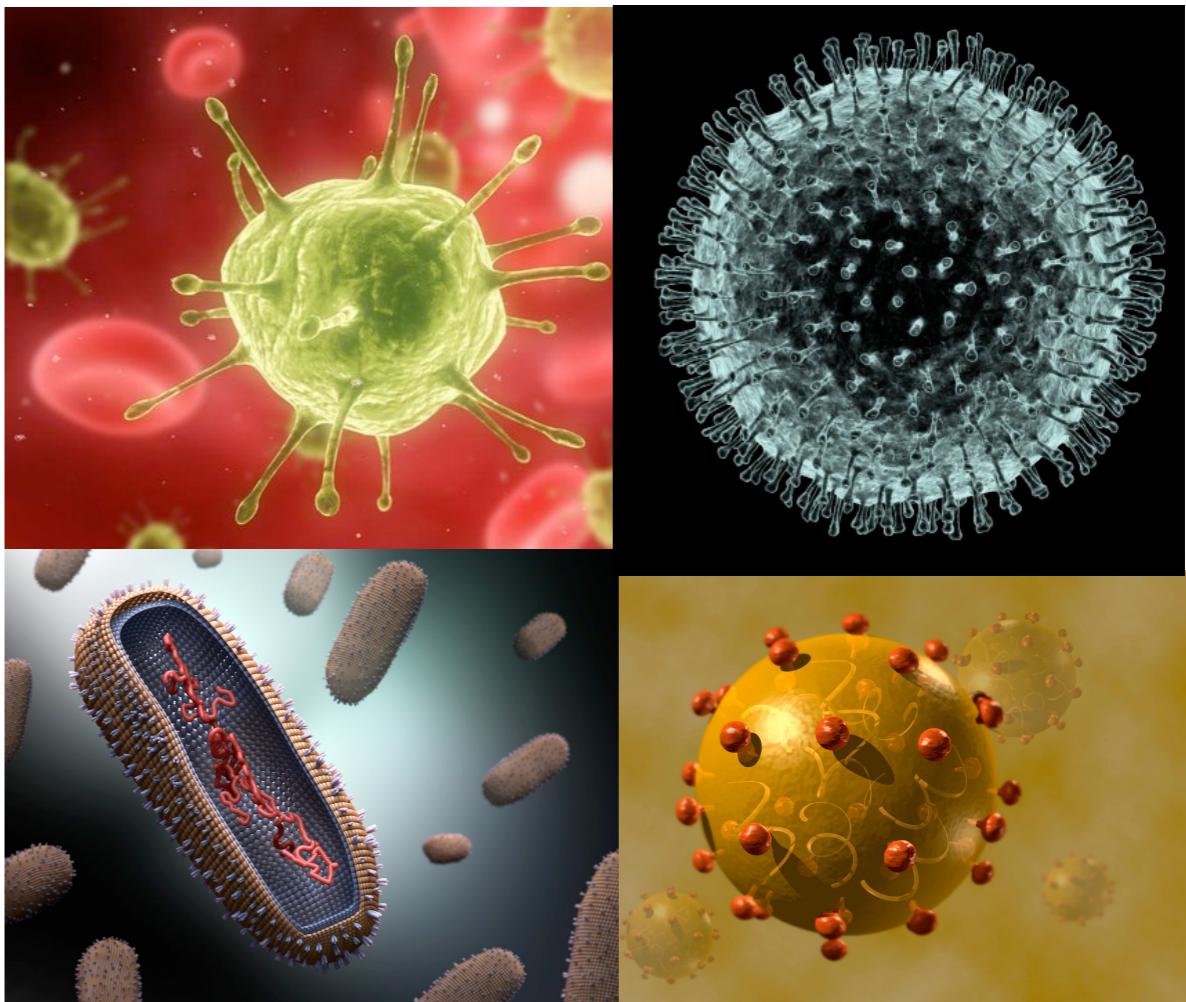


Figure 10.3 Pathogens.

A *pathogen*, a word rooted in the Greek words “suffering, passion” (*pathos*) and “producer of” (*genēs*), denotes an infectious agent or germ. Pathogens could be a microorganism, like a virus, a bacterium, a prion, or a fungus, that can cause a disease in its host. The figure shows several much-studied pathogens, like (a) the HIV virus, responsible for AIDS, (b) the SARS virus, responsible for SARS, (c) an influenza virus and (d) the hepatitis C virus. Image after (a) <http://www.livescience.com/18107-hiv-therapeutic-vaccines-promise.html> and (b-d) http://www.huffingtonpost.com/2014/01/13/deadly-viruses-beautiful-photos_n_4545309.html

tuberculosis, transmitted when two individuals breathe the air in the same room; contagious diseases and parasites transmitted when people touch each other; HIV and other sexually transmitted diseases passed on during sexual intercourse.

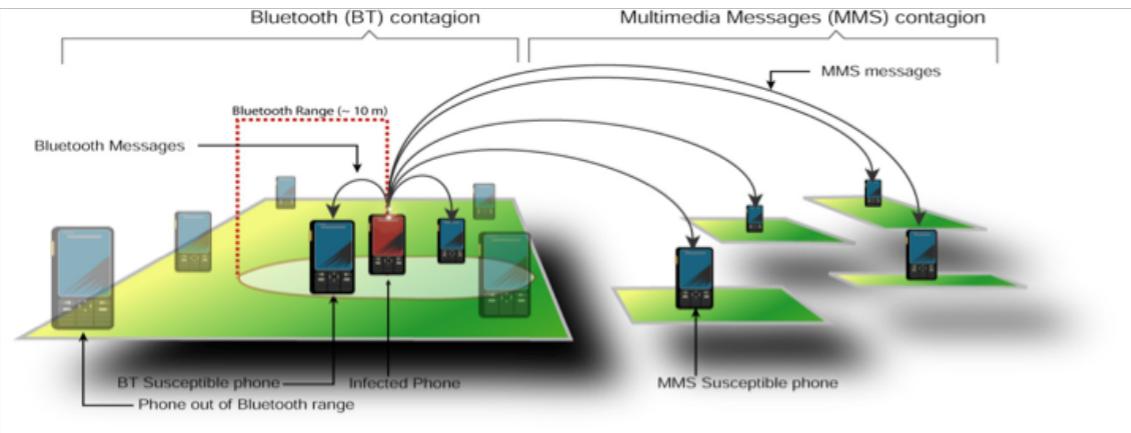


Figure 10.4 Mobile Phone Viruses.

The development of smart phones, capable of sharing programs and data with each other, turned mobile phones into a fertile ground for virus writers. Indeed, since 2004 hundreds of smart phone viruses have been identified, having reached a state of sophistication in a few years that took computer viruses about two decades to achieve [3]. Mobile viruses can be transmitted using two main communication mechanisms:

(i) A Bluetooth (BT) virus can infect all phones found within BT range from the infected phone. As physical proximity is essential for a BT connection, the transmission of a BT virus is determined by the owner's location and mobility patterns. Hence BT viruses follow a spreading pattern similar to influenza.

(ii) Viruses carried by multimedia messages (MMS) can infect all susceptible phones whose number is in the infected phone's phonebook. Hence MMS viruses follow a long-range spreading pattern that is independent of the infected phone's physical location, similar to that observed in computer viruses.

For more information on the spreading patterns of mobile phone viruses, see Ref. [4].

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

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	Bedbugs	Hotel - traveler network
Malaria	Plasmodium	Mosquito - Human network

Table 10.1 Networks and Agents

The spread of a particular agent, like a pathogen or a meme, is determined by the network on which it 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.

transmitted by humans. 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 10.4).

- **Social:** The role of the social and professional network in the acceptance and spread 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 the number of studies focusing on social spreading phenomena have exploded lately, 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: biological viruses, computer viruses, ideas or products; they spread on different types of networks, like a social network, a computer network or a professional network (Table 10.1); they follow widely different time scales and mechanisms of transmission. Despite this diversity, these spreading processes follow common patterns described by the same network-based theoretical and modeling framework. Our goal is to introduce the quantitative framework required to understand and model these spreading phenomena and to learn to exploit the predictive power of the developed network-based tools.

SECTION 2

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 into distinct states, based on the stage of the disease characterizing 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.
- **Infectious (I):** Contagious individuals who have contacted the pathogen and hence they can infect others.
- **Recovered (R):** Individuals who have been infected before, but have recovered from the disease, hence they are not infectious.

The modeling of some diseases requires additional states, like *immune* individuals, those who cannot be infected; *latent* individuals,

who have been exposed to the disease, but are not yet contagious; or *removed* individuals, that died as a result of the disease.

With time 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 too becomes infected. Eventually she will recover and develop immunity, hence from the perspective of the pathogen she will be removed from the population as she is not susceptible to the same strain of influenza any longer.

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 anyone can infect anyone else.

The purpose of this section is to introduce the traditional epidemic modeling framework built on these two hypotheses. We do so by exploring the dynamics of three frequently used epidemic models, the so-called SI, SIS and SIR models.

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 also 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 β . We wish to ask the following question: 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 $\langle k \rangle \frac{S(t)}{N}$ susceptible individuals per unit time. Since $I(t)$ infected individuals are transmitting the pathogen, each at rate β , 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 the rate of change of $I(t)$ follows

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

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 the infected population at time t , and for simplicity drop the (t) variable from $i(t)$ and $s(t)$. With this notation we can re-write (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

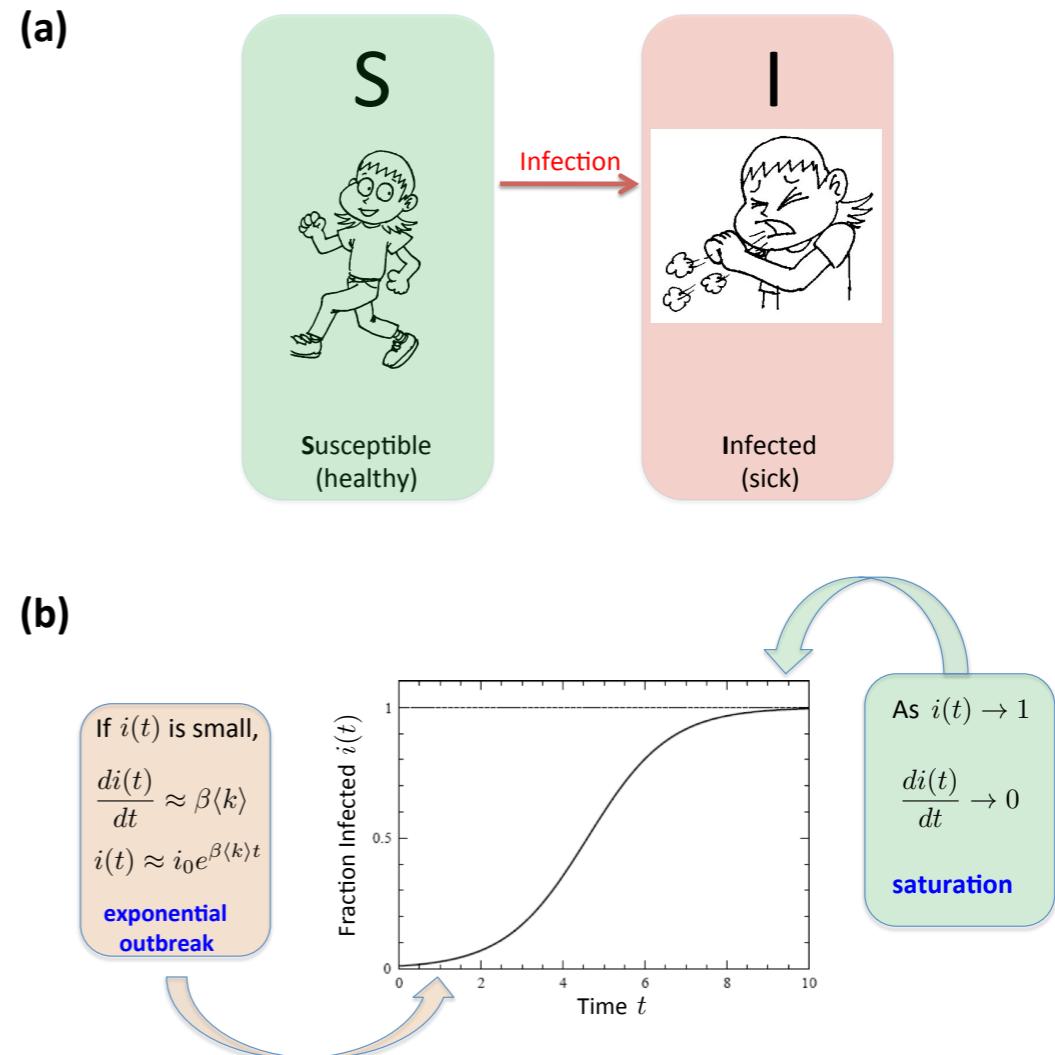


Figure 10.5 The Susceptible-Infected (SI) Model.

- (a) In the SI model an individual can be in one of two states: susceptible (healthy) or infected (sick). The arrow indicates that once an individual becomes infected, it will stay infected (i.e. it cannot recover).
- (b) The SI model assumes that if a susceptible individual comes into contact with an infected individual, it becomes infected at rate β . Hence at early times the fraction of infected individuals grows exponentially (left box). As eventually everyone becomes infected, at large times we have $i(\infty) = 1$ (right box). The plot in the middle shows the time evolution of the number of infected individuals, as predicted by (10.4). Image courtesy of Yong-Yeol Ahn.

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

Integrating both sides, we obtain

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

With the notation $C = i_0/(1 - i_0)$ and $i_0 = i(t=0)$ we obtain that the fraction of infected individuals increases in time as

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

This result predicts that:

- At the beginning the fraction of infected individuals increases exponentially (Figure 10.5b). Indeed, early on an infected individual encounters only susceptible individuals, hence it can spread the pathogen rather effectively.
- The characteristic time needed is to reach an $1/e$ fraction (about 63%) of all susceptible individuals is

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

Consequently (10.5) is the inverse of the speed with which the pathogen spreads through the population. It predicts that increasing the network density $\langle k \rangle$ or β enhances the speed with which the pathogen spreads.

- As the number of infected individuals increases, an infected individual encounters fewer and fewer susceptible individuals.

Hence the growth of i slows down for large t (Figure 10.5). The infection 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 drugs. To capture this fact we need to allow the infected individuals to recover, which means that they stop spreading the disease. This leads to 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 μ , becoming again susceptible (Figure 10.6a). 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 μ is the *recovery rate* and the μi term captures the rate at which individuals recover from the disease. The solution of (10.6) provides the fraction of infected individuals in function of time (Figure 10.6b)

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

In contrast with (10.4), that predicts that eventually everyone becomes infected, (10.7) predicts two possible outcomes for an epidemic:

- **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 (Figure 10.5). Yet, i does not saturate when everyone becomes infected ($i(\infty) \equiv i(t \rightarrow \infty) \neq 1$), but reaches a constant $i(\infty) < 1$ (Figure 10.6b). 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 term in the exponent in (10.7) becomes negative. Therefore, i decreases exponentially with time, indicating that the infection will gradually die out. The reason is that the number of individuals cured per unit time exceeds the number of newly infected individuals. Therefore with time everyone recovers and the pathogen disappears from the population.

The existence of these two outcomes suggests that some pathogens will spread while others die out shortly. To understand what governs the difference between these two outcomes we write the characteristic time of a pathogen as

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

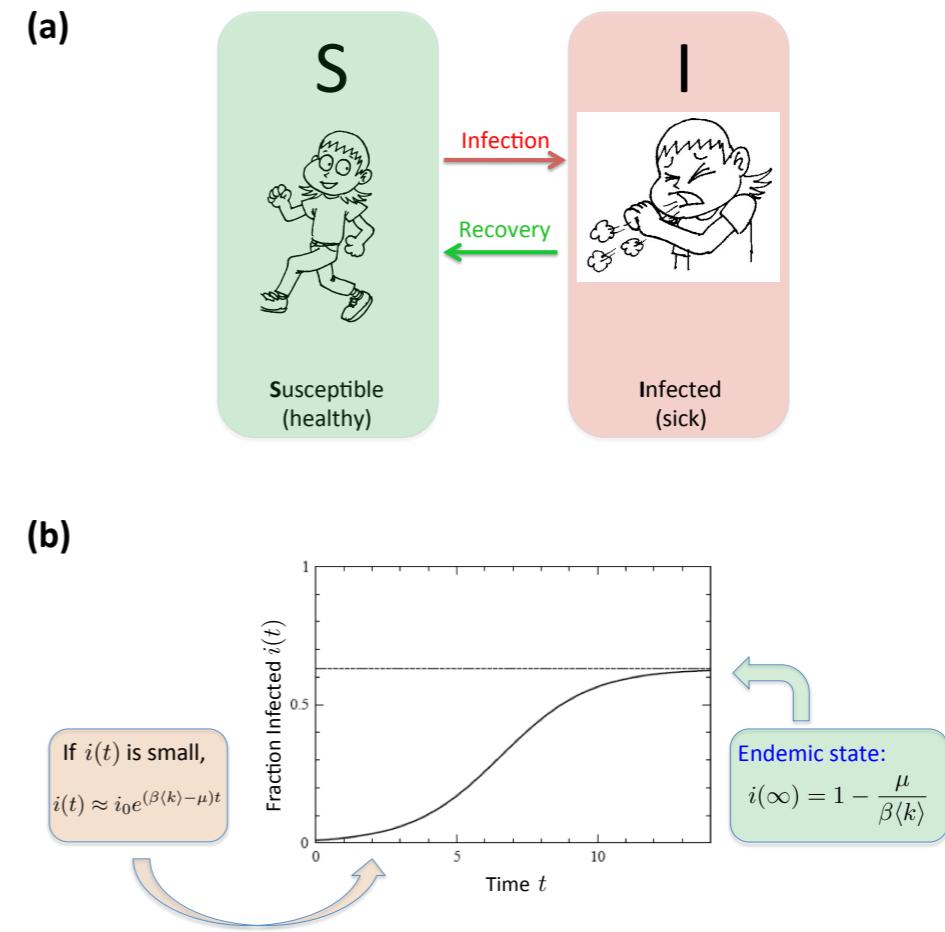


Figure 10.6 The Susceptible-Infected-Susceptible (SIS) Model

- (a) The SIS model has the same two states as the SI model: susceptible and infected. It differs from the SI model in that it allows recovery, i.e. infected individuals can become susceptible again at rate μ .
- (b) The time evolution of the fraction of infected individuals in the SIS model, as predicted by (10.6). As recovery is possible, 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, however, that for small μ the disease can also die out.

Images courtesy of Yong-Yeol Ahn.

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 pandemic 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 will infect, assuming that all its contacts are susceptible. If $R_0 < 1$ the pathogen will die out in the population, as the number of recovered individuals exceed the number of new infections. If $R_0 > 1$ the pathogen will spread. The higher is R_0 , the faster is the spreading process. The table lists R_0 for several well-known pathogens. After [7].

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 tells us the number of new infections each infected individual causes.

The basic reproductive number is valuable for its predictive power:

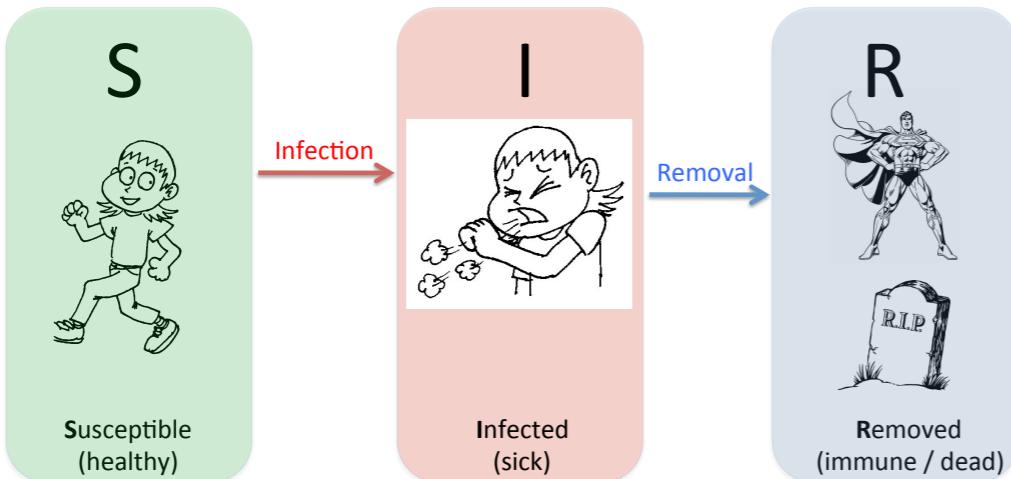
- If R_0 exceeds unity, τ is positive, hence the epidemic is in the endemic state. Indeed, if each infected individual infects more than one healthy person, the pathogen is poised to spread. The higher is R_0 , the faster is the spreading process.
- If $R_0 < 1$ then τ is negative and the epidemics dies 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 epidemiologists try to estimate for a new pathogen as it tells them the severity of the problem they face. For several well-studies pathogens R_0 listed in Table 10.2. The high R_0 of some of these pathogens indicates the dangers they pose: for example for measles each infected individual 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 matter any longer from the perspective of the pathogen as they cannot be infected, nor can they infect others. The SIR model, whose

(a)



(b)

$$\begin{aligned}\frac{ds(t)}{dt} &= -\beta \langle k \rangle i(t) [1 - r(t) - i(t)] \\ \frac{di(t)}{dt} &= -\mu i(t) + \beta \langle k \rangle i(t) [1 - r(t) - i(t)] \\ \frac{dr(t)}{dt} &= \mu i(t).\end{aligned}$$

(c)

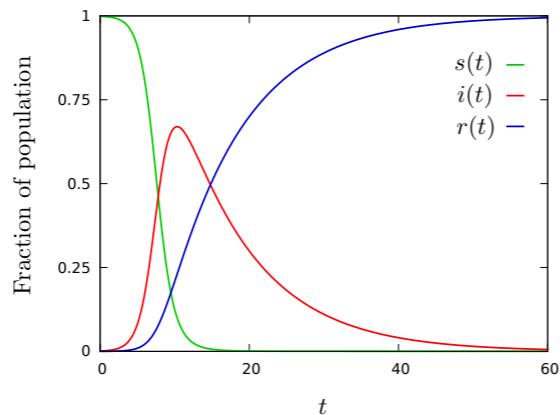


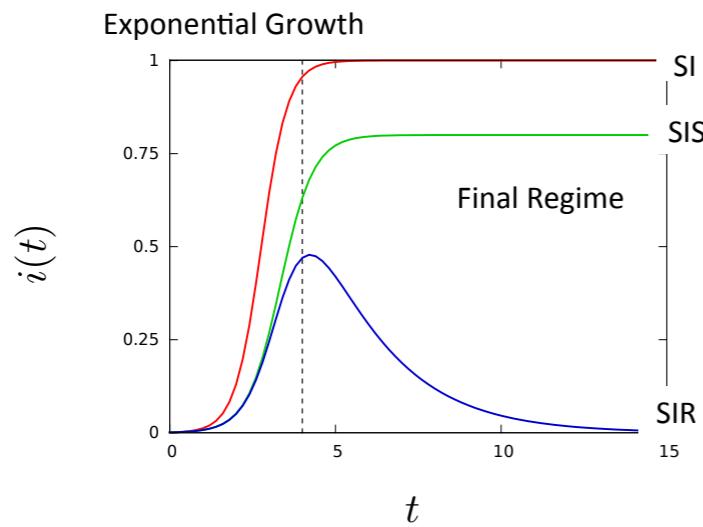
Figure 10.7 The Susceptible-Infected-Recovered (SIR) Model.

- 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. Image courtesy of Yong-Yeol Ahn.
- The differential equations governing the time evolution of the fraction of individuals in the susceptible $s(t)$, infected $i(t)$ and the removed $r(t)$ state.
- The time dependent behavior of s , i and r as predicted by the equations shown in (b). The model predicts that with time all individuals transition from a susceptible (healthy) state to the infected (sick) state and then to the recovered (immune) state.

properties are discussed in Figure 10.7, captures the dynamics of this process.

In summary, depending on the characteristics of the pathogen of interest, we need to use different quantitative models to capture the dynamics of an epidemic outbreak. As shown in Figure 10.8, 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 very small, the disease spreads freely and the number of infected individuals increases exponentially. The three models predict different outcomes 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 helps us predict the long-term fate of an epidemic: for $R_0 > 1$ the pathogen will spread, 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. To accurately predict the dynamics of an epidemic, we need to account for the precise role the contact network plays in epidemic phenomena. This is the goal of the next section.



SI

SIS

SIR

Exponential Growth:
Exponential growth
of infected individuals

$$i(t) = \frac{i_0 \exp(\beta t)}{1 - i_0 + i_0 \exp(\beta t)}$$

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

No closed
solution

Late behavior:
Saturation at $t \rightarrow \infty$

$$i(\infty) = 1$$

$$i(\infty) = 1 - \frac{\mu}{\beta}$$

$$i(\infty) = 0$$

Epidemic threshold:

Disease does not
always spread

No threshold

$$R_{0c} = 1$$

$$R_{0c} = 1$$

Figure 10.8 Comparing the SI, SIS and SIR Models.

The typical behavior of the fraction of infected individuals, i , in the SI, SIS and SIR models. Two different regimes stand out:

Exponential growth: All three models predict an exponential growth in the number of infected individuals for the early stages of the epidemic. For the same β the SI model predicts the fastest growth (smallest τ , see (10.5)), as for the SIS and SIR models the growth is slowed by recovery, resulting in a larger τ , as predicted by (10.8).

Final Regime: Asymptotically the three models predict rather different outcomes: in the SI model everyone becomes infected, $i(\infty) = 1$; in the SIS model a finite fraction of individuals are infected; in the SIR model everyone recovers, hence the number of infected individuals goes to zero. The table under the figure summarizes the properties of the three models.

SECTION 3

Network Epidemics

The ease of 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 10.9), 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 describe the spread of pathogens around the globe.

The traditional epidemic models discussed in the previous section do not explicitly incorporate the structure of the underlying 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 of these assumptions are false: in reality an individual can transmit a pathogen only to those it comes into contact with, hence pathogens spread on a complex contact network. As these contact networks are often scale-free (Section 10.4), $\langle k \rangle$ does not properly

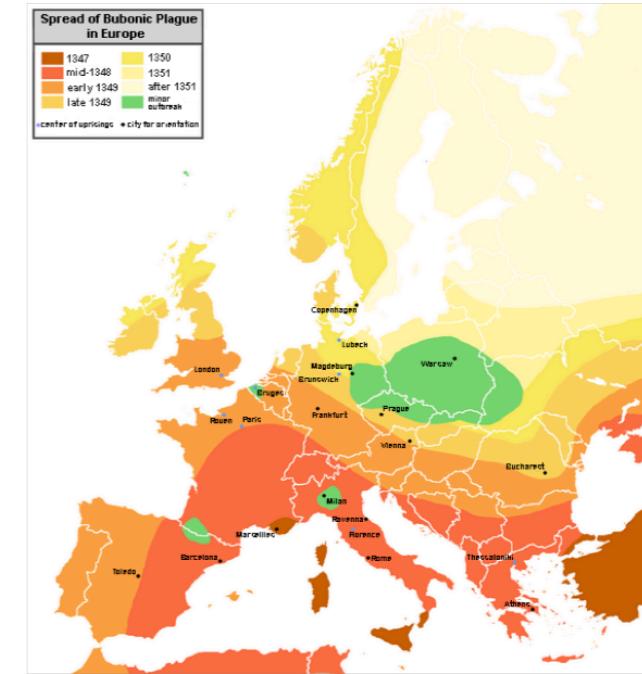


Figure 10.9 The Great Plague

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 by traveling along the Silk Road, it reached Crimea around 1346. From there, probably carried by Oriental rat fleas residing on the black rats that were regular passengers on merchant ships, spread throughout the Mediterranean and Europe. Its slow speed was due to the slow travel 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.

describe the network topology. The failure of these basic hypotheses prompted a fundamental change in epidemic modeling. 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]. The purpose

of this section is to introduce this new formalism, familiarizing ourselves with the funding concepts of *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 infected. Therefore the mathematical formalism must consider the degree of each node as an implicit variable. This is achieved using the *degree block approximation*, that distinguishes nodes based on their degree and assumes that nodes with the same degree are statistically equivalent (Figure 10.10). 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 sum of all infected degree- k nodes is the total fraction of infected 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:

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

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

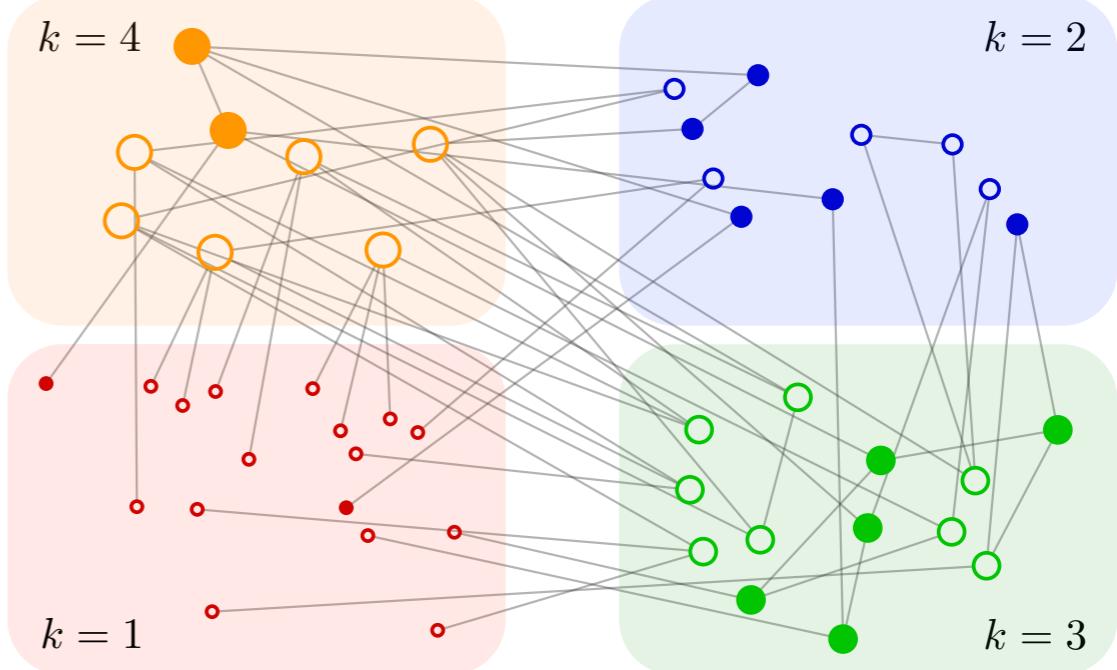


Figure 10.10 Degree Block Approximation

The epidemic models discussed in Section 10.2 group each node into compartments based on their infection 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, nodes with the same degree behave similarly. This allows us to write a separate rate equation for each (degree block), resulting in the degree block approximation of Eq. (10.13). Note that the degree block approximation does not eliminate the compartments based on the health state of individuals; independent of their degree an individual can be susceptible to the disease (empty circles) or infected (full circles).

- The average degree $\langle k \rangle$ in (10.3) is replaced with each node's actual degree k .
- The density function Θ_k represents the fraction of the neighbors of node k that are infected. In the homogenous mixing assumption Θ_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 .

- 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, a separate equation for each degree present in the network.

We start by exploring the early time behavior of i_k . This choice is 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. Lacking a cure, the only way to alter the course of an epidemic is to do so early, using quarantine-based measures and travel restrictions to slow its spread. To make the right decision about the nature, the timing and the magnitude of the intervention, we need to know 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 $\beta i_k k \Theta_k(t)$ can be neglected. Hence we can approximate (10.13) with

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

Using the $\Theta_k(t)$ function (10.40) derived in Advanced Topics 10.B, (10.14) becomes

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

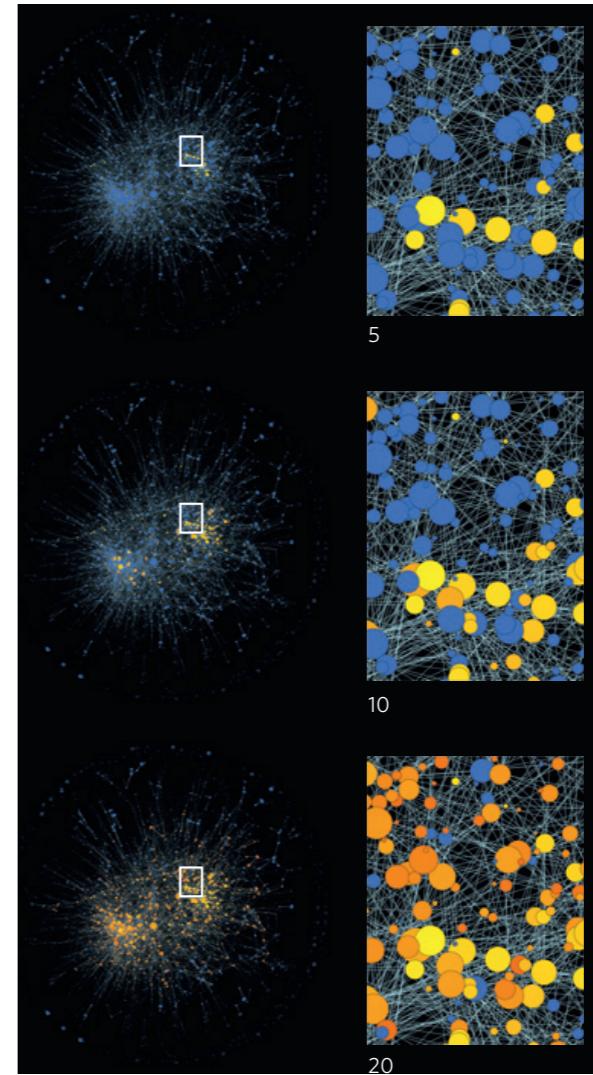


Figure 10.11 Progression of an Epidemic in the SI Model

The progression of the susceptible–infected (SI) epidemic in a scale-free network at three snapshots of the process, corresponding to times $t = 5$, 10 and 20. Susceptible nodes are shown in blue and infected nodes are colored from yellow to red according to the time of infection (red corresponding to later times). The size of each node is proportional to its degree. The figure illustrates that the large hubs are the first nodes to be infected. Then the epidemic follows a dynamical cascade from the large towards the smaller degree nodes, reaching at the end the numerous low-degree nodes as well. After [112].

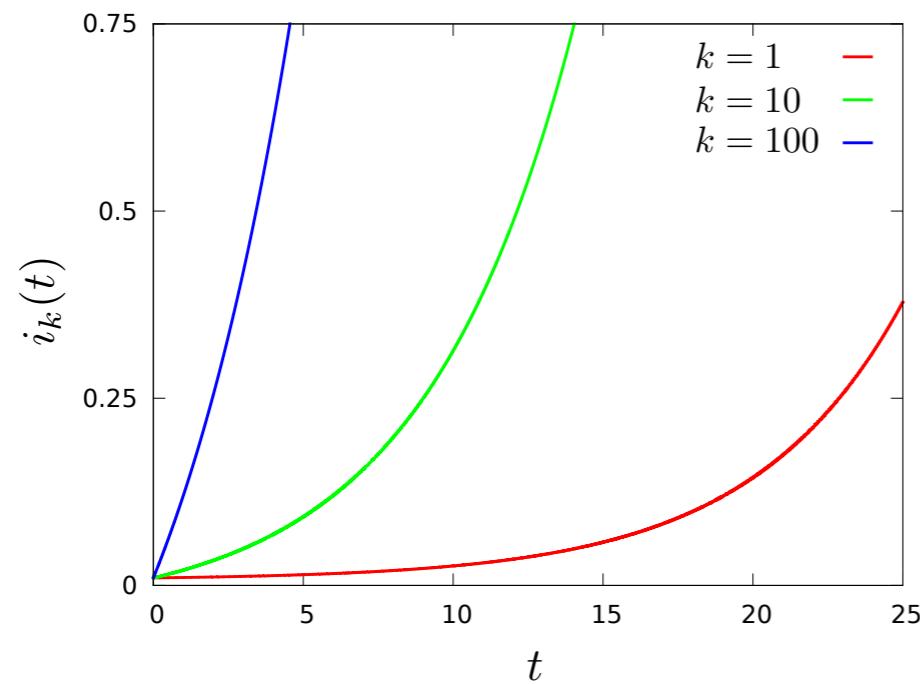


Figure 10.12 Fraction of Infected Nodes in the SI Model

Equation (10.17) predicts that the 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. This is illustrated in the figure above, that shows the fraction of infected nodes with degrees $k = 1, 10$, and 100 in an Erdős-Rényi network with average degree $\langle k \rangle = 2$ for a spreading rate $\beta = 0.1$ and initial condition $i_0 = 0.01$. As the figure indicates at $t = 5$ less than 1% of the $k = 1$ degree nodes are infected, but close to 10% of the $k = 10$ nodes and over 75% of the $k = 100$ nodes.

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

$$i_k = i_0 \left(1 + \frac{k\langle k \rangle - 1}{\langle k^2 \rangle - \langle k \rangle} (e^{t/\tau} - 1) \right). \quad (10.16)$$

A key prediction of (10.16) is that the higher the degree of a node, the more likely to be infected. Indeed, for any time t we can rewrite

(10.16) as $i_k = g(t) + kf(t)$, indicating that the group of nodes with higher degree has a higher fraction of infected nodes (Figure 10.11 and 10.12).

According to (10.12) the total fraction of infected nodes is

$$i = \int_{k_{\max}}^0 i_k P(k) dk = i_0 \left(1 + \frac{\langle k \rangle^2 - \langle k \rangle}{\langle k^2 \rangle - \langle k \rangle} (e^{t/\tau} - 1) \right). \quad (10.17)$$

Equations (10.16) and (10.17) provide the characteristic time for the spread of the pathogen on a network as

$$\tau = \frac{\langle k \rangle}{\beta(\langle k^2 \rangle - \langle k \rangle)}. \quad (10.18)$$

Therefore τ depends not only on $\langle k \rangle$, but also on the network's degree distribution through $\langle k^2 \rangle$. To understand the significance of this prediction, let us derive τ for different networks:

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

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

recovering the result (10.5) for homogenous networks.

- *Scale-free networks with $\gamma \geq 3$.* If the contract network on which the disease spreads is scale-free with a degree exponent $\gamma \geq 3$, both $\langle k \rangle$ and $\langle k^2 \rangle$ are finite. Consequently τ is also finite and the

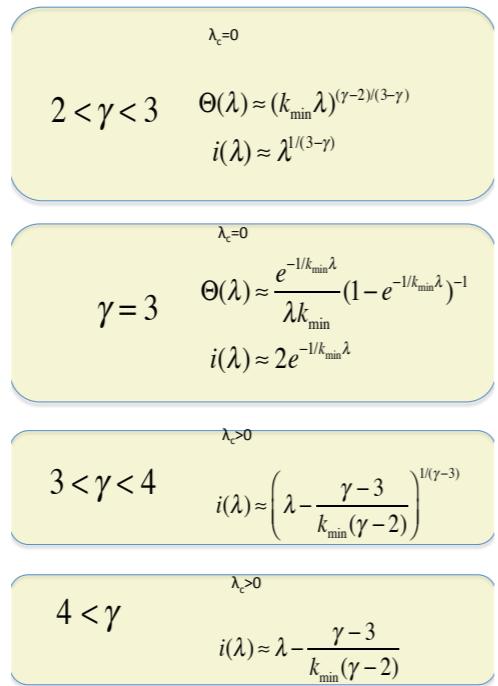


Figure 10.13 The Asymptotic Behavior of the SIS Model.

The endemic state of the SIS model on a network captures the stationary properties of an epidemic. The fraction of individuals who are infected in the endemic state, $i(\lambda) = i(t \rightarrow \infty)$, depends on the structure of the underlying network and the disease parameters β and μ . The figure summarizes the key properties of the predicted epidemic threshold λ_c , the density function $\Theta(\lambda)$ and $i(\lambda)$ for a scale-free network with degree exponent γ . The results indicate that only for $\gamma > 4$ does the epidemics on a scale-free network converge to the properties derived for a random network by the traditional epidemic models. The results are derived for a network with degree distribution (4.12). After [10].

spreading dynamics is similar to the behavior predicted for a random network but with an altered τ .

- Scale-free networks with $\gamma \leq 3$. For $\gamma < 3$ we encounter a qualitatively different behavior. Indeed, in the $N \rightarrow \infty$ limit $\langle k^2 \rangle \rightarrow \infty$ and (10.18) predicts $\tau \rightarrow 0$. In other words, the spread

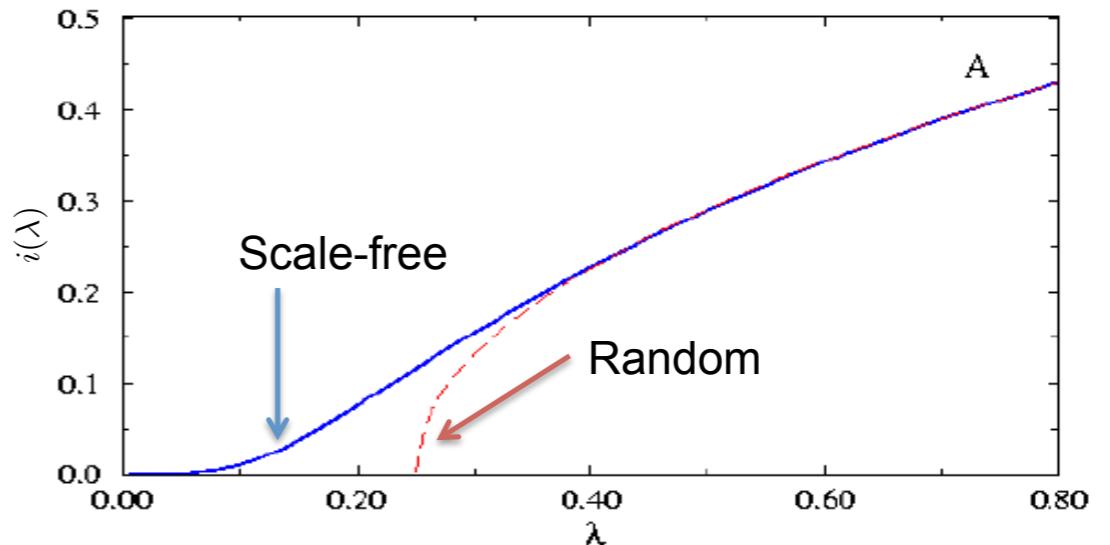


Figure 10.14 Epidemic Threshold

The fraction of infected individuals $i(\lambda) = i(t \rightarrow \infty)$ in the endemic state of the SIS model. The two curves are for a random (red curve) and a scale-free contact network (blue curve). The random network has a finite epidemic threshold, 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 λ_c , 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 λ can spread and persist in the population.

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 indicated by (10.16), 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 the super-spreaders encountered in Section 10.1.

- *Inhomogenous networks:* A network does not need to be strictly scale-free for the impact of the degree heterogeneity to be

detectable. Indeed, (10.18) predicts that as long as $\langle k^2 \rangle > \langle k \rangle(\langle k \rangle + 1)$, τ is reduced. Hence for any heterogenous network we expect an enhancement of the speed with which the pathogen spreads in the network.

In the SI model with time the pathogens reaches all individuals. Consequently the degree heterogeneity affects only the speed with which the disease sweeps through the population. To see the full impact of the network topology, we need to explore the behavior of the SIS and SIR models 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 (10.20) and (10.13) is the presence of the recovery term $-\mu i_k(t)$. This changes the characteristic time of the epidemic to (Advanced Topics 10.B)

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

For sufficiently large μ the characteristic time is negative, hence $i_k(t)$ decays exponentially. The condition for the decay depends on the network topology, through $\langle k \rangle$ and $\langle k^2 \rangle$. To obtain the condition for the pathogen to spread we define the *spreading rate*

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

which depends only on the biological characteristics of the pathogen, namely the probability of transmission (β) and the recovery rate μ . The higher is λ , the more likely that the disease will spread. But how high λ is sufficiently high? The answer is provided by the *epidemic threshold* λ_c that we calculate 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 for the pathogen to persist in the population we must have

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

Therefore, 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 10.14), with key consequences:

- If the spreading rate λ exceeds the epidemic threshold λ_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.

MODE L	MODEL	τ	λ_c
SI	$\frac{di_k(t)}{dt} = \beta[1 - i_k(t)]k\theta_k(t)$	$\frac{\langle k \rangle}{\beta(\langle k^2 \rangle - \langle k \rangle)}$	0
SIS	$\frac{di_k(t)}{dt} = \beta[1 - i_k(t)]k\theta_k(t) - \mu i_k(t)$	$\frac{\langle k \rangle}{\beta\langle k^2 \rangle - \mu\langle k \rangle}$	$\frac{\langle k \rangle}{\langle k^2 \rangle}$
SIR	$\frac{di_k(t)}{dt} = \beta S_k(t)\theta_k(t) - \mu i_k(t)$ $s_k(t) = 1 - i_l(t) - r_k(t)$	$\frac{\langle k \rangle}{\beta\langle k^2 \rangle - (\mu + \beta)\langle k \rangle}$	$\frac{1}{\frac{\langle k^2 \rangle}{\langle k \rangle} - 1}$

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 $\langle k^2 \rangle$, the corresponding characteristic τ and the epidemic threshold λ_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 τ and λ_c are derived in Advanced Topics 10.B.

- 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 by increasing the spreading rate λ is at the basis of most campaigns to stop a pathogen, as we discuss in Section 10.6.

Scale-free network: For a network with an arbitrary degree distribution we set $\tau^{SIS} \geq 0$ 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 (Figure 10.14). This means that even viruses that are hard to pass from individual to individual can spread successfully, representing 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. In a random network all nodes have comparable degree, $k \simeq \langle k \rangle$, hence if the spreading rate is under the epidemic threshold, the pathogen fails to spread. In a scale-free network, however, even if a pathogen is only weakly infectious, once it infects a hub, the hub can pass it on to a large number of other nodes, allowing it to persist in the population.

Taken together, 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 scale-free network $\tau \simeq 0$, which means that a virus can instantaneously reach most nodes.

- In a scale-free network $\lambda_c \simeq 0$, 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 (10.18) and (10.26) predict that τ and λ_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 $\langle k \rangle(\langle k \rangle + 1)$, we will observe an *enhanced spread*, resulting in a smaller τ and λ 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 as the underlying stochastic problem can be treated in its full mathematical complexity [105-108]. 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 [106-108]. Hence in reality the hubs play an even more important role than our earlier calculations indicate.

SECTION 4

Contact Networks

The network-based epidemic models predict that the spreading dynamics depends on the degree distribution of the contact network on which the pathogen spreads. Indeed, we found that $\langle k^2 \rangle$ affects both the correlation time τ and the epidemic threshold λ_c .

These predictions are of limited practical importance if the network on which diseases spread are random. Indeed, for random networks the predictions reduce to the known results of the traditional modeling framework (Section 10.2). If, however, $\langle k^2 \rangle$ of the contact network is higher than $\langle k \rangle(\langle k \rangle + 1)$, then the degree heterogeneity will lower λ_c , allowing even weakly infectious pathogens to spread. This will also lower τ , increasing the speed with which the pathogen sweeps through the population. In other words, the topology can fundamentally alter the outcome of an epidemic. The goal of this section is to inspect the structure of the contact networks encountered in epidemic processes. This confrontation with the real data is essential to understand the impact of the degree heterogeneity on the spread of a pathogen.

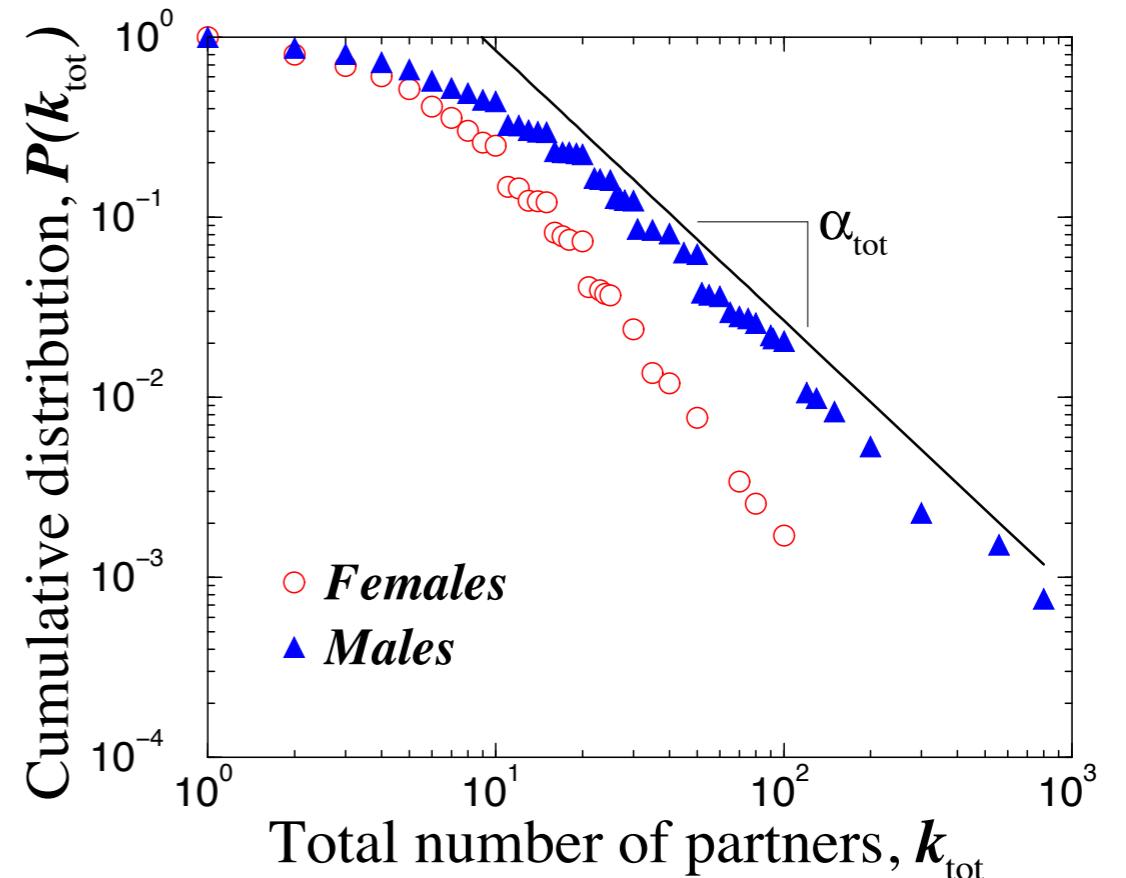


Figure 10.15 The Sexual Network

Cumulative distribution of the total number of sexual partners k since sexual initiation for individuals interviewed in the 1996 study focusing on sexual patterns in Sweden [12]. For women a power law fit to the tail indicates $\gamma = 2.1 \pm 0.3$ for $k > 20$; for men $\gamma = 1.6 \pm 0.3$ in the range $20 < k < 400$. Note that for men the average number of partners is higher than for women. This difference may be due to social expectations, prompting males to exaggerate the number of sexual partners they report. After [11].

Sexually Transmitted Diseases

The HIV virus, the pathogen responsible for AIDS, spreads mainly through sexual intercourse. Consequently, the relevant contact network needs to capture who had sexual relationship with whom, a rather confidential hence hard to collect information. The first

insight pertaining to the structure of this sexual network was provided by a survey that explored the sexual behavior of the Swedish population [12]. Through interviews and questionnaires, researchers collected information from 4,781 randomly chosen Swedes of ages 18 to 74. The study did not ask participants to reveal the identity of their sexual partners, but only to estimate the number of sexual partners they had during their lifetime. This number allowed the researchers to reconstruct the degree distribution of the sexual network [11]. As we have seen in the previous section, p_k is sufficient to predict many aspects of the spreading dynamics. The survey concluded that the degree distribution is well approximated with a power law (Figure 10.15), indicating that the sexual network has a scale-free topology. This study offered the first evidence of the relevance of the scale-free topology to the spread of pathogens. The finding was confirmed by data collected in Britain, US and Africa [13]. The scale-free nature of the sexual network indicates that most individuals have relatively few sexual partners. A few individuals, however, had sex with hundreds during their lifetime. Consequently the sexual network has a high $\langle k^2 \rangle$, which lowers both τ and λ_c .

Airborne Diseases

For airborne diseases, like influenza, SARS or HINI, the relevant contact network captures the set of individuals a person comes into physical proximity. The structure of this contact network is explored at two levels. First, the global travel network, capturing the global spread of a pathogen, is the input to virtually all large-scale epidemic prediction tools (Section 10.7). Second, recently measurements relying on digital badges have started to probe the

Box 10.1 The Hubs of the Sex Web

The scale-free nature of the sex web implies that while most individuals have only a few sexual links, a few have an exceptional number of sexual partners. But are such sex hubs real? Anecdotal evidence suggests that they are. Take for example Wild 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 [14]. “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 [15]. He is often called “patient zero”, as was one of the first homosexual men infected with HIV. Given his sexual habits, he became a super-spreader for AIDS within the homosexual community. Hubs are observed even in high school romantic networks, as shown in Figure 10.16.

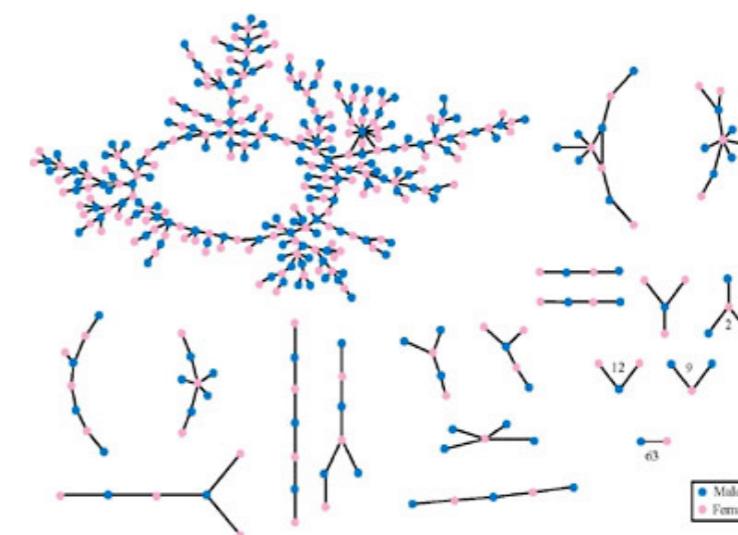


Figure 10.16 Romantic Links in a High School

The romantic and sexual links between high school students in a study focusing on over 800 adolescents living in a midwestern United States. Each circle represents a student and the links represent romantic relations reported to have taken place during the 6 months preceding the interview. The numbers indicate the number of times a particular subgraph was observed: for example there are 63 pairs of individuals (couples), that are isolated from the rest of the network. After [16].

Online Resource 10.1

A video introducing the RFA Tag technology and their use in mapping social interactions:

<http://vimeo.com/6590604>

Online Resource 10.2

Tracking Hospital Outbreaks, An Interactive Feature by Scientific American:

<http://www.scientificamerican.com/article.cfm?id=graphic-science-rfids-tags-track-possible-outbreak->

local properties of a contact network, i.e. the number of individuals a person directly interacts with. We discuss the conclusions of these investigations separately.

Global Travel Network

When it comes to the spread of airborne diseases, we must know how far is the pathogen taken once it is contacted by an individual. For this we need to explore the travel pattern of individuals. This topic has exploded with the emergence of mobile phones, that offer direct information about individual mobility patterns [17-20]. In the context of epidemic phenomena, the most relevant data comes from air travel, the mode of transportation that determines the speed with which pathogens spread around the globe. Consequently the air transportation network, that connects airports with direct flights between them, plays a key role in modeling and predicting the spread of pathogens [21-23]. As Figure

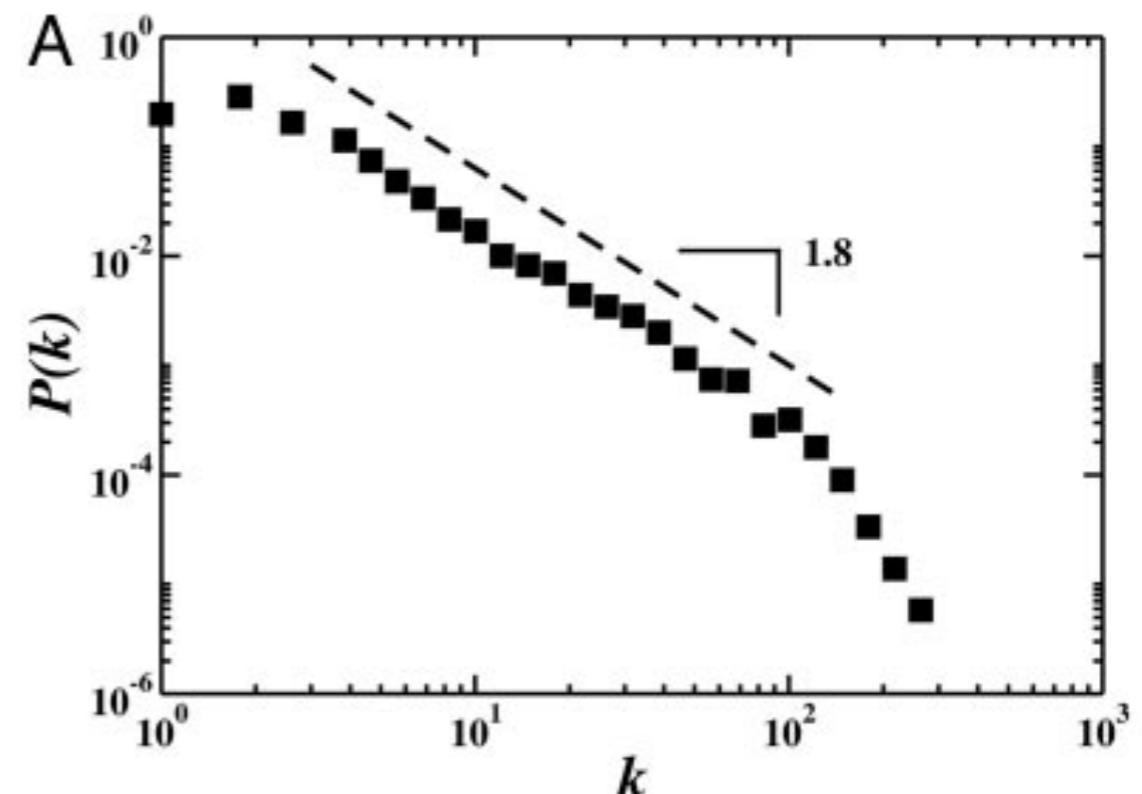


Figure 10.17 Air Transportation Network

The degree distribution p_k of the air transportation network is well approximated by a power-law for almost two decades 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 for the year 2002. The resulting worldwide air-transportation 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 [21].

10.17 shows, this network is scale-free, with degree exponent $\gamma = 1.8$. This remarkably low value is possible because there can be multiple flights between two airports, hence the network is not simple. A similar power law distribution is detected for the link weights, capturing the number of passengers that travel between two airports. As we discuss in Section 10.7, these heterogeneities

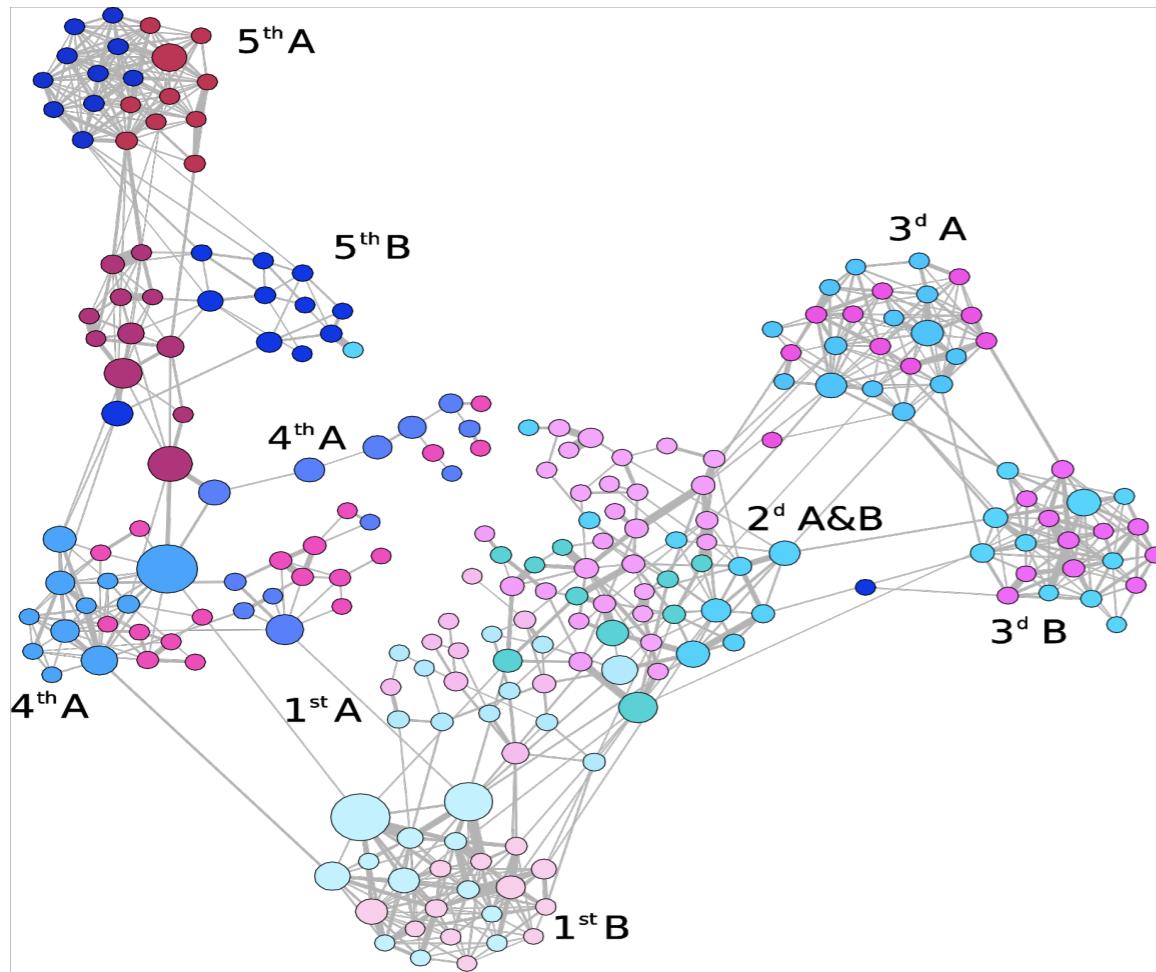


Figure 10.18 Face-to-face Interactions

The face-to-face contact networks mapped out using RFA tags in a school, capturing interactions between 232 students and 10 teachers across 10 classes [27]. The structure of the map obtained by RFID depends on the context in which it is collected: the school network reveals the presence of clear communities. In contrast, a study capturing the interactions between individuals that visited a museum reveal an almost linear network [25]; finally, a network of attendees of a small conference is rather dense, as most participants interact with most others [25]. After [27].

play a key role in predicting the spreading patterns of specific pathogens [21, 22].

Local Contact Patterns

Face-to-face interactions play a particularly important role in the spread of airborne diseases [24-27]. These interaction patterns can be monitored using Radio-Frequency Identification Devices (RFID) [25,27], mobile-phone based sociometric badges [109-110], or other wireless technologies [111].

RFID are digital badges that detect the proximity of other individuals that wear a badge (see Online Resource 10.2). 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 [25]. An RFID-mapped network is shown in Figure 10.18, describing the interactions in a high school during a two-day period [27]. Several results stand out:

- At any given moment an individual can be in face-to-face contact with only a few others. Furthermore, RFID tags detect interactions only with individuals wearing the same badge, considerably limiting the number of detected contacts. Consequently the contact networks mapped out in these studies have an exponential degree distribution.
- The duration of each face-to-face interaction follows a power law over several orders of magnitude. Therefore most contacts are brief, but there are a few lasting interactions, an uneven (bursty) temporal pattern [28] with key consequences for spreading (Section 10.5).

- The link weights, which capture the cumulative time two individuals have spent together, also follows a power law distribution. Therefore individuals spend most of their time with only a few others, with important implications for spreading phenomena (Section 10.5).
- While RFIDs capture only face-to-face interactions, 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 a SARS or HINI virus, but an RFID tag may not record this contact as an interaction.

In summary, while RFID tags provide remarkably detailed temporal and spatial information about local contacts, these datasets are so far of limited direct utility for epidemic modeling. To scale these studies up, one will need to rely on mobile-phone based technologies [29].

Location Networks

For many airborne pathogens the relevant contact network is the so-called *location network*, whose nodes are the locations connected by individuals that frequent both of them. Measurements combined with agent-based simulations indicate that the location network is fat tailed [30]: malls, airports, schools or supermarkets act as hubs, being linked to an exceptionally large number of smaller locations. Therefore, once the pathogen infects a hub, the disease can rapidly reach the other locations.

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. Some of the early computer viruses spread as email attachments. Once a user opened the attachment, the virus infected the user's computer and mailed a copy of itself to all email addresses found in the computer. In this case the relevant contact network is the email network, which, as we discussed in Table 4.1, is scale-free [31]. More recent computer viruses exploit various communication protocols, spreading on networks that reflect the pattern of interconnectedness characterizing the Internet, 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 10.4). An MMS virus sends a copy of itself to all phone numbers that a phone has in its 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. This co-location network is also highly heterogeneous [4].

In summary, we find that the contact networks that contribute to the spread of biological or digital viruses show significant degree heterogeneity. Many of these, like the email network, the internet, or sexual networks, have a well documented scale-free topology.

For others, like location or co-location networks, the degree distribution cannot be fitted with a simple power law, yet these networks have a high second moment, $\langle k^2 \rangle$. Therefore the analytical results obtained in the previous section are of direct relevance to pathogens spreading on these networks, allowing even weakly virulent viruses to easily spread through the contact network.

SECTION 5

Beyond the Degree Distribution

So far we have kept our models simple: we assumed that pathogens spread on an unweighted network 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 are not continuous, but have a finite temporal duration. The goal of this section is to explore the impact of these properties on the spread of a pathogen.

Temporal Networks

The existence of a link in a social network does not mean that two individuals are in constant interaction with each other. To the contrary, 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 of each interaction. The misleading conclusions one can reach by ignoring the timing is illustrated in

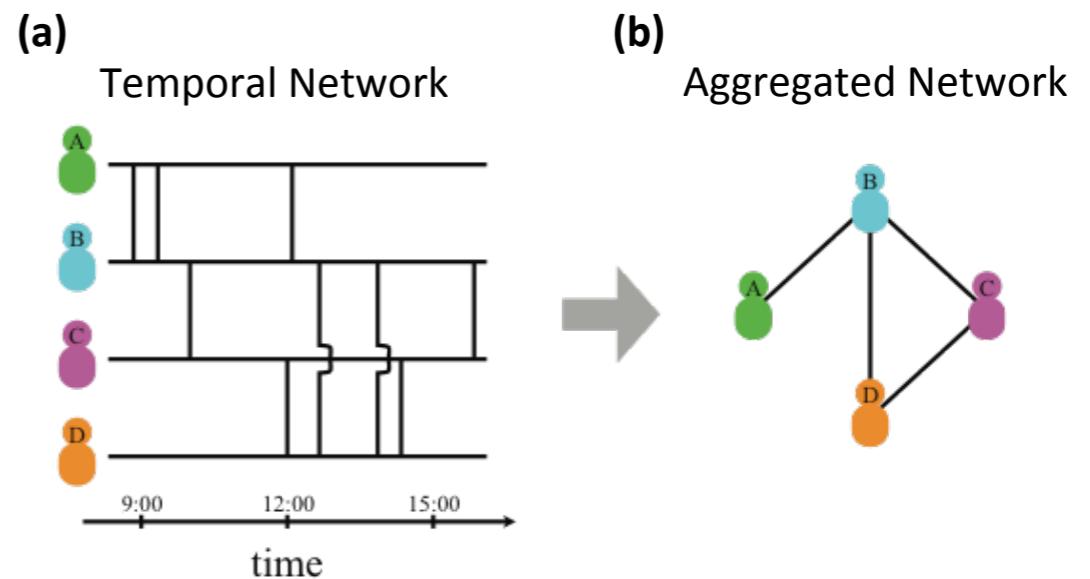


Figure 10.19 Temporal Networks

(a) *Temporal Network*: The timeline of the interactions between four individuals. Each vertical line shows the moment when two individuals come into contact with each other. Therefore, if A is the first to be infected, the pathogen can spread from A to B and then to C, finally 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 but there is no such path from D to A. (b) *Aggregated network* obtained by merging the temporal interactions shown in (a). After [33].

Figure 10.19 [32, 33, 34]. The static network of Figure 10.19b was obtained by aggregating the individual interactions shown on Figure 10.19a. 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 in Figure 10.19a, we realize that while an infection starting from A can spread to 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 [33, 34, 35]. By ignoring the temporality of these contact patterns,

we typically overestimate the speed and the extent of an outbreak [35, 36].

Busty Contact Patterns

An implicit assumption of the theoretical approaches discussed in the Sections 10.2 and 10.3 is that the timing of the interactions 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 10.20a-c). The measurements indicate otherwise: the interevent times in most social systems follow a power law distribution [28, 37] (Fig. 10.20d-f). This means that the sequence of contacts between two individuals is characterized by periods of frequent interactions, i.e. events following each other within a relatively short time frame. Yet, the power law also implies that occasionally there are very long time gaps. Therefore the contact patterns have an uneven, “busty” character in time (Figure 10.20d,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, bursty contact patterns alter the dynamics of the spreading process [36]. To be specific, power law interevent times increase the characteristic time τ , hence the number of infected individuals decays slower than predicted by a random contact pattern. For example, if the interevent times of email communications would follow a Poisson distribution, we would expect $i(t) \sim \exp(-t/\tau)$ with a decay time of $\tau \simeq 1$ day. In the real data, however, the decay time is $\tau \simeq 21$ days, a considerably slower process correctly predicted if we use power

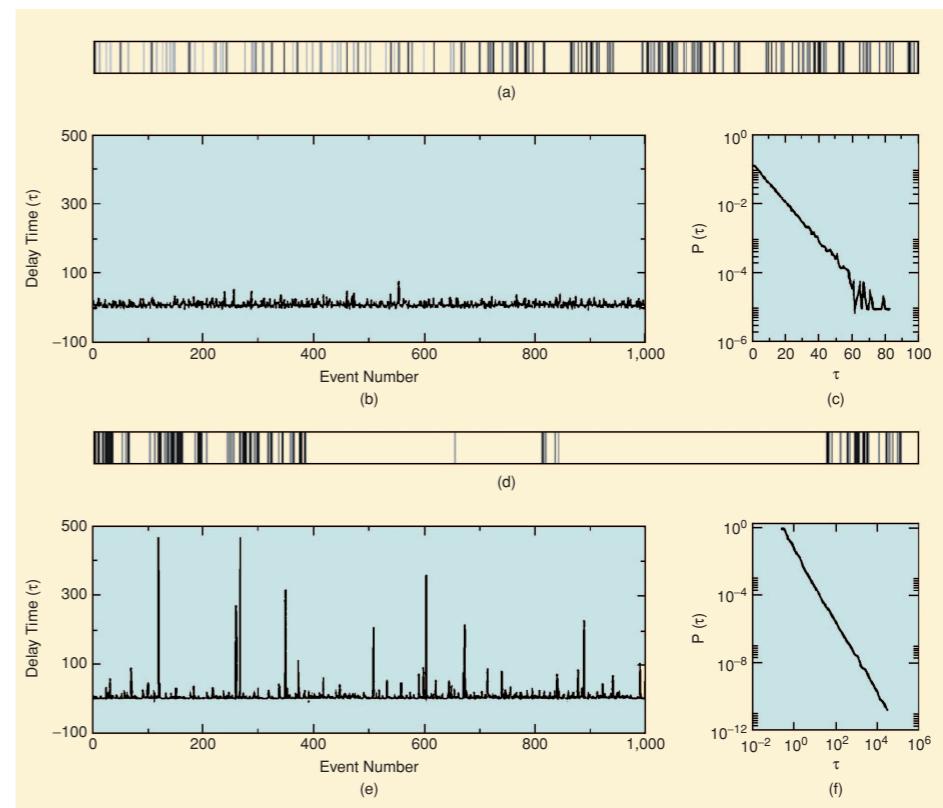


Figure 10.20 Bursty Interactions.

(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 absent. (b) The absence of long delays is visible on the plot showing the inter-event times τ for 1,000 consecutive 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 $P(\tau) = e^{-q\tau}$, shown on a log-linear plot. (d) The succession of events for a bursty pattern, when the interevent times follow a power-law distribution. While most events follow each other closely, there are a few very long interevent times, appearing as long delays in the contact pattern. The time sequence is not as uniform as in (a), but has a bursty character. (e) The waiting time τ 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). Note the presence of large spikes, which correspond to long delay times. (f) The delay time distribution $P(\tau) = \tau^{-2}$ for the bursty process shown in (d) and (e). After [38].

law interevent times [36].

Degree Correlations

As discussed in Chapter 7, many social networks are assortative, implying that high degree nodes are more likely to connect to other high degree nodes than expected by chance. Do such degree correlations affect the spread of a pathogen? The calculations indicate that degree correlations leave the key predictions of network epidemics in place, but alter the speed of an epidemic:

- Degree correlations alter the epidemic threshold λ_c : assortative correlations decrease λ_c and dissassortative correlations increase it [39, 40].
- Despite the changes in λ_c , for the SIS model the epidemic threshold vanishes for a scale-free network with diverging second moment, independent if the network is assortative, neutral or disassortative [41]. Hence the fundamental features of the results discussed in Section 10.3 are not affected by degree correlations.
- Given that hubs are the first to be infected in a network, assortativity is expected to accelerate 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 [42].

Link Weights and Communities

Throughout this chapter we assumed that all tie strengths are equal, hence we focused our attention on pathogens spreading on unweighted networks. 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 a similar fashion, we also ignored the community structure of the network on which the pathogen spreads. Yet, as we discussed in Chapter 9, real networks form detectable communities, which lead to repeated interactions between the nodes within the same community, potentially changing the spreading dynamics.

To explore the role of tie strengths and communities on spreading phenomena, we focus on the mobile phone network, where the strengths can be accurately determined [43]. Let us assume that at $t = 0$ we provide a randomly selected individual with some key information. At each time step the “infected” individual i will pass the information to his/her contact j with probability $p_{ij} \sim \beta w_{ij}$, where β is the spreading rate and w_{ij} is the number of minutes the two individuals have spent talking 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 precise role of the link weights in the spreading process, we also consider the control situation when the spreading takes place on the same network, but we make all tie strengths equal, $w = \langle w_{ij} \rangle$.

As Fig. 10.21a shows, information travels significantly faster on the control network (all weights are equal). The slower 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 [44]. Therefore once the information reaches a member of a community, it can rapidly spread to all other members of the same community, given the strong ties between them. Yet, as the ties between the communities are weak, the interaction 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 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 10.21b and c, that captures the spreading pattern in a small neighborhood of the mobile call network. In the control run the information typically follows the shortest path. When the link weights are taken into account, information flows along a longer backbone of strong ties. For example, the information rarely reaches the lower half of the network in Figure 10.21b but always reaches it in the control runs, as shown in (c).

In general, community structure has multiple effects on spreading, including the emergence of global cascades [45, 46], and altering the activity of individuals [44, 47].

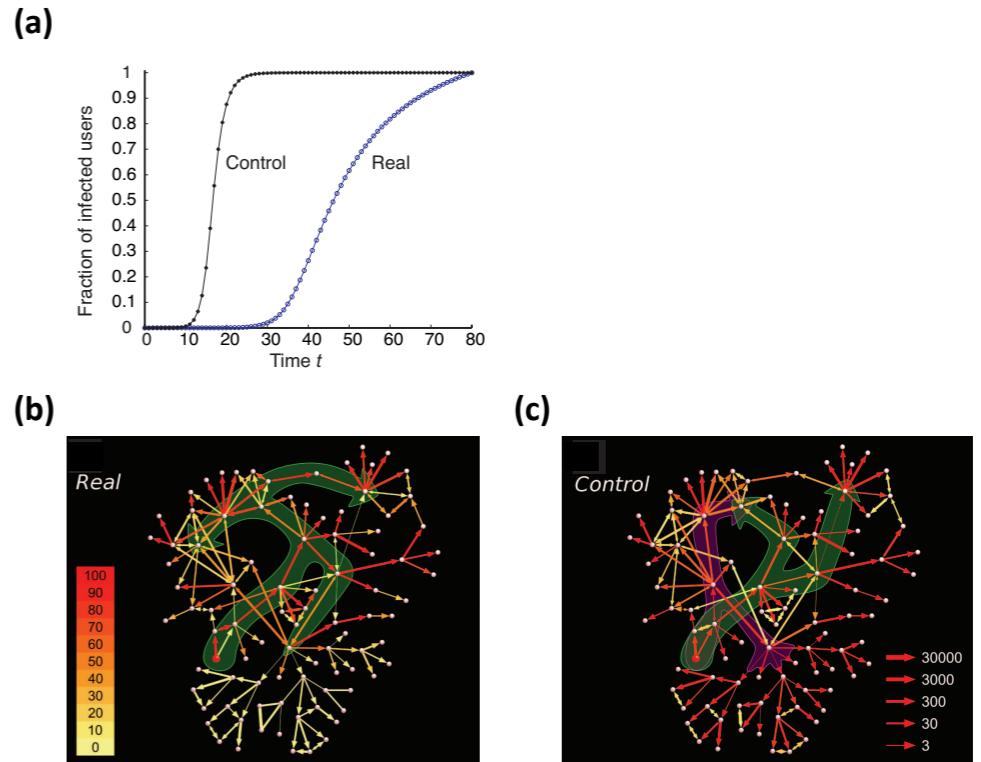


Figure 10.21 Information Diffusion in Mobile Phone Networks

The spread of information on a weighted mobile call graph. We assume that the information follows the SI model on a weighted network, so that the probability that a node passes information to one of its neighbors is proportional to the strength of the tie between them. (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 black symbols represent the control case, when all tie strengths are equal. (b) Spreading of the infection in a small network neighborhood using the real weights. The information is released from the red node, and the color of each arrow indicates the corresponding 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 numbers in the color scale refer to percentages of 1,000). The green background contours are guides to the eye, highlighting the difference in the direction the information follows in the real and the control simulations. (c) same in (b), but each link has the same weight $w = \langle w_{ij} \rangle$ (control simulation). After [43].

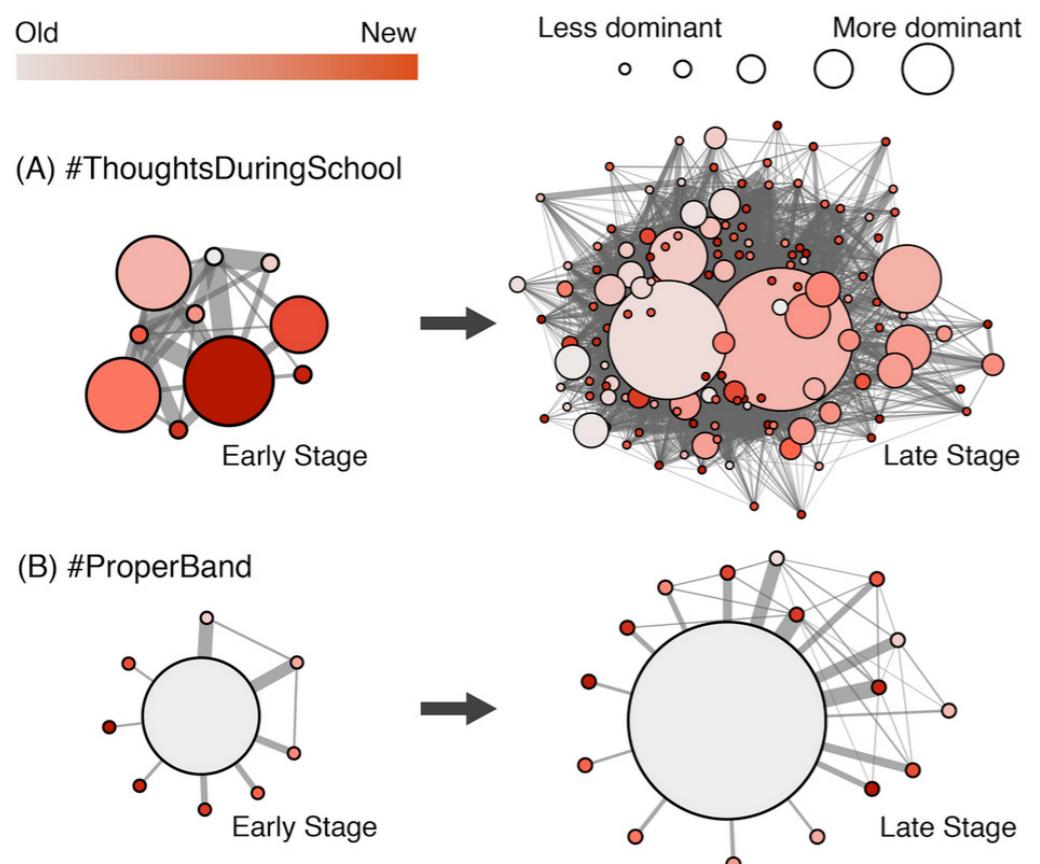


Figure 10.22 Simple vs. Complex Contagion

The figure shows the community structure of the Twitter follower network. The size of each node is proportional to the number of tweets produced by the 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 a viral meme (#ThoughtsDuringSchool hashtag) from the early stage (30 tweets, left) to the late stage (200 tweets, right). The meme easily jumps between communities, infecting many of them.

(b) Complex Contagion: The evolution of a non-viral meme (#ProperBand hashtag) from the early stage (left) to the final stage (65 tweets, right). The tweet tends to be trapped in a few communities, having difficulty to escape them. This is a signature of reinforcement, indicating that the meme follows complex contagion. After [48].

Online Resources 10.3

Spreading in Social Networks

http://www.ted.com/talks/nicholas_christakis_how_social_networks_predict_epidemics.html

“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.”

Nicholas Christakis, TED talk.

Complex Contagion

The diffusion of memes, representing ideas or behavior that spread from individual to individual, helps us better understand the role communities in spreading phenomena [48]. Meme diffusion has attracted considerable attention from marketing [5, 49], to network science [50, 51] communications [52], and social media [53-55]. To model meme diffusion we must realize that biological pathogens and memes can follow different spreading patterns, prompting us to systematically distinguish simple from complex contagion [48, 56, 57].

Simple contagion is the process we explored so far: to be infected by the pathogen it is sufficient to come into contact with an infected individual. *Complex contagion* describes the spread of behavior and is different in the sense that most individuals do not adopt a new behavioral pattern at the first contact. Rather, the adoption of a behavioral pattern often requires reinforcement [58], i.e. contact with several individuals who have already adopted. For example, in the early times of mobile phones, the higher was the fraction of a person's friends that had a phone, the more likely that she also bought one.

As we showed above, in simple contagion communities can trap an information or pathogen, slowing the spreading process. The effect goes in the opposite direction in complex contagion: because communities have more redundant ties, they offer social reinforcement, exposing an individual to multiple examples of adoption. Hence communities can enhance the diffusion of memes.

The difference between simple and complex contagion is illustrated by Twitter data. Tweets, or short messages, are often labeled with *hashtags*, which are keywords that can be seen 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 [48]. A high concentration of a meme within a certain community is evidence of reinforcement, making the meme interesting only to the members of that community. In contrast, viral memes can spread across communities, following a pattern similar to that encountered

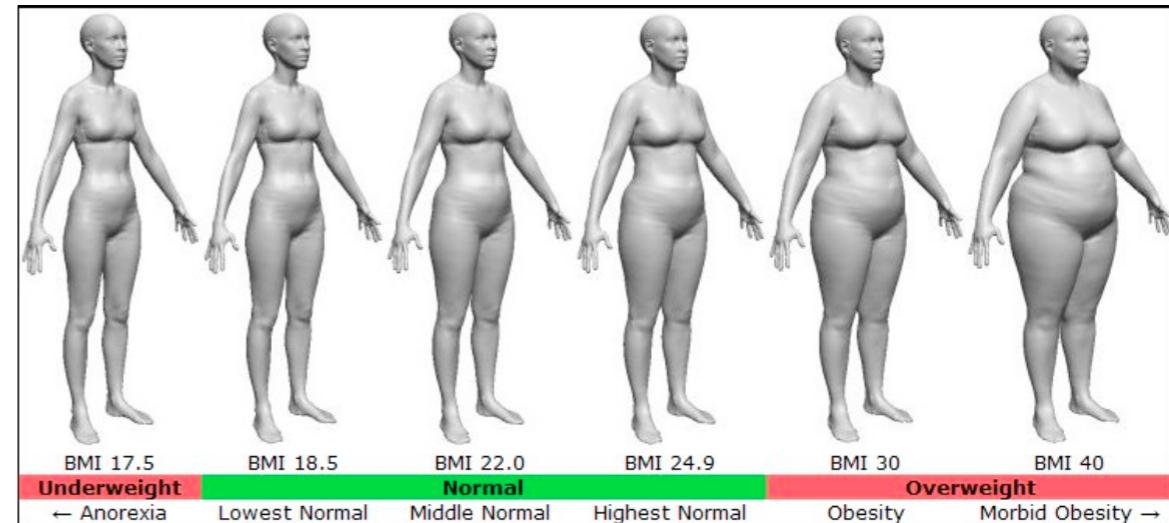


Figure 10.23 Body Mass Index

The body mass index (BMI) captures the shape of a human body based on an individual's mass and height. Proposed around 1850 by the Belgian scientist Adolphe Quetelet, BMI is defined as the individual's body mass divided by the square of the height, in units of kg/m^2 . The World Health Organization regards individuals with BMI of less than 18.5 as underweight, a value that may indicate malnutrition, an eating disorder, or other health problems. A BMI greater than 25 is considered overweight and above 30 is obese.

in biological pathogens. In general the more communities a meme reaches, the more viral it is (Fig. 10.22).

Do Our Friends Make Us Fat?

Infectious diseases, like influenza, SARS, or AIDS, are known to spread from individual to individual through the transmission of the appropriate pathogen. But could the social network influence the spread of noninfectious diseases as well? Recent measurements indicate that network effects are present in non-infectious diseases as well, impacting the spread of obesity, happiness, or behavior patterns, like giving up smoking [59, 60].

Obesity is diagnosed through an individual's body-mass index (BMI, Figure 10.23), which is determined by numerous factors, from genetics to diet and exercise. Yet, the social network also plays an important role. The analysis of the social network of 5,209 men and women enrolled in the Framingham Heart Study indicates that if one of your friends is obese, the risk that you too would gain weight in the next two to four years increases by 57% [59]. The risk triples if your best friend is the one that is overweight: in this case, your chances that you too increase your weight jumps by 171% (Figure 10.24). For all practical purposes, obesity appears to be just as contagious as influenza or AIDS, despite the lack of a pathogen that transmits it.

In summary, most network characteristics affect the spread of a pathogen in a network, from degree correlations to link weights, from the temporal patterns of the interactions to the burstiness of the contact pattern. Our goal in this section has been to briefly catalogue the impact of these network characteristics on spreading phenomena; given the diversity of the effects, for a more detailed discussion one needs to consult the primary literature.

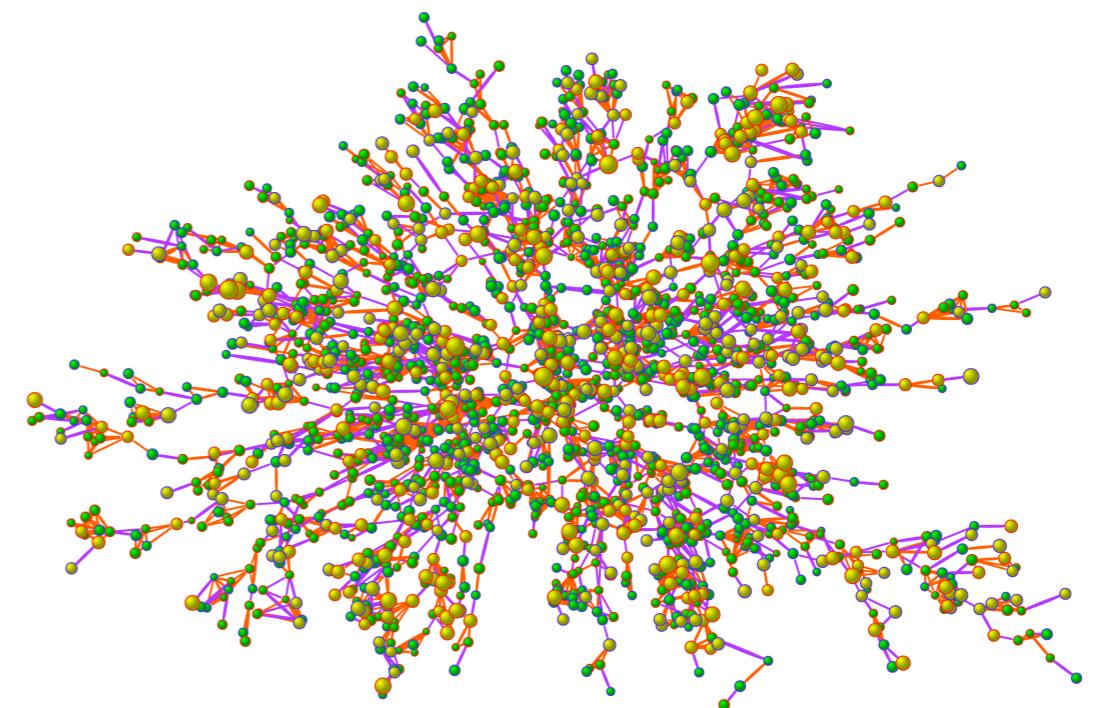


Figure 10.24 The Web of Obesity

The largest connected component of the social network capturing the friendship ties between 2,200 subjects of 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 and yellow nodes denote obese individuals ($BMI \geq 30$). The link colors indicate the nature of each relationship: purple links denote a friendship or marital tie and orange links denote family ties (e.g. siblings). Clusters of obese and non-obese individuals are visible in the network. The statistical analysis indicates that these clusters cannot be attributed to homophily, i.e. the fact that individuals of similar body size may preferentially be friends with each other. After [59]

SECTION 6

Immunization

Immunization strategies specify how vaccines, treatments or drugs are distributed in the population. Ideally, should a vaccine or treatment exist, everyone at risk of contracting the pathogen should receive it. This is often not feasible, full coverage being prohibited by cost considerations, the ability to reach all individuals at risk, real or perceived side effects of the treatment and the resistance of some individuals to accept the vaccine or the treatment. Given these constraints, immunization strategies aim to minimize the threat of a pandemic by most effectively distributing the available vaccines or treatments.

Most immunization strategies are guided by the prediction of the traditional epidemic models that if a pathogen's spreading rate λ is reduced under its critical threshold λ_c , the virus naturally dies out (Figure 10.14). The discovery that the epidemic threshold vanishes in scale-free networks has questioned the effectiveness of this strategy. Indeed, if the epidemic threshold vanishes, immunization strategies will have difficulty reducing λ under λ_c . In this section we

Box 10.2 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, like the HIV virus responsible for AIDS.

Contact-reducing interventions: For diseases with more severe health consequences officials can quarantine patients, close schools and limit access to frequently visited public spaces, like movie theaters or malls. These reduce the number of contacts between individuals, making the network sparser, 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. As we show in this section, vaccinations also reduce the spreading rate, enhancing the likelihood that the pathogen dies out.

discuss how to exploit our understanding of the network topology to design more effective network-based immunization strategies.

Random Immunization

The purpose of immunization is to protect the immunized individual from an infection. Equally important, however, is its secondary role: it reduces the spreading rate λ , hence slowing 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 start with a pathogen that follows the SIS model (10.5). If a g fraction of nodes are immunized, these nodes will not contribute to the spread of the disease, effectively becoming invisible to the pathogen. Only the remaining $(1 - g)$ fraction of the nodes are capable of contacting and spreading the pathogen. 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 $\lambda \equiv \beta/\mu$ to $\lambda' = \lambda(1 - g)$.

Random Networks: If the pathogen spreads on a random network, for a sufficiently high g the spreading rate λ' could fall below the epidemic threshold $\lambda_c = 1/(\langle k \rangle + 1)$ (10.25). The immunization rate g_c needed to achieve this is obtained by setting

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

obtaining

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

Consequently, if vaccination increases the fraction of immunized individuals above g_c , it pushes the spreading rate under the epidemic threshold λ_c . In this case τ becomes negative and the pathogen dies out naturally. This explains the importance that a high fraction of the population takes 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

Box 10.3 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 turned out to be ineffective in densely populated regions. The procedure that eventually stopped the transmission relied on simple network concepts: the authorities set as a goal to find and treat anyone who had been in contact with an infected individual, thereby preventing transmission to others. With that smallpox became the first disease to have been fought on a global scale and it was officially eradicated in 1980 (Figure 10.25).

Eradication is the complete elimination of a pathogen from the population. To select an infectious disease for eradication, the officials must make sure that the targeted pathogen does not have a non-human reservoir, so human vaccination can be effective against it and that there is an efficient and practical vaccine or drug to interrupt its transmission. So far eradication campaigns had a mixed success: while the eradication of smallpox and rinderpest was successful, 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 [62]

sexual network. Hence the classical epidemic models predict that for each of these viruses a sufficiently high immunization rate can revert the spreading process. The reason why a sufficiently high rate of immunization is difficult to achieve is discussed next.

Heterogenous Networks: Let us next consider the impact of random immunization on a disease that spreads on a network with high $\langle k^2 \rangle$. Random immunization changes λ to $\lambda(1 - g)$, hence now (10.26) allows us to determine the critical immunization g_c , using

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

obtaining

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

For a random network (10.29) reduces to (10.27). For a scale-free network with $\gamma < 3$ we have $\langle k^2 \rangle \rightarrow \infty$, so (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 eradication requires us to immunize 80%-100% of the population. 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 [61].

To illustrate the key role degree heterogeneity plays in immunization let us consider an email virus that spreads on the

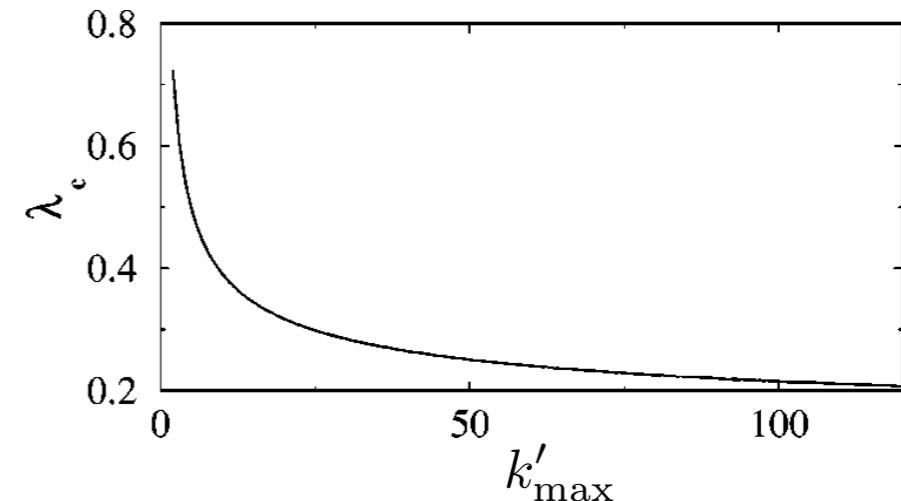


Figure 10.26 Immunizing the Hubs.

In heterogenous networks, like a scale-free network, we can eliminate a virus by immunizing the hubs, which increases the epidemic threshold. The figure shows the expected epidemic threshold if we immunize all hubs with degree larger than k_0 . The more hubs are immunized (i.e. the smaller is k_0), the larger is λ_c , i.e. the higher is the chance that the disease dies out. By immunizing the hubs, we effectively change the network on which the disease spreads, as the hubs are not visible any longer to the pathogen (Box 10.4).

email network. For $\langle k \rangle = 3.15$ (undirected case, Table 4.1), if the email network were to be random, (10.27) predicts $g_c = 0.75$. In other words, we could eradicate the virus by convincing 75% of computer users to update their antivirus software. Yet, the email network is scale-free with $\langle k^2 \rangle = 1,271$, hence (10.27) does not apply. In this case (10.29) predicts $g_c = 0.997$, i.e. more than 99.7% of the users must install the software to halt the email virus. This level of compliance is virtually impossible to achieve - many users simply ignore all warnings. This is the reason why many email viruses linger for years and disappear only after the operating systems that supports them is phased out [61].

Vaccination Strategies in Scale-free Networks

The problems encountered with random immunization are rooted in the vanishing epidemic threshold. Therefore, to successfully eradicate a pathogen in heterogenous networks, we must increase the epidemic threshold. This requires us to modify the underlying contact network to reduce its variance, $\langle k^2 \rangle$.

The hubs are responsible for the large variance of heterogenous networks. Therefore if we immunize the hubs, i.e. nodes whose degree exceeds some preselected value k'_{\max} , we decrease the variance and increase the epidemic threshold according to (10.26) [63, 64]. Indeed, if nodes with degrees $k > k'_{\max}$ are present, the epidemic threshold changes to (Advanced Topics 10.B)

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

Therefore, if $\gamma < 3$, the more hubs we cure (i.e. the smaller is k'_{\max}), the larger will be the epidemic threshold (Figure 10.26). By immunizing a sufficient fraction of the hubs we can increase λ above λ_c . This procedure is equivalent with altering the underlying network: by immunizing the hubs, we are fragmenting the contact network, making it more difficult for the pathogen to reach other nodes (Box 10.4).

This strategy represents a perspective change compared to the traditional immunization strategies: in a scale-free network, instead of trying to decrease the spreading rate using random immunization, we must alter the topology of the contact network, aiming to increase λ_c above the biologically determined $\lambda = \beta/\mu$.

Box 10.4 Network 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: we found that random immunization is not effective in eradicating a disease, but selective immunization, that targets the hubs, can restore a finite critical threshold, helping us to eradicate the disease. The analogy is not accidental, as the robustness and the immunization problem can be both linked to the diverging $\langle k^2 \rangle$, that measures the underlying degree heterogeneity. 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.A). Similarly, the reemergence of the epidemic threshold under hub immunization is equivalent with the finite percolation threshold characterizing a scale-free network under attack. Therefore, the attack and the immunization problems represent the two sides of the same coin.

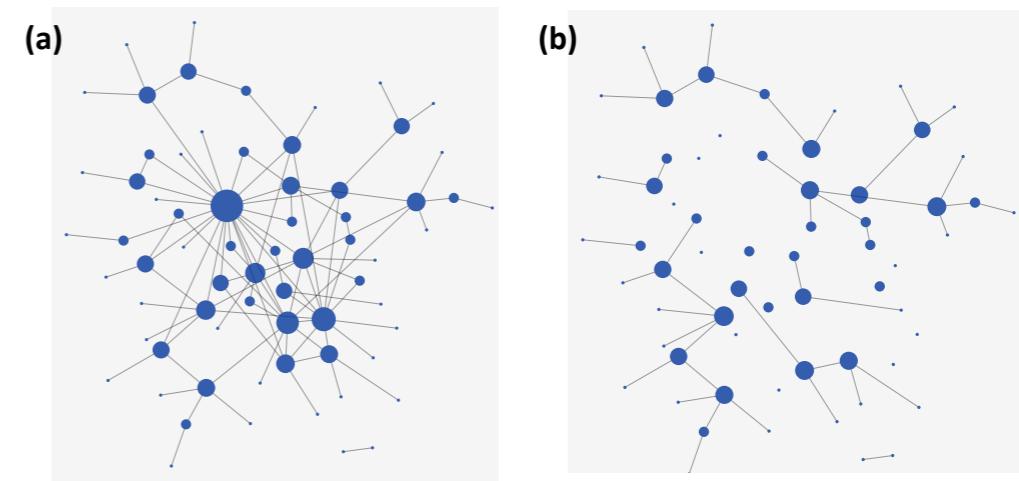


Figure 10.27 Percolation and Immunization

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

The problem with the hub-based strategy discussed above 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 of each individual in a population, nor can we accurately identify the super-spreaders during an influenza outbreak. Therefore we cannot identify the hubs to be immunized. Yet, we can still exploit our improved understanding of the network topology to design more efficient immunization strategies. To do so, we rely on the friendship paradox [65], 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 [66]:

- (i) Choose randomly a p fraction of nodes, like we do during random immunization. Call these nodes Group 0.
- (ii) Select randomly a link for each node in Group 0. We call Group I 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 the type of activity that could have resulted in the transmission of the pathogen. In the case of HIV, we need to ask them to name a sexual partner.
- (iii) Immunize the Group I 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 I is $k p_k / N \langle k \rangle$. Consequently the Group I

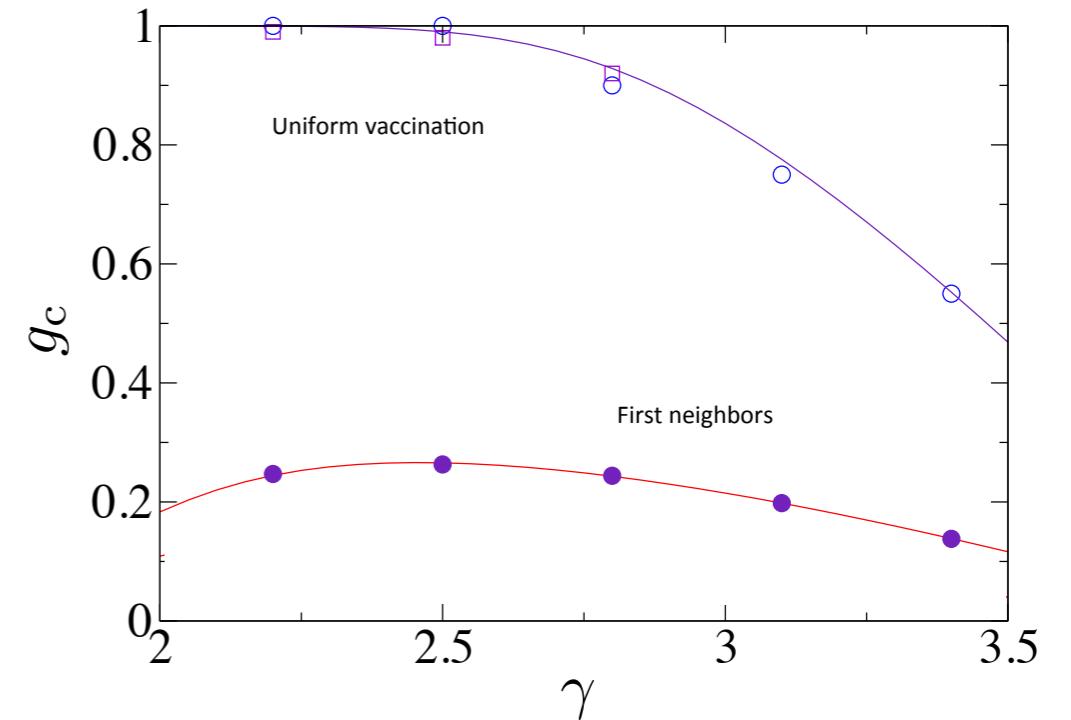


Figure 10.28 Selective Immunization of Scale-free Networks.

The plot shows the critical immunization threshold g_c in function of the degree exponent γ of the contact network on which the pathogen spreads following the SIS model. The curves correspond to two distinct immunization strategies: random immunization (top curve) and selective (acquaintance) immunization (lower curve). 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, $g_c < 1$ for random immunization even for $\gamma < 3$. Redrawn after [66].

individuals on average have higher degree than the Group 0 individuals. The implications of this bias are illustrated in Figure 10.28, which shows the critical threshold required to eradicate a pathogen for a scale-free network with degree exponent γ . The figure offers several key insights:

- (i) *Random immunization:* The top curve shows g_c for random immunization. For heterogeneous networks (small γ) we find

that $g_c \sim 1$, indicating that we must immunize all nodes to eradicate the disease. As γ approaches 3 the network develops a finite epidemic threshold and g_c drops. Hence in this case by immunizing a sufficiently high fraction of the population can eradicate the pathogen.

(ii) *Selective Immunization:* The biased strategy offers a dramatic improvement over random immunization. Indeed, now g_c is systematically under 30%. In other words by immunizing one neighbor of 30% of the nodes, we could eradicate the disease. The efficiency of this strategy depends only weakly on γ .

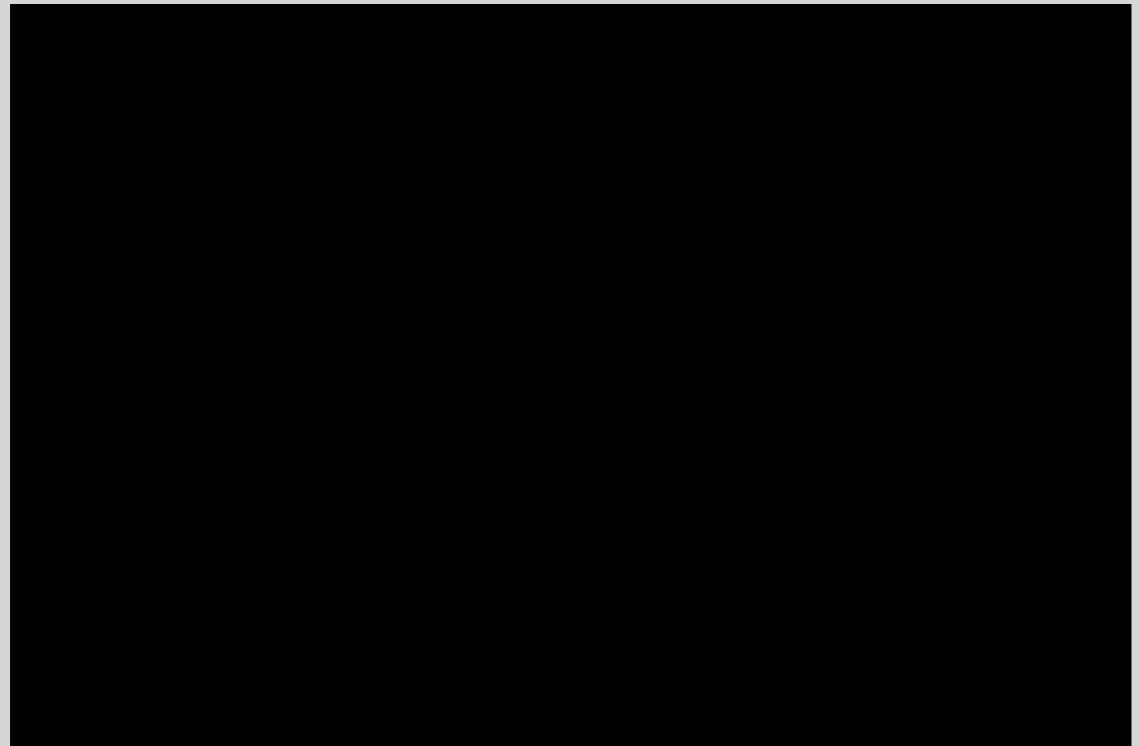
(iii) Selective immunization is more efficient than random immunization even for high γ , when hubs are less prominent.

In summary, if we have the resources to immunize everyone at risk of contacting a pathogen, we should do that - this has been the strategy of the traditional eradication campaigns (Box 10.3). When an extensive immunization is not feasible, we need to employ various immunization strategies to enhance the impact of our resources. The effectiveness of each immunization 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. Indeed, to succeed we need close to a 100% immunization, which is virtually impossible to achieve. In contrast, strategies that immunize the hubs have the highest efficiency of all possible vaccination campaigns. The difficulty with these is that they require a detailed knowledge of the node degrees, which is not available in most circumstances. Selective immunization, that immunizes the neighbors of randomly

selected nodes, can significantly enhance the effectiveness of immunization, without requiring an accurate map of the contact network. This strategy is efficient for both random and heterogeneous networks.

SECTION 7

Epidemic Prediction



Online Resources 10.4

Aaron Koblin's flight pattern, showing the real time flights across the world using data released by the Federal Aviation Administration. The global travel pattern captured by the movie is responsible for the rapid spread of viruses across continents, and the flight schedules represent the input to epidemic forecasts. While this video could easily be seen as a purely scientific illustration, it also serves as an art work. Indeed, the video is now in Media Art collection of the Museum of Modern Art (MoMA) in NY.

During much of its history humanity has been helpless when a pandemic swept a continent. The first vaccine was tested only in 1796 and the ability to systematically develop vaccines and cures against new pathogens became possible only in the late 1990s. Despite the spectacular advances in vaccine development, they remain effective only against a small number of pathogens. Consequently transmission-reducing and quarantine-based measures are the main tools of health professionals in their fight to reduce the impact of new pathogens. For the combination of vaccines, treatments and quarantine-based measures to be the most effective, we need to predict when and where the pathogen will emerge next, allowing local health officials to most effectively 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 [67] and by the 2003 SARS epidemic, which resulted in worldwide reporting guidelines that allows the collection and sharing of real time data pertaining to a pandemic [1], serving as input to modeling efforts. The 2009 HINI outbreak was the first test of this new framework, also becoming the first pandemic whose spread was predicted by modeling efforts in real time.

The discovery of a 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 it down, offering local agencies time to prepare for it?

These questions are addressed using powerful, large-scale epidemic simulators that consider as input demographic, mobility-related, and epidemiological data [68-70]. The algorithms behind these tools range from stochastic meta-population models [71-73] to agent-based computer simulations that account for the behavior and interactions of millions of individuals [74]. The goal of this section is summarize the features and the capabilities of these tools and to understand 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 [74, 75]. The first successful real time pandemic forecast based on network science relied on the Global Epidemic and Mobility (GLEAM) computational model [75] (Figure 10.29), which is a stochastic epidemic framework that uses an input high-resolution data on human demography and mobility on a worldwide scale. GLEAM employs a fully network-based computational model:

- In contrast with earlier real-space based approaches, GLEAM maps each geographic location into the nodes of a network.

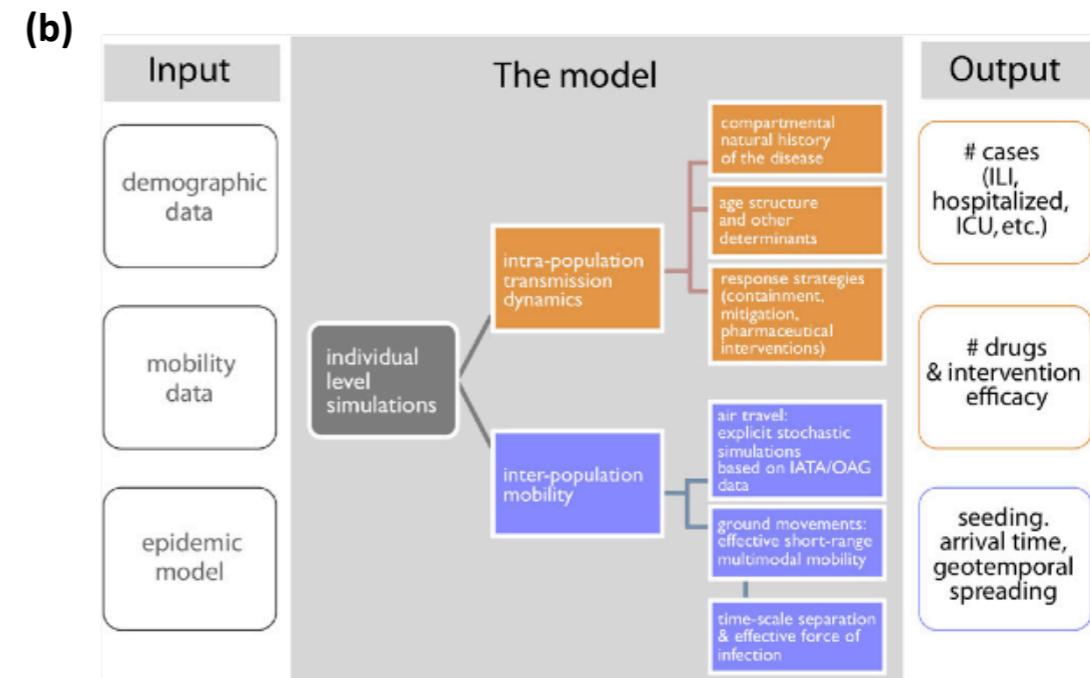
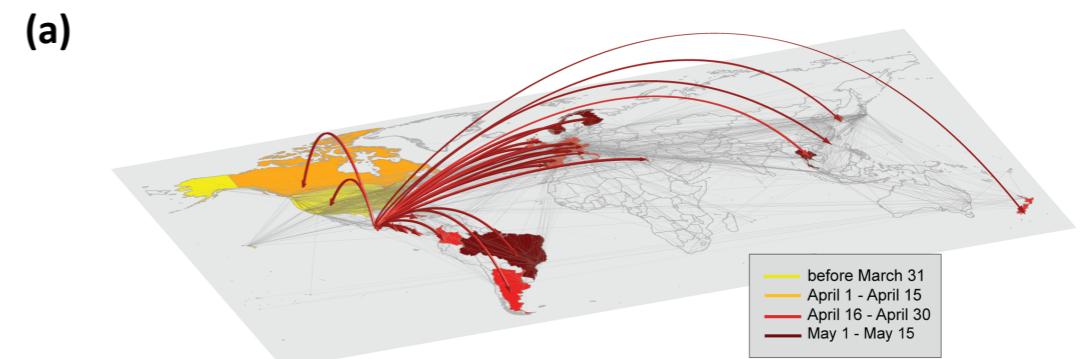


Figure 10.29 Modeling the 2009 H1N1 Pandemic.

(a) The spread of the 2009 H1N1 pandemic during the early stage of the outbreak. The arrows represent the arrival of the first infections to previously unaffected countries as infected individuals traveled from Mexico to those 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 the H1N1 virus. The left column (Input) represents the input databases capturing demographic, mobility and epidemiological information. The center column (model) describes the dynamic processes that are modeled at each time step. The right column (Output) offers examples of quantities the model can predict. After [77].

- Transport between these nodes, i.e. the links, are provided by global transportation data, like airline schedules.
- GLEAM estimates the epidemic parameters, like the transmission rate or reproduction number, using a network-based approach, relying on chronological data that captures the worldwide spread of the pandemic, rather than medical data [76].

GLEAM then implements the network-based epidemic framework encountered in the previous sections, generating a large number of potential outcomes of the pathogen's global progression for the coming months. The predictions for H1N1 were later compared with data collected from surveillance and virologic sources in 48 countries during the full course of the pandemic [75], resulting in several key findings:

- **Peak Time:** Peak time denotes the week when the largest number of individuals are infected in a particular country. Knowing the peak time can help decision makers decide the timing and the quantity of the vaccines or treatments they need to 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 10.30). In the remaining cases the difference between the real and the predicted peak was at most two weeks.
- **Early Peak:** The forecast predicted that the H1N1 epidemic will peak out in November, rather than in January or February, the typical peak time of most influenza-like viruses. This unexpected prediction turned out to be correct, confirming the model's

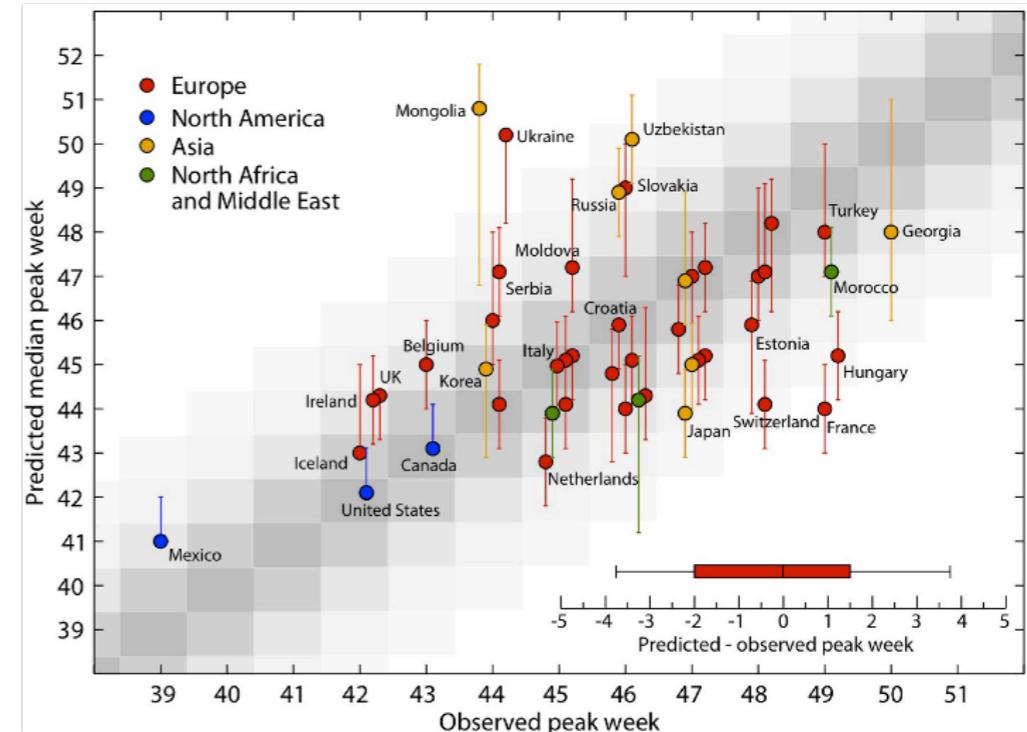


Figure 10.30 Activity Peaks

The predicted (vertical axis) and the observed (horizontal axis) activity peaks for H1N1 in several countries. The peak week denotes the week when most individuals are infected by the disease, and are shown in weeks after the beginning of the epidemic. The model predictions were obtained by analyzing 2,000 stochastic realizations of the outbreak. After [77].

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



Online Resources 10.5

A description of the GLEAM software package for epidemic prediction.

the vaccines after the November 2009 peak [78], too late to have a strong effect.

'What if' Analysis

By incorporating the time and nature of specific containment and mitigation procedures, simulations can help estimate the efficiency of specific contingency plans [68-70, 72, 79]. 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 the first instinct is to restrict travel. Yet, in a world where key resources travel by air, a travel ban can lead to economic collapse. Therefore before resorting to a travel ban, we must make sure that travel restrictions do impact the pandemic. First, we must factor in the fact 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

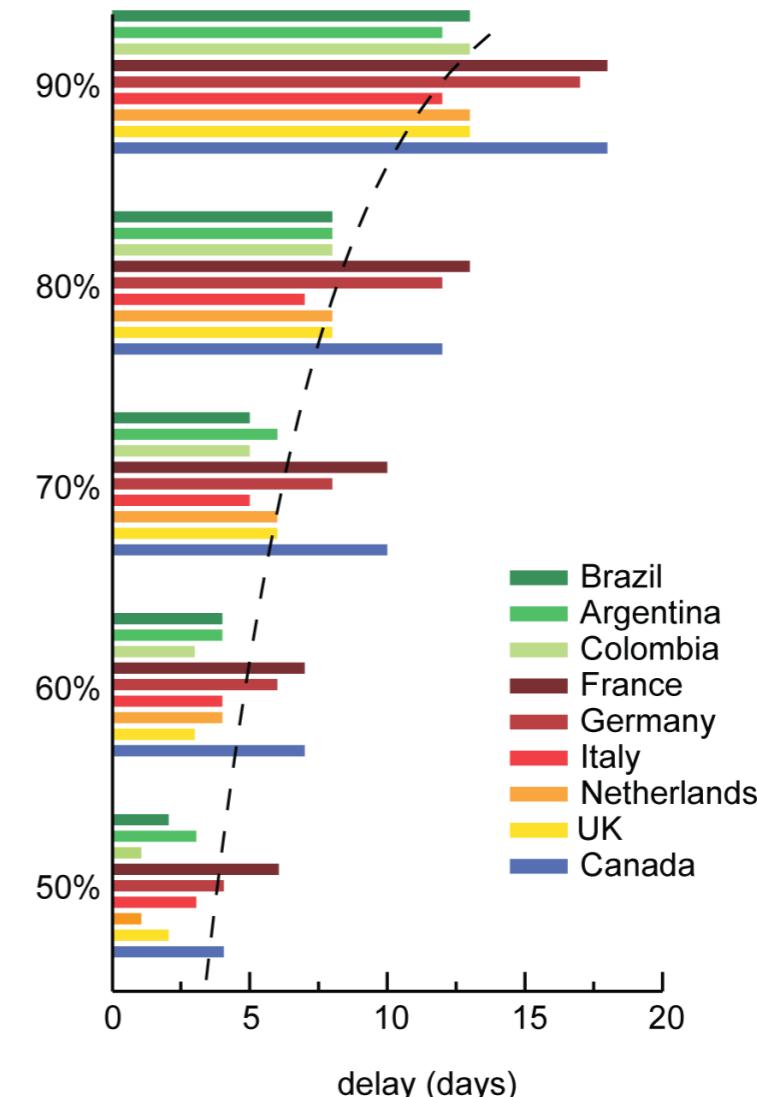


Figure 10.31 The Impact of Travel Reduction.

The predicted delay in the arrival time of the H1N1 virus from Mexico to the countries listed in the legend, 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].

individuals canceled non-necessary business and leisure activities in the infected region. The modeling indicates [75, 77] 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%, it delayed the peak time with less than 20 days (Figure 10.31). Note that travel restrictions do not decrease the number of infected individuals. They only offer local authorities more time to prepare for the pandemic. If this delay could increase local vaccination levels or help the deployment of cures, travel restrictions could be effective.

- **Antiviral treatment:** During the 2009 HINI pandemic Canada, Germany, Hong Kong SAR, Japan, the UK, and the USA distributed antiviral drugs to mitigate the impact of the disease [80]. 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 [81]. The simulations indicate that peak times would have been delayed with about 3 to 4 weeks, offering valuable time to immunize a larger fraction of the population before the pandemic reached its peak.

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 10.9), following a diffusive process described by simple reaction-diffusion models [82, 83]. As the next infection always emerged in the geographic proximity of the previous infections, there was a direct correspondence between the time of the outbreak and the physical distance from the origin of the outbreak.

Box 10.5 A Night at the Movies

For a fictionalized but plausible depiction of a major pandemic, watch *Contagion*, a 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. *Contagion* follows the desperate attempts of public health officials to stop a virus and panic from sweeping the globe, depicting the impact of both biological and social contagion. Similarly, 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 are forced to take to contain the spread of a deadly pathogen (Figure 10.32).



Figure 10.32 Outbreak: Fiction and Truth. The theatrical release posters of *Contagion* and *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 Poughkeepsie, NY, a village an hour 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 [84]. 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 and $p_{ij} > 0$ only if there is direct travel from i to j .

Given the multiple paths 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

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

If p_{ij} is small, implying that only a small fraction of travelers 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)

Box 10.6 The Speed of a Pandemic

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 10.29) 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 [84].

accounts for the fact that effective distances are additive, whereas probabilities along multi-step paths are multiplicative.

As Figure 10.33 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 strong contrast with the complex spreading pattern observed 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 10.34). We can therefore use the effective distance to determine the speed of a pathogen (Box 10.6).

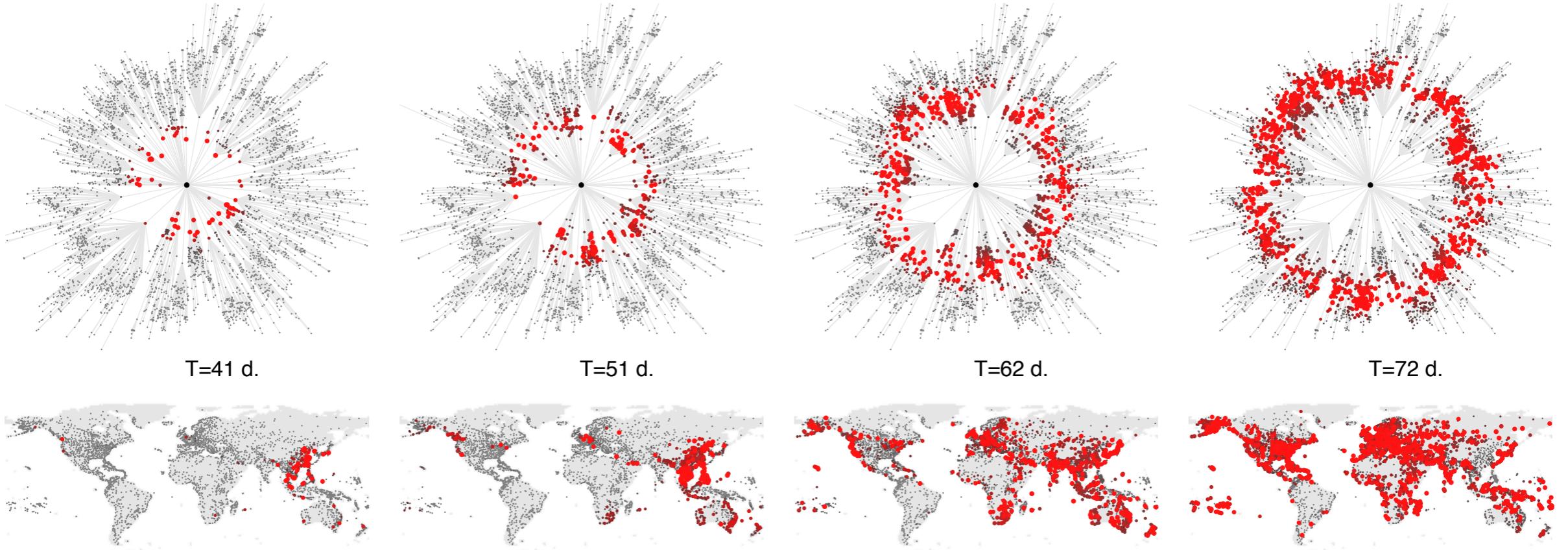


Figure 10.33 Effective Distance

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. After [84].

One of the most surprising aspect of the epidemic forecast literature is that the predictions of different models are rather similar, despite the fact that they use different mobility data (airline schedules [85] or dollar bill movement [86]) and rather different assumptions about the epidemic parameters (recovery rate, transmission rate, etc). The effective distance helps us understand why their predictions converge. Indeed, we can write the arrival time of a pathogen to location a as [84]

$$T_a = \underbrace{d_{eff}(P)/V_{eff}}_{eff, speed}(\beta, R_0, \gamma, \varepsilon) \quad (10.32)$$

Therefore the arrival time is the ratio of the effective distance d_{eff} and an effective speed V_{eff} . The effective speed is determined only by the epidemiological parameters, whereas the effective distance d_{eff} depends only on the topology of the mobility network encoded in p_{ij} . When confronted with a new outbreak, the disease-specific epidemiological parameters are unknown in the beginning. However, (10.32) predicts that the *relative arrival times are independent of the*

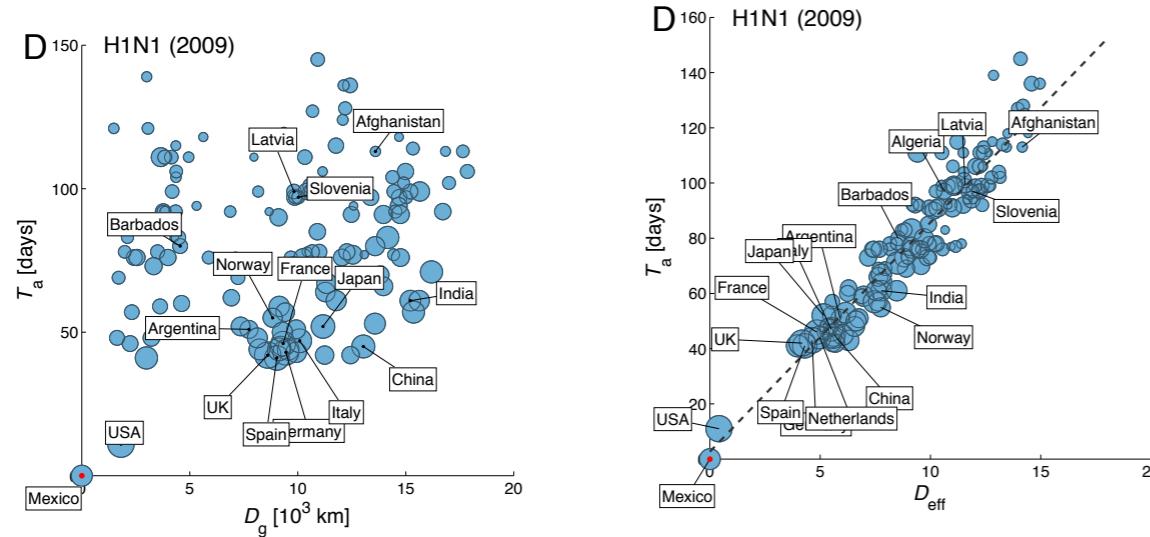


Figure 10.34 Effective Distance and Arrival Time.

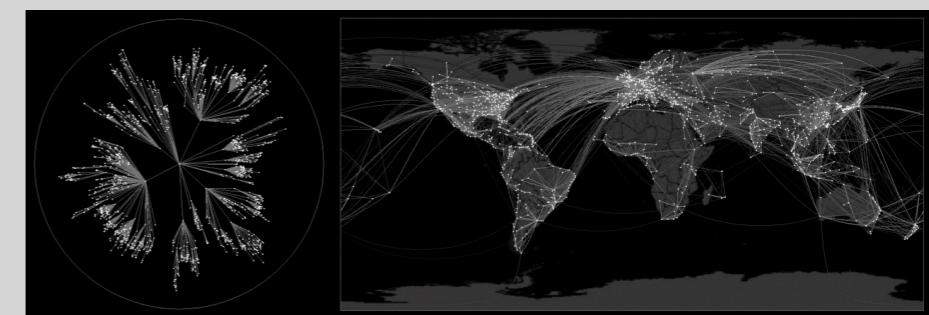
(a) Arrival times vs. geographic distance from its source (Mexico) for the 2009 H1N1 pandemic. Each circle represents one of the 140 affected countries and the symbol size indicates the total traffic in each country. Arrival times are the date of the first confirmed case in a given country after the beginning of the outbreak on March 17, 2009. In this representation the arrival time and the geographic distance are largely independent of each other ($R^2 = 0.0394$).

(b) Epidemic arrival time T_a vs. effective distance D_{eff} for H1N1, demonstrating the strong correlations between the effective distance (10.31) and the arrival time. After [84].

epidemiological parameters. For example, for an outbreak that starts at node i , the ratio of the arrival times to nodes j and l is

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

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 pattern around the world is unique and model-independent, the predictions of different models must converge, independent of the choice of the epidemiological



Online Resource 10.6

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

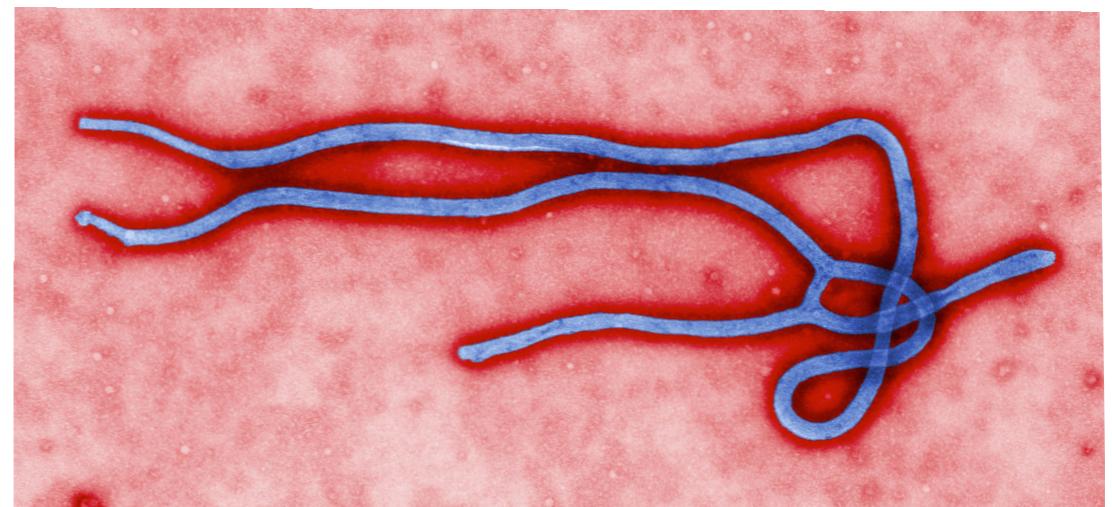


Figure 10.35 The Deadliest Outbreak.

With a fatality rate in the vicinity of 80%, the Ebola virus, seen under a microscope, 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 10.7 Identifying the Source of a Pandemic

Finding 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 [87] inspired a burst of research on the subject [88-95].

The difficulty in finding the source is rooted in the stochastic nature of the infection process. Indeed, 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 [87] (Figure 10.36).
- If we also have the time of infection of each node, we can reconstruct the dynamics of the epidemic, significantly enhancing our ability to detect the source.
- If we aim to detect the source of an infection 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 network generated by the Barabási-Albert model, monitoring the state of 18% of the highest degree nodes can offer a 90% success rate in detecting the source. In contrast, we need to monitor 41% of the nodes if we select them randomly to achieve the same level of accuracy [89].
- In the effective distance representation (Figure 10.33) the infection follows a circular pattern only if we use the right outbreak location. Otherwise the observed pattern is asymmetric. Therefore, to accurately detect the source, we need to find the location (node) from which the outbreak pattern shows the highest radial symmetry [84].



Figure 10.36 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 underlying mobility network.

parameters.

In summary, joint advances in data collection and epidemic modeling 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 the available treatments and vaccines.

The recent success of epidemic forecast is not due to the improved understanding of the underlying biology of the disease, but can be attributed to the lucky situation that when it comes to epidemic spreading patterns, 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 and the network-based epidemic framework described in this chapter allows us to turn this travel data into accurate predictions.

SECTION 8

Summary

The purpose of most networks is to facilitate transfer along its 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 develop a general theory of 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 nature and the dynamics of the spreading process, leading for example to different predictions for spreading on random and scale-free networks. This finding has laid the ground for a wider class of problems: the need to systematically understand the impact networks have on many dynamical processes [96], an increasingly active chapter of network science [97, 98].

Box 10.8 At a Glance: Network Epidemics

$$\text{Infection Rate: } \beta; \text{ Recovery Rate: } \mu; \text{ Spreading Rate: } \lambda \equiv \frac{\beta}{\mu}$$

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

$$\text{SI Model: } i(t) = \frac{i_0 \exp(\beta \langle k \rangle t)}{1 - i_0 + i_0 \exp(\beta \langle k \rangle t)}$$

$$\text{SIS Model: } i(t) = (1 - \frac{\mu}{\beta \langle k \rangle}) \frac{C e^{(\beta \langle k \rangle - \mu)t}}{1 + C e^{(\beta \langle k \rangle - \mu)t}}$$

Characteristic Time:

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

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

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

Epidemic Threshold:

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

$$\text{SIR: } \lambda_0 = \frac{1}{\frac{\langle k^2 \rangle}{\langle k \rangle} - 1}$$

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

Modeling the spread of pathogens also represents an important practical application of network science. The recent advances in this area were rather spectacular, giving birth to accurate epidemic predictions, 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 epidemic processes. 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 decoupling of the biological parameters and the network contributes to the accuracy of the observed predictive power. Consequently, an accurate forecast requires primarily an accurate knowledge of the mobility network.

The developed analytical framework has lead to 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 lasting theoretical and practical interest.

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 from a random network, the same strategy is suboptimal in a scale-free network. As most contact networks are heterogenous, this is a rather depressing conclusion. Yet, the introduced network-based selective immunization strategies can restore the epidemic

Box 10.9

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 systematically incorporates the properties of the underlying 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, the modeling package that offers real-time predictions for the spread of a pathogen, successfully tested during the H1N1 outbreak (Sect. 10.7).

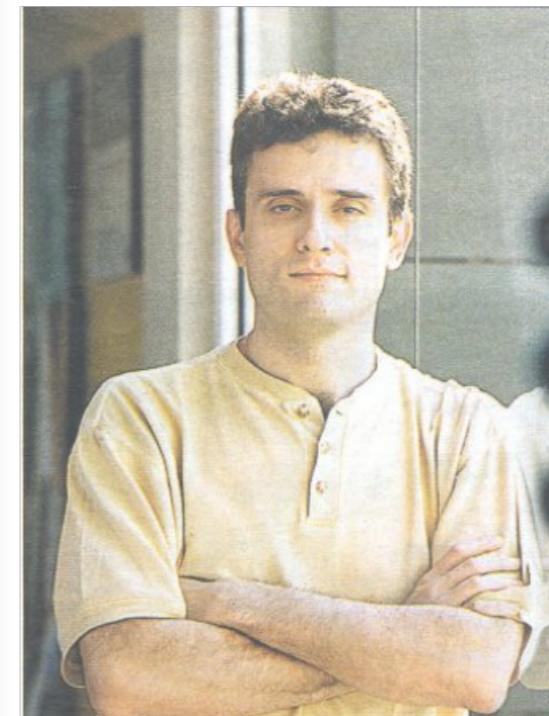


Figure 10.37 Romualdo Pastor-Satorras and Alessandro Vespignani. Both physicists by training, Pastor-Satorras was a postdoctoral associate with Vespignani at ICTP in Trieste when they discovered the impact of the scale-free property on the epidemic threshold. Subsequently both researchers had major contributions to network science, from the discovery of degree correlations (Chapter 7) to our understanding of weighted networks.

threshold and suppress the prevalence of a pathogen. They work by systematically altering the topology of the network on which a pathogen spreads.

The results of this chapter allow us to formulate our next law:

In scale-free networks the epidemic threshold and the characteristic time vanish. Therefore even weakly virulent viruses can spread and their spread is virtually instantaneous.

Evidence for this law is provided in Section 10.3. To recap in the context of the criteria A-C:

- A. *Quantitative Formulation*: Equation (10.19) predicts that the epidemic threshold depends inversely on $\langle k^2 \rangle$, hence it vanishes for networks with $\gamma < 3$ in the $N \rightarrow \infty$ limit. Similarly, the characteristic time also depends on $\langle k^2 \rangle$, vanishing for large variance.
- B. *Universality*: Section 10.4 offers evidence that many contact networks have a large $\langle k^2 \rangle$, which lowers λ_c and τ .
- C. *Non-random character*: Random networks have a finite $\langle k^2 \rangle$, hence for these the epidemic threshold and the characteristic time is always finite.

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 [97]. Consider a susceptible individual in contact with an infected individual, which means that the susceptible individual becomes infected with probability βdt during the time interval dt . Therefore 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 be responsible for infecting it. Therefore the probability that it avoids infection is

$(1 - \beta dt)^{k_i}$. Finally the total probability that node i becomes infected in a unit time 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

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

In a random network the typical node has approximately $\langle k \rangle$ neighbors. Replacing k_i with $\langle k \rangle$ in (10.33) we arrive to the first term of the continuum equation (10.3), capturing the rate of growth in the number of infected individuals. If we do not replace k_i with $\langle k \rangle$, we obtain to the first term of (10.13), capturing the diffusion of a pathogen in a heterogeneous network.

Epidemic Threshold and Network Topology

A key result of Section 10.3 connects the network topology to the epidemic threshold λ_c , a result derived using the continuum theory. We can arrive at the same result using a mechanistic argument, that helps us better understand the connection between the epidemic threshold and the network topology.

Consider a pathogen that is transmitted at a rate β . Therefore in a unit time an infected node with degree k will infect βk neighbors. If each infected node recovers at rate μ , 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 will gradually die 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 $\beta \langle k \rangle / \mu = 1$ allows us to calculate the epidemic threshold. Using $\lambda \equiv \beta / \mu$ we obtain $\lambda_c = 1 / \langle k \rangle$, or 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 (β and μ) 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, there are considerable number of nodes with sufficiently large k so that for them $\beta k / \mu > 1$. If such a high degree node is infected, even if the spreading rate λ 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 degree heterogeneity $\langle k^2 \rangle$.

Advanced Topics 10.B: Analytical Solution of the SI, SIS and SIR Models

The goal of this section is to solve the SI, SIS and SIR models on a network, deriving the results summarized in Table 10.3, namely the characteristic spreading time τ and the epidemic threshold λ_c .

The Density Function

The density function $\Theta_k(t)$ provides the fraction of nodes in the vicinity of a node with degree k . As discussed in Section 10.3, to calculate $i_k(t)$, we must first determine $\Theta_k(t)$. 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 simply 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

$$\Theta_k(t) = \frac{\sum_{k'} (k' - 1)p_{k'} i_{k'}(t)}{\langle k \rangle} \equiv \Theta(t). \quad (10.34)$$

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

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

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

SI Model

Using (10.11) we obtain

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

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

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

This has the solution

$$\Theta(t) = Ce^{t/\tau}, \quad (10.38)$$

where

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

Using the initial conditions $\Theta(t = 0) = C = i_0 \frac{\langle k \rangle - 1}{\langle k \rangle}$ we obtain the time dependent $\Theta(t)$ as

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

We insert this into (10.11) to arrive at (10.13).

SIR Model

In the SIR model the density of infected nodes is given by:

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

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

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

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

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

The solution of this equation is

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

where the characteristic time for the SIR model is

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

A global outbreak is possible only if $\tau > 0$, so that the number of infected nodes grows exponentially with time. This yields the critical spreading rate

$$\lambda \equiv \frac{\beta}{\mu} > \frac{\langle k \rangle}{\langle k^2 \rangle - \langle k \rangle}, \quad (10.46)$$

providing the epidemic threshold for the SIR model as (Table 10.4)

$$\lambda_c = \frac{\langle k \rangle}{\langle k^2 \rangle - \langle k \rangle}. \quad (10.47)$$

SIS Model

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

$$\frac{di_k}{dt} = \beta(1 - i_k)k\Theta_{SIS}(t) - \mu i_k. \quad (10.48)$$

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 (10.34) definition. However, in the SIS model the 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

$$\Theta_k^{SIS}(t) = \frac{\sum_{k'} k' p_{k'} i_{k'}}{\langle k \rangle} \equiv \Theta_{SIS}(t). \quad (10.49)$$

Again keeping only the first order terms we obtain

$$\frac{di_k}{dt} = \beta k \Theta^{SIS}(t) - \mu i_k. \quad (10.50)$$

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

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

This again has the solution

$$\Theta(t) = C e^{t/\tau}, \quad (10.52)$$

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 critical spreading rate

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

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 for the SIS and SIR Models

The purpose of this section is to 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 $k \geq k_{\min}$ and $c \approx (\gamma - 1)/k_{\min}^{-\gamma+1}$. In Section 10.10 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

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

Hub immunization means that we immunize all nodes whose degree is larger than k_0 . For epidemic spreading this is equivalent with removing these 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 already addressed in the Advanced Topics 8.D, where we studied the robustness of a network under attack. We have seen that hub removal has two effects:

- (i) The maximum degree of the network changes to k_0 .
- (ii) 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

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

According to equation (8.32) and (8.33) this yields

$$\langle k' \rangle = (1 - \tilde{f}) \langle k \rangle,$$

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

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 (8.38), 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

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

For the SIR model a similar calculation yields

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

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

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

Advanced Topics 10.D: The SIR Model and Bond Percolation

The SIR model is a dynamical model that captures the time dependent spread of an infection in a network. Yet, it can be mapped into a static bond percolation problem [99-102]. This mapping offers analytical tools that help us predict the model's behavior.

Consider an epidemic process on a network, so that each infected node transmits a pathogen to each of its neighbors with rate β , and recovers after a recovery time $\tau = 1/\mu$. The infection is typically viewed as a Poisson process, consisting of series of random contacts with average interevent time $\beta\tau$. Therefore, the probability that a neighboring node is not infected in a unit time is $e^{-\beta\tau}$.

This process is equivalent with bond percolation on the same network, where each directed link is occupied with probability $p_b = 1 - e^{-\beta\tau}$ (Figure 10.38). If β and τ are the same for each node, the network can be considered undirected. Although this mapping

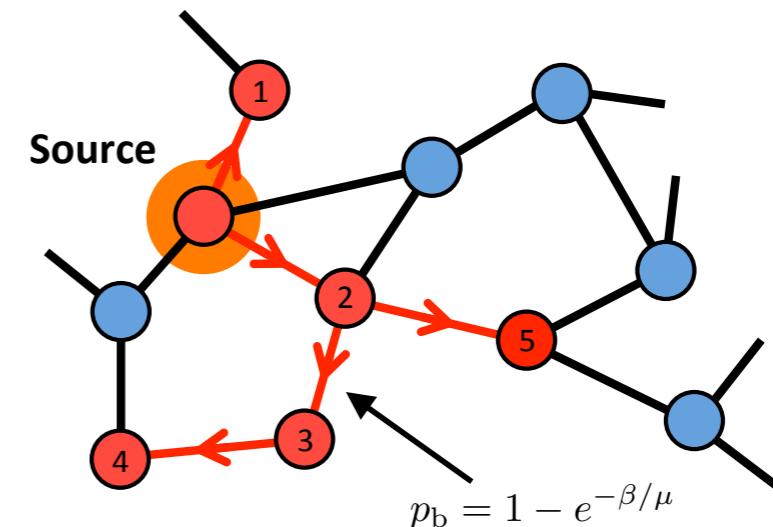


Figure 10.38 Mapping Epidemics into Percolation.

Consider the contact network on which the epidemic spreads. To map the spreading process into percolation, we leave in place each link with probability, $p_b = 1 - e^{-\beta/\mu}$, a probability determined by the biological characteristics of the pathogen. Therefore links are removed with probability $e^{-\beta/\mu}$. The cluster size distribution of the remaining network can be mapped exactly into the outbreak size. For large β/μ we will likely have a giant component, indicating that we could face a global outbreak. β/μ corresponds to a virus that has difficulty spreading and we end up with numerous small clusters, indicating that the pathogen will likely die out.

loses the temporal dynamics of the epidemic process, it has several advantages:

- The total fraction of infected individuals in the endemic state maps into the size of the giant component of the percolation problem.
- The probability that a single disease event dies out before reaching the endemic state equals the fraction of finite components in percolation.

- We can determine the epidemic threshold by exploring the known properties of bond percolation. For example, to find the threshold for bond percolation, we consider the average number of links outgoing from a node that can be reached by a link. This way we retrace the course of the epidemic: if an infected individual infects on average at least one other individual, then the epidemic can reach an endemic state. Since a node can be reached by one of its k links, its probability to be reached is $kp_k/N\langle k \rangle$. The probability of each of its $k - 1$ outgoing links infecting 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

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

As an endemic state can be reached only if $\langle R_i \rangle > 1$, we obtain the condition for the epidemic as [103, 104]

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

Equation (10.62) agrees with the result (10.46) derived earlier from the dynamical models: as scale-free networks with $\gamma \leq 3$ we have a divergent second moment, hence such networks undergo a percolation transition only at $p_b \rightarrow 0$ [103, 104]. That is, a virus can spread on this network regardless of how small is its infection probability β or how small is the recovery time τ .

SECTION 13

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. Superspreading 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] 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.
- [12] 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.
- [13] 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.
- [14] W. Chamberlain. *A View from Above*. Villard Books, New York, 1991.

- [15] R. Shilts. *And the Band Played On*. St. Martin's Press, New York, 2000.
- [16] 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.
- [17] M. C. González, C. A. Hidalgo, and A.-L. Barabási. Understanding individual human mobility patterns. *Nature* 453: 779-782, 2008.
- [18] C. Song, Z. Qu, N. Blumm, and A.-L. Barabási. Limits of Predictability in Human Mobility. *Science* 327: 1018-1021, 2010.
- [19] F. Simini, M. González, A. Maritan, and A.-L. Barabási. A universal model for mobility and migration patterns. *Nature* 484: 96-100, 2012.
- [20] D. Brockmann, L. Hufnagel, and T. Geisel. The scaling laws of human travel. *Nature* 439: 462–465, 2006.
- [21] 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.
- [22] L. Hufnagel, D. Brockmann, and T. Geisel. Forecast and control of epidemics in a globalized world. *PNAS* 101: 15124, 2004.
- [23] R. Guimerà, S. Mossa, A. Turtschi, and L.A. N. Amaral. The worldwide air transportation network: Anomalous centrality, community structure, and cities' global roles. *PNAS* 102: 7794, 2005.
- [24] C. Cattuto, W. Van den Broeck, A. Barrat, V. Colizza, J.-F. Pinton, et al. Dynamics of Person-to-Person Interactions from Distributed RFID Sensor Networks. *PLoS ONE* 5: e11596, 2010.
- [25] 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. *J. Theor. Biol.* 271:166-180, 2011. <http://arxiv.org/abs/1006.1260>
- [26] K. Zhao, J. Stehle, G. Bianconi, and A. Barrat. Social network dynamics of face-to-face interactions. *Phys. Rev. E* 83: 056109, 2011.
- [27] J. Stehlé, N. Voirin, A. Barrat, C. Cattuto, L. Isella, J.-F. Pinton, M. Quaggiotto, 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(8):e23176, 2011.
- [28] A.-L. Barabási. The origin of bursts and heavy tails in human dynamics. *Nature*, 435:207-11, 2005.
- [29] V. Sekara, and S. Lehmann. Application of network properties and signal strength to identify face-to-face links in an electronic dataset. *arXiv:1401.5836*, 2014.
- [30] S. Eubank, H. Guclu, VSA. Kumar, MV Marathe, A Srinivasan, Z. Toroczkai, and N. Wang. Modelling disease outbreaks in realistic urban social networks. *Nature* 429: 180-184, 2004.
- [31] H. Ebel, L-I. Mielsch, and S. Bornholdt. Scale-free topology of e-mail networks. *Phys. Rev. E* 66: 035103, 2002.

- [32] M. Morris, and M. Kretzschmar. Concurrent partnerships and transmission dynamics in networks. *Soc Netw*, 17:299–318, 1995.
- [33] N. Masuda and P. Holme. Predicting and controlling infectious diseases epidemics using temporal networks. *F1000Prime Rep* 5:6, 2013.
- [34] P. Holme, and J. Saramäki. Temporal networks. *Physics Reports* 519: 97–125, 2012.
- [35] 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. *Phys Rev E* 83: 025102(R), 2011.
- [36] A. Vazquez, B. Rácz, A. Lukács, and A.-L. Barabási. Impact of non-Poissonian activity patterns on spreading processes. *Phys Rev Lett*, 98:158702, 2007.
- [37] A. Vázquez, J. G. Oliveira, Z. Dezsö, K.-I. Goh, I. Kondor, and A.-L. Barabási. Modeling bursts and heavy tails in human dynamics. *Phys Rev E* 73: 036127, 2006.
- [38] A.-L. Barabási. The architecture of complexity. *IEEE Control Systems Magazine* 27:33, 2007.
- [39] A.V. Goltsev, and S. N. Dorogovtsev, and J. F. F. Mendes. Percolation on correlated networks. *Phys Rev E*.78: 051105, 2008.
- [40] 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 - Condensed Matter and Complex Systems*, 76: 643, 2010.
- [41] 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.
- [42] Y. Moreno, J. B. Gómez, and A.F. Pacheco. Epidemic incidence in correlated complex networks. *Phys. Rev. E* 68, 035103, 2003.
- [43] 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.
- [44] M. S. Granovetter. The strength of weak ties. *American Journal of Sociology* 78: 1360–1379, 1973.
- [45] A. Galstyan, and P. Cohen. Cascading dynamics in modular networks. *Physical Review E* 75: 036109, 2007.
- [46] J. P. Gleeson. Cascades on correlated and modular random networks. *Physical Review E* 77: 046117, 2008.
- [47] 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.
- [48] L Weng, F Menczer and Y-Y Ahn. Virality Prediction and Community Structure in Social Networks. *Scientific Reports* 3: 2522, 2013.
- [49] 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.

- [50] J. Leskovec, L. Adamic, and B. Huberman. The dynamics of viral marketing. *ACM Trans. Web* 1, 2007.
- [51] L. Weng, A. Flammini, A. Vespignani, and F. Menczer. Competition among memes in a world with limited attention. *Sci. Rep.* 2: 335, 2012.
- [52] J. Berger, and K. L. Milkman. What makes online content viral? *Journal of Marketing Research* 49: 192–205, 2009.
- [53] 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.
- [54] G. Szabó and, B.A. Huberman. Predicting the popularity of online content. *Communications of the ACM* 53: 80–88, 2010.
- [55] 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.
- [56] D. Centola. The spread of behavior in an online social network experiment. *Science* 329: 1194–1197, 2010.
- [57] 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.
- [58] M. Granovetter. Threshold Models of Collective Behavior. *American Journal of Sociology* 83:1420–1443, 1978.
- [59] 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.
- [60] N.A. Christakis, J. H. Fowler. The collective dynamics of smoking in a large social network. *New England Journal of Medicine* 358: 2249-2258, 2008.
- [61] R Pastor-Satorras, and A Vespignani. *Evolution and structure of the Internet: A statistical physics approach*. Cambridge University Press, Cambridge, 2007.
- [62] F. Fenner et al. *Smallpox and its Eradication*. WHO, Geneva, 1988. <http://www.who.int/features/2010/smallpox/en/>
- [63] Z. Dezső and A-L. Barabási. Halting viruses in scale-free networks. *Phys. Rev. E* 65: 055103, 2002.
- [64] R. Pastor-Satorras and A. Vespignani. Immunization of complex networks. *Phys. Rev. E* 65: 036104, 2002.
- [65] S. L. Feld. Why Your Friends Have More Friends Than You Do. *Am. J. Soc.* 96: 1464, 1991.
- [66] R. Cohen, I. S. Havlin, and D. ben-Avraham. Efficient Immunization Strategies for Computer Networks and Populations. *Phys. Rev. Lett.* 91: 247901, 2003.
- [67] L.A. Rvachev, and I. M. Longini Jr. A mathematical model for the global spread of influenza. *Mathematical Biosciences* 75: 3-22, 1985.

- [68] A. Flahault, E. Vergu, L. Coudeville, and R. Grais. Strategies for containing a global influenza pandemic. *Vaccine* 24:6751-6755, 2006.
- [69] 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.
- [70] I.M. Longini Jr, A. Nizam A, S X, K. Ungchusak, W. Hanshaoworakul, D. Cummings, and M. Halloran. Containing pandemic influenza at the source. *Science* 309:1083-1087, 2005.
- [71] 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.
- [72] 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.
- [73] 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.
- [74] I. M. Hall, R. Gani, H. E. Hughes, and S. Leach. Real-time epidemic forecasting for pandemic influenza. *Epidemiol Infect* 135:372-385, 2007.
- [75] 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.
- [76] 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.
- [77] P. Bajardi, C. Poletto, J. J. Ramasco, M. Tizzoni, V. Colizza, et al. Human Mobility Networks, Travel Restrictions, and the Global Spread of 2009 HINI Pandemic. *PLoS ONE* 6:e16591, 2011.
- [78] 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/ HINI influenza in the Northern Hemisphere. *EHT Journal* 2:e11, 2009.
- [79] 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. *Proc Natl Acad Sci USA* 105:4639-44, 2008.
- [80] G. M. Leung, A. Nicoll. Reflections on Pandemic (HINI) 2009 and the international response. *PLoS Med* 7:e1000346, 2010.
- [81] A. C. Singer, B. M. Howard, A. C. Johnson, C. J. Knowles, S. Jackman, C. Accinelli, A. B. Caracciolo, I. Bernard, S. Bird, T. Boucard, A. Boxall, J. V. Brian, E. Cartmell, C. Chubb, J. Churchley, S. Costigan, M. Crane, M. J. Dempsey, B. Dorrington, B. Ellor, J. Fick, J. Holmes, T. Hutchinson, F. Karcher, S. L. Kelleher, P. Marsden, G. Noone, M. A.

- Nunn, J. Oxford, T. Rachwal, et al. Meeting report: risk assessment of Tamiflu use under pandemic conditions. *Environ Health Perspect* 116:1563-1567, 2008.
- [82] R. Fisher. The wave of advance of advantageous genes. *Ann. Eugen.* 7: 355–369, 1937.
- [83] J.V. Noble. Geographic and temporal development of plagues. *Nature* 250, 726–729, 1974.
- [84] D. Brockmann and D. Helbing. The Hidden Geometry of Complex, Network-Driven Contagion Phenomena. *Science* 342: 1337-1342, 2014.
- [85] 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.
- [86] D. Brockmann, L. Hufnagel, and T. Geisel. The scaling laws of human travel. *Nature* 439: 462-5, 2006.
- [87] 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.
- [88] A.Y. Lokhov, M. Mezard, H. Ohta, L. Zdeborová. Inferring the origin of an epidemic with dynamic message-passing algorithm. *arXiv:1303.5315v1*, 2013.
- [89] P. C. Pinto, P. Thiran, M. Vetterli. Locating the Source of Diffusion in Large-Scale Networks. *Phys. Rev. Lett.* 109: 068702, 2012.
- [90] C. H. Comin and L. da Fontoura Costa. *Phys. Rev. E* 84: 056105, 2011.
- [91] D. Shah and T. Zaman. *IEEE Trans. Inform. Theory* 57, 5163, 2011.
- [92] K. Zhu and L. Ying. Information source detection in the sir model: A sample path based approach. *arXiv:1206.5421*, 2012.
- [93] B.A. Prakash, J. Vreeken, and C. Faloutsos, in ICDM'12; Proceedings of the IEEE International Conference on Data Mining, 2012.
- [94] V. Fioriti and M. Chinnici. Predicting the sources of an outbreak with a spectral technique. *arXiv:1211.2333 [math-ph]*, 2012.
- [95] W. Dong, W. Zhang and C.W.Tan. Rooting out the rumor culprit from suspects. *arXiv:1301.6312 [cs.SI]*, 2013.
- [96] B. Barzel, and A.-L. Barabási. Universality in network dynamics. *Nature Physics* 9: 673, 2013.
- [97] A. Barrat, M. Barthélémy and A. Vespignani. *Dynamical Processes on Complex Networks*. Cambridge University Press, 2012.
- [98] S. N. Dorogovtsev, A.V. Goltsev, J. F. F. Mendes. Critical phenomena in complex networks. *Reviews of Modern Physics* 80, 1275, 2008.
- [99] R. Cohen and S. Havlin. *Complex Networks - Structure, Robustness and Function*. Cambridge University Press, 2010.

- [100] P. Grassberger. On the critical behavior of the general epidemic process and dynamical percolation. *Mathematical Biosciences* 63: 157, 1983.
- [101] M. E. J. Newman. The spread of epidemic disease on networks. *Physical Review E* 66: 016128, 2002.
- [102] C. P. Warren, L. M. Sander, and I. M. Sokolov. Firewalls, disorder, and percolation in networks. *Mathematical Biosciences* 180: 293, 2002.
- [103] R. Cohen, K. Erez, D. ben-Avraham, and S. Havlin. Resilience of the Internet to random breakdown. *Physical Review Letters* 85: 4626–4628, 2000.
- [104] 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.
- [105] 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.
- [106] R. Durrett. Some features of the spread of epidemics and information on a random graph. *PNAS* 107:4491-4498, 2010.
- [107] 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.
- [108] C. Castellano, and R. Pastor-Satorras. Thresholds for epidemic spreading in networks. *Physical Review Letters* , 2010.
- [109] B.N. Waber, D. Olgun, T. Kim, and A. Pentland. Understanding Organizational Behavior with Wearable Sensing Technology. *Academy of Management Annual Meeting*. Anaheim, CA. August, 2008.
- [110] 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.
- [111] M. Salathé, M. Kazandjieva, J.W. Leeb, P. Levis, M.W. Feldman, and J.H. Jones. A high-resolution human contact network for infectious disease transmission. 107: 22020–22025, 2010.
- [112] A. Vespignani. Modelling dynamical processes in complex socio-technical systems. *Nature Physics*, 8:32-39, 2012