V5 Epidemics on networks

Epidemic models attempt to capture the dynamics in the spreading of a disease (or of an idea, a computer virus, or the adoption of a product).

Central questions the epidemic models try to answer are:

- How do contagions (*dt. ansteckende Krankheiten*) spread in populations?
- Will a disease become an epidemic?
- Who are the best people to vaccinate? (What do you think?)
- Will a given YouTube video go viral?
- What individuals should we market for maximizing product penetration?

http://www.lsi.upc.edu/~CSN/slides/11epidemic.pdf

Different types of disease epidemics

Infectious diseases spread over networks of contacts between individuals.

Airbourne diseases like influenza or tuberculosis are communicated when 2 people **breathe** the **air** in the same room.

Contagious diseases and parasites can be communicated when people **touch**.

HIV and other sexually transmitted diseases are communicated when people have sex.

The **patterns** of such **contacts** can be represented as (social) **networks**.

Classic epidemic models = „fully mixed“

Before we will discuss the modelling of epidemics in social networks,
we will introduce some **classic mathematical models** of epidemics.

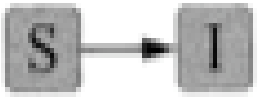
Mathematical modeling of epidemics has started much earlier than studying network topologies!

The traditional approaches make use of
- a **fully mixed approximation** („**mean field**“ in the physics world)

where every individual has an equal chance per unit time of coming into contact with every other.

The SI model (susceptible / infected)

In the simplest mathematical representation of an epidemic, there are just 2 states, susceptible and infected.



An individual in the **susceptible** state does not have the disease yet but could catch it once he/she gets in contact with an infected person.

An **infected** individual has the disease and can potentially pass it on to other susceptible persons once they get into contact.

The SI model

Let us consider a disease spreading through a population of individuals.

S(t) : average (or expected) number of **susceptible individuals** at time *t*

X(t) : average (or expected) number of **infected people** at time *t*.

In the following, we drop the explicit time-dependence of *S(t)* and *X(t)* .

Assume that each individual has, on average, β **contacts** with randomly chosen other people **per unit time.**

Note that the disease is only transmitted when an infected person has contact with a susceptible person.

The SI model

Let the total population consist of n people, $n = S + X$

Average probability that a person you meet at random is susceptible is S / n .

→ an infected person has contact with on average $\beta\ S / n$ susceptible people per unit time.

On average, there are X infected individuals in total.

→ the **average rate of new infections** is $\beta\ S\ X / n$.

The rate of change of X is thus $$\frac{dX}{dt} = \beta \frac{SX}{n}$$

The number of susceptible people goes down at the same rate: $$\frac{dS}{dt} = -\beta \frac{SX}{n}$$

The SI model

It is convenient to define variables representing the fractions of susceptible and infected individuals $s = \frac{S}{n}, \quad x = \frac{X}{n}$

Then, the differential equations become

$$\frac{ds}{dt} = -\beta s x \qquad \frac{dx}{dt} = \beta s x$$

Since $S + X = n$ and thus $s + x = 1$, we don't need both equations.

We can e.g. eliminate s from the equations by replacing $s = 1 - x$

This gives $\frac{dx}{dt} = \beta(1 - x)x$

The SI model - solution

This equation $\frac{dx}{dt} = \beta(1-x)x$ occurs in many places in biology, physics and elsewhere. It is called the **logistic growth equation**.

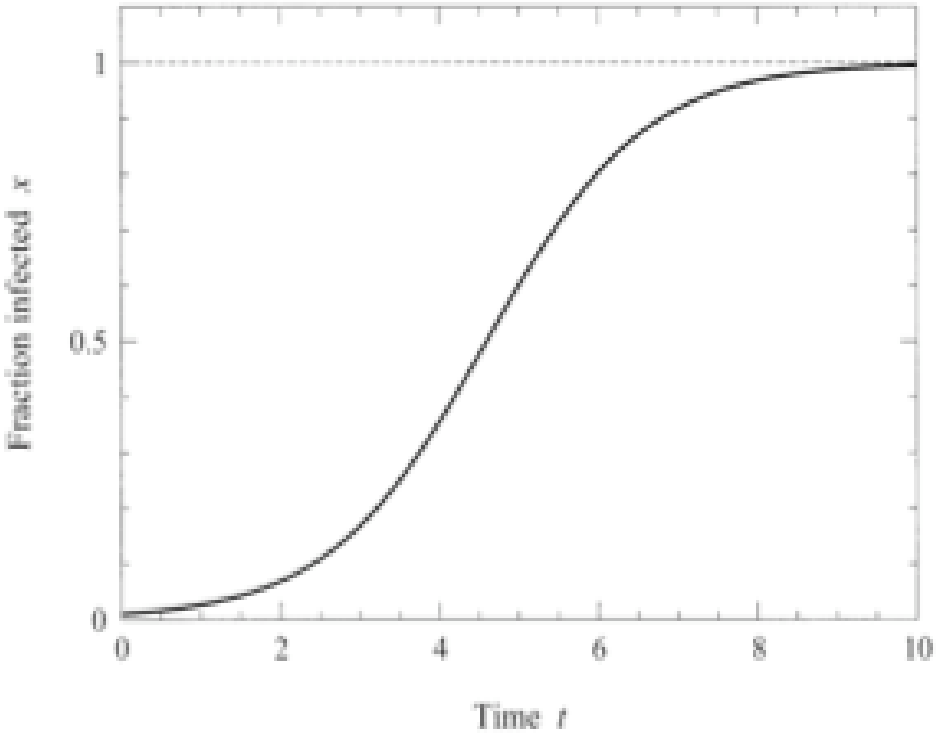
Its solution is $x(t) = \frac{x_0 e^{\beta t}}{1 - x_0 + x_0 e^{\beta t}}$

where x_0 is the value of x at time $t = 0$.

Generally, this produces an S-shaped „logistic growth curve“ for the fraction of infected individuals at time t.

The SI model is the simplest possible model of infection.

An initial burst phase is followed by saturation when all people are infected.



The SIR model: susceptible – infected – recovered/removed

There are many ways to extend the SI model to make it more realistic or more appropriate as a model of a specific disease.

One common extension includes **recovery** from disease.

In the SI model, infected individuals remain infected (and infectious) forever.

For many real diseases, however, people recover from infection after a certain time because their immune system fights off the agent causing the disease.

The SIR model: susceptible – infected – recovered/removed

Furthermore, people often retain their immunity to the disease after such a recovery such that they cannot catch it again.

For this, we need a third state, the **recovered state *R***.

For some other diseases, people do not recover but **die** instead.

In epidemiological terms, such „removal“ is the „same thing“ as „recovery“. (This is sarcastic …)

We can treat both scenarios with the same S – I – R model.

The SIR model: 2 stages

The SIR model was introduced in 1927 by W. O. Kermack and A. G. McKendrick

The dynamics of the fully mixed SIR model has 2 stages.

In stage 1, susceptible individuals become infected when they have contact with infected individuals. Contacts happen at an average rate β as before.

In stage 2, infected individuals recover (or die) at some constant average rate γ.

τ : time that an infected individual is likely to remain infected before they recover.

Probability of recovering in any time interval $\delta\tau$ is $\gamma\ \delta\tau$.
Probability of not recovering in the same interval: $1 - \gamma\ \delta\tau$

The SIR model: susceptible – infected – recovered/removed

→ probability that the individual is still infected after a total time τ:

$$\lim_{\delta\tau\to 0}(1-\gamma\,\delta\tau)^{\tau/\delta\tau} = e^{-\gamma\tau}$$

(remember that $e^x = \lim_{n\to\infty}\left(1+\frac{x}{n}\right)^n$; $n\to\infty$ is the same thing as $\delta\tau\to 0$)

The probability $p(\tau)d\tau$ that the individual remains infected this long and then recovers in the interval between τ and $\tau + d\tau$ is this quantity times $\gamma\, d\tau$:

$$p(\tau)d\tau = \gamma e^{-\gamma\tau}d\tau$$

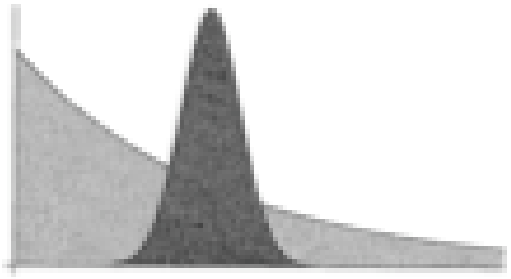
This is a standard exponential distribution.

Thus an infected person is most likely to recover directly after becoming infected, but might in theory remain in the infected state for quite a long time.

The SIR model: comparison to real diseases

This behavior is not very realistic for most real diseases where most victims remain infected for about the same length of time (one ore several weeks).

Few stay in the infected state for much longer or shorter than the average.



Distribution of times for which an individual remains infected is typically narrowly peaked around some average value (dark curve) for real diseases, quite unlike the exponential distribution assumed by the SIR model (grey curve).

This is one thing that we will improve when we return to look at network models of epidemics.

The SIR model: mathematical solution

In terms of the fractions *s, x,* and *r* of individuals in the 3 states, the equations for the SIR model are:

$$\frac{ds}{dt} = -\beta s x$$

$$\frac{dx}{dt} = \beta s x - \gamma x$$

$$\frac{dr}{dt} = \gamma x$$

In addition, the 3 variables satisfy $s + x + r = 1$.

To solve these equations, we eliminate *x* by inserting the 3rd eq. into the 1st one

$$\frac{1}{s}\frac{ds}{dt} = -\frac{\beta}{\gamma}\frac{dr}{dt}$$

The SIR model: mathematical solution

$$\frac{1}{s}\frac{ds}{dt} = -\frac{\beta}{\gamma}\frac{dr}{dt}$$

To solve this, we integrate both sides with respect to *t:*

$$\int_{t=0}^{t'} \frac{1}{s}\frac{ds}{dt}\,dt = \int_{t=0}^{t'} -\frac{\beta}{\gamma}\frac{dr}{dt}\,dt$$

$$\ln\, s = -\frac{\beta}{\gamma}r + r_0$$

$$s = e^{-\frac{\beta}{\gamma}r + r_0}$$

$$s = s_0 e^{-\frac{\beta}{\gamma}r}$$

Here s_0 is the value of s at $t = 0$ and we have chosen the constant of integration so that there are no individuals in the recovered state at $t = 0$.

The SIR model: numerical solution

Now we put $x = 1 - s - r$ into $\frac{dr}{dt} = \gamma x$ and use $s = s_0 e^{-\frac{\beta}{\gamma} r}$ to get

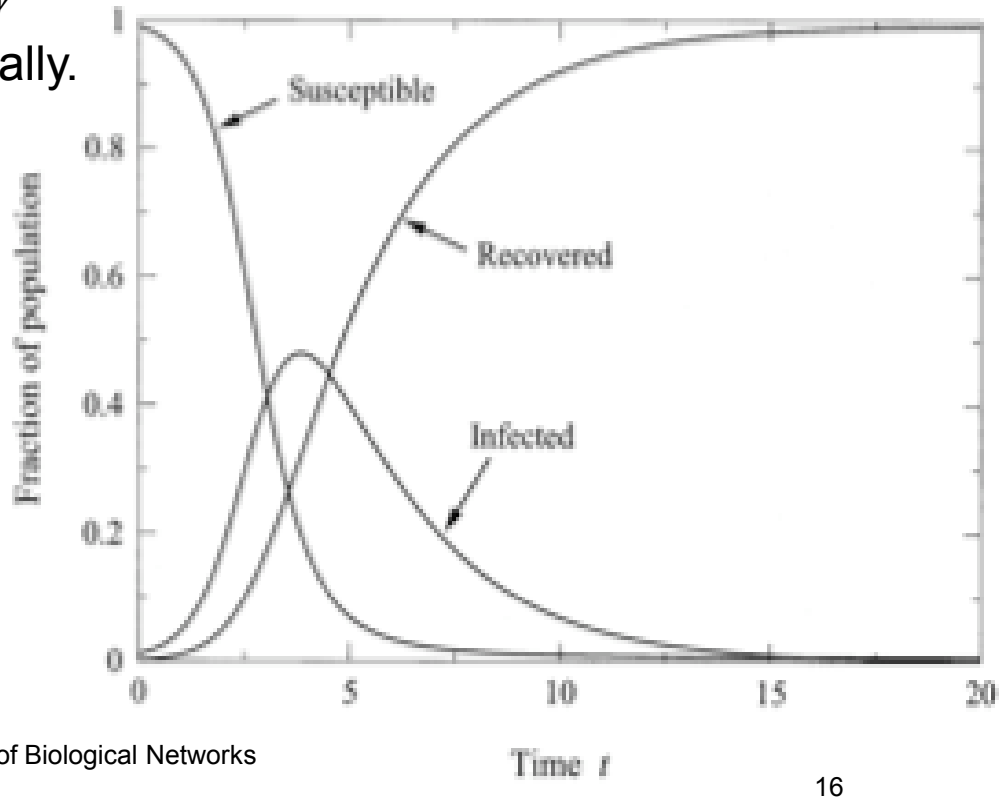
$\frac{dr}{dt} = \gamma \left(1 - r - s_0 e^{-\frac{\beta}{\gamma} r}\right)$ or $\frac{1}{\gamma} \frac{1}{1 - r - s_0 e^{-\frac{\beta}{\gamma} r}} dr = dt$

The solution of this is $t = \frac{1}{\gamma} \int_0^r \frac{du}{1 - u - s_0 e^{-\frac{\beta}{\gamma} u}}$.

This integral can be evaluated numerically.

From *r* we then get *s* and *x*.

The figure shows an example case.



The SIR model: initial behavior

In the most common case, the disease starts either with a single infected individual or with a small number *c* of individuals.

→ the initial values of the variables are

$$s_0 = 1 - \frac{c}{n}, \qquad x_0 = \frac{c}{n}, \qquad r_0 = 1$$

In the limit of large population sizes $n \to \infty$, we can write $s_0 \cong 1$.

Then, the final value of *r* satisfies $r = 1 - e^{-\beta r/\gamma}$

If $\beta \leq \gamma$ there will be no epidemic.

In that case, infected individuals recover faster than susceptible individuals become infected.

The SIR model: epidemic transition

The transition between the epidemic and non-epidemic regimes happens at the point $\beta = \gamma$ and is called the **epidemic transition**.

An important quantity in the study of epidemics
is the **basic reproduction number** R_0 .

This is defined as the average number of additional susceptible people
to which an infected person passes the disease before the person recovers.

If each person catching the disease passes it onto 2 others on average,
then $R_0 = 2$.

If half of them pass it on to just one person and the rest to none then $R_0 = 0.5$

The SIR model: epidemic threshold

If we had $R_0 = 2$, the number of new cases would double at each round, thus grow exponentially.

Conversely if $R_0 = 0.5$ the disease would die out exponentially.

The point $R_0 = 1$ separates the growing and shrinking behavors.

This is the **epidemic threshold.**

We can calculate R_0 straightforwardly for the SIR model.

If an individual remains infectious for a time τ, then the expected number of others they will have contact with during that time is $\beta\tau$.

The SIR model: epidemic threshold

In a „naive population“ at the start of a disease (where only a few individuals are infected and the other susceptible) all of the people with whom one has contact will be susceptible.

Then we average over the distribution of τ $\quad p(\tau) = \gamma e^{-\gamma\tau}$
to get the average number R_0 :

$$R_0 = \beta\gamma \int_0^\infty \tau e^{-\gamma\tau} d\tau = \cdots = \frac{\beta}{\gamma}$$

after using some integration tricks.

The epidemic threshold of the SIR model is thus $\beta = \gamma$ as we have derived before.

The SIS model

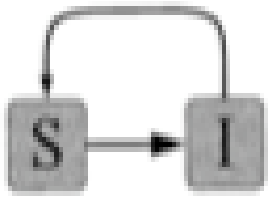
A different extension of the SI model is one that allows for **reinfection**.

For diseases that do not confer immunity to their victims after recovery, individuals can be infected more than once.

The simplest such model is the **SIS model**.

It has the 2 states susceptible and infected.

Infected individuals move back into the susceptible state after recovery.



$$\frac{ds}{dt} = \gamma x - \beta s x$$

$$\frac{dx}{dt} = \beta s x - \gamma x$$

with $s + x = 1$

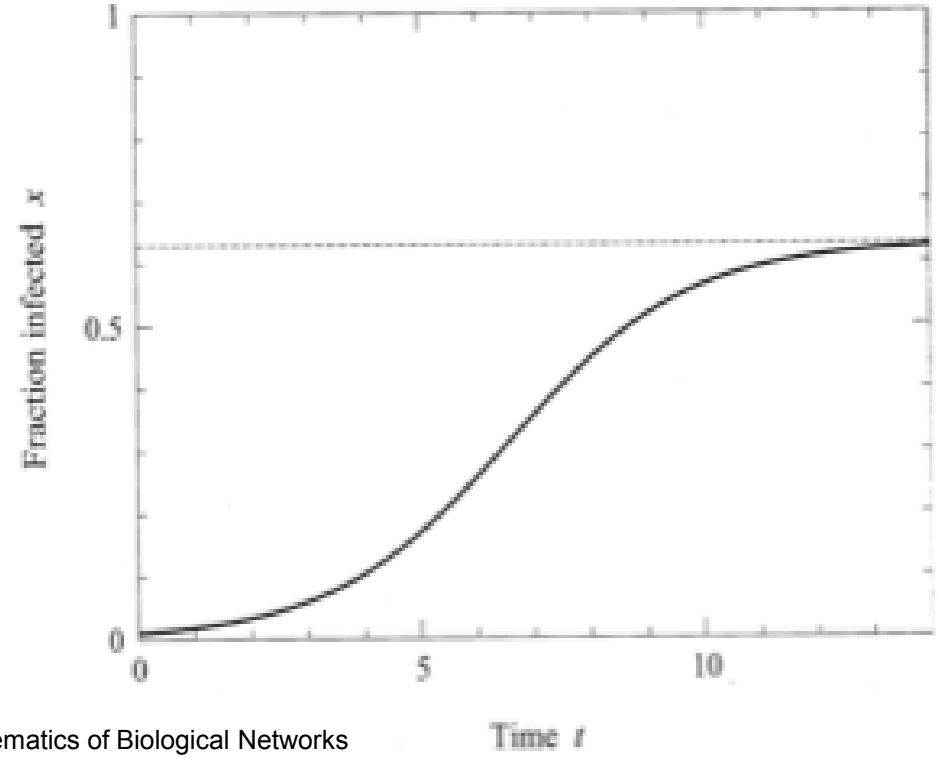
The SIS model

Putting $s = 1 - x$ gives $\frac{dx}{dt} = (\beta - \gamma - \beta x)x$

Which has the solution $x(t) = \left(1 - \frac{\gamma}{\beta}\right) \frac{Ce^{(\beta-\gamma)t}}{1+Ce^{(\beta-\gamma)t}}$

In the case of a large population and a small number of initial carriers, and $\beta > \gamma$ this produces a logistic growth curve.

In this model, we never have the whole population infected with the disease.



The SIRS model

In the SIRS model, individuals recover from infection and gain temporary immunity.

After a certain time they become susceptible again with an average rate δ.

$$\frac{ds}{dt} = \delta r - \beta s x$$

$$\frac{dx}{dt} = \beta s x - \gamma x$$

$$\frac{dr}{dt} = \gamma x - \delta r$$

and

$$s + x + r = 1$$

This model cannot be solved analytically.

Review

The SIR model is appropriate for infectious diseases that confer lifelong immunity, such as measles or whooping cough.

The SIS model is predominantly used for sexually transmitted diseases (STDs), such as chlamydia or gonorrhoea, where repeated infections are common.

Keeling & Eames, J R Soc Interface (2005) 2: 295–307.

Epidemic Models on Networks

Sofar all approaches introduced have assumed „full mixing“ of the population.

In this case each individual can potentially have contact with any other at a level sufficient to transmit the disease.

In the real world, however, the set of a person‘s contacts can be represented as a network.

The structure of that network can have a strong effect on the way a disease spreads through the population.

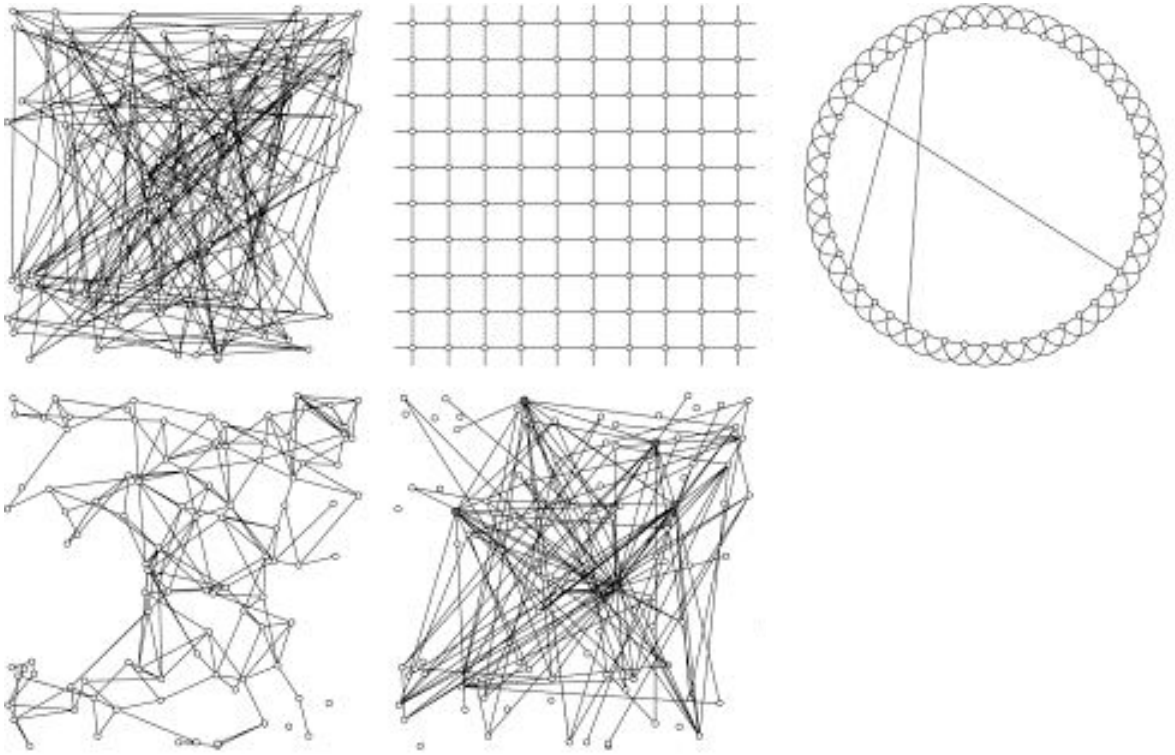
Epidemic Models on Networks

We will define the **transmission rate** β (or infection rate) as the probability per unit time that an infected individual will transmit the disease to a susceptible individual to whom he/she is connected by an edge in the appropriate network.

The transmission rate is a property of a particular disease but also a property of the social and behavioral parameters of the population.

Epidemics on idealized networks

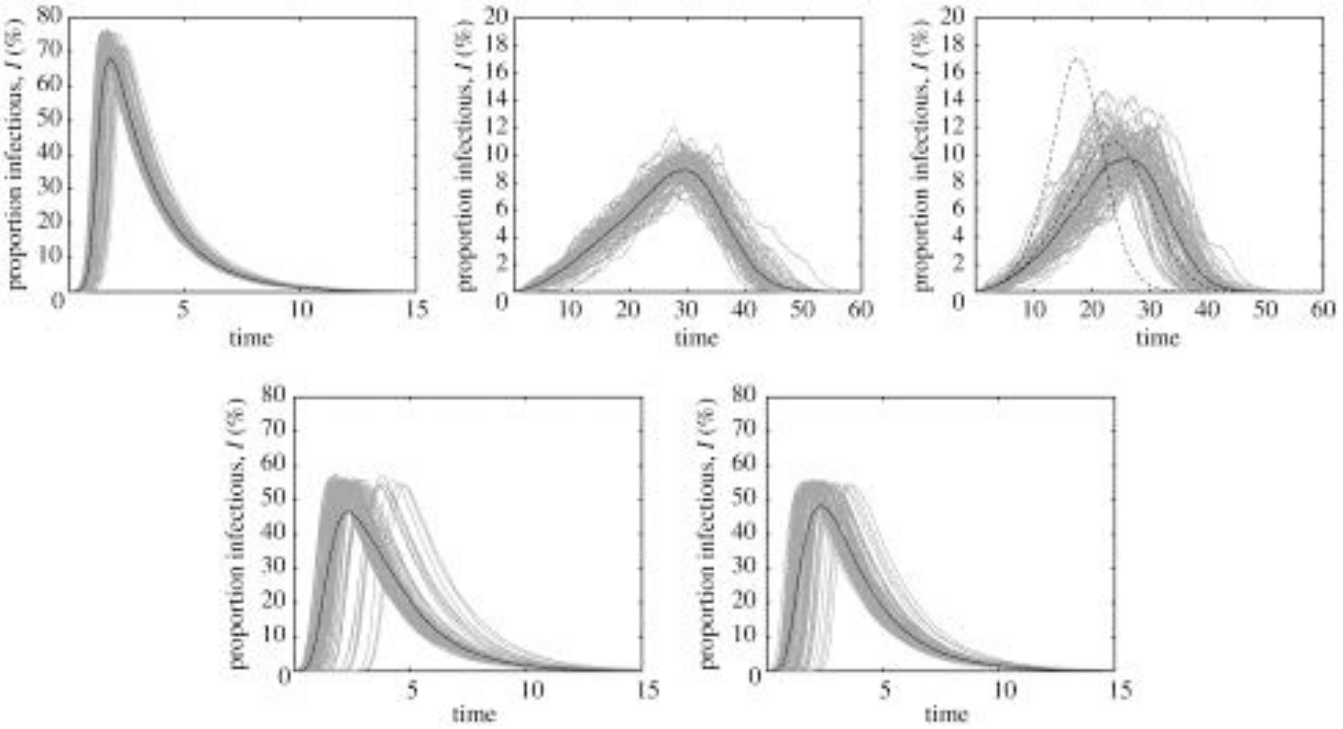
Shown are 5 distinct network types containing 100 individuals.
These are from left to right: random, lattice, small world (top row), spatial and scale-free (bottom row).



In all 5 graphs, the average number of contacts per individual is approximately 4.

Keeling & Eames, J R Soc Interface (2005) 2: 295–307.

Dynamic spreading on different network architectures



Typical SIR epidemics on the 5 network types.

→ the square lattice (top, middle) shows the slowest dynamics

→ highly connected “hub” nodes accelerate spreading of disease

Keeling & Eames, J R Soc Interface (2005) 2: 295–307.

Time-dependent properties of Epidemic Networks

An SI outbreak starting with a single randomly chosen vertex somewhere eventually spreads to all members of the component containing that vertex.

Let us assume that vertex *i* belongs to the giant component.

With probability s_i , vertex *i* is **susceptible**.

To become infected an individual must catch the disease from a neighboring individual *j* that is already infected.

The probability for *j* being **infected** is

$$x_j = 1 - s_j$$

The **transmission** of the disease during the time interval *t* and *t + dt* occurs with probability βdt.

Time-dependent properties of Epidemic Networks

Multiplying these probabilities and then summing over all neighbors of *i* , yields the total probability of *i* becoming infected :

$$\beta s_i \sum_j A_{ij} x_j$$

where A_{ij} is an element of the adjacency matrix.

Thus, the s_i obey a set of *n* non-linear differential equations

$$\frac{ds_i}{dt} = -\beta s_i \sum_j A_{ij} x_j = -\beta s_i \sum_j A_{ij}(1 - s_j)$$

From $s_i + x_i = 1$ we get the complementary equation for x_i.

$$\frac{dx_i}{dt} = \beta s_i \sum_j A_{ij} x_j = \beta (1 - x_i) \sum_j A_{ij} x_j$$

We will assume again that the disease starts either with a single vertex or a small randomly selected number *c* of vertices.

Thus $x_i = c/n \rightarrow 0$, $s_i = 1 - c/n \rightarrow 1$ in the limit for large *n*.

Time-dependent properties of Epidemic Networks

The equation $\frac{ds_i}{dt} = -\beta s_i \sum_j A_{ij} x_j$=- $\beta s_i \sum_j A_{ij}(1 - s_j)$ is not solvable in closed form for general A_{ij}.

By considering suitable limits, we can calculate some features of its behavior.

Let us e.g. consider the behavior of the system at early times.

For large n and the given initial conditions, x_i will be small in this regime.

By ignoring terms of quadratic order, we can approximate

$$\frac{dx_i}{dt} = \beta s_i \sum_j A_{ij} x_j \approx \beta \sum_j A_{ij} x_j$$

or in matrix form

$$\frac{\mathrm{d}\mathbf{x}}{\mathrm{dt}} = \beta \mathbf{A}\mathbf{x}$$ where **x** is the vector with elements x_i .

Time-dependent properties of Epidemic Networks

Write **x** as a linear combination of the eigenvectors of the adjacency matrix

$$\mathbf{x}(t) = \sum_{r=1}^{n} a_r(t)\mathbf{v}_r$$

where $\mathbf{v}_r$ is the eigenvector with eigenvalue σ_r . Then

$$\frac{d\mathbf{x}}{dt} = \sum_{r=1}^{n} \frac{da_r}{dt}\mathbf{v}_r = \beta\mathbf{A}\sum_{r=1}^{n} a_r(t)\mathbf{v}_r = \beta\sum_{r=1}^{n} \sigma_r a_r(t)\mathbf{v}_r$$

By comparing terms in $\mathbf{v}_r$ we get $\frac{da_r}{dt} = \beta\sigma_r a_r$

This has the solution $a_r(t) = a_r(0)e^{\beta\sigma_r t}$

Substituting this expression back gives $\mathbf{x}(t) = \sum_{r=1}^{n} a_r(0)e^{\beta\sigma_r t}\mathbf{v}_r$

The fastest growing term in this expression is the term corresponding to the largest eigenvalue σ_1 .

Time-dependent properties of Epidemic Networks

Assuming this term dominates over the others we will get

$$\mathbf{x}(t) \sim e^{\beta \sigma_1 t} \mathbf{v}_1$$

So we expect the number of infected individuals to grow exponentially, just as in the fully mixed version of the SI model, but now with an exponential constant that depends not only on β but also on the leading eigenvalue of the adjacency matrix.

The probability of infection in this early period varies from vertex to vertex roughly as the corresponding element of the leading eigenvector $\mathbf{v}_1$.

In lecture V1, the elements of the leading eigenvector of the adjacency matrix were termed the **eigenvector centrality**.

Thus eigenvector centrality is a crude measure of the probability of early infection of a vertex in an SI epidemic outbreak.

Review (V1): Eigenvector Centrality

This limiting vector of the eigenvector centralities is simply proportional to the leading eigenvector of the adjacency matrix.

Equivalently, we could say that the centrality **x** satisfies

$$\mathbf{A}\,\mathbf{x} = k_1\,\mathbf{x}$$

This is the **eigenvector centrality** first proposed by Bonacich (1987).

The centrality x_i of vertex i is proportional to the sum of the centralities of its neighbors:

$$x_i = {k_1}^{-1} \sum_j A_{ij} x_j$$

This has the nice property that the centrality can be large either because a vertex has many neighbors or because it has important neighbors (or both).