Network Science (VU) (706.703)

Epidemics

Denis Helic

ISDS, TU Graz

January 18, 2019
Outline

1. Introduction
2. SI Model
3. SIR Model
4. Epidemics on Networks
5. Time-Dependent Properties
Models of the spread of disease

- Mathematical representation of epidemics predates network science
- In a typical model we classify individuals into one of the categories
  1. **Susceptible** (S): someone who does not have disease but can catch it if they come in contact with an infected person
  2. **Infected** (I): someone who has the disease and can pass it on if they come in contact with a susceptible person
  3. **Recovered** (R): someone who was infected and has cured the disease and is now immune to it
Models of the spread of disease

- These states abstract all the biological details of catching a disease
- However, they capture some of the gross diseases dynamics
- Numbers of S, I, R persons in time
- The time component
- The models also abstract the individual tendency to catch the disease by assuming that each individual catches the disease if it comes into contact with an infected person
Models of the spread of disease

- Models are probabilistic, i.e. there is a chance that an individual comes in contact with another individual.

- All the numbers are therefore random, and if the disease spreads more than once the numbers will be a bit different each time.

- Thus, we are interested in expected or average numbers of S, I, R persons in time.

- Additional simplification for the beginning: complete networks.

- Everyone is connected to everyone and has the equal chance of getting in contact to every other person per unit times.

- Fully mixed, mass-action approximation, or mean-field theory.
**SI model**

- We have only susceptible (S) and infected (I) individuals
- Let $S(t)$ be the number of susceptible individuals at time $t$
- Let $X(t)$ be the number of infected individuals at time $t$
- Because of randomness we are interested in expectations and denote them with $S$ and $X$
The number of infected individuals goes up when susceptible individuals contract the disease from the infected ones.

Suppose that in a fully mixed model people meet with other people at random with a rate $\beta$.

This means that each individual has, on average, $\beta$ contacts per unit time with randomly chosen other individuals.

The disease is transmitted only when a susceptible person has a contact with an infected person.
The total population consists of \( n \) people

The average probability of meeting a susceptible person: \( \frac{S}{n} \)

An infected person meets therefore \( \beta \frac{S}{n} \) susceptible people per unit time

We have on average \( X \) infected persons

Thus, the overall average rate of new infections per unit time is given by \( \beta \frac{SX}{n} \)
SI Model

The rate of change of $X$ in time is then given by the following differential equation:

$$\dot{X} = \beta \frac{SX}{n}$$

At the same time the number of susceptible individuals goes down at the same rate:

$$\dot{S} = -\beta \frac{SX}{n}$$
Let us define the variables representing the fractions of S and I individuals: $s = \frac{S}{n}$ and $x = \frac{X}{n}$

We can then rewrite the previous equations:

\[
\begin{align*}
\dot{x} &= \beta sx \\
\dot{s} &= -\beta sx
\end{align*}
\]
SI model

- We have $S + X = n$ or $s + x = 1$
- Thus, $s = 1 - x$, and after eliminating $s$ we obtain:

$$\dot{x} = \beta(1 - x)x$$

- This is *logistic growth equation*
SI model

\[ \dot{N} = rN(1 - \frac{N}{K}) \]

\[ \dot{x} = \beta(1 - x)x \]

- \( K \) carrying capacity now equals 1 (all individuals are infected)
- \( r \) growth rate is now \( \beta \)
Logistic growth curve

The logistic growth curve is a model that describes the growth of a population in a limited environment. It is often used to model the spread of diseases in a population. The equation for the logistic growth curve is given by:

\[ \frac{dx}{dt} = \beta x (1 - x) \]

where \( x \) is the fraction of the population that is infected, \( t \) is time, and \( \beta \) and \( x_0 \) are parameters of the model.

In the graph, different curves represent different values of \( x_0 \) and \( \beta \). For example, the blue curve represents \( x_0 = 0.01, \beta = 1 \), the green curve represents \( x_0 = 0.01, \beta = 2 \), the red curve represents \( x_0 = 0.10, \beta = 1 \), and the cyan curve represents \( x_0 = 0.10, \beta = 2 \).
SI model: numerical solution
SIR model

- Now we have susceptible (S), infected (I), and recovered (R) individuals.
- Recovered individuals retain their immunity and can not get infected anymore.
- As before, let \( S(t) \) be the number of susceptible individuals at time \( t \).
- Let \( X(t) \) be the number of infected individuals at time \( t \).
- Let \( R(t) \) be the number of recovered individuals at time \( t \).
- Because of randomness we are interested in expectations and denote them with \( S \), \( X \), and \( R \) respectively.
The dynamics of the fully mixed SIR model has two stages:

- In the first stage S individuals become I individuals when they have contact with I individuals.
- Contacts between individuals happen at average rate $\beta$ per person as before.
- In the second stage infected individuals recover at some constant average rate $\gamma$. 
When do individuals recover?

- Probability to recover in any time interval $\Delta t$: $\gamma \Delta t$
- Probability of not recovering in any time interval $\Delta t$: $1 - \gamma \Delta t$
- Probability that an individual is still infected after a total time $t$:

$$\lim_{\Delta t \to 0} (1 - \gamma \Delta t)^{\frac{t}{\Delta t}} = e^{-\gamma t}$$
The probability that an individual remains infected for time $t$ and then recovers in an interval between $t$ and $t + dt$ is:

$$p(t)dt = \gamma e^{-\gamma t} dt$$

This is a standard exponential distribution: an individual is most likely to recover just after becoming infected, but might in theory remain infected for a long times.

Not really realistic: typically you would expect a normal distribution of recovery times with the average time depending on the disease.

It makes the math simpler ;)}
In terms of the fractions $s$, $x$, and $r$ we can write our equations:

\[
\begin{align*}
\dot{s} & = -\beta sx \\
\dot{r} & = \gamma x \\
\dot{x} & = \beta sx - \gamma x
\end{align*}
\]

$s$, $x$, and $r$ satisfy:

\[s + x + r = 1\]
SIR model

- We eliminate $x$ between the first two equations:
  \[
  \frac{1}{s} \frac{ds}{dt} = -\frac{\beta}{\gamma} \frac{dr}{dt}
  \]

- We integrate both sides with respect to $t$ to get (with $s_0$ being the number of individuals in $S$ state at time $t = 0$ and we have chosen that there are no individuals in $R$ state at time $t = 0$):
  \[
  s = s_0 e^{-\frac{\beta}{\gamma} r}
  \]
Now we put $x = 1 - s - r$ in the second differential equation and obtain:

$$\dot{r} = \gamma (1 - r - s_0 e^{-\frac{\beta}{\gamma} r})$$

The solution is then:

$$t = \frac{1}{\gamma} \int_{0}^{r} \frac{du}{1 - u - s_0 e^{-\frac{\beta}{\gamma} u}}$$

The last integral can not be solved analytically: numerical evaluation
Time evolution of the SIR model

\[ s_0 = 0.99, \beta = 1, \gamma = 0.3 \]
Time evolution of the SIR model

- The fraction of $S$ individuals decreases monotonically
- $S$ individuals become infected
- $I$ individuals recover at constant rate
- The number of $R$ individuals increases monotonically
- The fraction of $I$ individuals goes up at first, and then down as people recover
Time evolution of the SIR model

- The number of S individuals does not go to zero
- When $x \to 0$ there are no infected people to pass the disease
- Any individuals who do not get the disease until late will most probably never get the disease at all
- Similarly, the fraction of R individuals does not quite reach 1
Outbreak size

- The asymptotic value of $r$ has important practical interpretation.
- It is the total number of persons who ever catch the disease.
- It can be calculated as the value for which $\dot{r} = 0$ from:

$$\dot{r} = \gamma (1 - r - s_0 e^{-\frac{\beta}{\gamma} r})$$

- This gives: $r = 1 - s_0 e^{-\frac{\beta}{\gamma} r}$, which can be solved numerically or graphically.
Outbreak size

Saturation in SIR model; $s_0 = 0.99, \beta = 1, \gamma = 0.3, r_{max} = 0.9596$
Outbreak size

- The size of the epidemics is the function of $\beta$ (contact rate) and $\gamma$ (recovery rate).
- The size goes continuously to zero as $\beta/\gamma$ approaches 1 from above.
- Also, for $\beta/\gamma \leq 1$ there is no epidemic at all.
- The simple explanation is that if $\beta \leq \gamma$ then infected individuals recover faster than they meet with susceptible individuals.
- The number of infected persons starts small and goes down, and the disease dies out instead of spreading out.
Python Notebook

- Check SIR examples from python notebook
- http://kti.tugraz.at/staff/denis/courses/netsci/dynamics.ipynb
SIR model: numerical solution
Further extensions of the SI model

- A different extension would be to allow for reinfection
- E.g. after recovering the individuals go into susceptible S state
- SIS model (with some constant rate $\gamma$ the people become susceptible again (e.g. bacteria)
- Another possibility the individuals recover only temporarily
- SIRS model (with some constant rate $\delta$ the people move from recovered to susceptible)
Python Notebook

- Check SIS and SIRS examples from python notebook
- http://kti.tugraz.at/staff/denis/courses/netsci/dynamics.ipynb
**SIS model: numerical solution**

![Time series: SIS Model (Fully Mixed) $x_0 = 0.01, \beta = 1.0, \gamma = 0.3$](image)

Denis Helic  (ISDS, TU Graz)  
NetSci  
January 18, 2019  
32 / 59
SIRS model: numerical solution

Time series: SIRS Model (Fully Mixed) $x_0 = 0.01, \beta = 1.0, \gamma = 0.3, \delta = 0.2$

- $S$ (blue)
- $I$ (green)
- $R$ (red)
Summary: Epidemic models

- Fully mixed models: no network involved
- Everybody can meet anybody else
- Depending on the model various behaviors
  - E.g. SIR model: large outbreak possible depending on the ratios of parameters
  - Some people do not get infected at all
  - E.g. SIRS model: damped waves of small outbreaks, endemic states
- Behavior dependent only on the parameters of the dynamics
In the real world, the networks have strong effect on the way a disease spreads through the population.

E.g. through regular acquaintances, friends, neighbors, etc.

Network models of epidemics take into account the underlying network.

But in principle they work in the same way as fully mixed models.
Epidemic models on networks

- We will define \textit{transmission rate} or \textit{infection rate} to be the probability per unit time that the infection will be transmitted from an infected to a susceptible individual, who are connected by a link in the network.
- The transmission rate is the property of the disease.
- But it is also a property of the social and behavioral parameters of the population.
- Given a value of transmission rate $\beta$ and the initial number of infected individuals $x_0$, we can generalize all the models to the network case.
- However, it is difficult to solve these models for a general network.
- Typically, we simulate the spread on a computer.
Late-time properties of epidemics on networks

- There is one respect in which the model is straightforward
- In the case $t \to \infty$ (late-time properties)
- For example, SI model on networks
- Since nobody recovers, in the end (regardless how small is $\beta$), everyone connected to the initially infected nodes will get infected
- Everyone in the component to which the initial carriers belong
- Most networks have one large component that typically contains a significant fraction of nodes
- Thus, if we start with a single infected individual and that node belongs to the large component we will have a large outbreak
Late-time properties of epidemics on networks

- That is the difference between fully mixed and network models of epidemics
- Fully mixed: the size of the outbreak determined by the model parameters
- Network: disease behavior determined by
  - Model parameters
  - Network structure and the position of the initial carrier
We are interested in the average (expected) probabilities $s_i(t)$, $x_i(t)$, or $r_i(t)$ of a node $i$ being in a given state.

Given the adjacency matrix we can write down the equations for these quantities.

An SI outbreak starting on a randomly chosen node will eventually spread to all members of the component.

We are now interested only in the case of giant component.

All other outbreaks have non-significant effects and die out.
Time-dependent properties of the SI model

- Suppose $i$ is in the giant component
- What is the probability that $i$ gets infected between times $t$ and $t + dt$?
- To become infected $i$ has to catch the disease from an infected neighbor $j$
- Probability that $j$ is already infected: $x_j = 1 - s_j$
- Probability that the disease is transmitted: $\beta dt$
- Additionally we require that $i$ is in S state beforehand (the probability of this: $s_i$)
Time-dependent properties of the SI model

- Multiplying these probabilities and summing over all neighbors of $i$
- The total probability of becoming infected is given by:

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

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

- Minus: the probability of being susceptible goes down when nodes become infected
Time-dependent properties of the SI model

- Similarly:

\[ \dot{x}_i = \beta s_i \sum_j A_{ij} x_j = \beta (1 - x_i) \sum_j A_{ij} x_j \]

- Those two equations are really same equations since \( s_i + x_i = 1 \)
Time-dependent properties of the SI model

- We use the same initial conditions as before.
- We start with a single infected node chosen uniformly at random, or a small fraction of nodes.
- In the limit of large network size the initial conditions are: $x_i = 0$, and $s_i = 1$.
- The equations are coupled set of $n$ non-linear differential equations.
- Can not be solved in closed form for general $A_{ij}$.
- Let us therefore consider suitable limits.
Time-dependent properties of the SI model

\[ \dot{x}_i = \beta (1 - x_i) \sum_j A_{ij} x_j = \beta \sum_j A_{ij} x_j - x_i \sum_j A_{ij} x_j \]

- For large \( n \) and assuming initial conditions as above \( x_i \) will be small
- We can ignore terms of quadratic order in small quantities

\[ \dot{x}_i = \beta \sum_j A_{ij} x_j \]
Time-dependent properties of the SI model

\[ \dot{x} = \beta A x \]

- Linear system, which we know how to solve!
- We write \( x \) as the linear combination of the eigenvectors, where \( v_r \) is the eigenvector with eigenvalue \( \kappa_r \)

\[ x(t) = \sum_{r=1}^{n} a_r(t) v_r \]
Time-dependent properties of the SI model

\[ \dot{x} = \sum_{i=1}^{n} \frac{d a_r}{d t} v_r = \beta A \sum_{i=1}^{n} a_r(t) v_r = \beta \sum_{i=1}^{n} \kappa_r a_r(t) v_r \]

Then comparing terms in \( v_r \):

\[ \dot{a}_r = \beta \kappa_r a_r \]
Time-dependent properties of the SI model

- Substituting in the previous equation:

\[ x(t) = \sum_{i=1}^{n} a_r(0)e^{\beta \kappa_r t}v_r \]

- The fastest growing term corresponds to \( \kappa_1 \) and assuming that it dominates over the others:

\[ x(t) \sim e^{\beta \kappa_1 t}v_1 \]
Time-dependent properties of the SI model

- We expect the number of $I$ individuals to grow exponentially.
- Similarly to the fully mixed model.
- Now, the exponential constant depends not just on $\beta$.
- It also depends on the leading eigenvalue of the adjacency matrix (structure of the network).
- Moreover, the probability infection in this early period varies from node to node roughly as the corresponding element of the leading eigenvector.
What are the corresponding elements of the leading eigenvector?

Eigenvector centrality!

Thus, eigenvector centrality is an approximate measure of the probability of the early infection of a node.

At late times this probability tends to 1.

At early times the nodes with higher eigenvector centralities become infected faster.
Python Notebook

- Check SI network examples from python notebook
- [http://kti.tugraz.at/staff/denis/courses/netsci/dynamics.ipynb](http://kti.tugraz.at/staff/denis/courses/netsci/dynamics.ipynb)
SI network model: numerical solution
We can extend these techniques to the SIR model. Again, we concentrate on outbreaks taking place in the giant component. We define $s_i$, $x_i$, and $r_i$ to be the probabilities that node $i$ is in S, I, or R state. The evolution of $s_i$ is (approximately) governed by the same equation as before:

$$
\dot{s}_i = -\beta s_i \sum_j A_{ij} x_j
$$
Time-dependent properties of the SIR model

- $x_i$ ($\gamma$ is the recovery rate):

\[ \dot{x}_i = \beta s_i \sum_j A_{ij} x_j - \gamma x_i \]

- $r_i$:

\[ \dot{r}_i = \gamma x_i \]
Time-dependent properties of the SIR model

- We use the same initial conditions as before
- We start with a single infected node chosen uniformly at random, or a small fraction of nodes
- In the limit of large network size the initial conditions are: $x_i = 0$, and $s_i = 1$
- This simplifies the equation for $x_i$:

$$\dot{x}_i = \beta \sum_j A_{ij} x_j - \gamma x_i = \sum_j (\beta A_{ij} - \gamma \delta_{ij}) x_j$$
Time-dependent properties of the SIR model

\[ \dot{x} = \beta Mx \]

- Linear system, which we know how to solve!
- \( M \) is a symmetric \( n \times n \) matrix:

\[ M = A - \frac{\gamma}{\beta} I \]
Time-dependent properties of the SIR model

- We write $x$ as the linear combination of the eigenvectors of $M$, where $\mathbf{v}_r$ is the eigenvector with eigenvalue $\kappa_r$.
- Since $M$ differs from $A$ only in a multiple of $I$ it has the same eigenvectors as $A$

\[
M\mathbf{v}_r = A\mathbf{v}_r - \frac{\gamma}{\beta}I\mathbf{v}_r = (\kappa - \frac{\gamma}{\beta})\mathbf{v}_r
\]

- The eigenvalue has shifted by $\frac{\gamma}{\beta}$
Time-dependent properties of the SIR model

- We can solve as previously:

\[ x(t) = \sum_{i=1}^{n} a_r(0) e^{(\beta \kappa_r - \gamma) t} v_r \]

- The exponential constant depends on \( \beta \) (infection rate), \( \gamma \) (recovery rate), and \( \kappa \) (network structure).

- The faster people recover the less chance they have to pass the disease on the others.

- Again, the fastest growing term corresponds to \( \kappa_1 \) and assuming that it dominates over the others:

\[ x(t) \sim e^{(\beta \kappa_1 - \gamma) t} v_1 \]
Time-dependent properties of the SIR model

- Again, eigenvector centrality is an approximate measure of the probability of the early infection of a node.
- However, it is possible now for $\gamma$ to be sufficiently large that the exponential constant in the leading term is negative.
- Then, the term **decays** exponentially rather than grows.
- The total number of infected individuals will decay over time.
- The disease dies out.
Time-dependent properties of the SIR model

• The point at which this happens:

\[ \beta \kappa_1 - \gamma = 0 \]
\[ \frac{\beta}{\gamma} = \frac{1}{\kappa_1} \]

• The position of epidemic threshold depends on the leading eigenvalue of the adjacency matrix.

• If the leading eigenvalue is small then \( \beta \) must be large, or \( \gamma \) must be small for an outbreak.

• In other words: small value of \( \kappa_1 \) makes it harder for disease to spread.

• Sparse networks have smaller \( \kappa_1 \), dense networks have larger \( \kappa_1 \).