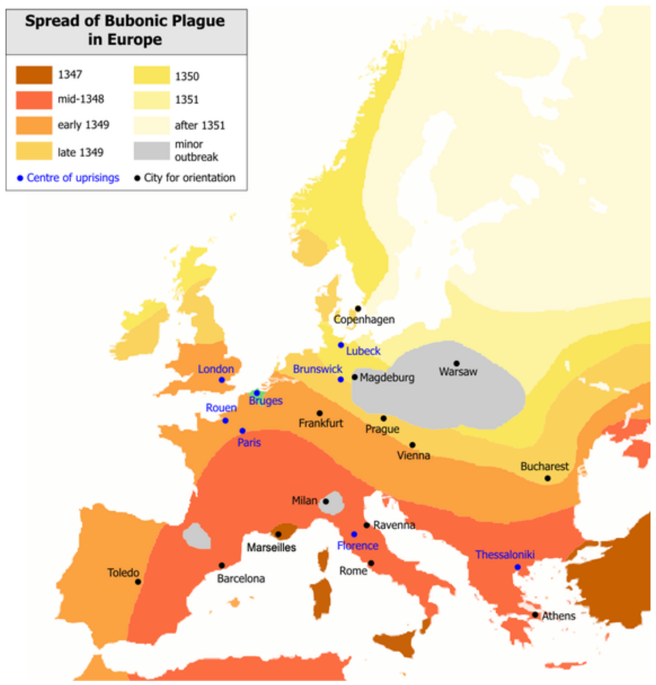
DIFFUSION ON NETWORKS

Spreading processes

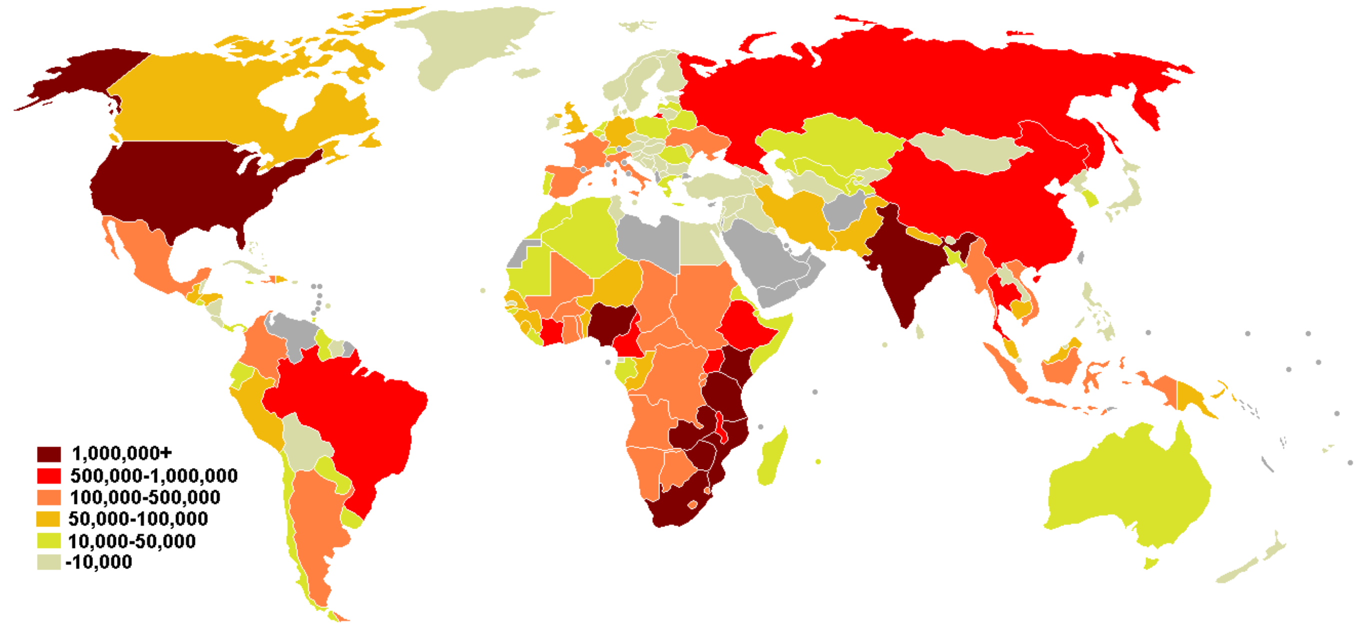
Dynamic ON networks

Spreading processes

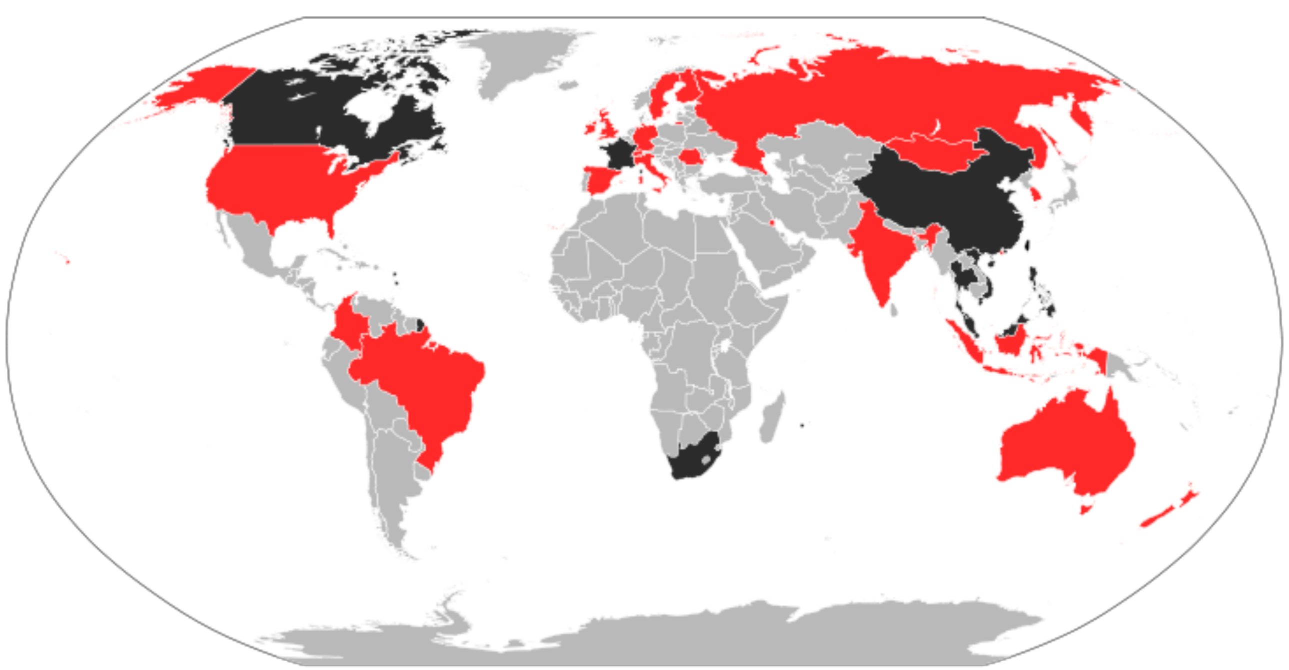
Biological epidemic spreading



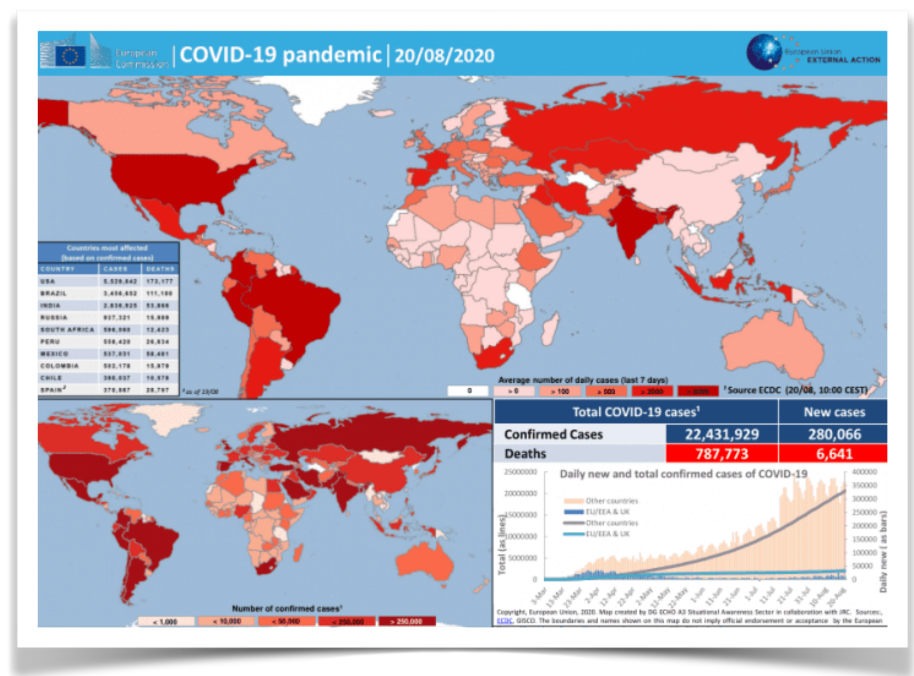
The great plague (14th century)



HIV (2008)



SARS (2008)

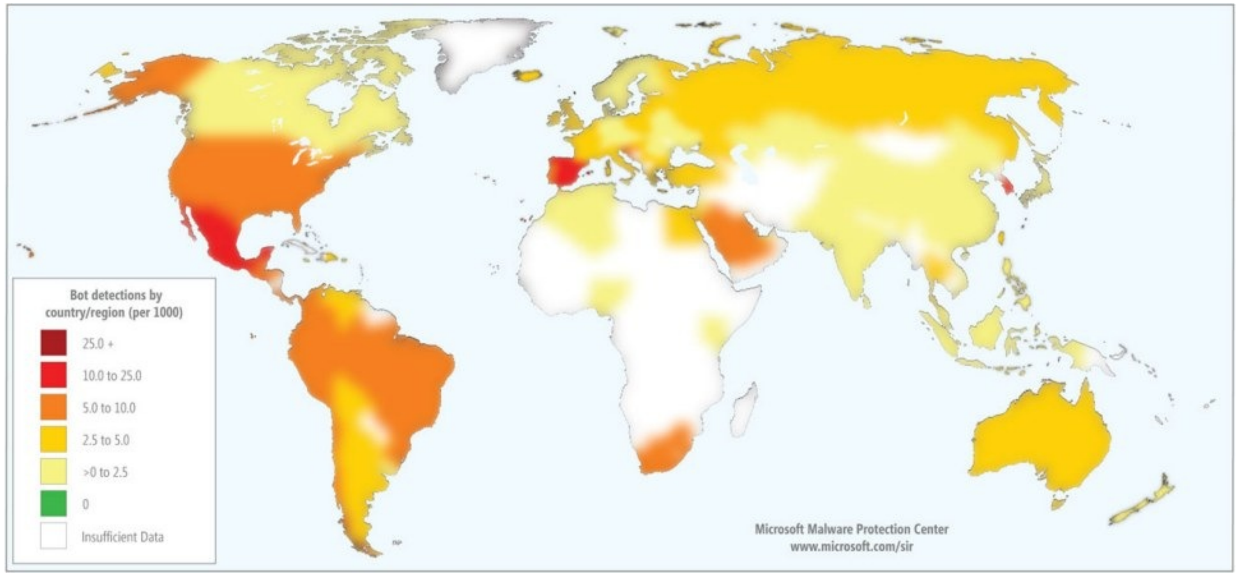


Covid-19

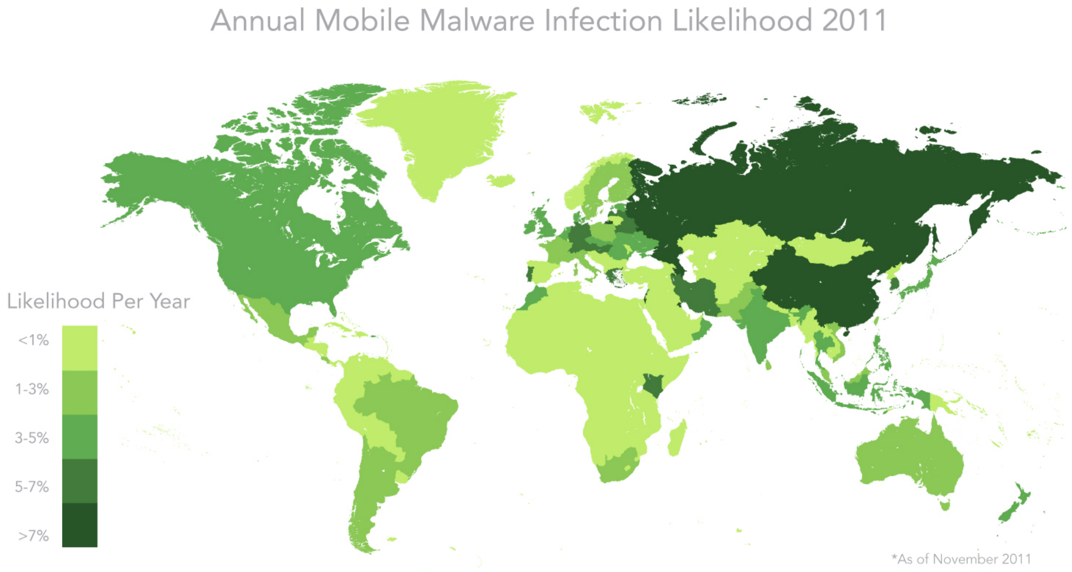
Wikipedia

Spreading processes

Malware spreading



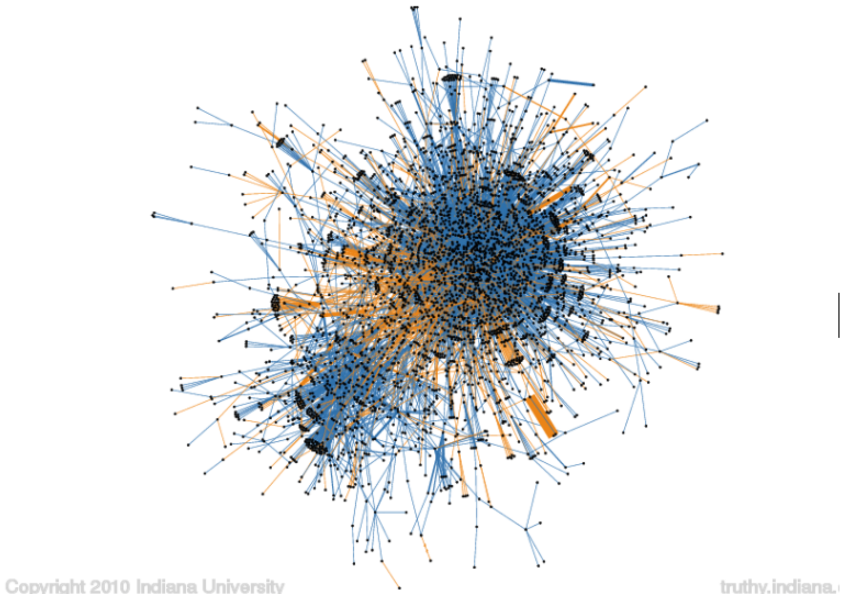
Botnet infections (2010)



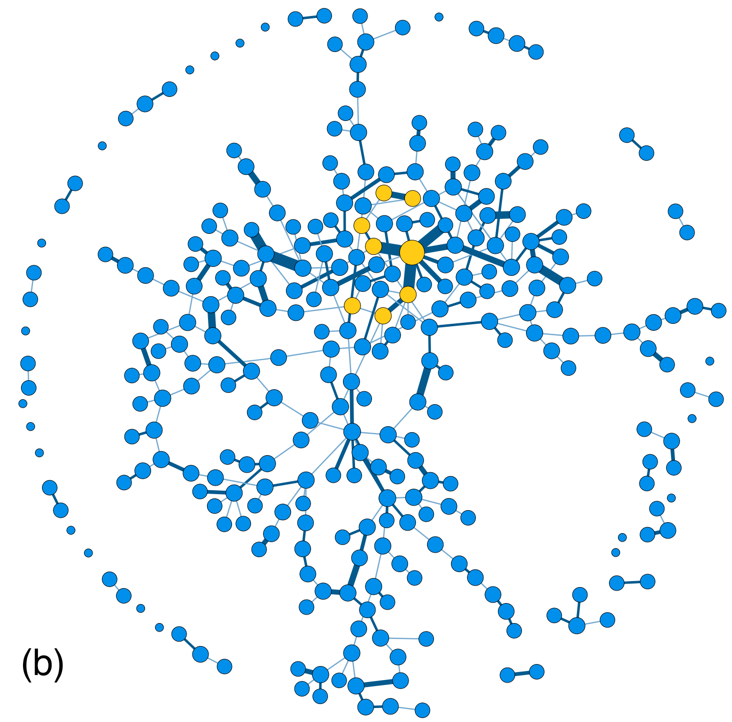
Mobile malware (2011)

Spreading processes

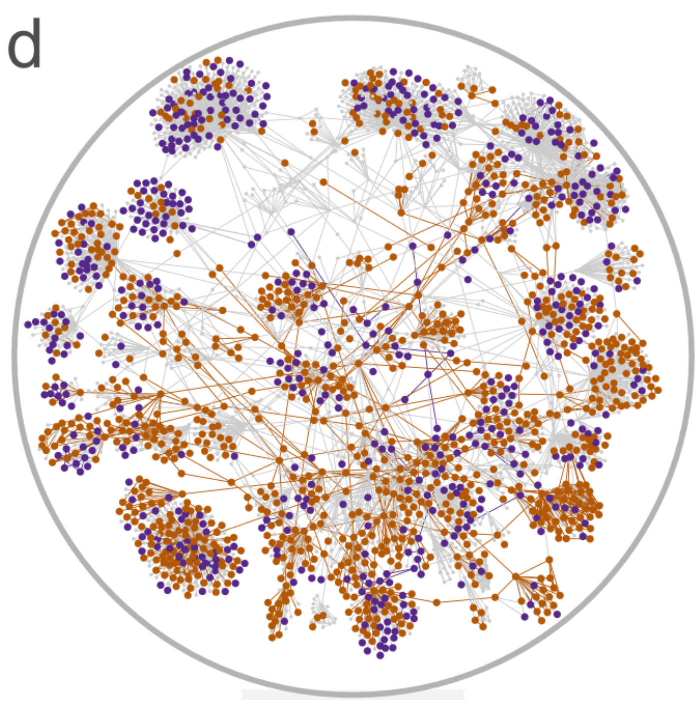
Social contagion



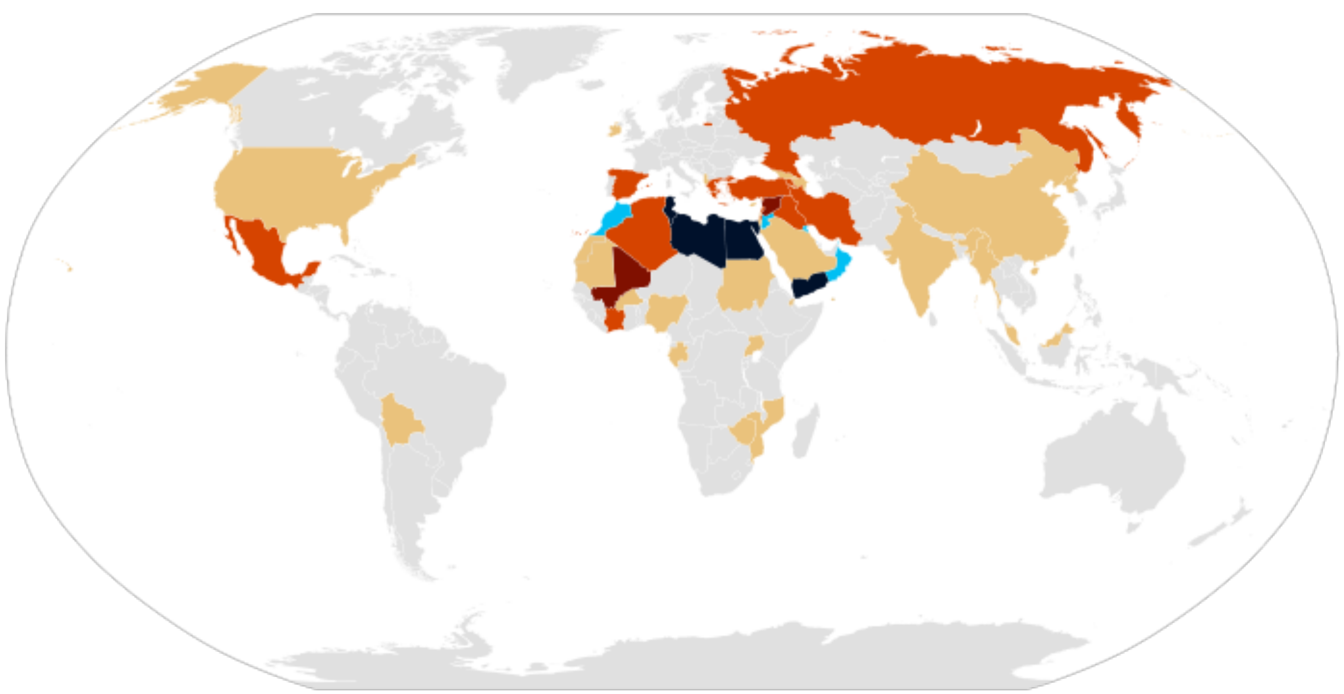
Information spreading



Rumour spreading
Karsai et.al. (2014)



Adoption spreading (Skype)
Karsai et.al. (2014)



Wikipedia

Protest diffusion (Arabian spring)

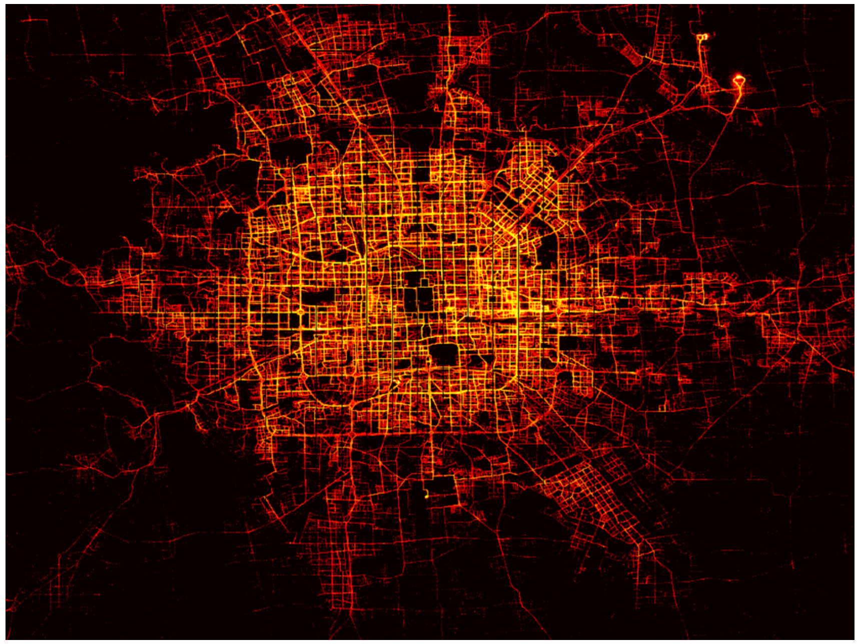
Spreading processes

Why on networks?

- Spreading usually happen through interactions between agents
    - Geographic vicinity
    - Physical connection
    - Social interaction
    - etc.
- Network structure critically influence the dynamics of spreading processes

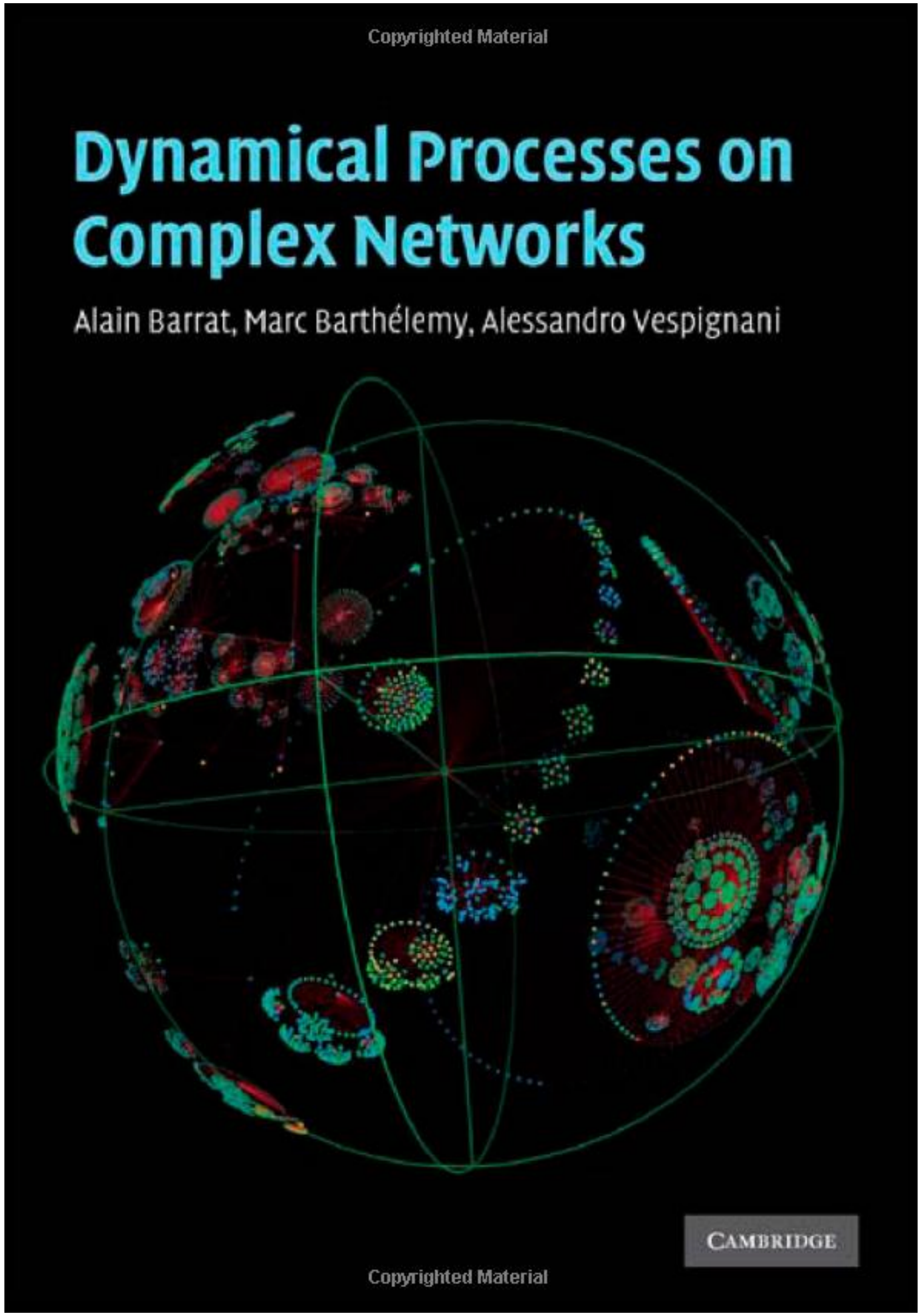
PPD blog, Jooseery (2011)

Freese (2009)



Unknown

Literature



I'm not an epidemiologist!

Only an introduction,
Trust the experts

Simple spreading

processes

Spreading processes

SI - SIR - SIS

Three of the most popular models of diffusion in epidemiology are the **SI**, **SIR** and **SIS** models. Letters correspond to the states in which individuals can be according to the model:

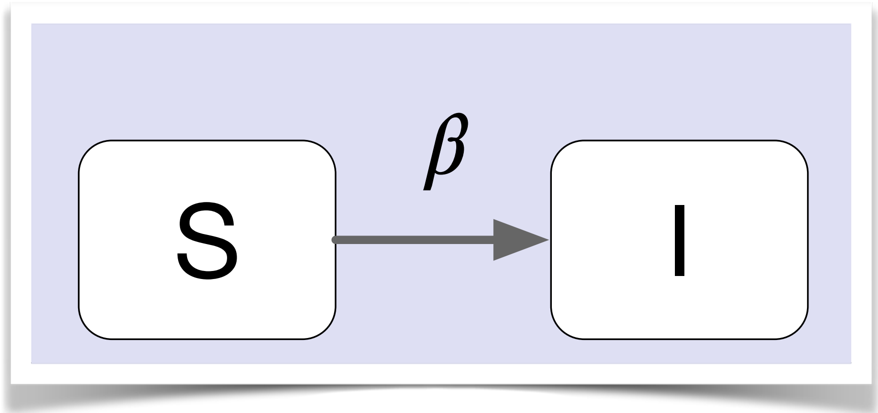
- **Susceptible**: Individual is not Infected
- **Infected**: Individual is Infected
- **Recovered/Removed**. Individual cannot be infected again (Considered cured or dead)

All individuals are in one of the states allowed by the model, and we define:

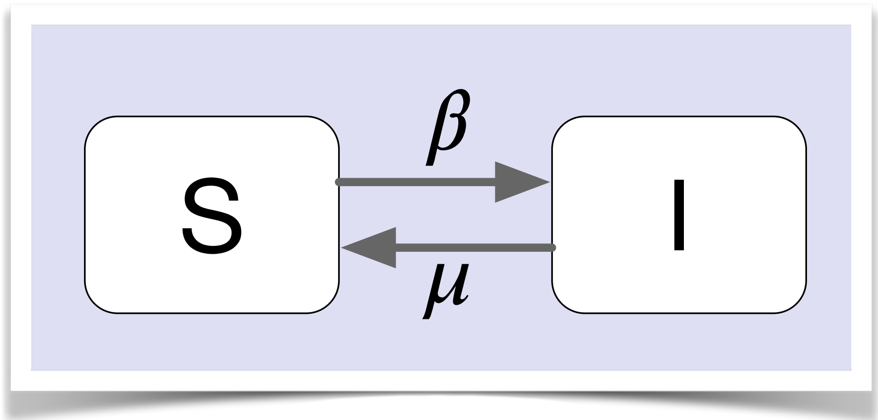
$s(t)$	Fraction of individuals in Susceptible state at time t
$i(t)$	Fraction of individuals in Infected state at time t
$r(t)$	Fraction of individuals in Recovered state at time t
i_0	Initial($t = 0$) fraction of infected individuals

Spreading processes

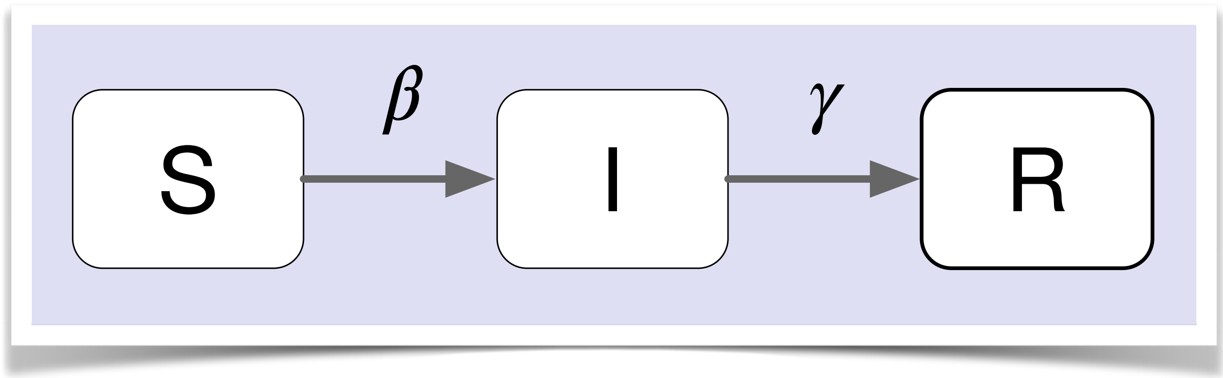
SI



SIS



SIR



Homogeneous mixing

Non-network approach

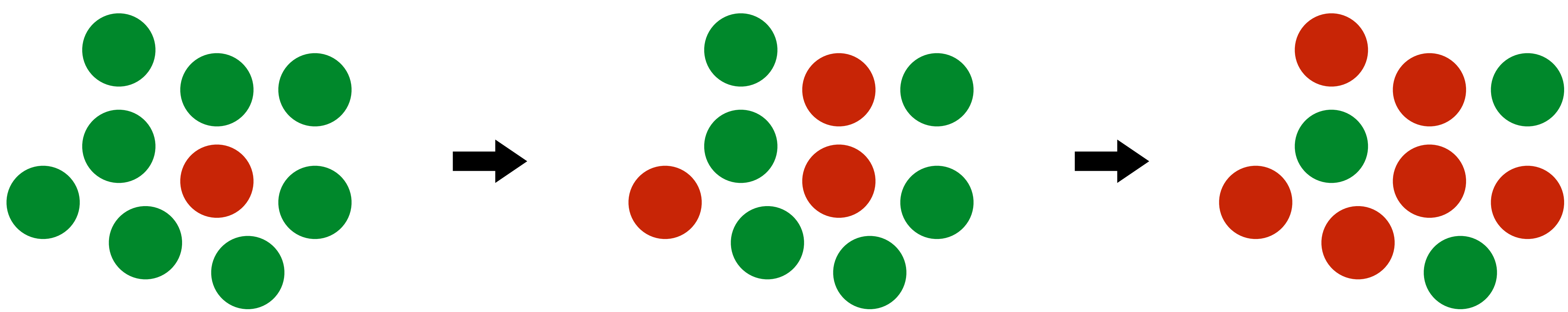
- Any individual can interact with any other
- The population has a finite size
- Individuals have an average number of contacts per unit of time

Kermack, W. O., & McKendrick, A. G. (1927). A contribution to the mathematical theory of epidemics. *Proceedings of the royal society of london. Series A, Containing papers of a mathematical and physical character*, *115*(772), 700-721.

SI model

τ	**Infectivity**: probability that the contact between an *Infected* individual and a *Susceptible* one results in the infection of the Susceptible.
$\hat{c}$	*Contact rate*: average number of contact per person per time
β	**Effective contact rate**, $\beta = \tau\hat{c}$, number of newly infected individuals by each infected individual in a population in which everyone else is susceptible.

$$\boxed{S} \xrightarrow{\boldsymbol{\beta}} \boxed{I}$$



The SI model

SI - characteristics

Each of the i infected individuals infects in average β contacts, but only $s = (1 - i)$ of its contacts are indeed susceptible. More formally using differential equations:

$\frac{di}{dt}$	**Rate of new infection**: $\frac{di}{dt} = \beta is = \beta(1 - i)i$
$i(t)$	**Infected fraction**[a]: $i(t) = \frac{i_0 e^{\beta t}}{1 - i_0 + i_0 e^{\beta t}}$
$s(t)$	**Susceptible fraction**: $1 - i(t)$

The SI model

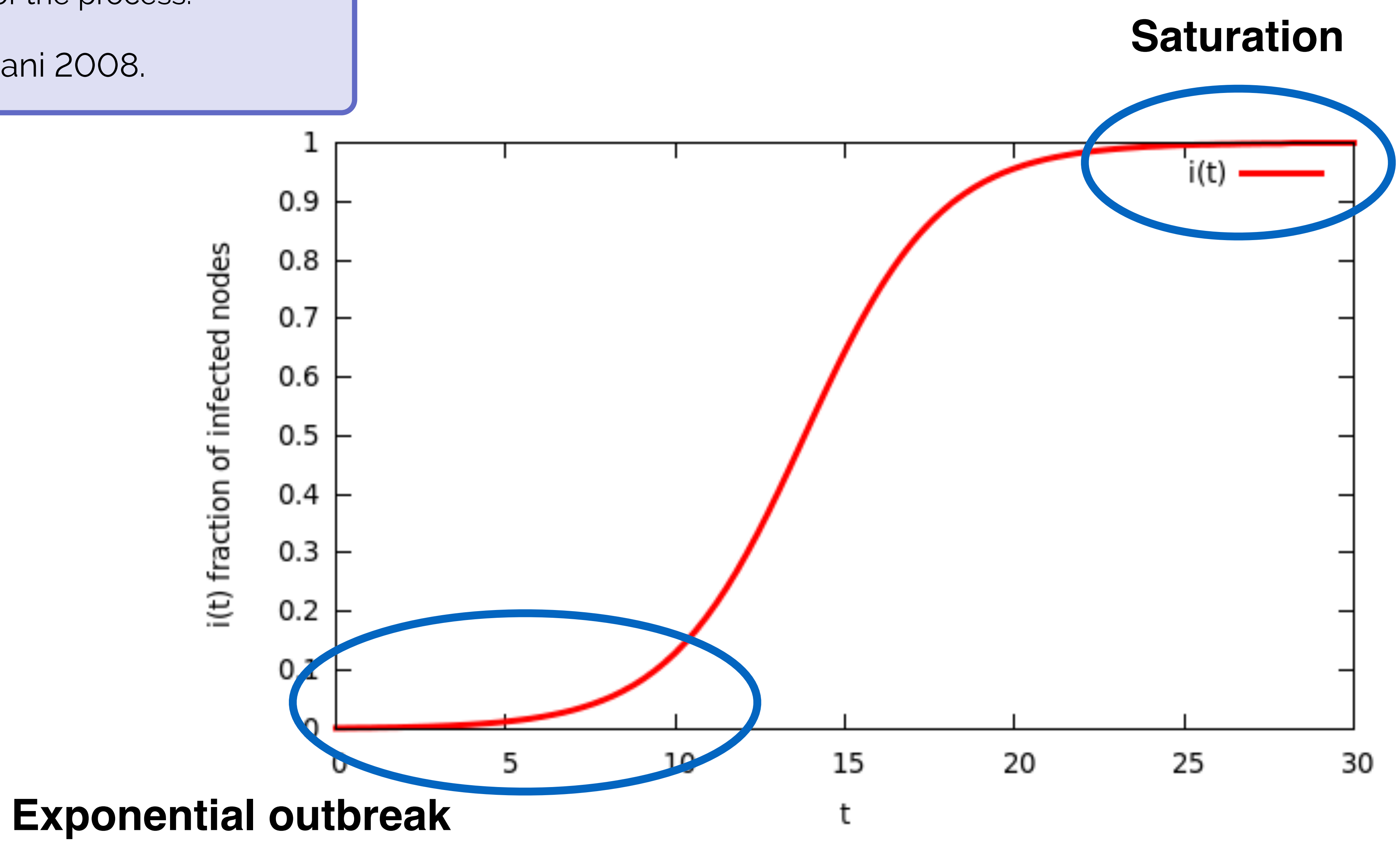
The process can be separated in three steps:

- At first, the fraction of infected individuals **Grows exponentially** until a large fraction of the population is infected. (i is small, $\frac{di}{dt} \approx \beta i \Rightarrow$ exponential)
- Due to **saturation**, the infection of the last individuals is slow
- The growth is faster and faster until half the population is infected ($\mathrm{argmax}_{x,y}(x(1-x)) : x = y = 0.5$).

If $\beta > 0$, everyone is infected at the end of the process.

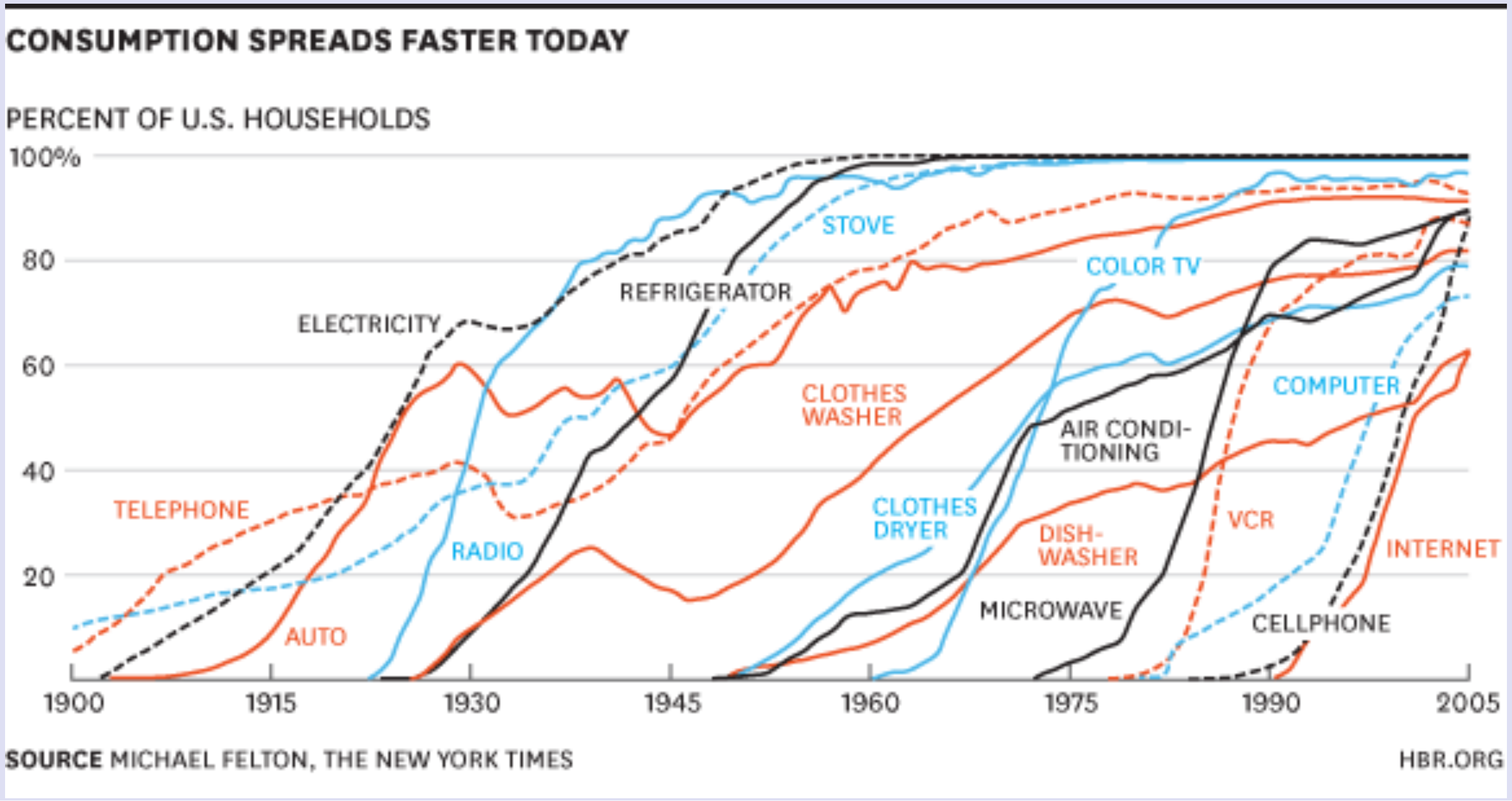
---

[a] Barrat, Barthelemy, and Vespignani 2008.



The SI model

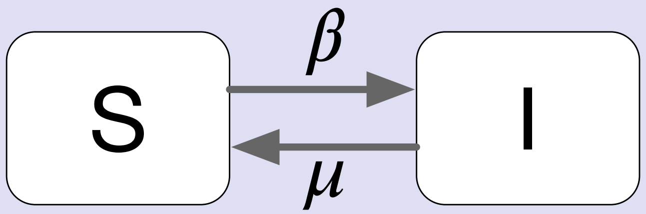
Example: technology adoption



The SIS model

Additionally to β, the SIS model requires another parameter:

μ	**recovery rate**: probability that an *Infected* individual go back to the susceptible state per unit of time.



The SIS model

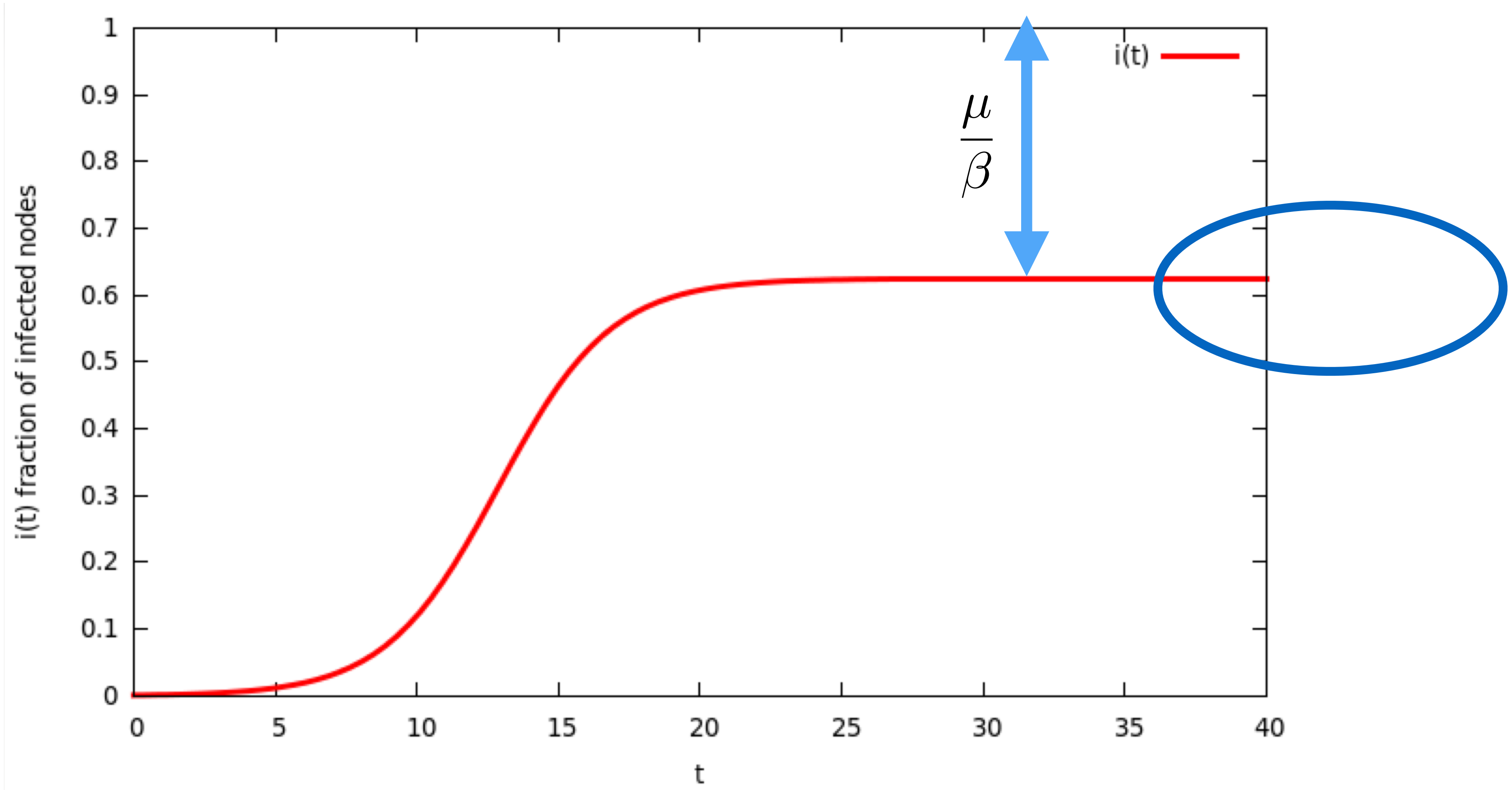
SIS - characteristics

Intuitively, the fraction of infected individuals is now reduced by those switching to the susceptible state, more formally:

$\frac{di}{dt}$	**Rate of new infection**: $\beta i(1-i) - \mu i = i(\beta - \mu - \beta i)$
$i(t)$	**Infected fraction**[a]: $\left(1 - \frac{\mu}{\beta}\right) \frac{Ce^{(\beta-\mu)t}}{1+Ce^{\beta-\mu)t}}$

For large times, $i(t) \to 1 - \frac{\mu}{\beta}$, i.e., the fraction of infected individuals stabilize around a value which depends only of parameters μ and β.

[a] Barrat, Barthelemy, and Vespignani 2008.



The SIS model

λ ratio or (R_0)

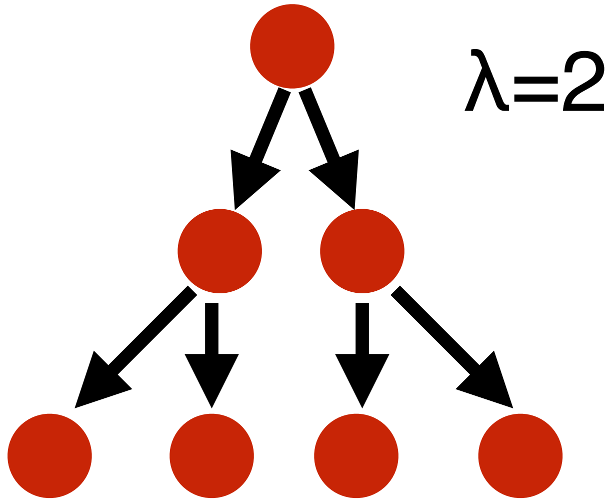
In the SIS model, an important notion is the λ ratio, also called R_0.

$$R_0 = \frac{\beta}{\mu}$$

R_0 can be understood as the average number of individuals that will be infected by an infected individual, **in a population in which all other nodes are Susceptible**. R_0 is a property of the model and **do not change with time**.
Looking at the R_0 is important in the early stage of the epidemic:

- if $R_0 > 1$, **there will be an outbreak**
- if $R_0 < 1$, **the epidemic will disappear naturally**.

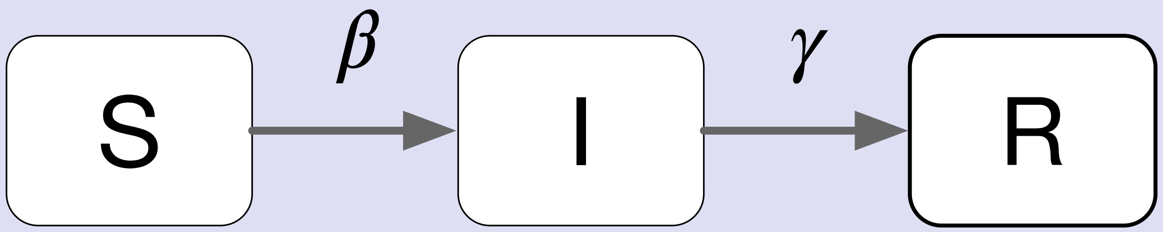
If R_0 is just above 1, the outbreak also can stop naturally by chance in the early stage.



The SIR model

Additionally to β, the SIR model requires another parameter:

γ	**recovery rate**: probability that an *Infected* individual switch to the Recovered state per unit of time.



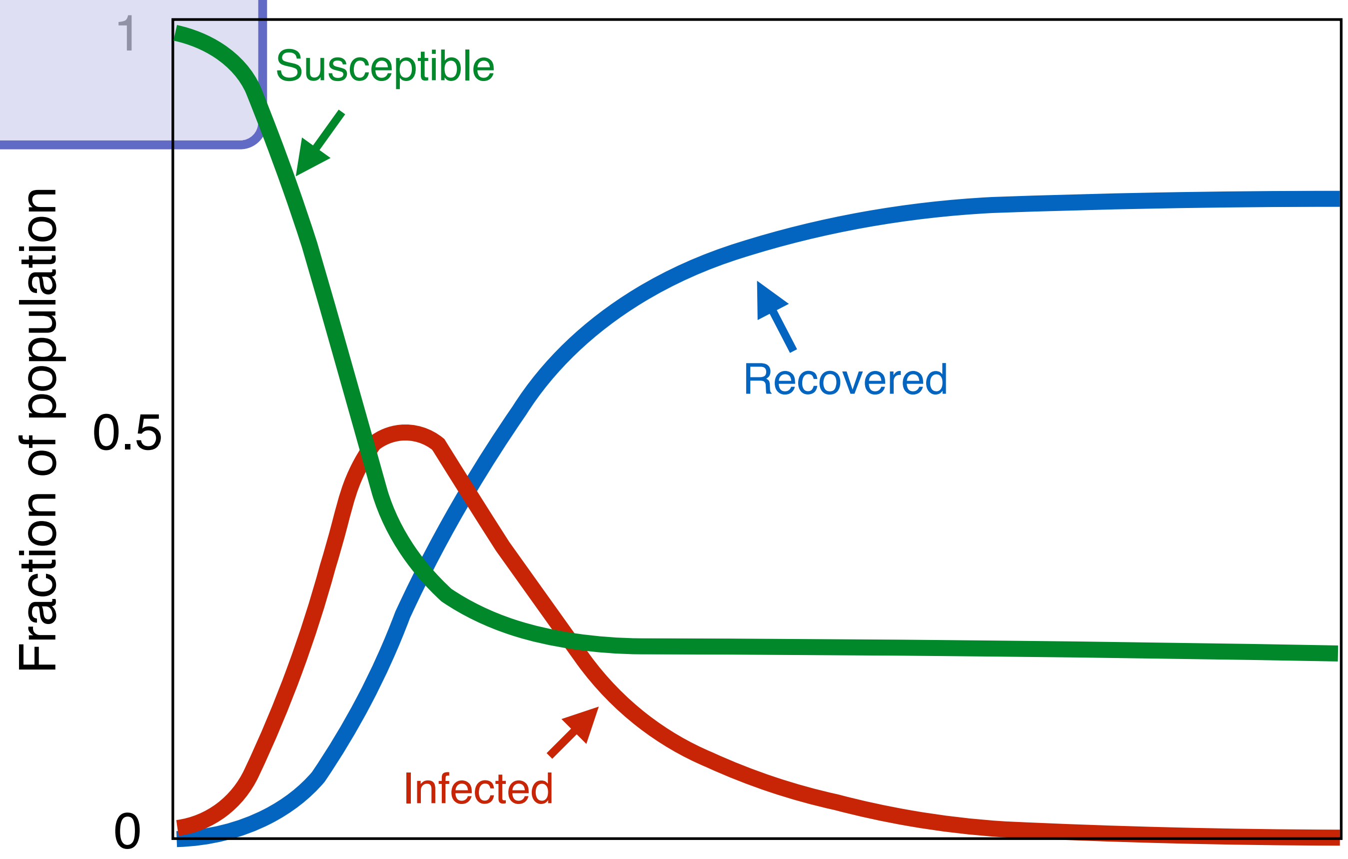
Spreading processes

SIR - characteristics

Intuitively, the fraction of infected individuals is now reduced by those switching to the recoved state, more formally:

$$\frac{ds}{dt} = -\beta is, \frac{di}{dt} = \beta is - \gamma i, \frac{dr}{dt} = \gamma i$$

- The initial steps of the outbreak still follow an exponential growth
- The fraction of infected nodes reach a peak and then decreases
- The fraction of recovered saturates below 1
- The fraction of susceptible do not necessarily reach 0
- The λ ratio is defined as $\lambda = \frac{\beta}{\gamma}$



Spreading processes

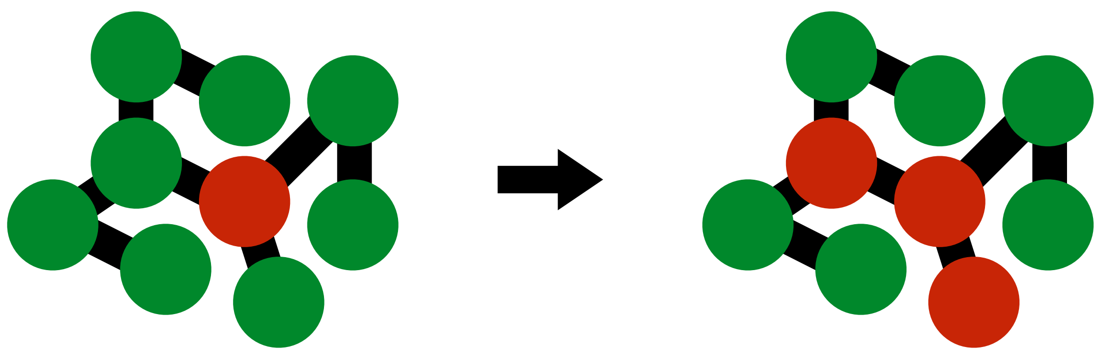
Many other models exist:
SIRD, MSIR, SEIR
SEIS, MSEIRS
Variable contact rate
Voter
Majority rule
Etc.

Check for instance:
https://ndlib.readthedocs.io/en/latest/reference/reference.html#diffusion-models

Spreading on Networks

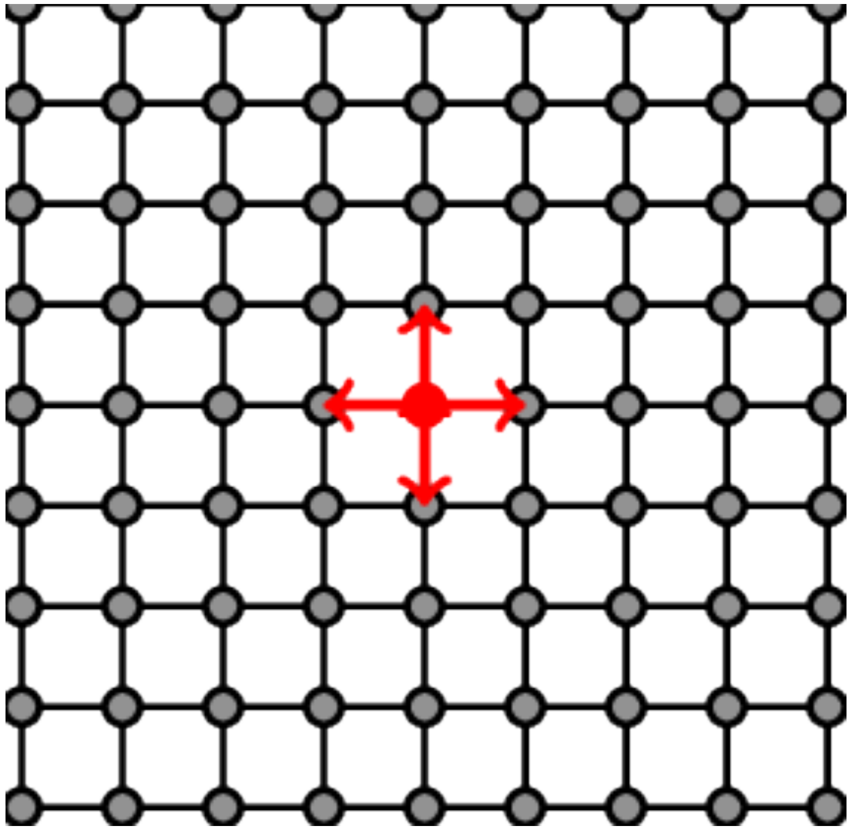
Epidemic spreading on networks

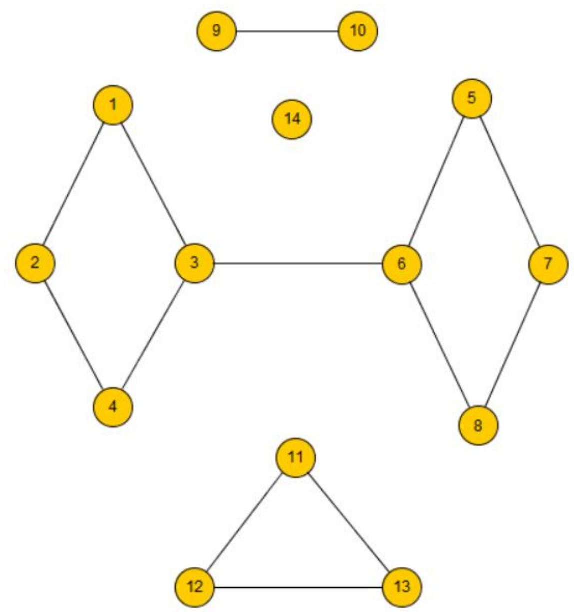
The homogeneous mixing approach is clearly unrealistic: interactions are organized in networks

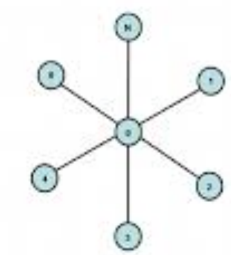


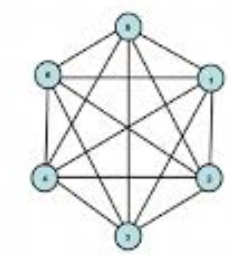
How much does it affect spreading?

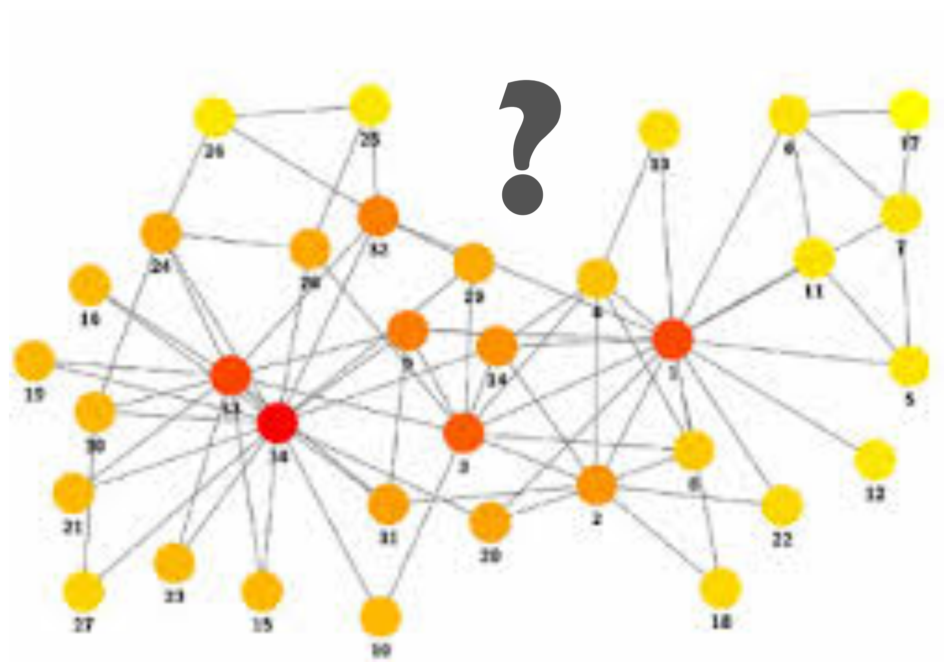
Epidemic spreading on networks











Epidemic spreading on networks

Notation change on networks

$\hat{c}$ has no meaning in networks (its role is played by the network structure), so by convention we use $\beta = \tau$: the probability for a node to infect each of its neighbor at each step.

On Networks $\qquad \beta = \tau \qquad$ On homogeneous mixing

Homogeneous networks

Homogeneous Mixing

$\frac{di}{dt}$ | **Rate of new infection**: $\frac{di}{dt} = \beta is = \beta(1-i)i$

Homogeneous Networks

If we consider an **homogeneous random network** in which all nodes have degree exactly k, then we can consider the spreading on this network as similar to the non-network models, with $\hat{c} = k$. For instance, the SI model becomes:

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

ER random graph =>approximation still holds,

$(k \approx \langle k\rangle)$

Homogeneous networks

R_0 on networks

In homogeneous or ER networks, R_0 is naturally defined as $\frac{\beta \langle k \rangle}{\mu}$
Another way to express the same thing is that, if we define $R_0 = \frac{\beta}{\mu}$, then the epidemic threshold is not equal to 1 but to $\frac{1}{\langle k \rangle}$

(Just a notation change)

Epidemic spreading on heterogeneous networks

- In degree heterogeneous networks the $k \simeq \langle k \rangle$ approximation does not hold

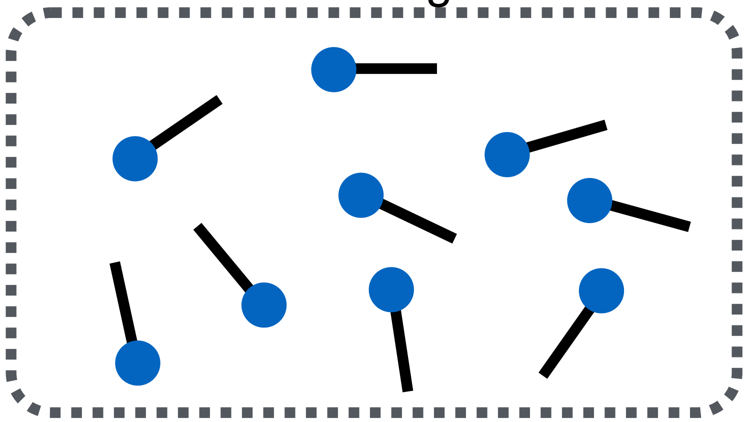
- Solution: Degree Block Approximation
  - Assumption: all nodes with the same degree are statistically equivalent
  - Look for infection/susceptible node densities in the degree groups

$$i_k$$
$$s_k$$

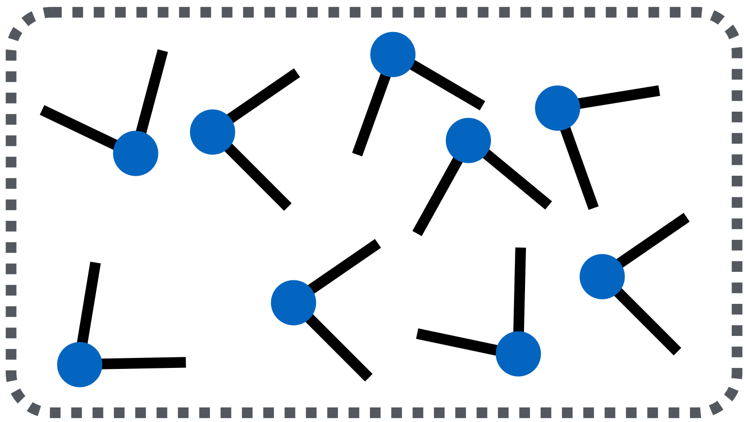
  - Calculate the global average by a sum considering the degree distribution

$$i = \sum_k P(k) i_k \qquad s = \sum_k P(k) s_k$$

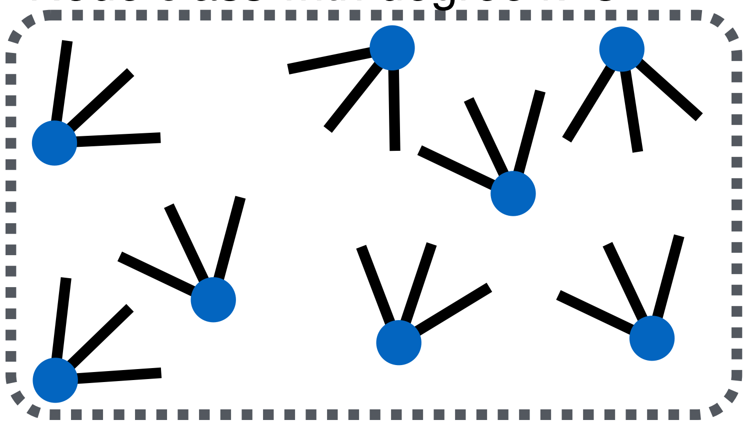
Node class with degree k=1



Node class with degree k=2



Node class with degree k=3



Epidemic spreading on heterogeneous networks

Homogeneous Networks

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

Heterogeneous Degrees - SI

For the SI model, we know that all nodes are infected in the end, but what may vary is **speed** of the process.
The speed of diffusion by degree block can be expressed as:

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

with Θ_k being the fraction of infected neighbors of a node with degree k.

Epidemic spreading on heterogeneous networks

- Due to the friendship paradox, nodes are more likely to be connected to large nodes than to small ones

$$\Theta_k = \sum_{k'} P(k'|k) i_{k'}$$

- Assume: no degree-degree correlations in the network

Number of stubs of degree k'/ Total Number of stubs (normalized by nb node)

$$P(k'|k) = \frac{k' P(k')}{\sum_{k''} k'' P(k'')} = \frac{k' P(k')}{\langle k \rangle}$$

And:

$$\Theta_k = \Theta = \frac{\sum_{k'} k' P(k') i'_k}{\langle k \rangle}$$

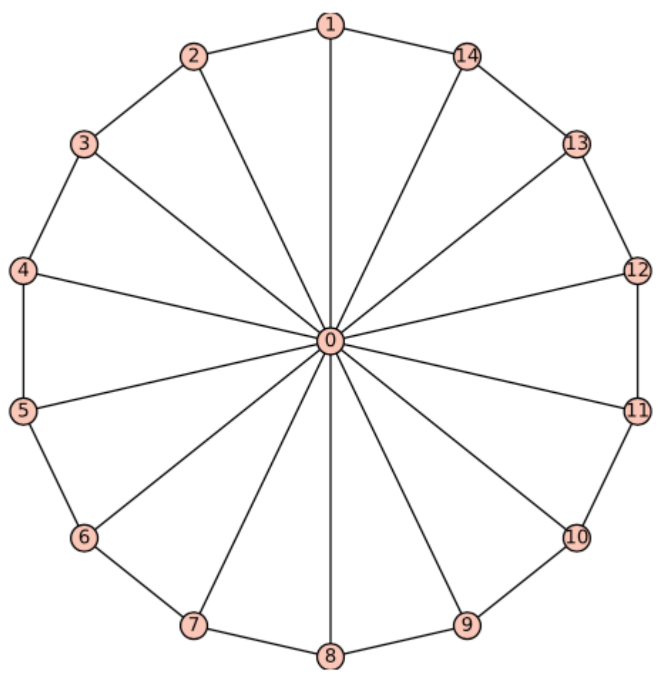
Barrat, Barthelemy, Vespignani book

SI process on heterogeneous networks

Heterogeneous Degrees - SI - time scale

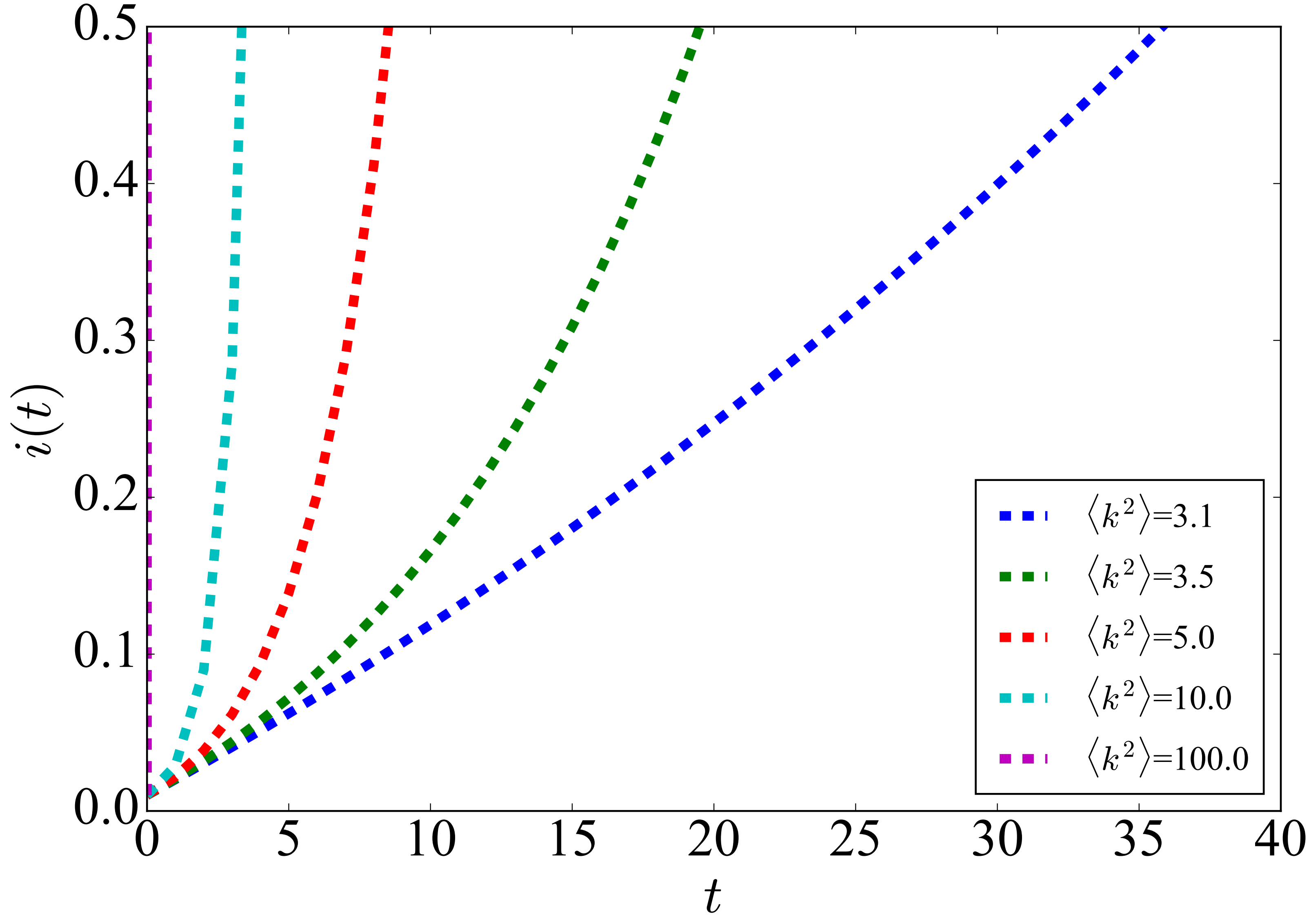
From previous equations, it can be shown[a] that the **time scale** τ of the process, i.e., a measure inversely proportional to its speed, is $\tau = \frac{\langle k \rangle}{\beta(\langle k^2 \rangle - \langle k \rangle)}$.
Thus, for a given average degree $\langle k \rangle$ and a given β, **the more heterogeneous the degrees, the faster the diffusion**.

If the degree distribution follows a power law of exponent $\alpha < 3$, we have seen that $\langle k^2 \rangle$ diverge towards infinity, thus τ tends toward 0, thus the diffusion is nearly instantaneous.



Barrat, Barthelemy, Vespignani book

SI process on heterogeneous networks



SIS process on heterogeneous networks

Heterogeneous Degrees - λ

For SIS and SIR models, it can also be shown[a] that the epidemic threshold λ (or R_0) is not reached when $\lambda = \frac{\beta\langle k\rangle}{\mu} > 1$ as in homogeneous networks, but when $\lambda > \frac{\langle k\rangle^2}{\langle k^2\rangle}$.

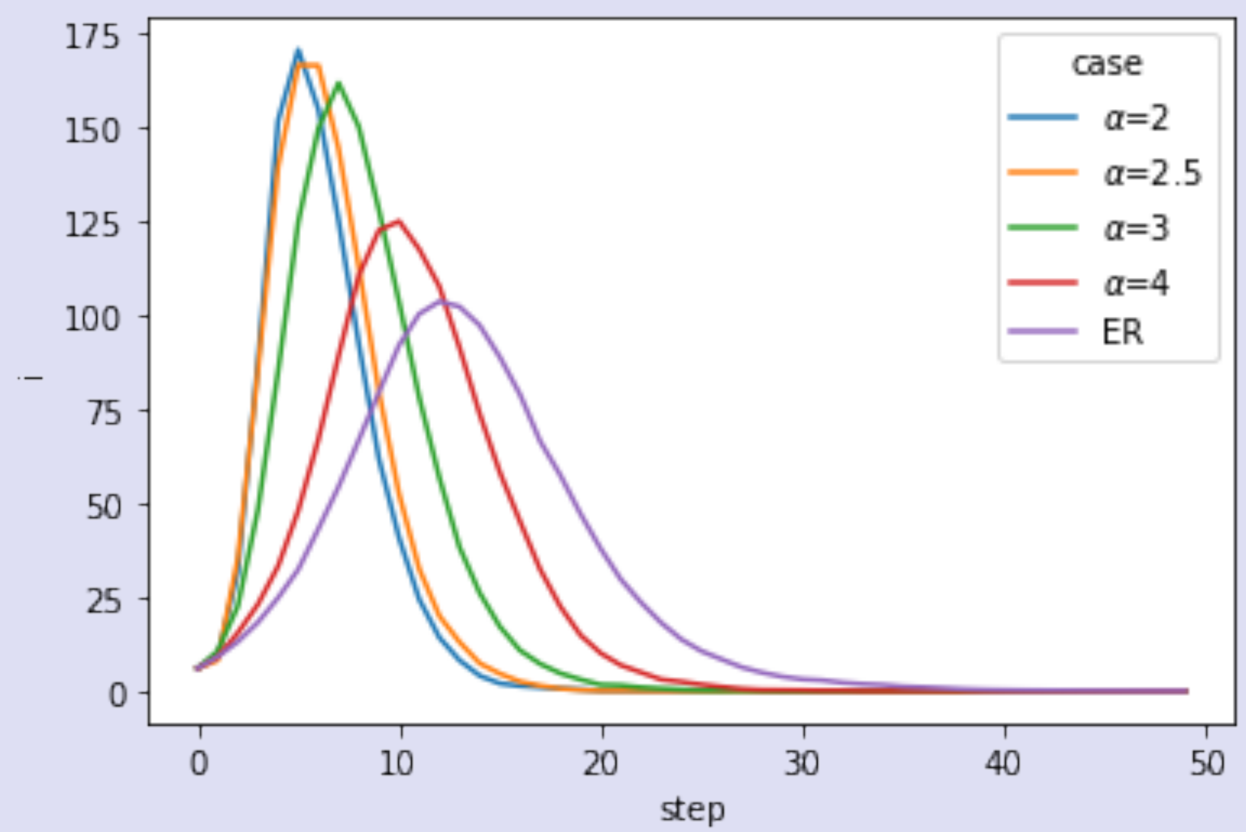
This means that in a very heterogeneous network, **an outbreak can start even if λ is very small, and below 1**. Intuitively, even if

Barrat, Barthelemy, Vespignani book

Experiments

SIR - Scale Free

In this experiment, we compare an ER network to Configuration Models with power law degree distributions.
Network parameters: $n = 1000$, $\langle k \rangle = 5$. We vary the exponent of the distribution, while keeping $\langle k \rangle = 5$ constant.
SIR parameters: $\theta = 0.2$, $\gamma = 0.5$. The initial number of infected nodes is 5, all of them in the same community structure.



The highest the exponent of the degree distribution, the faster is the diffusion.

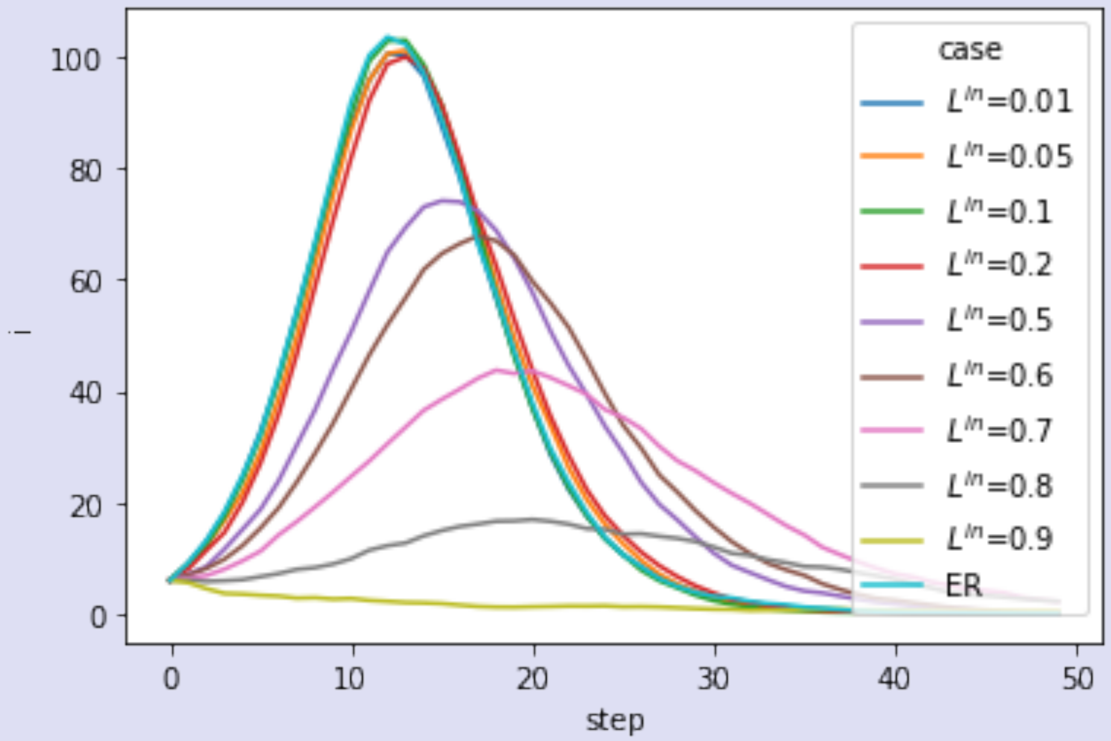
SIR - Community Structure

In this experiment, we compare an ER network to Stochastic Block Models.

Network parameters: $n = 1000, \langle k \rangle = 5$.

SBM parameters Number of blocks $|C| = 100$. We vary L^{in}, the fraction of all edges that are inside blocks. When $L^{in} = 0.01, p^{in} \approx p^{out} = 0.005$. When $L^{in} = 0.9, p^{in} = 0.5, p^{out} \approx 0.0005$

SIR parameters: $\theta = 0.2$,$\gamma = 0.5$. The initial number of infected nodes is 5, all of them in the same community structure.



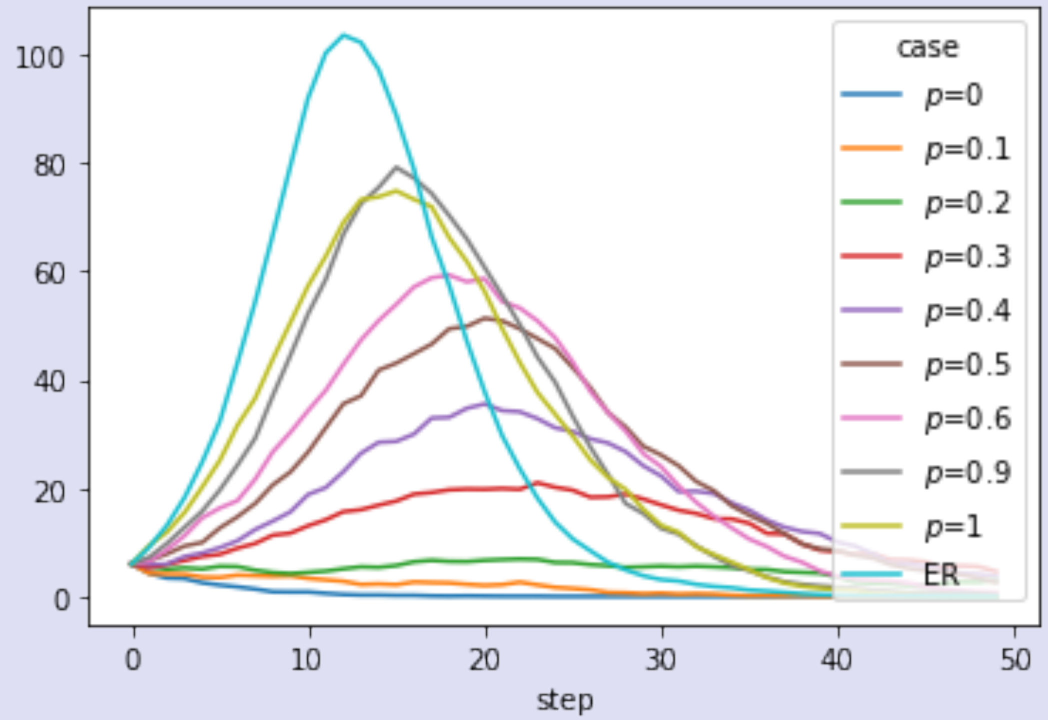
We observe that the more marked the communities, the less efficient the spreading process.

SIR - Spatial effect - WS

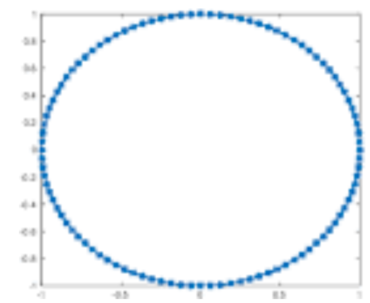
In this experiment, we compare an ER network to Watts Strogatz random graphs, varying the probability of rewiring edges. It can be understood as a model of spatial proximity: with $p = 0$, each node is connected only to its direct neighbors in the 1 dimensional space. If $p = 1$, each node is connected to exactly k random nodes.

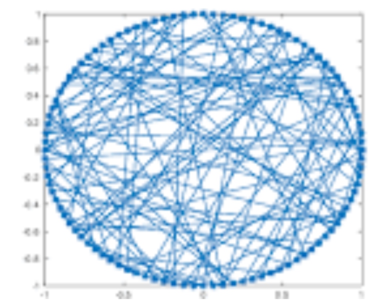
Network parameters:$n = 1000$, $\langle k \rangle = 5$

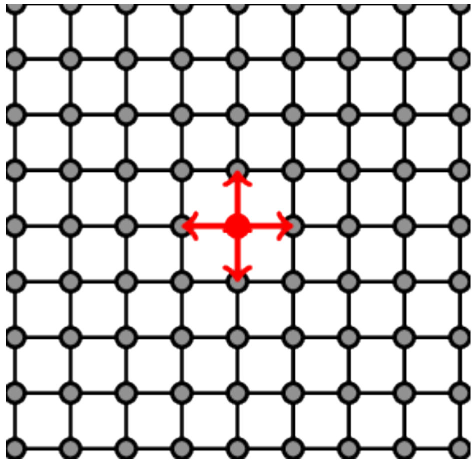
SIR parameters: $\theta = 0.2$,$\gamma = 0.5$. The initial number of infected nodes is 5, being 5 direct neighbors.



The more nodes tend to be connected to direct neighbors in space, the slower the diffusion.







Applications

Applications

- Model fitting (to better know an observed diffusion)
- Predicting future trends
- Epidemic control
  - Vaccine, etc. => Which nodes/edges to target?

- Example of strategy: friend paradox
  - Vaccine contacts of random nodes instead of random nodes

[a] Cohen, Havlin, and Ben-Avraham 2003.

OTHER MODELS

- Many other diffusion models
  - Contagious but without symptoms state
  - Propagation of information information
  - Opinion dynamics (states correspond to opinion, e.g., red/blue), diffusion rules can vary a lot
    - Majority rule: your opinion change to the one of the majority around you
    - Repeated exposition rule: each time you are exposed to an idea, you are likely to change your opinion
    - Etc.