IDENTITIES

Enlightening
Interdisciplinarity
in STEM
for Teaching

Three approaches to epidemiological modelling and simulation

Co-funded by the
Erasmus+ Programme
of the European Union

Grant Agreement n°2019-1-IT02-KA203-063184

Equation-based models

(Almost) 100 years of epidemic models

- (Kermack & McKendrick, 1927) first mathematical model that described the evolution of an epidemic

A Contribution to the Mathematical Theory of Epidemics

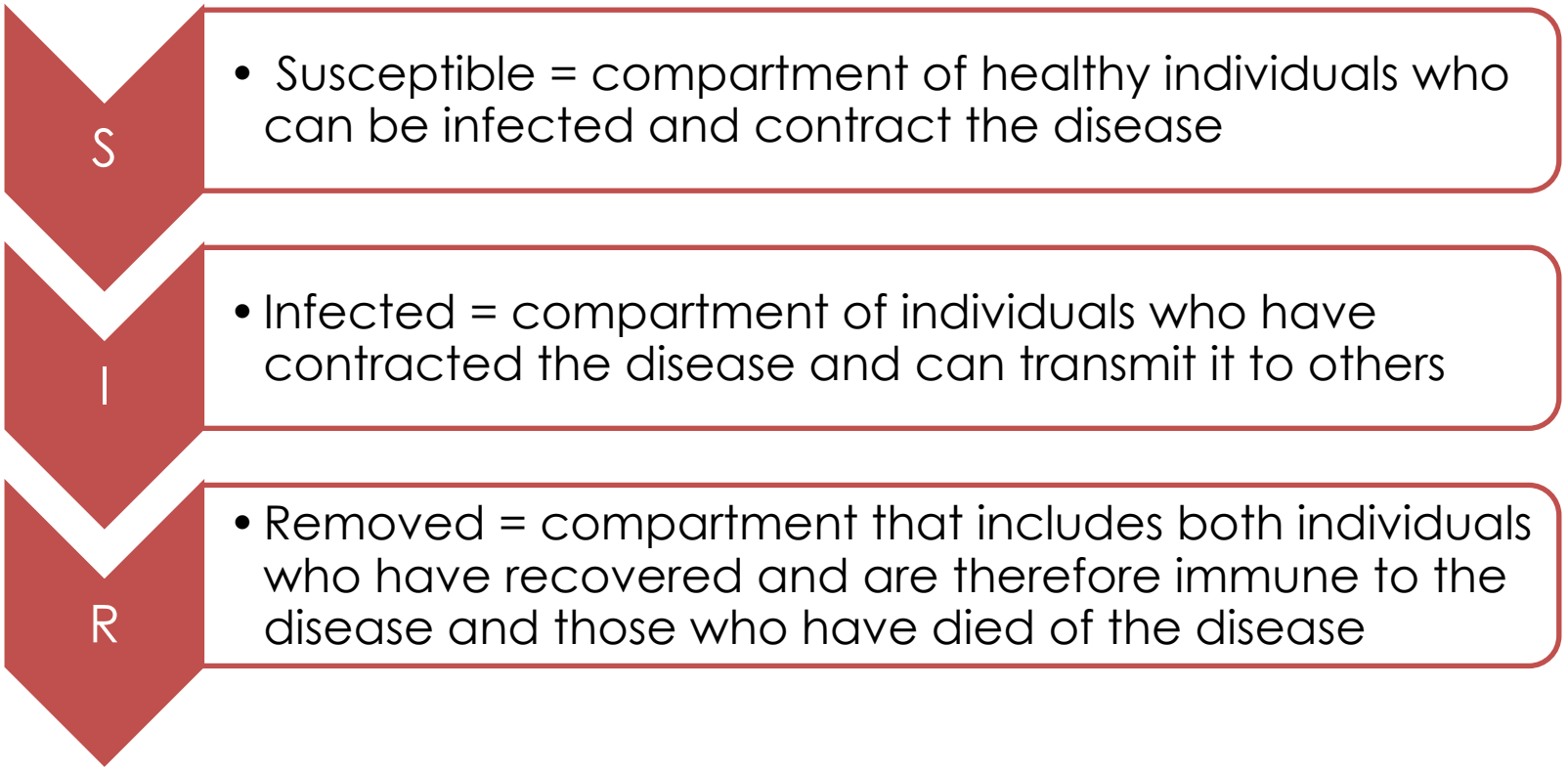
W. O. Kermack, A. G. McKendrick

Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character, Volume 115, Issue 772 (Aug. 1, 1927), 700-721.

- Numerous versions of this model to suit various epidemiological contexts (also COVID-19)
- The basic idea, however, remained that of Kermack and McKendrick's 1927 paper

The SIR model (Kermack & McKendrick, 1927)

A population is divided into three compartments



The SIR model (Kermack & McKendrick, 1927)

The time evolution of these three compartments is described by a system of three differential equations

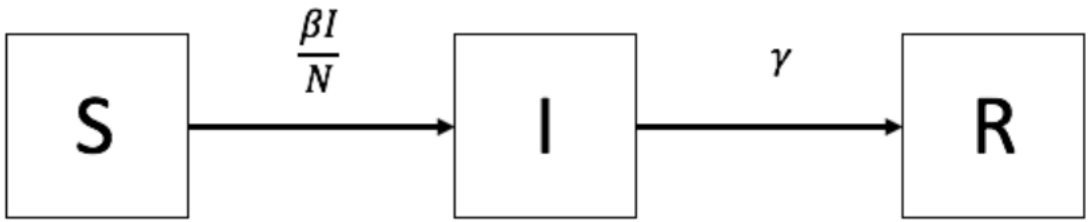
$$\frac{dS(t)}{dt} = -\frac{\beta I(t) S(t)}{N}$$

$$\frac{dI(t)}{dt} = \frac{\beta I(t) S(t)}{N} - \gamma I(t)$$

$$\frac{dR(t)}{dt} = \gamma I(t)$$

$$S(t) + I(t) + R(t) = N = cost$$

$$\frac{dS}{dt} + \frac{dI}{dt} + \frac{dR}{dt} = 0$$



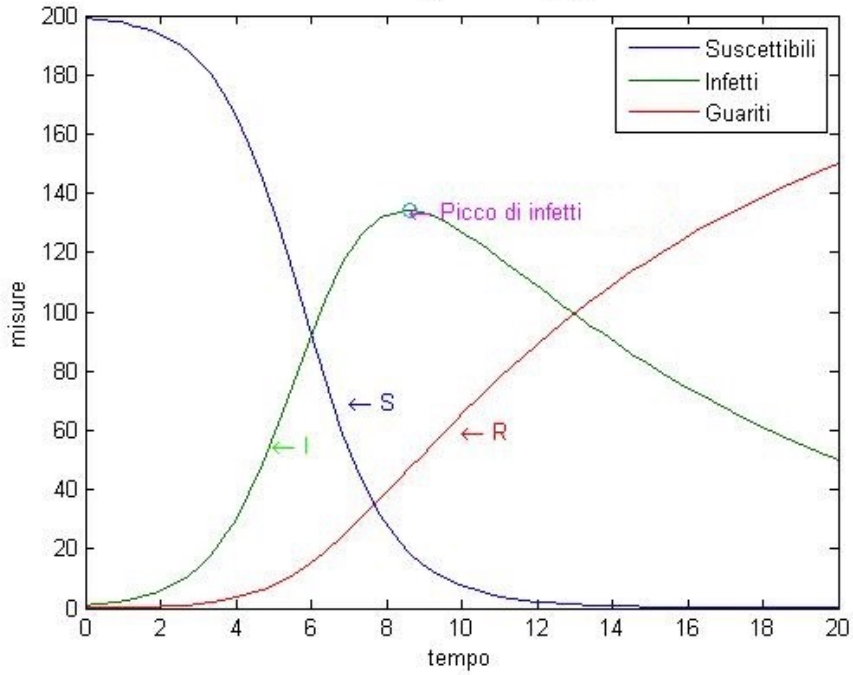
The assumptions underlying the model

- The population **is isolated**
- **Birth or death dynamics** (from causes other than illness) are neglected: the sum of individuals in the compartments is constant
- The population is completely **mixed (***full mixing* hypothesis): individuals with whom a susceptible person has contact are chosen at random in the population
- All individuals have the same **number of** contacts with other individuals in the same time period
- The number of contacts between individuals does not vary according to the state of the disease: an infected individual has the same contacts as a susceptible or cured individual
- All infected individuals transmit the disease with **equal probability**
- All susceptible individuals are equally likely to contract the disease
- The disease has no **incubation period**
- Once cured, an individual is immediately **immune to the** disease

The results of applying the model

- At the beginning of the epidemic, the number of Ss will gradually decrease due to the contagions, while the class of Is will swell for the same reason.
- As the I's increase, the greater the probability of an S being infected, and therefore the increase in I's will initially tend to accelerate

- At some point, however, some individuals will begin to move from the I-class to the R-class because they have healed or died in the meantime.
- The number of susceptible is always decreasing and the number of removed is always increasing



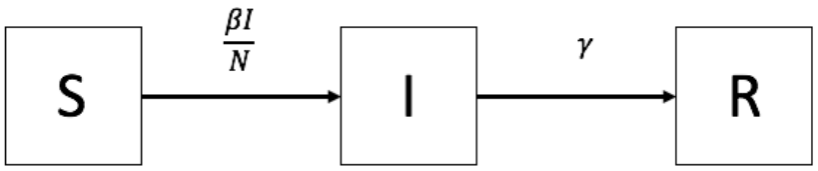
http://misterpalomar.blogspot.com/2020/02/la-matematica-delle-epidemie-parte-prima.html

How would you modify the model (the equations) to distinguish between the dead and the healed?

$$\frac{dS(t)}{dt} = -\frac{\beta I(t) S(t)}{N}$$

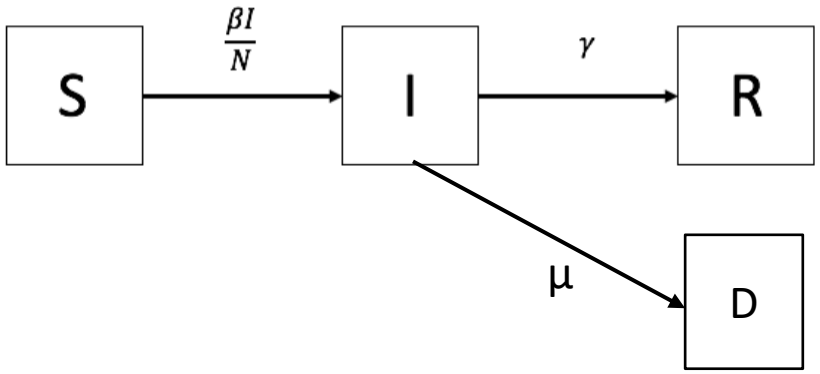
$$\frac{dI(t)}{dt} = \frac{\beta I(t) S(t)}{N} - \gamma I(t)$$

$$\frac{dR(t)}{dt} = \gamma I(t)$$



The SIRD model

$$\frac{dS}{dt} = -\beta SI; \quad \frac{dI}{dt} = \beta SI - \gamma I - \mu I = (\beta S - \gamma - \mu)I; \quad \frac{dR}{dt} = \gamma I; \quad \frac{dD}{dt} = \mu I$$



Other variations

https://www.epicx-lab.com/uploads/9/6/9/4/9694133/inserm-covid-19_report_lockdown_idf-20200412.pdf

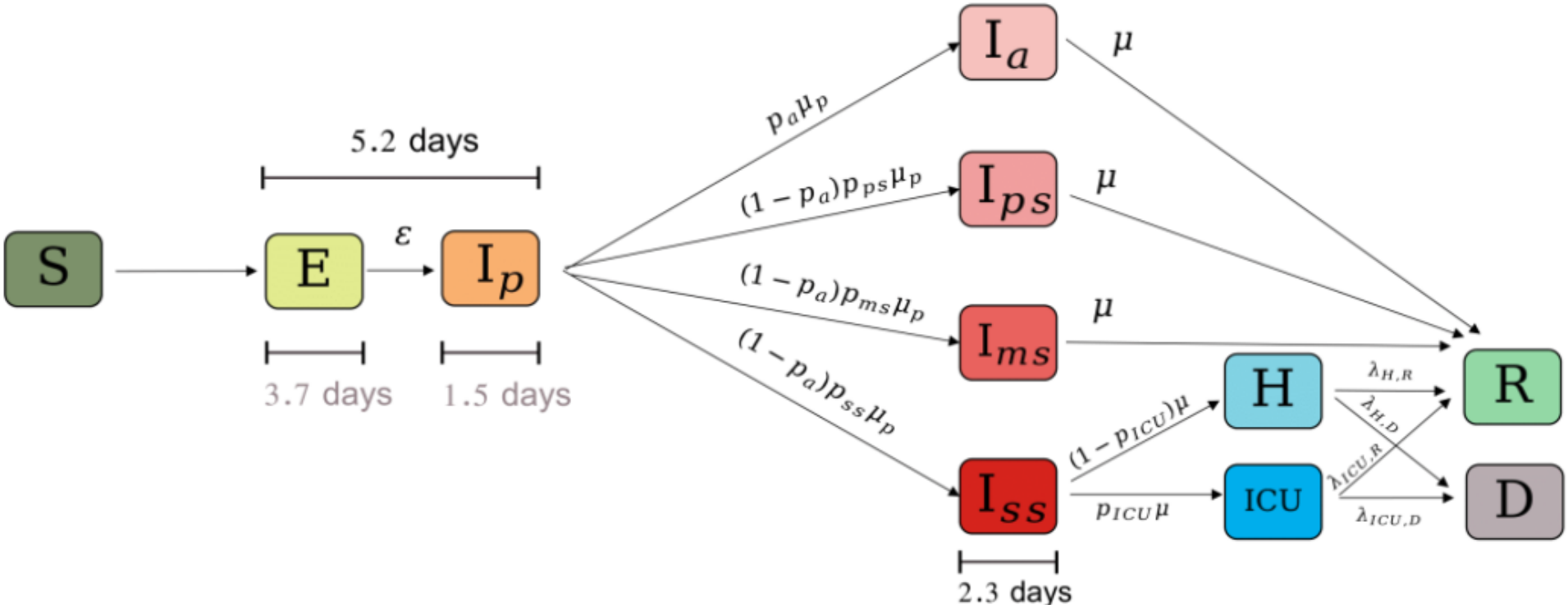


Figure 2. Compartmental model. S=Susceptible, E=Exposed, I_p= Infectious in the prodromic phase (the length of time including E and I_p stages is the incubation period), I_a=Asymptomatic Infectious, I_{ps}=Paucysymptomatic Infectious, I_{ms}=Symptomatic Infectious with mild symptoms, I_{ss}=Symptomatic Infectious with severe symptoms, ICU=severe case admitted to ICU, H=severe case admitted to the hospital but not in intensive care, R=Recovered, D=Deceased.

The SIR model as an equation-based model

- The model is already expressed as a system of differential equations
- Due to their non-linearity, it is very complex to solve them analytically in an exact manner, i.e. one cannot express S(t), I(t) and R(t) as functions of t, S(0), I(0) and R(0)
- Resolution is through numerical integration methods

Simulating the SIR model: approach by integration

- 3 requirements for each numerical integration method:
  - Equations to be integrated
  - Initial conditions
  - Increasing sequence of discrete time values

METHOD OF INTEGRATION

Solutions: sequence of values of unknowns at specified time values

Simulating the SIR model: approach by integration

```
betac = 1
betai = 0.8
gamma = 0.3
N = 1000
```

Initialisation of model parameters

```
def dX_dt(X, t=0):
    return array([ - betac*betai*X[0]*X[1]/N ,
                   betac*betai*X[0]*X[1]/N - gamma*X[1] ,
                   gamma*X[1] ])

from scipy import integrate
t = linspace(0, 50, 50)

X0 = array([N-1, 1, 0])

X = integrate.odeint(dX_dt, X0, t)
```

Simulating the SIR model: approach by integration

```
betac = 1
betai = 0.8
gamma = 0.3
N = 1000
```

```
def dX_dt(X, t=0):
    return array([ - betac*betai*X[0]*X[1]/N ,
                   betac*betai*X[0]*X[1]/N - gamma*X[1] ,
                   gamma*X[1] ])
```

Writing the model equations. X is a three-component vector:
X[0] = S,
X[1]=I,
X[2]=R

```
from scipy import integrate
t = linspace(0, 50, 50)

X0 = array([N-1, 1, 0])

X = integrate.odeint(dX_dt, X0, t)
```

Simulating the SIR model: approach by integration

```
betac = 1
betai = 0.8
gamma = 0.3
N = 1000

def dX_dt(X, t=0):
    return array([ - betac*betai*X[0]*X[1]/N ,
                   betac*betai*X[0]*X[1]/N - gamma*X[1] ,
                   gamma*X[1] ])
```

```
from scipy import integrate
t = linspace(0, 50, 50)
```

Setting a **discrete** time interval over which to calculate the evolution of populations
t = [0,1,2,3,4,5,6,...,50]

```
X0 = array([N-1, 1, 0])

X = integrate.odeint(dX_dt, X0, t)
```

Simulating the SIR model: approach by integration

```
betac = 1
betai = 0.8
gamma = 0.3
N = 1000

def dX_dt(X, t=0):
    return array([ - betac*betai*X[0]*X[1]/N ,
                   betac*betai*X[0]*X[1]/N - gamma*X[1] ,
                   gamma*X[1] ])

from scipy import integrate
t = linspace(0, 50, 50)

X0 = array([N-1, 1, 0])

X = integrate.odeint(dX_dt, X0, t)
```

Setting of model initial conditions: S(0) = N-1; I(0) = 1; R(0) = 0

Simulating the SIR model: approach by integration

```
betac = 1
betai = 0.8
gamma = 0.3
N = 1000

def dX_dt(X, t=0):
    return array([ - betac*betai*X[0]*X[1]/N ,
                   betac*betai*X[0]*X[1]/N - gamma*X[1] ,
                   gamma*X[1] ])

from scipy import integrate
t = linspace(0, 50, 50)

X0 = array([N-1, 1, 0])

X = integrate.odeint(dX_dt, X0, t)
```

Method for numerical integration of differential equations over a specified discrete interval

Simulating the SIR model: approach by integration

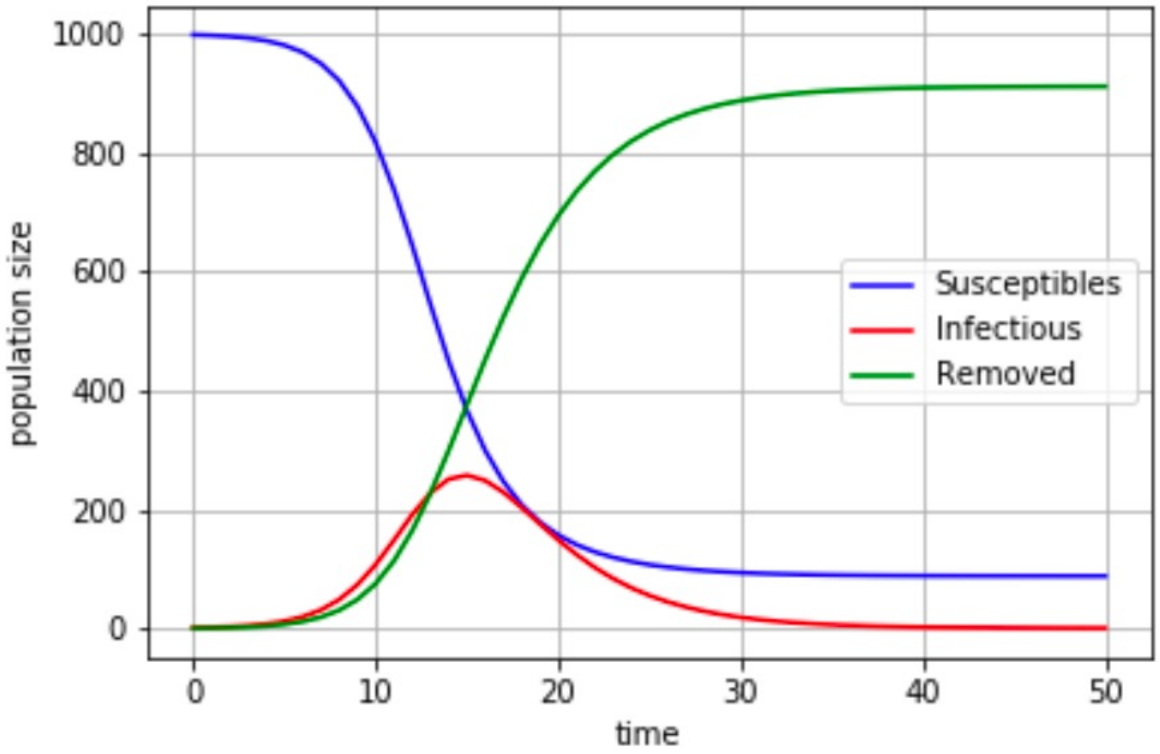


Figure 3. Results of the simulation of the computational equation-based model with odeint integration method: in blue the evolution of susceptible population size, in red the infectious, in green the removed ($\beta = 0.8, \gamma = 0.3, N = 1000, S(0) = 999, I(0) = 1, R(0) = 0, time = [0, 1, \dots, 50]$).

The SIR model expressed by finite differences

Discretisation of the original equations leads to a reformulation of the differential equations as finite difference equations

$$\frac{dS}{dt} = \frac{S_{t+1} - S_t}{\Delta t} \rightarrow S_{t+1} = S_t + \frac{dS}{dt} \cdot \Delta t$$

$$S_{t+1} = S_t - \frac{\beta S_t I_t}{N} \Delta t$$

$$I_{t+1} = I_t + \left[\frac{\beta S_t I_t}{N} - \gamma I_t \right] \Delta t$$

$$R_{t+1} = R_t + \gamma I_t \Delta t$$

Simulating the SIR model: finite difference approach

```
S = [N-1].
I = [1]
R = [0]
```

I initialise three vectors: N-1

I 1

R 0

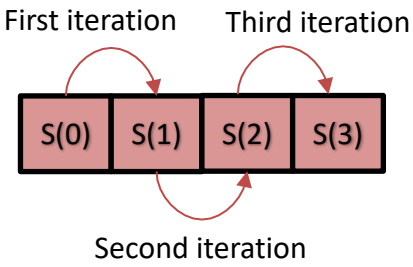
```
def population_t1():
    susceptibles_t0 = S[-1].
    infected_t0 = I[-1].
    recovered_t0 = R[-1]
    susceptibles_t1 = susceptibles_t0 - beta*infected_t0*susceptibles_t0/N
    infected_t1 = infected_t0 + beta*infected_t0*susceptibles_t0/N - gamma*infected_t0
    recovered_t1 = recovered_t0 + gamma*infected_t0
    S.append(susceptibles_t1)
    I.append(infected_t1)
    R.append(recovered_t1)
    return S, I, R

for i in range(t):
    SIR = population_t1()
    S = SIR[0].
    I = SIR[1].
    R = SIR[2].
```

Simulating the SIR model: finite difference approach

```
S = [N-1].
I = [1]
R = [0]

def population_t1():
    susceptibles_t0 = S[-1].
    infected_t0 = I[-1].
    recovered_t0 = R[-1]
    susceptibles_t1 = susceptibles_t0 - beta*infected_t0*susceptibles_t0/N
    infected_t1 = infected_t0 + beta*infected_t0*susceptibles_t0/N - gamma*infected_t0
    recovered_t1 = recovered_t0 + gamma*infected_t0
    S.append(susceptibles_t1)
    I.append(infected_t1)
    R.append(recovered_t1)
    return S, I, R
```




```
for i in range(t):
    SIR = population_t1()
    S = SIR[0].
    I = SIR[1].
    R = SIR[2].
```

You populate the vectors one component at a time
Each new component is calculated taking the last existing component as the initial condition

Simulation with finite differences

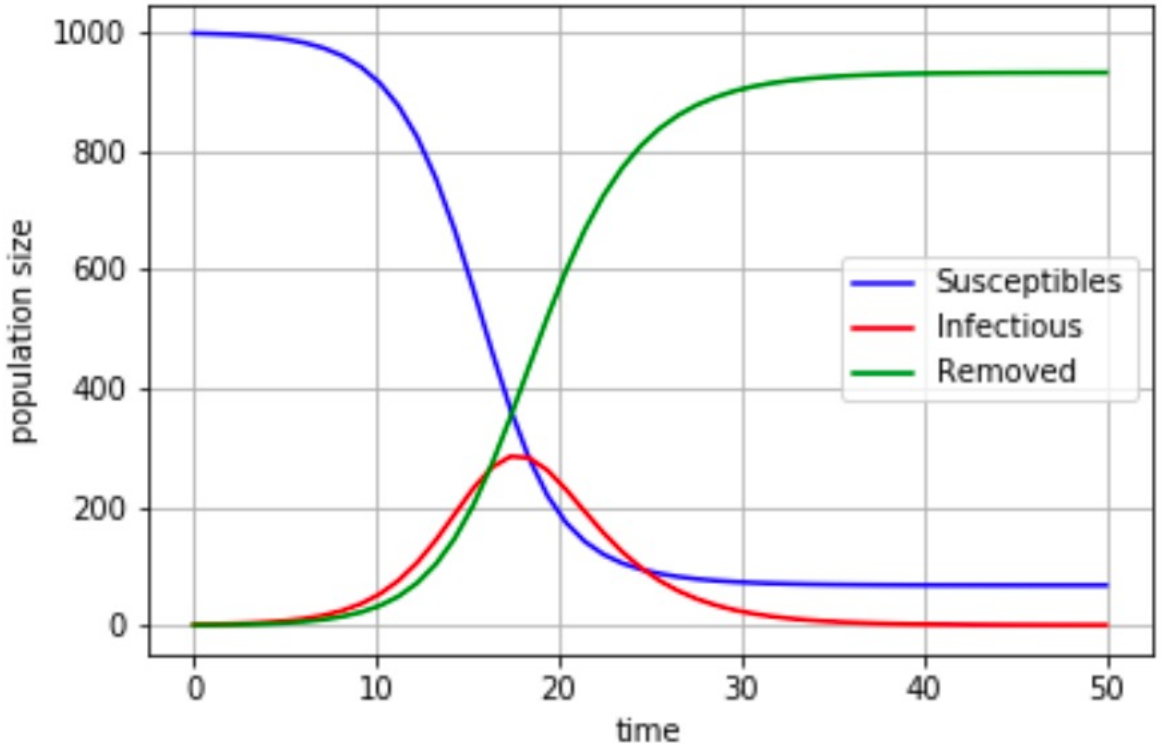


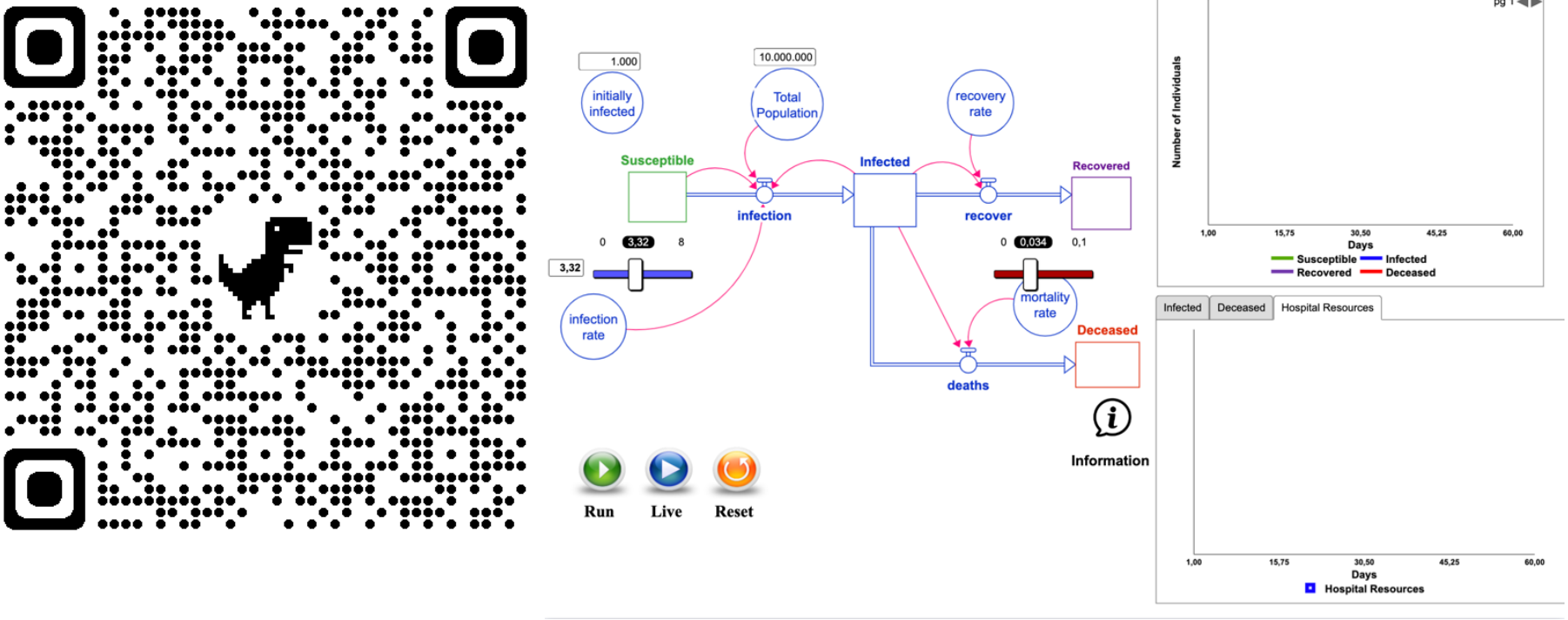
Figure 5. Results of the simulation of the computational equation-based model with finite difference integration method: in blue the evolution of susceptible population size, in red the infectious, in green the removed ($\beta = 0.8, \gamma = 0.3, N = 1000, S(0) = 999, I(0) = 1, R(0) = 0, time = [0, 1, \dots, 50]$).

Two equation-based approaches

- The approaches seen so far are both equation-based
- The first simulated the model **by integrating the system of differential equations** using a numerical method (odeint from scipy, Python library)
- The second simulated the model by making an **a priori discretisation of the equations**, making them finite difference equations and calculating the populations at each instant based on the values at the previous time instant
- Both approaches to the simulation of the SIR model follow naturally from the formulation of the model itself
- In neither case is the individual agent traceable (population variations)

An equation-based simulation for the SIR model

https://sites.google.com/site/biologydarkow/physiology/covid-19-sir-simulation

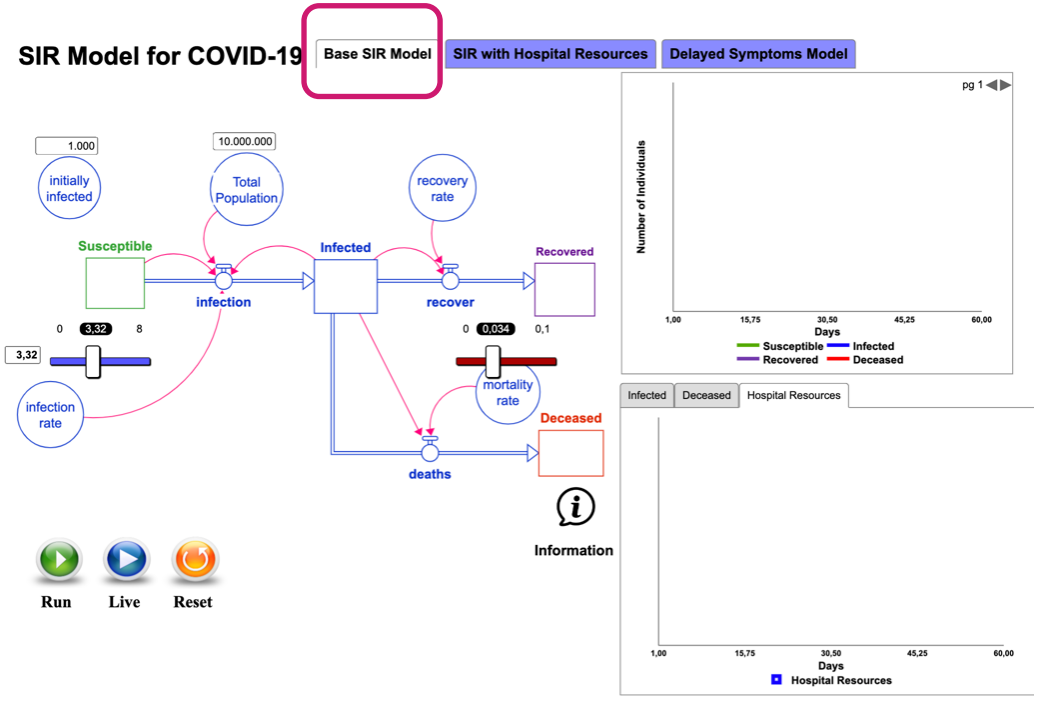


https://docs.google.com/document/d/1LHka1z6DObjBhiLv2EInMO0d86bSqazdyF-DRLmG5RM/edit#

An equation-based simulation for the SIR model

Examine the basic model:

- Which variable/parameter would **social distancing** directly impact on in the model? Why?
- Will social distancing increase or decrease this variable/parameter? Run several simulations and describe how social distancing affects the patterns of system evolution.



Agent-based models

How to simulate the SIR model with an agent-based approach?

$$\frac{dS}{dt} = -S\,P(susceptible\ becomes\ infectious) =$$
$$= -S\,P(susceptible\ contacts\ infectious)\,P(infectious\ infects\ susceptible) =$$
$$= -S\,[mean\ number\ of\ contacts\ for\ individual * P(the\ contact\ is\ infectious)] * \beta_i =$$
$$= -S\left[\beta_c \frac{I}{N}\right]\beta_i = -S\beta_i\beta_c \frac{I}{N} = -\frac{\beta IS}{N} \qquad (16)$$

The probability that a susceptible becomes infectious ($P(susceptible\ becomes\ infectious)$) depends on the probability that a susceptible meets an infectious ($P(susceptible\ contacts\ infectious)$), and on the probability that a susceptible becomes infectious after the contact ($P(infectious\ infect\ susceptible)$)

How to simulate the SIR model with an agent-based approach?

$$\frac{dR}{dt} = IP(infectious\ is\ removed) = \gamma I = \frac{1}{d} I$$

the recovery-or-death rate γ can be modelled as the reciprocal of the average duration of infectivity d. This is also equivalent to the assumption that the duration of time spent by an individual in the infectious state is a random variable with an exponential distribution (Sterman, 2000; p. 305).

How to simulate the SIR model with an agent-based approach?

$$\frac{dI}{dt} = -\frac{dS}{dt} - \frac{dR}{dt} = \frac{\beta IS}{N} - \gamma I = \frac{\beta IS}{N} - \frac{1}{d} I$$

For deriving the infectious compartment, we can apply the boundary condition

Agent-based approach to the SIR model

What has changed compared to before?

The interpretation of the variables in the equations underlying the model has changed

EQUATION-BASED	**AGENT-BASED**
Frequency/rate of infection and removal	Contacts of individuals + probability of infection + duration of the infectivity period
Big eye on population changes	A small eye on what happens to agents

Agent-based approach to the SIR model

The agent-based approach is based on:

- Creating a number of agents
- Associate these agents with the appropriate variables representing their state during the evolution of the system
- Codifying rules of behaviour for each agent over time

Possibly,

- Carrying out aggregate measurements on the system to facilitate the presentation of results

- Equal to the size of the population of interest

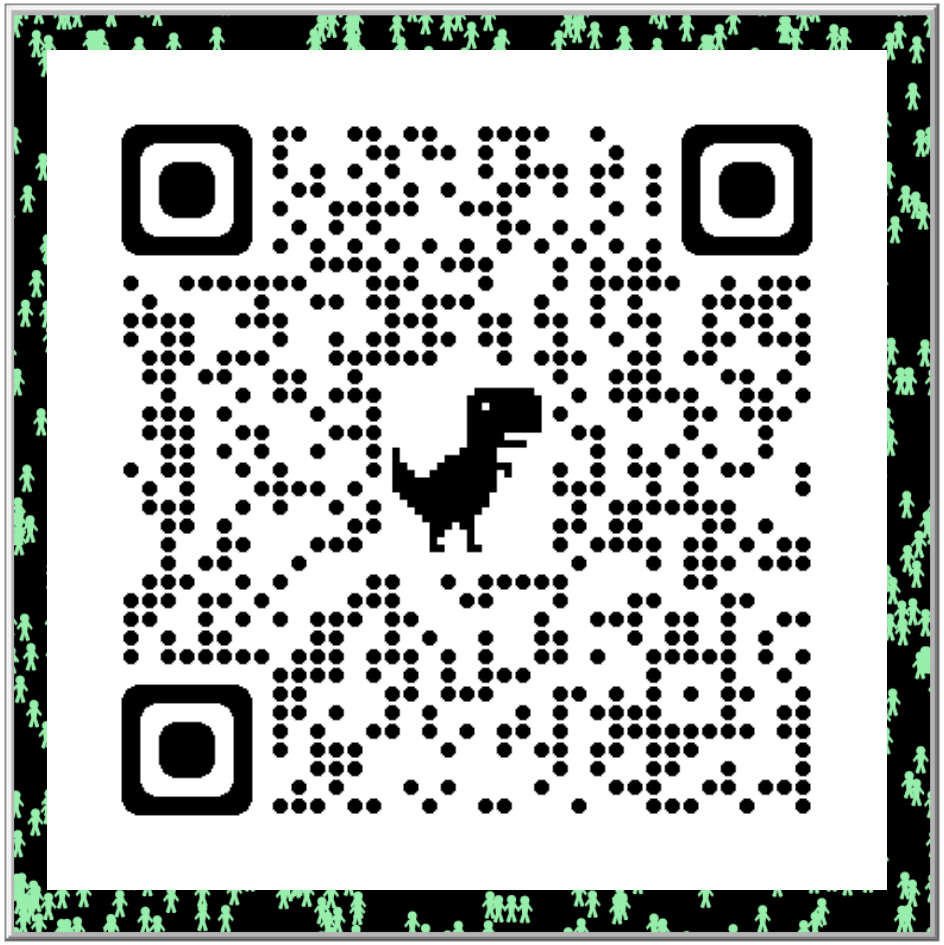
- Any agent can be susceptible, infected or cured and may have become infected at some time

- Next slide

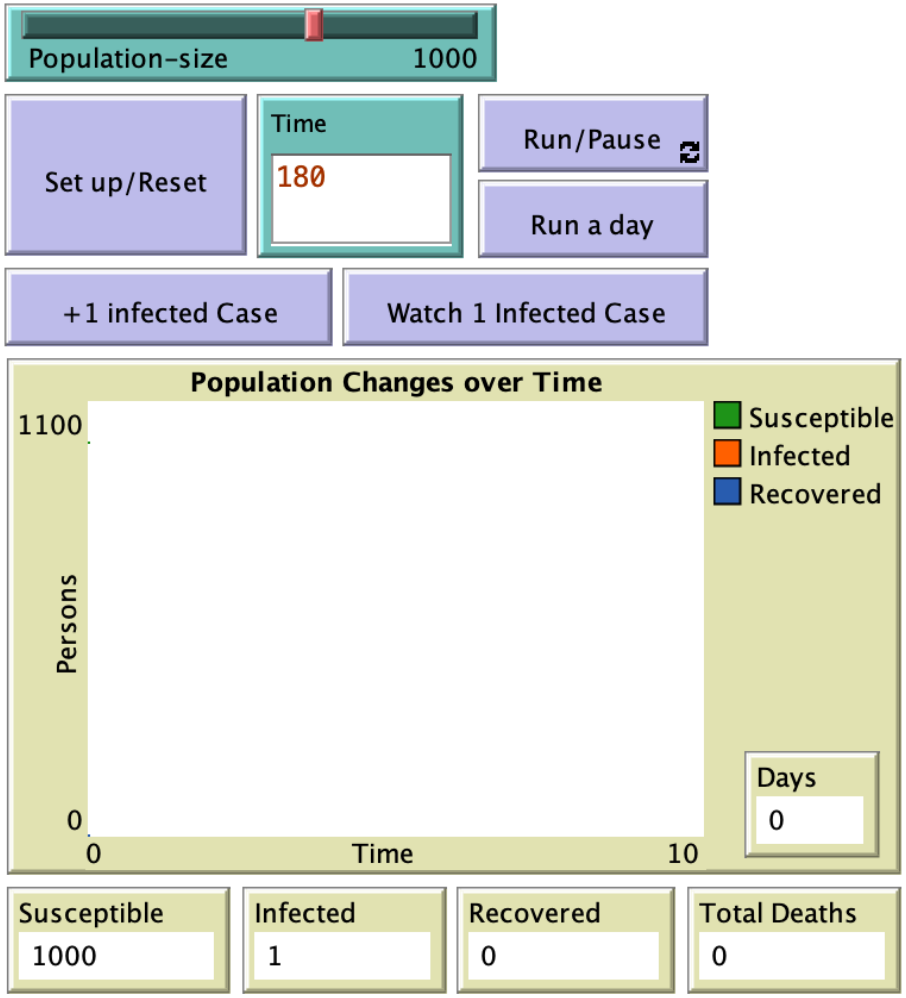
- Keep information on the total number of susceptible, infected and cured

How can we model the spread of a virus in a population using an agent-based approach?

What kind of agents?
What characteristics should they have?
How would they behave?



Green: Susceptible |Orange: Infected |Blue: Recovered



http://modelingcommons.org/browse/one_model/6279#model_tabs_browse_nlw

Simulation with agent-based approach

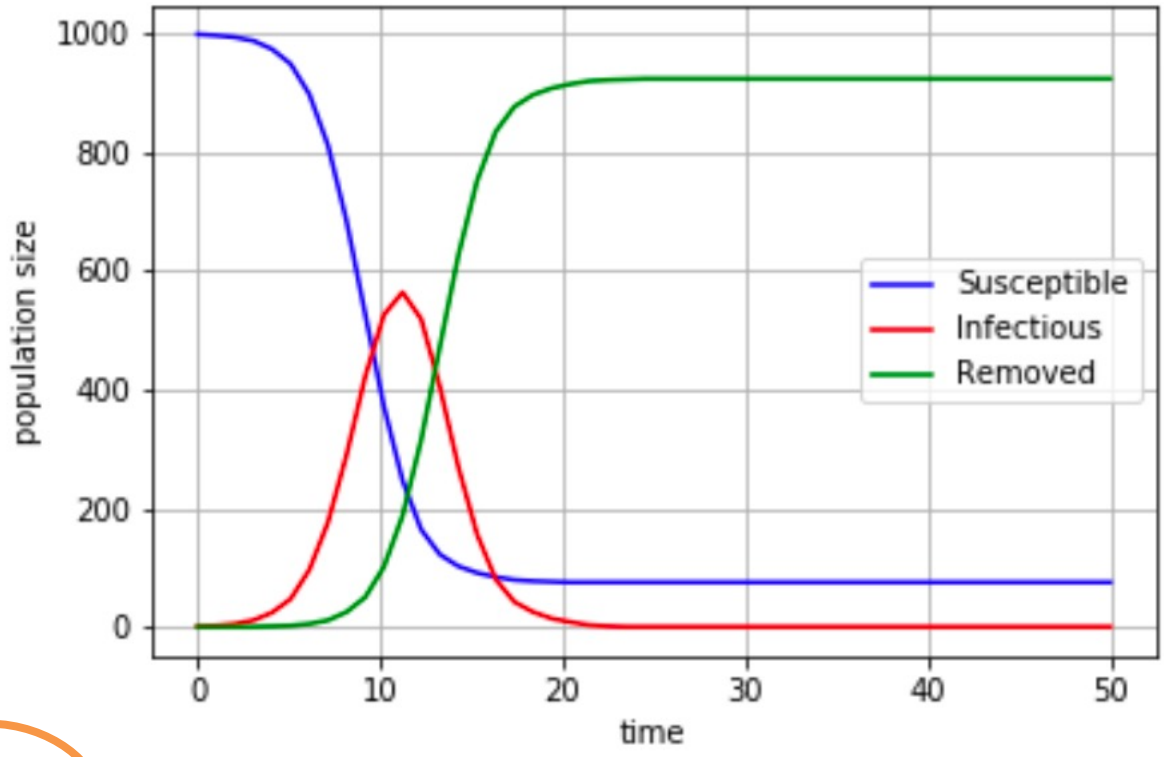
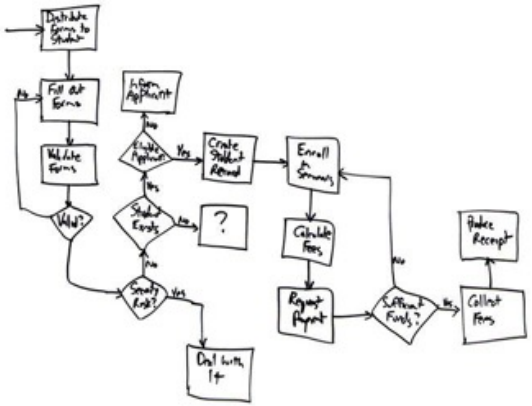


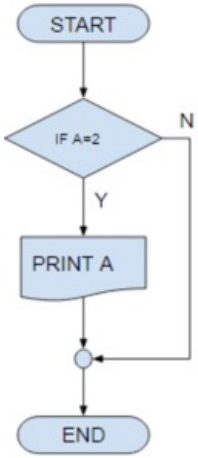
Figure 8. Average results of 100 simulation runs of the computational agent-based model: in blue the evolution of susceptible population size, in red the infectious, in green the removed ($\beta_i = 0.8$, $\beta_c = 1, \gamma = 0.3, N = 1000, S(0) = 999, I(0) = 1, R(0) = 0, time = [0, 1, \dots, 50]$).

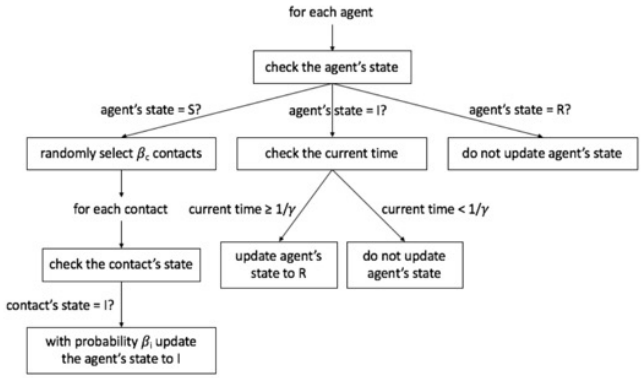
Analysing the simulation (graphical interface and code), schematise in the form most congenial to you the rules of behaviour of the agents and the functioning of the simulation itself

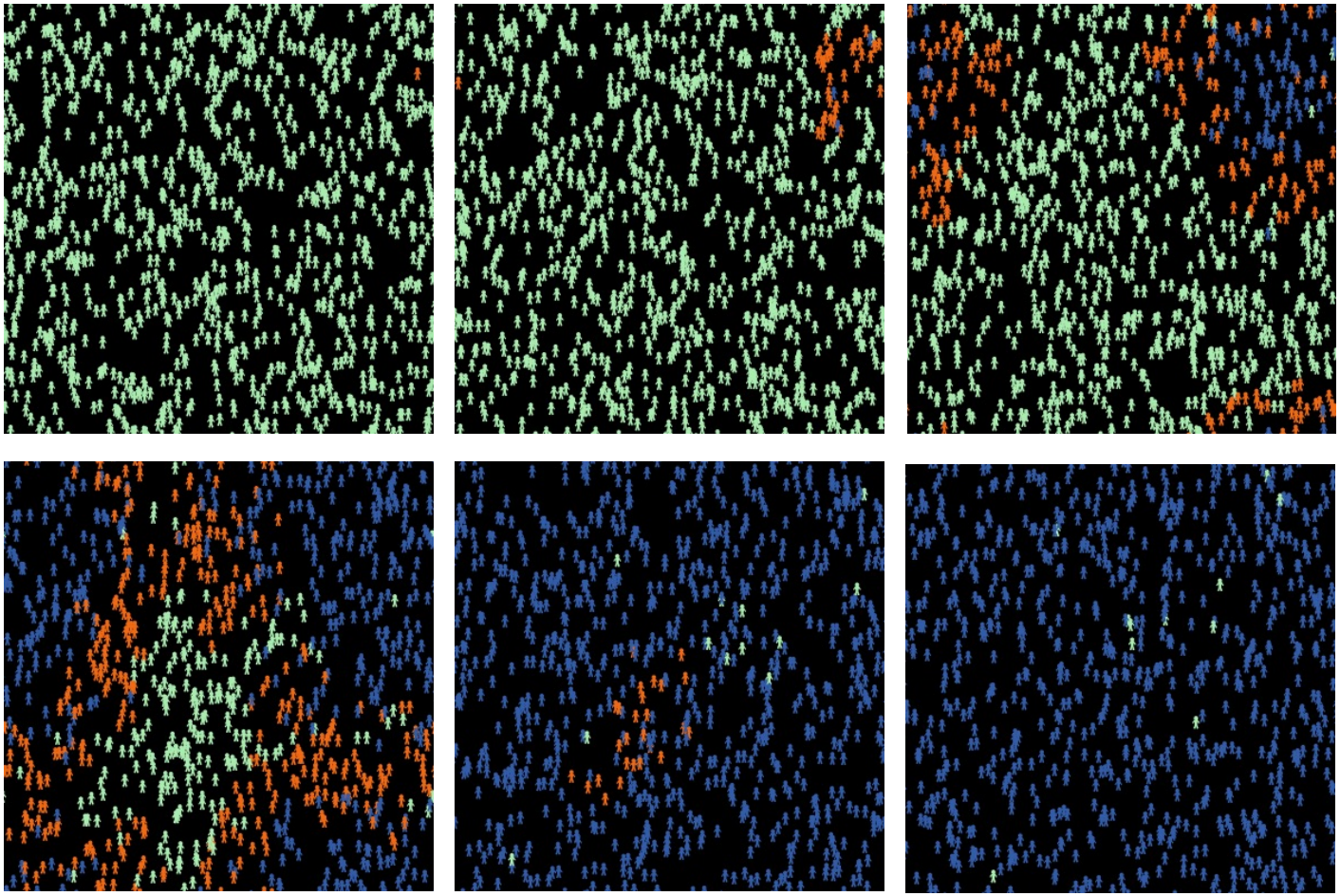
When in doubt, the answer is one and only one:
NetLogo dictionary https://ccl.northwestern.edu/netlogo/docs/dictionary.html

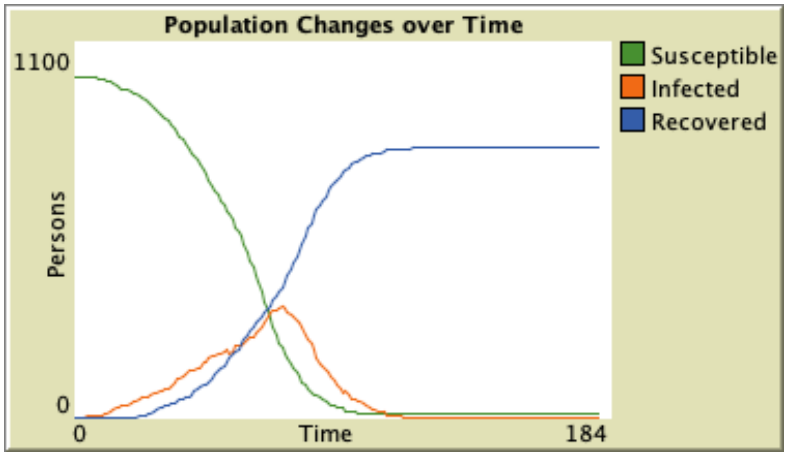
Name of the procedure	Scope of the procedure
	Setup procedures
`setup`	General setup settings: clear plots, create the grid
`setup-turtles`	Agent settings: create agents, position them in the grid
`setup-culture-max`	Assigning a value to the culture with maximum traits values
`setup-agent-culture`	Assigning a random culture to each agent
`setup-agent-culture-color`	Setting the color to the agent according to its culture
	Main procedure
`go`	Execute all the local and graph procedures in their order until there are active agents
	Local procedures
`cultural-interaction`	Agents look for a neighbor to interact with
`overlap_between`	Reporting overlap between two agents (range from 0 to F)
`culturally_interacting [target_turtle neighbor_turtle]`	Interaction between a target agent and its selected neighbor



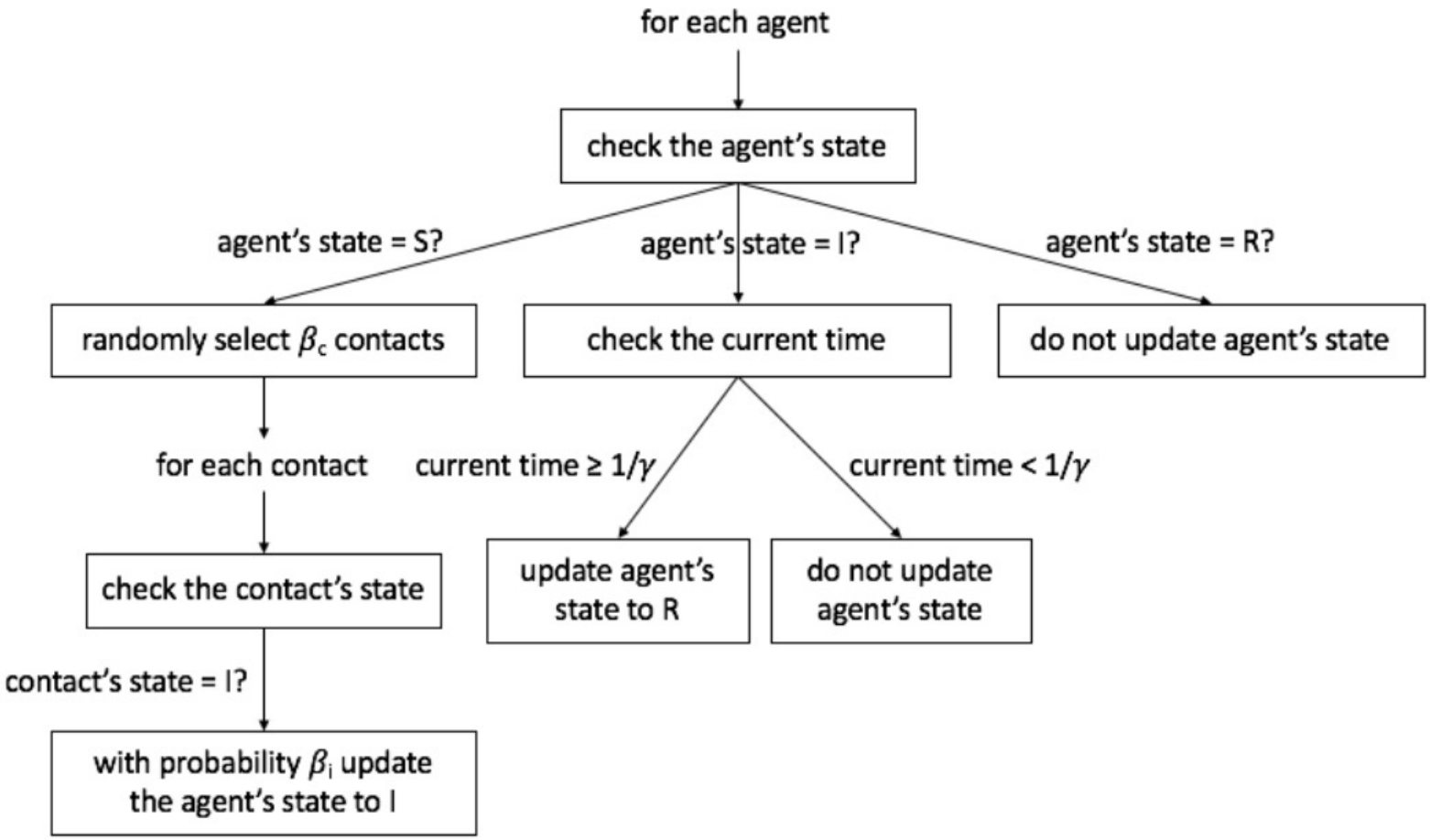


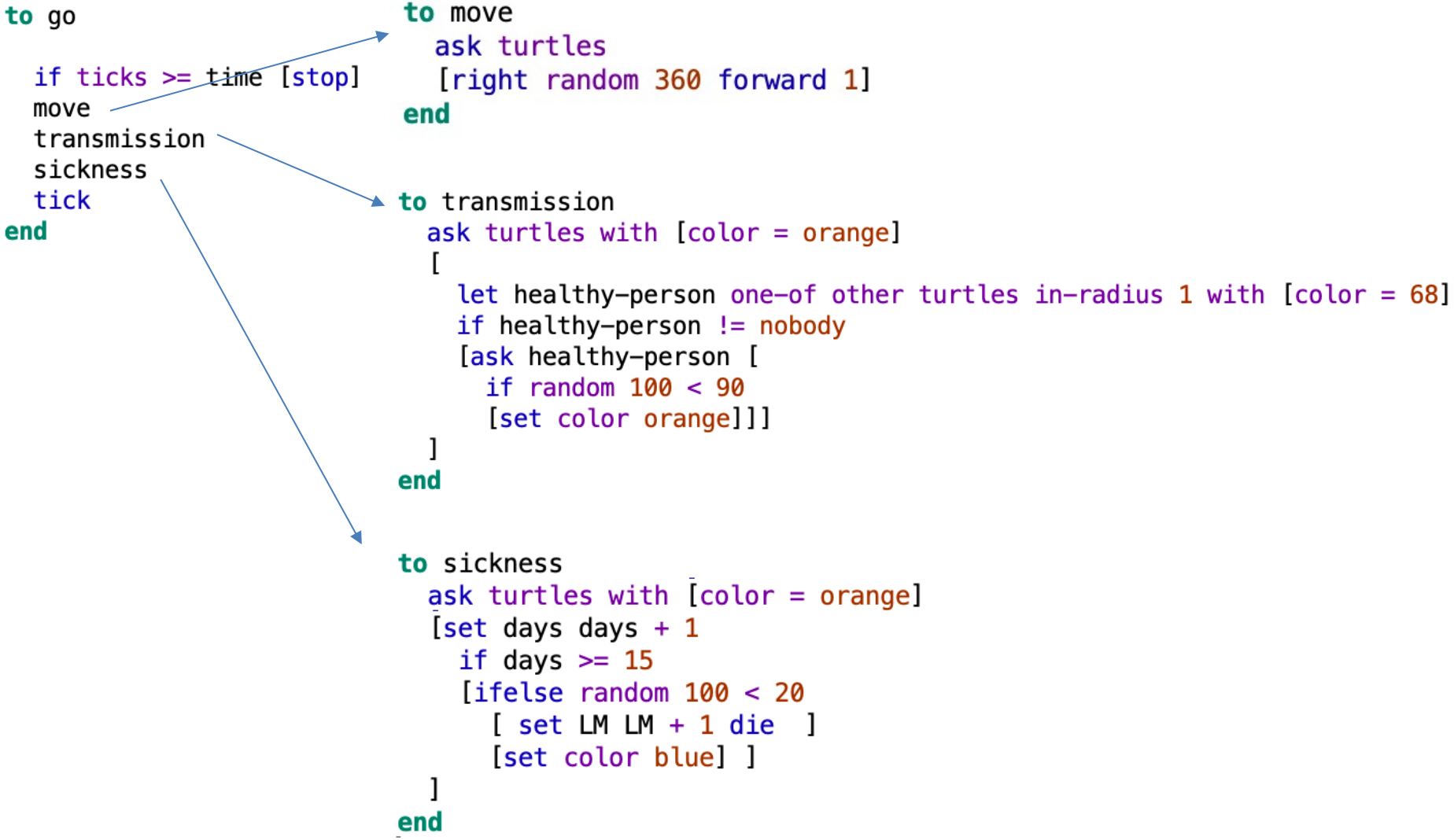





Population Changes over Time
1100
0
Persons
0
Time
184
Susceptible
Infected
Recovered

The rules of each agent according to the agent-based approach to the SIR model




to go
if ticks >= time [stop]
move
transmission
sickness
tick
end
to move
ask turtles
[right random 360 forward 1]
end
to transmission
ask turtles with [color = orange]
[
let healthy-person one-of other turtles in-radius 1 with [color = 68]
if healthy-person != nobody
[ask healthy-person [
if random 100 < 90
[set color orange]]]
]
end
to sickness
ask turtles with [color = orange]
[set days days + 1
if days >= 15
[ifelse random 100 < 20
[set LM LM + 1 die]
[set color blue]]
]
end

Comparison of approaches

- The results obtained from the different modelling approaches are **similar** but **not exactly equivalent**
- The agent model produces different results each time it is run
- This is not the case for equation approaches
- Equation approaches contain information about the probability but this **probability** appears in the form of frequencies that are parameterised in the model as constants

```
betac = 1
betai = 0.8
gamma = 0.3
N = 1000
```

Parameter for average contacts
Parameter for average probability of contagion
Parameter for average probability of recovery

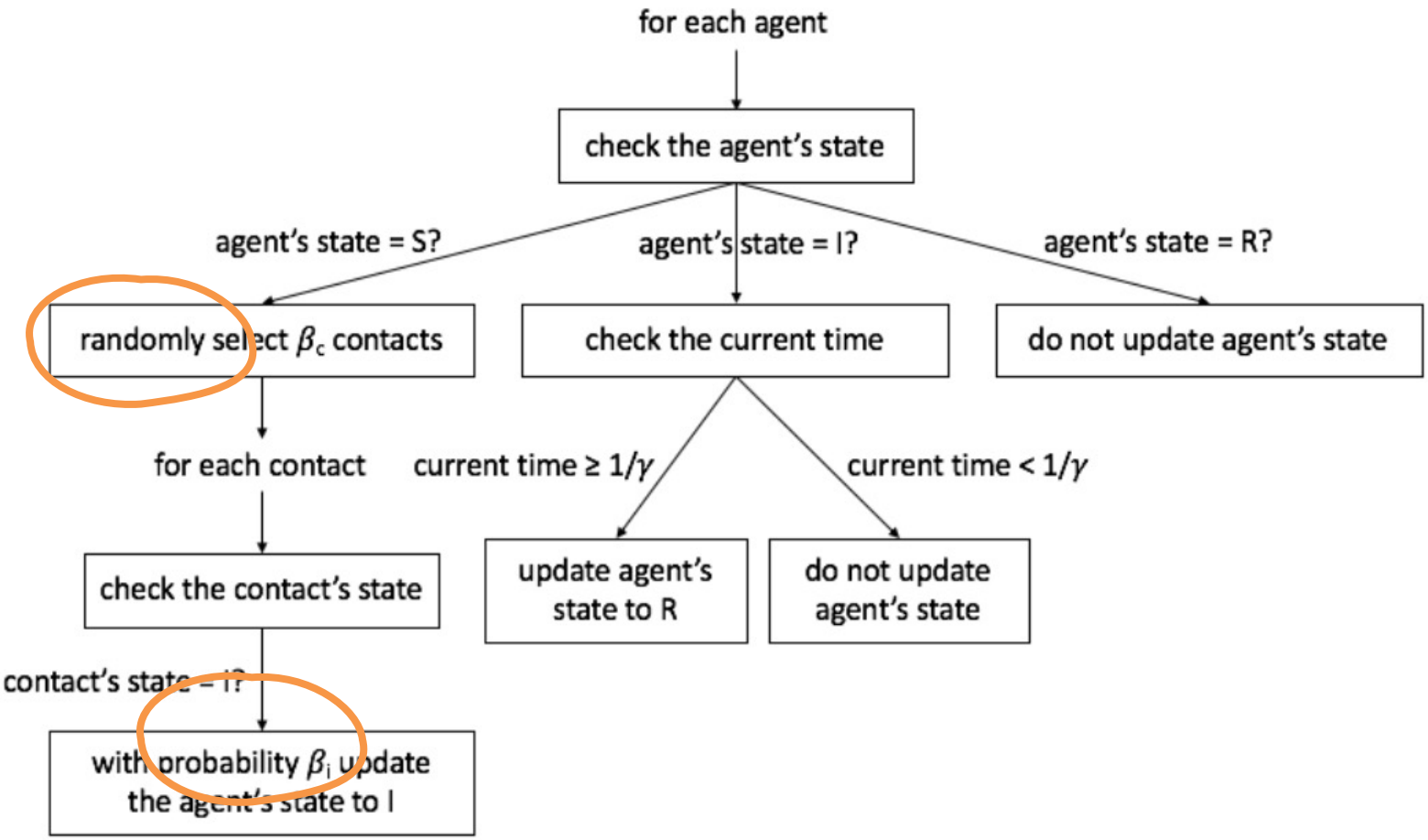
```
def dX_dt(X, t=0):
    return array([ - betac*betai*X[0]*X[1]/N ,
                   betac*betai*X[0]*X[1]/N - gamma*X[1] ,
                   gamma*X[1] ])

X = integrate.odeint(dX_dt, X0, t)
```

Comparison of approaches

- The results obtained from the different modelling approaches are **similar** but **not exactly equivalent**
- The agent model produces different results each time it is run
- This is not the case for equation approaches
- Equation approaches contain information about the probability but this **probability** appears in the form of frequencies that are parameterised in the model as constants
- The agent model is the only one that genuinely includes the **stochasticity of** the system in its formulation
- Going **beyond the determinism** of differential equations

The rules of each agent according to the agent-based approach to the SIR model



In summary

Comparison of approaches: top-down vs bottom-up

- The equation-based model is **top-down** because the equations already include expert knowledge of the problem.
- The agent model is bottom-up in that **the overall behaviour of the population emerges from measures of the dynamics of individual agents** and their probability of becoming infected or recovering.

- The equation-based model **examines the population as a whole**, at most compartmentalised
- The agent-based model **inspects each agent** and provides details on its status.

Comparison of approaches: determinism vs probability

- The agent-based model produces **different results each time** it is run
- The equation-based model **always** gives the same **result from** the same initial conditions.

- Equation-based approaches also include information about the disease that relates to probability. However, these probabilities only appear as parameterised **frequencies** in the model as **constants**. After defining them, they become constant values, numerical parameters and lose their original meaning as probabilistic measures.
- The agent approach genuinely includes the **stochasticity of** the system in its formulation, overcoming the determinism of differential equations. When agents, not aggregated populations, are considered, each individual has a probability that is specified by the given parameters.

Comparison of approaches: continuous vs. discrete

- The equation-based model is formulated with **continuous** differential **equations in both population and** time
- Temporal **discretisation is the** result of the various **integration** steps, which in turn are only due to our need to calculate and simulate

- The agent model is **inherently discrete in both population and** time
- The minimum components of the system are the agents, which are discrete by definition
- Time steps (ticks) are finite and discrete

Classic vs. complex... (i.e. where we started)

The 3 categories of comparison can describe not only the differences between **equation-based and agent-based simulations,** but also between **classical and complex systems**.

At the root of complexity is the presence of a **large but finite number of individual elements** in a system. These elements, called agents, **interact locally** according to **non-linear** (sometimes probabilistic) **relationships.** In turn, these rules, when applied to all agents, give the system **emergent global properties** that can only partially be explained from the behaviour of the individual components (see definition by Cilliers, 2002).

The **linearity** of interactions between components - which was a paradigm in Newtonian physics - ensures the **superposition** of effects on the system from the combination of local behaviour ("the whole is the sum of its parts"). Classical systems are also **deterministic** since, given differential equations describing their dynamics and initial conditions, their evolution in time is determined in any future state with arbitrary precision.

Describing models: equation or agent?

is in one sense more general. The problem may be summarised as follows: One (or more) infected person is introduced into a community of individuals, more or less susceptible to the disease in question. The disease spreads from the affected to the unaffected by contact infection. Each infected person runs through the course of his sickness, and finally is removed from the number of those who are sick, by recovery or by death. The chances of recovery or death vary from day to day during the course of his illness. The chances that the affected may convey infection to the unaffected are likewise dependent upon the stage of the sickness. As the epidemic spreads, the number of unaffected members of the community becomes reduced. Since the course of an epidemic is short compared with the life of an individual, the population may be considered as remaining constant, except in as far as it is modified by deaths due to the epidemic disease itself. In the course of time the epidemic may come to an end. One of the most important probems in epidemiology is to ascertain

Describing models: equation or agent?

is in one sense more general. The problem may be summarised as follows: One (or more) infected person is introduced into a community of individuals, more or less susceptible to the disease in question. The disease spreads from the affected to the unaffected by contact infection. Each infected person runs through the course of his sickness, and finally is removed from the number of those who are sick, by recovery or by death. The chances of recovery or death vary from day to day during the course of his illness. The chances that the affected may convey infection to the unaffected are likewise dependent upon the stage of the sickness. As the epidemic spreads, the number of unaffected members of the community becomes reduced. Since the course of an epidemic is short compared with the life of an individual, the population may be considered as remaining constant, except in as far as it is modified by deaths due to the epidemic disease itself. In the course of time the epidemic may come to an end. One of the most important probems in epidemiology is to ascertain

Describing models: equation or agent?

is in one sense more general. The problem may be summarised as follows: One (or more) infected person is introduced into a community of individuals, more or less susceptible to the disease in question. The disease spreads from the affected to the unaffected by contact infection. Each infected person runs through the course of his sickness, and finally is removed from the number of those who are sick, by recovery or by death. The chances of recovery or death vary from day to day during the course of his illness. The chances that the affected may convey infection to the unaffected are likewise dependent upon the stage of the sickness. As the epidemic spreads, the number of unaffected members of the community becomes reduced. Since the course of an epidemic is short compared with the life of an individual, the population may be considered as remaining constant, except in as far as it is modified by deaths due to the epidemic disease itself. In the course of time the epidemic may come to an end. One of the most important probems in epidemiology is to ascertain

Describing models: equation or agent?

- In the introduction to the original paper, the first approach is an **agent** approach
- In order to make the 'sense' of the model, the authors feel the need to give a description that follows individuals in their transition from susceptible to infected to cured
- Having clarified the *mechanism* on agents, we move on to the formulation in terms of population compartments and their variations
- From there, one is ready to move on to the **equation-based** formulation of the model

Research in science **education** has also focused heavily on learning through equation and agent models

- Equation models have a more familiar mathematical structure, especially for older students
- Agent models satisfy more the demand for a sense of mechanism, of understanding 'what is going on'.

Network models

Back to the assumptions of the SIR model

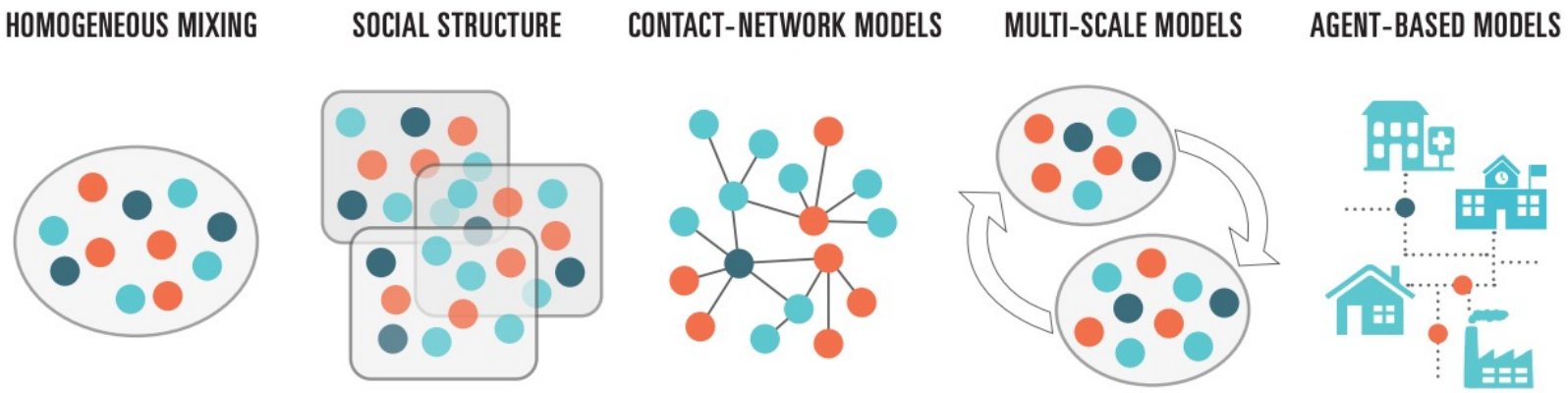
- The population **is isolated**
- **Birth or death dynamics** (from causes other than illness) are neglected: the sum of individuals in the compartments is constant
- The population is completely **mixed (***full mixing* hypothesis): individuals with whom a susceptible person has contact are chosen at random in the population
- All individuals have the same **number of** contacts with other individuals in the same time period
- The number of contacts between individuals does not vary according to the state of the disease: an infected individual has the same contacts as a susceptible or cured individual
- All infected individuals transmit the disease with **equal probability**
- All susceptible individuals are equally likely to contract the disease
- The disease has no **incubation period**
- Once cured, an individual is immediately **immune to the** disease

Problems with the homogeneity hypothesis

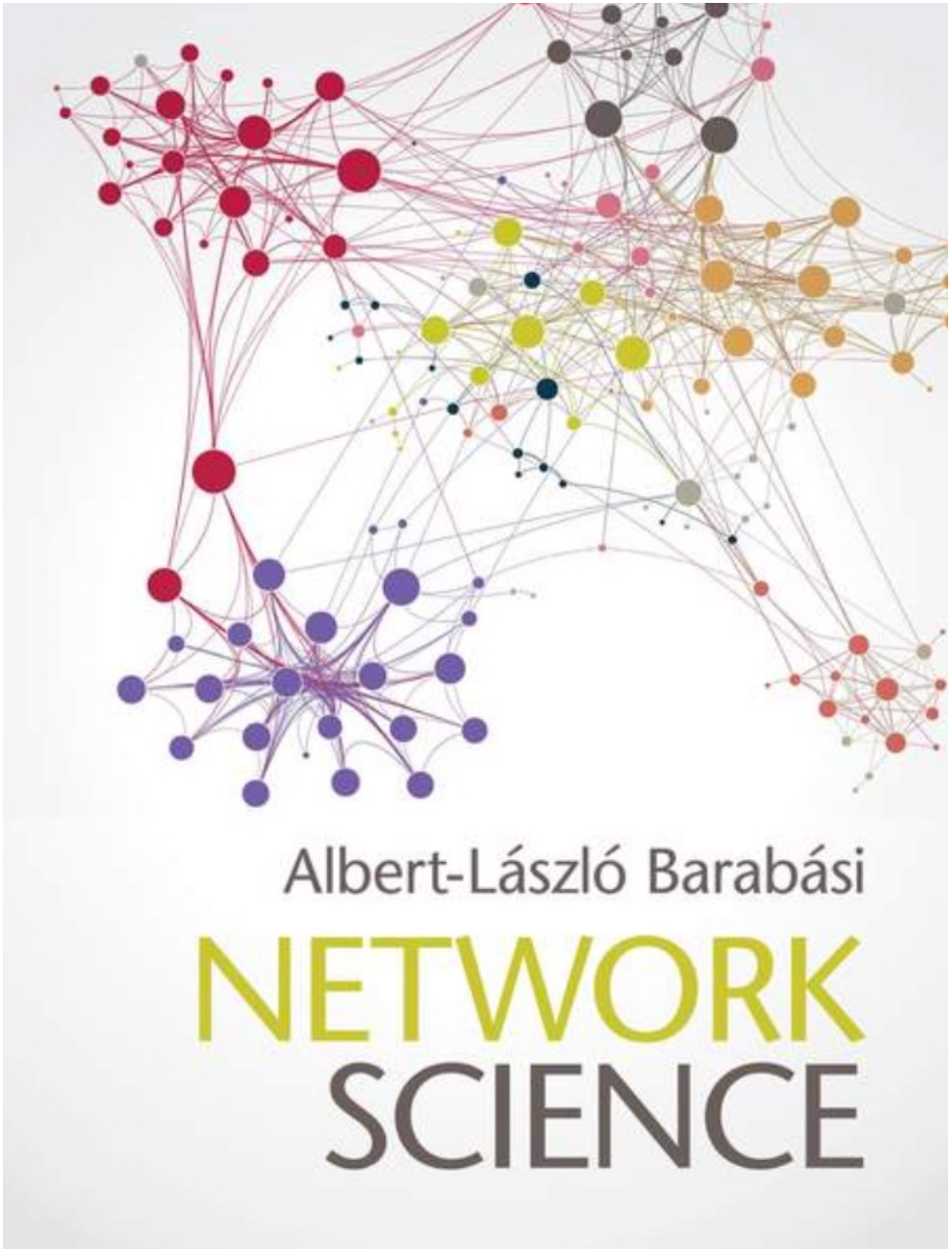
The homogeneity assumption used in the SIR model equations may be inadequate in various real-world situations in which individuals:

- have great heterogeneity in contact rates
- have specific interaction patterns
- are in contact with only a small part of the population.

A wide range of social and biological contagion processes requires capturing **the contact pattern structure** of individuals in modelling (Pastor-Satorras, Castellano, Van Mieghem & Vespignani, 2015).



Charting the Next Pandemic: Modeling Infectious Disease Spreading in the Data Science Age
https://link.springer.com/book/10.1007/978-3-319-93290-3



http://networksciencebook.com/

Networks: the heart of complex systems

Behind every complex system is an intricate network that encodes the interactions between the system's components:

- The network that encodes the interactions between genes, proteins and metabolites integrates these components into the cells that thus become alive. The very existence of this **cellular network** is a prerequisite for life.
- The electrical pattern that captures the connections between neurons, called the **neural network**, is the key to our understanding of how the brain works and our consciousness.
- The sum of all professional, friendship and family ties, often called the **social network**, is the fabric of society and determines the dissemination of knowledge, behaviour and resources.
- **Communication Networks**
- **Electrical Networks**
- **Commercial Networks**
- **Internet networks**

Network: the heart of complex systems

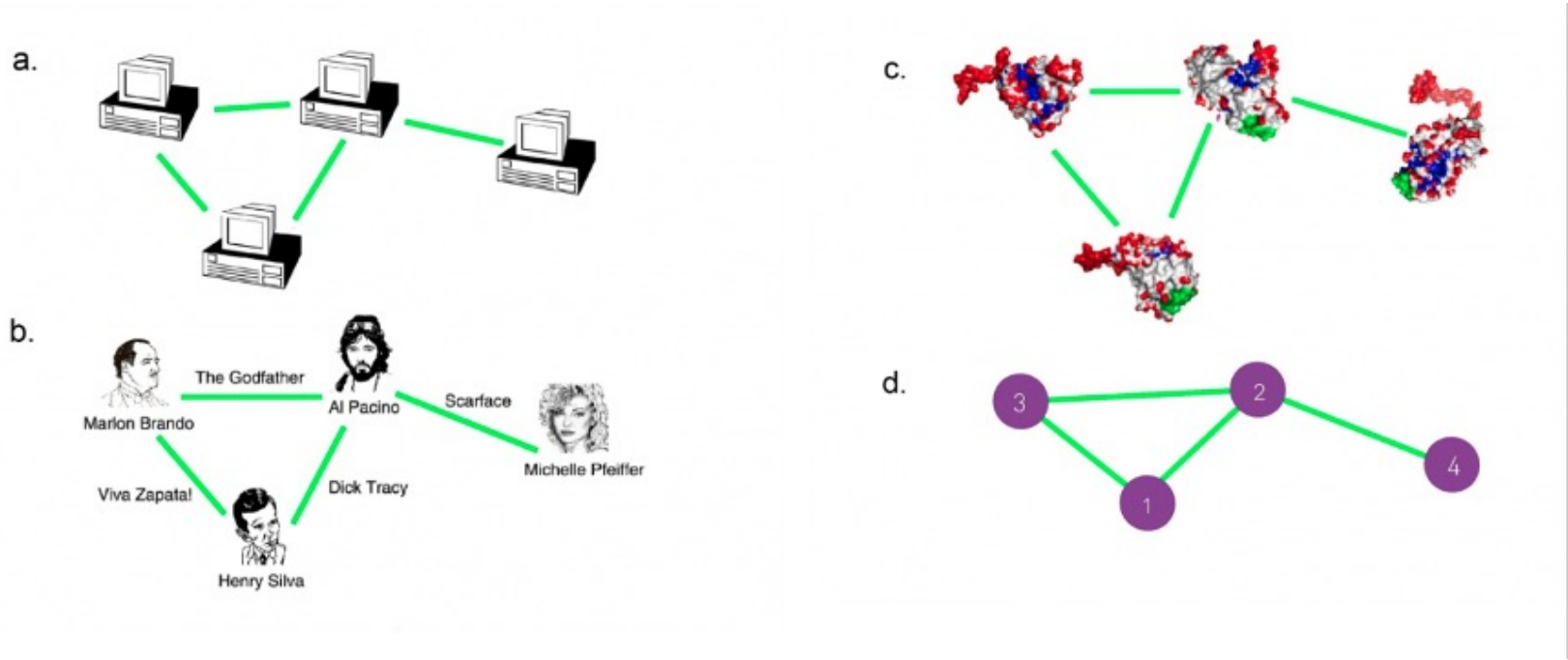
- The interest in network science that exploded during the first decade of the 21st century is rooted in the discovery that, despite **the obvious diversity of complex systems**, the structure and evolution of the networks behind each system is guided by a **common set of fundamental laws and principles**.
- Thus, despite the incredible differences in the shape, size, nature, age and scope of real networks, most networks are guided by **common organisational principles**.
- Once the nature of the components and the precise nature of the interactions between them are disregarded, the resulting networks are more similar than different.

Network science 101

- Networks are described mathematically as graphs
- A graph is a collection of points, called vertices (nodes in physics, actors in the social sciences).
- These points are joined by a set of connections, called edges, links and ties, in mathematics, physics and social sciences respectively.

- Each edge denotes the presence of a *relationship* or *interaction* between vertices that unites
- Bidirectional interaction (undirected networks) or unidirectional interaction (directed networks)

Different networks, same graph



The figure shows a small subset of (a) the Internet, where routers are connected to each other; (b) the Hollywood actors' network, where two actors are connected if they starred in the same film; (c) a protein-protein interaction network, where two proteins are connected if there is experimental evidence that they can bind to each other in the cell. Although the nature of the nodes and links differs, these networks have the same graphical representation, consisting of N = 4 nodes and L = 4 links, shown in (d).

Different networks, same properties

- **scale-freeness**, a criterion Barabási and Albert use to characterise complex networks: as a network grows, new nodes connect to the most connected nodes, a phenomenon known as **preferential attachment**. The resulting network is called scale-free because it exhibits the same properties on multiple scales.
- This phenomenon is **pervasive**: it is empirically observed in many different situations

- *Why do we observe the same thing in contexts that a priori have nothing in common?*
- *Is scale-freeness a sign of a universal law (nomothetic approach) or is it only an empirical characterisation (ideographic approach?*

» ***'A clash of two cultures'***

(Jacomy, 2020)
https://journals.sagepub.com/doi/full/10.1177/2053951720949577

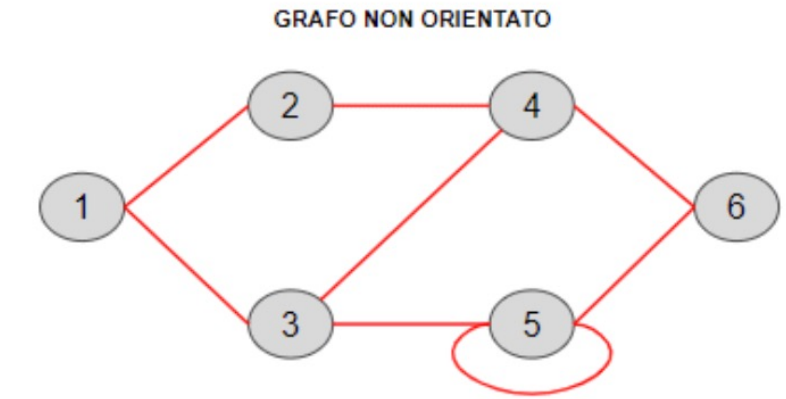
From graph to adjacency matrix

A compact way to specify all connections in a graph of dimension N (i.e. with N vertices) is the adjacency matrix A of dimension N × N

a_{ij} = 1 if there is an edge connecting nodes i and j

a_{ij} = 0 otherwise

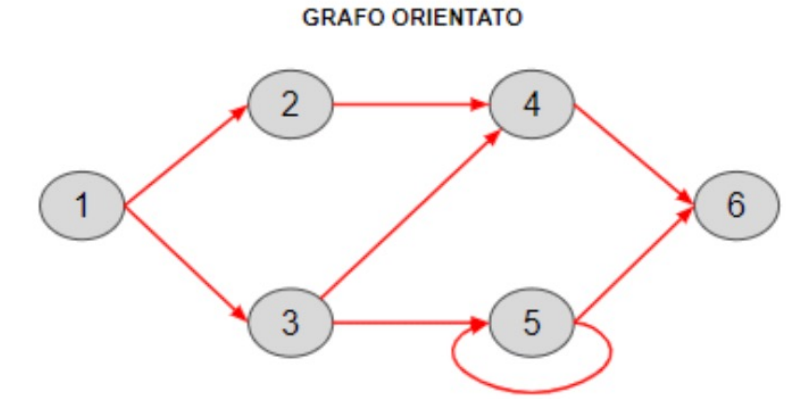
A is symmetrical in undirected graphs and asymmetrical in oriented graphs



nodi destinazione

nodi di partenza	1	2	3	4	5	6
1	0	1	1	0	0	0
2	1	0	0	1	0	0
3	1	0	0	1	1	0
4	0	1	1	0	0	1
5	0	0	1	0	1	1
6	0	0	0	1	1	0

E' una matrice trasposta

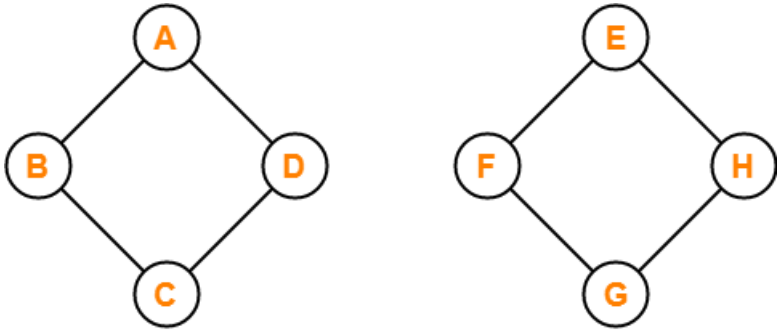


nodi destinazione

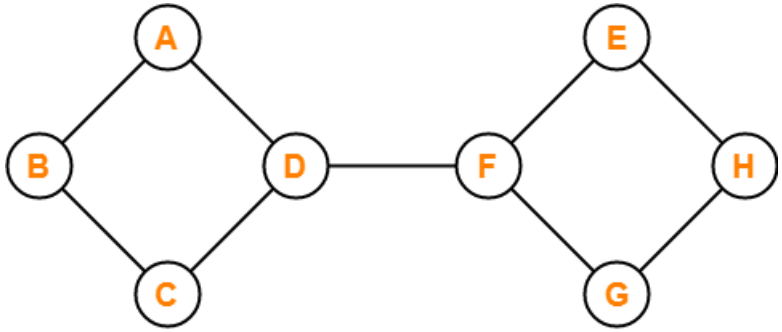
nodi di partenza	1	2	3	4	5	6
1	0	1	1	0	0	0
2	0	0	0	1	0	0
3	0	0	0	1	1	0
4	0	0	0	0	0	1
5	0	0	0	0	1	1
6	0	0	0	0	0	0

Path and connection

- A path $P_{i0,in}$ connecting vertices i_0 and i_n is a succession of connected edges {(ij, ij+1)} with j = 0, ..., n-1.
- n is the number of edges crossed, also called the path length
- A graph is **connected** if there is a path connecting any two vertices in the graph
- A C-component of a graph is defined as a connected subgraph



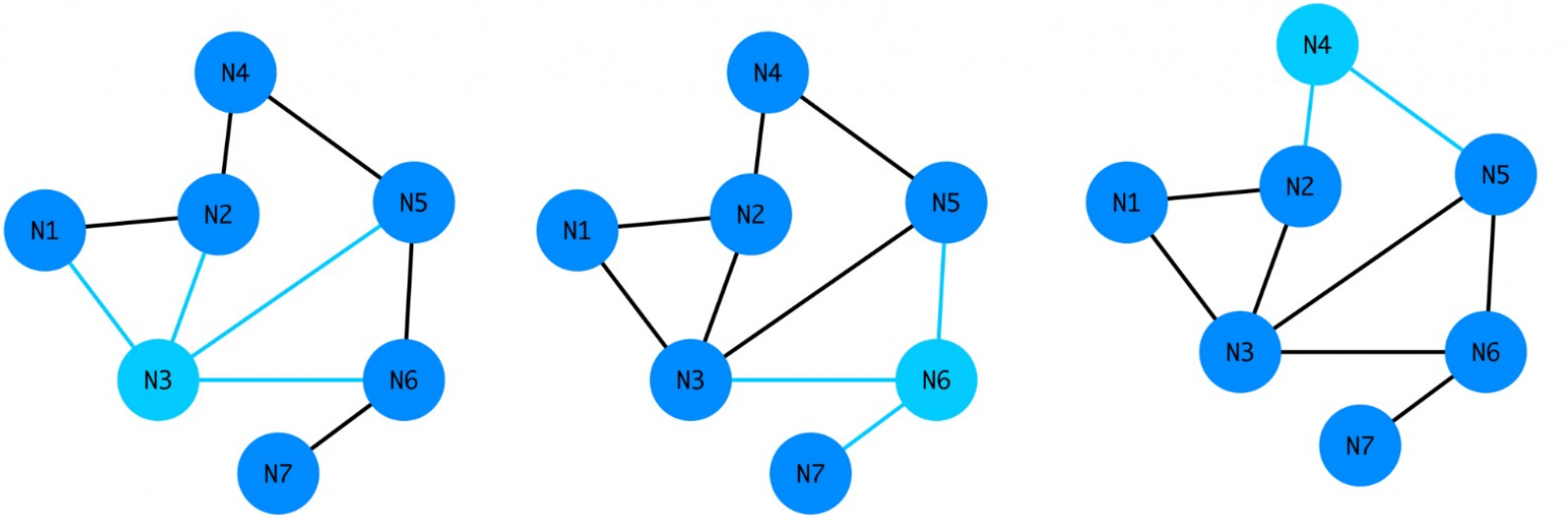
Example of Disconnected Graph



Example of Connected Graph

Degree of a node

The degree of a node is the number of edges (links) the node possesses

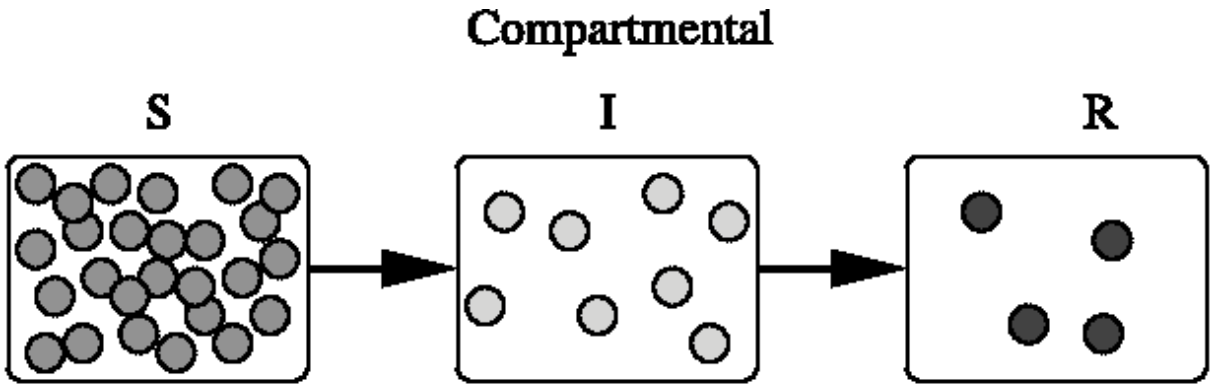


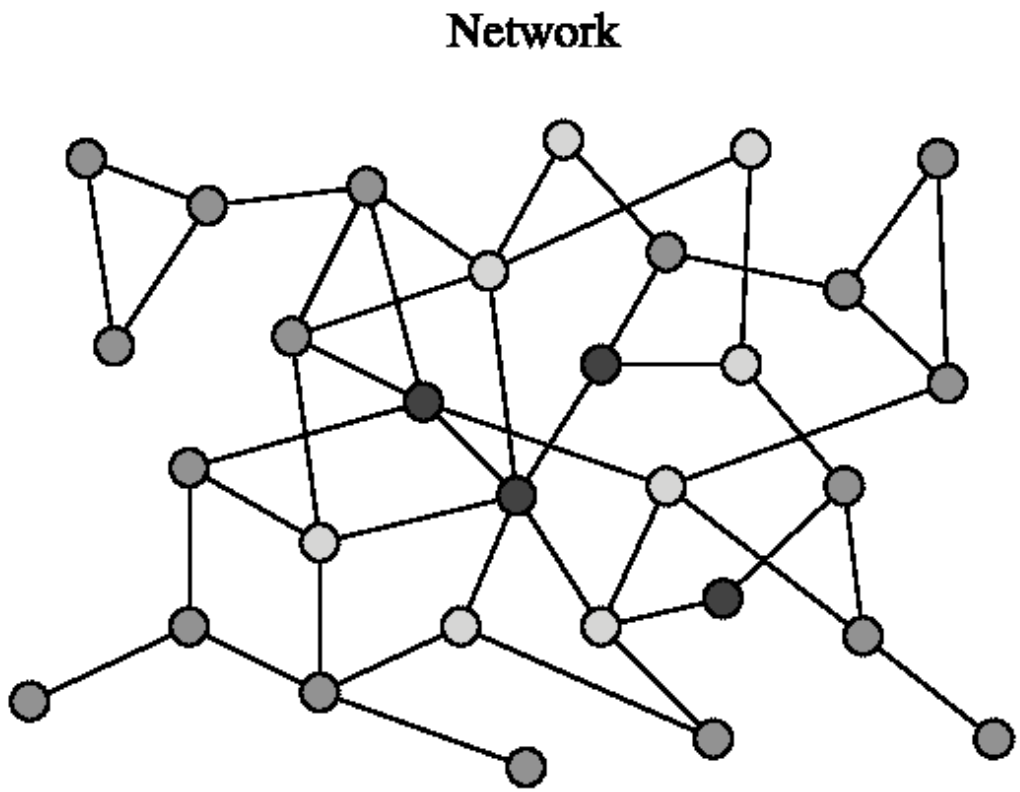
$k_3 = 4$ $k_6 = 3$ $k_4 = 2$

With these notions, how would you construct a network to model an epidemiological phenomenon?

Vertices, links, directed/undirected, components...

Help yourself graphically





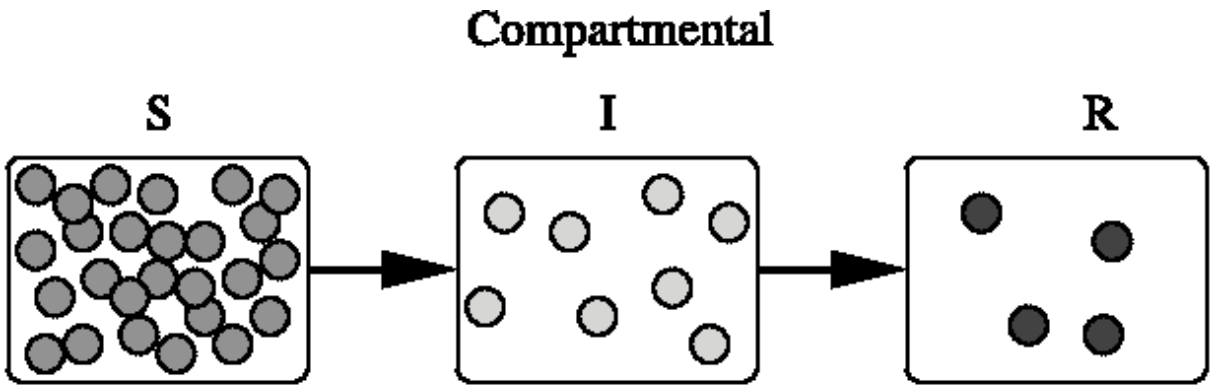
In short, the **nodes** are the **people**, while the **links** are the **potential transmission paths of** a virus

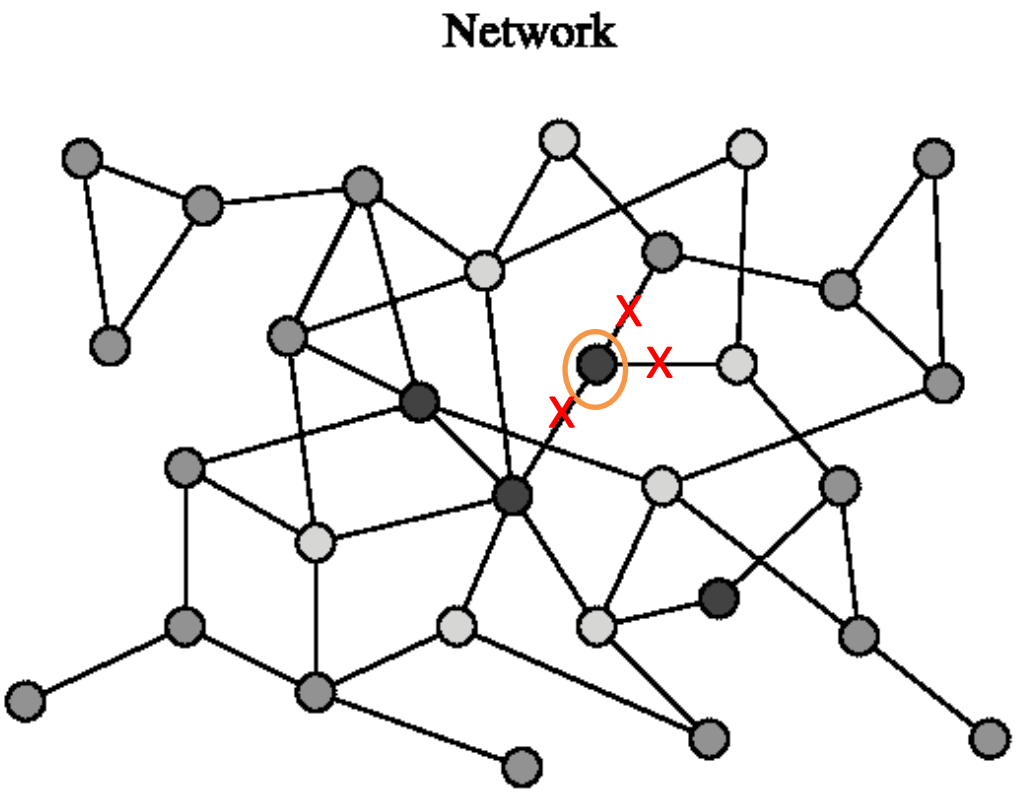
Undirected graph

What happens when you vaccinate?

Reason both with respect to the model you previously formulated in the group and to the one now proposed

Reason both locally and globally





When an individual gets **vaccinated**, his or her node is **disconnected** from the network (because it can no longer acquire or transmit a disease).

Vaccination then fades into the net.

Three strategies for vaccinating a population

Strategy A

- a certain number of nodes are chosen at random and vaccinated

Strategy B

- a number of the most connected nodes are identified and vaccinated

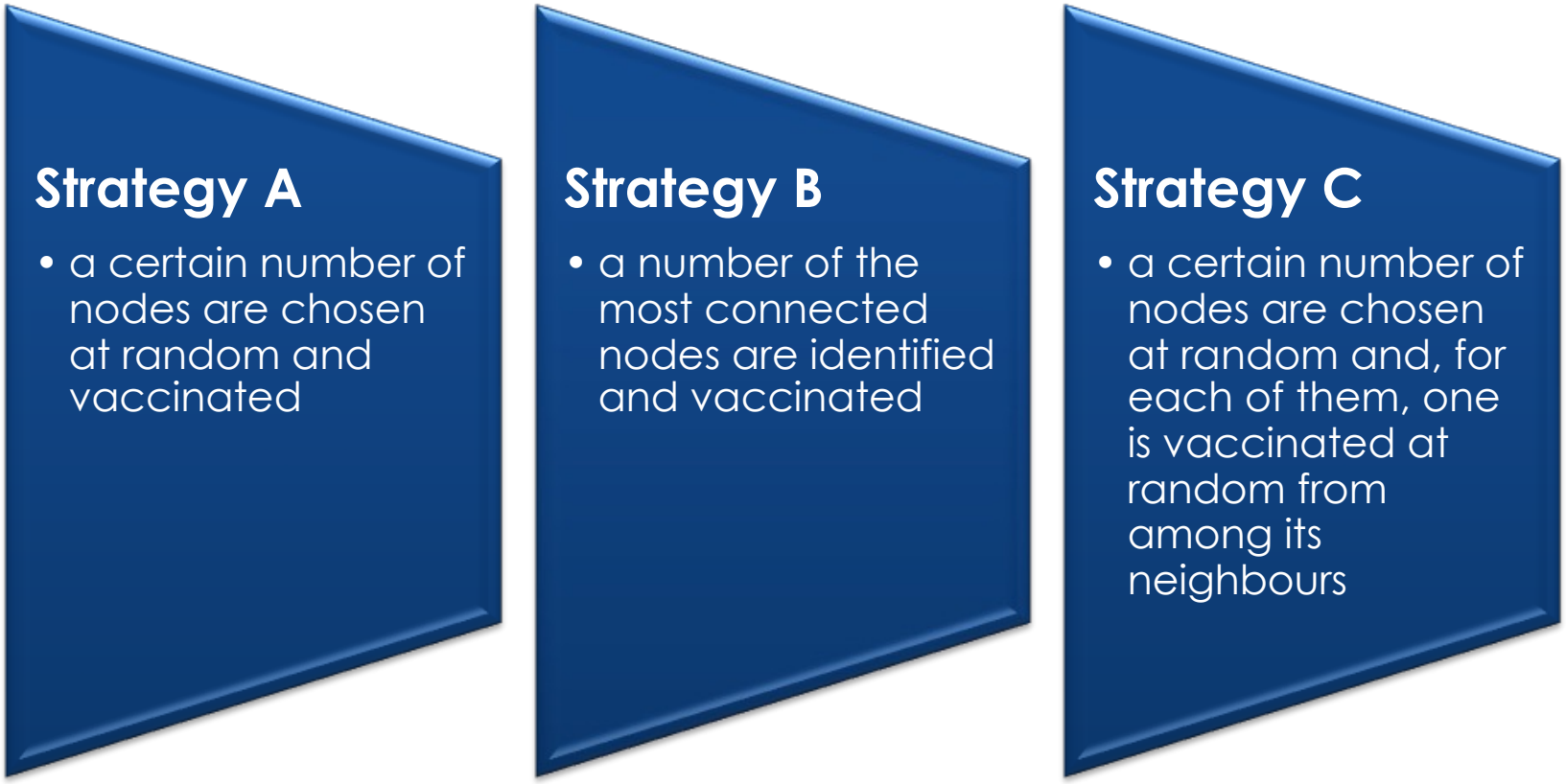
Strategy C

- a certain number of nodes are chosen at random and, for each of these, one is vaccinated at random from among its neighbours

Which strategy do you consider most effective? According to which 'parameters'? Why?

If you had to select a second choice, what would it be? Why?

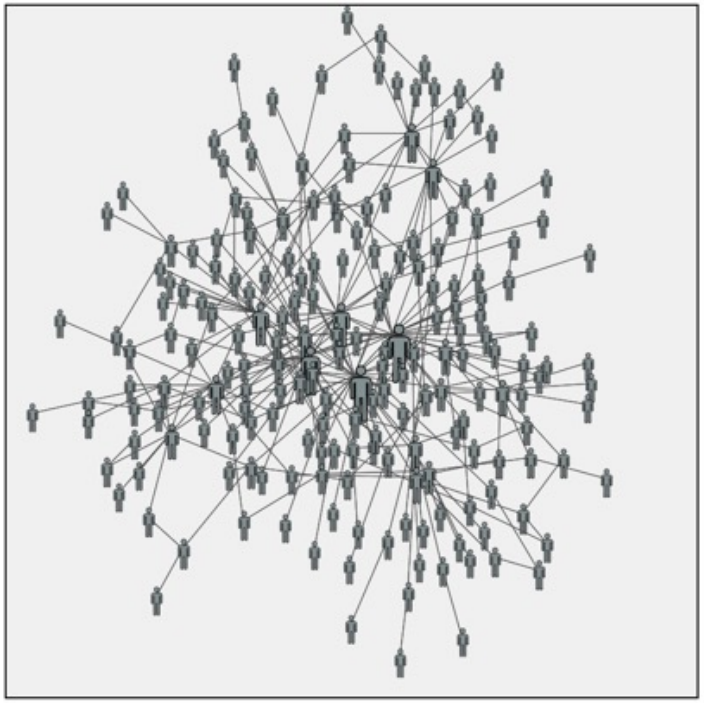
An unexpected ending

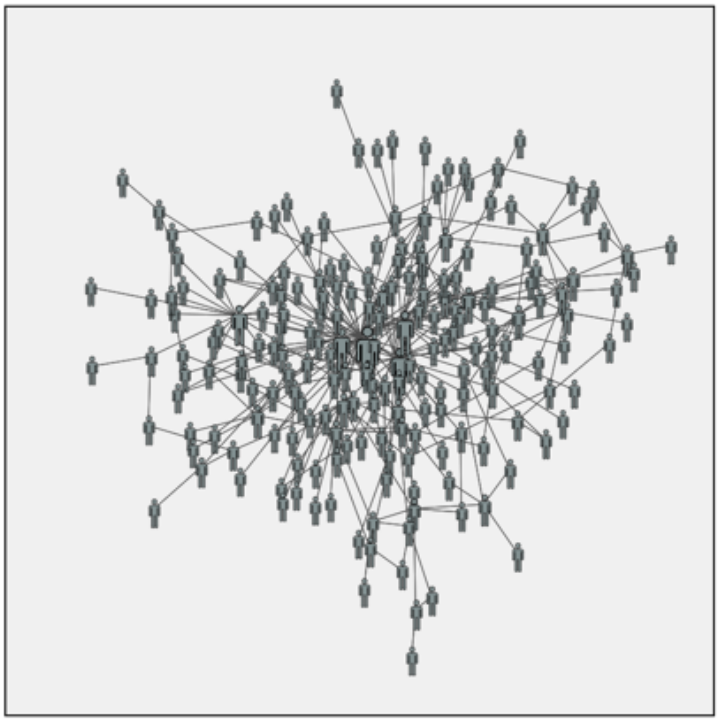


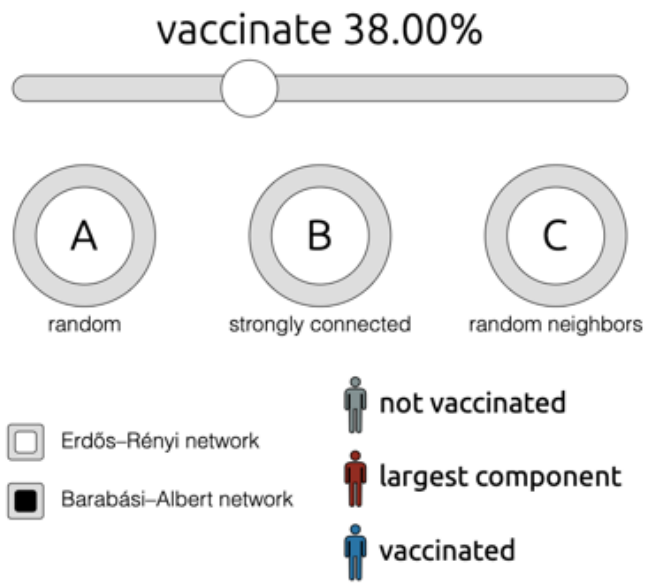
- Intuitively, we expect **strategy B** to be **more efficient** than A and C, and indeed it is.
- Strategies **A and C would appear to be equivalent** BUT they are not
- Strategy **C is more efficient than strategy A**

Let us explore the three strategies with a simulation

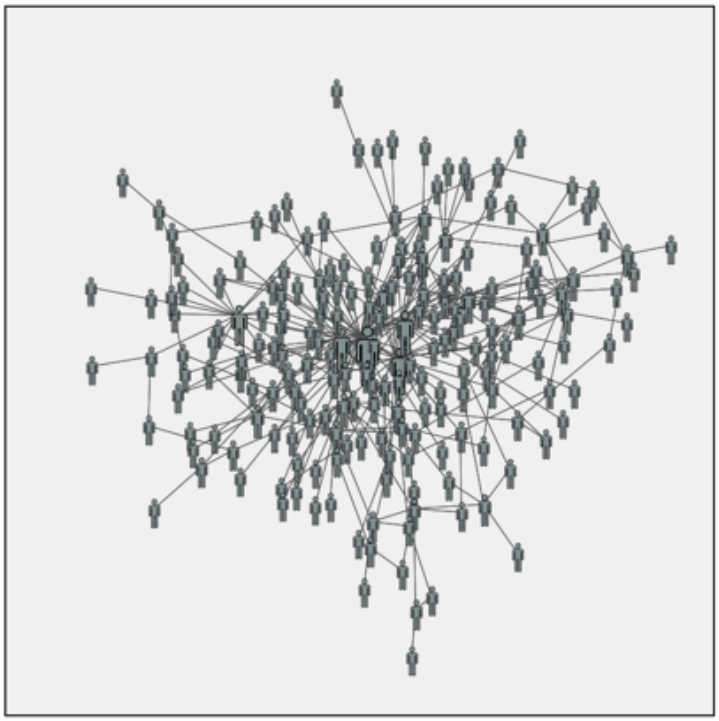
https://www.complexity-explorables.org/slides/facebooked-flu-shots/

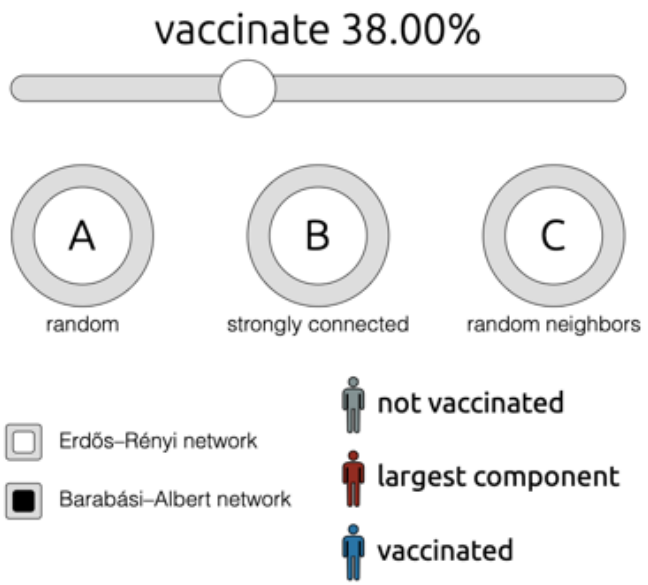






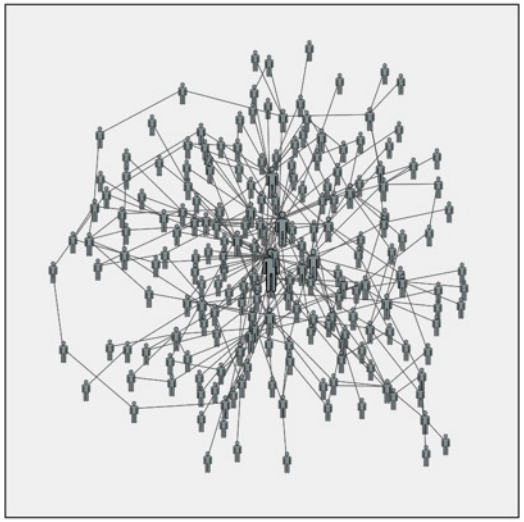
The network, consisting of 200 people (nodes), has only one component, so if a contagion process were to propagate through this network, eventually every node could be affected. Highly connected individuals are displayed somewhat larger than those with a small number of connections. The **connectivity of** a node is measured by its **degree**, i.e. the number of connections (neighbours) it has.



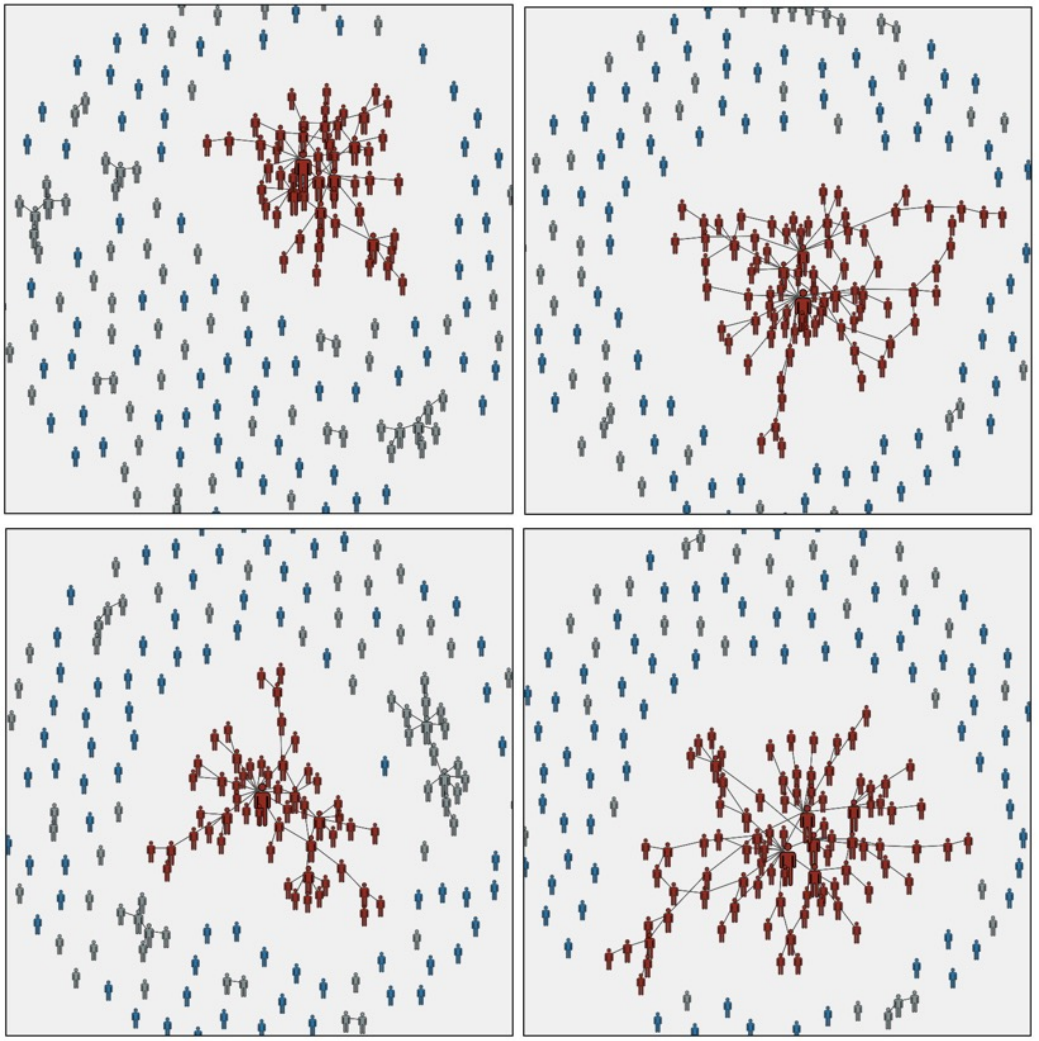


When vaccinating a certain fraction of the population, a certain number of nodes will be disconnected from the network according to strategies A, B or C.

Strategy A

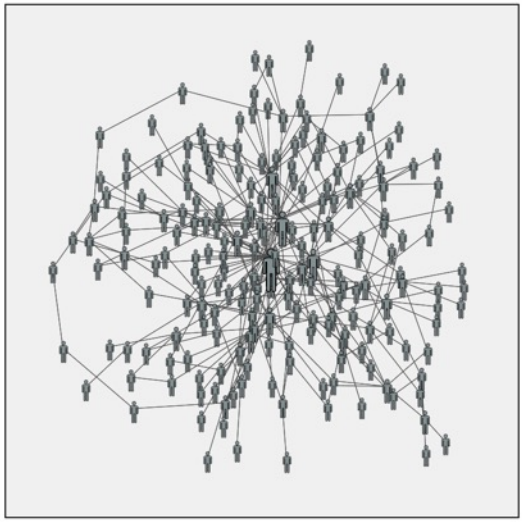


All vaccinated individuals are now isolated and moving to the periphery. However, **a considerable part of the network is still in a large component**. The network has not really disconnected, because 38% is too low for this strategy.



4 different simulations
Same percentage of vaccinated (38%)
Same strategy
Same type of network

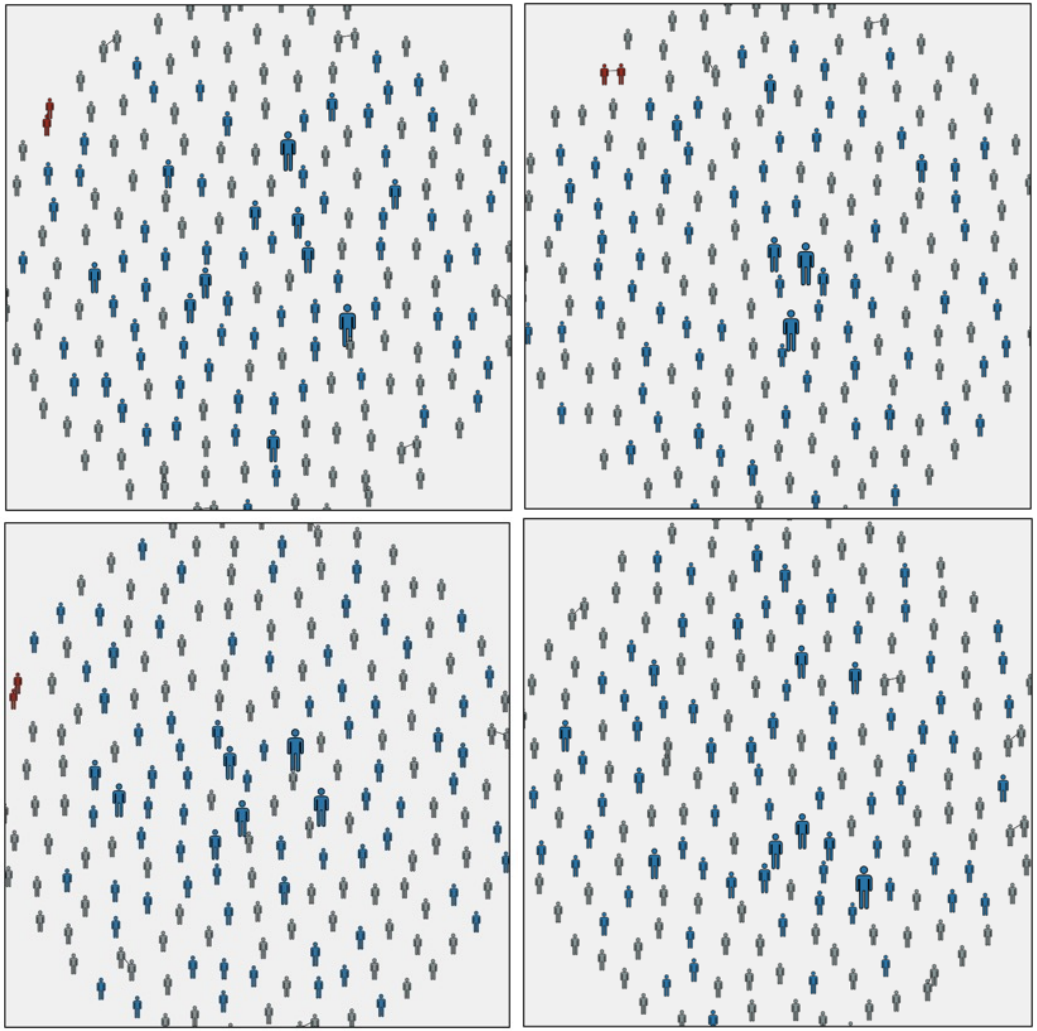
Strategy B



The network becomes very sparse and the largest component is also very small.

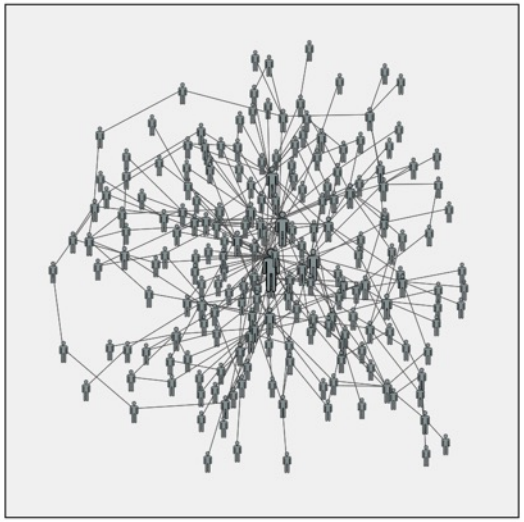
By removing high degree nodes, we effectively remove many more links. The network falls apart.

Among the isolated nodes, many are not vaccinated (herd immunity)

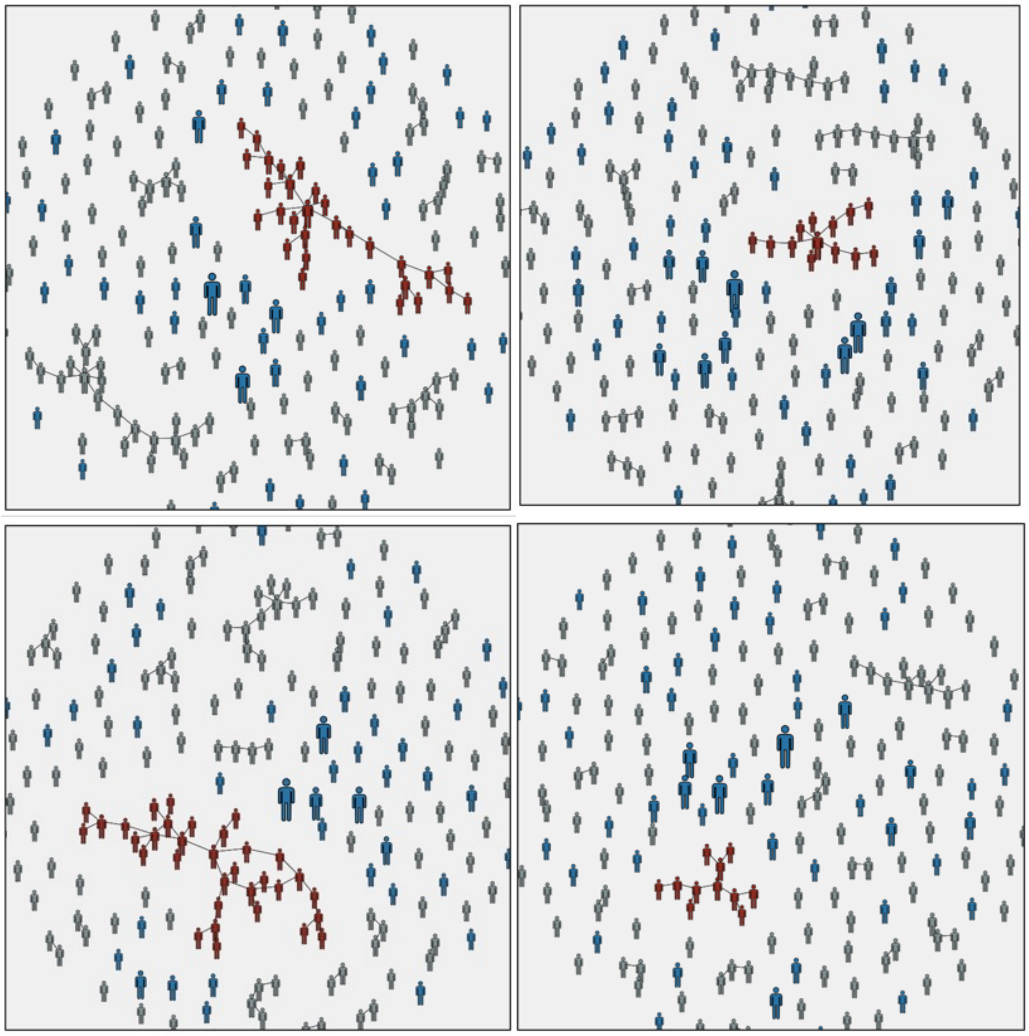


4 different simulations
Same percentage of vaccinated (38%)
Same strategy
Same type of network

Strategy C



However, comparing the size of the largest component in strategy A and C, we see that typically this **giant component is significantly smaller for C than for A**. Therefore, strategy C is more effective!



4 different simulations
Same percentage of vaccinated (38%)
Same strategy
Same type of network

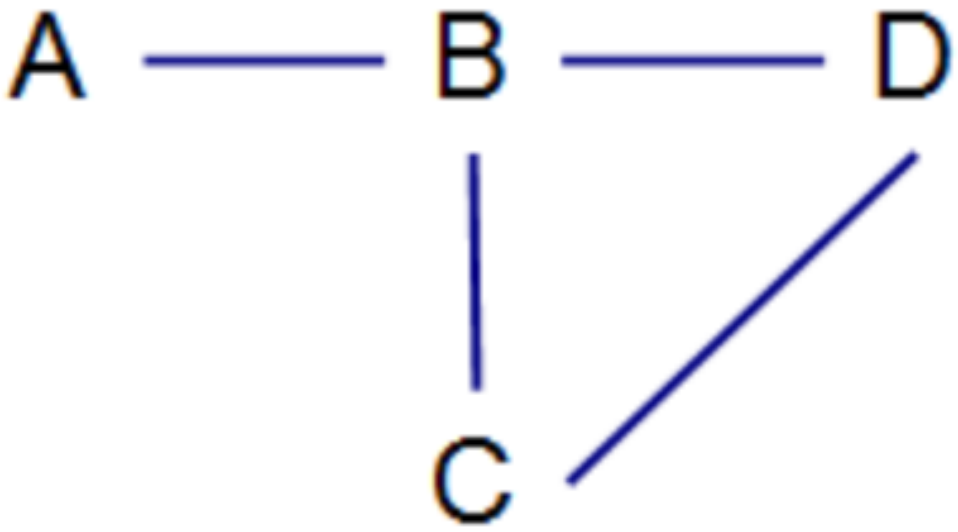
Why the difference between A and C?
The paradox of friendship

- A peculiar property of complex networks, especially those with heterogeneous connectivity at nodes, is that on ***average* the degree of a node's neighbours is greater than the degree of the node itself**

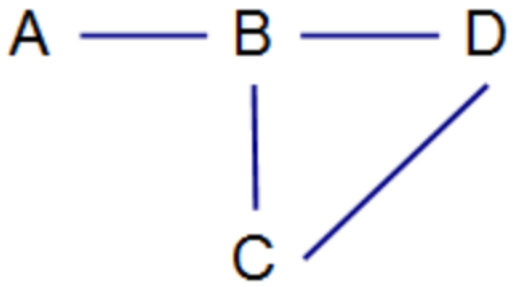
- This is known as the **friendship paradox**

- Why should my 'friend' show different properties to mine?
- After all, am I not also a friend of my friend?

The paradox of friendship



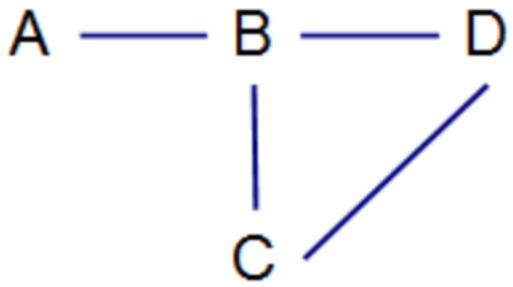
The paradox of friendship



Person	Number of friends	Total number of friends of friends	Mean number of friends of friends
A			
B			
C			
D			

Total:
Mean:

The paradox of friendship



Person	Number of friends	Total number of friends of friends	Mean number of friends of friends
A	1	3	3
B	3	5	1.67
C	2	5	2.5
D	2	5	2.5

Total: 8 18 9.67

Mean: **2 2.25 2.42**

out of 4 persons

out of 8 connections/friendships

out of 4 persons

The paradox of friendship

"the mean number of friends of friends is always greater than the mean number of friends of individuals".

- The secret is hidden in the term 'on average' and that we are comparing different averages
- In one case we are averaging **over nodes**, in the other we are **averaging over** links

- When we choose a random set of nodes (strategy A), the probability of choosing node n among N nodes is the same for all nodes (1/N)
- When we choose a random neighbour of a random node, the probability of choosing a node is proportional to the degree q of the target node: we are no longer choosing between nodes uniformly
- We are more likely to choose a node with a higher degree.

The paradox of friendship

- **It can be shown** that on average a neighbour's degree is:

$$q_0 = (1 + \sigma^2)k_0$$

average degree of a node

variance in the degree of nodes

- On average, the degree of the neighbouring node is always greater than the average degree of the node
- The effect is strongest for networks that have high variance node distributions (e.g. in Barabasi-Albert networks)

Demonstration: https://mindyourdecisions.com/blog/2012/09/04/why-your-friends-have-more-friends-than-you-the-friendship-paradox

Why Your Friends Have More Friends than You Do[1]

Scott L. Feld
State University of New York at Stony Brook

(Feld, 1991)

https://www.jstor.org/stable/2781907

It is reasonable to suppose that individuals use the number of friends that their friends have as one basis for determining whether they, themselves, have an adequate number of friends. This article shows that, if individuals compare themselves with their friends, it is likely that most of them will feel relatively inadequate. Data on friendship drawn from James Coleman's (1961) classic study *The Adolescent Society* are used to illustrate the phenomenon that most people have fewer friends than their friends have. The logic underlying the phenomenon is mathematically explored, showing that the mean number of friends of friends is always greater than the mean number of friends of individuals. Further analysis shows that the proportion of individuals who have fewer friends than the mean number of friends their own friends have is affected by the exact arrangement of friendships in a social network. This disproportionate experiencing of friends with many friends is related to a set of abstractly similar "class size paradoxes" that includes such diverse phenomena as the tendencies for college students to experience the mean class size as larger than it actually is and for people to experience beaches and parks as more crowded than they usually are.