Diffusion et propagation de l’information

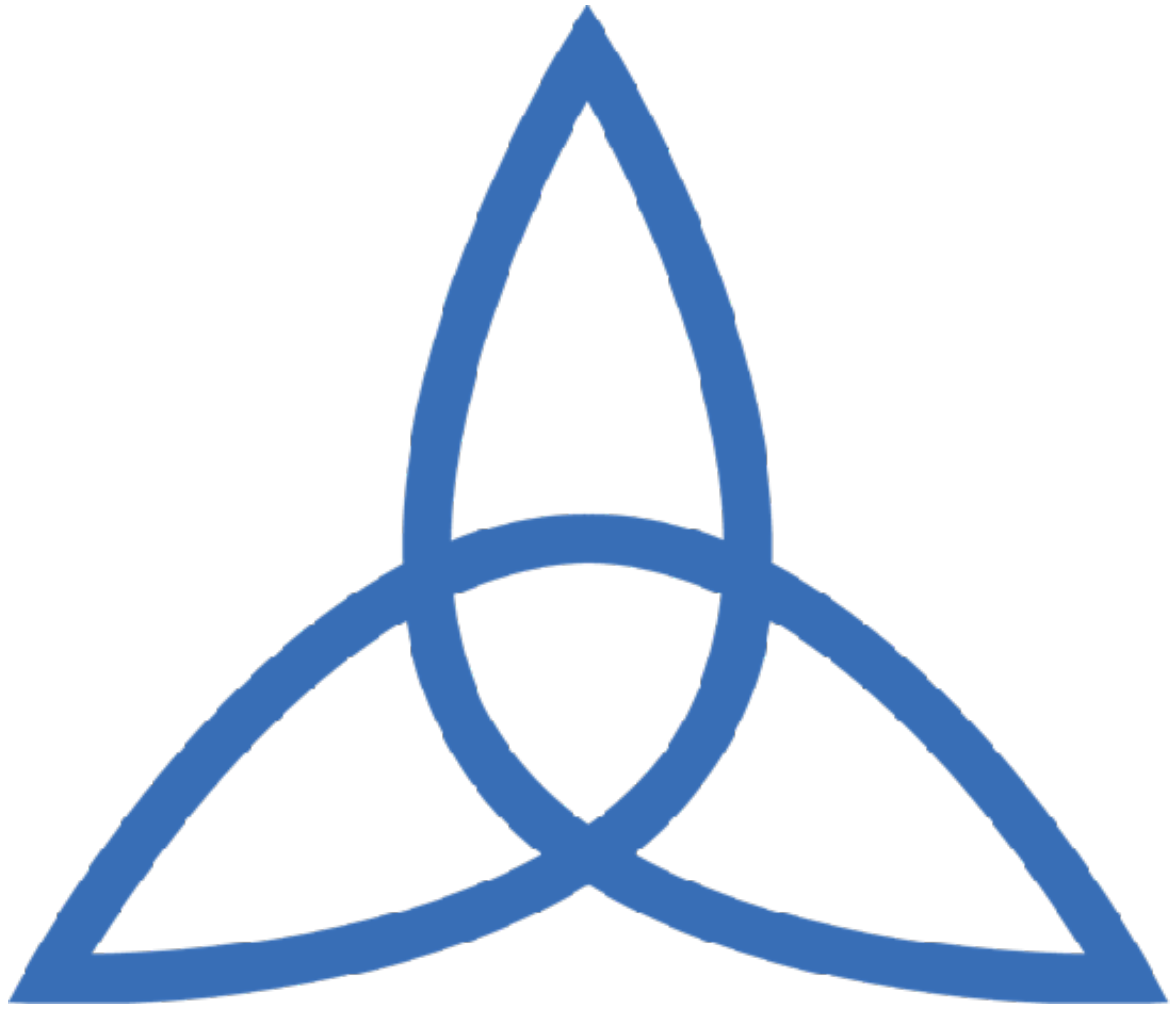
17ème école d’été de Peyresq
en traitement du signal et des images

Francesco Pinotti

francesco-pinotti.netlify.app

francesco.pinotti@biology.ox.ac.uk

UNIVERSITY OF
OXFORD

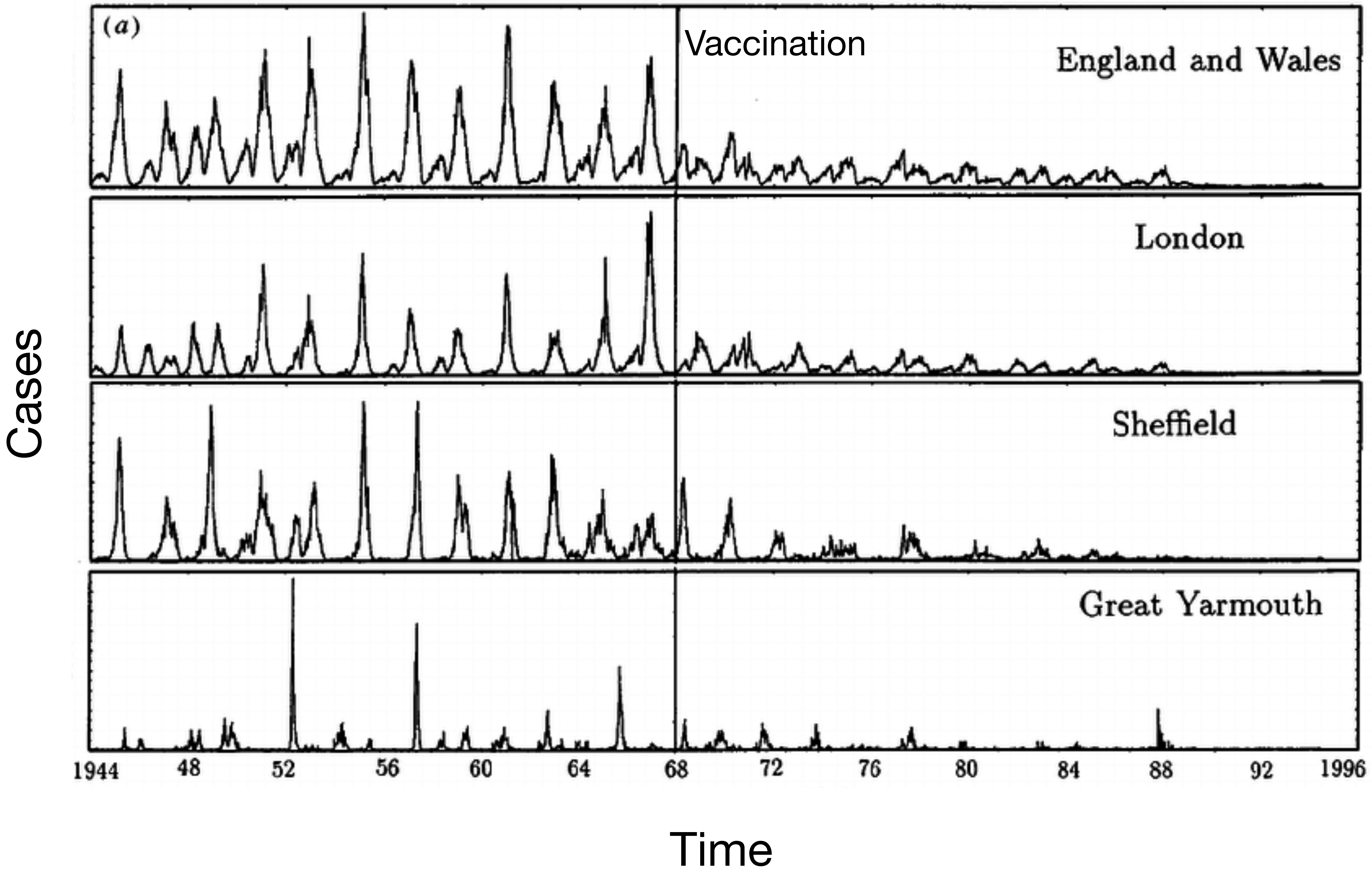


Outline

- Modelling epidemics
- Epidemics on networks
- Spatial and metapopulation models
- Agent-based models

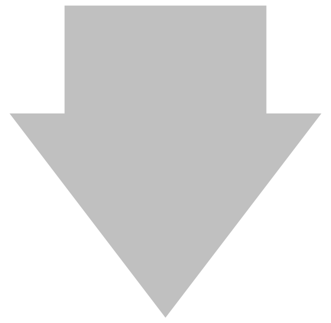
Modelling epidemics

Measles (1944-1996)



Potential questions

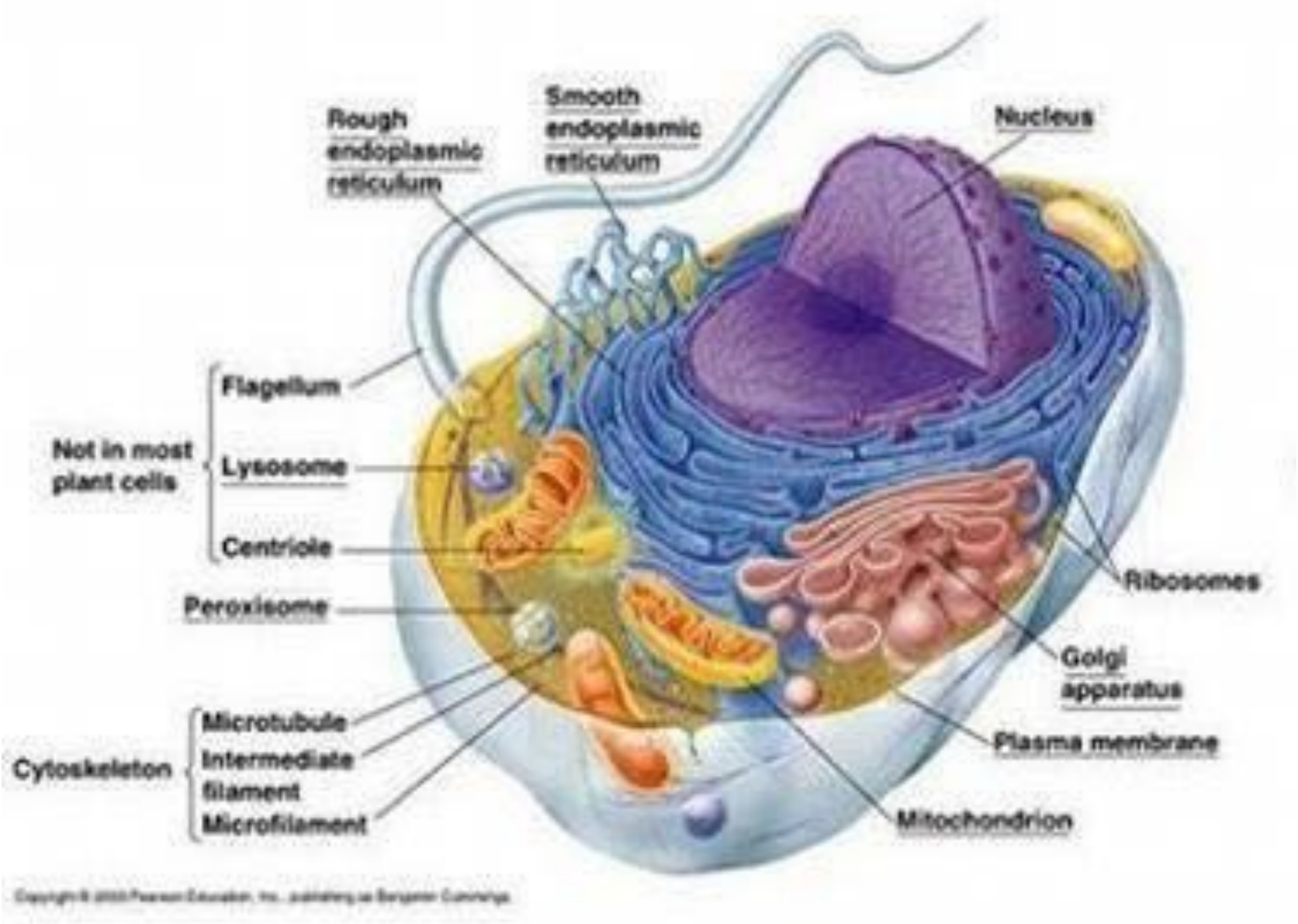
- Why 2 year gaps before vaccination?
- National vs subnational epidemics?
- Impact of vaccination?



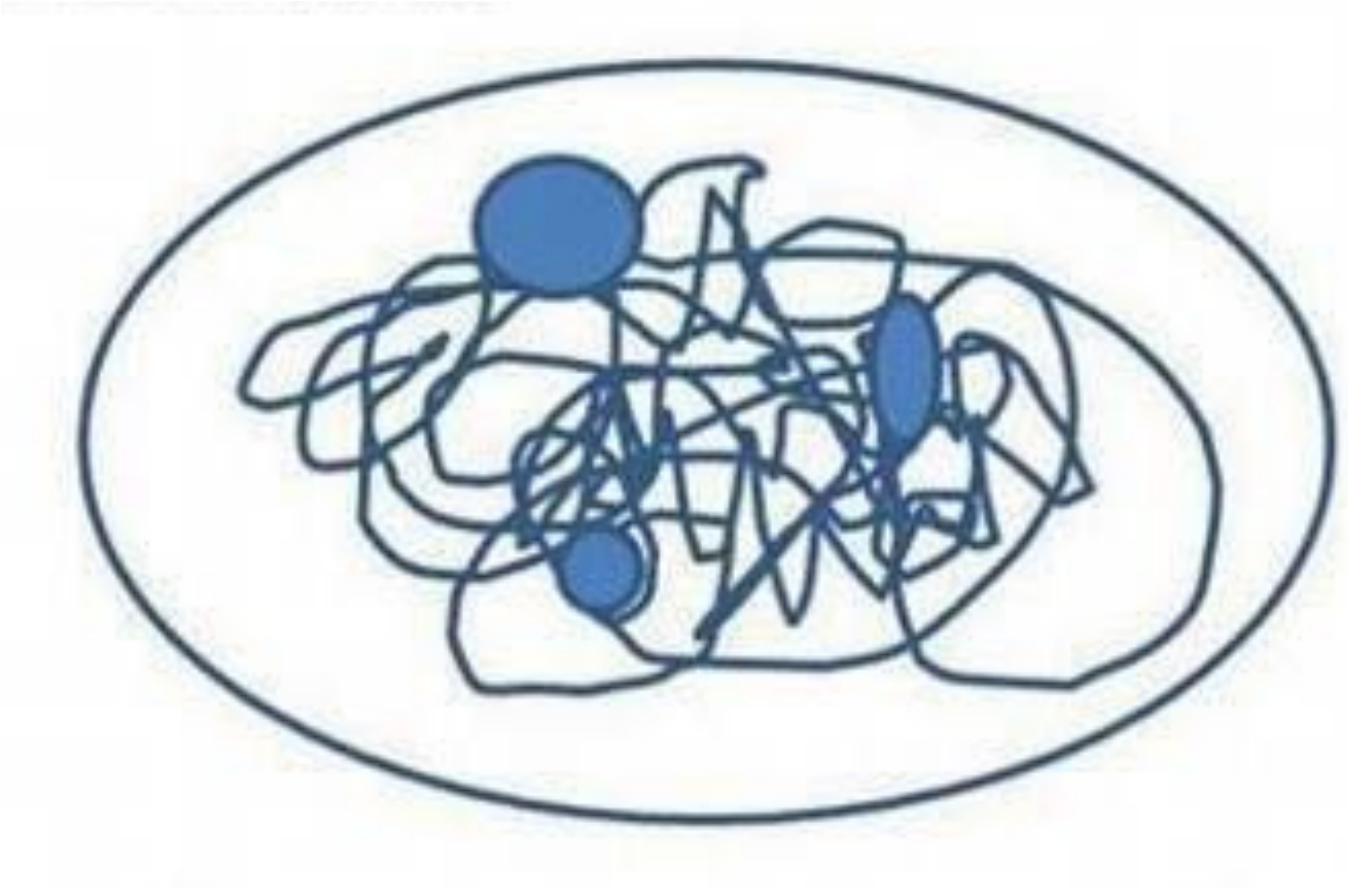
Mathematical modelling

Finkenstädt et al, PRSB, 1998

The view of a cell by:



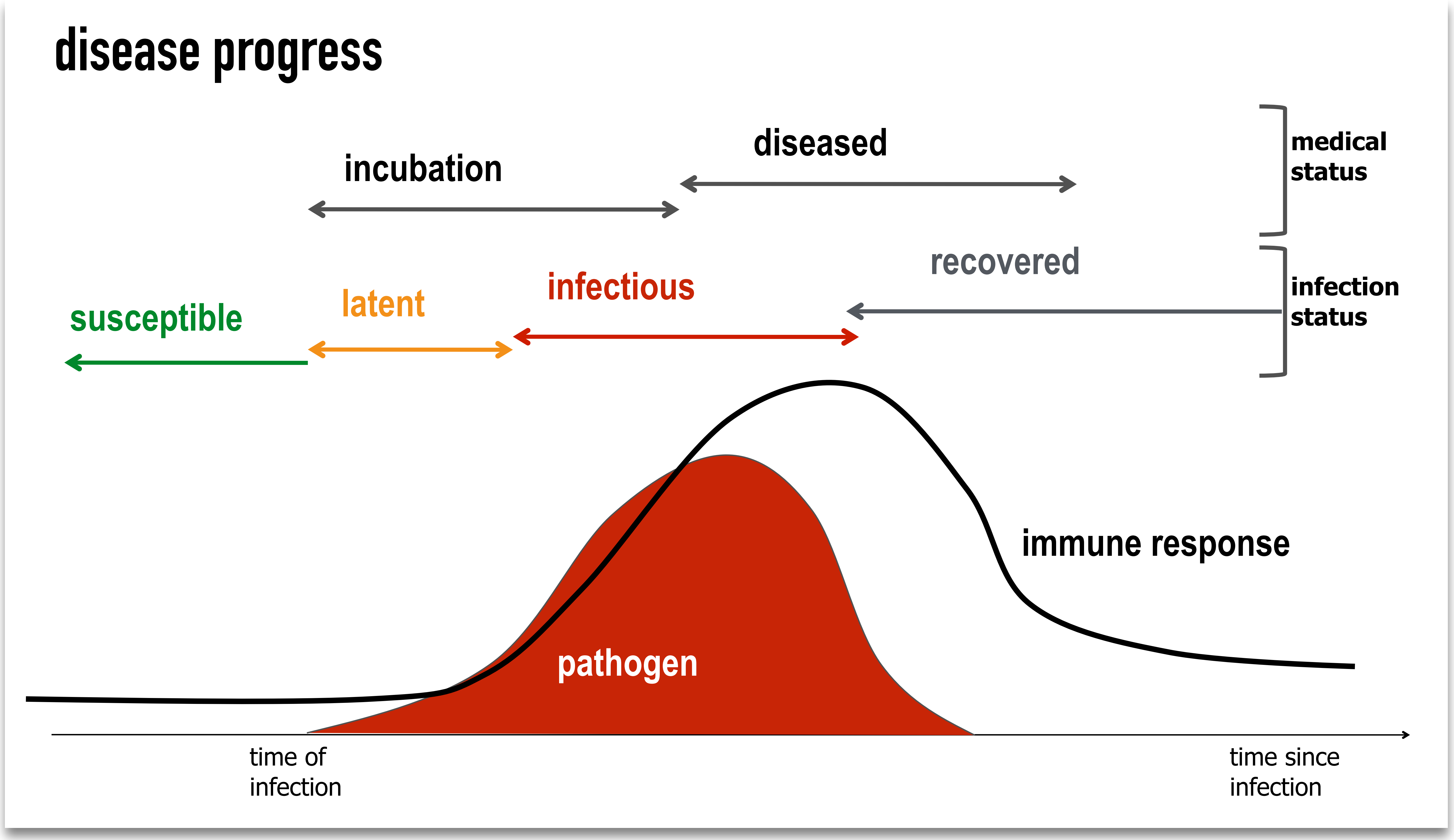
a biologist



a chemist

a physicist

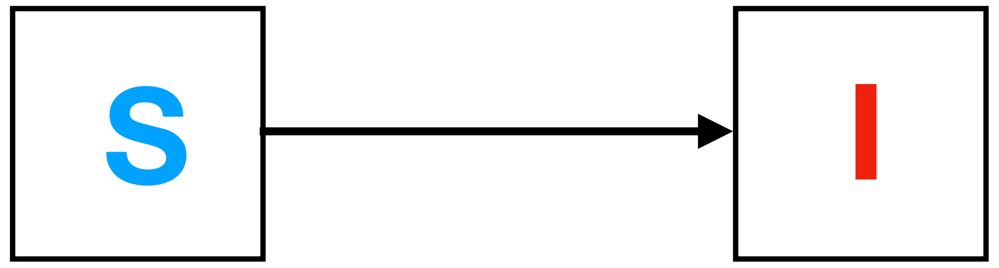
How infection works



Credits: Chiara Poletto

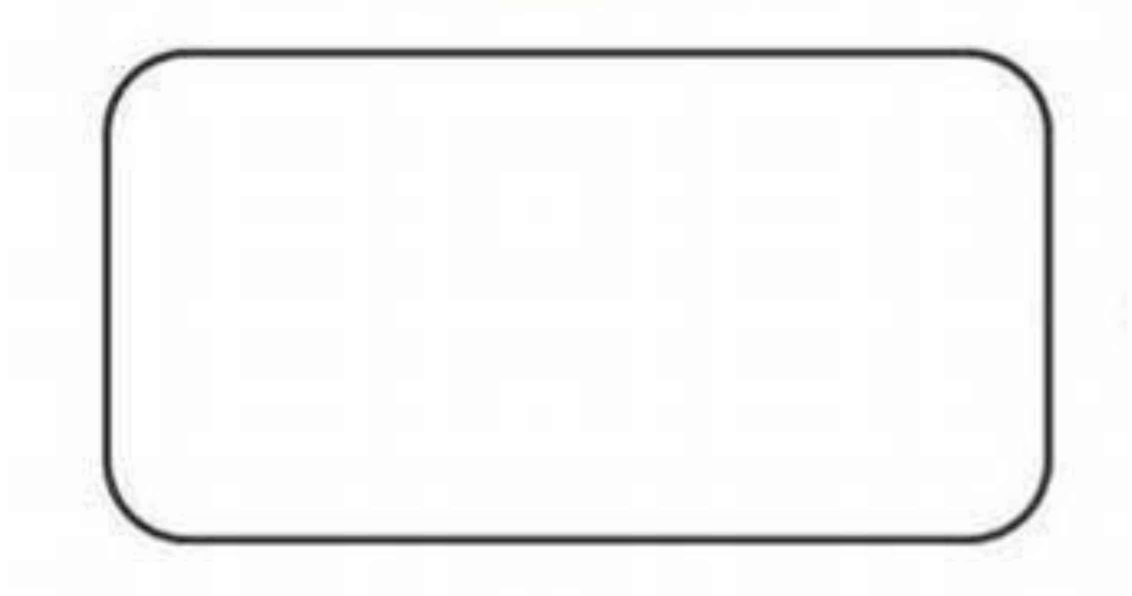
Compartmental models

Susceptible-Infected model (SI)



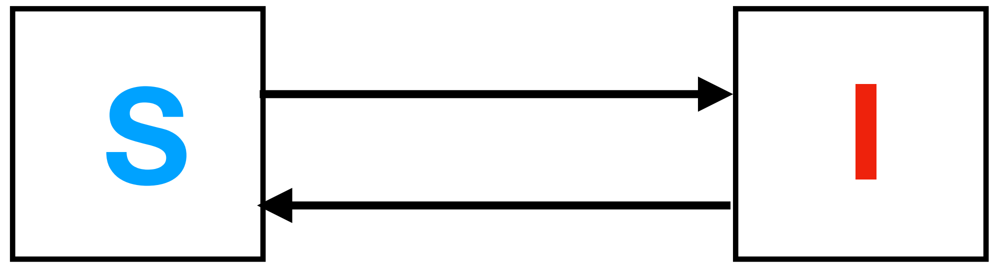
S

I



A cell as seen by a physicist

Susceptible-Infected-Susceptible model (SIS)



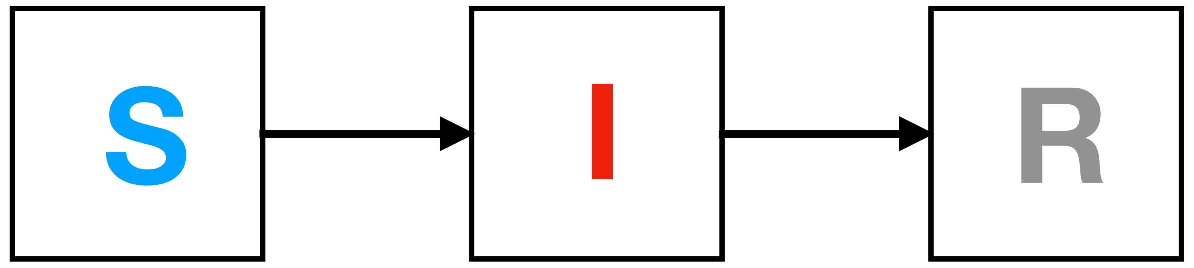
S

I

Infected can recover

Individuals are sorted into different compartments according to infection status

Susceptible-Infected-Recovered model (SIR)



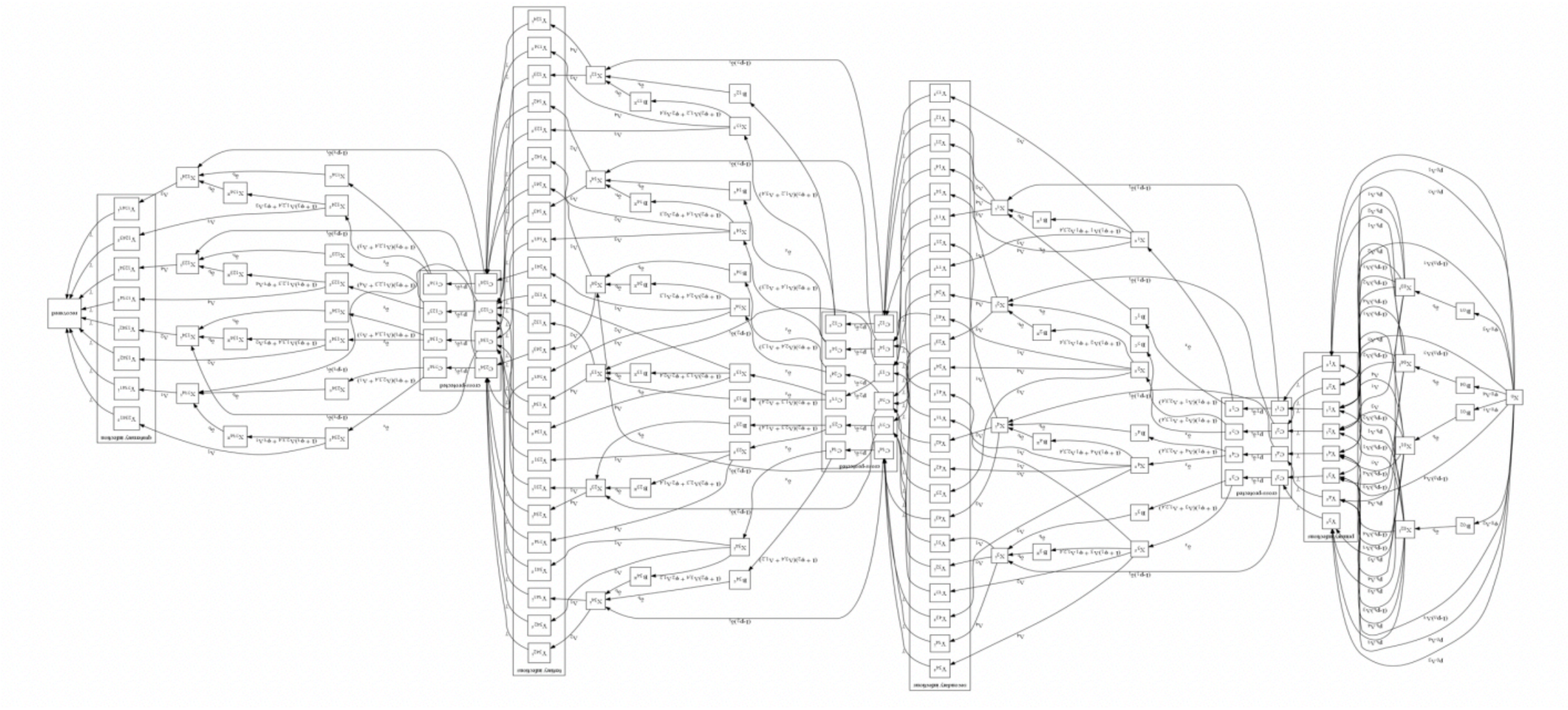
S

I

R

Infected recover and become immune

Things can escalate quickly: a model for Dengue virus



Alexander et al, PNAS, 2021

One-body reactions

Require a single entity

$$\mathbf{A} \xrightarrow{\alpha} \mathbf{B}$$

*“An entity of type **A** spontaneously morphs into **B** at rate α"*

Examples: disease progression

The number of **A**’s evolves according to:

$$\frac{d}{dt}n_A = -\alpha \cdot n_A$$

Reaction rate $\propto n_A$

Note: time to reaction is exponentially distributed: $Pr(\text{no reaction up to } t) = e^{-\alpha t}$

Two-body reactions

Require two entities

$$\mathbf{A} + \mathbf{B} \xrightarrow{\beta} \mathbf{A} + \mathbf{A}$$

"An entity of type ***B*** *meets an entity of type* ***A*** *and morphs into* ***A*** *at rate β"*

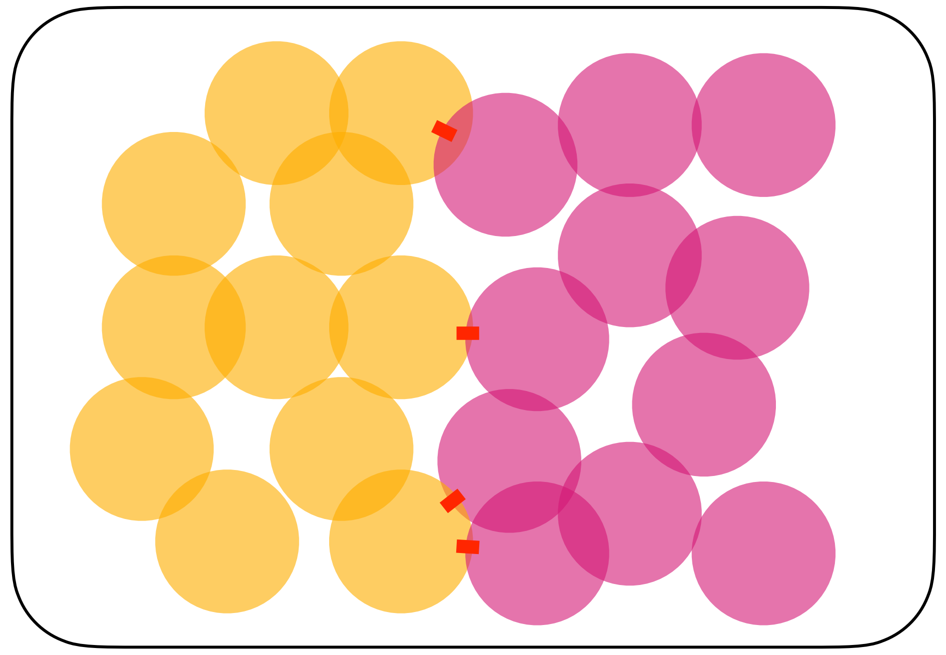
Example: pathogen transmission

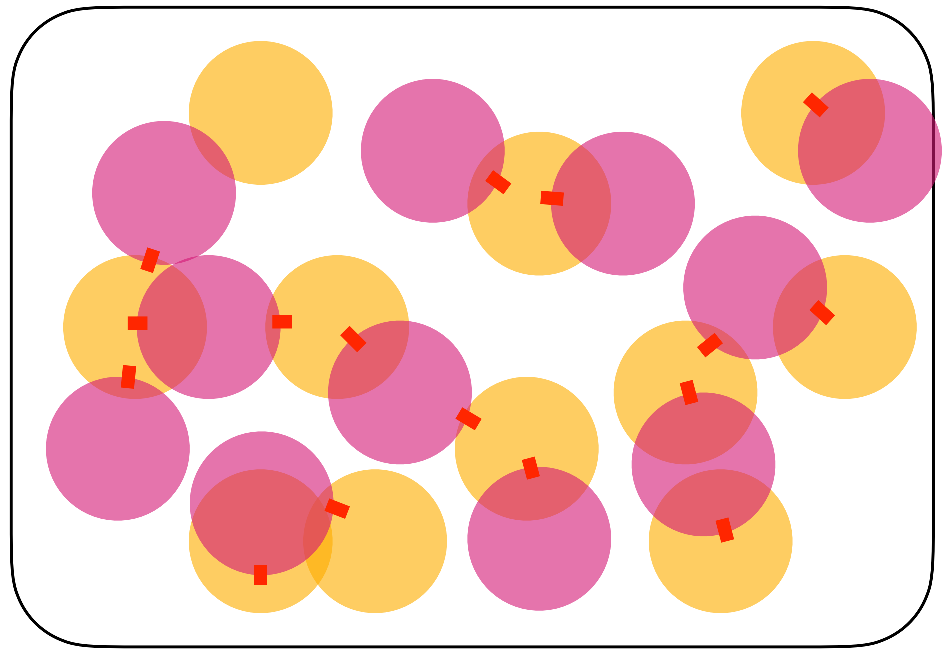
The number of **A**'s evolves according to:

$$\frac{d}{dt}n_A = \beta \cdot \boxed{c_{AB}} \longrightarrow$$

Mean number of encounters between types **A** and **B**

$c_{A,B}$ depends on how **A** and **B** are arranged with respect to each other





A simplification: homogeneous mixing

Key idea: individuals are all equivalent and mix completely at random

- Host equivalence: all hosts make the same number of contacts $\langle k \rangle$
- Contact equivalence: same probability to bump into Barack Obama or Francesco Pinotti

$$c_{AB} = n_A \cdot \langle k \rangle \cdot \frac{n_B}{N}$$

Nr. of **A** types

Avg. contacts per individual

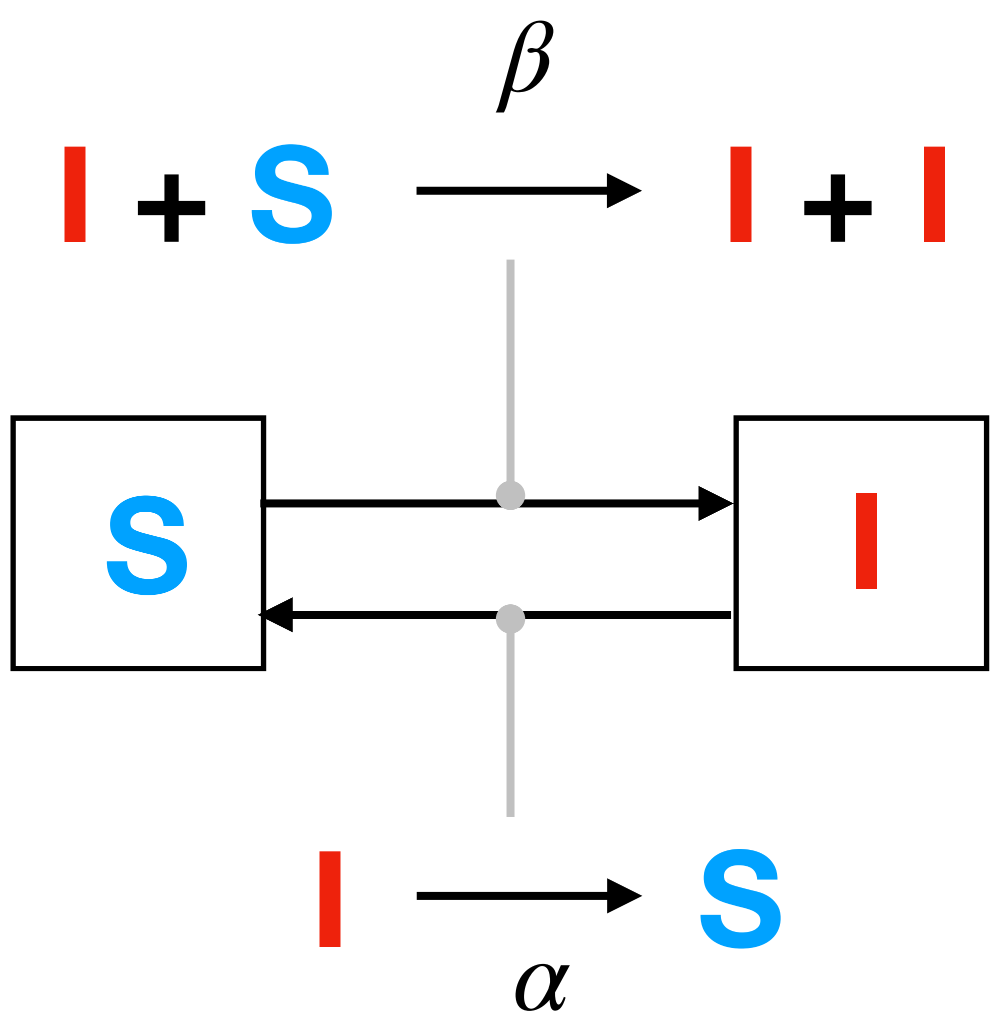
Probability to meet a **B** type

Total contacts of **A** types

$$\mathbf{A} + \mathbf{B} \xrightarrow{\beta} \mathbf{A} + \mathbf{A}$$

$$\frac{d}{dt} n_A = - \beta \langle k \rangle \frac{n_A n_B}{N}$$

SIS dynamics: homogeneous mixing



$$I + S \xrightarrow{\beta} I + I$$

$$I \xrightarrow[\alpha]{} S$$

SIS equations

$$\frac{d}{dt}n_S = -\beta \cdot \frac{n_I}{N} \cdot n_S + \alpha n_I$$

$$\frac{d}{dt}n_I = +\beta \cdot \frac{n_I}{N} \cdot n_S - \alpha n_I$$

Population size is constant $\frac{d}{dt}(n_S + n_I) = 0 \longrightarrow n_S = N - n_I$

One equation is enough $\frac{d}{dt}n_I = +\beta \cdot \frac{n_I(N - n_I)}{N} - \alpha n_I$

Define $I = n_I/N$ $\frac{dI}{dt} = +\beta \cdot I(1 - I) - \alpha I \equiv F(I)$

Fixed point analysis

Solve $F(I^*) = 0$

$I_1^* = 0$ Extinction

$I_2^* = 1 - \frac{\alpha}{\beta}$ Endemicity

SIS dynamics: homogeneous mixing

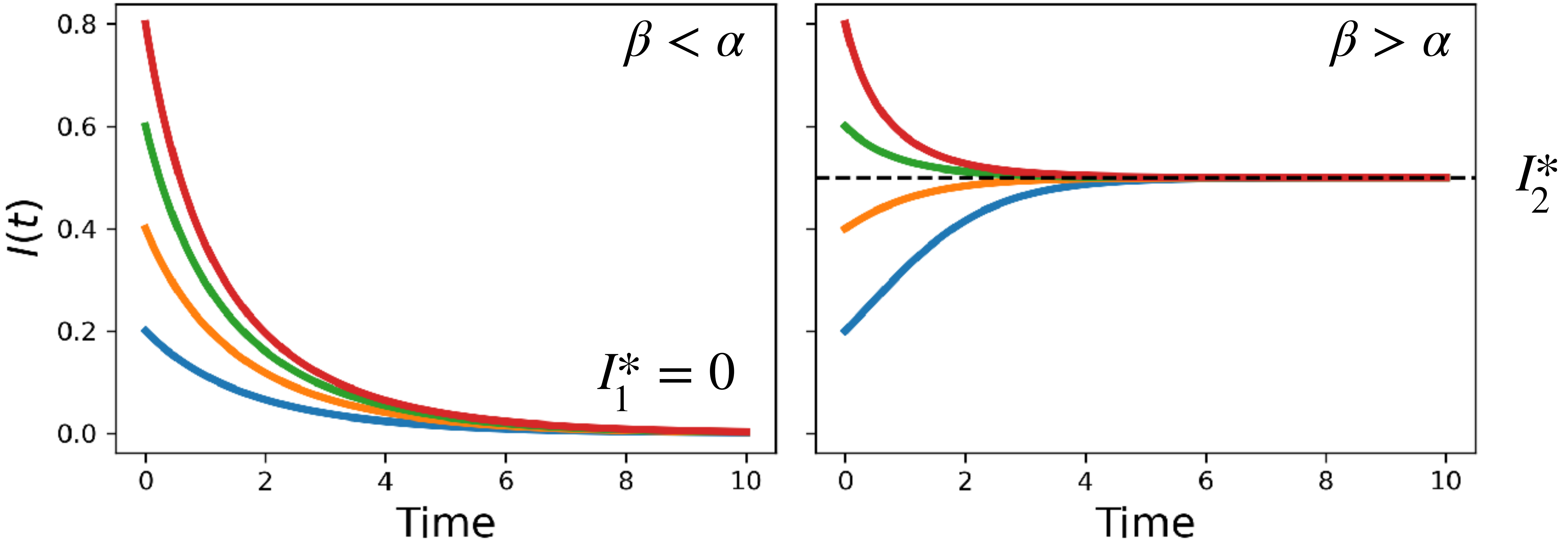
Linear stability analysis

Assume $I(t = 0) \approx 0$

$$\frac{dI}{dt} \approx (\beta - \alpha)I$$

$\beta < \alpha$: I decays to I_1^* (**extinction**)

$\beta > \alpha$: I flows to I_2^* (**endemicity**)



The basic reproductive number

The condition for endemicity can be expressed as:

$$\boxed{R_0 \equiv \frac{\beta}{\alpha}} > 1$$

Basic reproductive number

R_0 is the mean number of secondary infections generated by an infected in an entirely susceptible population

$$R_0 = \beta \langle k \rangle \times \frac{1}{\mu}$$

We reintegrated $\langle k \rangle$

Average number of infected contacts

X

Average infection duration

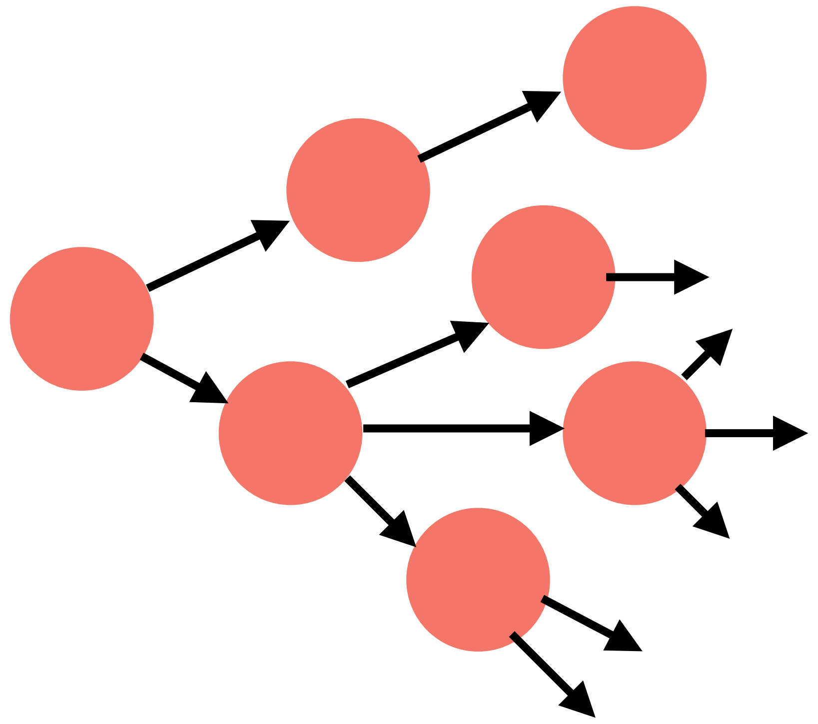
The basic reproductive number

The condition for endemicity can be expressed as: $\boxed{R_0 \equiv \frac{\beta}{\alpha}} > 1$

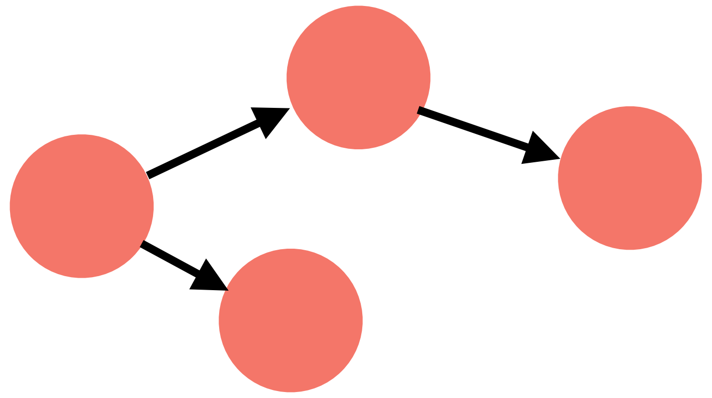
Basic reproductive number

R_0 is the mean number of secondary infections generated by an infected in an entirely susceptible population

$R_0 > 1$: Exponential growth

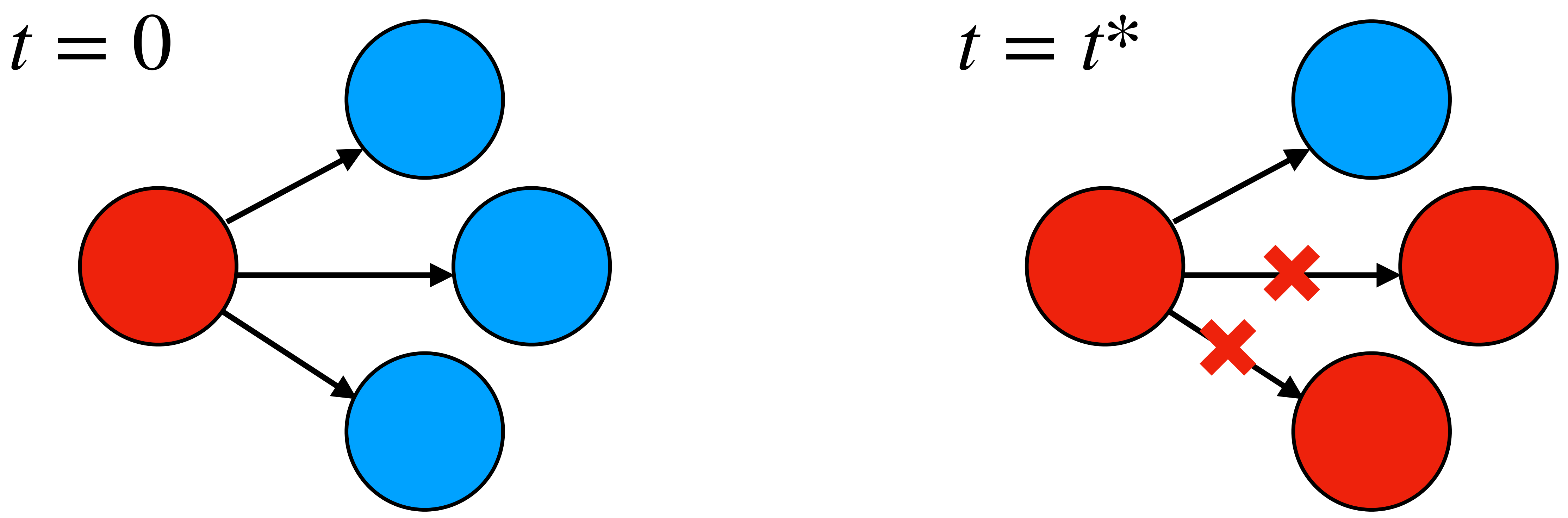


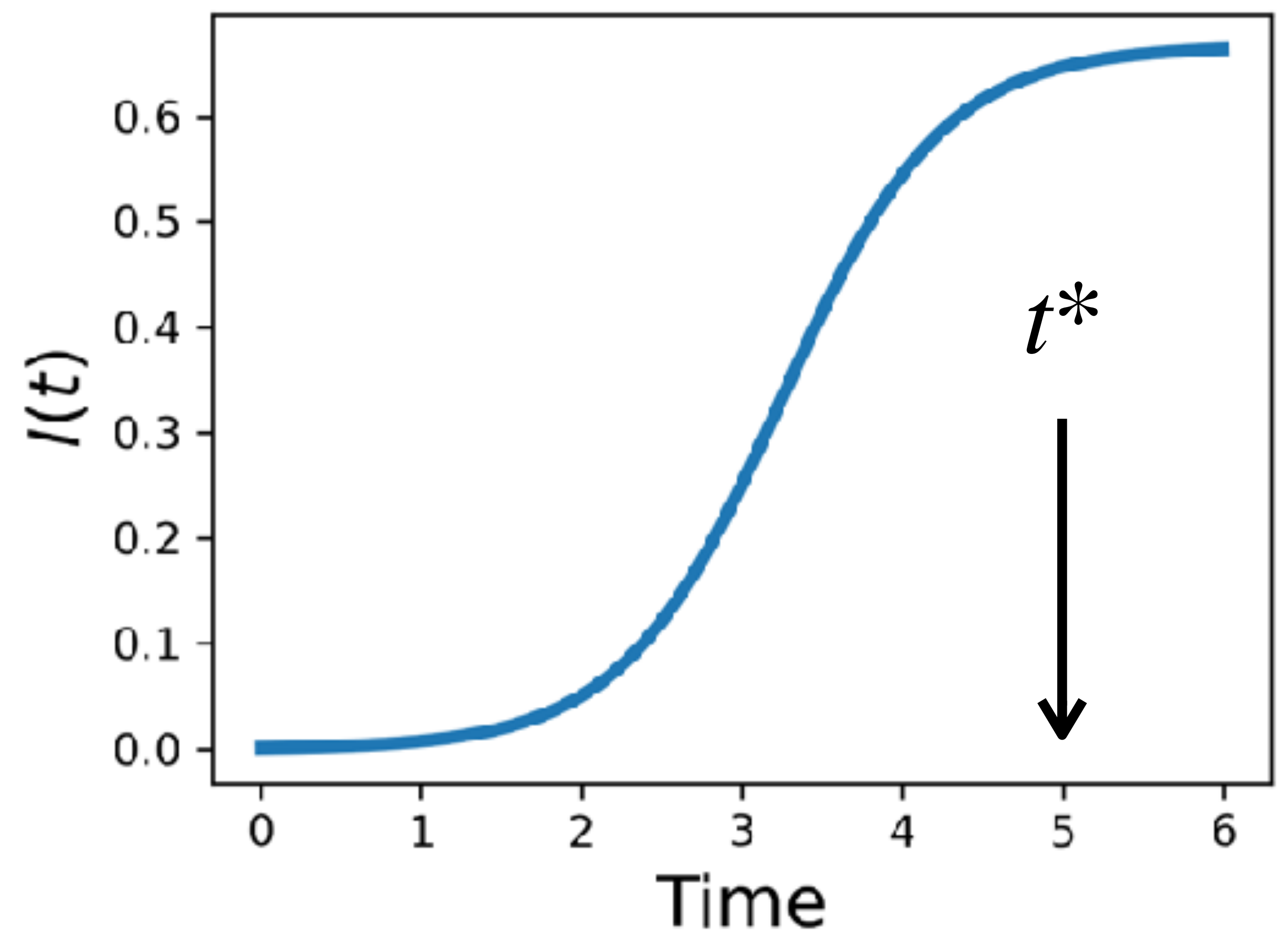
$R_0 < 1$: Stuttering chains



$R_0 > 1$: unlimited exponential growth?

Epidemic slows down with fewer susceptible people





Effective reproductive number:

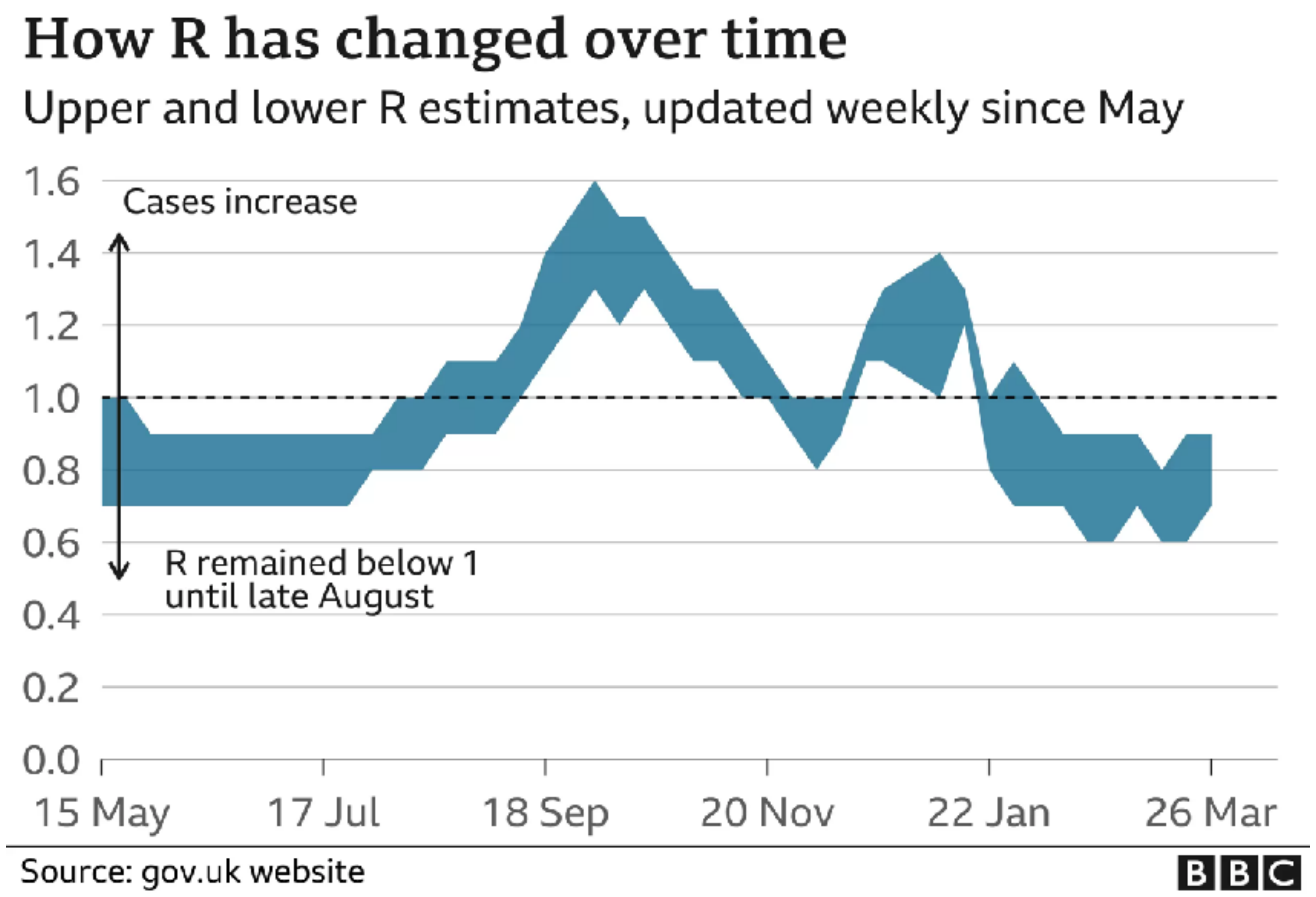
$$R_t = R_0 \cdot S(t)$$

Reproductive potential

Pop-level susceptibility

- $R_t > 1$ epidemic grows
- $R_t < 1$ epidemic grows
- $R_t = 1$ equilibrium

R_t for SARS-CoV2 in England


How R has changed over time

Upper and lower R estimates, updated weekly since May

Source: gov.uk website

Effective reproductive number:

$$R_t = R_0 \cdot S(t)$$

Reproductive
potential

Pop-level
susceptibility

- $R_t > 1$ epidemic grows
- $R_t < 1$ epidemic grows
- $R_t = 1$ equilibrium

Vaccination

Immunising people reduces S and hence R_t . If $R_t < 1$ no epidemic occurs.

We vax a fraction u of people (everyone is susceptible initially): $R_t = R_0(1 - u)$

Herd-immunity threshold

$$R_t < 1 \quad \Longleftrightarrow \quad u > \boxed{u^* = 1 - \frac{1}{R_0}}$$

+ No need to vax everyone

— Vaccine is not perfect

— Immunity waning

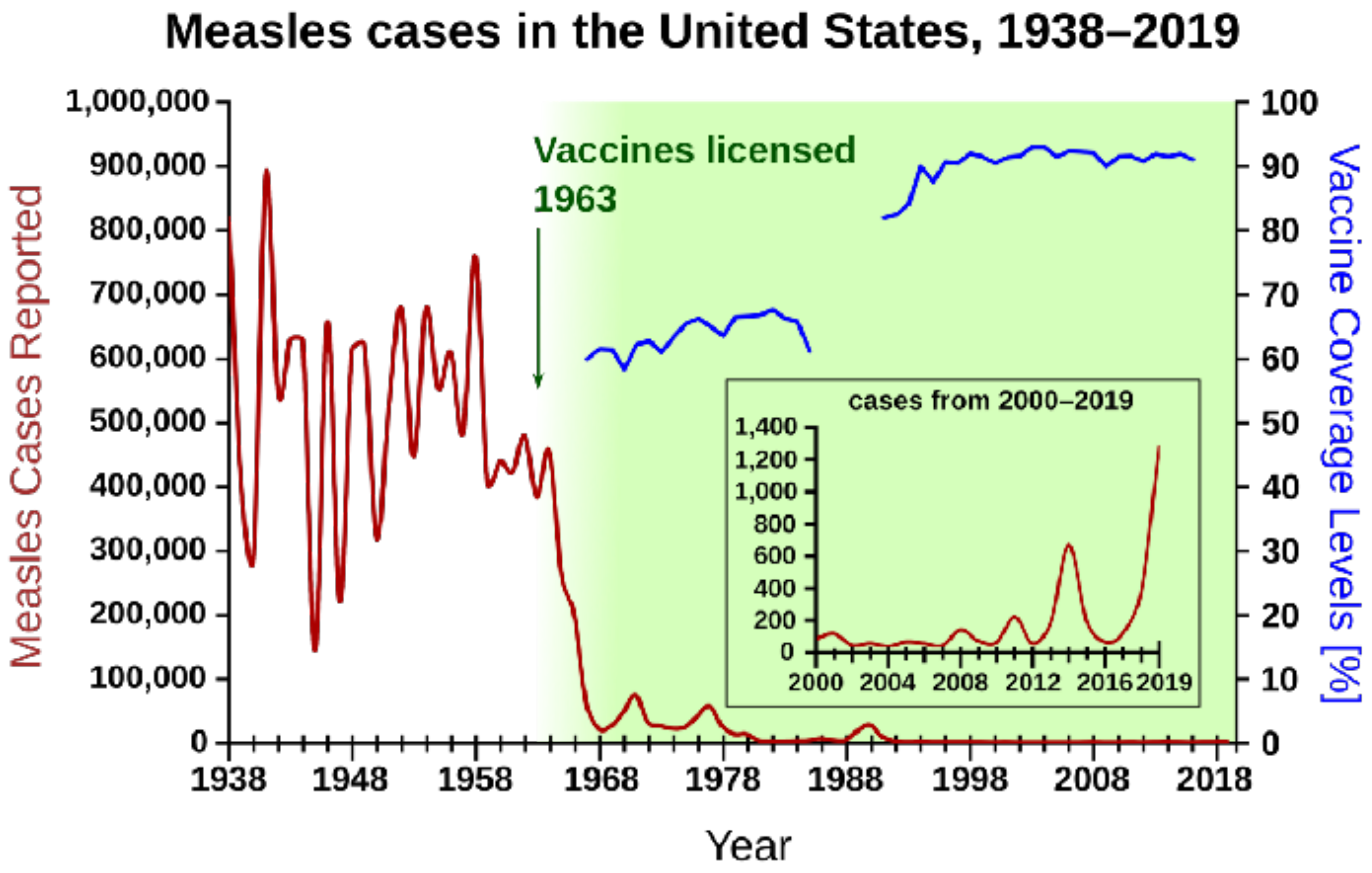
— Demographic turnover

— Vaccine hesitancy

Measles:

$R_0 \approx 15$

$u^* \approx 0.93$



From Wikipedia

Stochastic epidemics

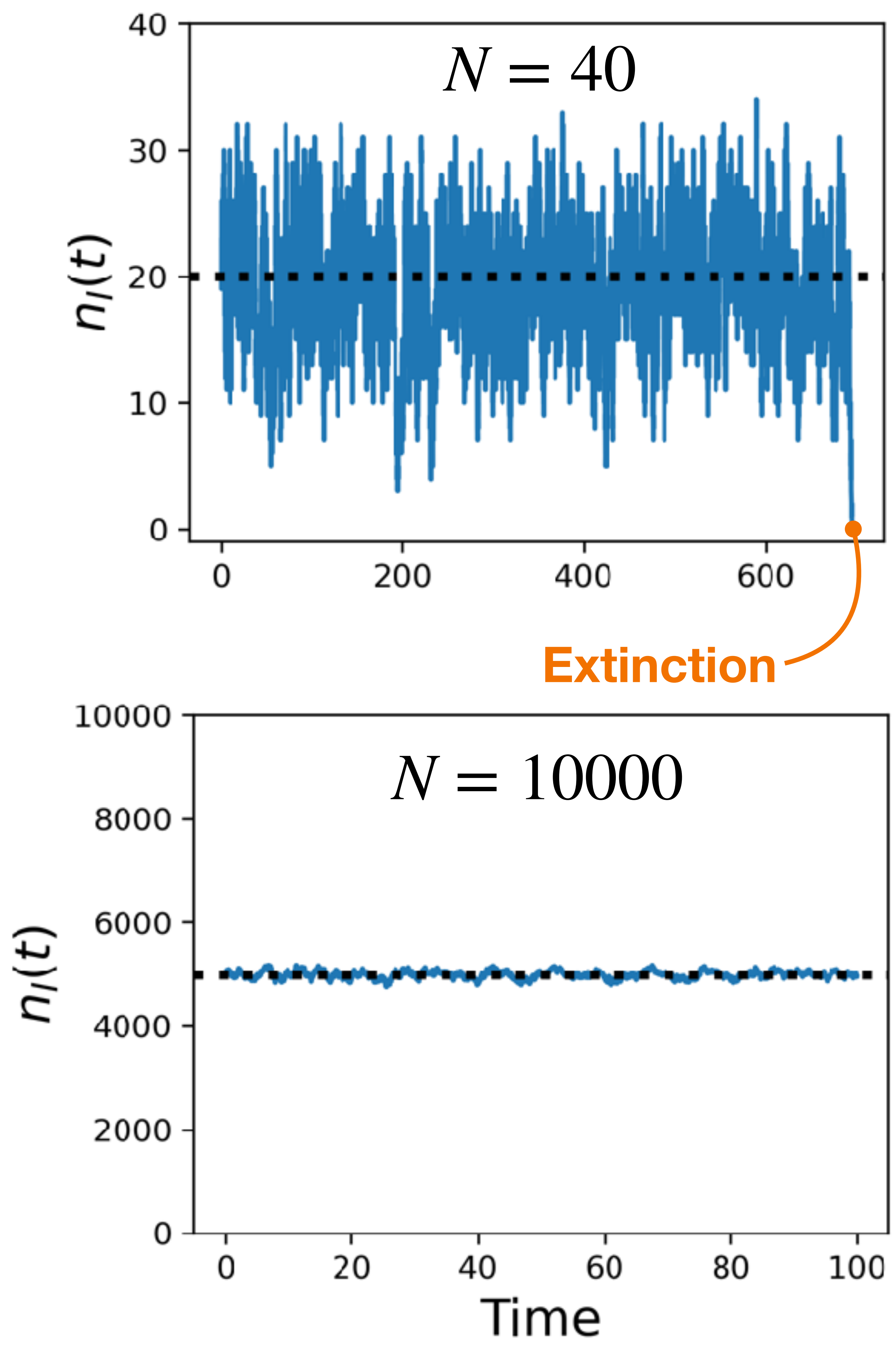
- Deterministic ODEs describe the **average** behaviour of an outbreak.
- Outbreaks are inherently random and discrete
- What does $I(t) = 0.0001$ mean when there are only 40 individuals?
- Deterministic ODEs fail to describe rare extinction events

Master equation

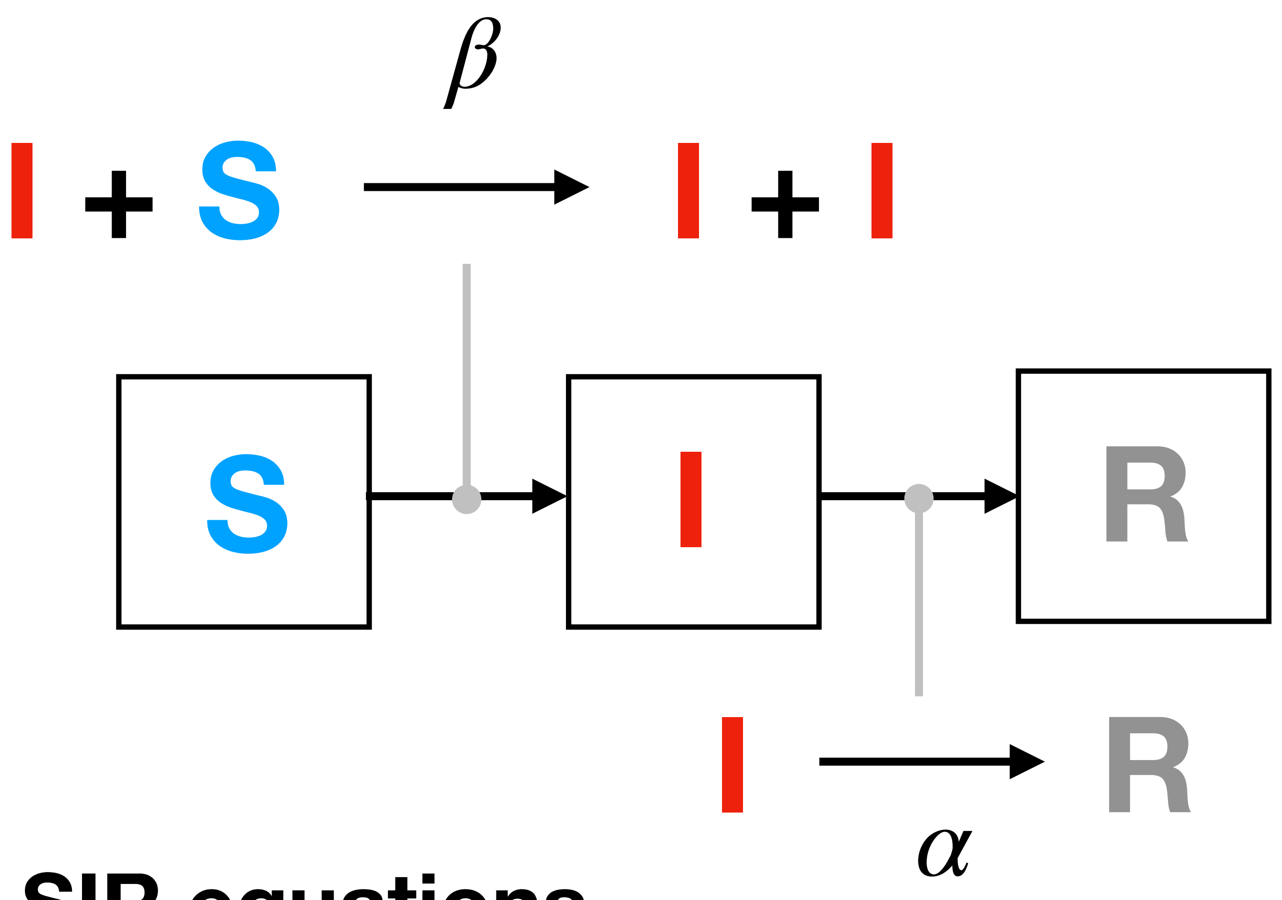
$$\frac{d}{dt}P_t(n_I) = w_+(n_I - 1) \cdot P_t(n_I - 1) + \\ + w_-(n_I + 1) \cdot P_t(n_I + 1) - \\ - [w_+(n_I) + w_-(n_I)] \cdot P_t(n_I)$$

$$w_+(n_I) = \frac{\beta n_I (N - n_I)}{N} \qquad w_-(n_I) = \mu n_I$$

$R_0 = 2$ $\quad$ $I^* = 0.5$



SIR dynamics: homogeneous mixing



$I + S \xrightarrow{\beta} I + I$

$I \xrightarrow{\alpha} R$

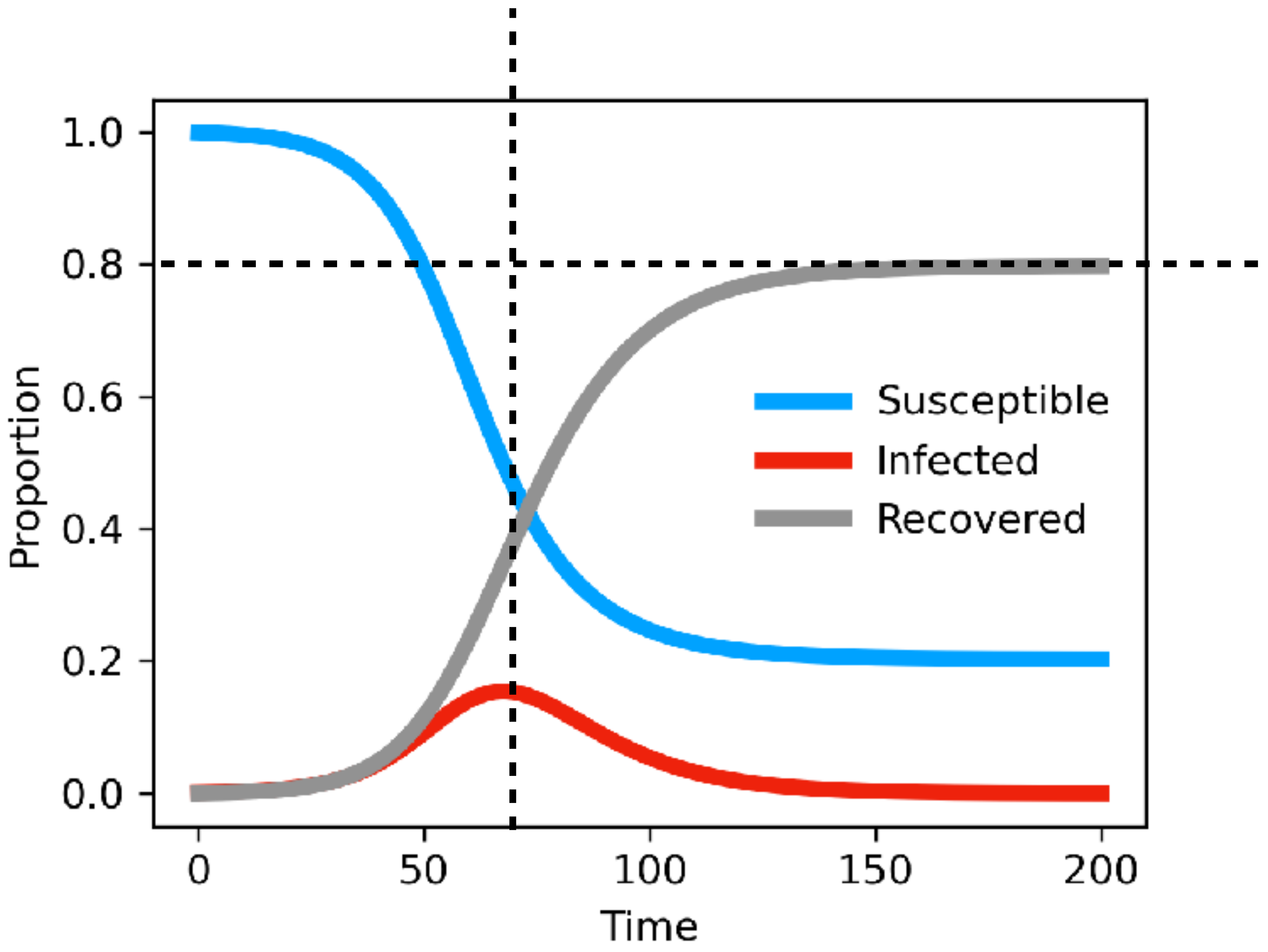
SIR equations

$$\frac{d}{dt}S = -\beta SI$$

$$\frac{d}{dt}I = +\beta SI - \alpha I$$

$$\frac{d}{dt}R = -\alpha I$$

This model describes a single outbreak when $\beta > \alpha$.
The accumulation of immunity leads to the end of the epidemic.



Epidemic peak

Final size R^*

SIR dynamics: final size formula

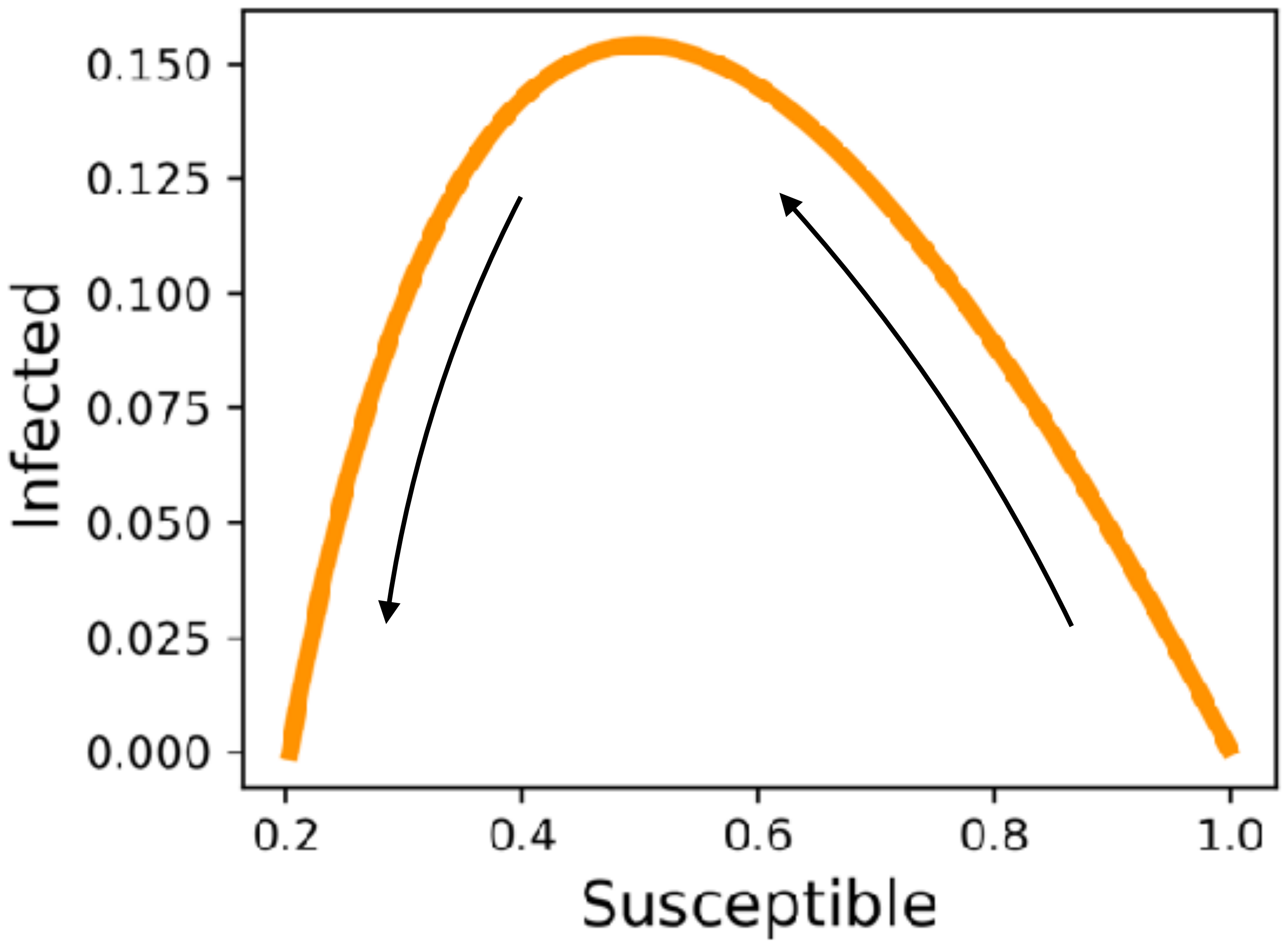
How many people are infected during a single outbreak? We may not know $I(t)$ or $S(t)$ but can derive $S(I)$

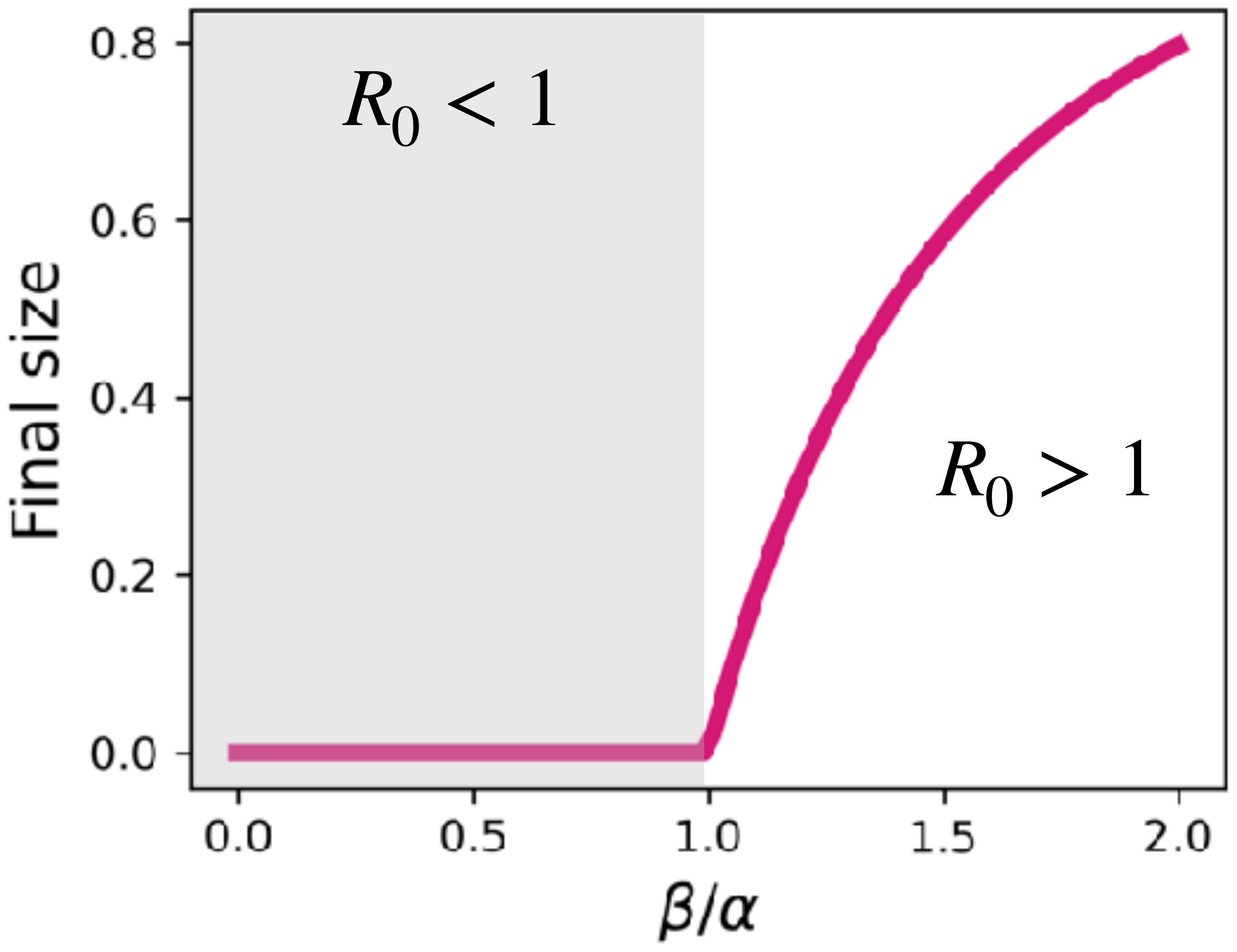
$$\frac{dS}{dI} = \frac{\beta S}{\beta S - \alpha}$$

$$R^* = 1 - S^*$$

$$1 - R^* = \exp\left(-\frac{\beta}{\alpha}R^*\right)$$

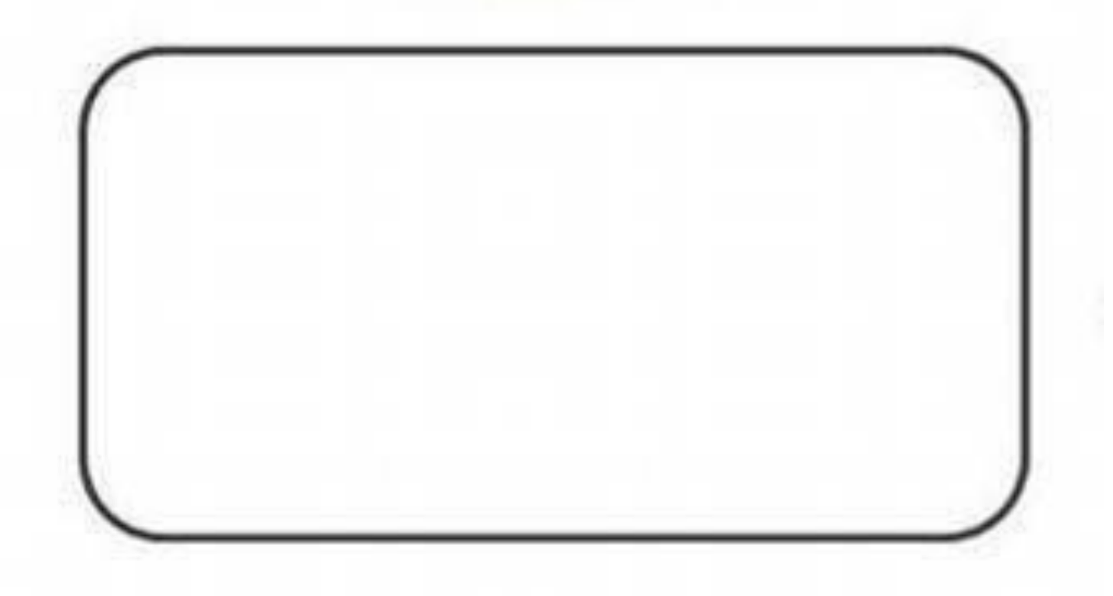
Final size formula





(Some) limitations of simple models

- Determinism → Use stochastic version
- Fixed parameters → Use time-varying parameters $\beta \to \beta(t)$
- Constant population → Include vital dynamics
- Exponential transition times → Method of stages/integro-differential equations
- Not accounting for population heterogeneities → Ad-hoc compartments, e.g. $I(t) \to I(\text{age}, t)$
- Simplistic contact hypotheses → Use **network**, **spatial** and **agent-based** models



A cell as seen by a physicist

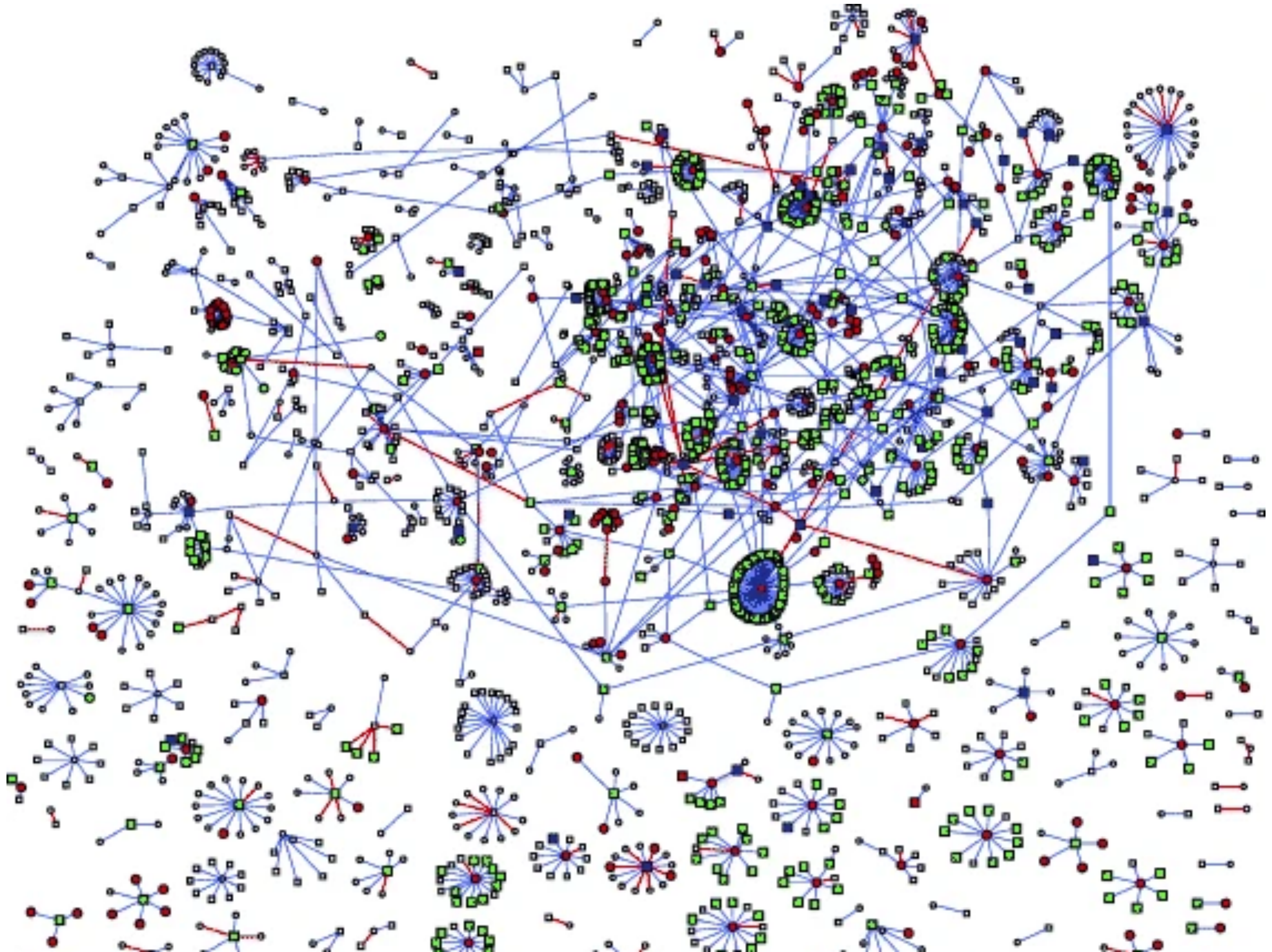
Epidemics on networks

Network epidemiology

Homogeneous mixing assumptions

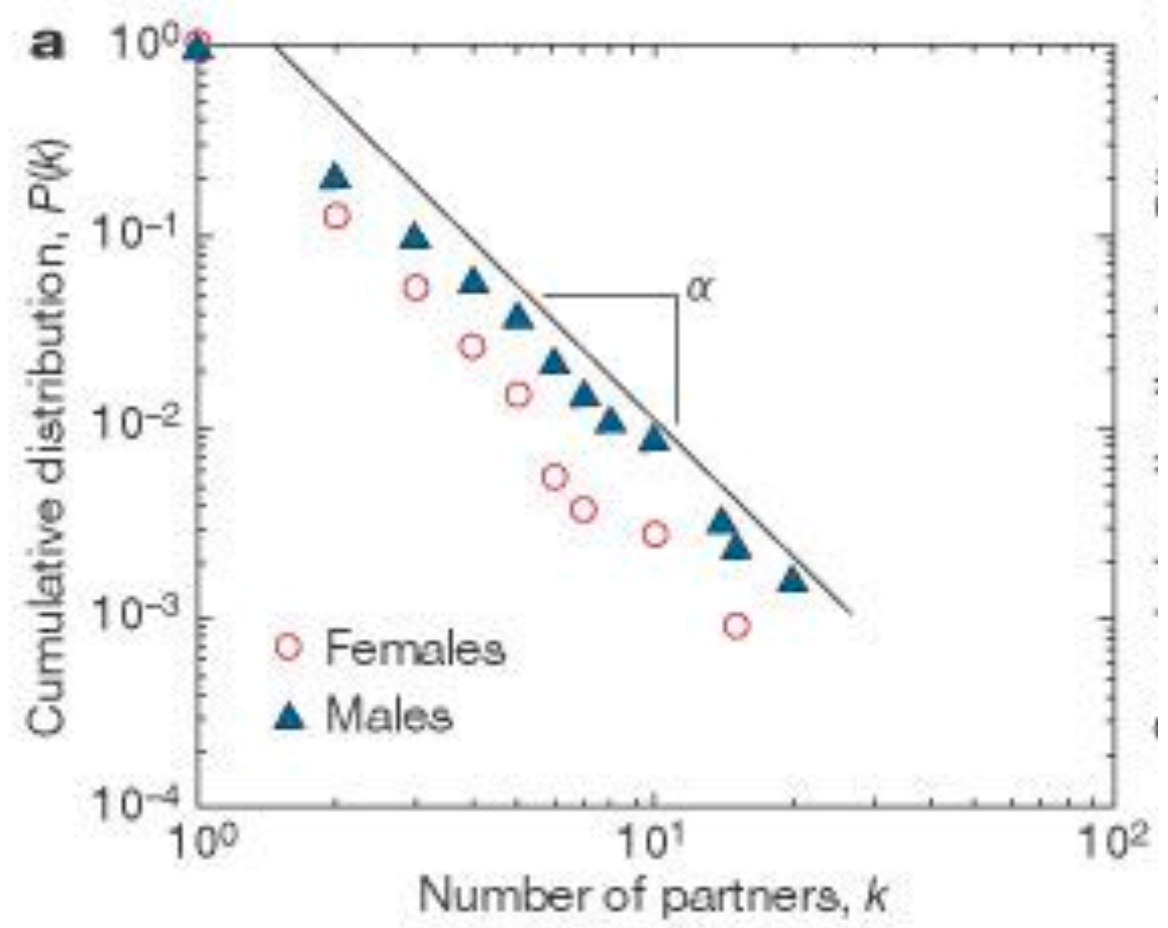
1. Host equivalence → Contact heterogeneities
2. Contact equivalence → Some contacts more likely than others

Sexual network
in a US town



Liljeros et al, Nature, 2001
Potterat et al, 2004
Liljeros, 2009

# of partners in the last year



$$P(k) \propto k^{-\alpha}, \quad \alpha \simeq 2.3$$

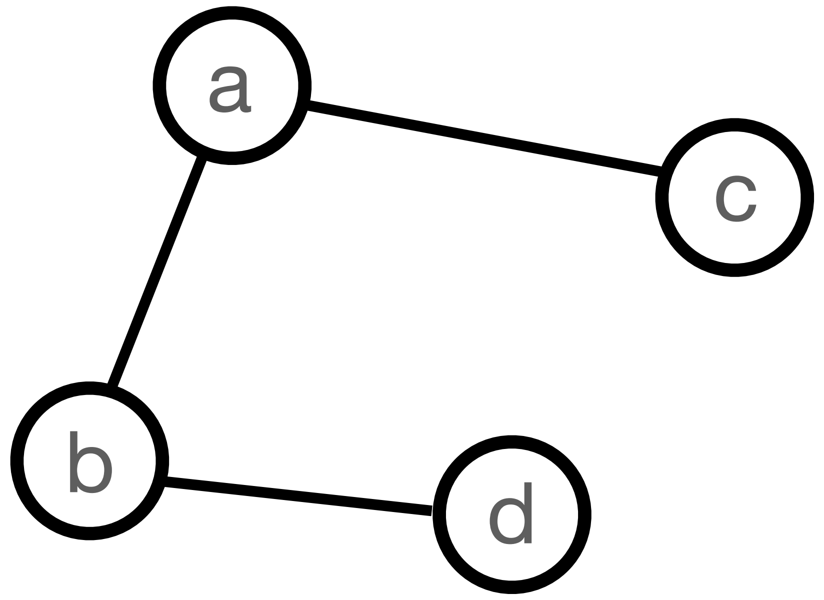
Risk is heterogeneous!!!

Not just sexual contacts: epidemiological relevance of contacts depends on host & pathogen

Network basics

We are concerned with simple graphs:

- Undirected edges
- No self-edges
- No multi-edges



Node set

$$V = \{a, b, c, d\}$$

Edge set

$$E = \{(a, b), (a, c), (b, d)\}$$

Adjacency matrix

$$A_{i,j} = 1 \text{ if } (i, j) \in E$$
$$A_{i,j} = 0 \text{ otherwise}$$

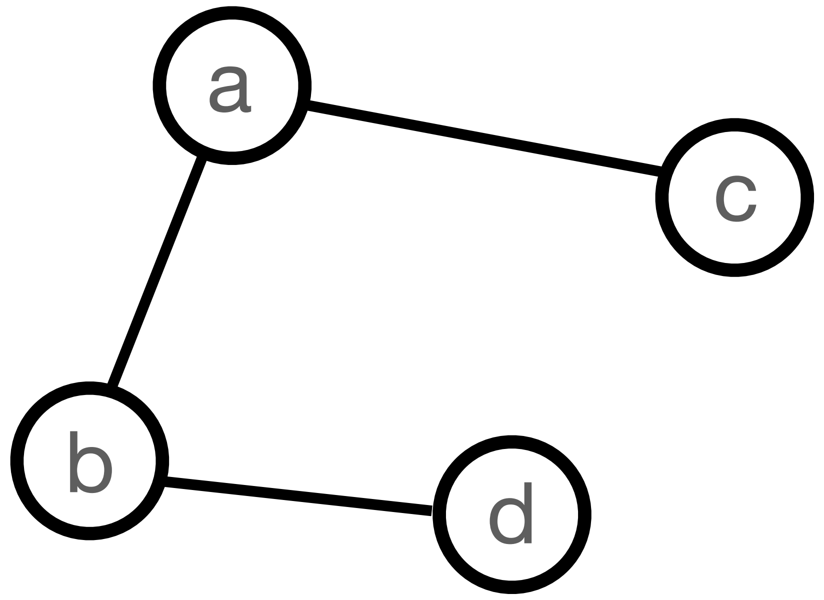
$$A = \begin{array}{cccccc} & a & b & c & d & \\ & 0 & 1 & 1 & 0 & a \\ & 1 & 0 & 0 & 1 & b \\ & 1 & 0 & 0 & 0 & c \\ & 0 & 1 & 0 & 0 & d \end{array}$$

Network basics

Node degree:

$$k_i = \sum_j A_{i,j}$$

$$k_a = k_b = 2 \qquad k_c = k_d = 1$$



Node set

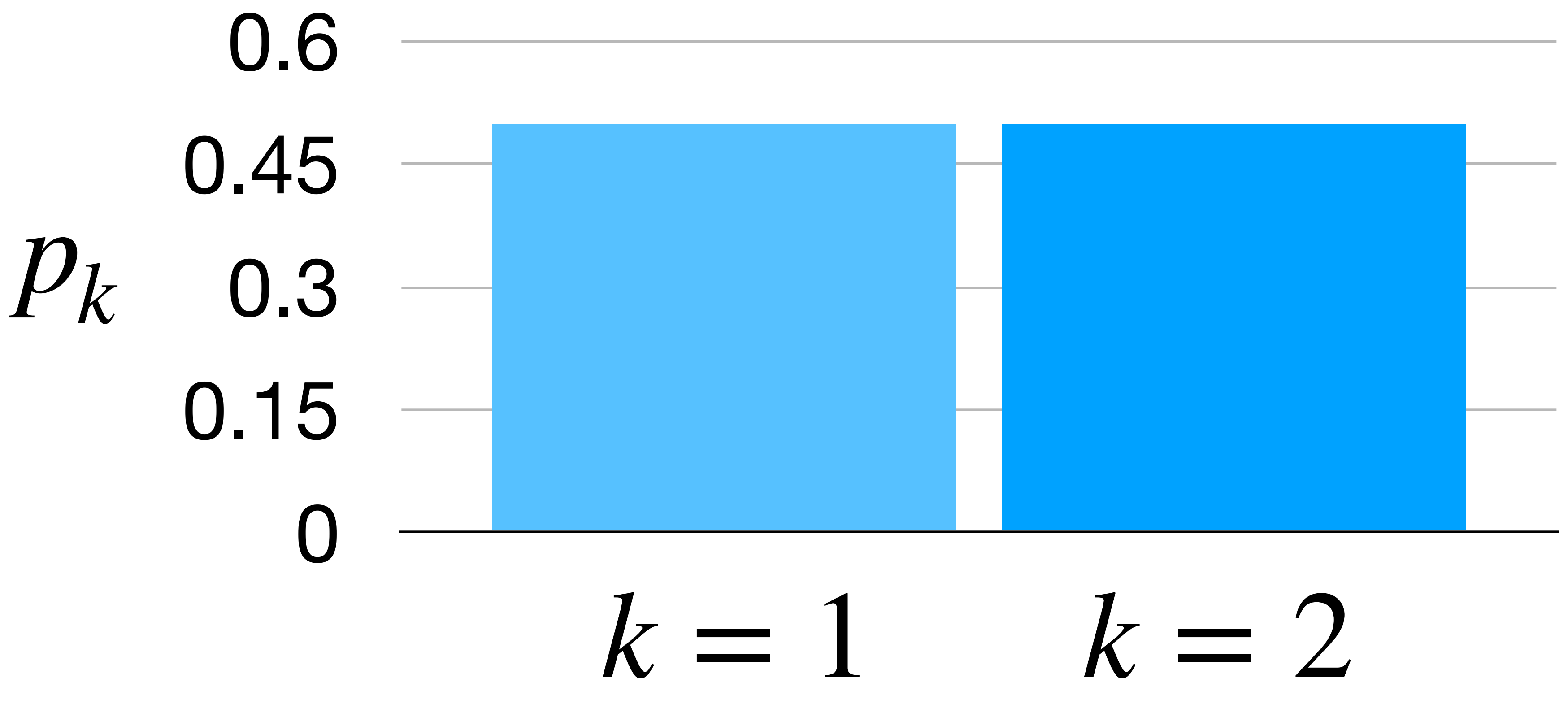
$$V = \{a, b, c, d\}$$

Edge set

$$E = \{(a,b), (a,c), (b,d)\}$$

The **degree distribution** is defined as

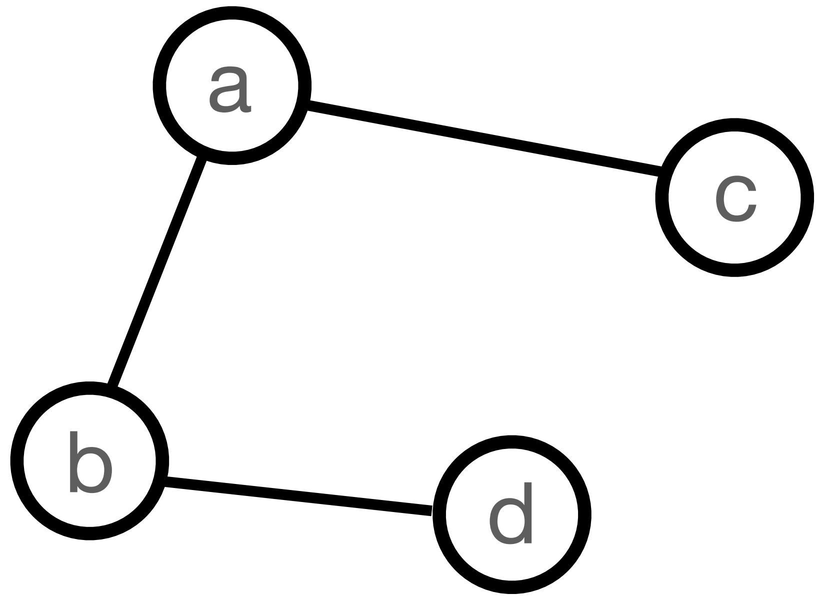
$$p_k = \frac{n_k}{N}$$



Degree moments

Degree moments

$$\langle k^n \rangle = \sum_{k=1}^{\infty} k^n p_k$$



Node set

$$V = \{a, b, c, d\}$$

Edge set

$$E = \{(a, b), (a, c), (b, d)\}$$

Average degree

$$\langle k \rangle = \sum_{k=1}^{\infty} k p_k$$

Degree variance

$$Var(k) = \langle k^2 \rangle - \langle k \rangle^2$$

Note that $\langle k^2 \rangle \geq \langle k \rangle^2$

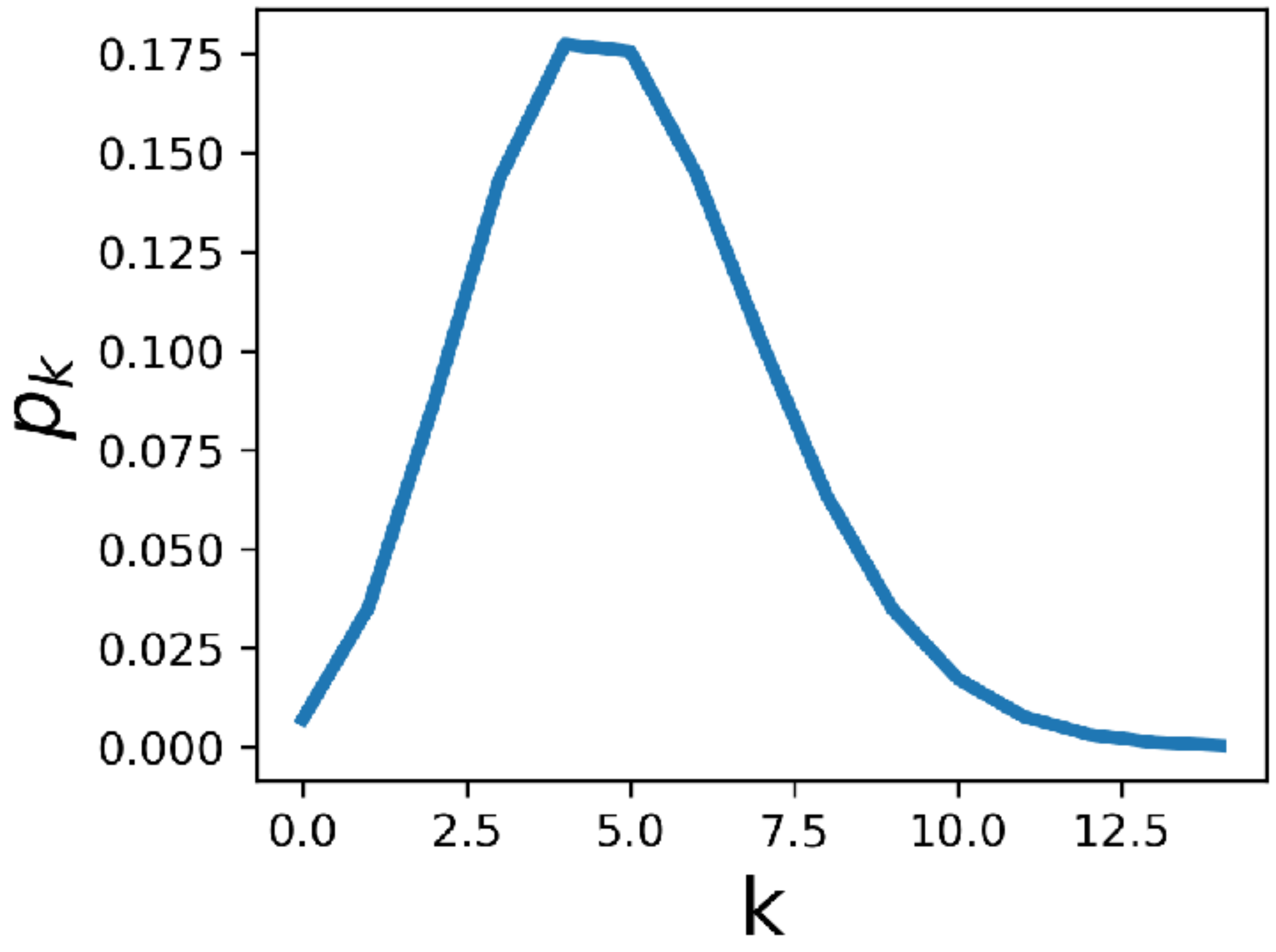
Poisson degree distributions

Poisson degree distribution

$$p_k = e^{-\langle k \rangle} \frac{\langle k \rangle^k}{k!}$$

- It's narrow ($Var(k) = \langle k \rangle$)
- Very small tail because of $1/k!$

Example: Erdős-Rényi model



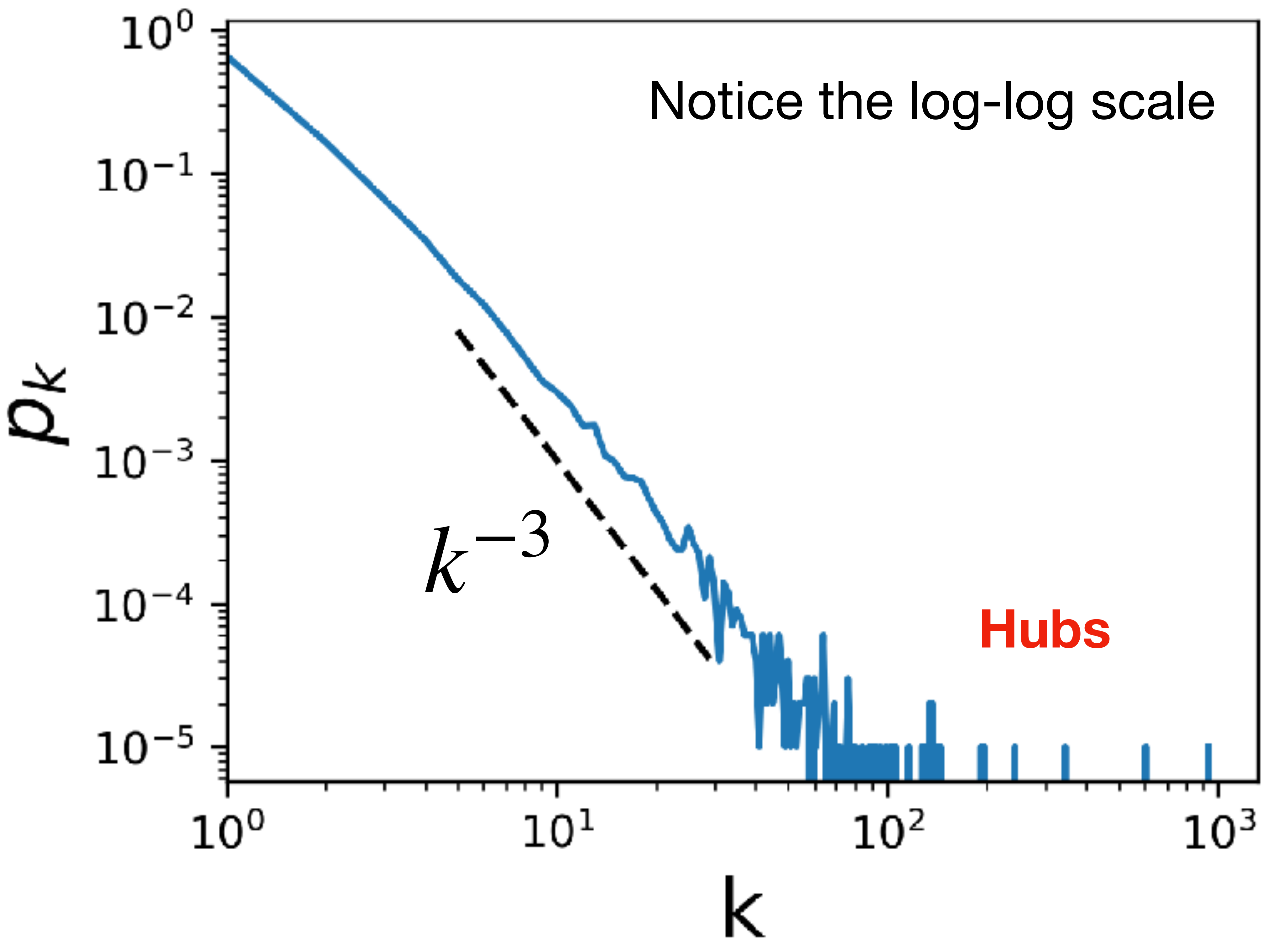
Power-law degree distributions

Power-law degree distribution

$$p_k \propto k^{-\gamma}$$

- Slower decay than exponential
- Wide range of degree values
- Infinite variance when $\gamma < 3$

Example: Barabasi-Albert model (has $\gamma = 3$)



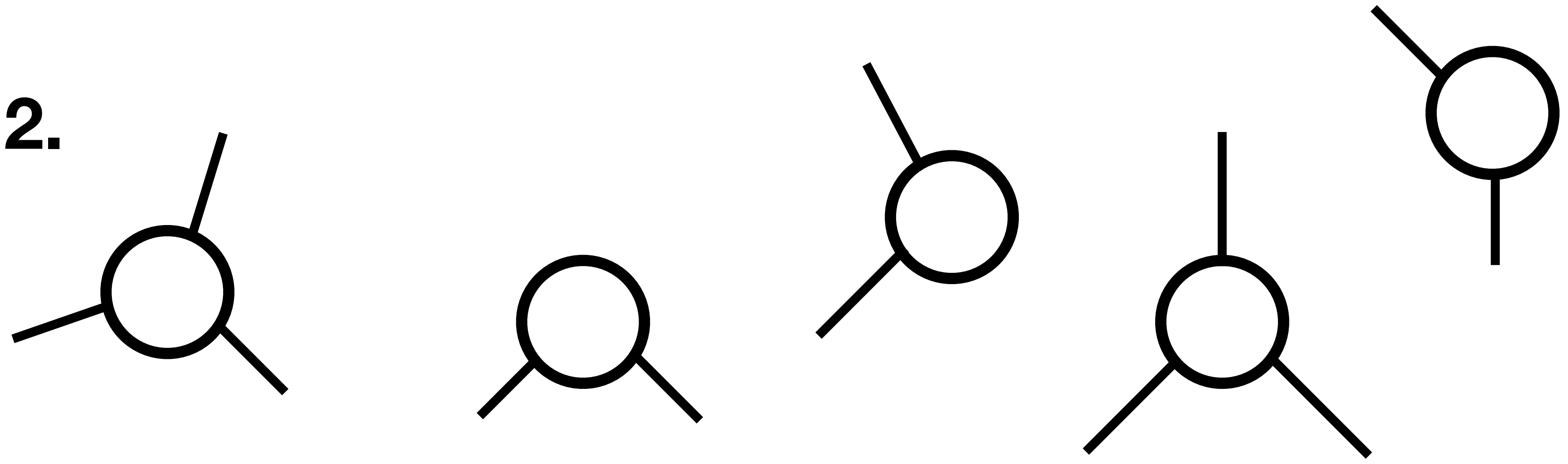
The configuration model

Generate networks with the **desired degree distribution**

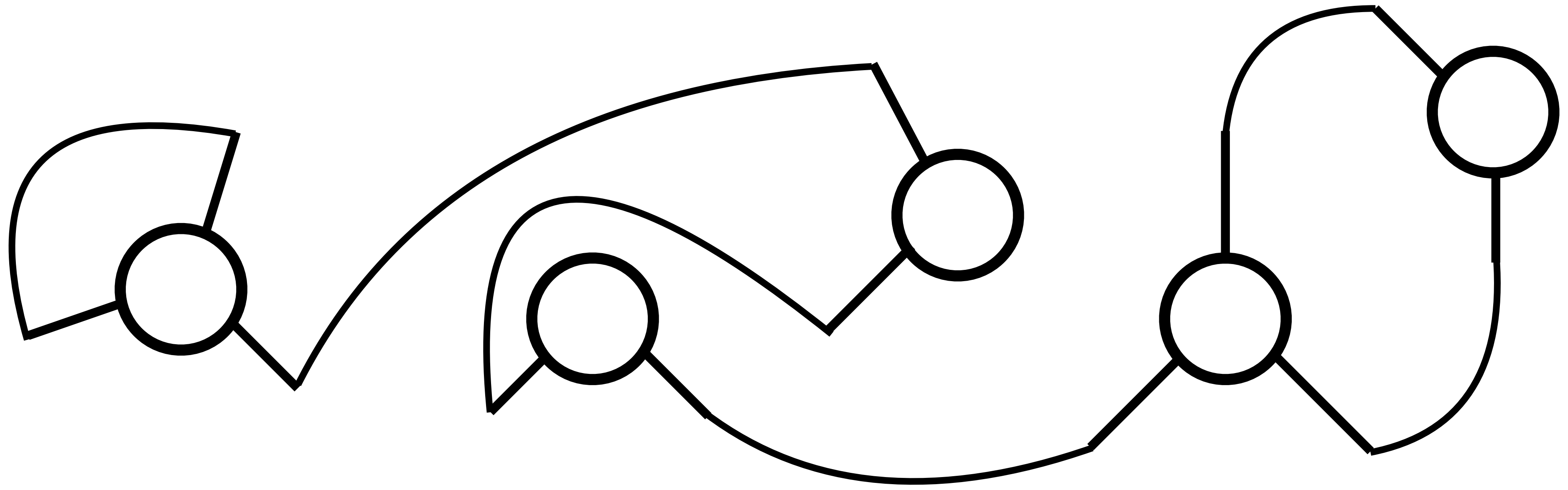
1. Start with a list of degrees $\overrightarrow{k} = \{k_1, k_2, \ldots, k_N\}$
2. Create N “stars" with $k_1, k_2, \ldots, k_N$ stubs each
3. Match stubs at random to form edges
4. Remove **self-** and **multi-**edges or repeat process

Number of self- and multi-edges decreases with network size

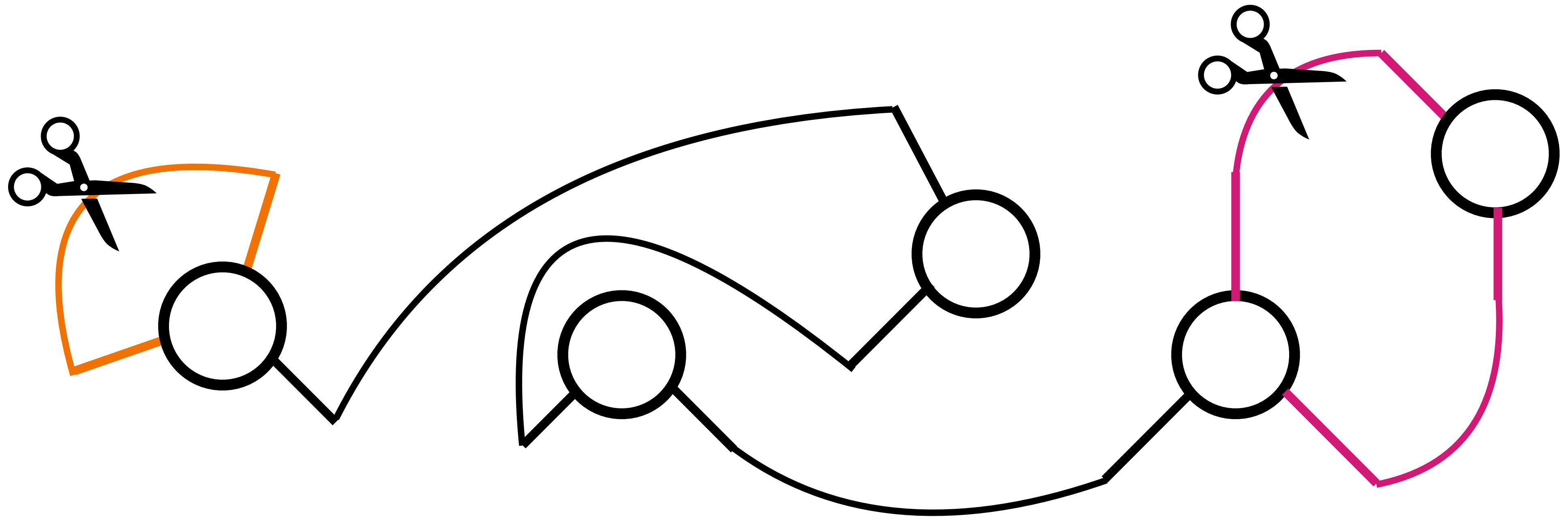
1-2.



3.

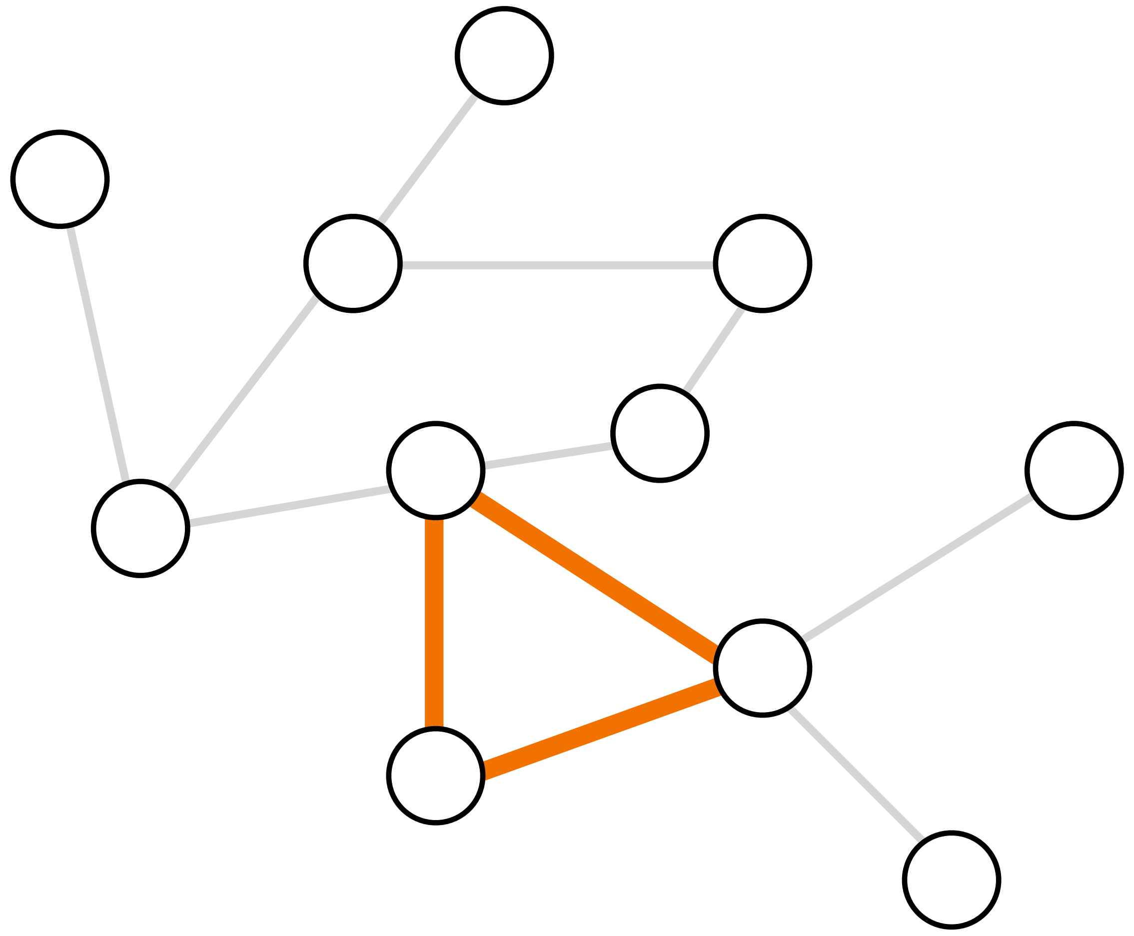


4.



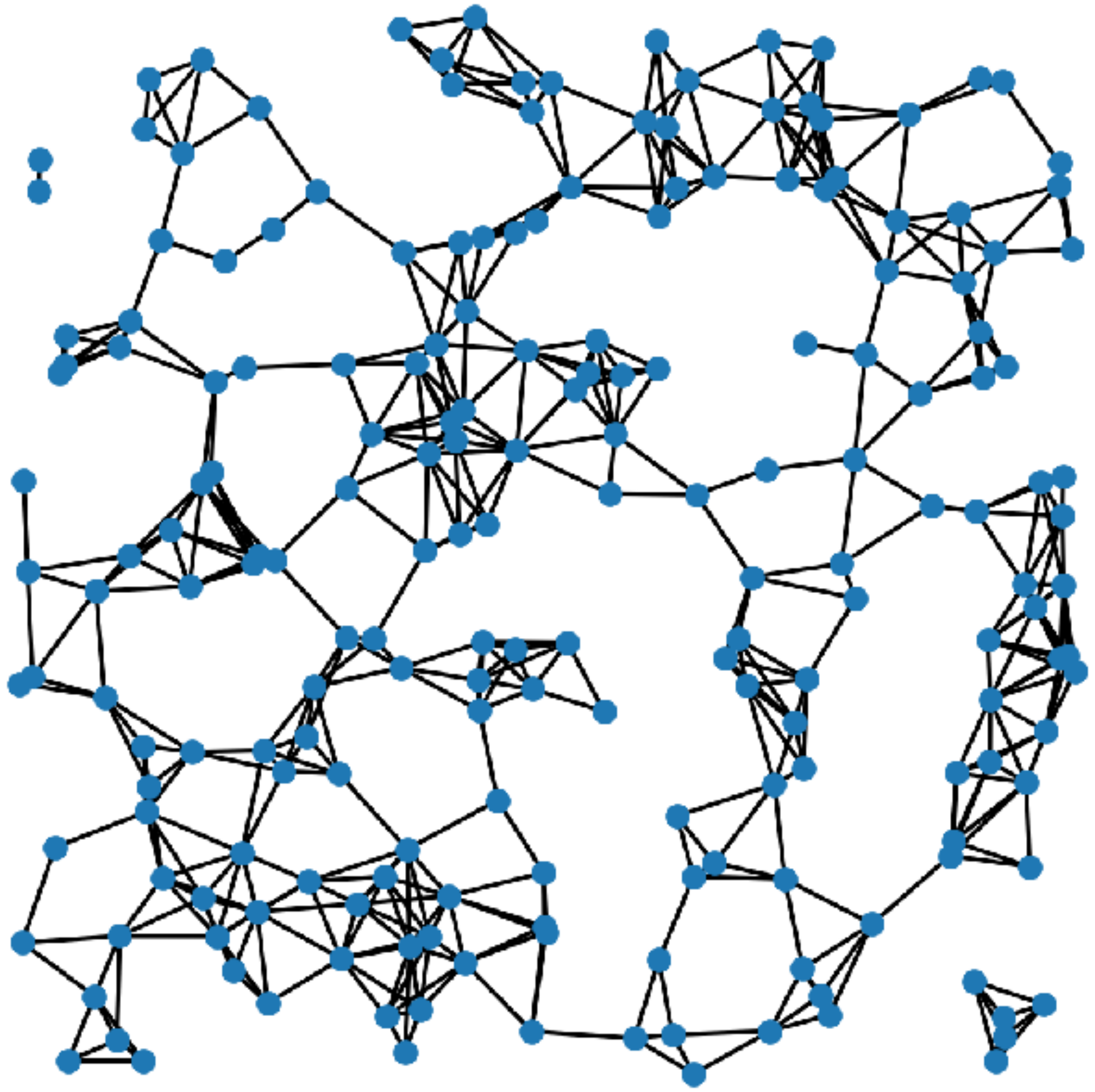
Is the degree distribution all there is to networks?

The CM fails to generate networks with many triangles



CM networks display small clustering

A spatial network



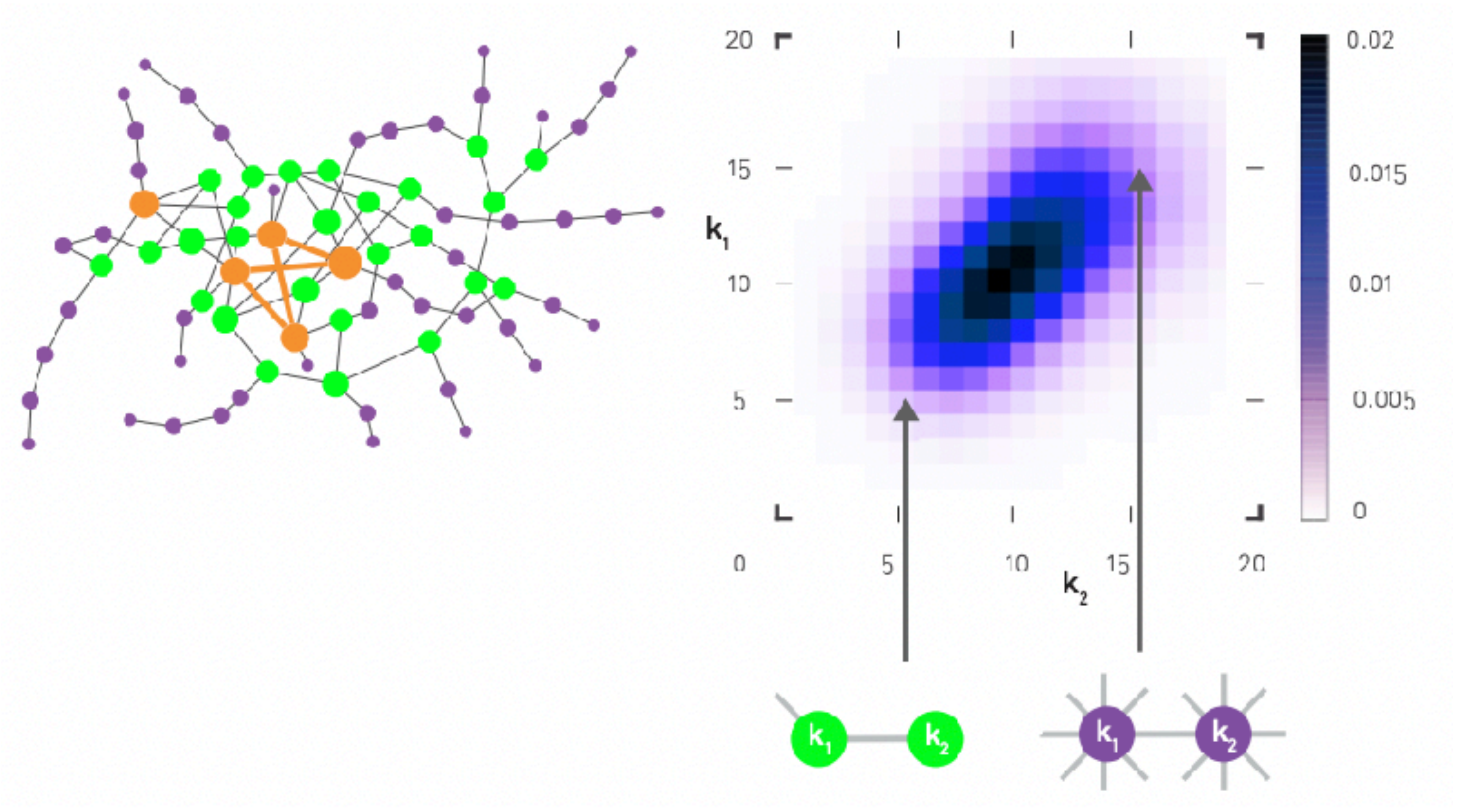
Is the degree distribution all there is to networks?

The CM implies a connection probability:

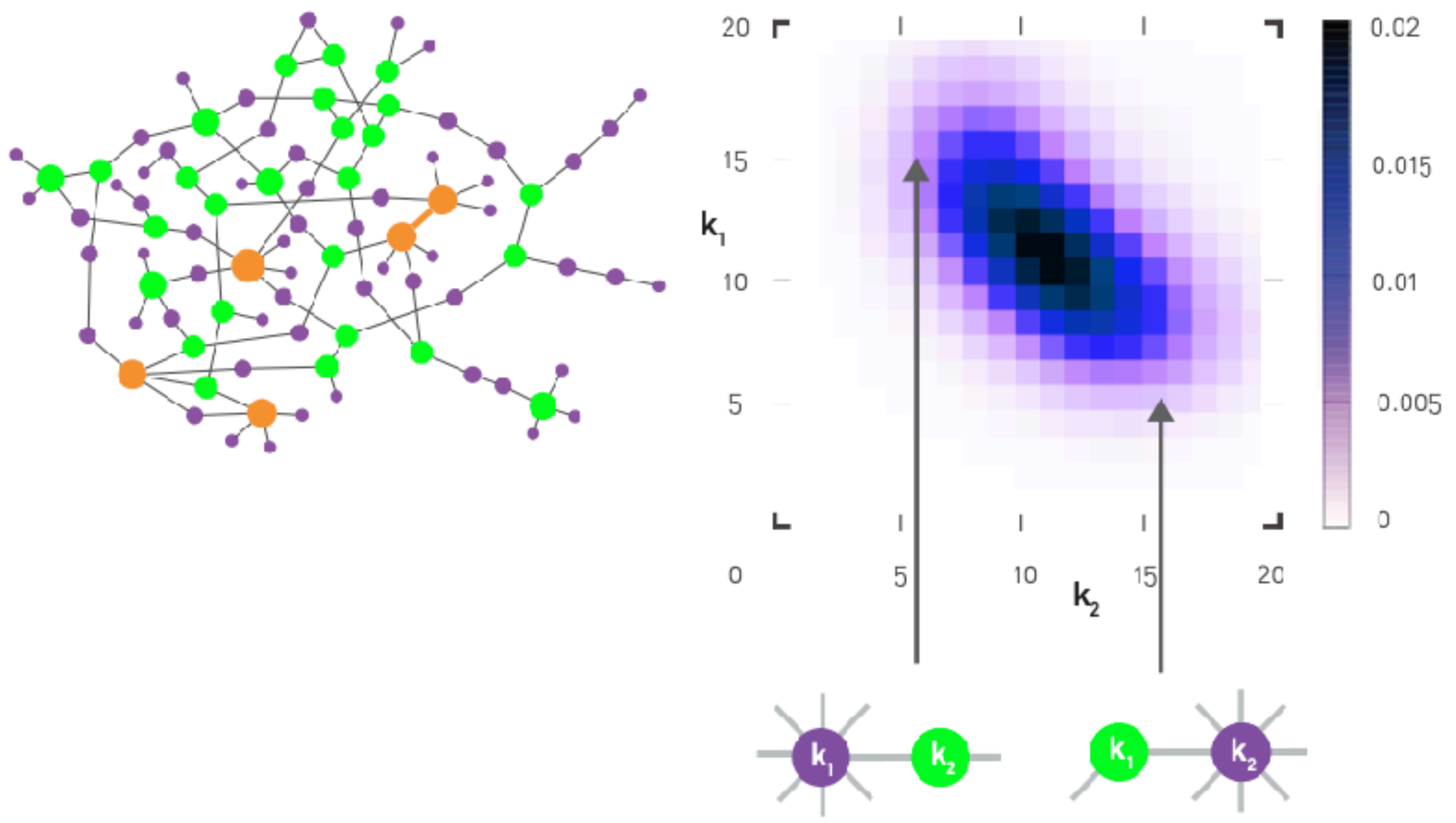
$$p_{i,j} = \frac{k_i k_j}{2|E|}$$

i.e. nodes choose their connections at random

Assortative networks: nodes with similar degree tend to be neighbours



Disassortative networks: nodes with dissimilar degree tend to be neighbours



Barabási, Network Science 2016

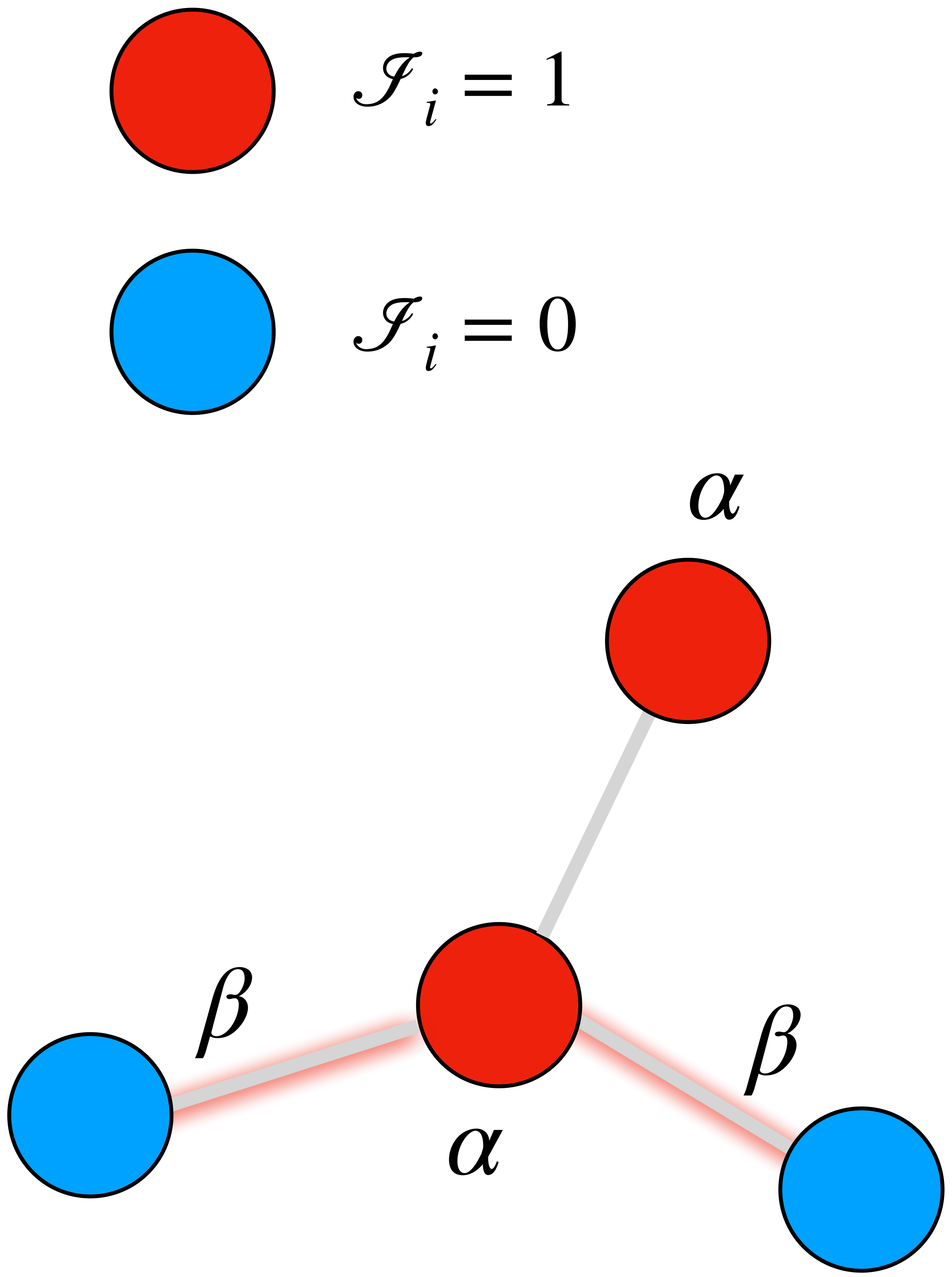
SIS model on networks

Each individual/node is either susceptible ($\mathcal{I}_i = 0$) or infected ($\mathcal{I}_i = 1$).

Infection propagates along edges at rate β while recovery occurs at rate α

We define the probability $\rho_i(t) = \Pr\{\mathcal{I}_i(t) = 1\}$ that node i is infected at time t

$$\frac{d}{dt}\rho_i(t) = -\alpha\rho_i + \beta\sum_{j\neq i} A_{i,j}\cdot\underbrace{\Pr\{\mathcal{I}_i(t) = 0,\mathcal{I}_j(t) = 1\}}_{\text{Two-point correlation}}$$

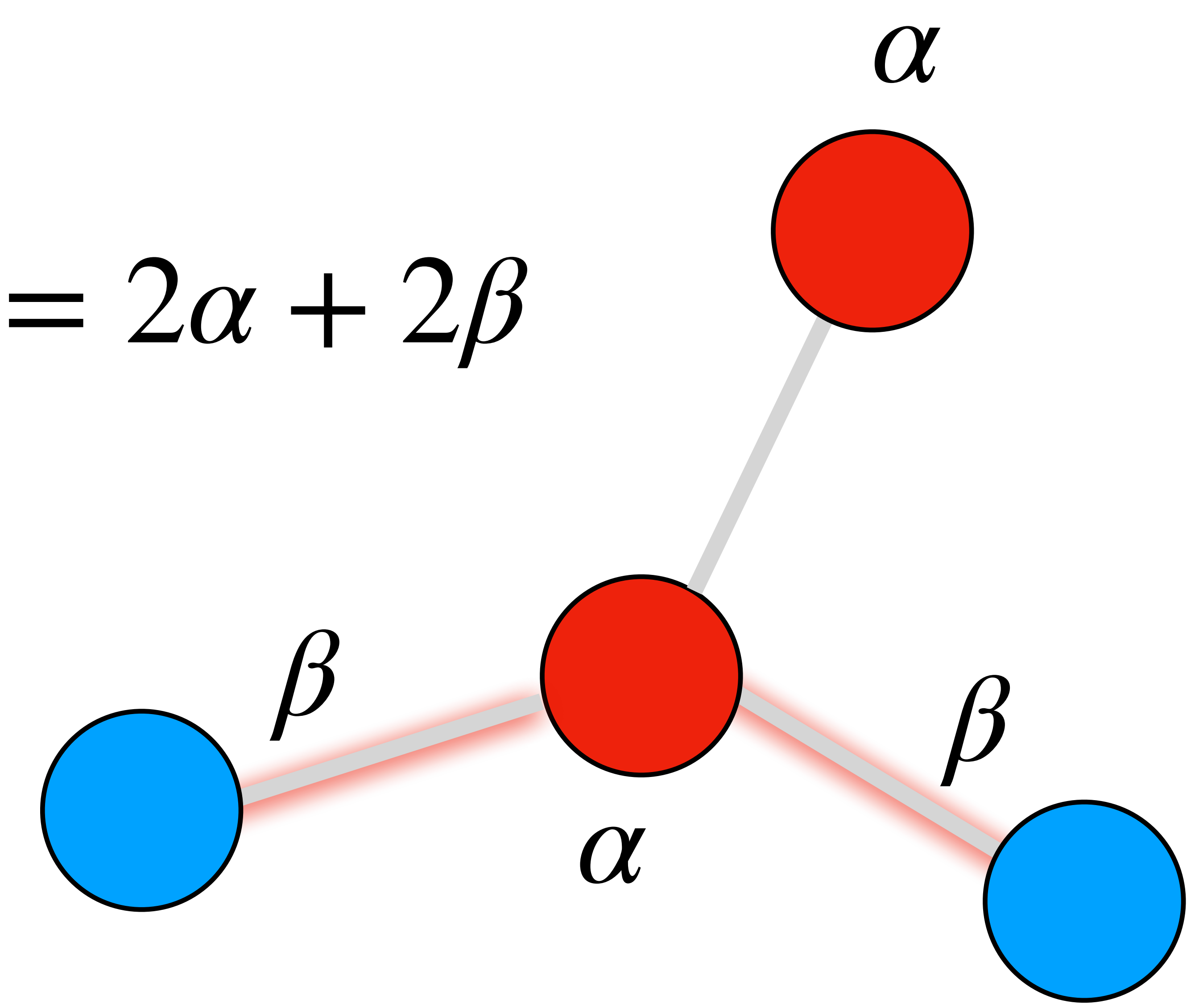


Gillespie-SIS simulations

Key idea: recovery and infection times are exponentially distributed so the time to the next event (recovery or **infection**) is also exponentially distributed with rate:

$$\lambda = \alpha n_I(t) + \beta \sum_{i,j} A_{i,j} \mathcal{I}_i(t)[1 - \mathcal{I}_j(t)]$$

$$\lambda = 2\alpha + 2\beta$$



Then keep repeating the following steps:

- Draw time to next event $\tau_{next} \sim exp(\lambda)$
- Select an event proportionally to its rate
- Update the system according to chosen event

Individual-based mean-field theory

Problem: the hierarchy of n-points correlations

$$\Pr\{\mathcal{I}_i(t) = a\} \longleftarrow \Pr\{\mathcal{I}_i(t) = a, \mathcal{I}_j(t) = b\} \longleftarrow \Pr\{\mathcal{I}_i(t) = a, \mathcal{I}_j(t) = b, \mathcal{I}_k(t) = c\} \longleftarrow$$

Two-point correlation **Three-point correlation**

IBMF: assuming independence between nodes stops the hierarchy

$$\Pr\{\mathcal{I}_i(t) = 0, \mathcal{I}_j(t) = 1\} \approx \Pr\{\mathcal{I}_i(t) = 0\} \cdot \Pr\{\mathcal{I}_j(t) = 1\} = [1 - \rho_i(t)] \cdot \rho_j(t)$$

IBMF equations

$$\boxed{\frac{d}{dt}\rho_i(t) = -\alpha\rho_i + \beta(1-\rho_i)\sum_{j\neq i} A_{i,j}\rho_j}$$

Just N equations!

IBMF: epidemic threshold

Initially ρ_i is very small

$$\frac{d}{dt}\rho_i(t) \approx -\alpha\rho_i + \beta\sum_{j\neq i} A_{i,j}\rho_j = \sum_{j} J_{i,j}\cdot\rho_j$$

Jacobian matrix

$$\boxed{J_{i,j} = -\alpha\delta_{i,j} + \beta A_{i,j}}$$

The eigenvalues Λ_n^J of J decide the fate of initial perturbations:

$$\vec{\rho} \approx \sum_{n} \vec{u}_n \cdot e^{\Lambda_n^J\cdot t}$$

- If all the Λ_n^J are negative, $\vec{\rho}$ shrinks to 0, hence no equilibrium.
- If at least one Λ_n^J (the largest) is positive, the epidemic takes off.

Epidemic threshold

$$\max\{\Lambda_n^J\} = \beta\cdot\max\{\Lambda_n^A\} - \alpha \geq 0$$

$$R_0 = \frac{\beta}{\alpha}\cdot\max\{\Lambda_n^A\} \geq 1$$

Degree-based mean-field approximation

Uses degree-aggregated variables, e.g. **density of infected nodes with degree k**

$$I_k(t) \equiv \frac{n_{I,k}(t)}{N} = \frac{\sum_{i\in k} \rho_i(t)}{N}$$

$\sum\limits_{i\in k}$ **runs over nodes with degree k**

Evolves according to:

$$\frac{d}{dt} I_k(t) = -\alpha I_k + \beta(p_k - I_k) \sum_{k'} \frac{kk'}{\langle k \rangle} I_{k'}$$

Connection probability $\propto kk'$

Hypotheses:

- Neglects neighbour correlations (like IBMF)
- Nodes with the same degree are dynamically equivalent
- Assumes that the pathogen spreads slowly with respect to the dynamics of pathogens

From IBMF to DBMF

DBMF involves an approximation on top of IBMF.

Sum over nodes with some degree and then over all degrees

$$\sum_j = \sum_{k'} \sum_{j \in k'}$$

$$\frac{d}{dt} \sum_{i \in k} \rho_i(t) = -\alpha \sum_{i \in k} \rho_i + \beta \sum_{i \in k} (1 - \rho_i) \sum_{j \neq i} A_{i,j} \rho_j$$

Annealed approximation:
Nodes 'see' only the average connectivity

$$A_{i,j} \rightarrow p_{i,j} = \frac{k_i k_j}{N \langle k \rangle}$$

N.B. no degree correlations

$$\frac{d}{dt} n_{I,k} = -\alpha n_{I,k} + \beta \sum_{k'} \sum_{i \in k} \sum_{j \in k'} (1 - \rho_i) \cdot A_{i,j} \cdot \rho_j$$

$$\frac{d}{dt} n_{I,k} = -\alpha n_{I,k} + \beta \sum_{k'} \sum_{i \in k} (1 - \rho_i) \sum_{j \in k'} \rho_j \cdot \frac{k_i k_j}{N \langle k \rangle}$$

$$N \langle k \rangle = 2 |E|$$

From IBMF to DBMF

DBMF involves an approximation on top of IBMF.

$$\frac{d}{dt}n_{I,k} = -\alpha n_{I,k} + \beta \sum_{k'} \sum_{i \in k} (1-\rho_i) \sum_{j \in k'} \rho_j \cdot \frac{k_i k_j}{N\langle k \rangle}$$

But $k_i = k$ and $k_j = k'$ because $i \in k$ and $j \in k'$

$$\frac{d}{dt}n_{I,k} = -\alpha n_{I,k} + \beta k \sum_{i \in k} (1-\rho_i) \sum_{k'} \frac{k' \sum_{j \in k'} \rho_j}{N\langle k \rangle}$$

The sums $i \in k$ and $j \in k'$ can now be done explicitly

$$\frac{d}{dt}n_{I,k} = -\alpha n_{I,k} + \beta k (n_k - n_{I,k}) \sum_{k'} \frac{k' I_{k'}}{\langle k \rangle}$$

Divide by N to get the DBMF equation

DBMF: epidemic threshold

Let's find the stationary solution I_k^*

$$-I_k^* + \lambda\Theta^* \cdot k(p_k - I_k^*) = 0$$

$$\Theta(t) = \frac{1}{\langle k \rangle}\sum_k kI_k(t)$$

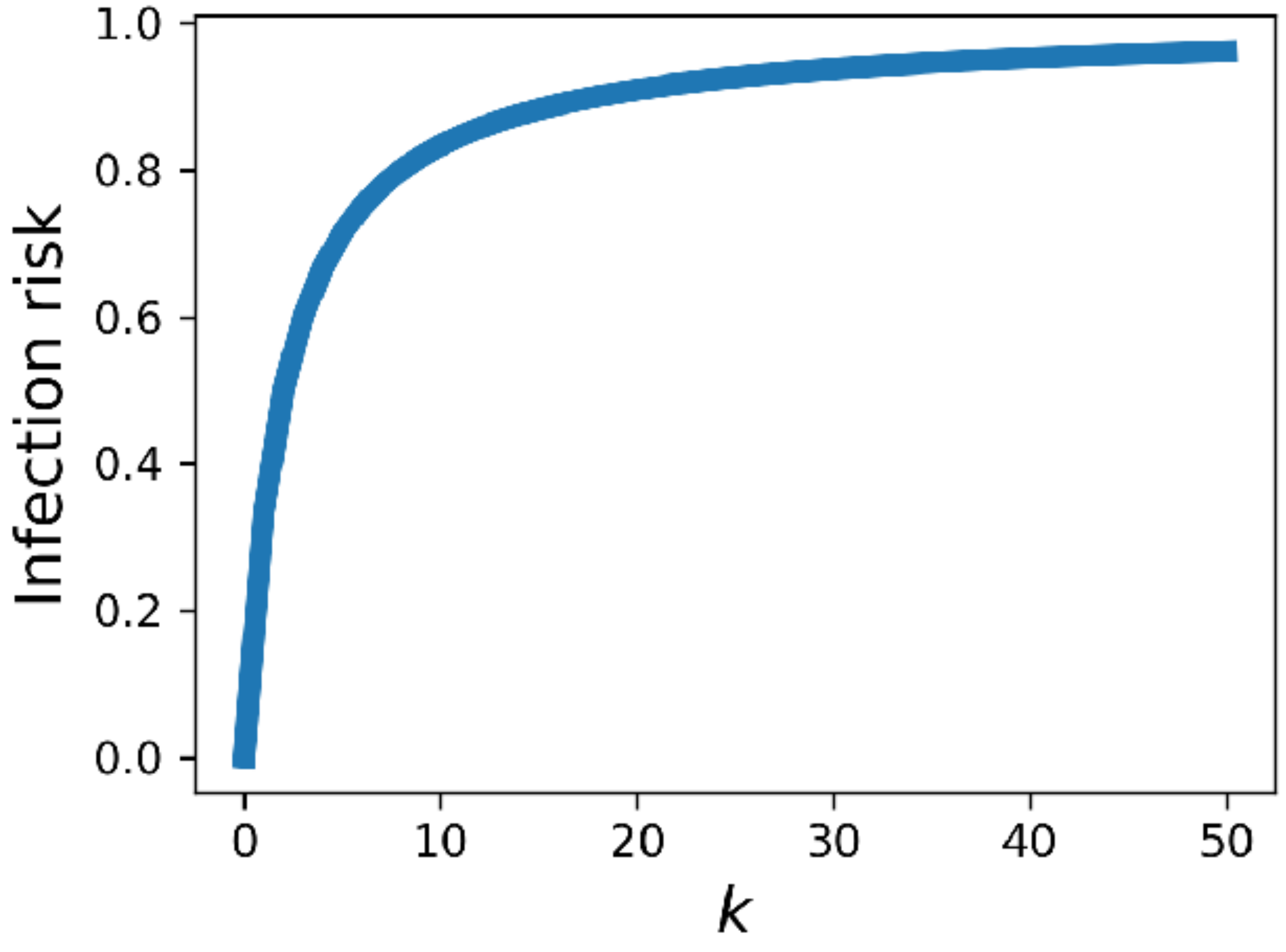
$$\lambda = \frac{\beta}{\alpha}$$

In terms of Θ^* (still unknown):

$$I_k^* = \frac{\lambda\Theta^* k}{1 + \lambda\Theta^* k} \cdot p_k$$

➡

$$\frac{I_k^*}{p_k} = \frac{\lambda\Theta^* k}{1 + \lambda\Theta^* k}$$



Probability of infection increases with degree k

DBMF: epidemic threshold

Θ^* is found using self-consistency. We are interested in the endemic solution $\Theta * > 0$.

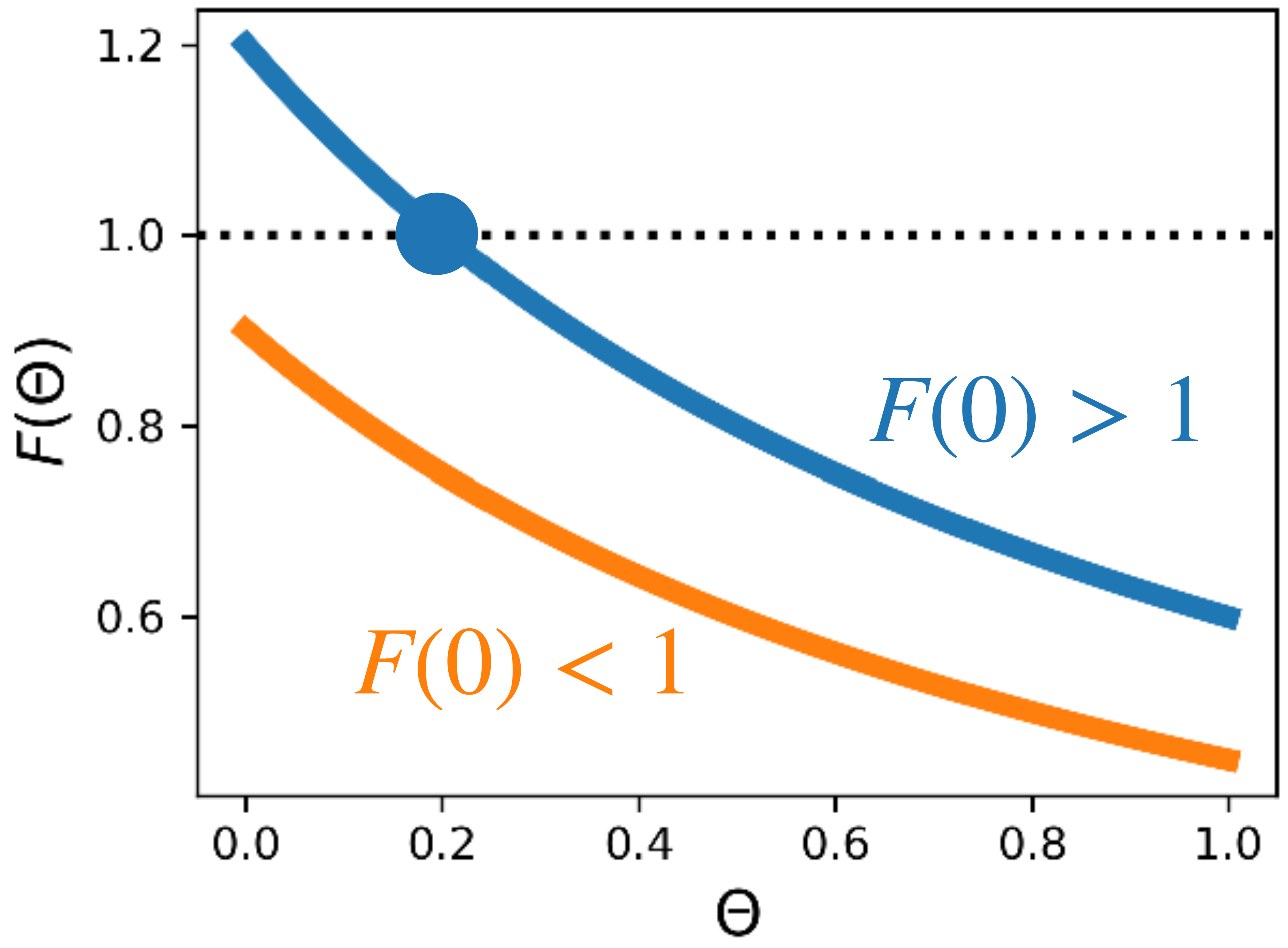
$$\Theta^* = \frac{1}{\langle k\rangle}\sum_k k I_k^* = \frac{1}{\langle k\rangle}\sum_k \frac{\lambda\Theta^* k^2}{1+\lambda\Theta^* k} p_k$$

$$\underbrace{\boxed{\frac{1}{\langle k\rangle}\sum_k \frac{\lambda k^2\cdot p_k}{1+\lambda\Theta^* k}}}_{F(\Theta^*)} = 1$$

Has a solution only if $F(0) \geq 1$

$$F(0) = \frac{\lambda}{\langle k\rangle}\sum_k k^2 p_k = \boxed{\frac{\beta}{\alpha}\frac{\langle k^2\rangle}{\langle k\rangle} \geq 1}$$

We just found $R_0 = \frac{\beta}{\alpha}\frac{\langle k^2\rangle}{\langle k\rangle}$



Impact of contact heterogeneities

1. Contact heterogeneities facilitate transmission with respect to homogeneous mixing

$$\frac{\beta}{\alpha}\frac{\langle k^2\rangle}{\langle k\rangle} \geq \boxed{\frac{\beta}{\alpha}\langle k\rangle}$$

Homogeneous mixing R_0

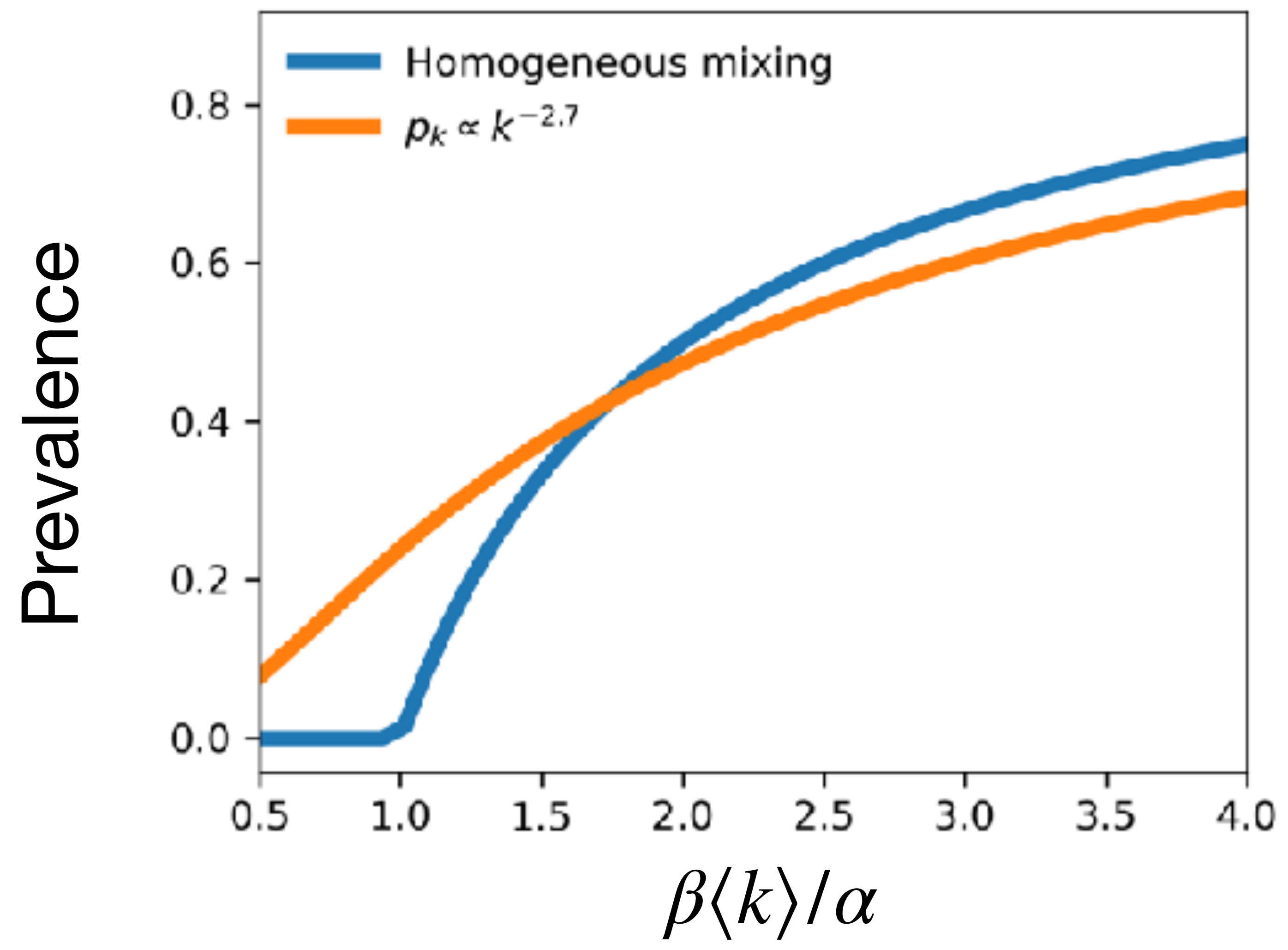
2. Epidemic threshold can vanish in very heterogeneous networks

Assume $p_k = Ck^{-\gamma}$ → $\langle k^2\rangle = \sum_k k^2 p_k = C\sum_k k^{-\gamma+2}$

R_0 diverges with $\langle k^2\rangle$ when $\gamma < 3$

3. Smaller prevalence in networks

Low degree nodes do not contribute much to transmission: infection transmits in a part of the Network



Comparison with IBMF

IBMF

$$R_0^{IBMF} = \frac{\beta}{\alpha} \cdot \Lambda_m^A$$

DBMF

$$R_0^{DBMF} = \frac{\beta}{\alpha} \cdot \frac{\langle k^2 \rangle}{\langle k \rangle}$$

Remember: DBMF makes a further approximation wrt IBMF

General result for random graphs

$$\Lambda_m^A = \max\{\sqrt{k_{max}}, \frac{\langle k^2 \rangle}{\langle k \rangle}\}$$

k_{max} : typical maximum degree

Assume $p_k = Ck^{-\gamma}$ in a finite network of size N

$$\Lambda_m^A = \begin{cases} \sqrt{k_{max}} & \text{if } \gamma > 2.5 \\ \frac{\langle k^2 \rangle}{\langle k \rangle} & \text{if } \gamma < 2.5 \end{cases}$$

Chung et al, PNAS, 2003
Boguña et al, Eur Phys J B, 2004
Castellano et al, PRL, 2010

Vaccination in heterogeneous networks

Which nodes should we vax?

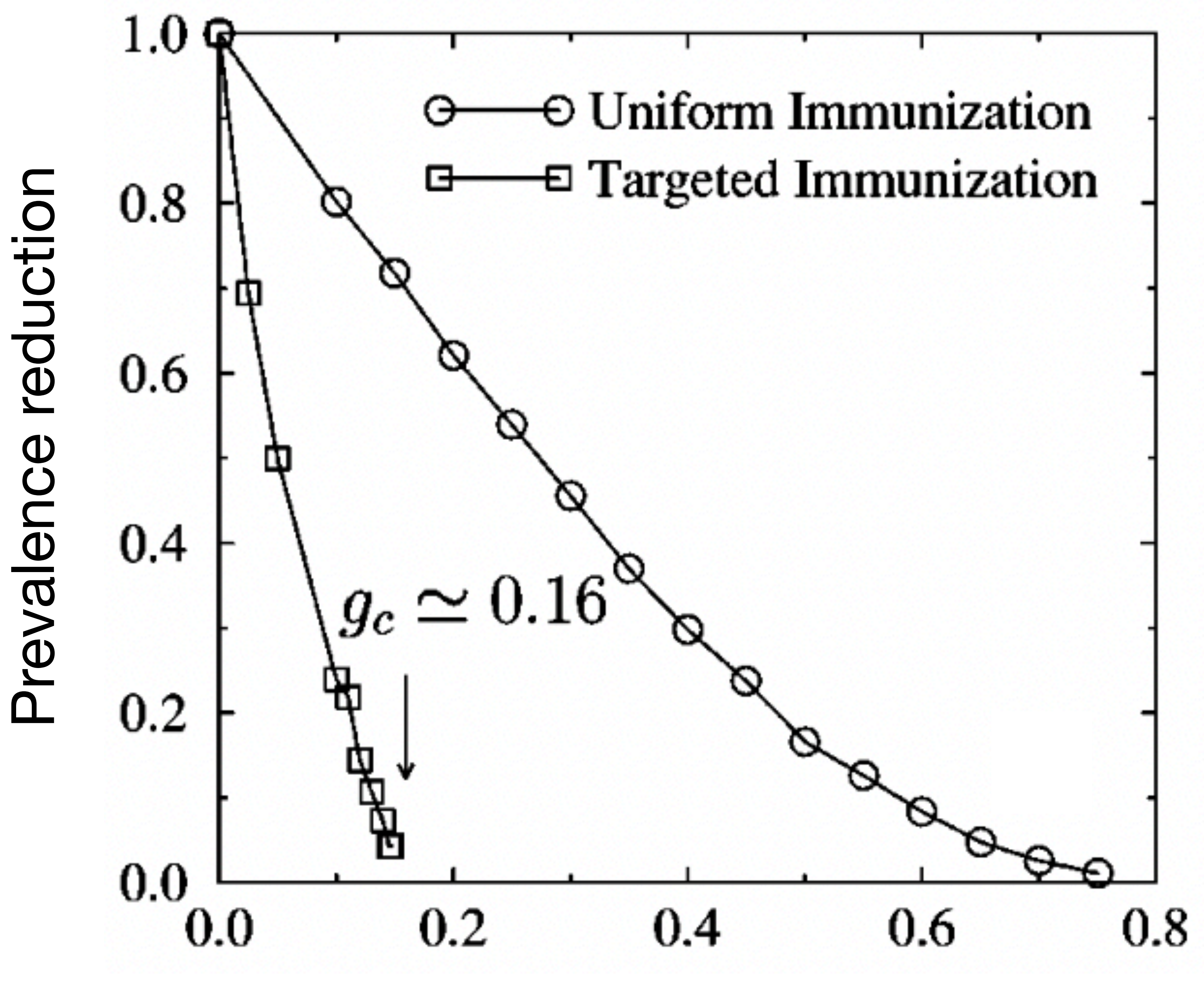
- **Uniform vaccination**: select nodes at random
- **Targeted vaccination**: start from most connected nodes

The **targeted** strategy is better because it removes nodes that contribute the most to transmission (and their contacts)

$$R_0^{DBMF} = \frac{\beta}{\alpha} \cdot \frac{\langle k^2 \rangle}{\langle k \rangle}$$

Suggests increasing returns from vaxing high-degree nodes

SIS model on network with $p_k \propto k^{-3}$



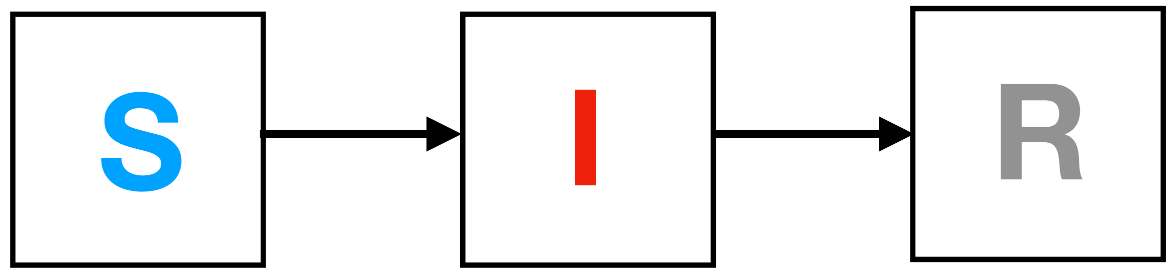
Proportion vaxxed

Pastor-Satorras & Vespignani, PRE, 2002

DBMF analysis of SIR dynamics

Nodes with larger degrees are depleted faster

$$\frac{d}{dt}S_k(t) = -\beta k S_k \Theta \quad \Rightarrow \quad S_k(t) \approx e^{-\beta k \int_0^t \Theta(u) du}$$



Same epidemic threshold of SIS model

$$\frac{d}{dt}I_k(t) = -\alpha I_k + \beta k S_k \Theta \quad \Rightarrow \quad R_0 = \frac{\beta}{\alpha}\frac{\langle k^2 \rangle}{\langle k \rangle}$$

Event-driven SIR simulations

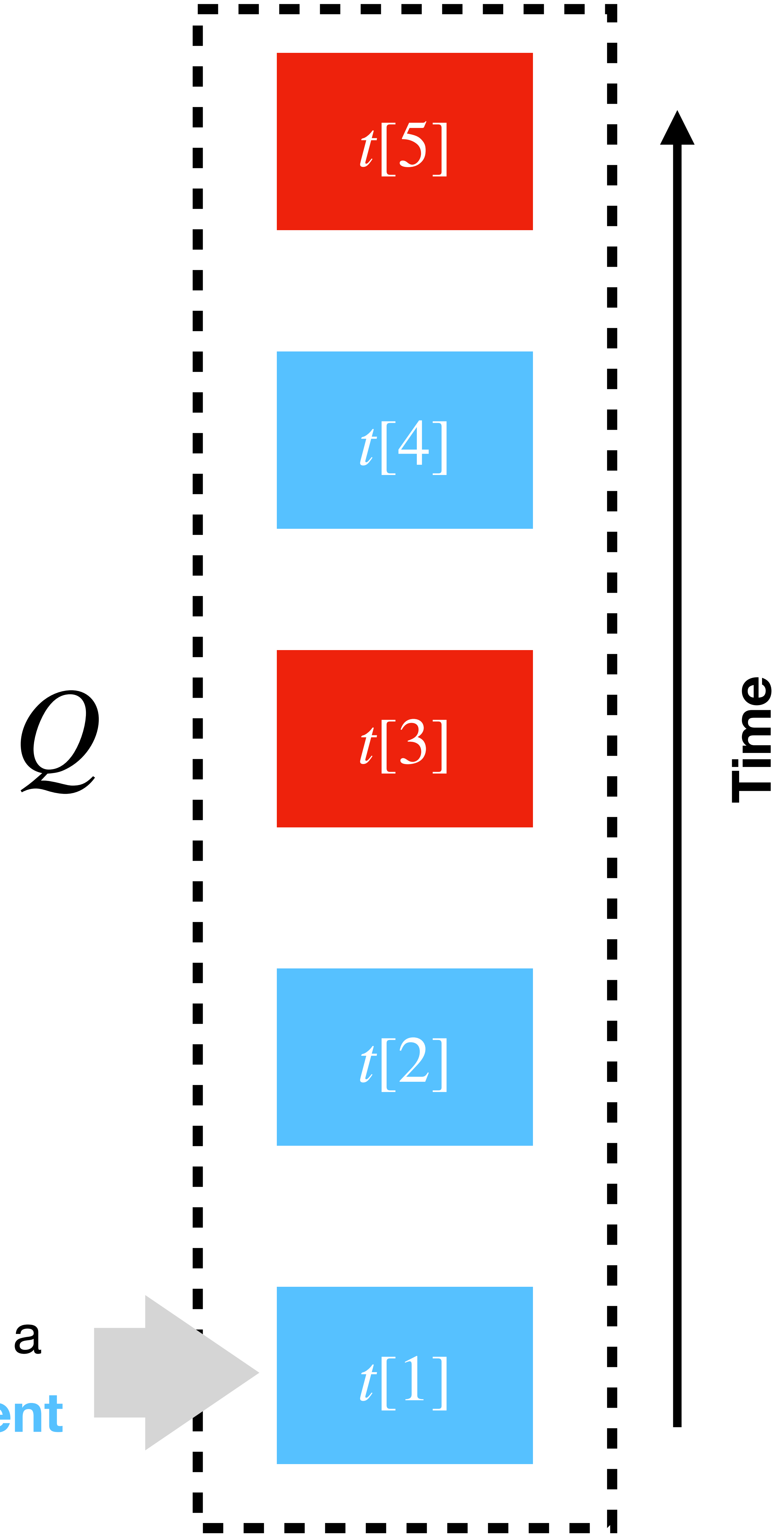
Keep a time-ordered list Q of **recovery** and **infection** events.

Apply events sequentially. Note that an event may spawn new events and thus modify Q itself.

Assume node i is infected at time t:

- Draw an infectious period $\tau \sim exp(\alpha)$ and push a **recovery event** into Q
- For each susceptible neighbour j draw an infection time $\tau_j \sim exp(\beta)$, but push an **infection event** only if $\tau_j < \tau$ (i.e. before it recovers)

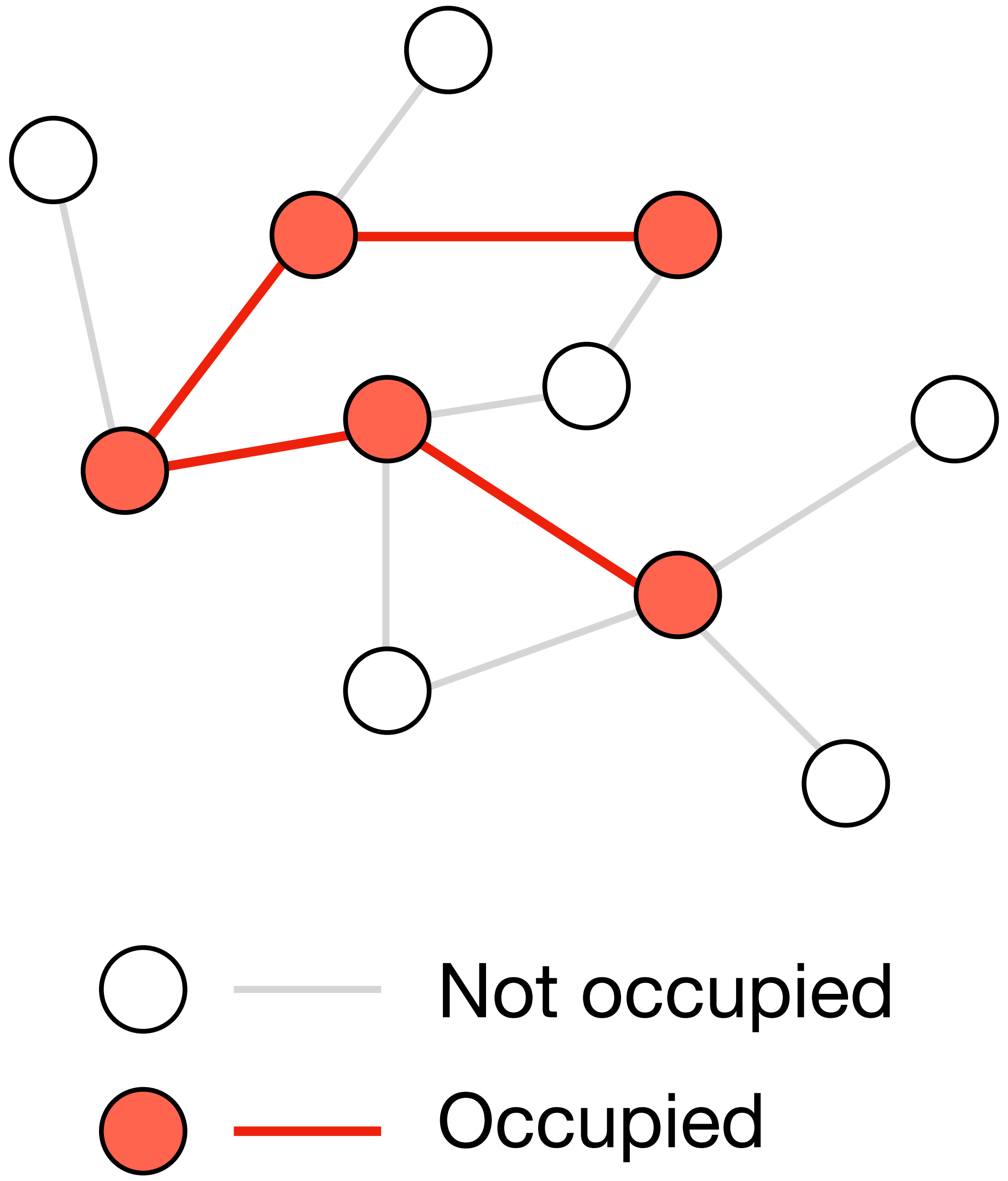
Discard infection events hitting infected or recovered nodes



Next event is a **recovery event**

Percolation analysis of SIR dynamics

Key idea: retrospectively, an **outbreak** looks like the result of **percolation**



Infection travels along one edge with probability $\boxed{T = \frac{\beta}{\alpha + \beta}}$ **Bond occupation probability**

A large outbreak corresponds to an **infinitely large** percolating cluster

When percolation occurs mean cluster size $\langle s \rangle$ diverges

Desired result (don't know T_c yet)

- If $T < T_c$ then $\langle s \rangle$ is finite (**no percolation**/**small outbreaks only**)
- If $T > T_c$ then $\langle s \rangle$ is infinite (**percolation**/**large outbreak**)

Newman, PRE, 2002

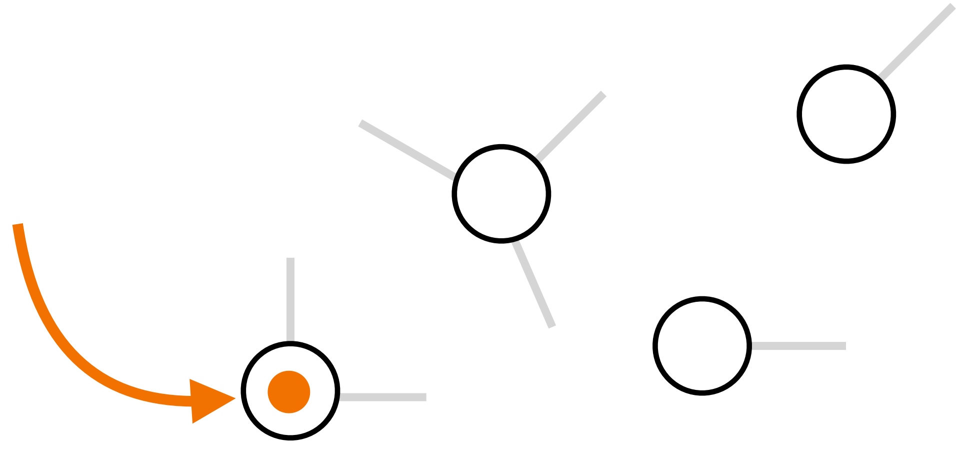
Interlude: choosing a random node

What is the probability that a randomly chosen node has degree k? Depends on how the node is selected!

Strategy A:

Pick a random node label

Answer: p_k



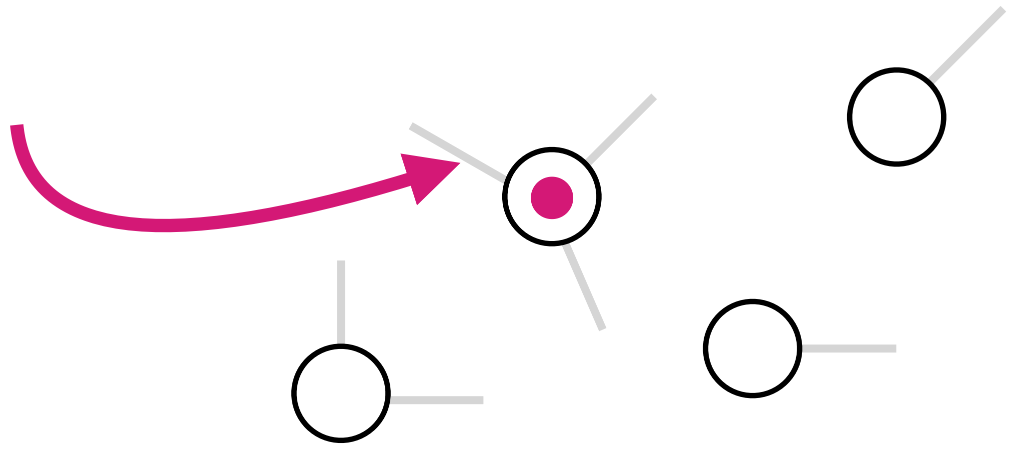
$$\langle k \rangle_A = \langle k \rangle$$

Strategy B:

First pick a random edge and then one node at its ends (or pick a stub)

Answer: $q_k = \dfrac{kp_k}{\langle k \rangle}$

A $k = 10$ node is 10x more likely to be chosen than a $k = 1$ node because 10 edges lead to it.



Friendship paradox

$$\langle k \rangle_B = \frac{\langle k^2 \rangle}{\langle k \rangle}$$

Newman, PRE, 2002

Interlude: probability generating functions

Let p_n denote a discrete probability distribution. We define the PGF $G(x) = \sum_n p_n x^n$

Properties

1. $G(x)$ encodes p_n
2. $G(x)$ allows to calculate the moments of p_n
3. Convolution: if $Z = X + Y$, then $G_Z(u) = G_X(u)G_Y(u)$
4. Composition: if $Z = X_1 + X_2 + \ldots + X_K$ where K itself is random, then $G_Z(u) = G_K(G_X(u))$

Moments

$$\boxed{G(x=1) = \sum_n p_n = 1 \qquad \frac{d}{dx}G(x=1) = \sum_n np_n = \langle n \rangle}$$

Newman, PRE, 2002

Examples

1. **Bernoulli** *"One egg hatches with probability T"*

Let $X = \begin{cases} 0 \text{ with prob. } 1 - T \\ 1 \text{ with prob. } T \end{cases}$

$$G_X(u) = (1 - T) \cdot u^0 + T \cdot u^1 = 1 - T + Tu$$

(PGF definition)

2. **Binomial** *"One hen lays a fixed number M of eggs, each hatching with probability T"*

$Y \sim Binomial(M, T)$. Because $Y = X_1 + \ldots + X_M$:

$$G_Y(u) = G_X^M(u) = (1 - T + Tu)^M$$

(Property 3: addition)

3. **Binomial with random number of draws** *"One hen lays a random number M of eggs, each hatching with probability T"*

$$G_Y(u) = G_M(G_X(u)) = G_M(1 - T + Tu)$$

(Property 4: composition)

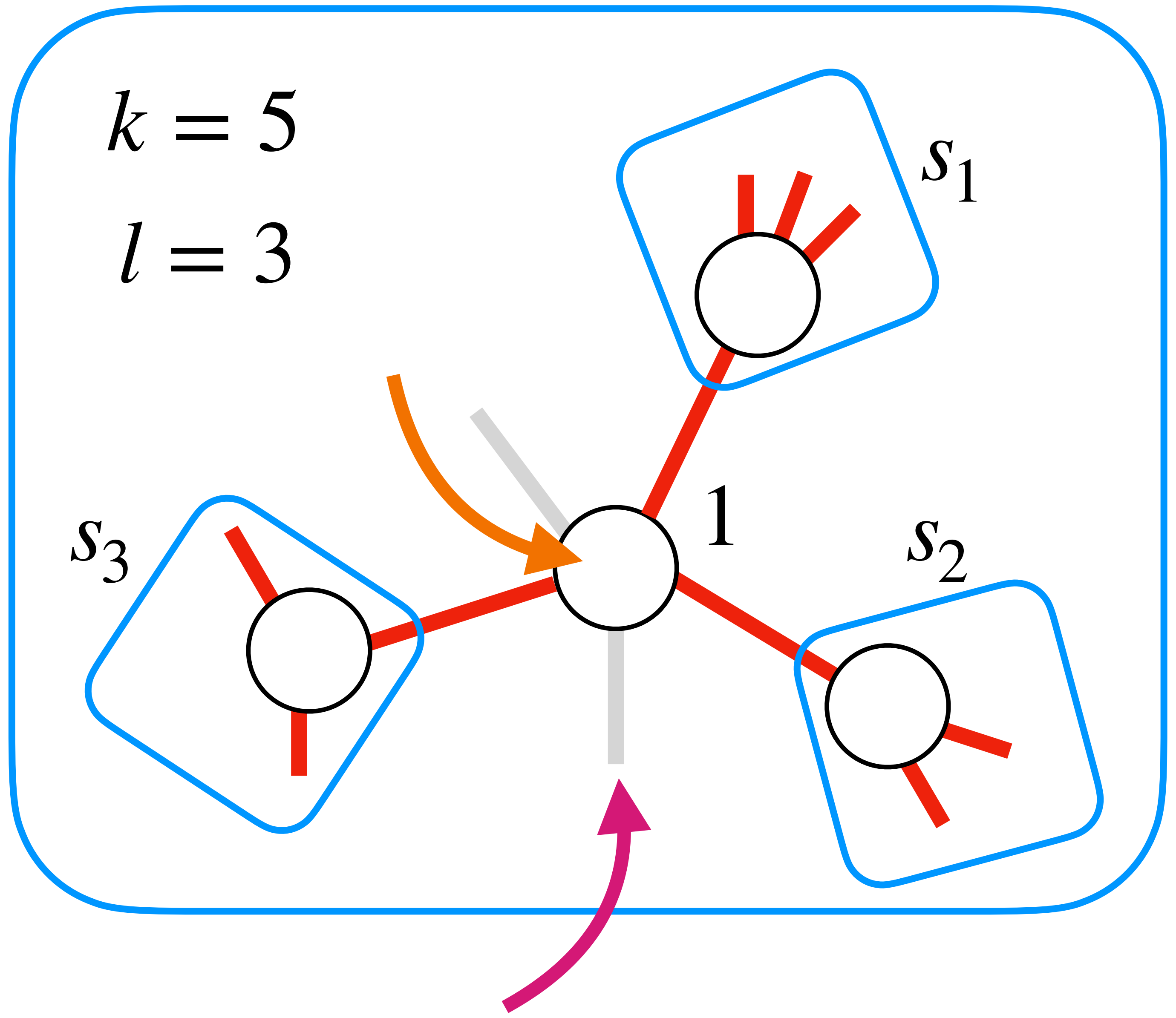
Newman, PRE, 2002

Back to SIR and percolation

Pick a random node. How large is the cluster containing it?

Assume node has degree k , but only $l \leq k$ edges are occupied

$$s = 1 + s_1 + s_2 + s_3$$



Cluster size can be found recursively (exact on **trees**)

$$s = 1 + \sum_{i=1}^{l} s_i$$

Newman, PRE, 2002

Back to SIR and percolation

Pick a random node. How large is the cluster containing it?

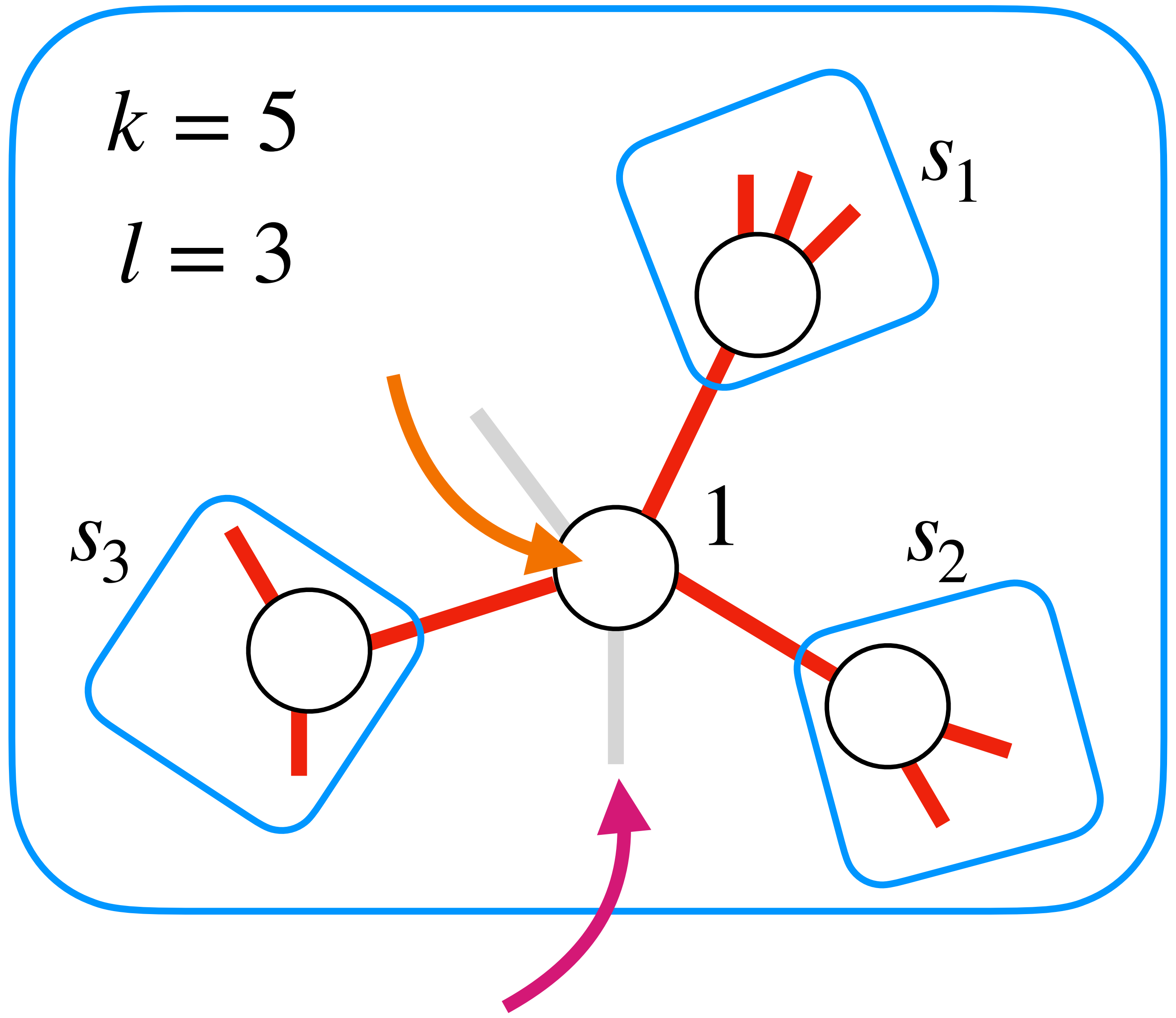
Assume node has degree k , but only $l \leq k$ edges are occupied

Define:

- The PGF of cluster sizes when the first node is picked with **strategy A**: $H_A(u)$
- The PGF of cluster sizes when the first node is picked with **strategy B**: $H_B(u)$
- The PGF of the degree distribution $G_A(u)$ (needed with **strategy A**)
- The PGF of the excess degree distribution $G_B(u) = \sum_k \frac{(k+1)p_{k+1}}{\langle k \rangle} u^k$ (needed with **strategy B**)

Cluster size can be found recursively (exact on **trees**) $s = 1 + \sum_{i=1}^{l} s_i$

$$s = 1 + s_1 + s_2 + s_3$$



$$\boxed{\langle s \rangle = H_A'(1)}$$

Newman, PRE, 2002

Cluster size PGF

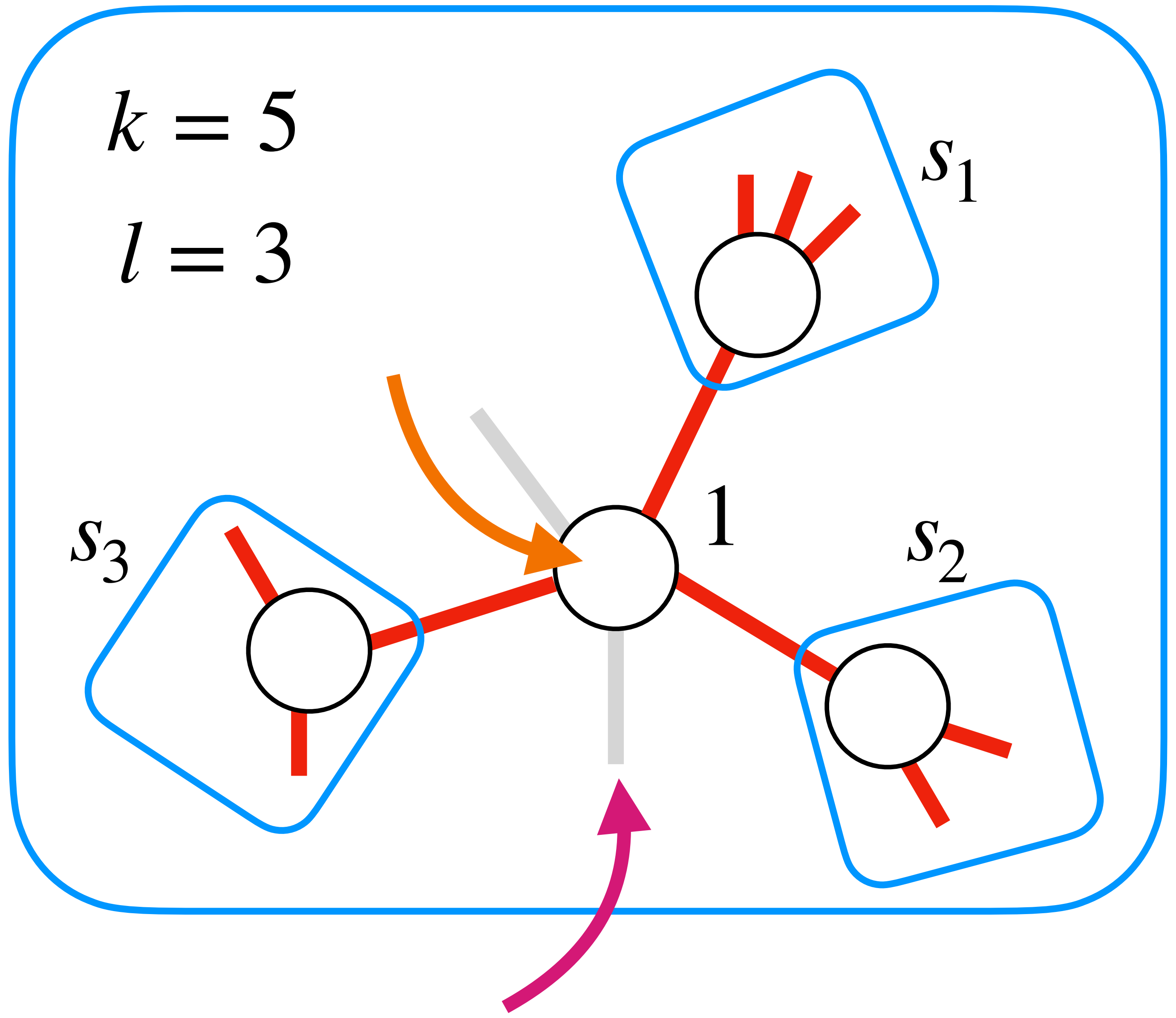
Pick a random node. How large is the cluster containing it?

Assume node has degree k , but only $l \leq k$ edges are occupied

Cluster size can be found recursively (exact on **trees**)

$$s = 1 + \sum_{i=1}^{l} s_i$$

$$s = 1 + s_1 + s_2 + s_3$$



Using **strategy A** to pick the initial node:

$$H_A(u) = u \cdot \overbrace{G_A}^{\text{Prop. 3}}\overbrace{[1 - T + TH_B(u)]}^{\text{Prop.4}}$$

Initial node contributes 1 to cluster size, $G(u) = u$

Initial degree comes from p_k due to **strategy A**

Random occupation contributes $1 - T + T(\cdot)$

Sub-clusters are entered by following an edge (**strategy B**)

Newman, PRE, 2002

Cluster size PGF

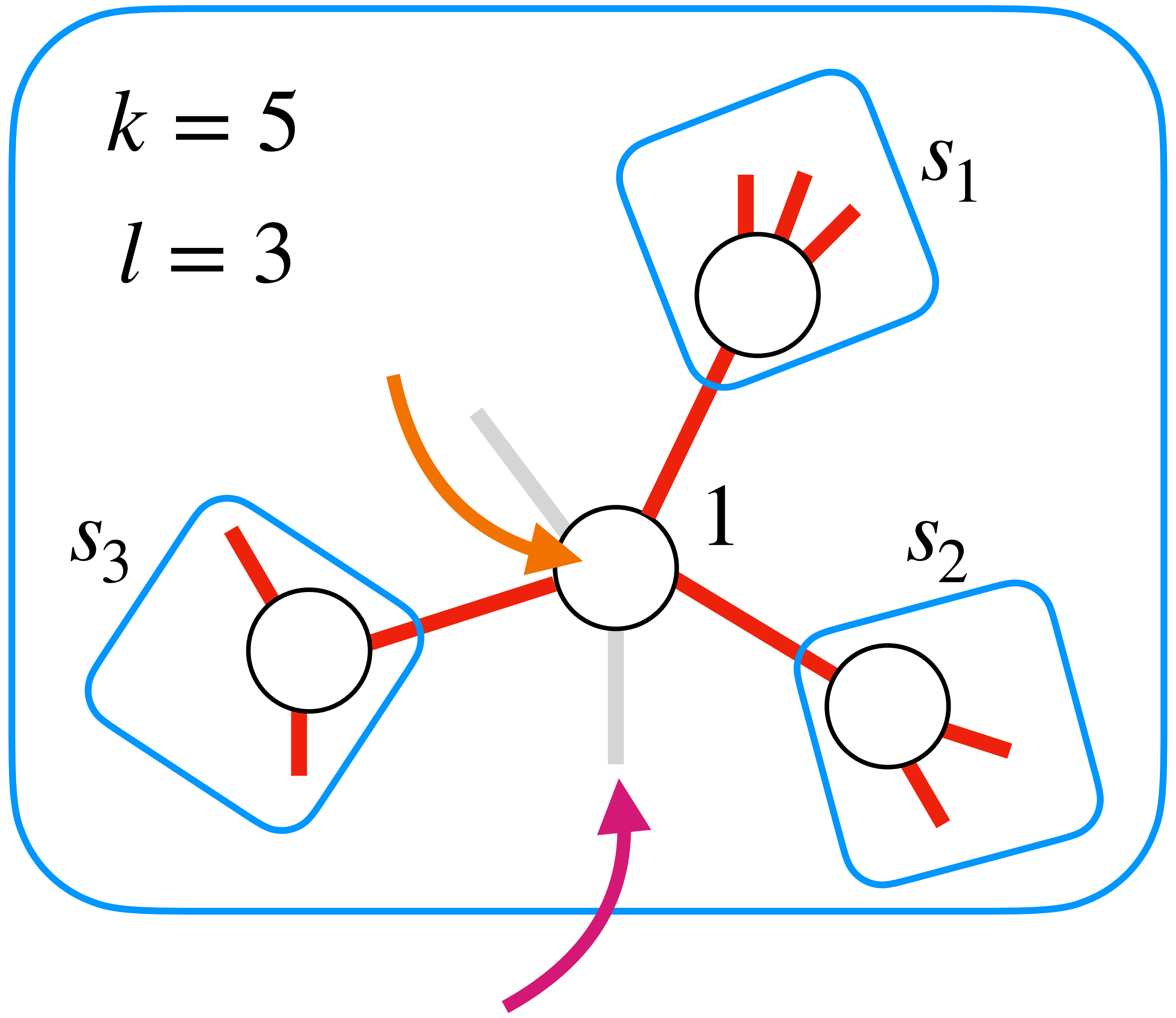
Pick a random node. How large is the cluster containing it?

Assume node has degree k , but only $l \leq k$ edges are occupied

Cluster size can be found recursively (exact on **trees**)

$$s = 1 + \sum_{i=1}^{l} s_i$$

$$s = 1 + s_1 + s_2 + s_3$$



Using **strategy B** to pick the initial node:

$$H_B(u) = u \cdot \overset{\text{Prop. 3}}{\overline{G_B}}\overset{\text{Prop.4}}{\overline{[1 - T + TH_B(u)]}}$$

Initial node contributes 1
to cluster size, $G(u) = u$

Initial degree comes from
$kp_k/\langle k \rangle$ due to **strategy B**

Random occupation
contributes
$1 - T + T(\cdot)$

Sub-clusters are entered
by following an edge
(**strategy B**)

Newman, PRE, 2002

Mean cluster size

Finally, use **property 2** to obtain mean cluster size from PGFs

$$H_A(u) = u \cdot G_A[1 - T + TH_B(1 - T + Tu)] \longrightarrow \langle s \rangle = H_A'(1) = 1 + TG_A'(1)H_B'(1)$$

$$H_B(u) = u \cdot G_B[1 - T + TH_B(1 - T + Tu)]$$

$$H_B'(1) = \frac{1}{1 - TG_B'(1)}$$

Mean cluster size:

$$\langle s \rangle = 1 + \frac{TG_A'(1)}{1 - TG_B'(1)}$$

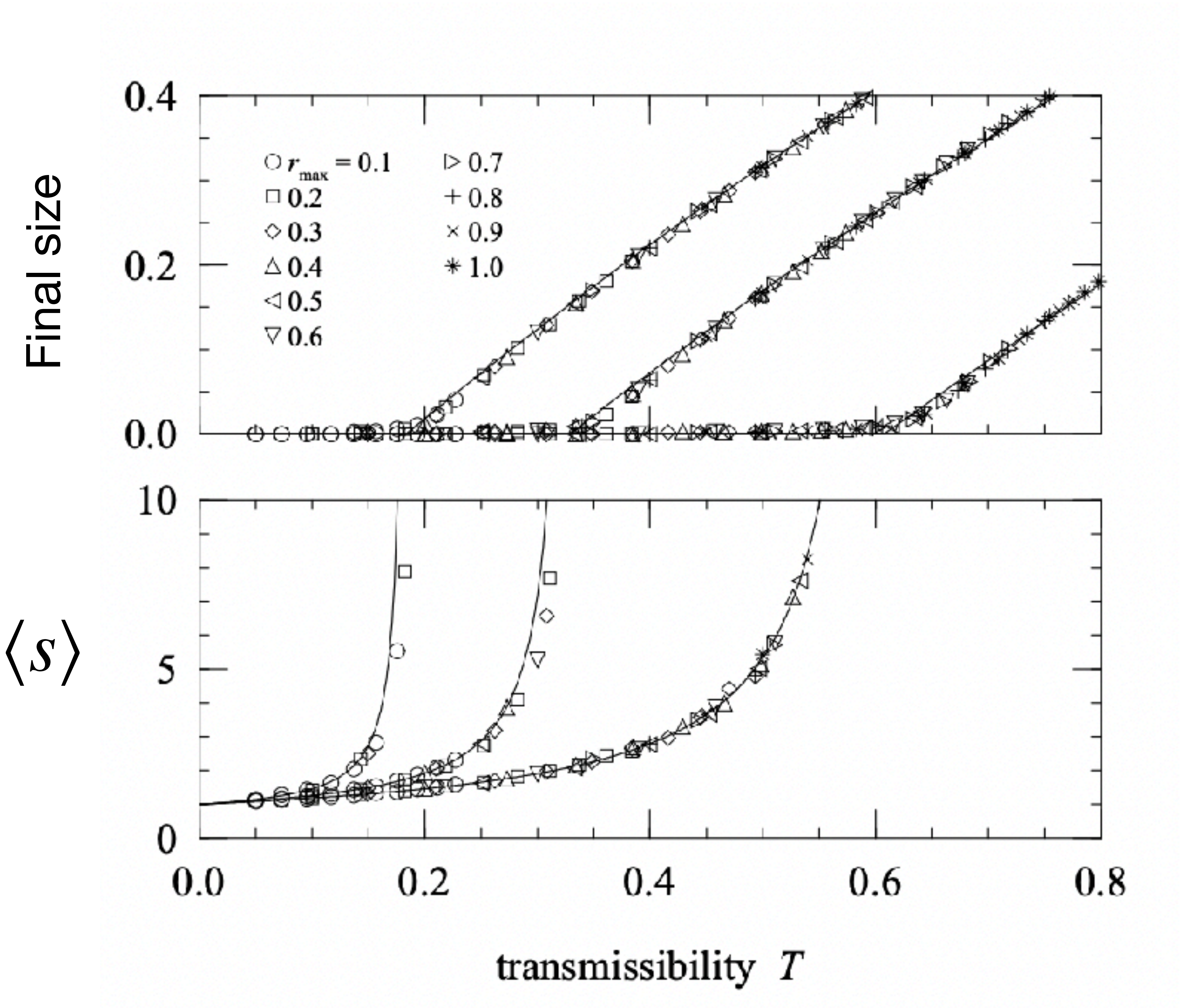
Newman, PRE, 2002

Percolation threshold

Using $T = \frac{\beta}{\beta + \alpha}$ a large outbreak occurs when:

$$\frac{\beta}{\alpha} \geq \frac{\langle k \rangle}{\langle k^2 \rangle - 2\langle k \rangle}$$

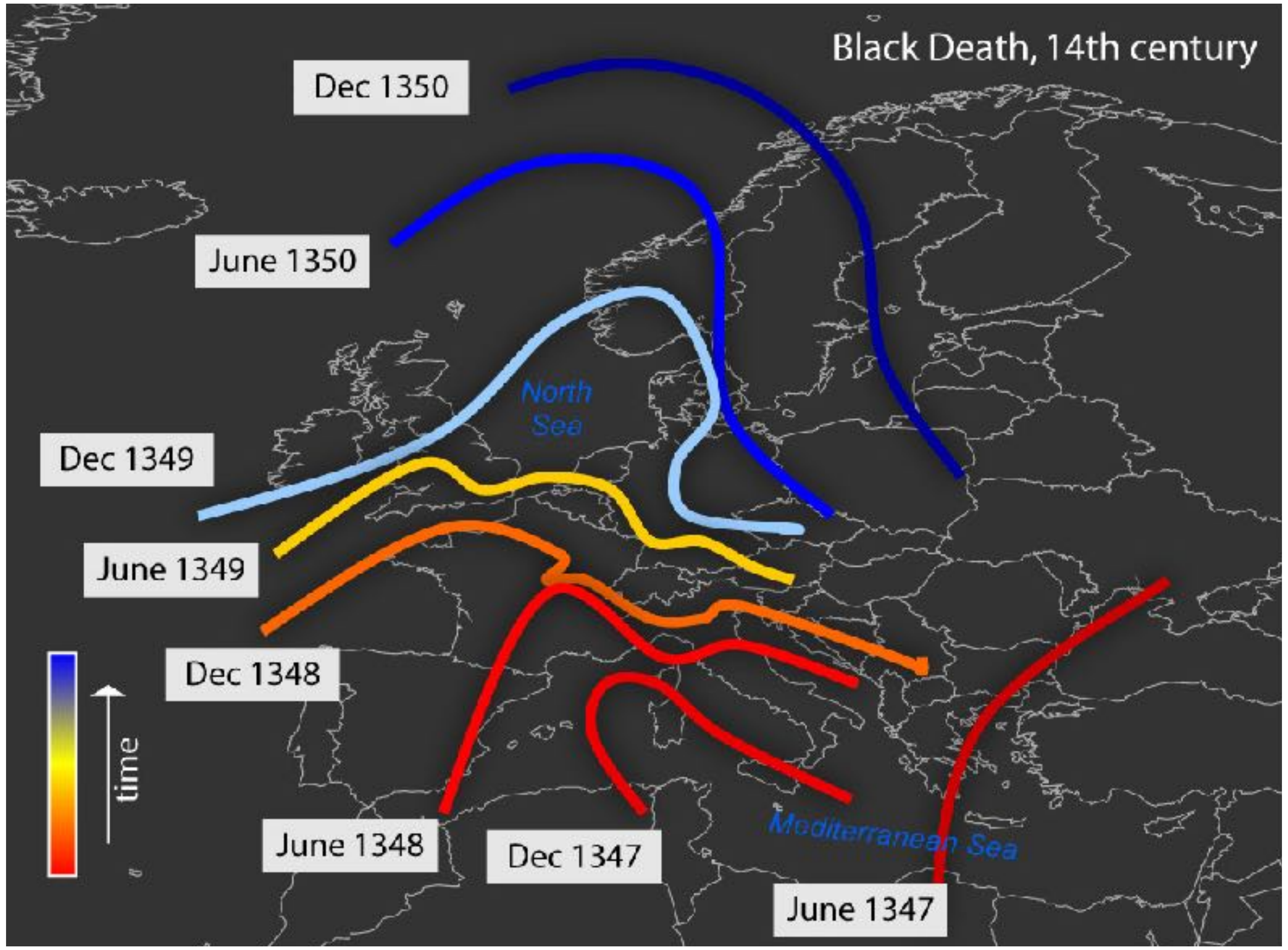
Exact on trees, works well with loopy networks



Newman, PRE, 2002

Spatial and metapopulation models

The Black Death (1347-1350)



Credits: Langer, Sci Amer 1964;
Chiara Poletto

Front velocity:

$$v = 2\sqrt{D(R_0 - 1)\alpha} \approx 140\,\text{miles} \cdot \text{y}^{-1}$$

Noble, *Nature*, 1974

Where:

R_0 : Basic reproductive number

α : Recovery/death rate

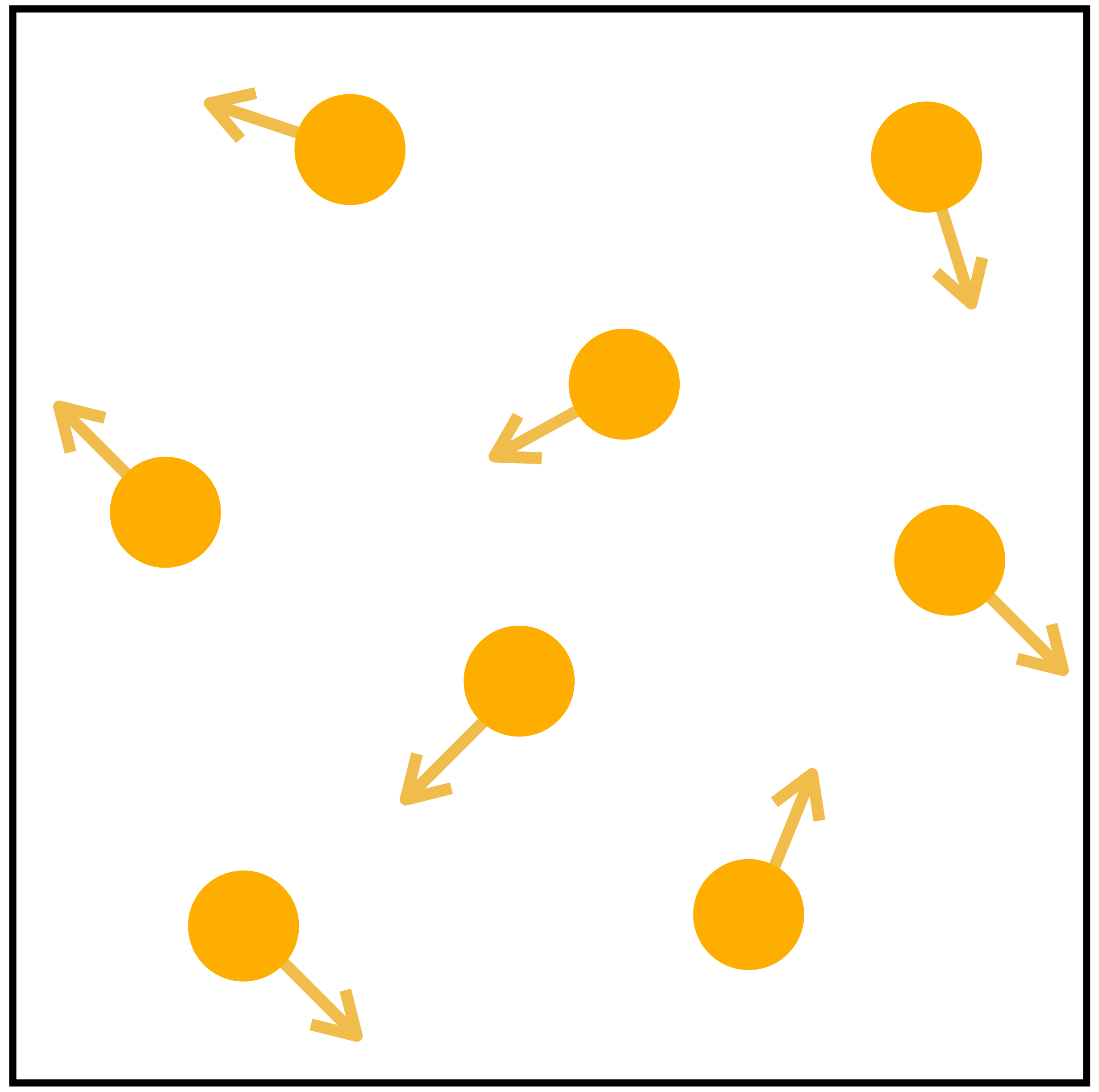
$\boxed{D : \text{Diffusion constant}}$

↖

Relates to host **dispersal**

Diffusion in space

Let $f(\vec{r}, t)$ denote the density of hosts in space and time



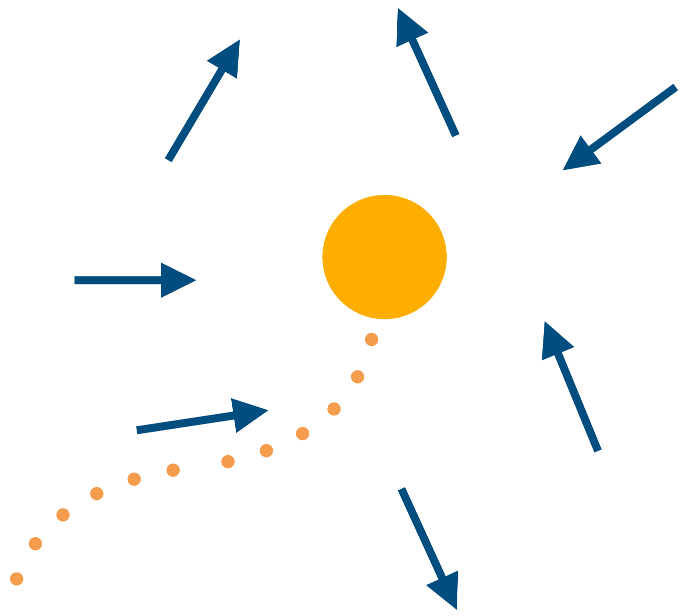
$$\frac{\partial f(\vec{r}, t)}{\partial t} = \underline{D\frac{\partial^2 f(\vec{r}, t)}{\partial x^2} + D\frac{\partial^2 f(\vec{r}, t)}{\partial y^2}}$$

2D Laplacian: $D\nabla^2 f$

This PDE describes hosts moving as **random walkers**:

It is equivalent to an equation of motion with a **random force** $\vec{\eta}$

$$\frac{d\vec{r}}{dt} = \vec{\eta}$$



$\vec{\eta}$ pulls the walker in random directions all the time

Reaction-diffusion equations

- **Diffusion** alone is boring
- **Diffusion** & **local reactions** create interesting patterns

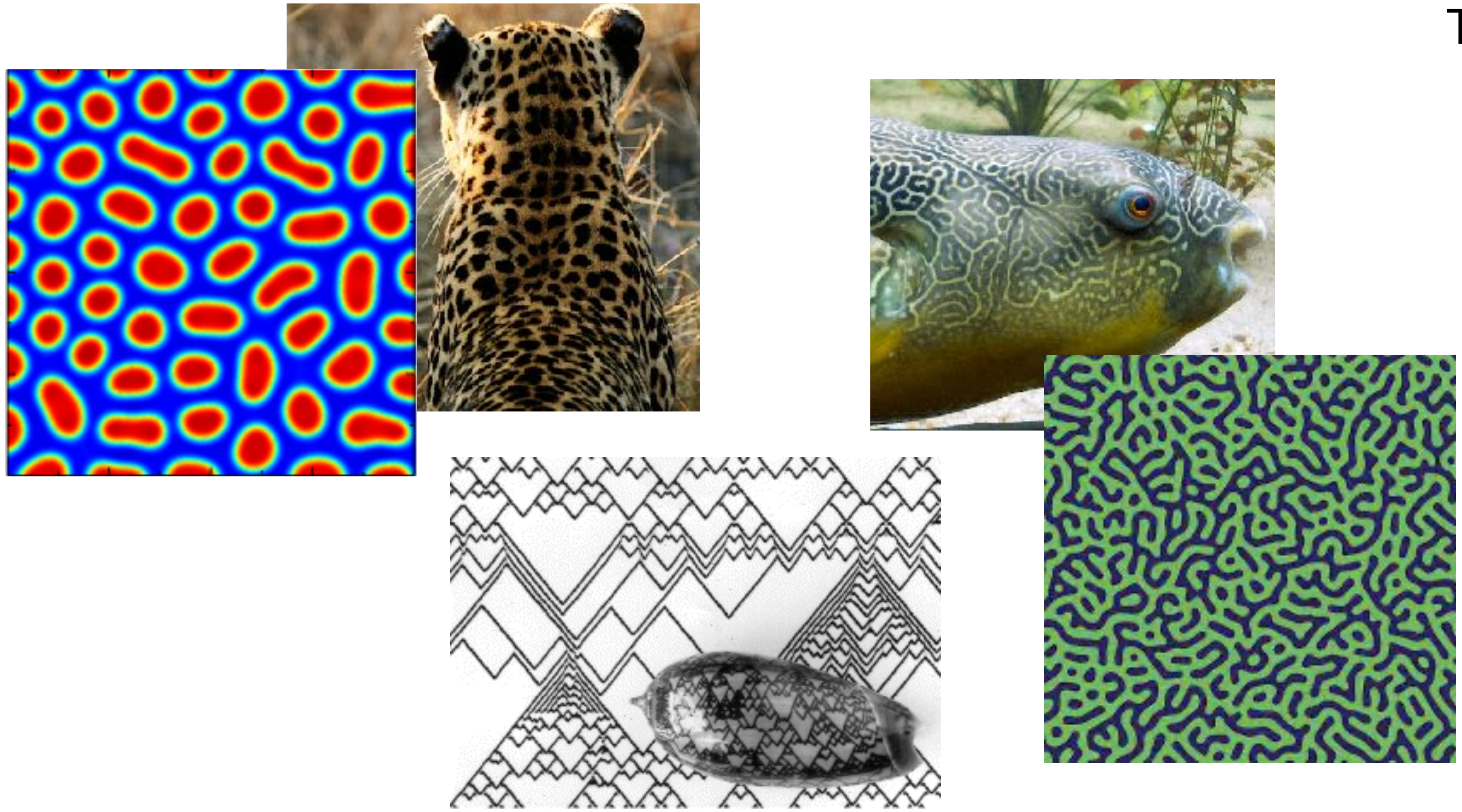
$$\frac{\partial f}{\partial t} = D\nabla^2 f + \boxed{R(f)}$$

Reaction term

E.g. local death:

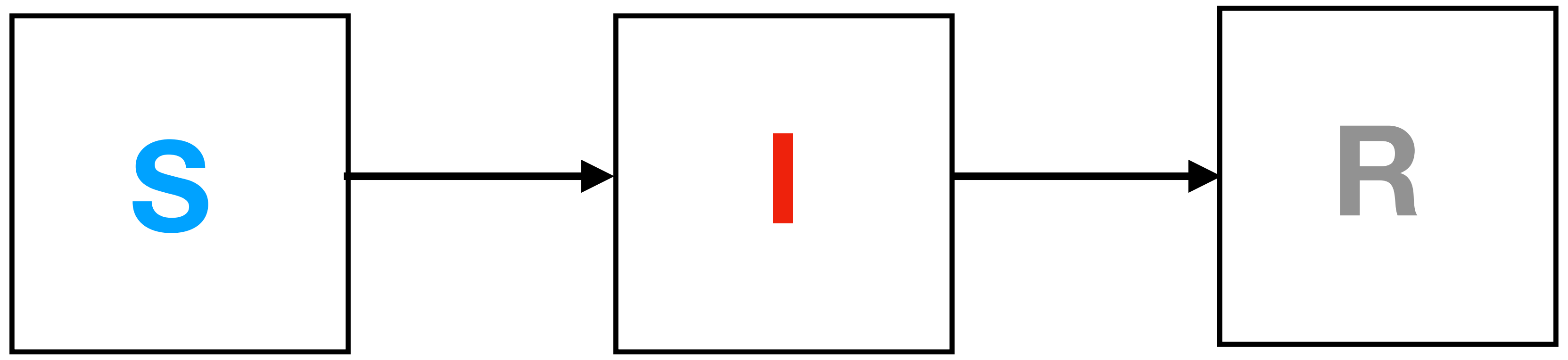
$R(\overrightarrow{x}, t) = -\alpha f(\overrightarrow{x}, t)$

Turing patterns are found everywhere!

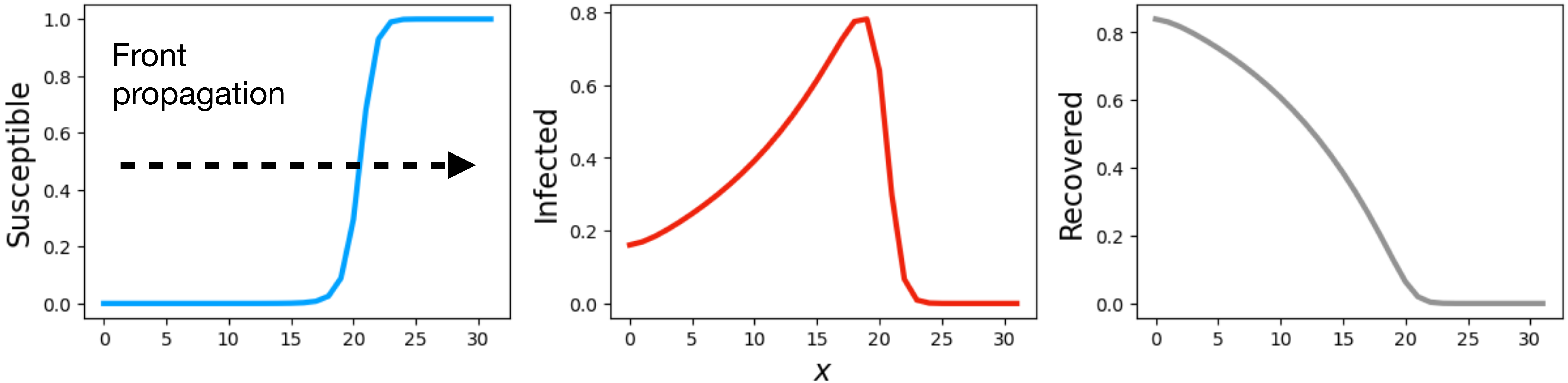


SIR model in 1D

$$\frac{\partial S}{\partial t} = D\frac{\partial^2 S}{\partial x^2} - \beta S I \qquad \frac{\partial I}{\partial t} = D\frac{\partial^2 I}{\partial x^2} + \beta S I - \alpha I$$



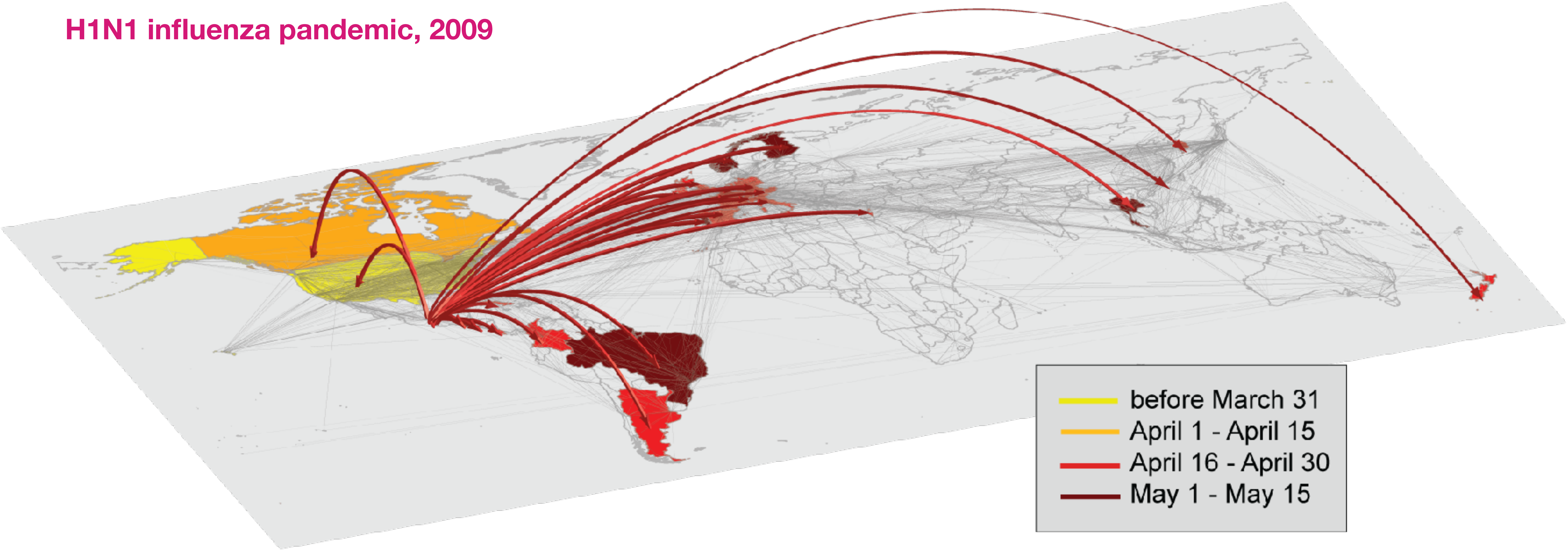
Travelling wave solution: $I(x,t) \to I(z = x - vt)$ with velocity: $v = 2\sqrt{D(\beta - \alpha)}$



Murray, Mathematical Biology: I an Introduction, 2007

Pathogens now fly first class

H1N1 influenza pandemic, 2009



n,

Geographic distance can't explain pathogen arrival times

Credits: Chiara Poletto

Mobility network data

Air travel: data collected by the International Air Transport Association (IATA)

How many people travelled daily on a given route?

n,

Hufnagel et al PNAS 2004
Colizza et al PLoS Med 2007
Balcan et al PNAS 2009

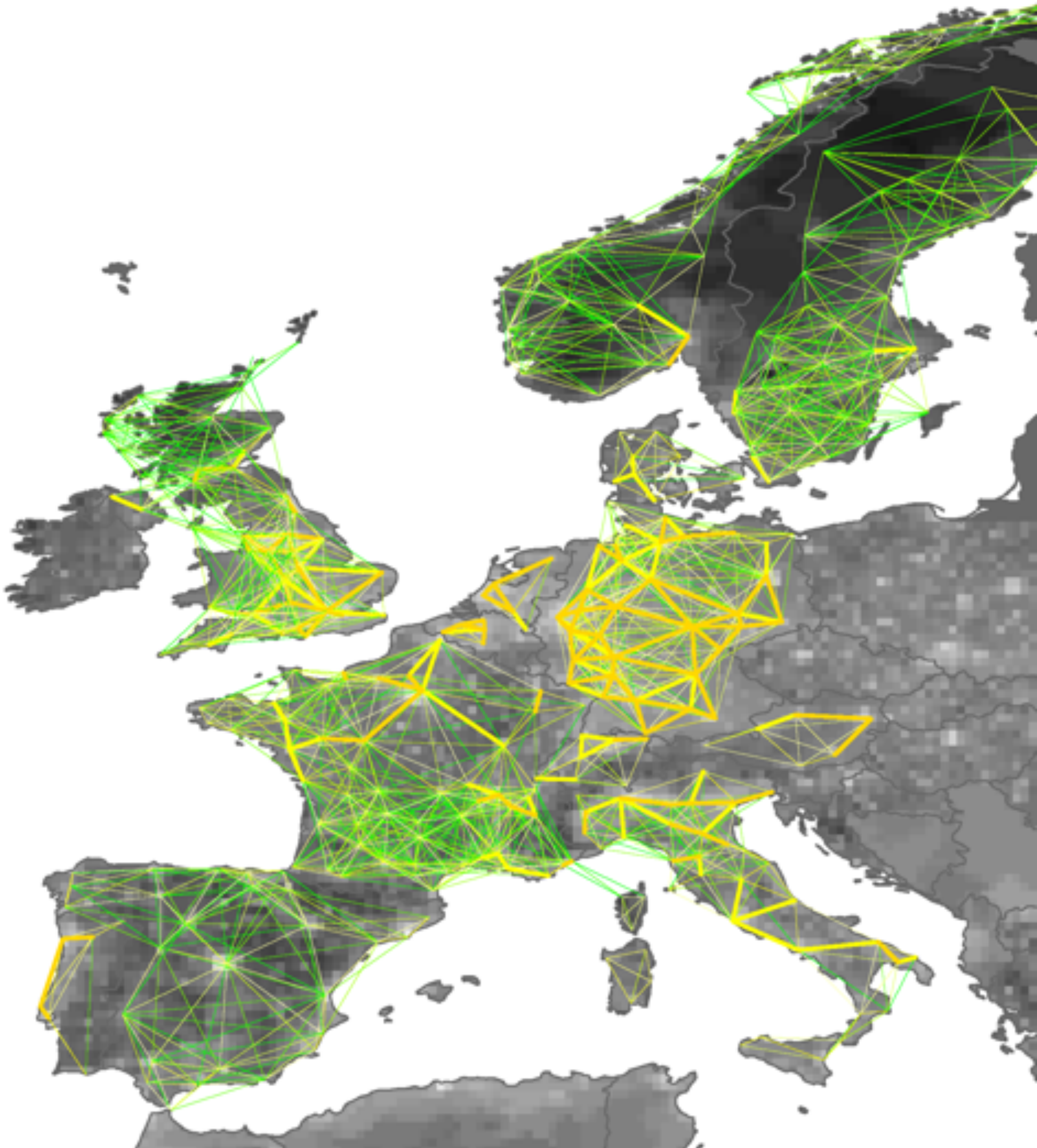
Mobility network data

Commuting: census data from different countries

How many people living at one location commute to another location?

n,

Balcan et al PNAS 2009



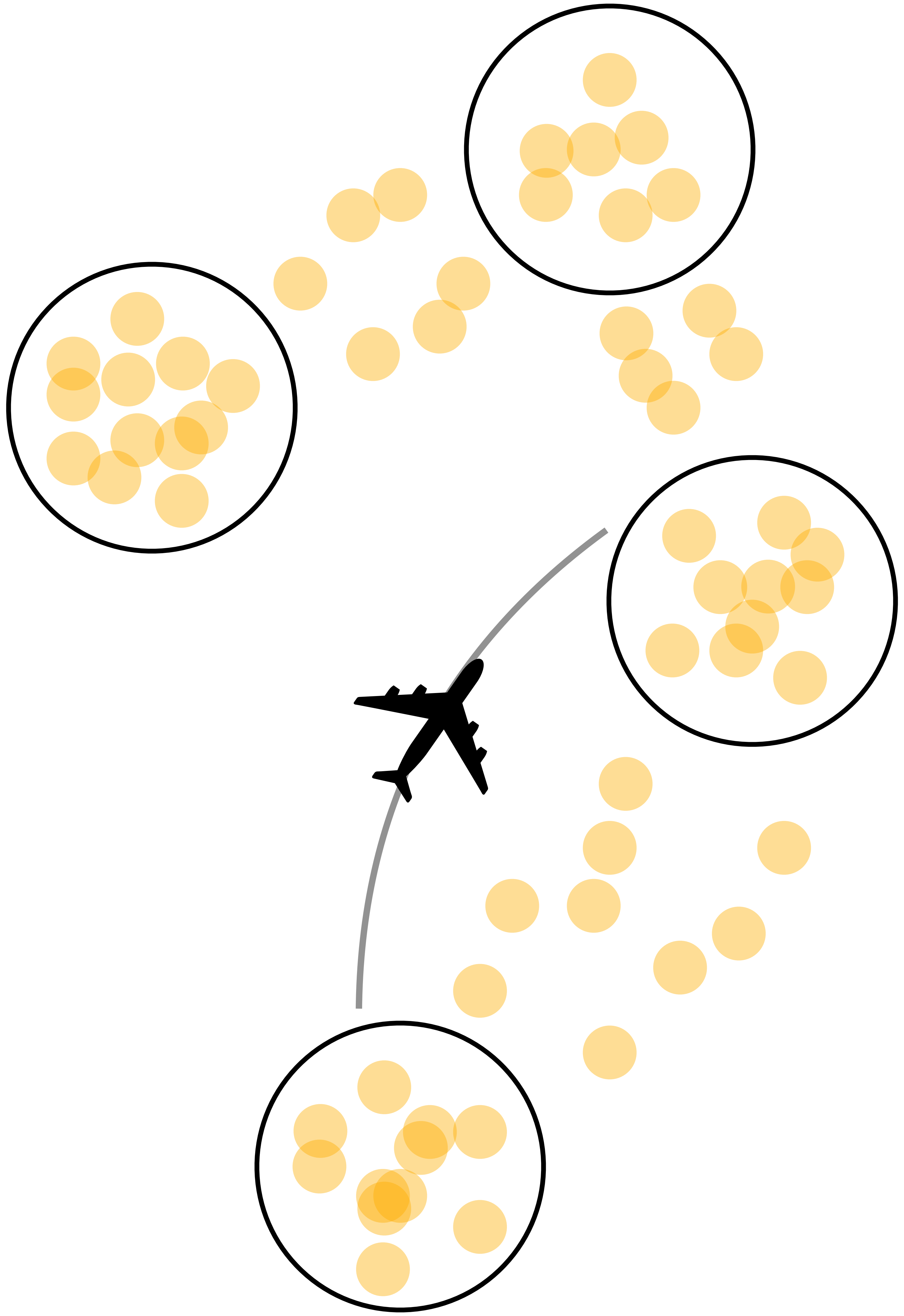
Metapopulations

A **metapopulation** is a population that has been partitioned in smaller subpopulations contained in discrete **patches**

Subpopulations may interact with each other through individual movements (**coupling**/**flows**)

An example relevant to infectious disease epidemiology:

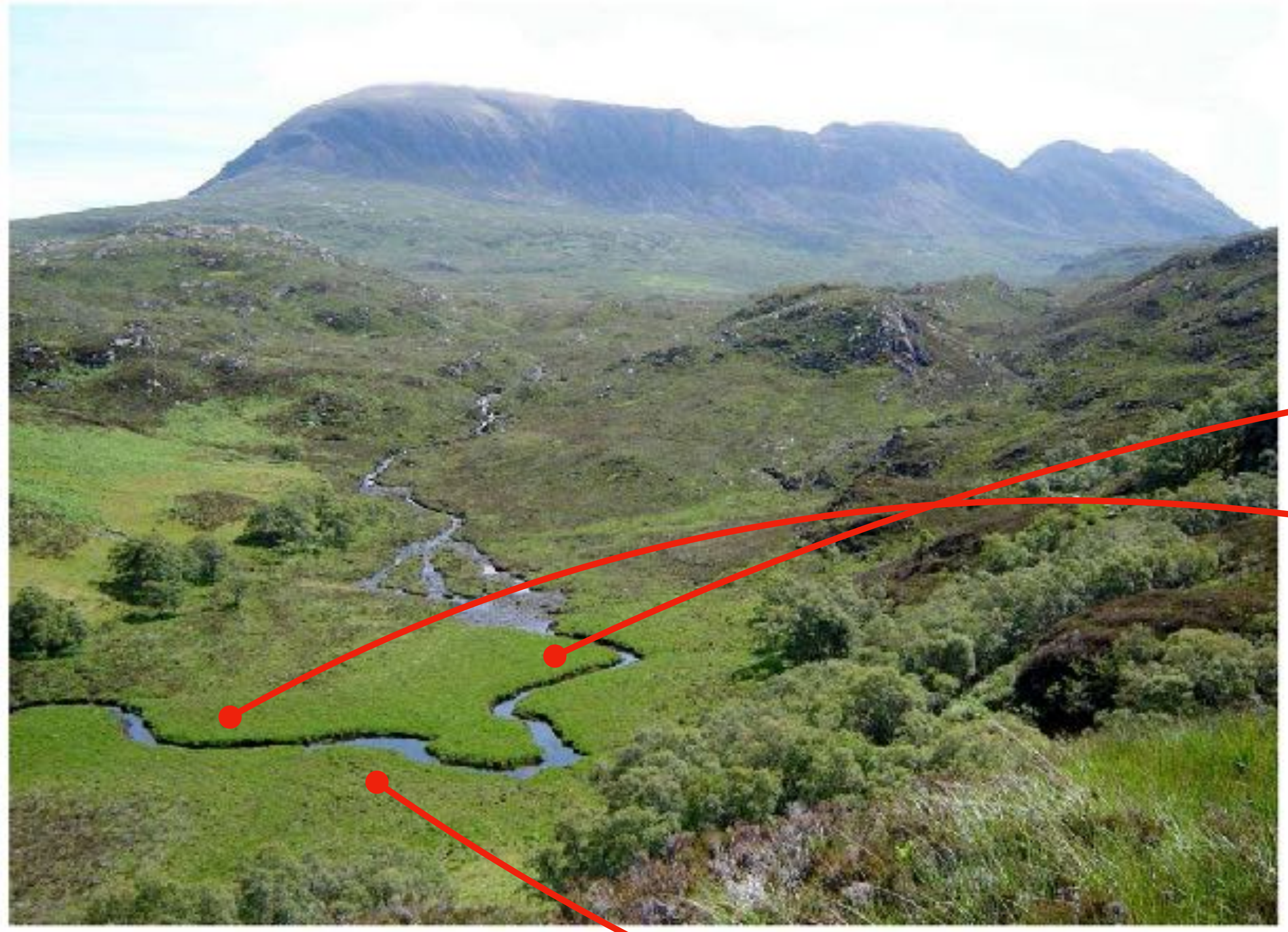
Patches: airports and surrounding areas
Coupling: air travel fluxes



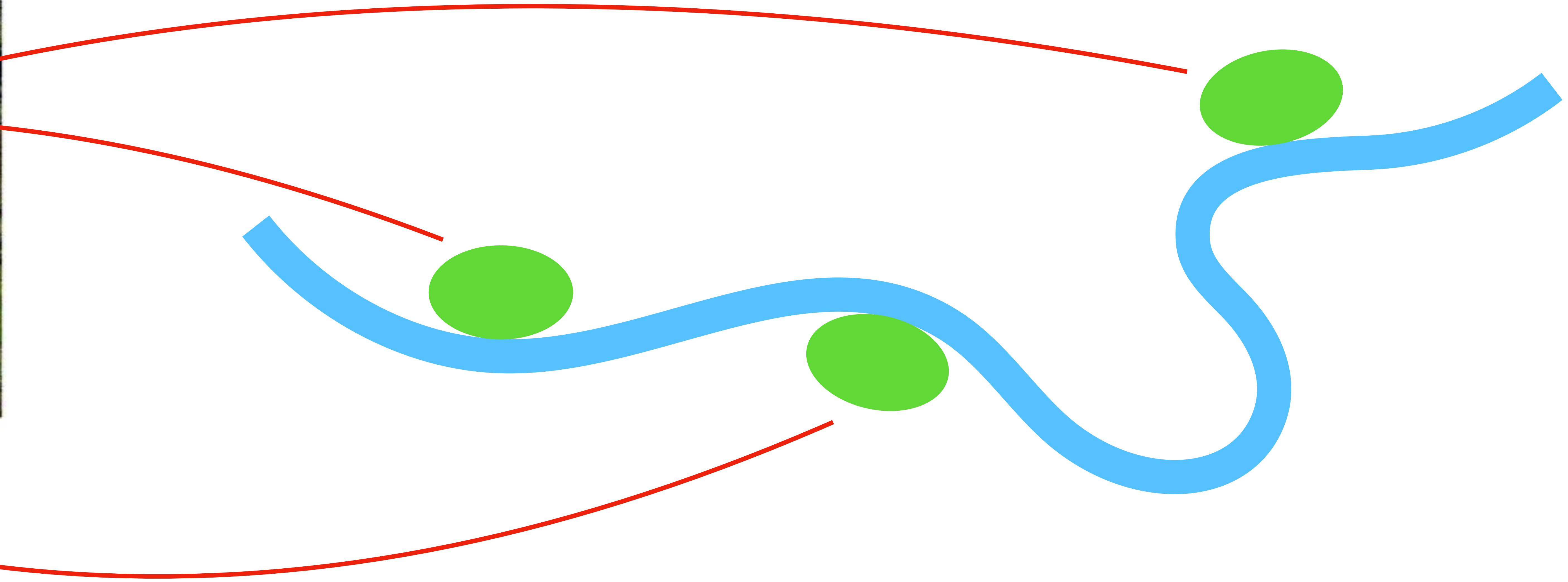
An example from ecology

Water voles occupy a **fragmented habitat** made of short stretches of grassland along rivers

Water voles can migrate between patches



Sutherland et al, Ecology 2012
Photo credit: Rosalie Burdon

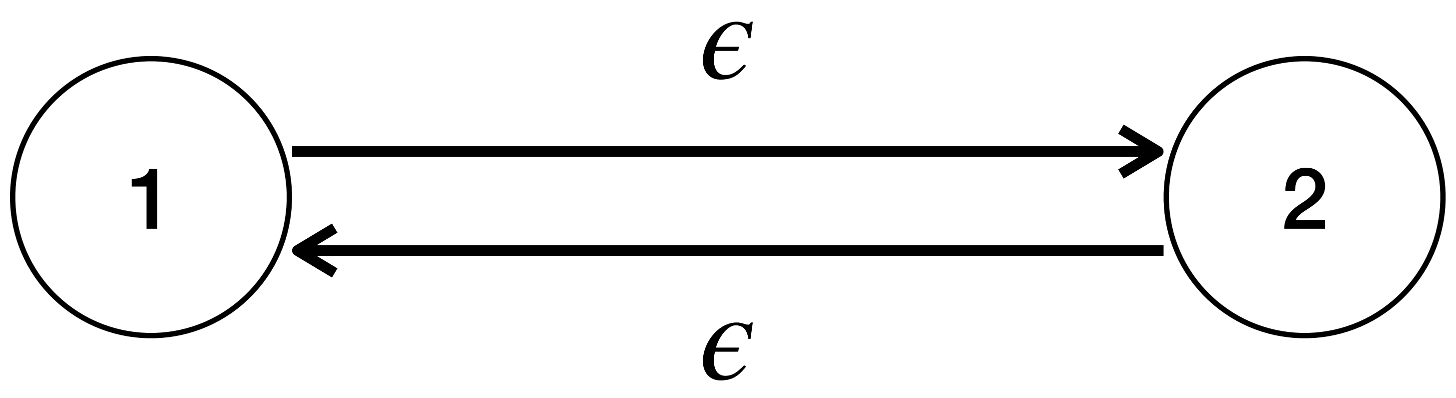


Diffusion on networks

Let us consider 2 subpopulations (nodes) with simple diffusion

$$\frac{dN_1}{dt} = \epsilon N_2 - \epsilon N_1$$

$$\frac{dN_2}{dt} = \epsilon N_1 - \epsilon N_2$$

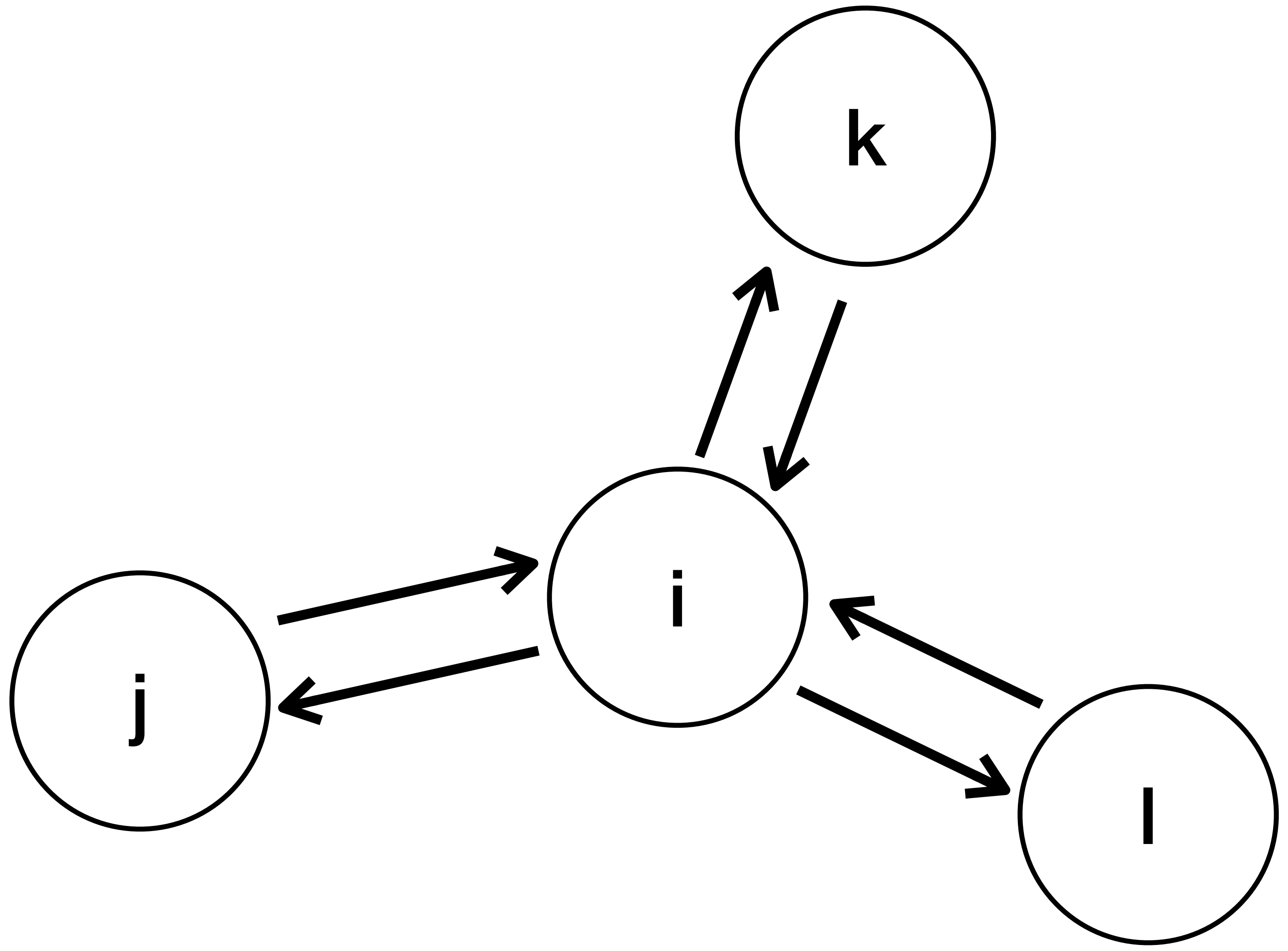


It is assumed that any individual travels at rate ϵ (**coupling**)

Diffusion on networks

When considering multiple nodes:

$$\frac{dN_i}{dt} = \epsilon(N_j + N_k + N_l) - 3\epsilon N_i$$



We find a general pattern:

$$\frac{dN_i}{dt} = -\epsilon \sum_j L_{i,j} \cdot N_j(t)$$

Graph Laplacian

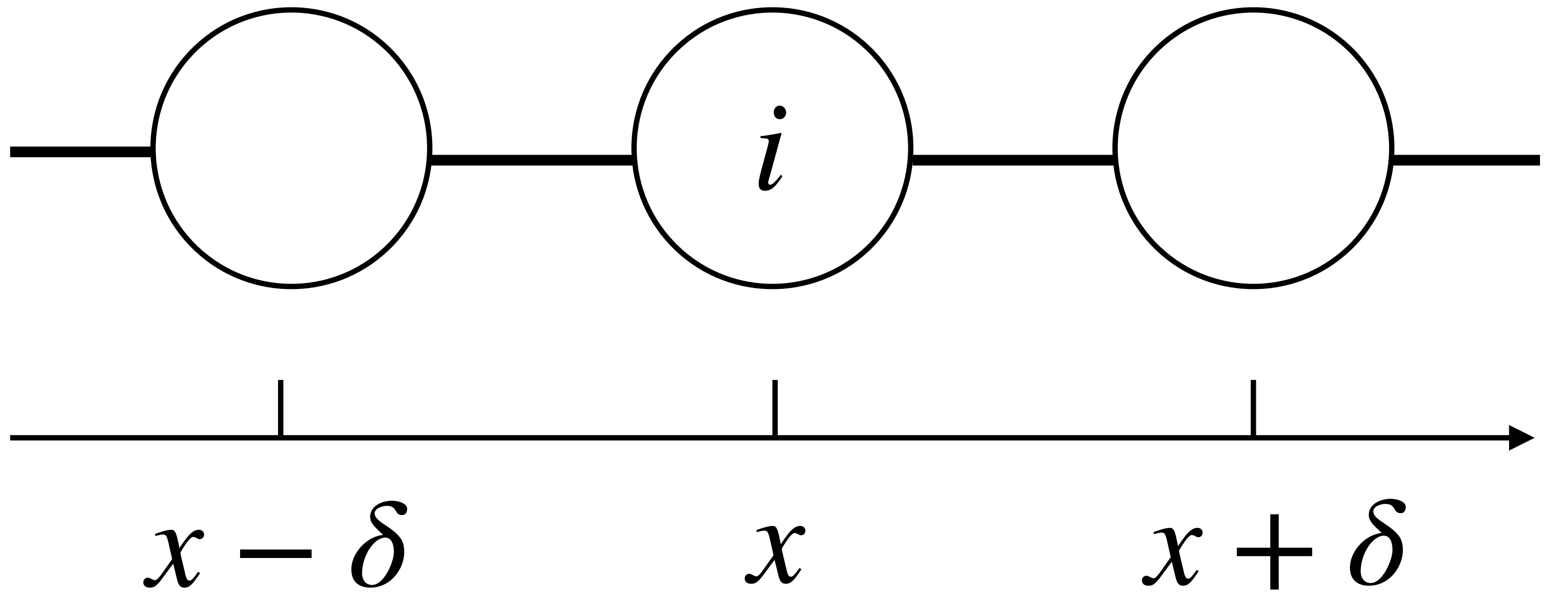
$$L_{i,j} = \begin{cases} k_i & \text{if } i = j \\ -A_{i,j} & \text{if } i \neq j \end{cases}$$

Recovering spatial diffusion

The graph laplacian extends the operator ∇^2 to networks.

Let us consider a special graph: a 1D array of evenly spaced (spacing: δ) nodes

$$-\sum_j L_{i,j} \cdot f_j = f(x-\delta) + f(x+\delta) - 2f(x)$$



Expanding: $$f(x \pm \delta) \approx f(x) \pm \delta f'(x) + \frac{\delta^2}{2} f''(x)$$

Hence: $$-\sum_j L_{i,j} \cdot f_j \approx \delta^2 \frac{\partial^2 f}{\partial x^2}$$

We obtain the Laplace operator!

Heterogeneous rates

Host at some locations are more likely to travel than others. Also, some destinations are more popular than others.

$$\frac{dN_i}{dt} = -\epsilon_i N_i + \sum_j \epsilon_{j,i} \cdot N_j$$

$\epsilon_{i,j}$ **: rate of travelling from i to j**

$\epsilon_i = \sum_j \epsilon_{i,j}$ **: total rate of travelling from i**

If an average of $w_{i,j}$ individuals travel from i to j in one day, we can estimate these rates:

$$\hat{\epsilon}_i = \frac{\sum_j w_{i,j}}{\bar{N}_i}$$

Total people leaving in i

$$\epsilon_{i,j} = \hat{\epsilon}_i \cdot \boxed{\frac{w_{i,j}}{\sum_j w_{i,j}}}$$

Probability of choosing j as a destination when leaving from i

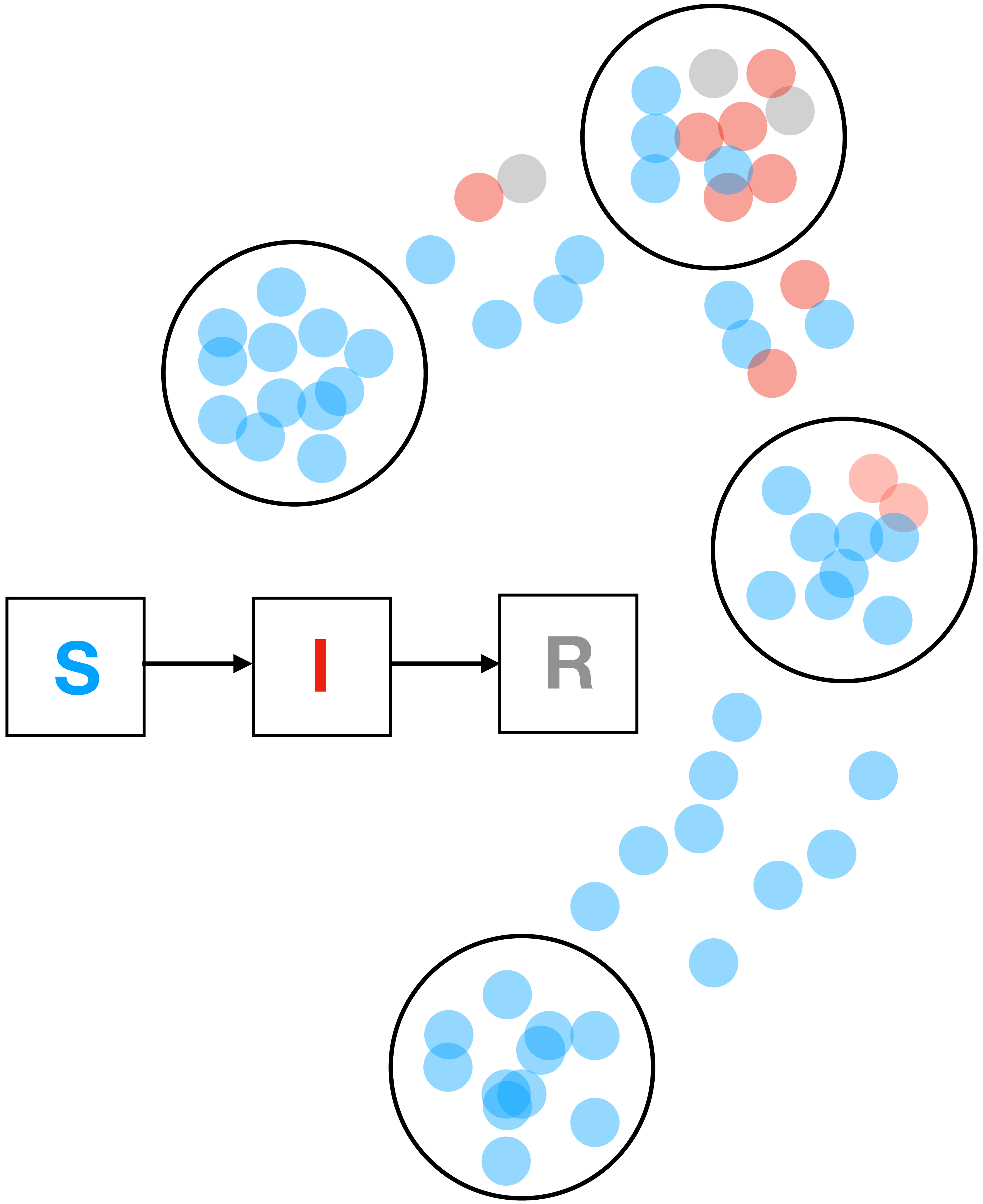
SIR metapopulation model

Still a reaction-diffusion system

$$\frac{dS_i}{dt} = -\beta \frac{S_i I_i}{N_i} - \epsilon_i S_i + \sum_j \epsilon_{j,i} \cdot S_j$$

$$\frac{dI_i}{dt} = \beta \frac{S_i I_i}{N_i} - \alpha I_i - \epsilon_i I_i + \sum_j \epsilon_{j,i} \cdot I_j$$

$$\frac{dR_i}{dt} = \alpha I_i - \epsilon_i R_i + \sum_j \epsilon_{j,i} \cdot R_j$$



Global invasion threshold

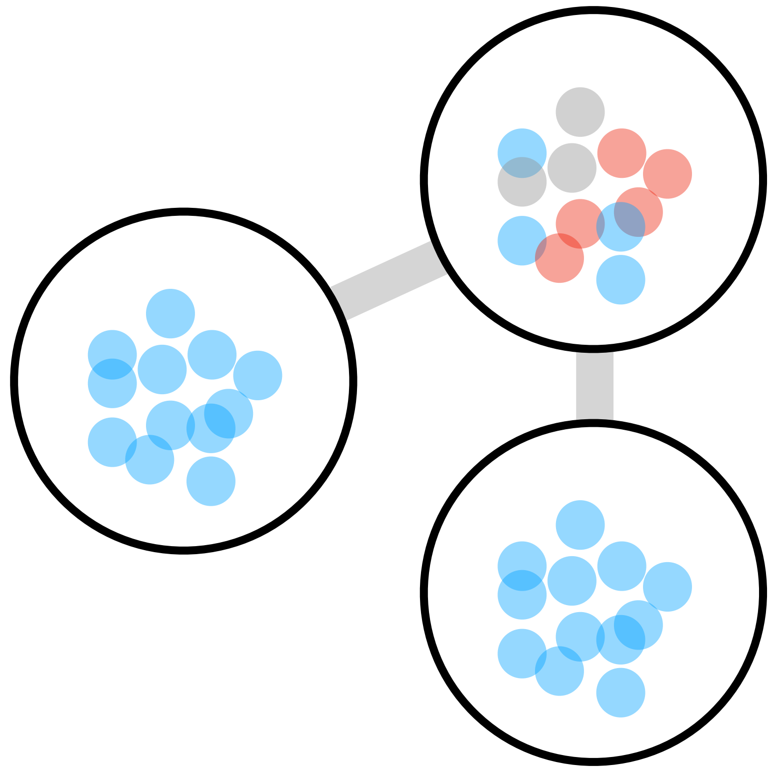
Local spread occurs when $R_0 = \beta/\alpha > 1$ but does not guarantee global spread (many patches invaded)

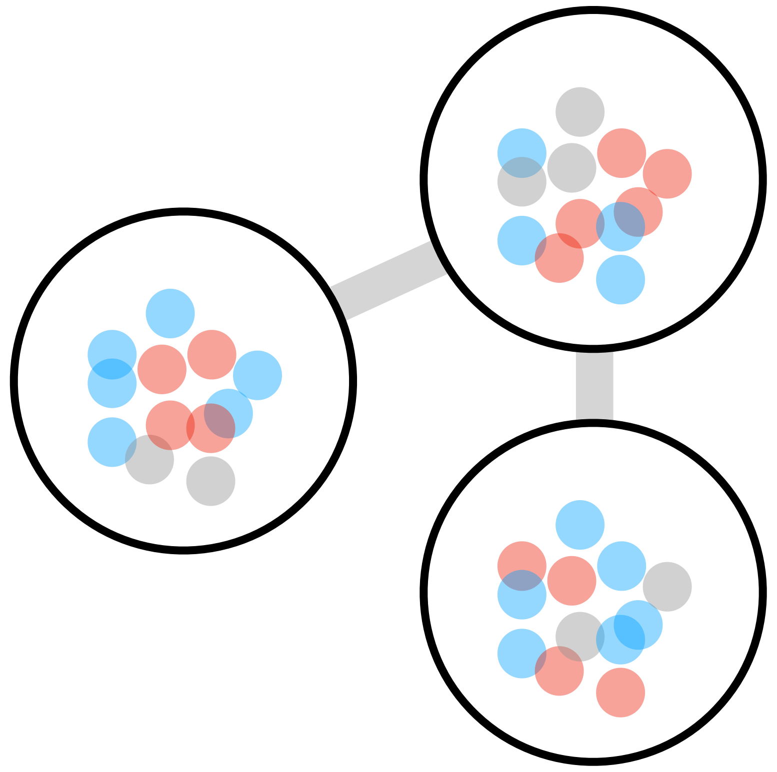
$\epsilon = 0$

$\epsilon \to \infty$

No global spread

Global spread

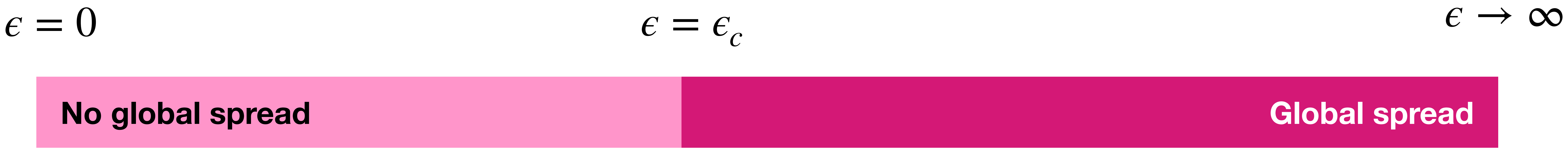




Colizza & Vespignani, PRL 2007, JTB 2008

Global invasion threshold

Local spread occurs when $R_0 = \beta/\alpha > 1$ but does not guarantee global spread (many patches invaded)



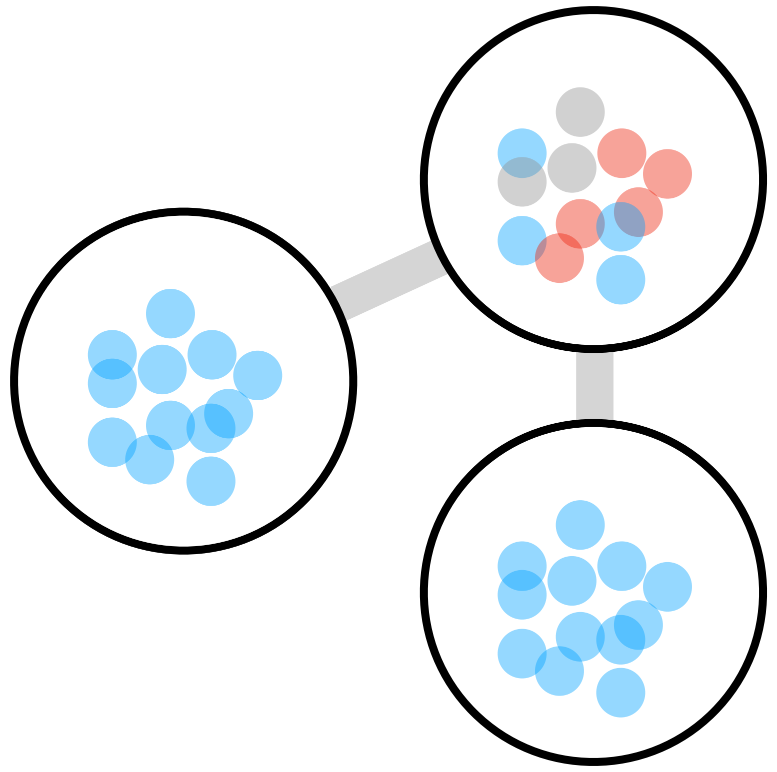
$\epsilon = 0$

$\epsilon = \epsilon_c$

$\epsilon \to \infty$

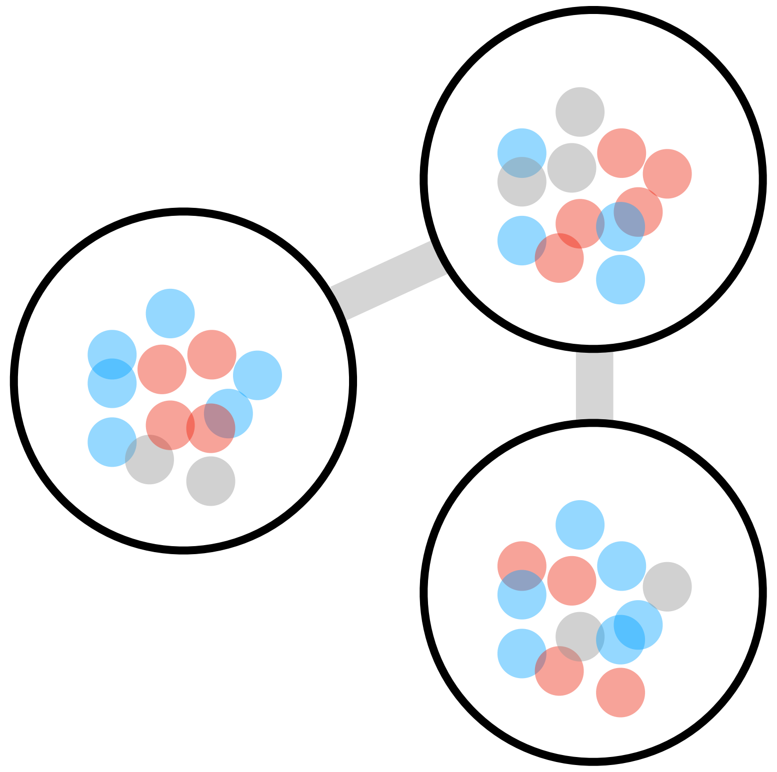
No global spread

Global spread



Global invasion requires $\epsilon > \epsilon_c$ for fixed R_0

Equivalently: $R_g(R_0, \epsilon) > 1$



Colizza & Vespignani, PRL 2007, JTB 2008

Global invasion threshold

Let D_k^n be the number of patches with degree k invaded at generation n

$$D_k^n = \sum_{k'} D_{k'}^{n-1} \cdot (k' - 1) \cdot \frac{kP_k}{\langle k \rangle} \cdot \left(1 - \frac{\sum_{m=0}^{n-1} D_k^m}{V_k} \right) \cdot \left[1 - \left(\frac{1}{R_0} \right)^{\lambda_{k,k'}} \right]$$

1. Invaded patches at $t - 1$ with degree k'

2. $k' - 1$ possible channels

3. Prob connecting to a degree k patch

4. Prob uninvaded degree k patch

V_k : #patches with degree k

5. Prob successful invasion given $\lambda_{k,k'}$ infected travellers

Colizza & Vespignani, PRL 2007, JTB 2008

Global invasion threshold

Initially almost all patches are uninvaded and $\lambda_{k,k'} \approx \bar{\lambda} \ll 1$:

$$D_k^n = \sum_{k'} D_{k'}^{n-1} \cdot (k'-1) \cdot \frac{kP_k}{\langle k \rangle} \cdot \bar{\lambda} \left(R_0 - 1\right)$$

1. Prop infected

3. Prop selecting edge

$$\bar{\lambda} = \bar{N} \frac{2(R_0-1)}{R_0^2} \cdot \frac{\epsilon}{\alpha} \cdot \frac{1}{\langle k \rangle}$$

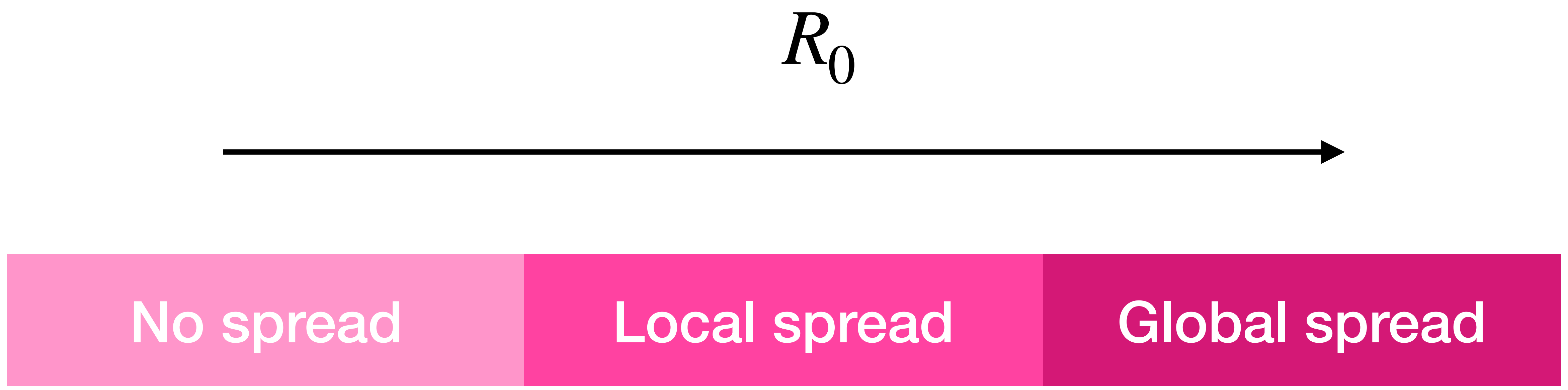
2. Prop travelling while infectious

Define $\theta^n = \sum_k D_k^n (k-1)$

$$\theta^n = \underbrace{\boxed{\frac{\langle k^2 \rangle - \langle k \rangle}{\langle k \rangle} \cdot \frac{2(R_0-1)^2}{R_0^2} \frac{\epsilon \bar{N}}{\alpha}}}_{R_g(R_0, \epsilon)} \cdot \theta^{n-1}$$

Colizza & Vespignani, PRL 2007, JTB 2008

Global invasion threshold



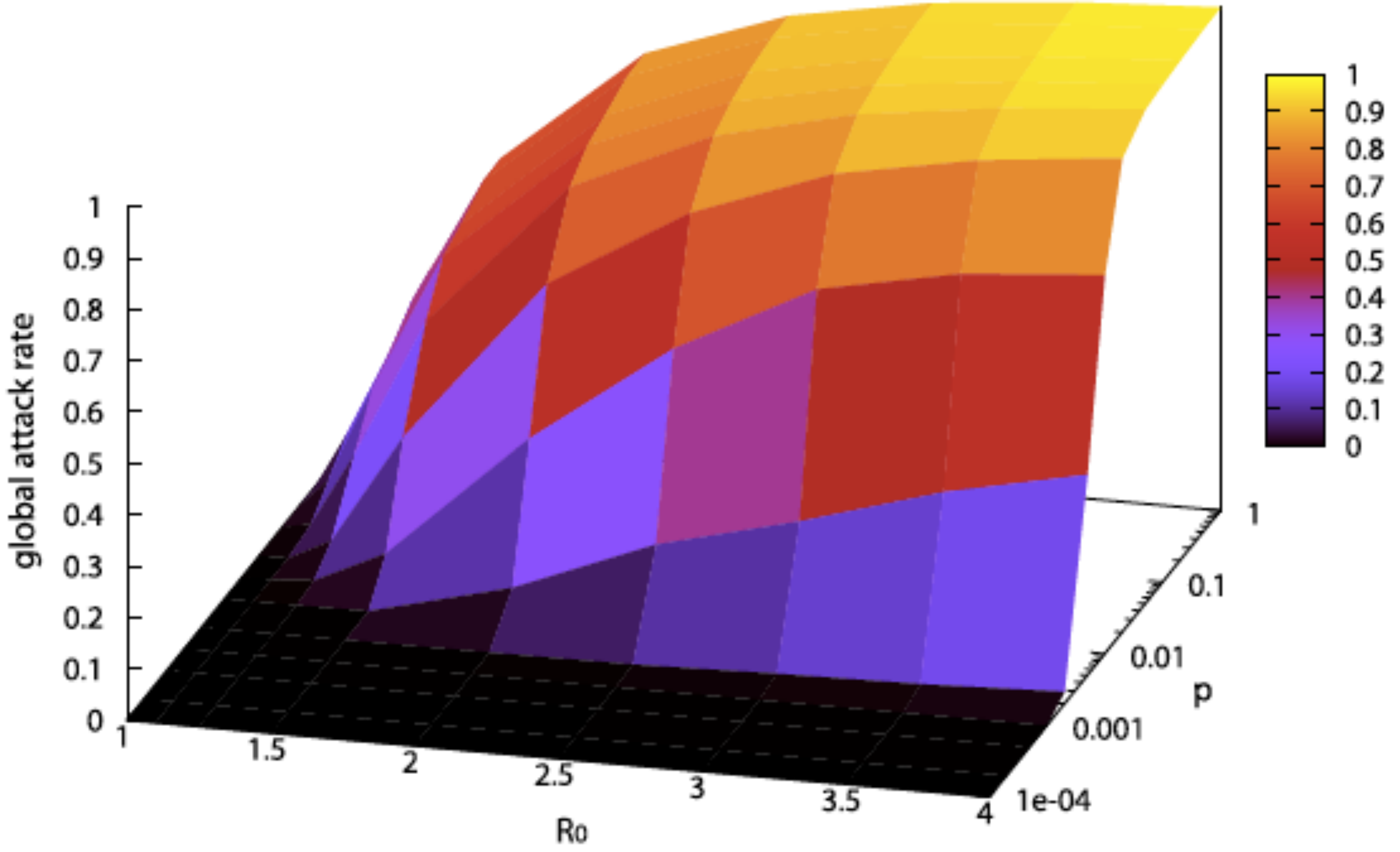
R_0

No spread

Local spread

Global spread

$$R_g = \frac{\langle k^2 \rangle - \langle k \rangle}{\langle k \rangle} \cdot \frac{2(R_0 - 1)^2}{R_0^2} \frac{\epsilon \bar{N}}{\alpha}$$



Colizza & Vespignani, PRL 2007, JTB 2008

Effective distance

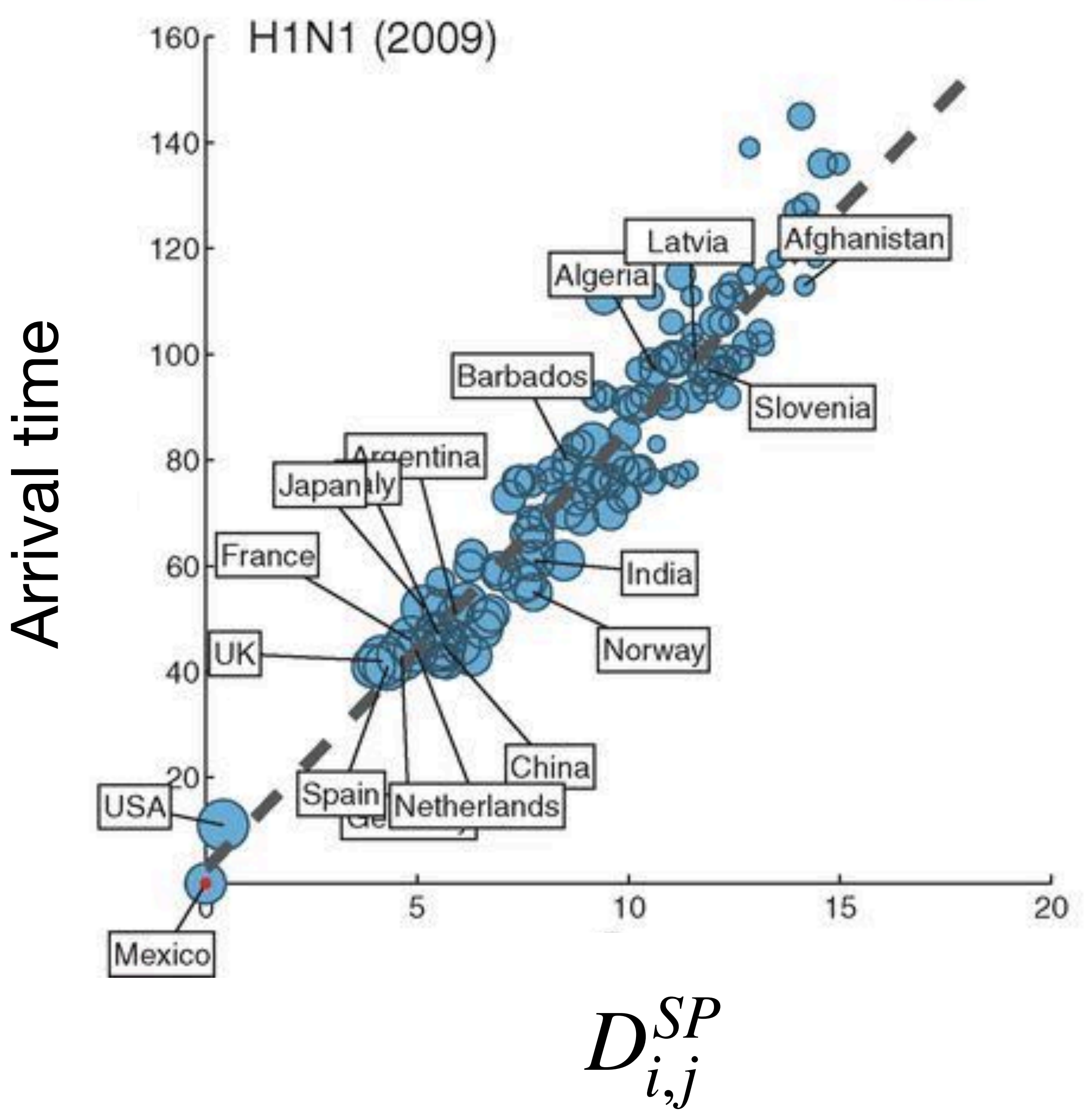
Passenger flows, not geographic distance, predict arrival times

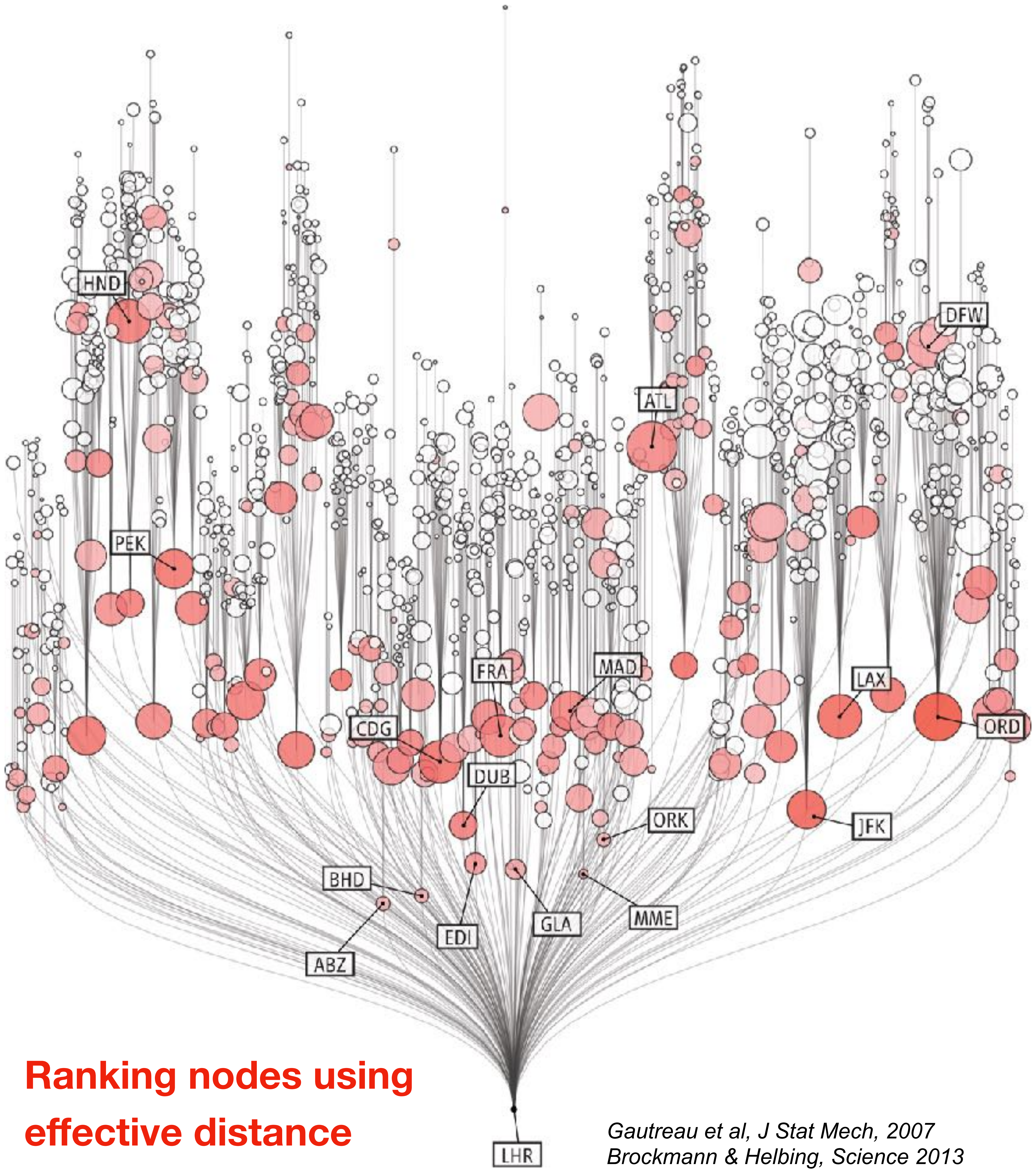
Shortest path distance

$$D_{i,j}^{SP} = \sum_{(m,n)\in SP_{i,j}} D_{m,n}$$

Effective distance

$$D_{i,j} = 1 - ln\left(\frac{\epsilon_{i,j}}{\sum_j \epsilon_{i,j}}\right)$$



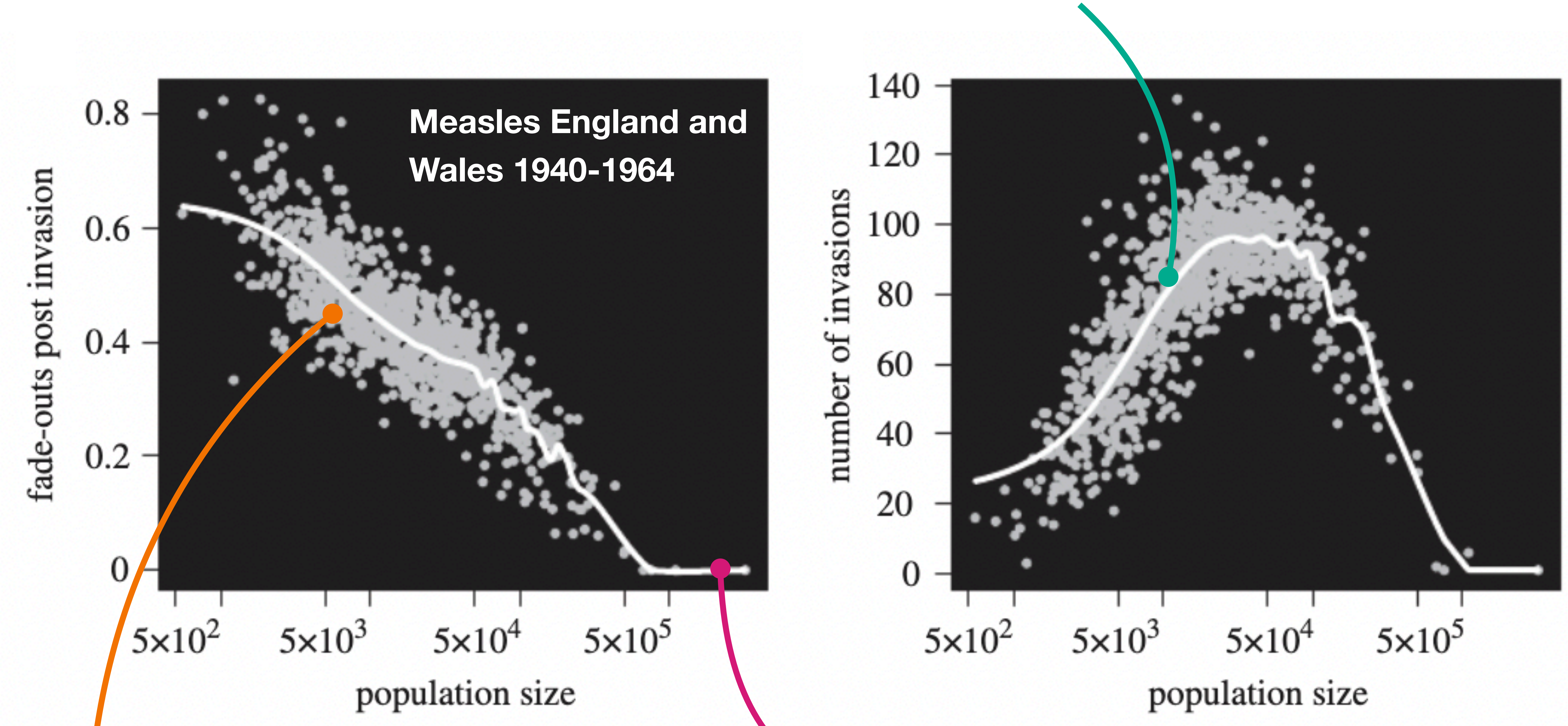


Ranking nodes using effective distance

Gautreau et al, J Stat Mech, 2007
Brockmann & Helbing, Science 2013
Iannelli et al, PRE, 2017

Persistence in a metapopulation

Local extinction is possible but global extinction is unlikely due to re-invasions



Measles fadeouts in small towns

No fadeouts in large towns (e.g. London)

Conlan et al, JRSI, 2009

Recurrent mobility

People are not random walkers: they move for a wide range of reasons (commuting, leisure, buying groceries..) but then return to their house after some time

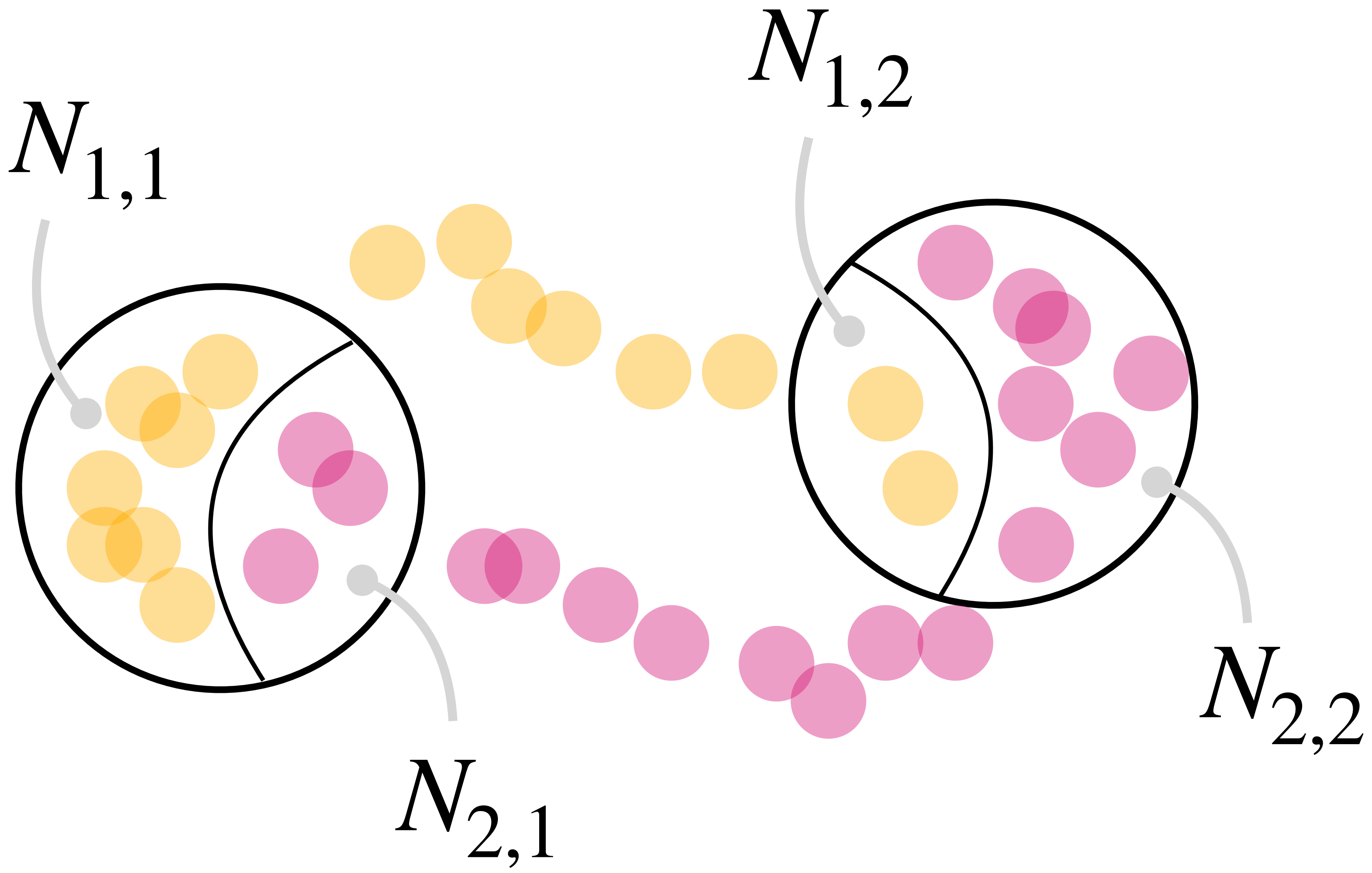
This can be accounted for in a metapopulation framework by tracking who lives where

$N_{i,j}$: # of hosts living in i but currently in j

$$\frac{dN_{i,i}}{dt} = -\epsilon_i N_{i,i} + \tau^{-1} \sum_j N_{i,j}$$

$$\frac{dN_{i,j}}{dt} = -\tau^{-1} N_{i,i} + \epsilon_{i,j} N_{i,i}$$

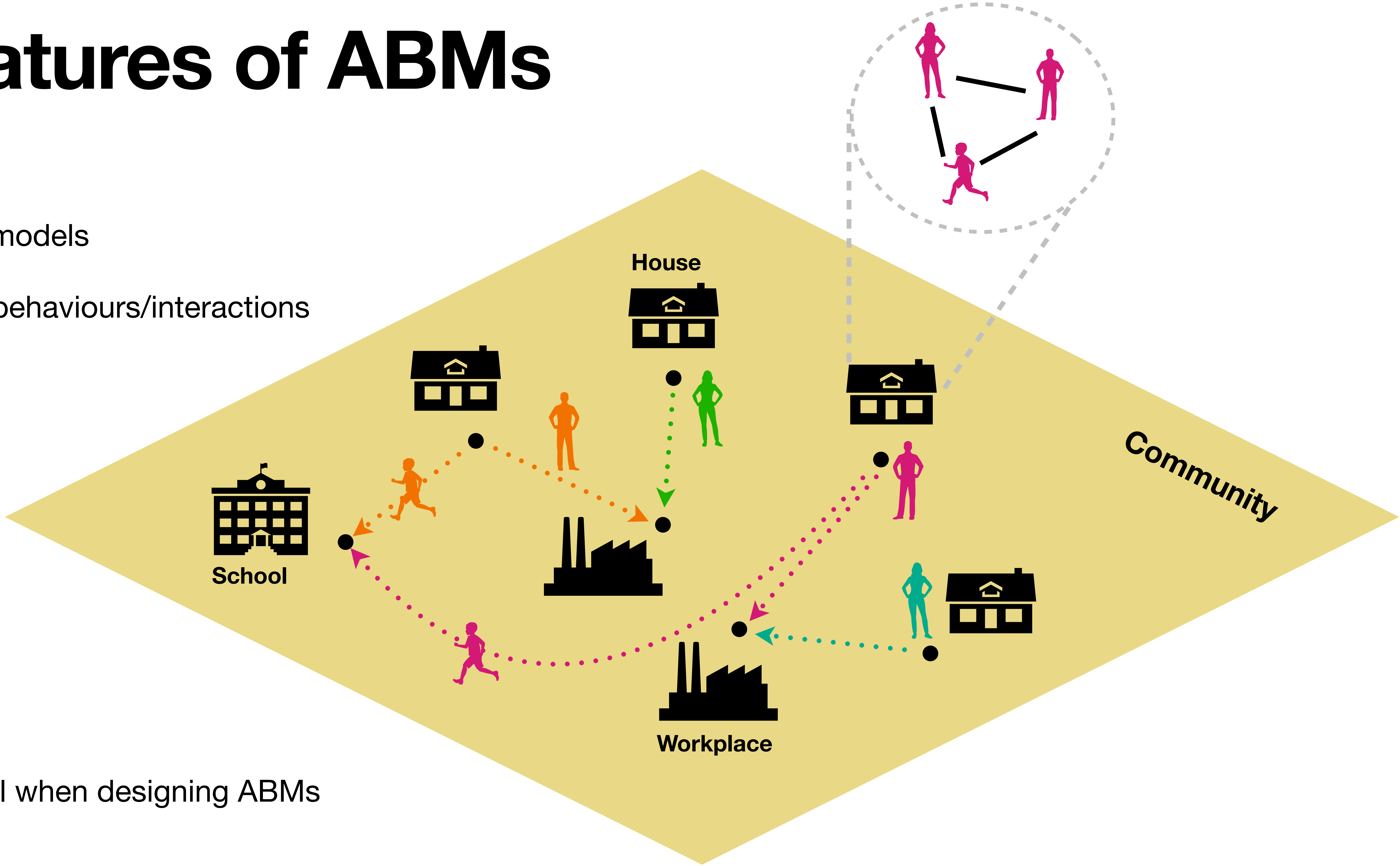
τ : **Time spent at destination**



Agent-based models

Some features of ABMs

- Computer simulation models
- Agents with complex behaviours/interactions
- Multi-scale
- Flexible & extendable
- Difficult to fit

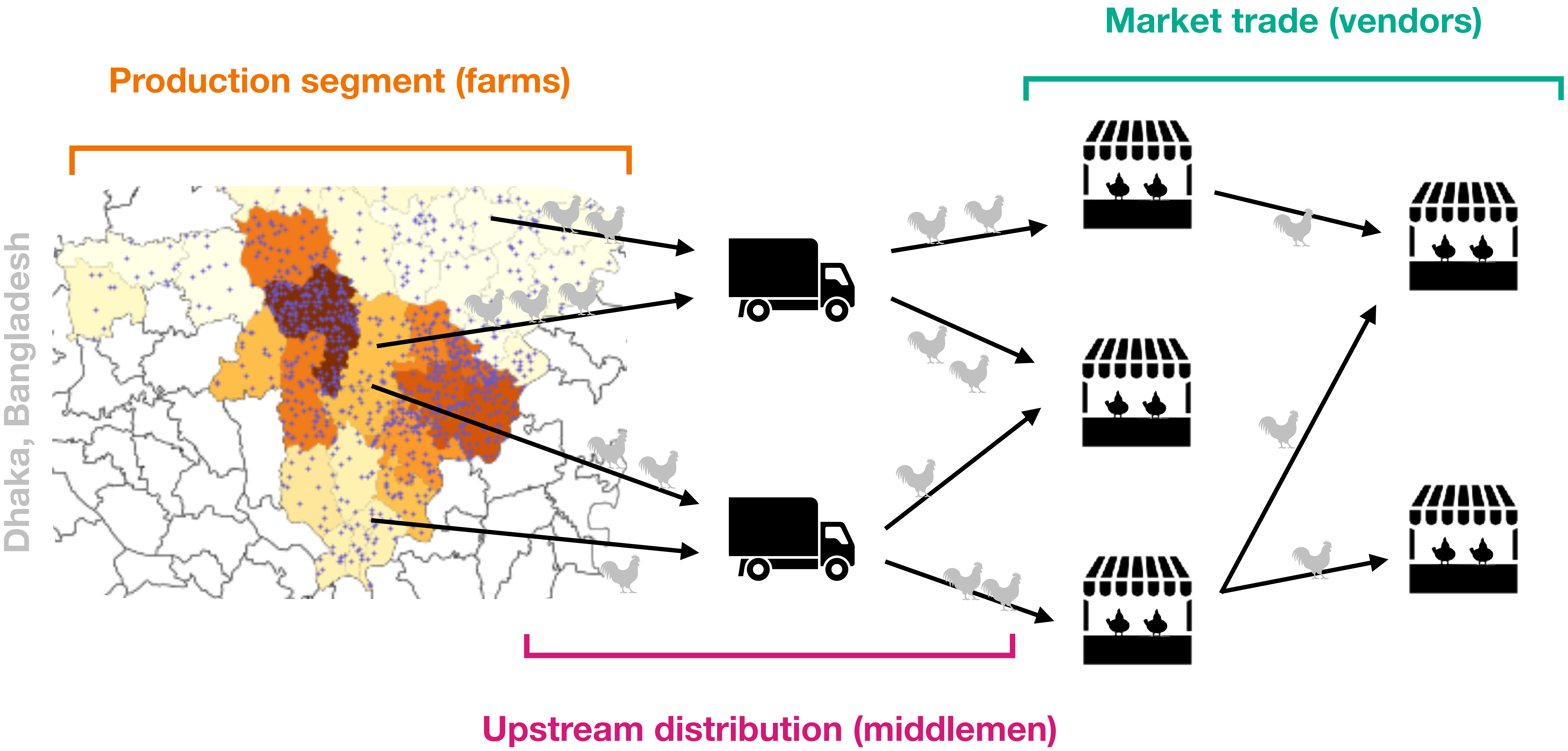


Networks can be useful when designing ABMs

An ABM for poultry movements

ONE HEALTH POULTRY HUB

An ABM to simulate transmission in a poultry supply chain

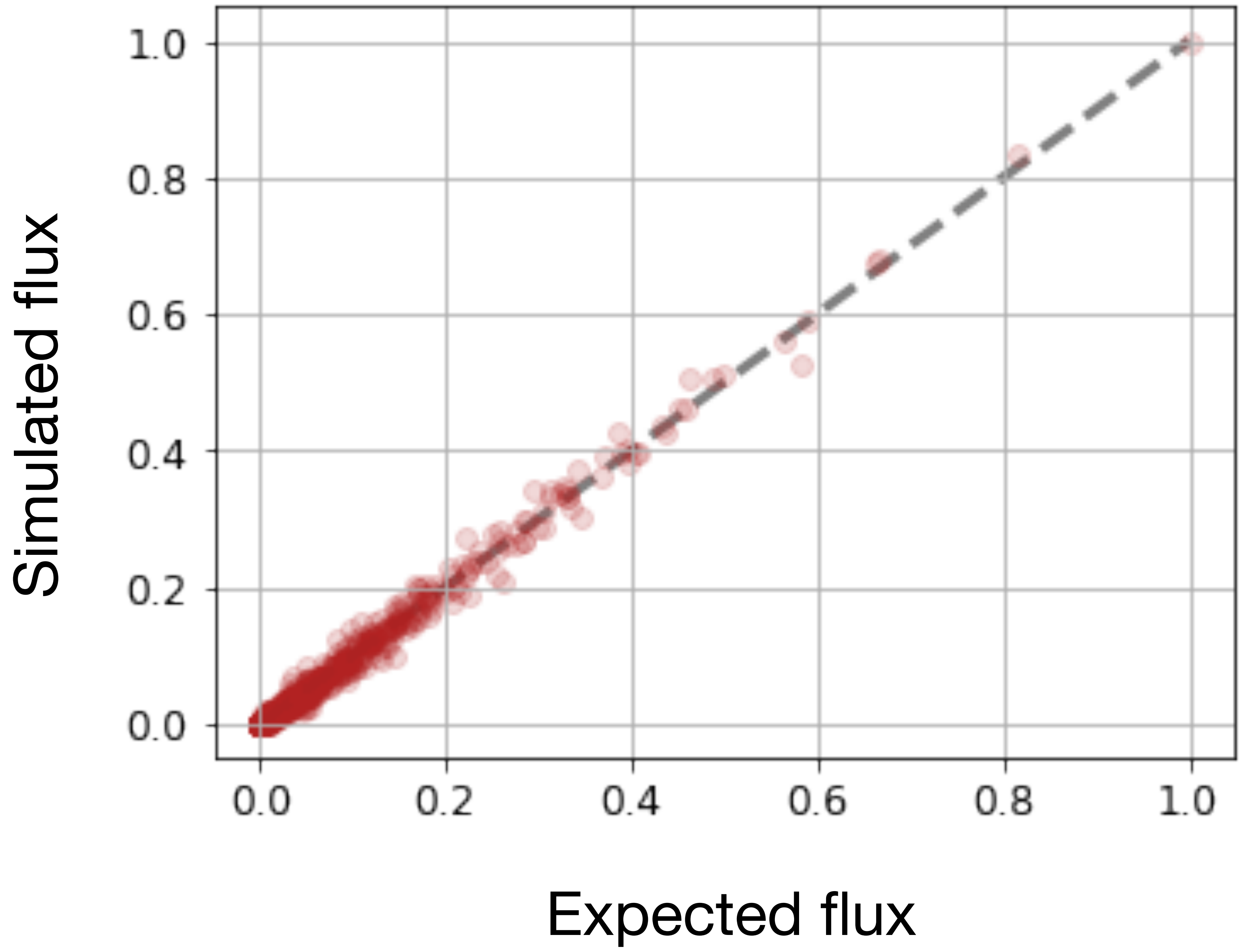


Realistic poultry movements

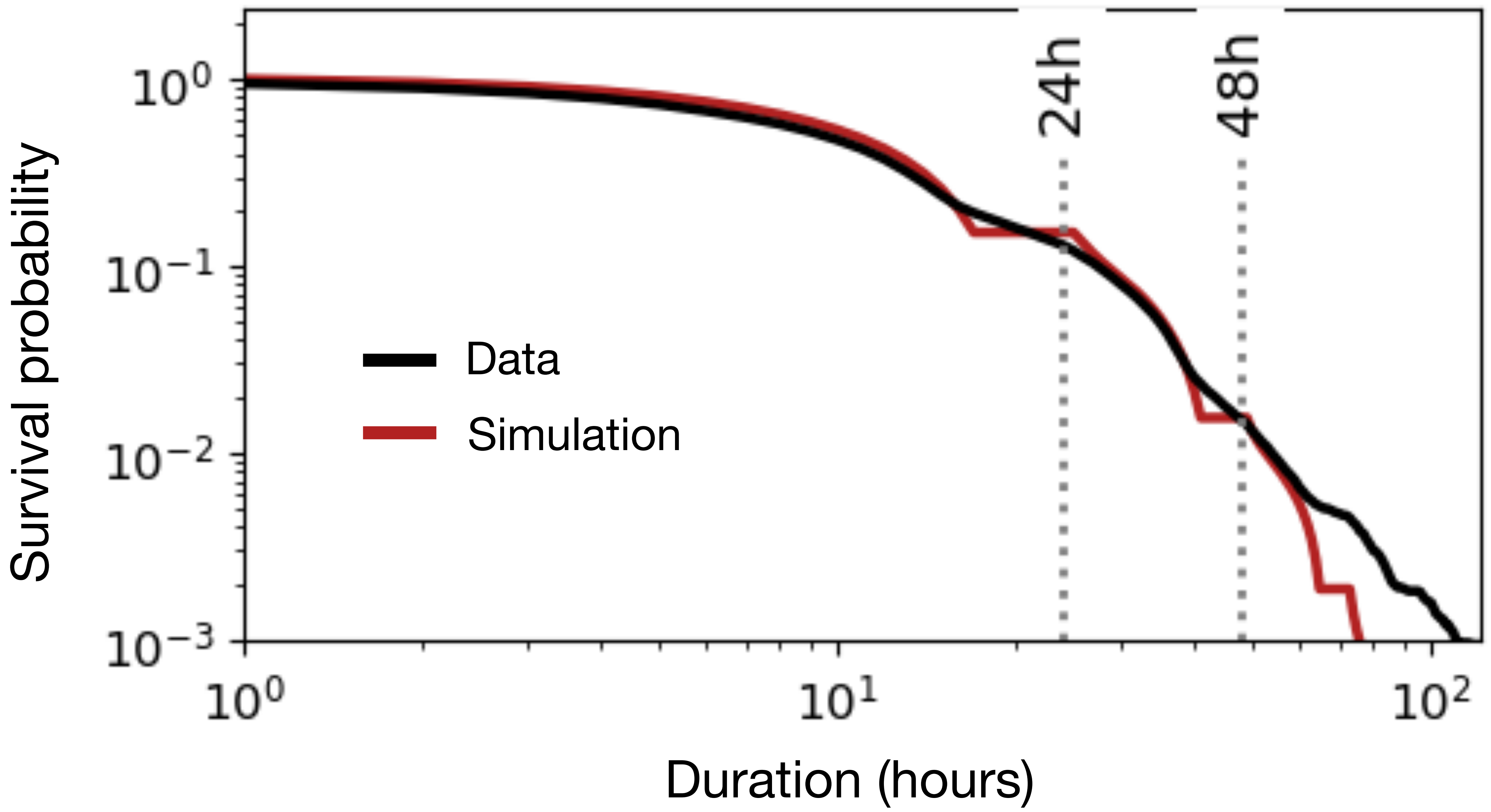
ONE HEALTH POULTRY HUB

ABM is informed by extensive field data

Region-to-market poultry flux
Where do chickens come from?



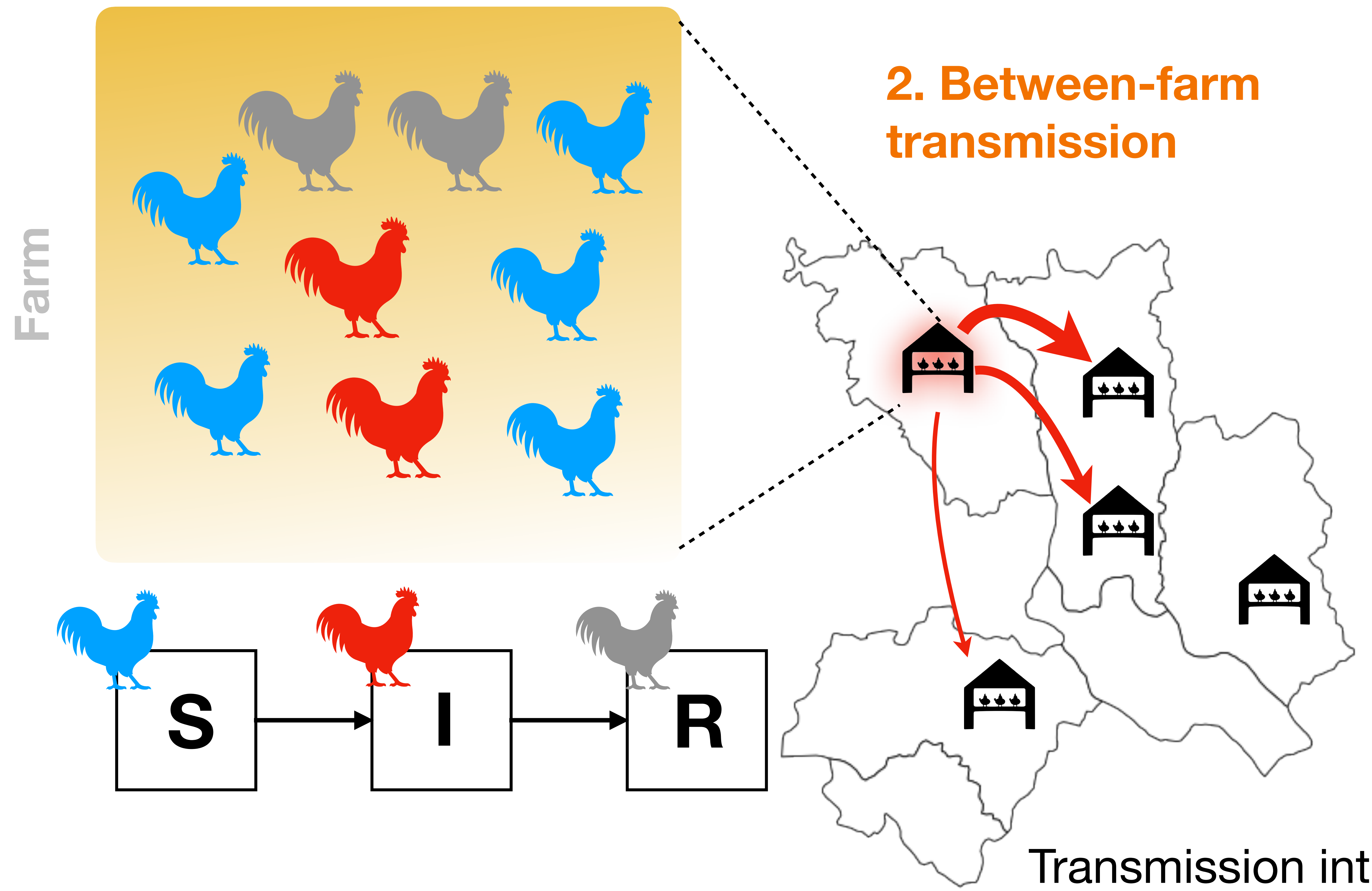
Length of stay in markets
How long before a chicken is sold?



Simulating Avian Influenza transmission

1. Within-flock transmission

2. Between-farm transmission



Transmission intensity decays with distance

Additional elements

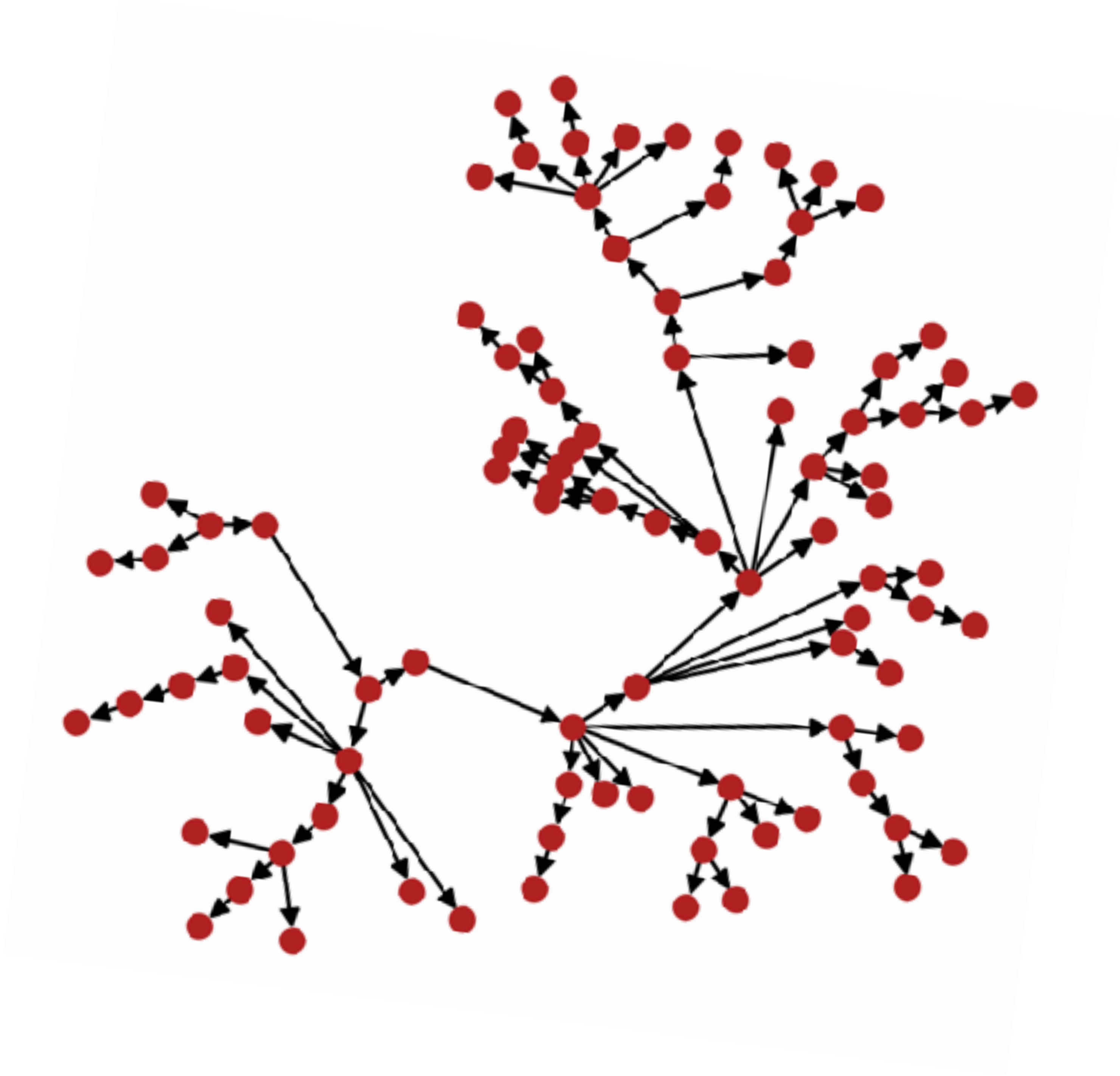
- Latent stage
- Age- and type-specific susceptibility
- Disease-induced mortality
- Multiple strains with interactions

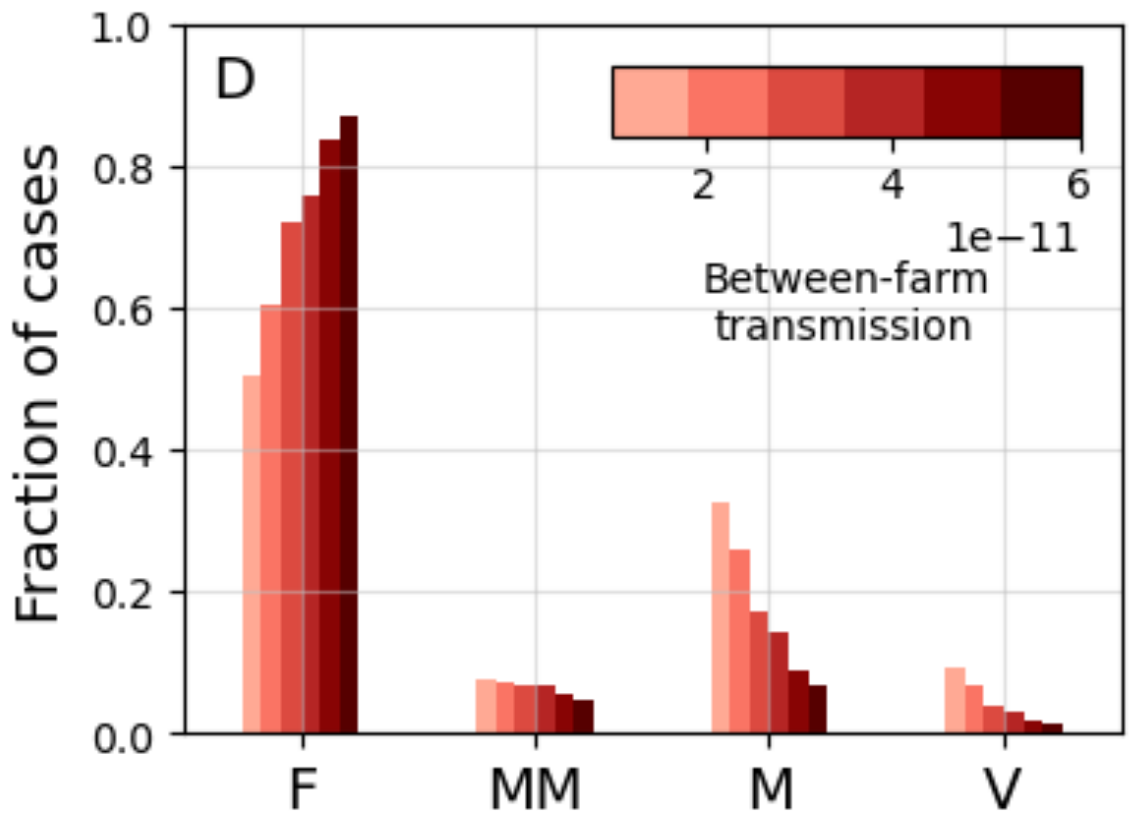
Simulating Avian Influenza transmission

ONE HEALTH POULTRY HUB

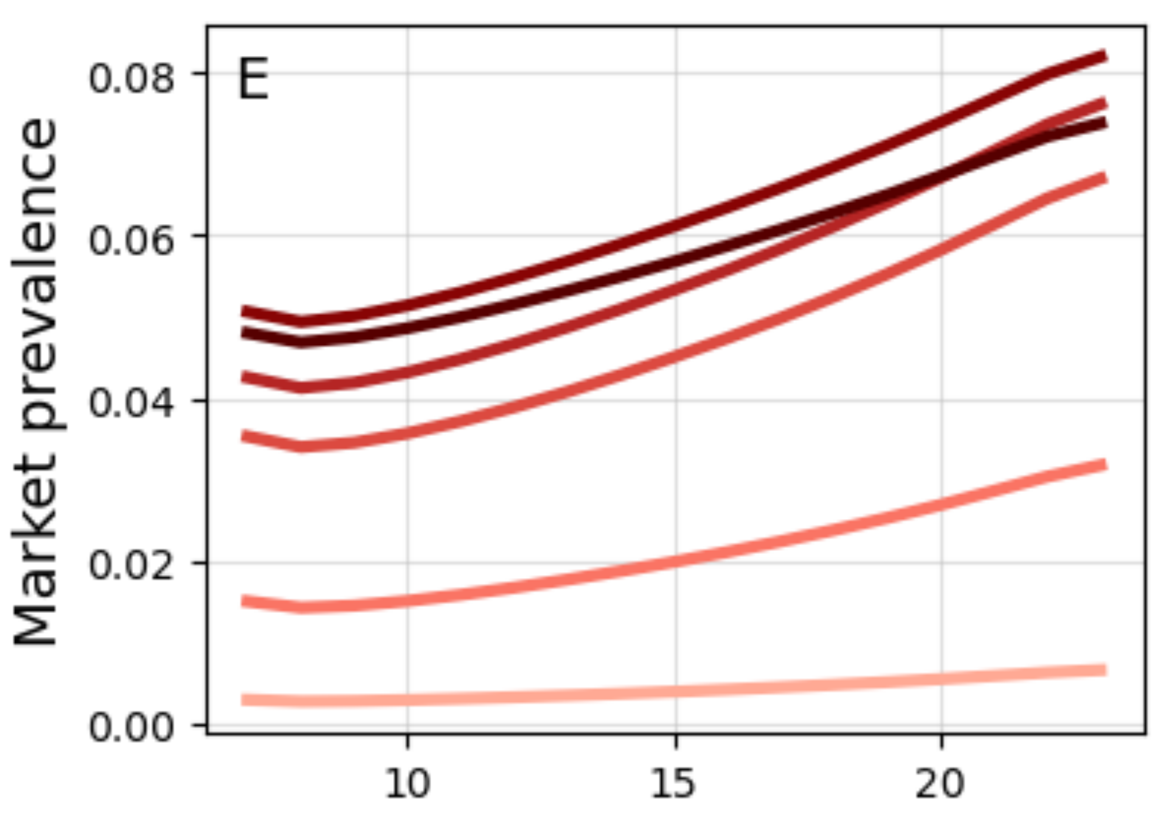
Transmission chain

Which chicken infected which chicken?

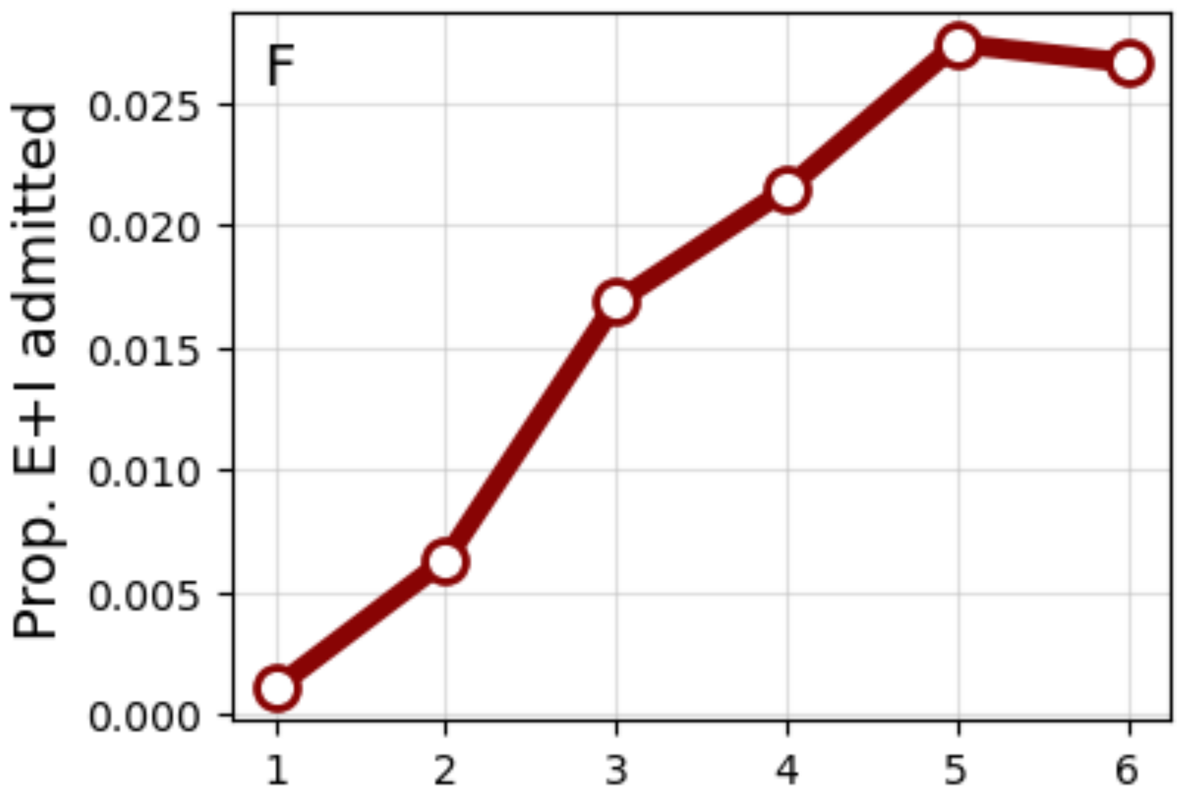




Infection location

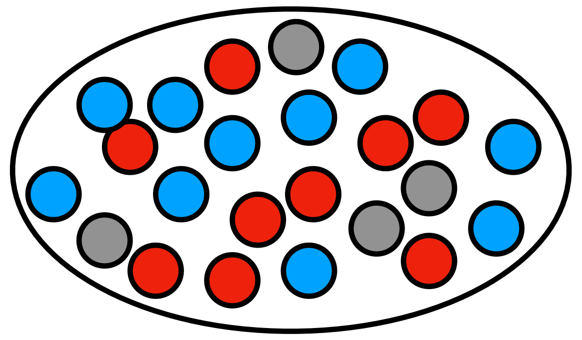


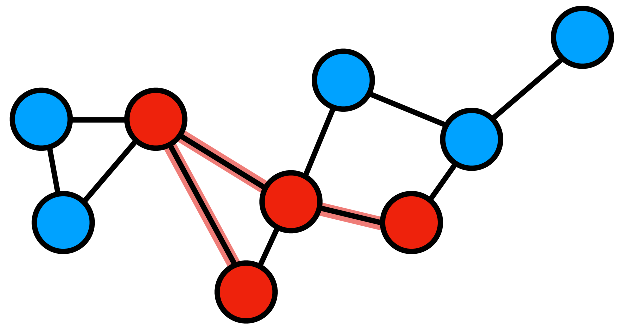
Time of day (hour)

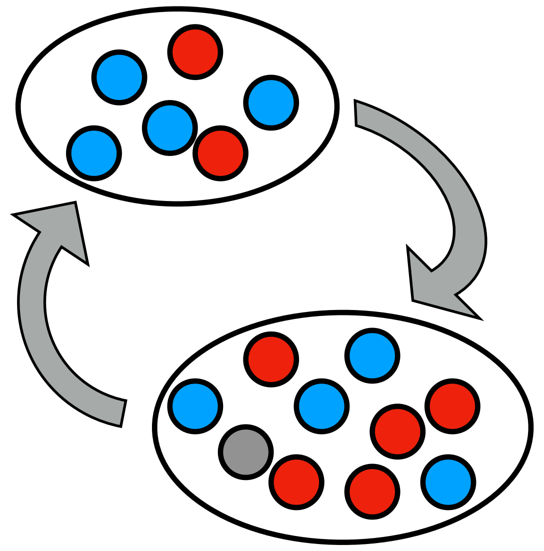


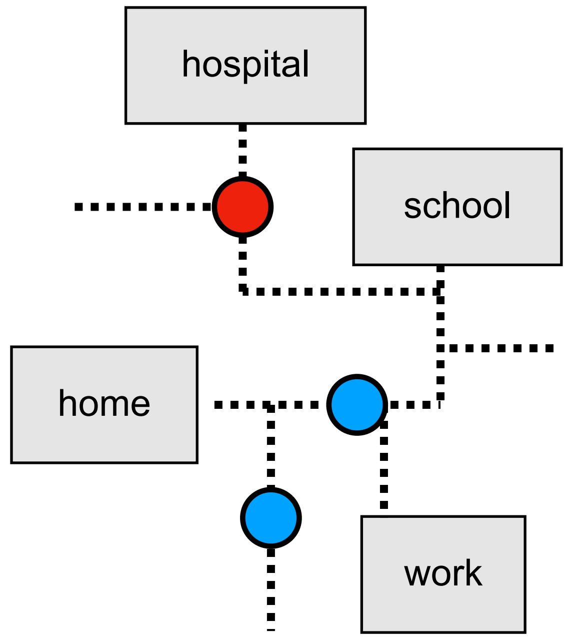
Between-farm transmission rate

Model complexity


Homogeneous mixing


Network

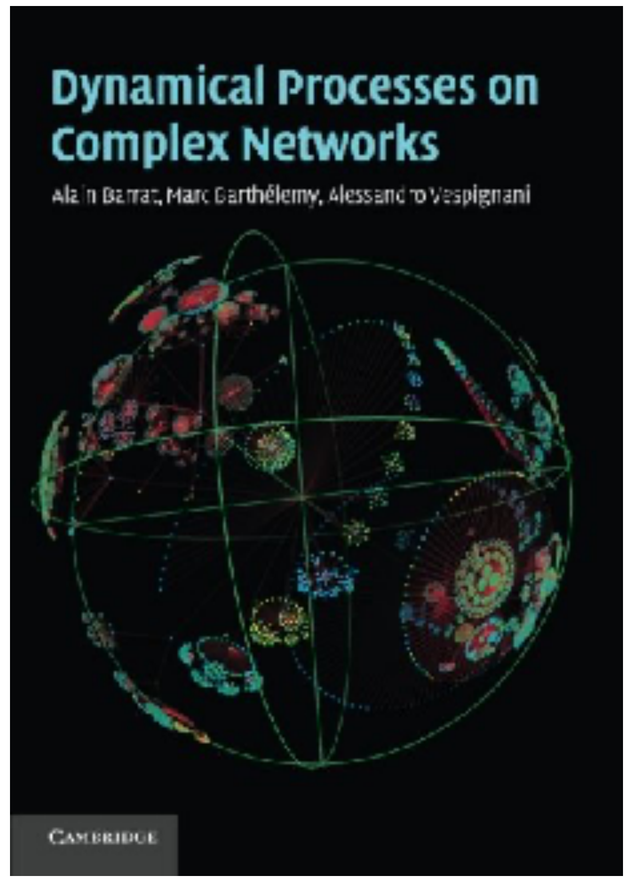

Metapopulation

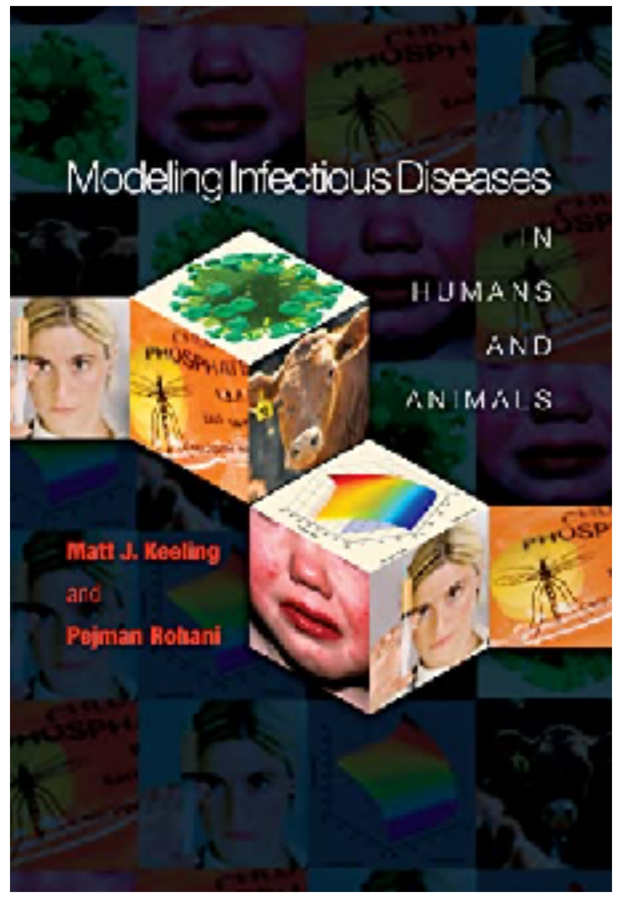


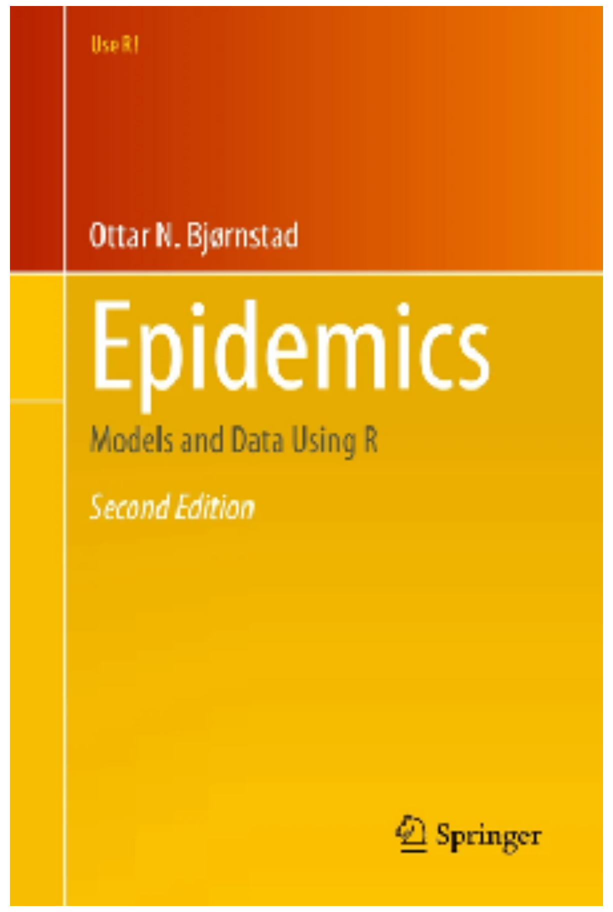
Agent-based

Resources

- (Paper) R. Pastor Satorras et Al., Epidemic processes in complex networks, Rev. Mod. Phys., 2015
- (Book) A. Barrat et Al., Dynamical Processes on Complex Networks, Cambridge University Press, 2012
- (Book) M. J. Keeling & P. Rohani, Modeling Infectious Diseases in Humans and Animals, Princeton University Press, 2007
- (Book) O. N. Bjørnstad, Epidemics: Models and Data Using R, Springer 2023







Thank you!

francesco-pinotti.netlify.app

@fpinotti92

francesco.pinotti@biology.ox.ac.uk

francescopinotti92

UNIVERSITY OF OXFORD

ONE HEALTH POULTRY HUB

Credits: Chiara Poletto