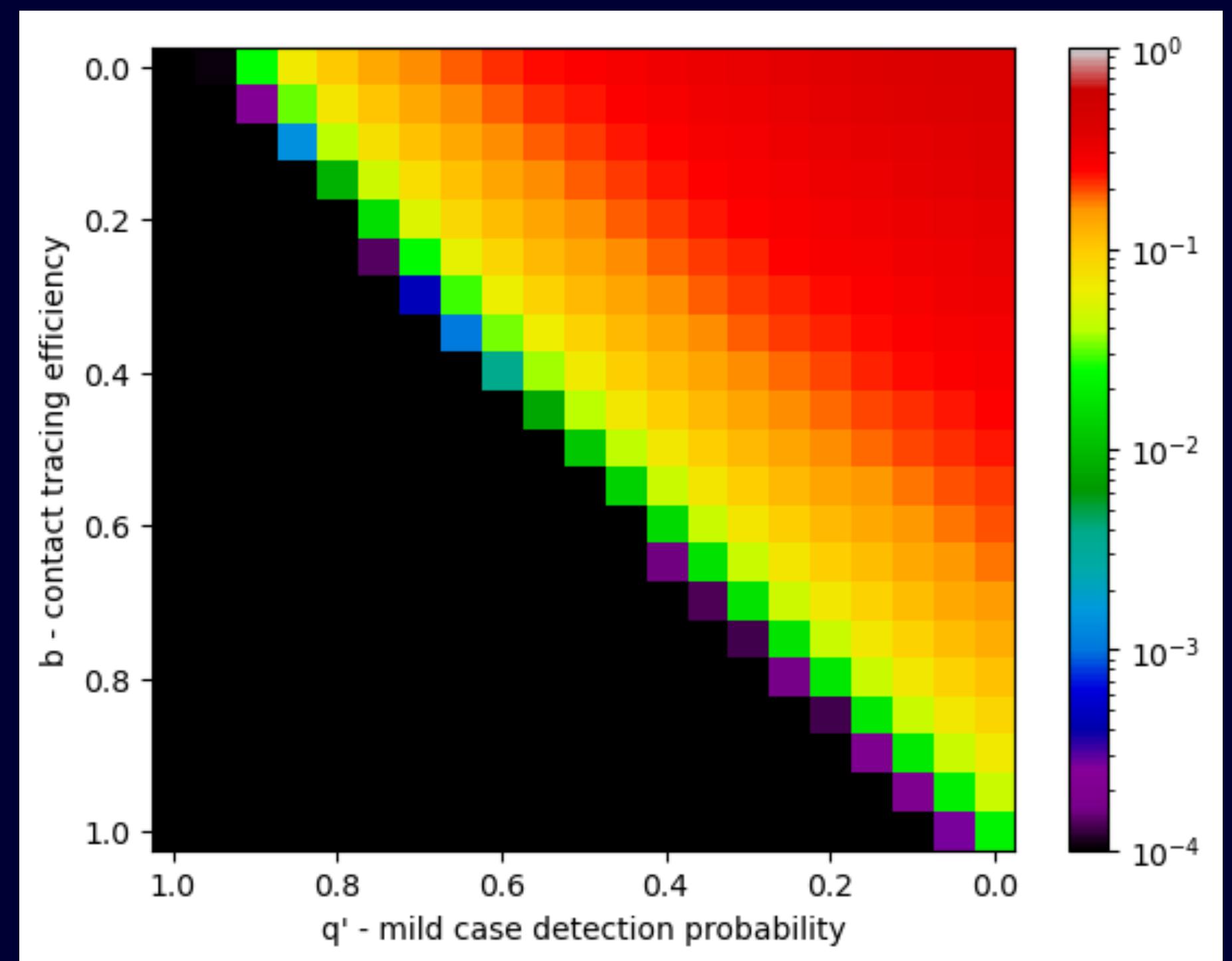
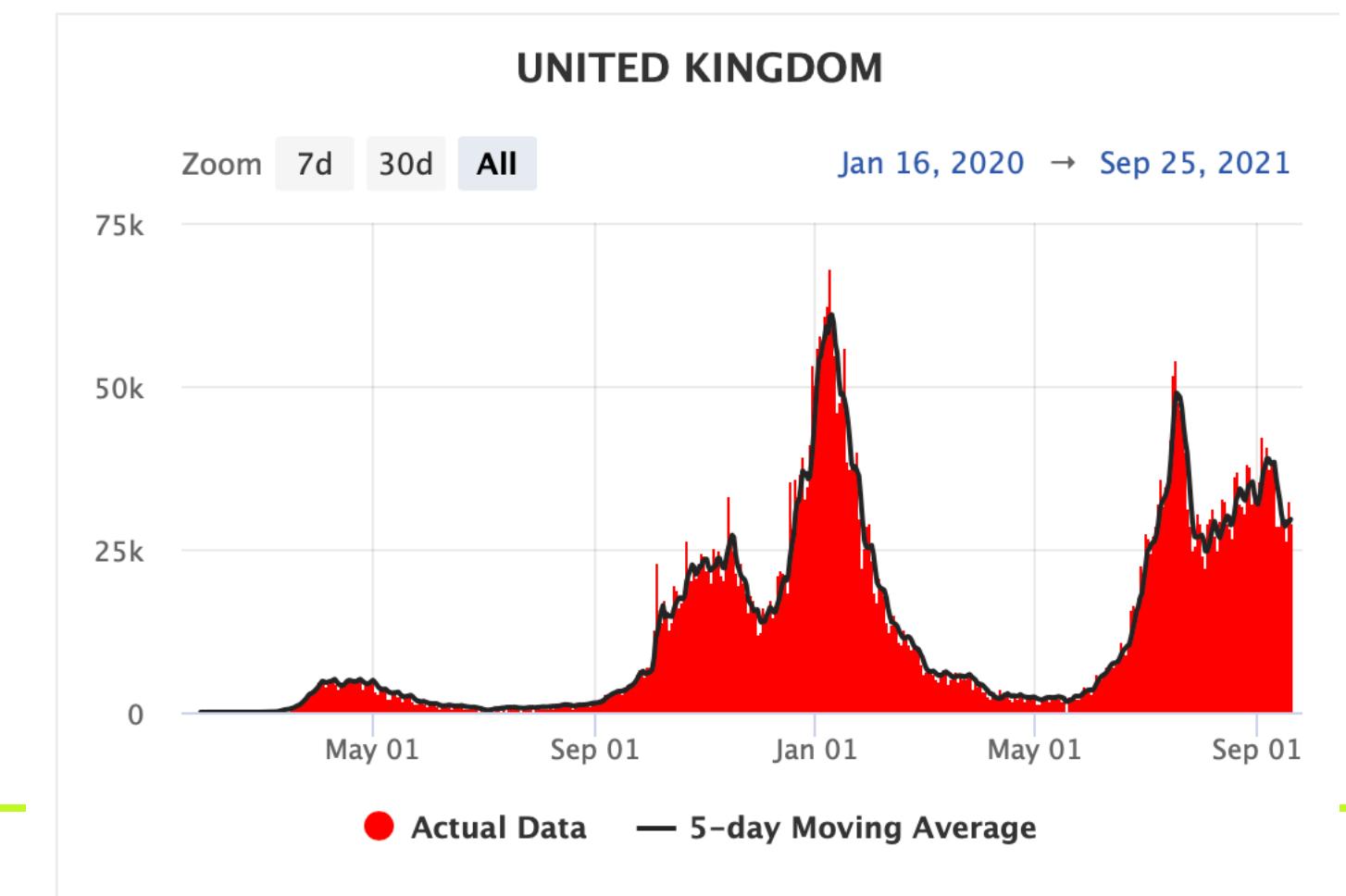
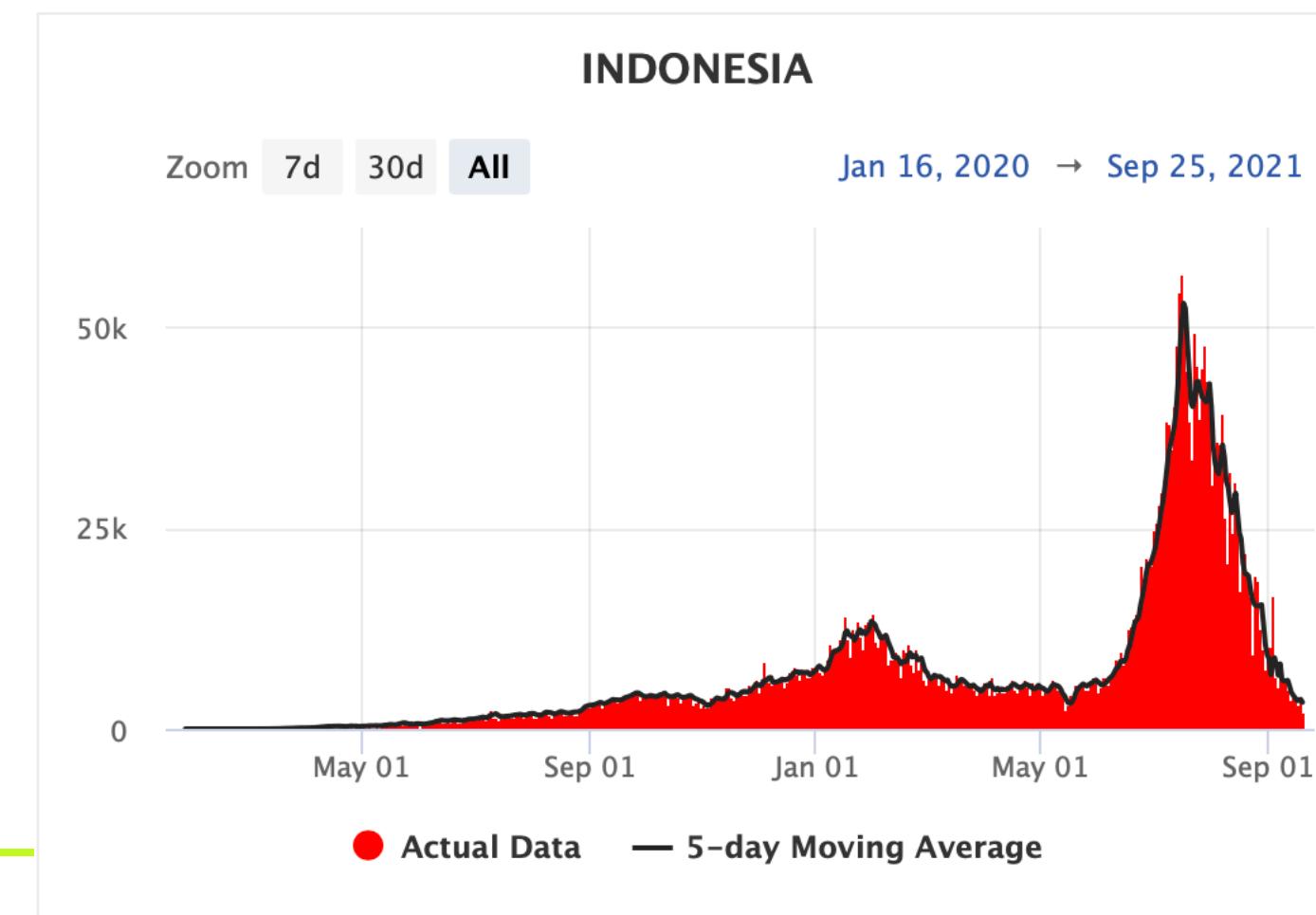
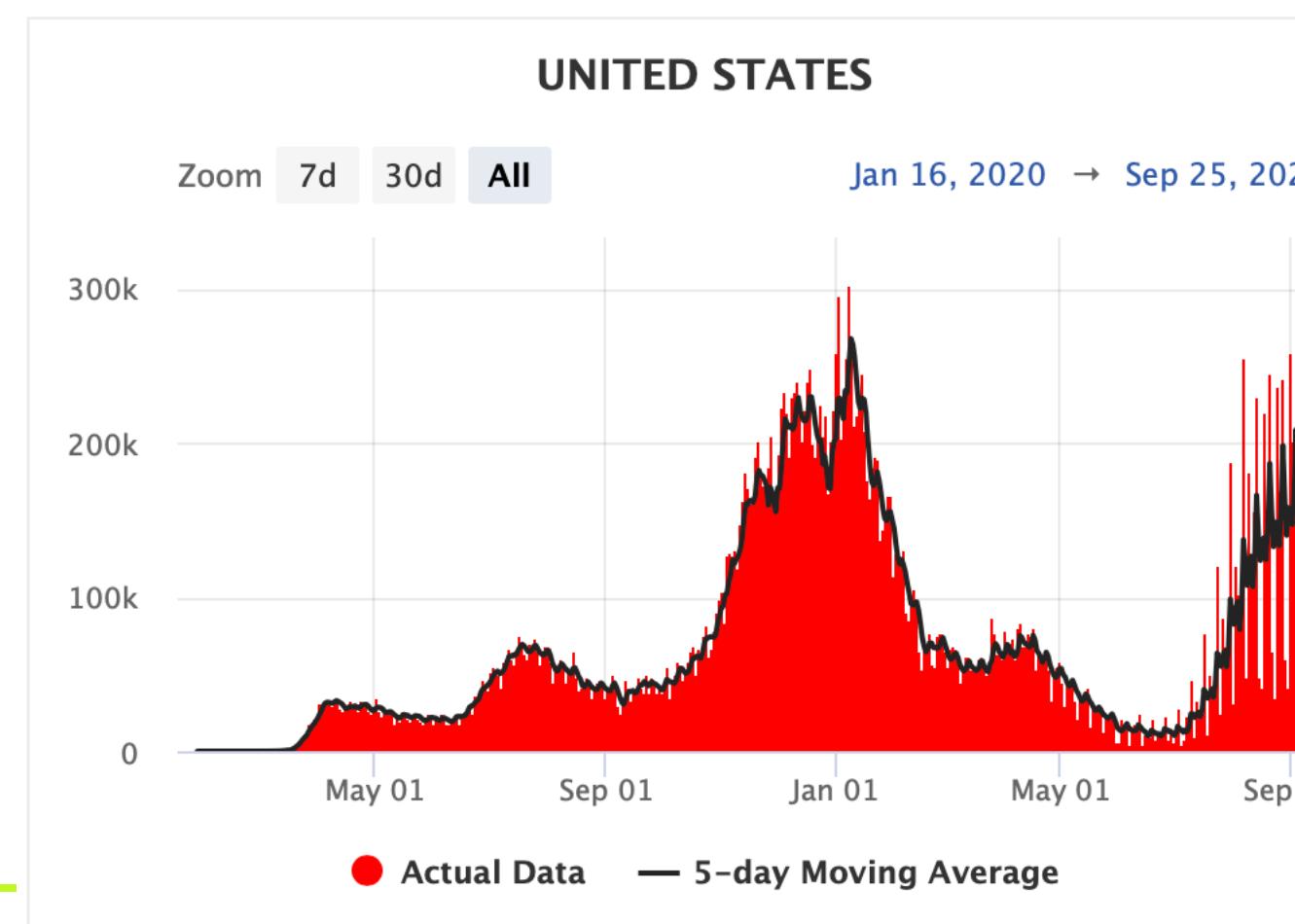
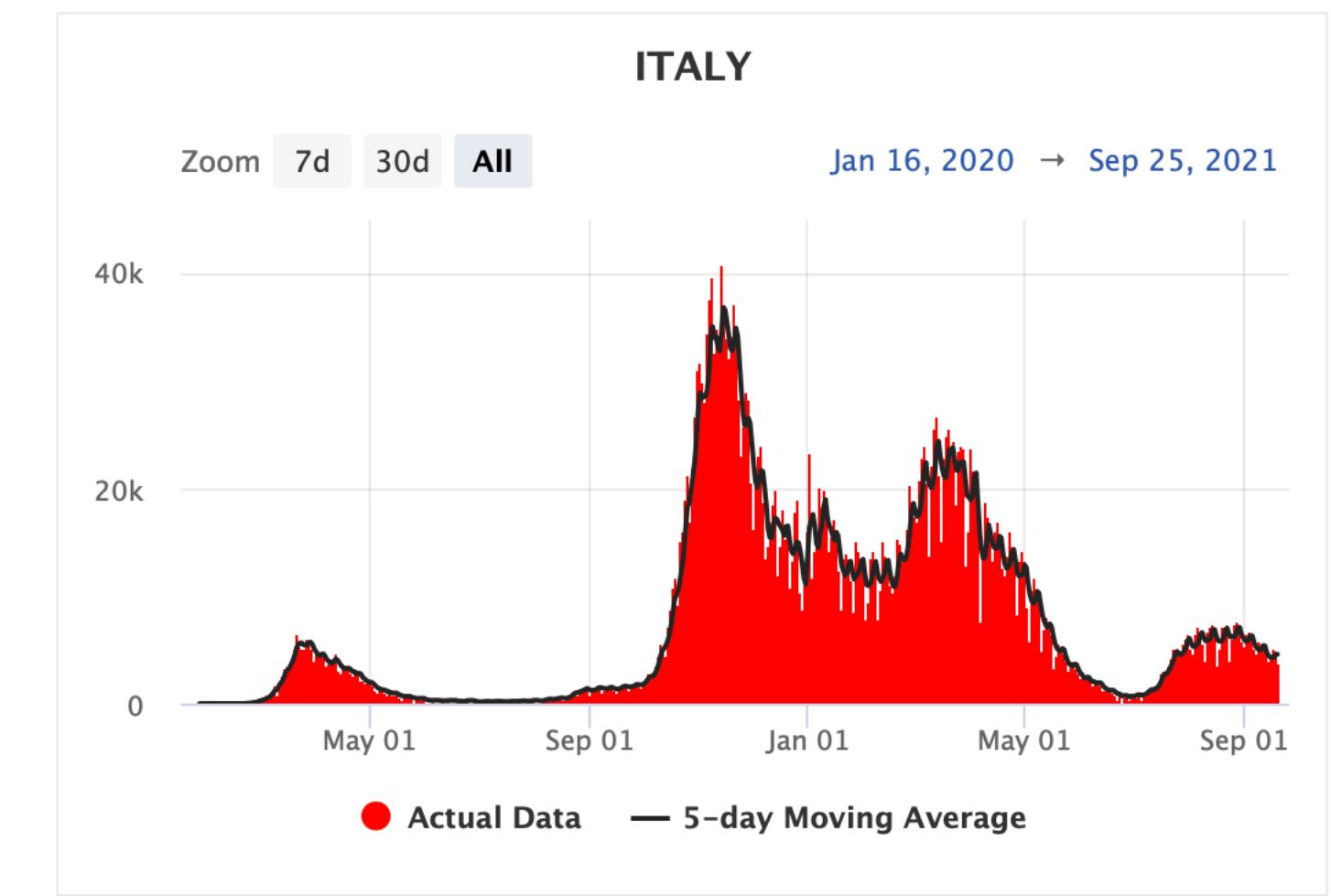
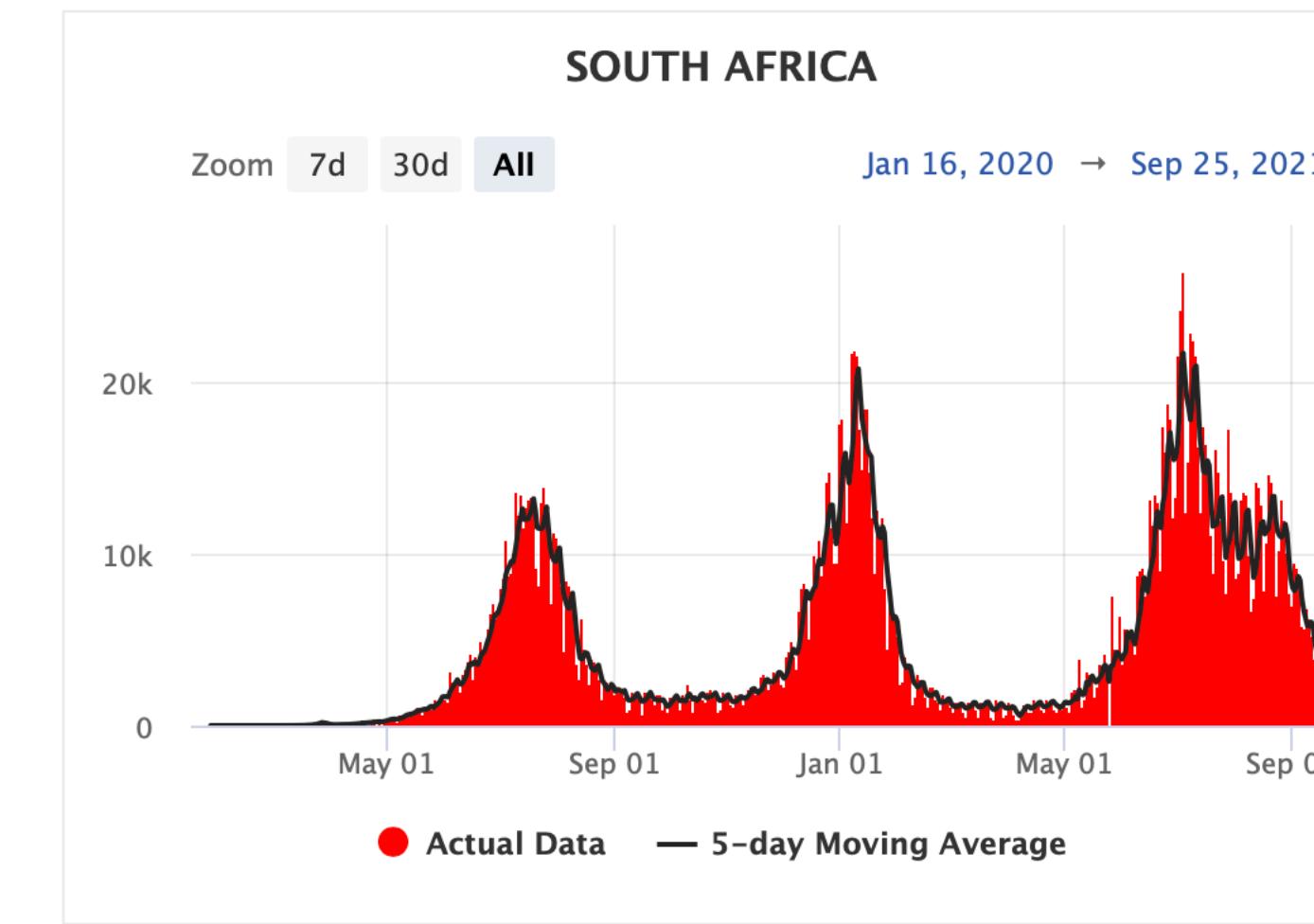
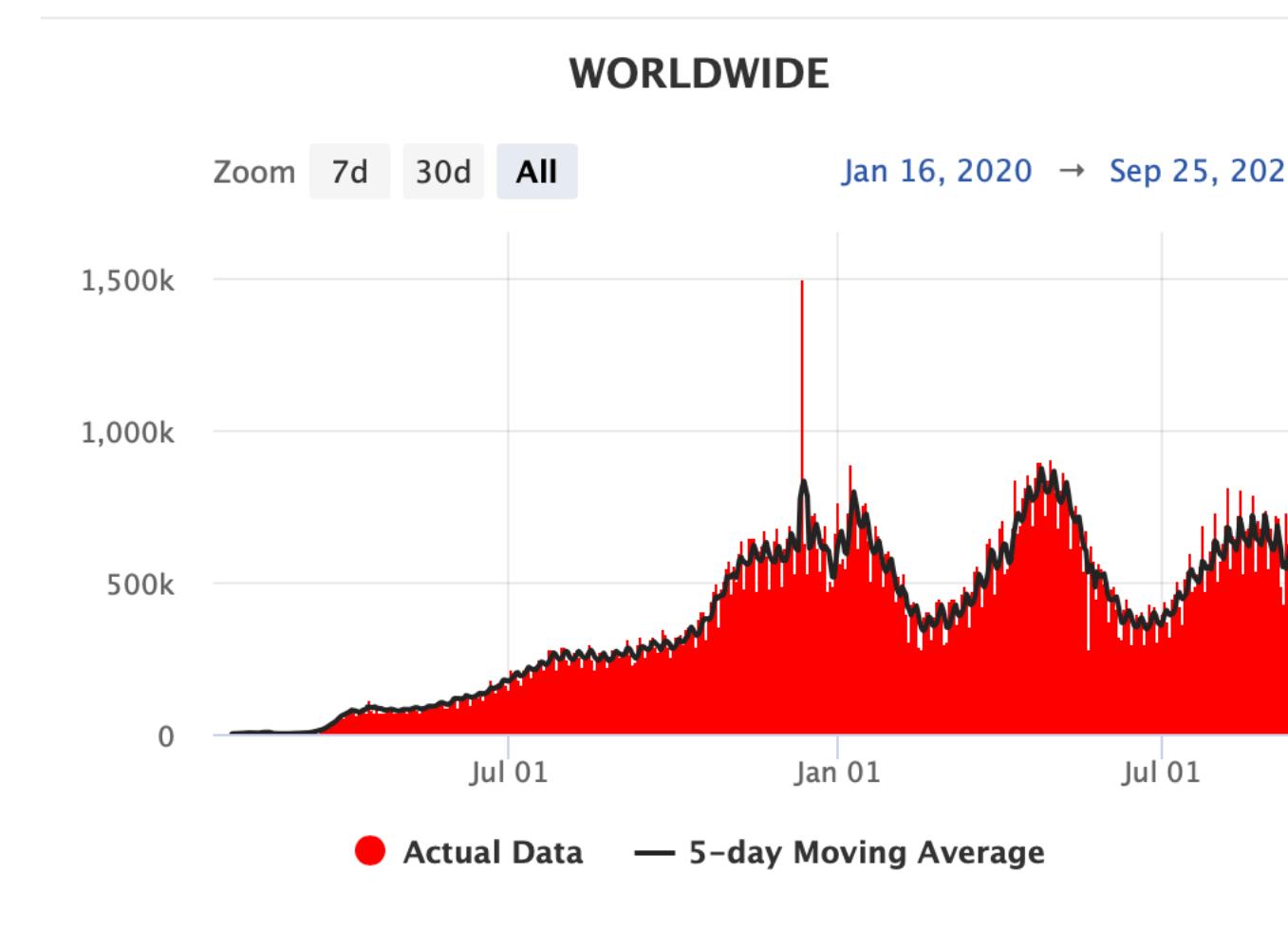


Predicting the unpredictable : mathematical models and the COVID 19 pandemic

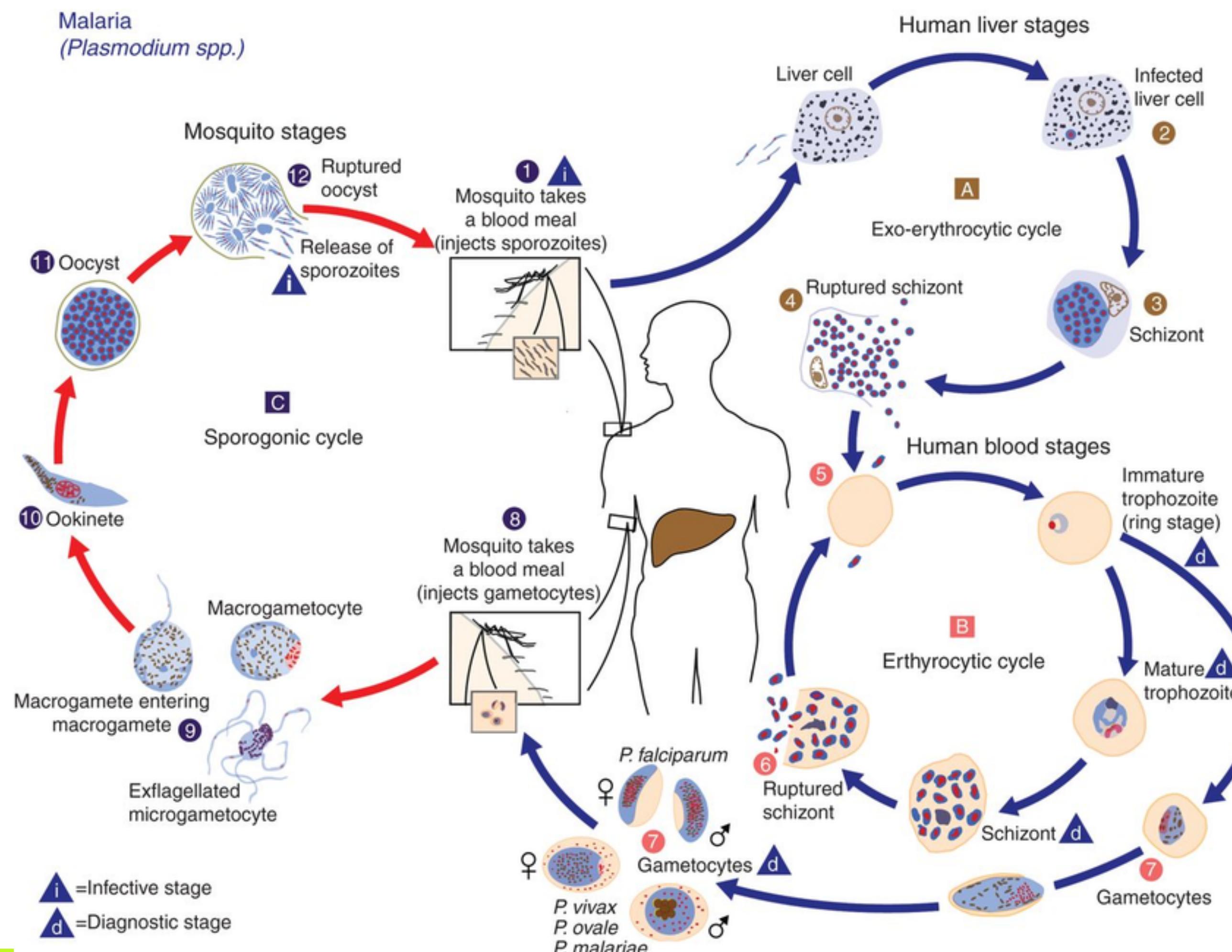
“It is very hard to predict, especially the future”
Niels Bohr (rephrasing an old danish proverb)



What should be the focus of predictions?



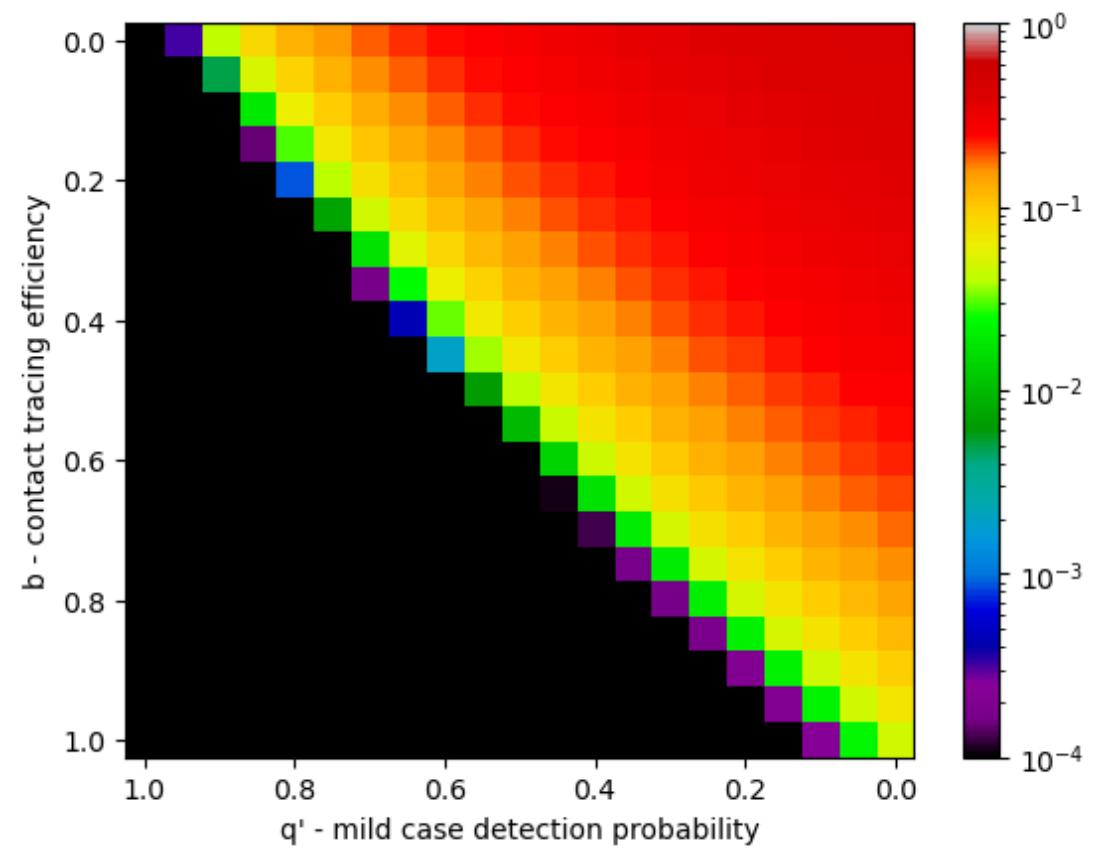
First mathematical models for infectious diseases : Malaria (Ross 1911 and Lotka 1923)



what is the most effective way to control malaria?

Different prediction categories of mathematical models

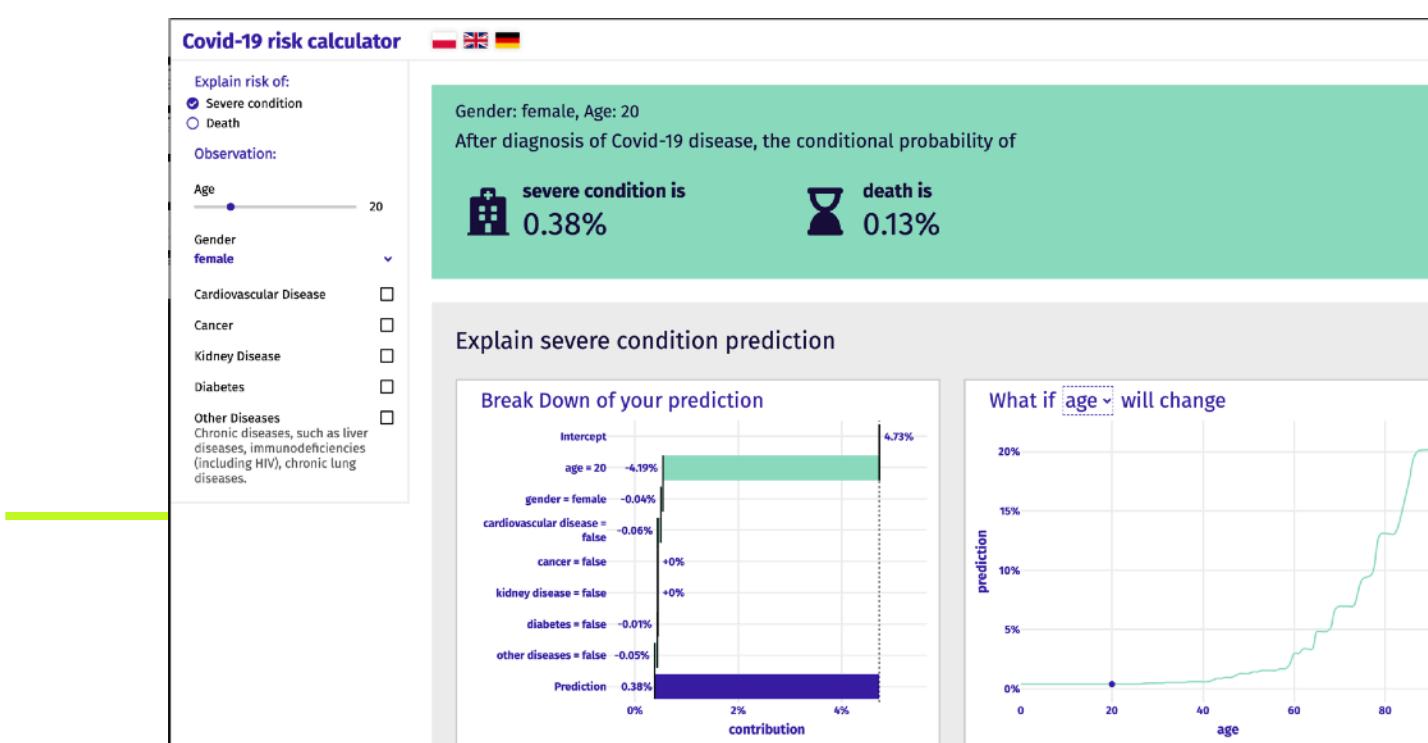
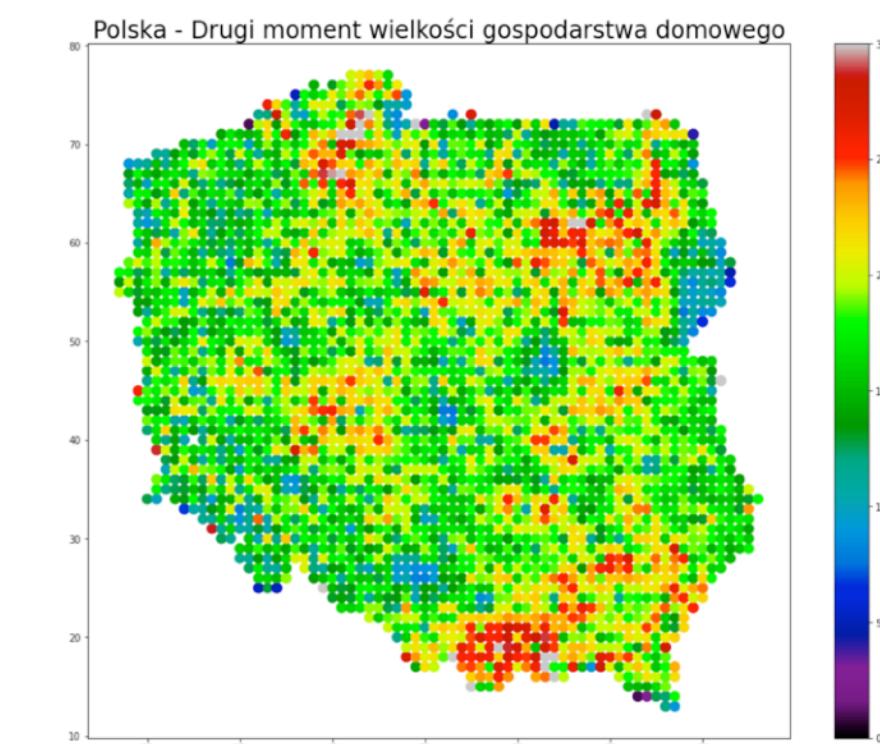
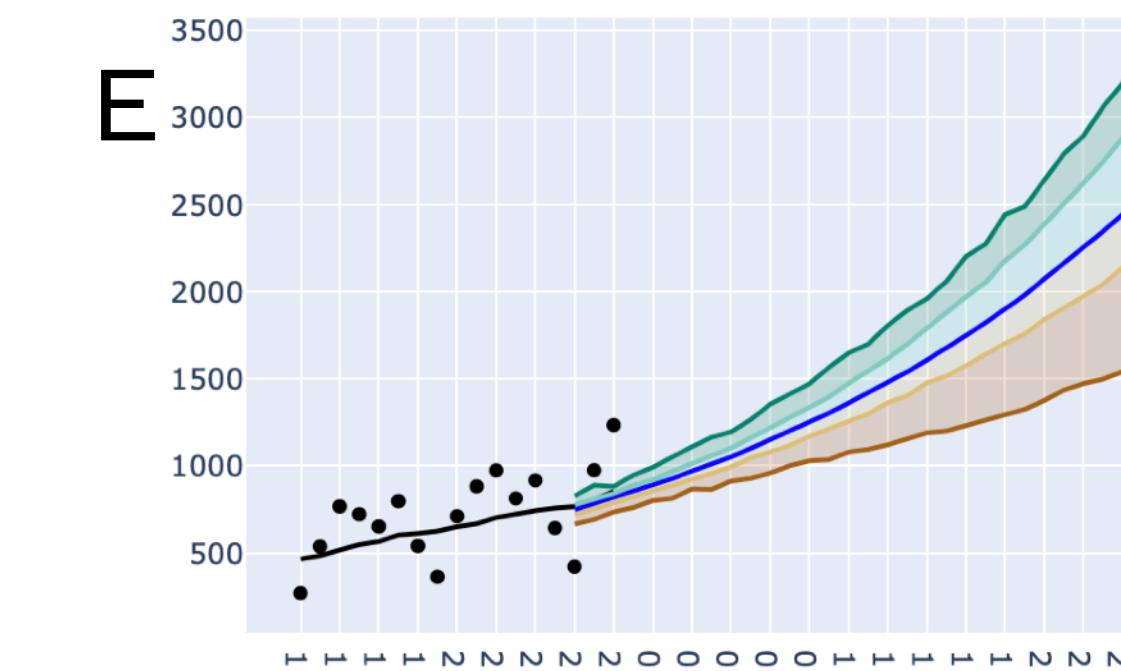
A. Forecasts long and short (like weather) : difficult to predict when epidemic dynamic changes



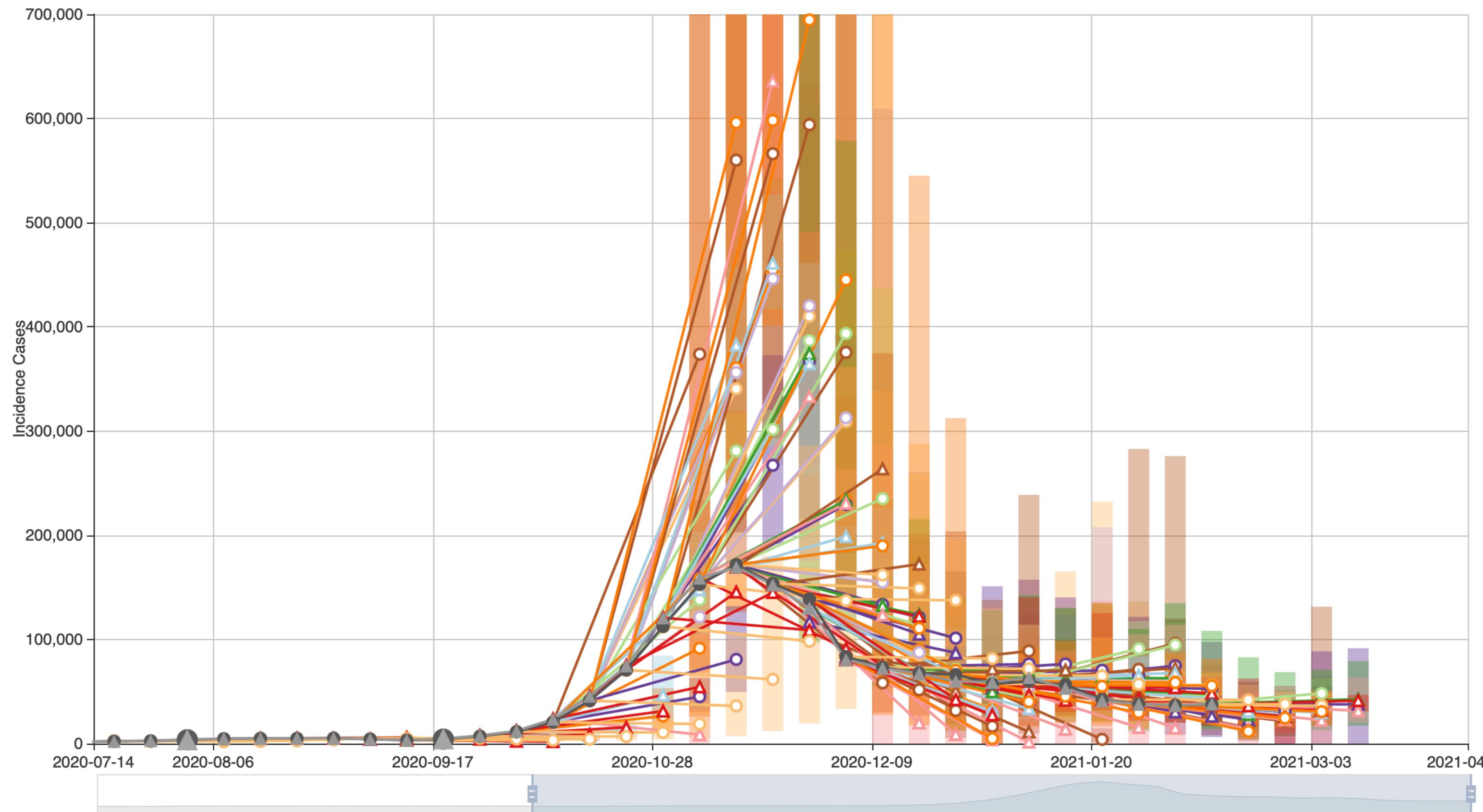
B. Analysis of the effect of countermeasures : especially contact reduction, testing level and speed , contact tracing efficiency (NPI's)

C. Understanding structural effects e.g the impact of households or mobility(workplace)

D. Patient outcome



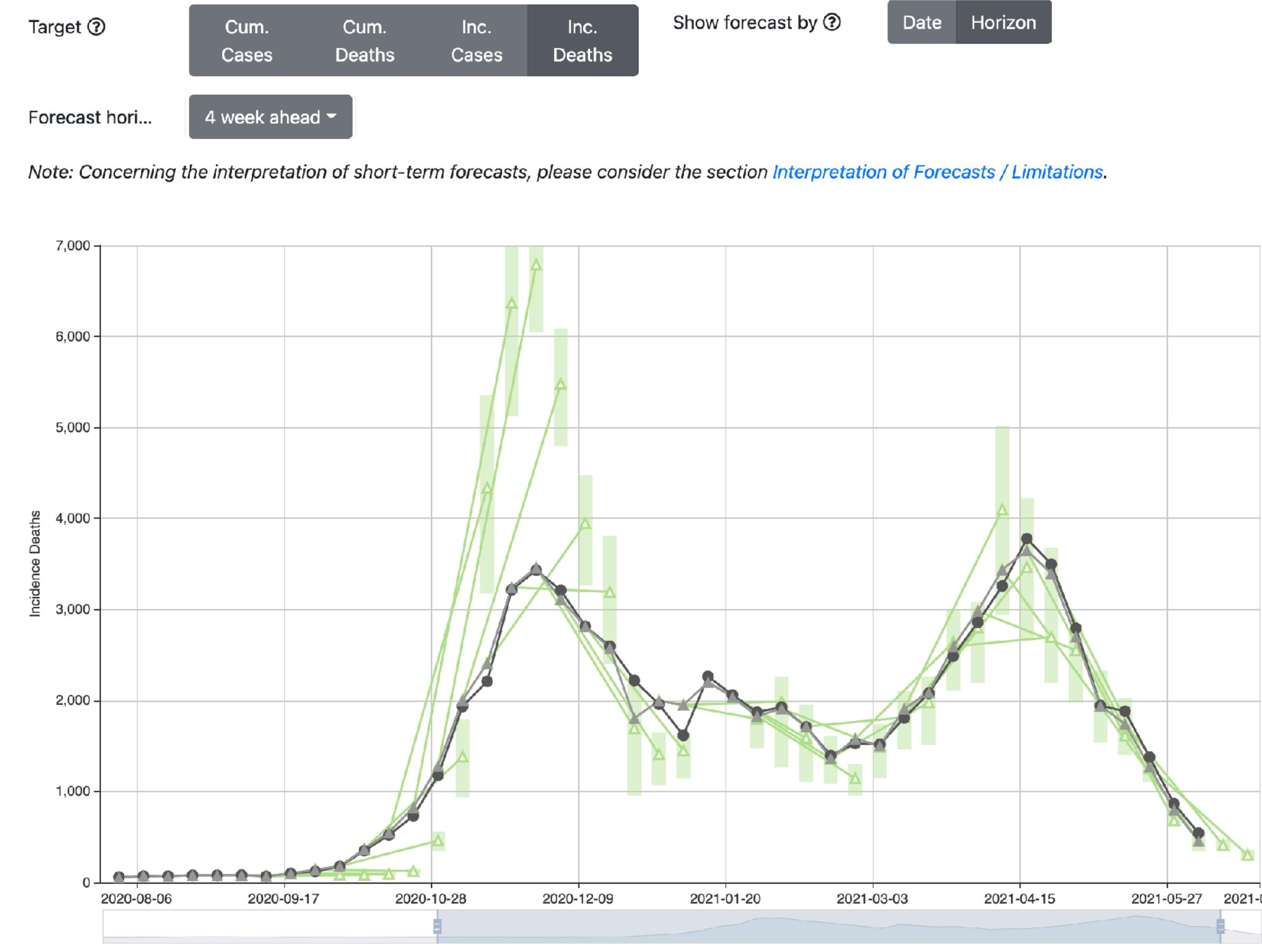
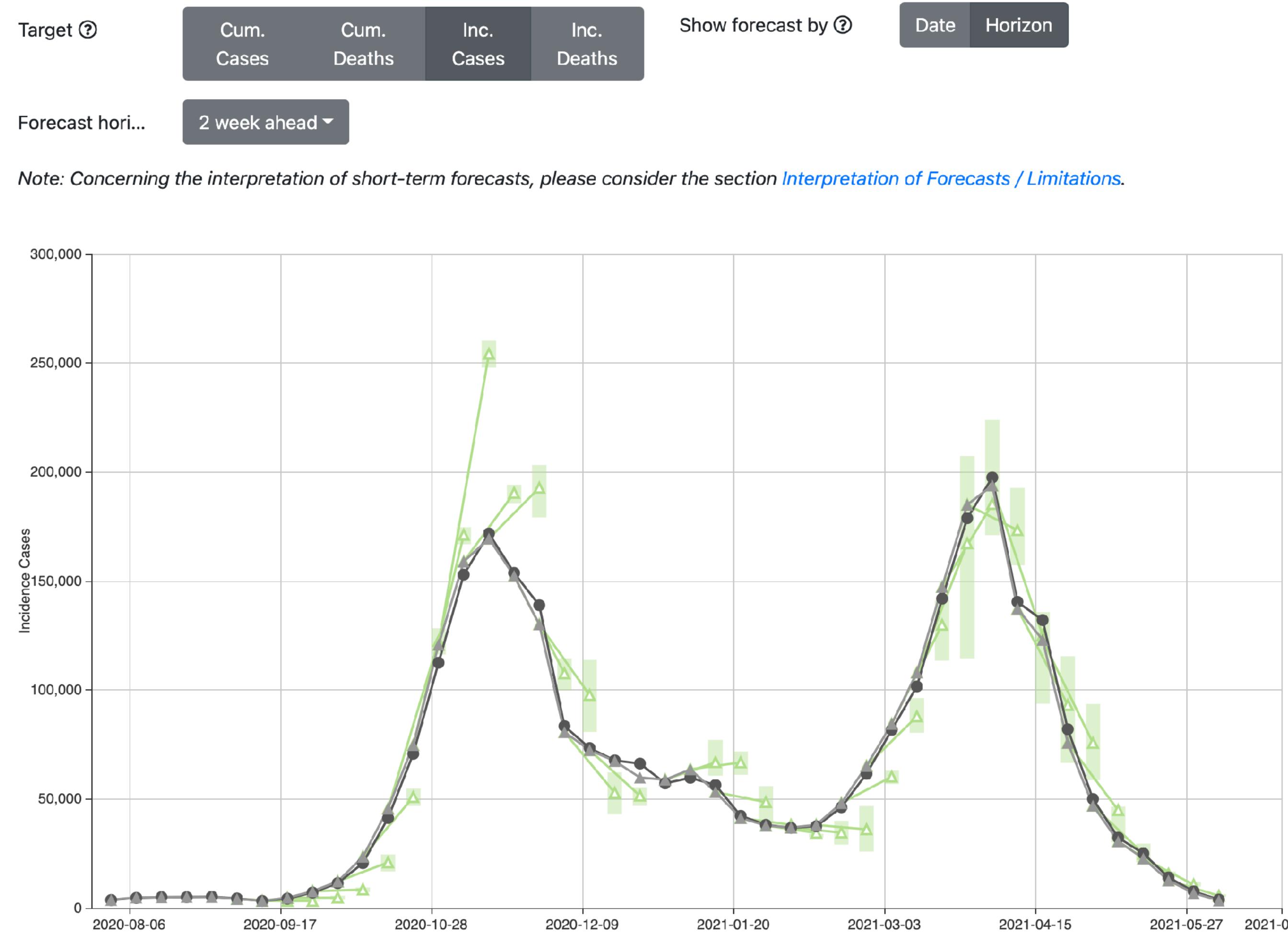
Predicting the timeline of epidemic is difficult: The German-Polish/European -COVID 19 forecast - hub, (US forecast hub)



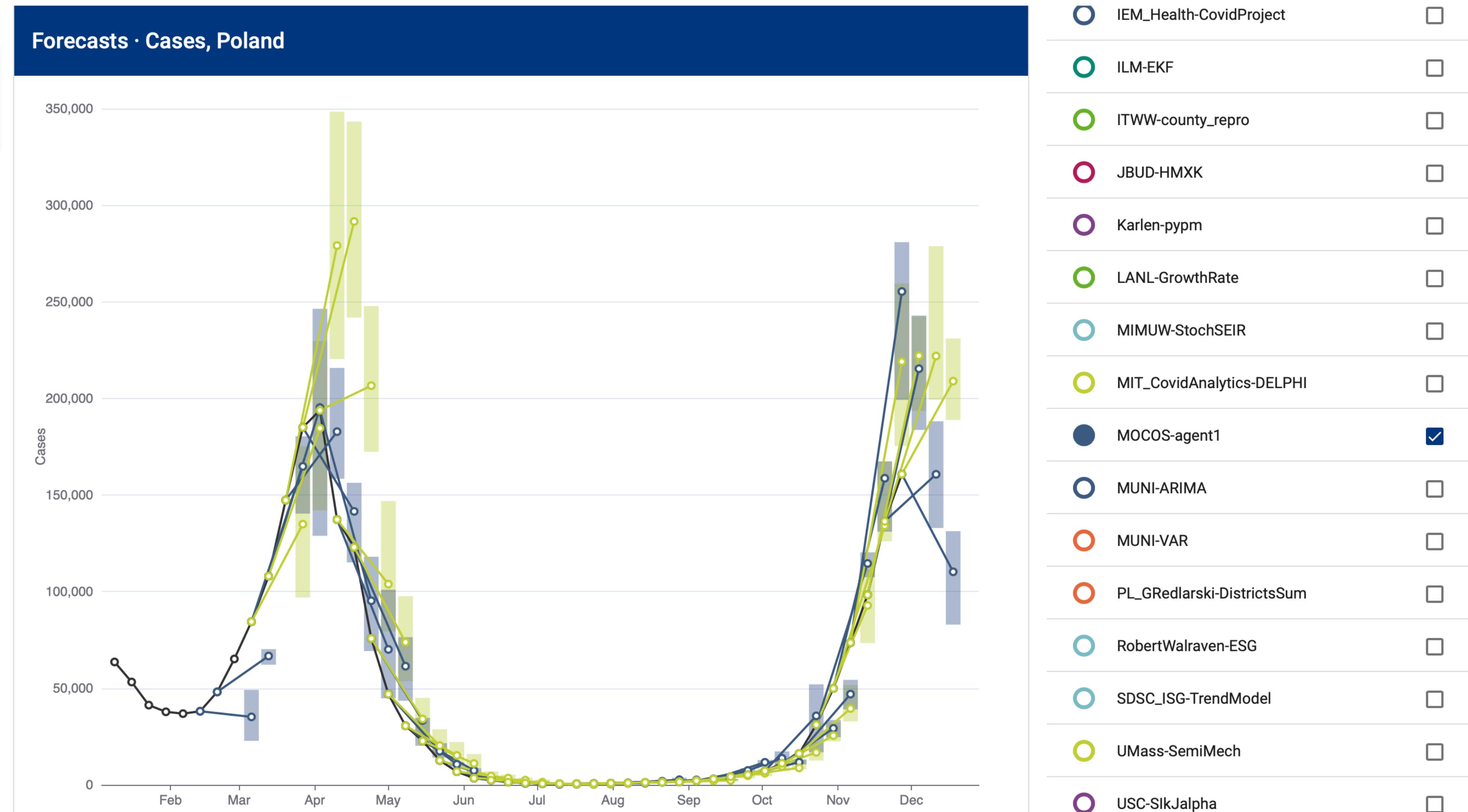
- About 40 teams try to predict the number of deaths and infected within next month for Poland and Germany
- Model types : a) statistical , b) compartment (e.g. ODE) c) individual based
- Baseline prediction : number/ trend stays as now + noise
- Ensemble prediction is better than most of individual models

Accuracy of predictions depend on quality of data and modelling skills (and luck)

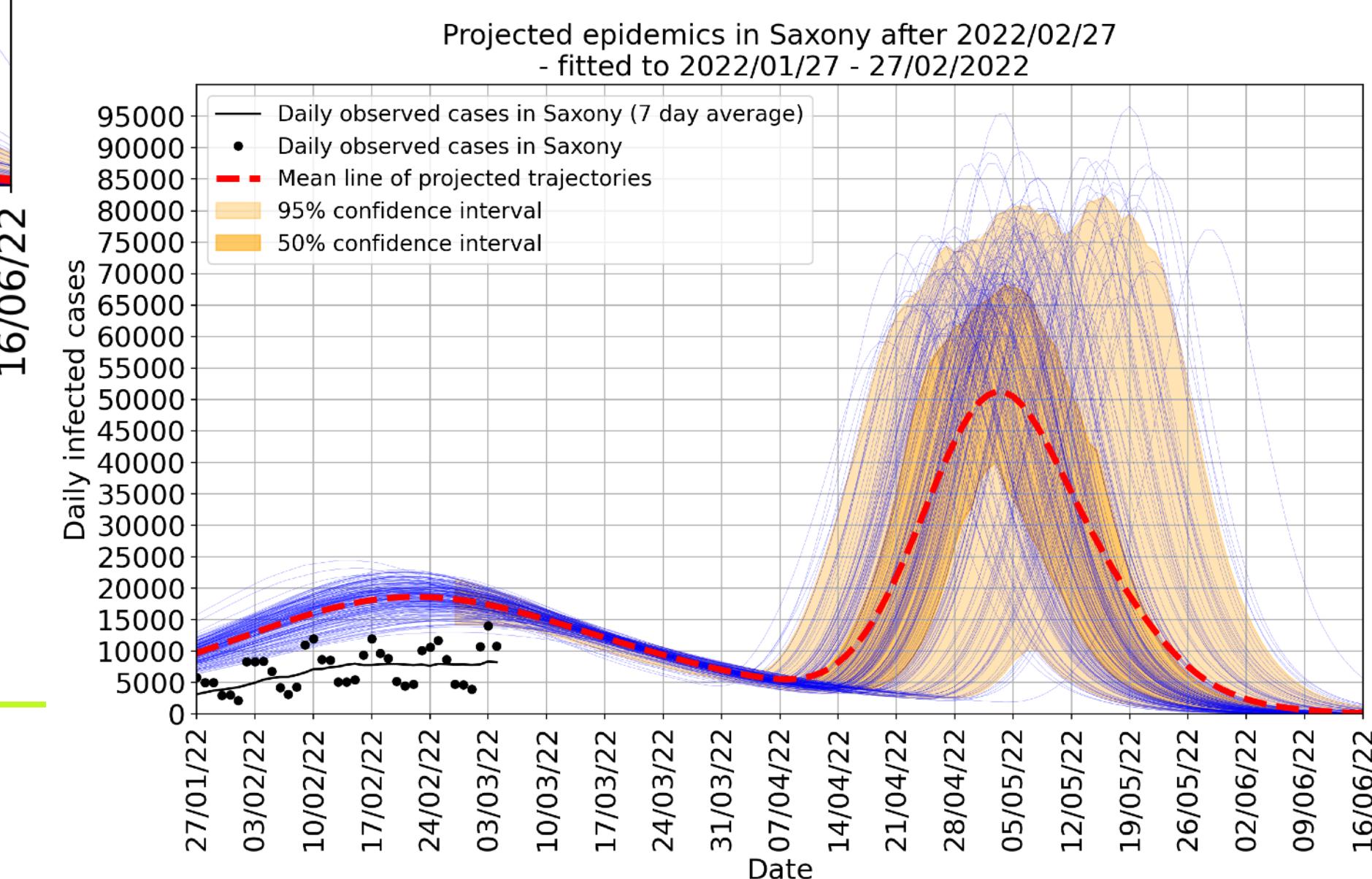
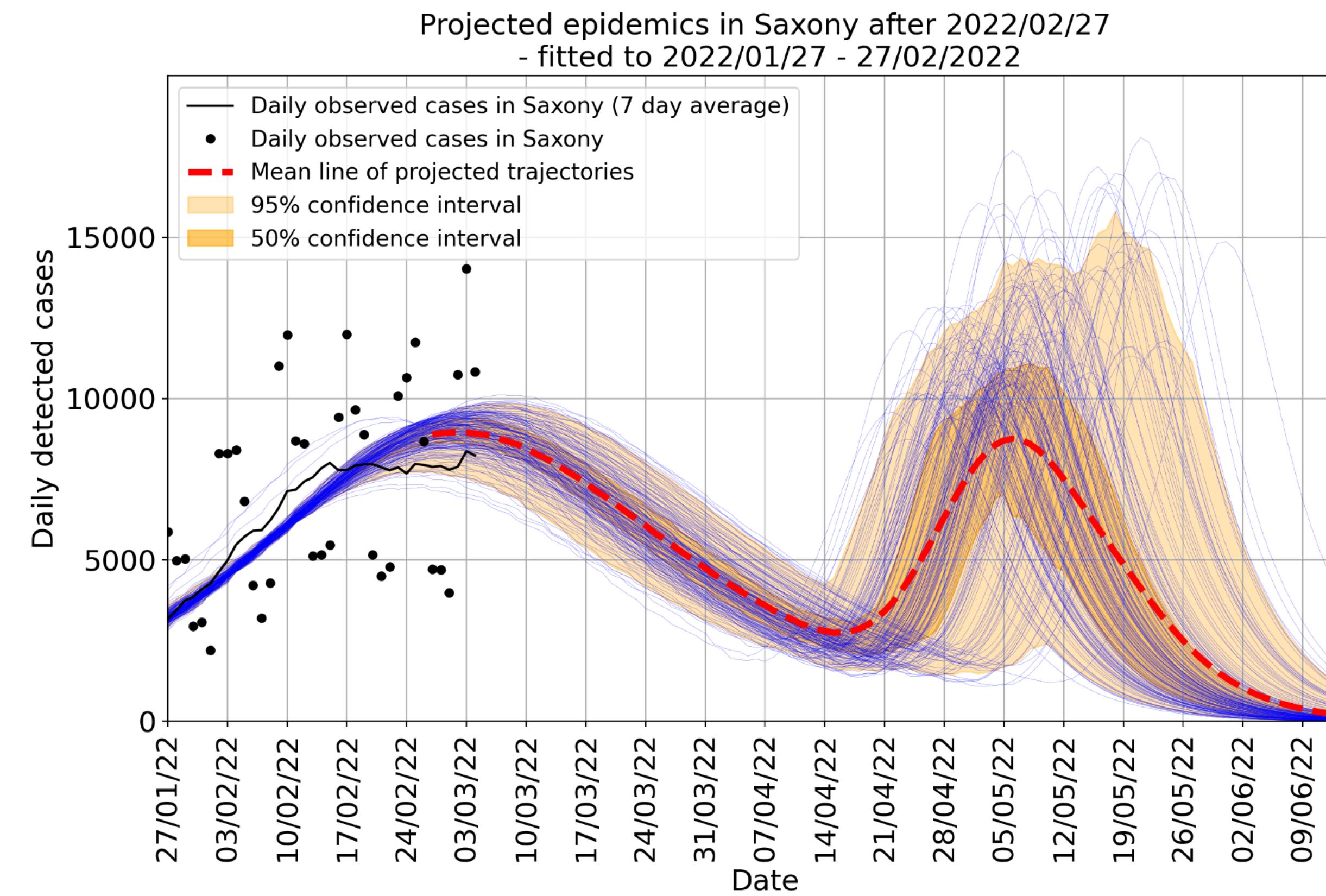
MOCOS predictions Poland during 2. and 3. Wave



MOCOS and ensemble predictions Poland (3rd and 4th wave)



Ensembles of trajectories (MOCOS forecast Saxony)



Types of mathematical models

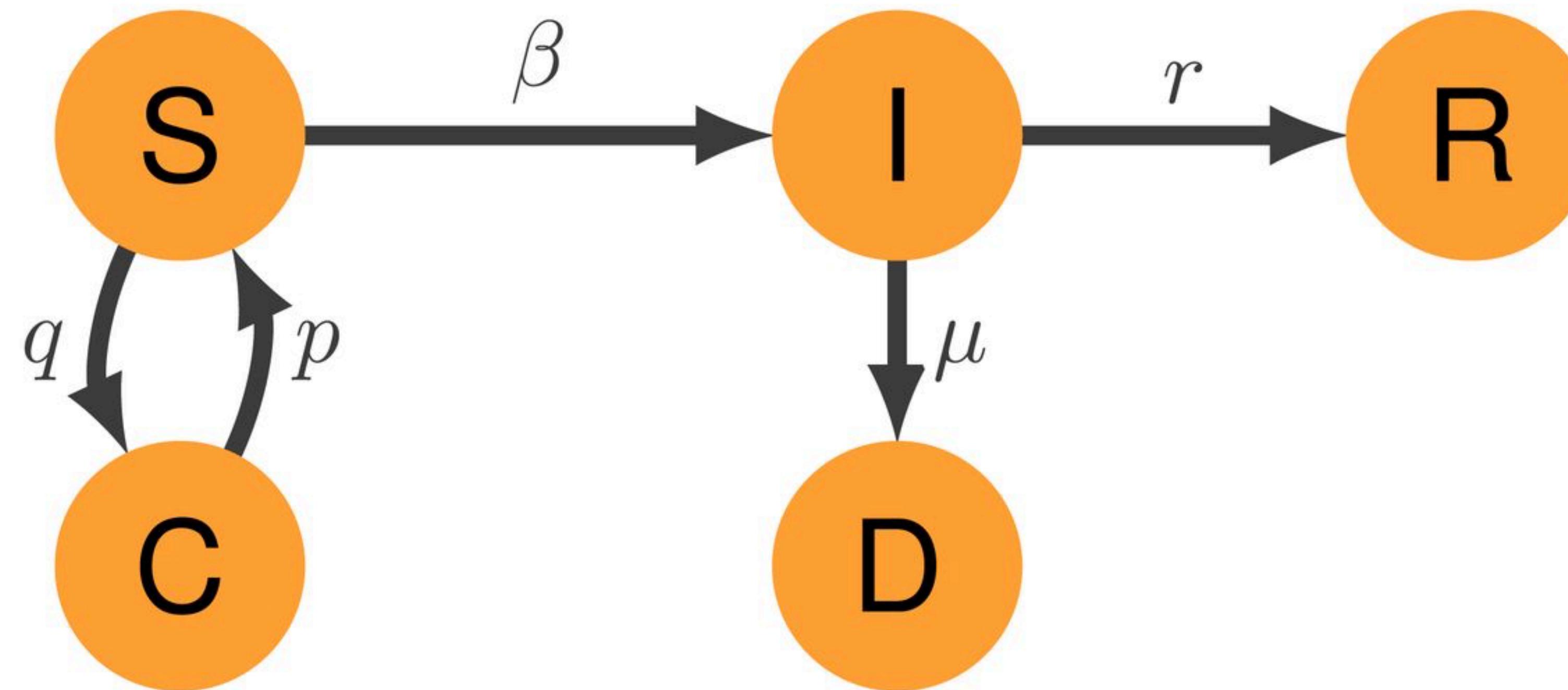
- For most infectious diseases there are
 - ODE models and stochastic ODE models
 - PDE and SPDE models (e.g. influenza, malaria)
 - Agent based models / discrete stochastic models on networks (e.g. sexually transmitted diseases like HIV, influenza, COVID 19)
- aim:
 - uncovering conditions for outbreaks (epidemic thresholds)
 - how to control the spread of a disease - sensitive parameters
 - prediction of incidence and prevalence

Prototype ODE model

$$\dot{S} = -\frac{\beta S I}{N} - qS + pC$$

$$\dot{I} = \frac{\beta S I}{N} - (r + \mu)I$$

$$\dot{R} = rI$$



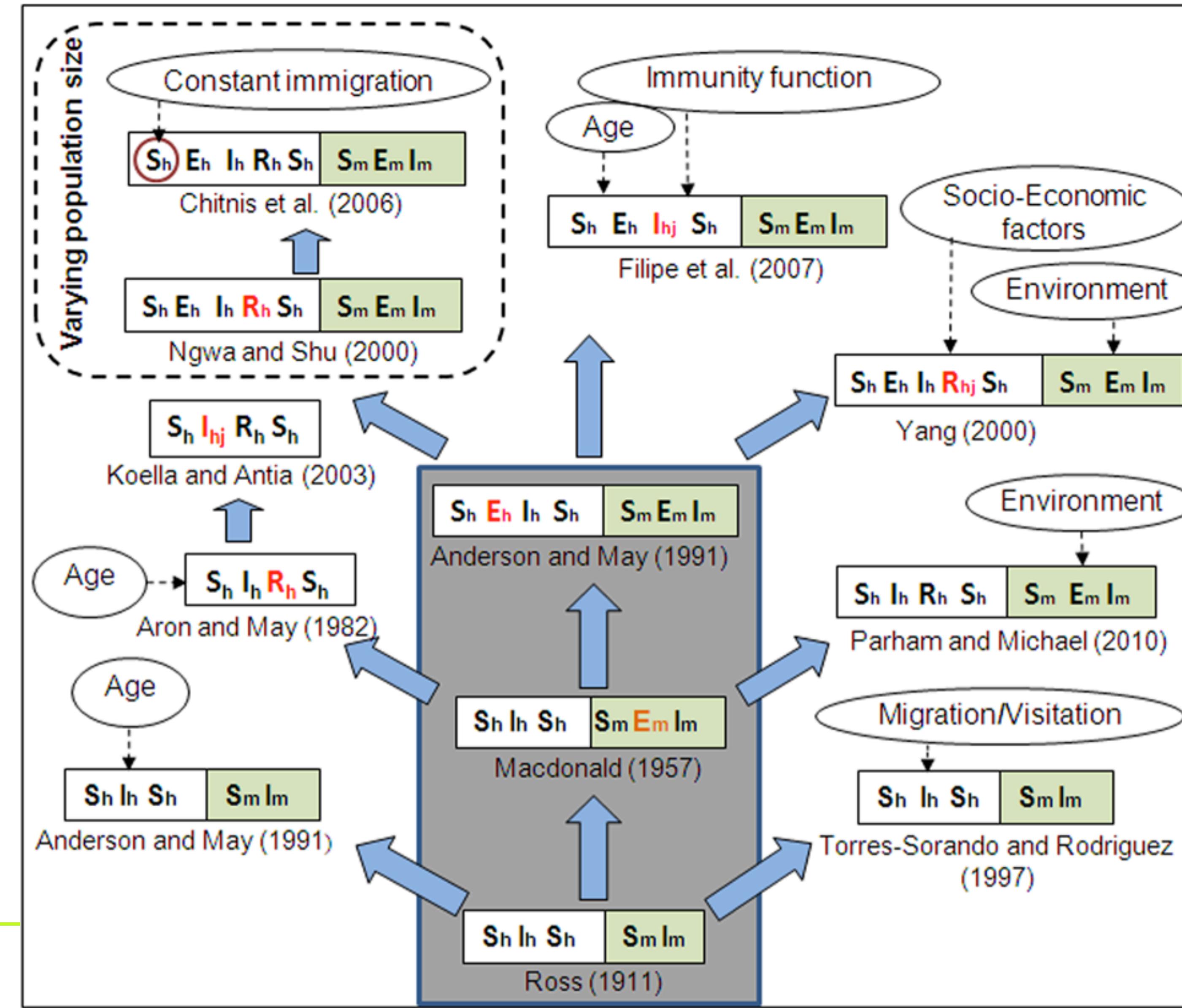
$$\dot{C} = qS - pC$$

$$\dot{D} = \mu I$$

Classes of epidemic processes

- SI: Susceptibles (noninfected which can become infected) and infected
 - HIV, Herpes,
- SIR: Susceptibles, infected, removals (recovered and immune or death)
 - measles, chickenpox, influenza to some extend, HIV, COVID-19
- SIRS: removals are removed only for a certain time, then they can become again infected
 - influenza, flu, COVID-19
- SIS: immediate reinfection after recovering possible
 - parasitic infections, mycosis
- SEIR: in medical literature often the class E - exposed - is added: infected but not yet infectious (SEIR - standard model for COVID 19 up to now)

Evolution of the Malaria model



Plus and minus :ODE versus Agent based

ODE

- + Precise model description easy
- + Easy to simulate
- + relative easy to analyse (R, endemic states, stability,...)
- difficult to add heterogeneity (e.g in form of distributions)

Agent based

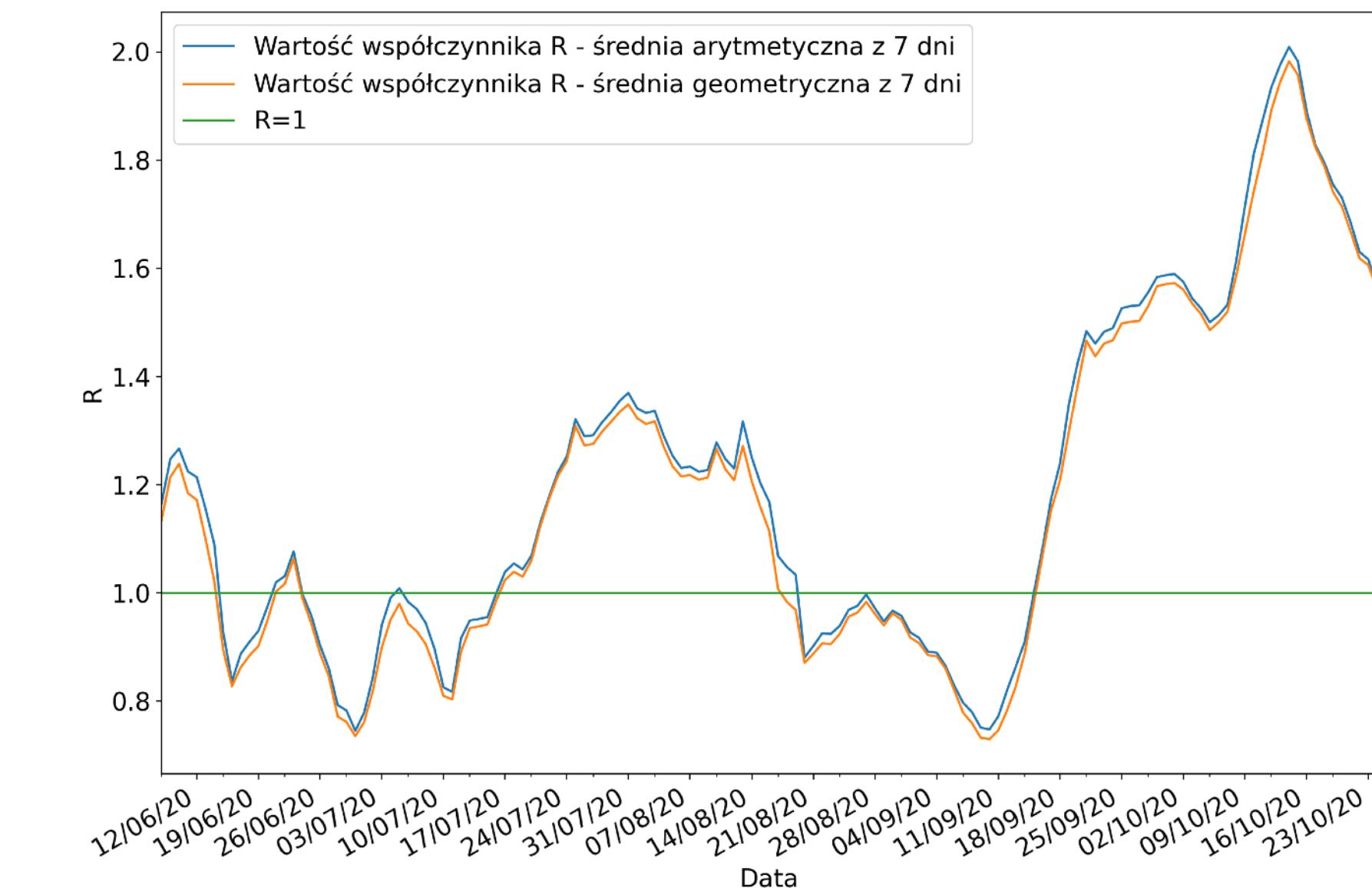
- + easy to add all types of heterogeneity (including spatial structure)
- Many parameters
- Difficult (impossible?) to analyse
- Computational expensive

But: ODE are limit models of agent based models

The Reproduction number R in Epidemiology

- How to find a criterion whether an infection is:
 - a) subcritical (dies out)
 - b) overcritical (positive fraction of the population becomes infected)
- $R_0 =_{def}$ mean number of secondary cases caused by a typical initial infected individual
 - $R_0 > 1 \iff$ overcritical
 - $R_0 < 1 \iff$ subcritical
 - $R_0 = 1 \iff$ critical
- Threshold problem: find the parameter set α_c for which $R_0 = 1$
- simple estimation of stationary infection size I^* :

$$1 = R_0 \left(1 - \frac{I^*}{n} \right) \Rightarrow I^* = n \left(1 - \frac{1}{R_0} \right) \quad (4)$$

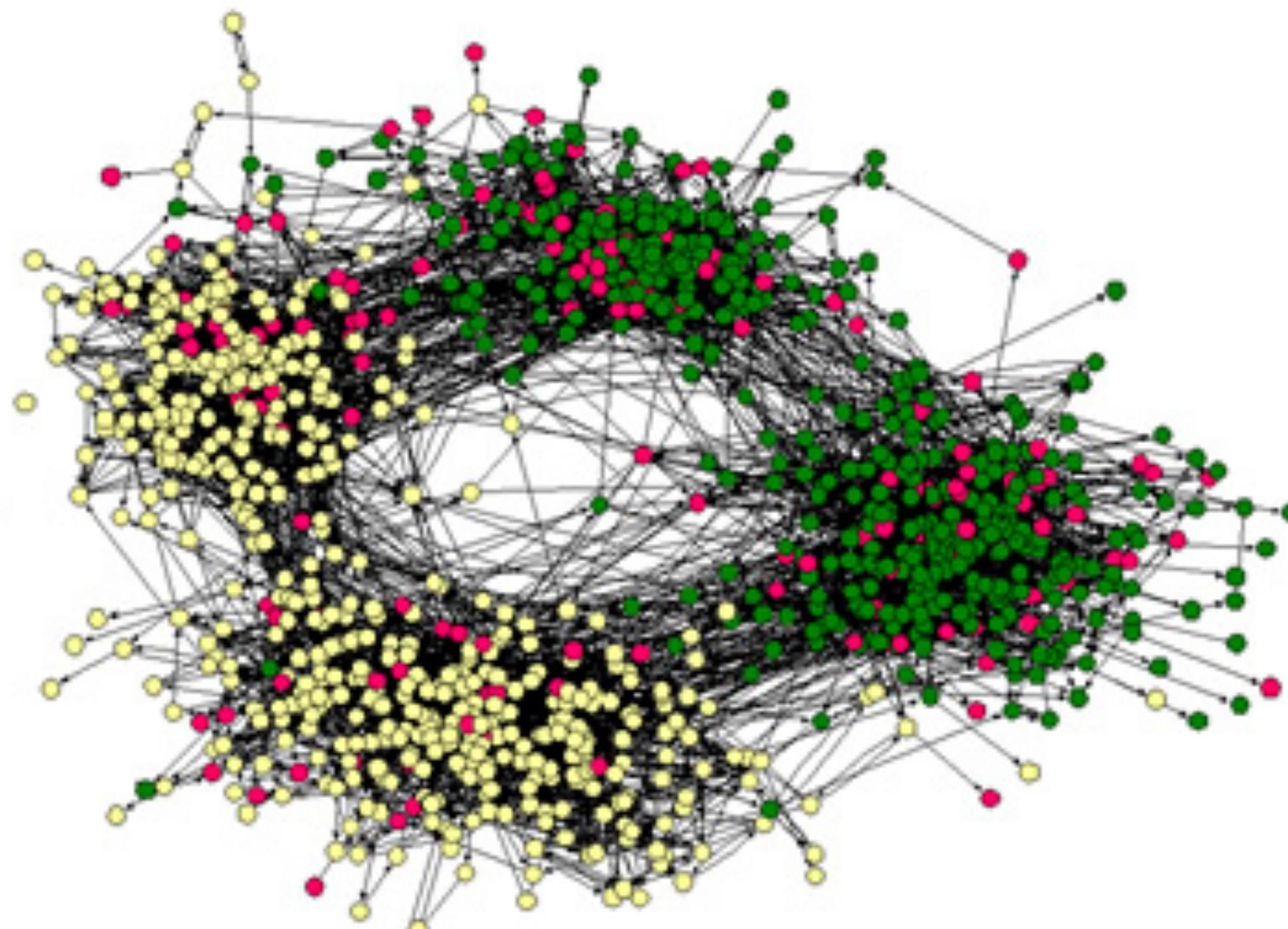


- Is it really true and if yes why?
- $R_{eff}(t)$: effective R is the true mean number of secondary

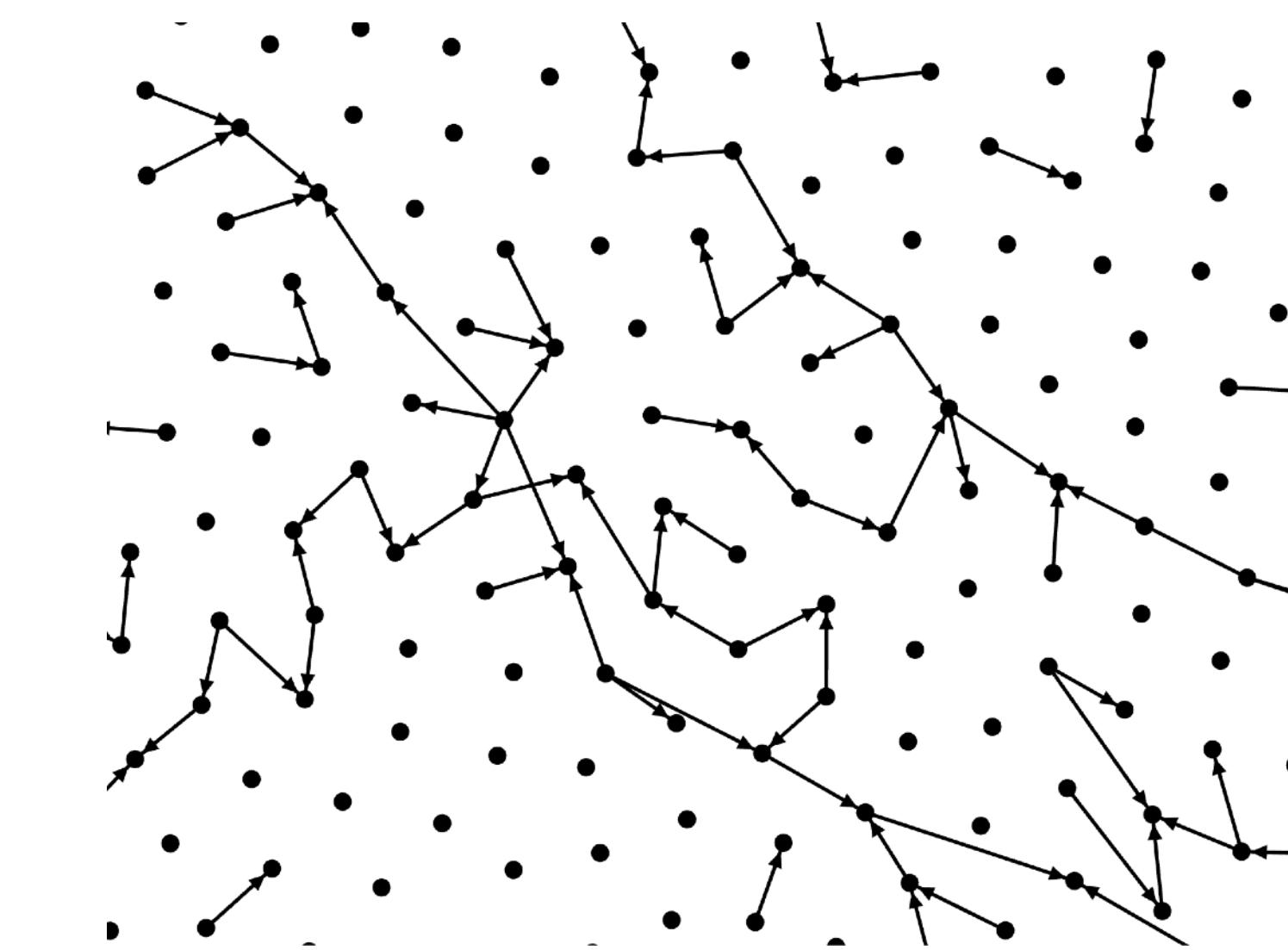
Epidemic models on networks (agent based models)

- agents with individual properties
- agent type dependent contact structure
- random, individual progression of disease
- time and process dependent change of environment
- Targetet complex counter-measures (NPI's e.g. contact tracing

**Epidemic takes place on network of contacts :
hence to understand the spread of infectious diseases one has to
understand the network of contacts and peoples behaviour**



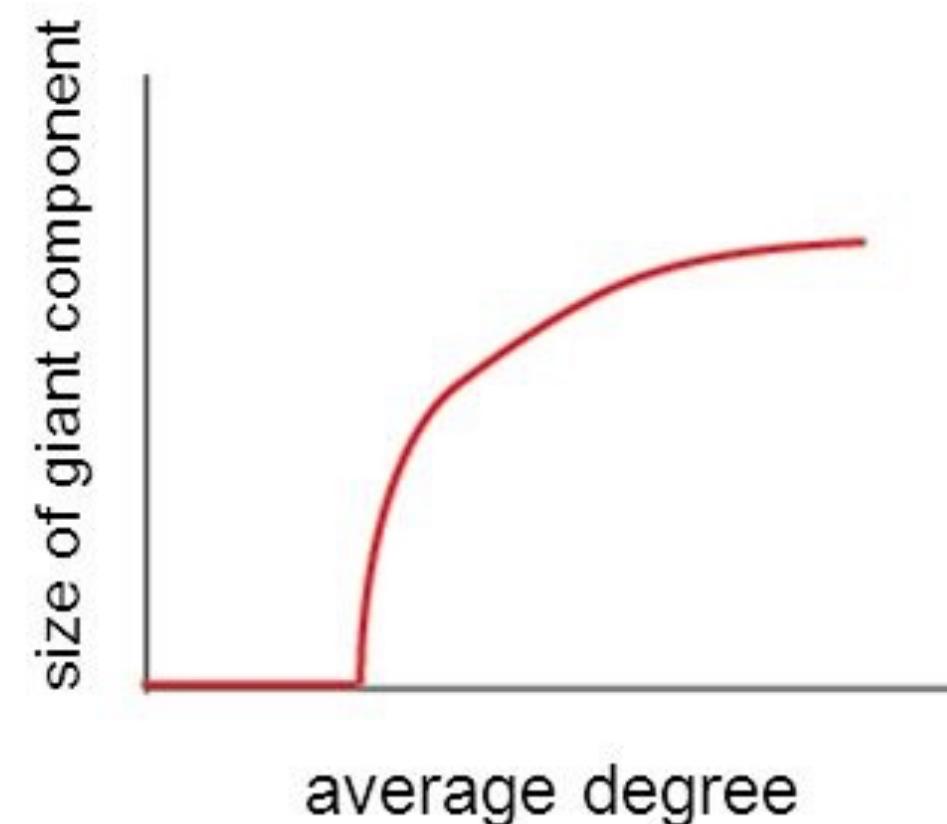
Real contact-networks can be very complex



**Infection graph is a subgraph of the
contact graph**

A remarkable thing happens at connectivity $R = 1$

Percolation threshold in Erdos-Renyi Graphs



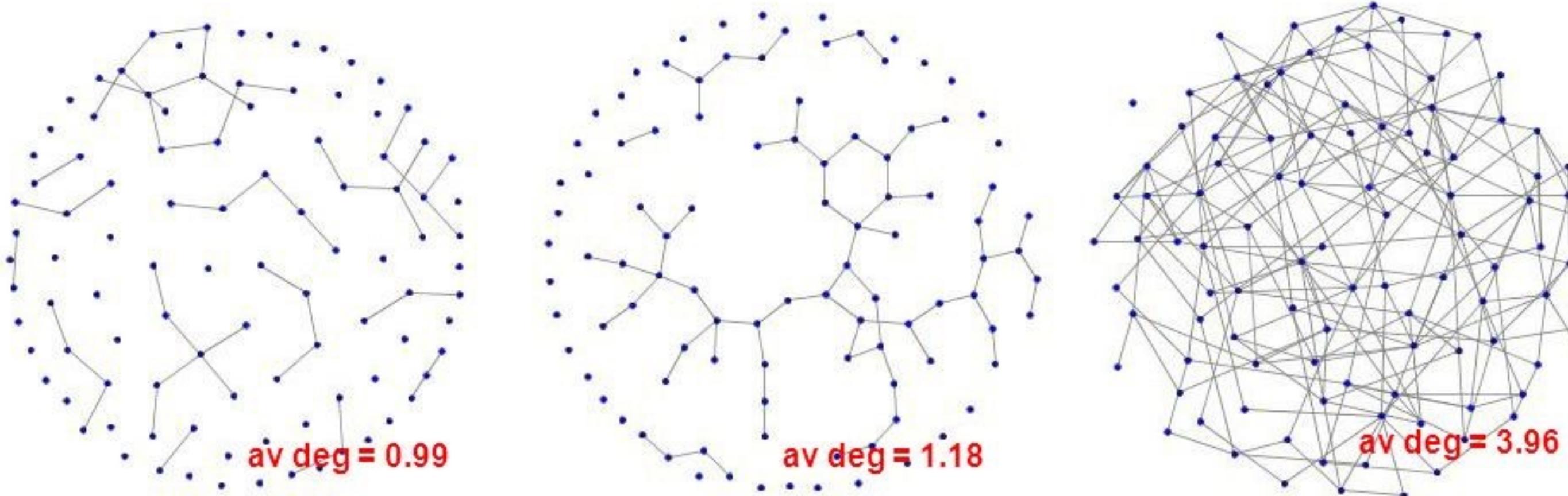
Percolation threshold: how many edges need to be added before the giant component appears?

As the average degree increases to $z = 1$, a giant component suddenly appears



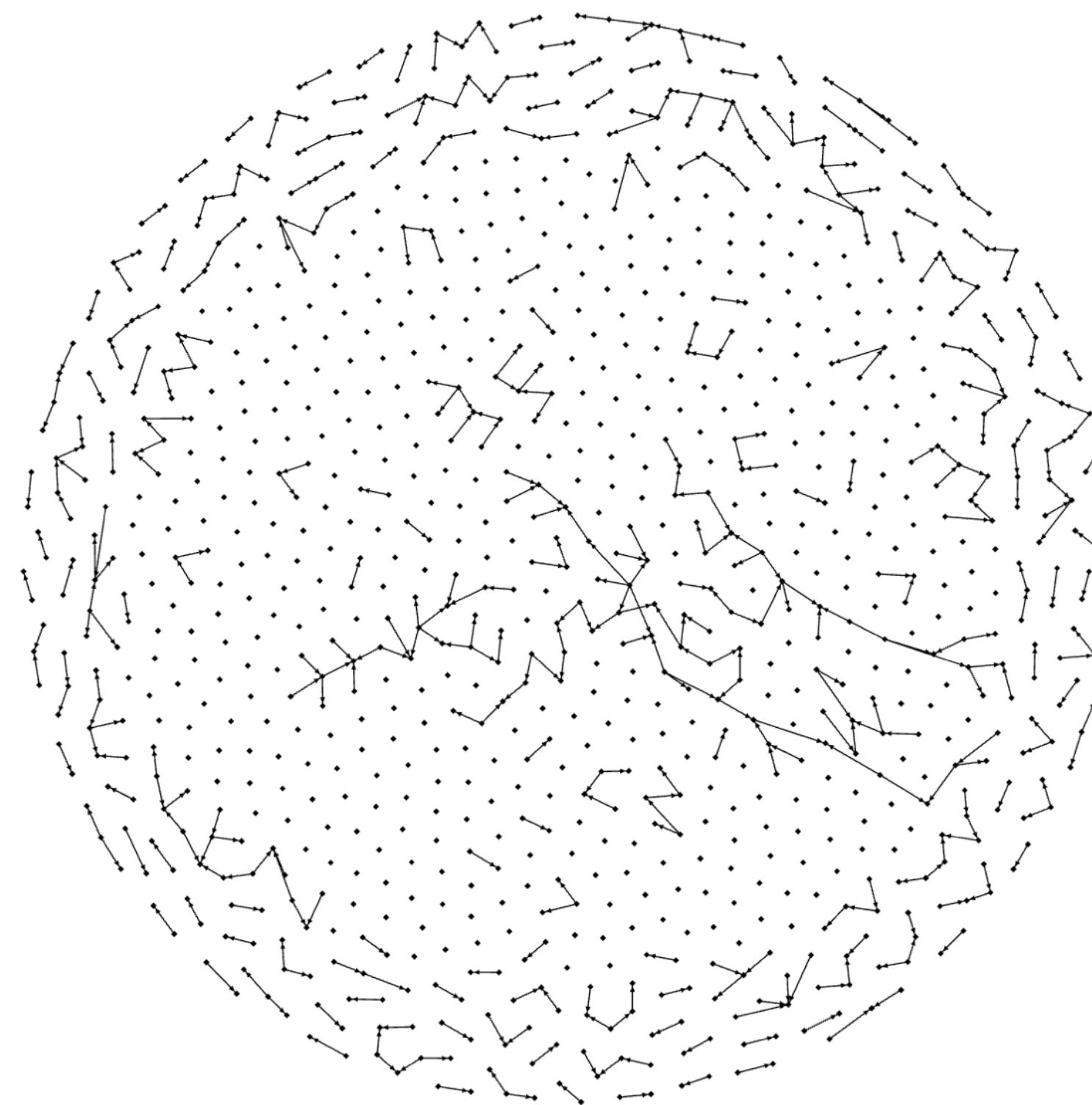
Paul Erdős (1913-1996)

Alfréd Rényi (1921-1970)

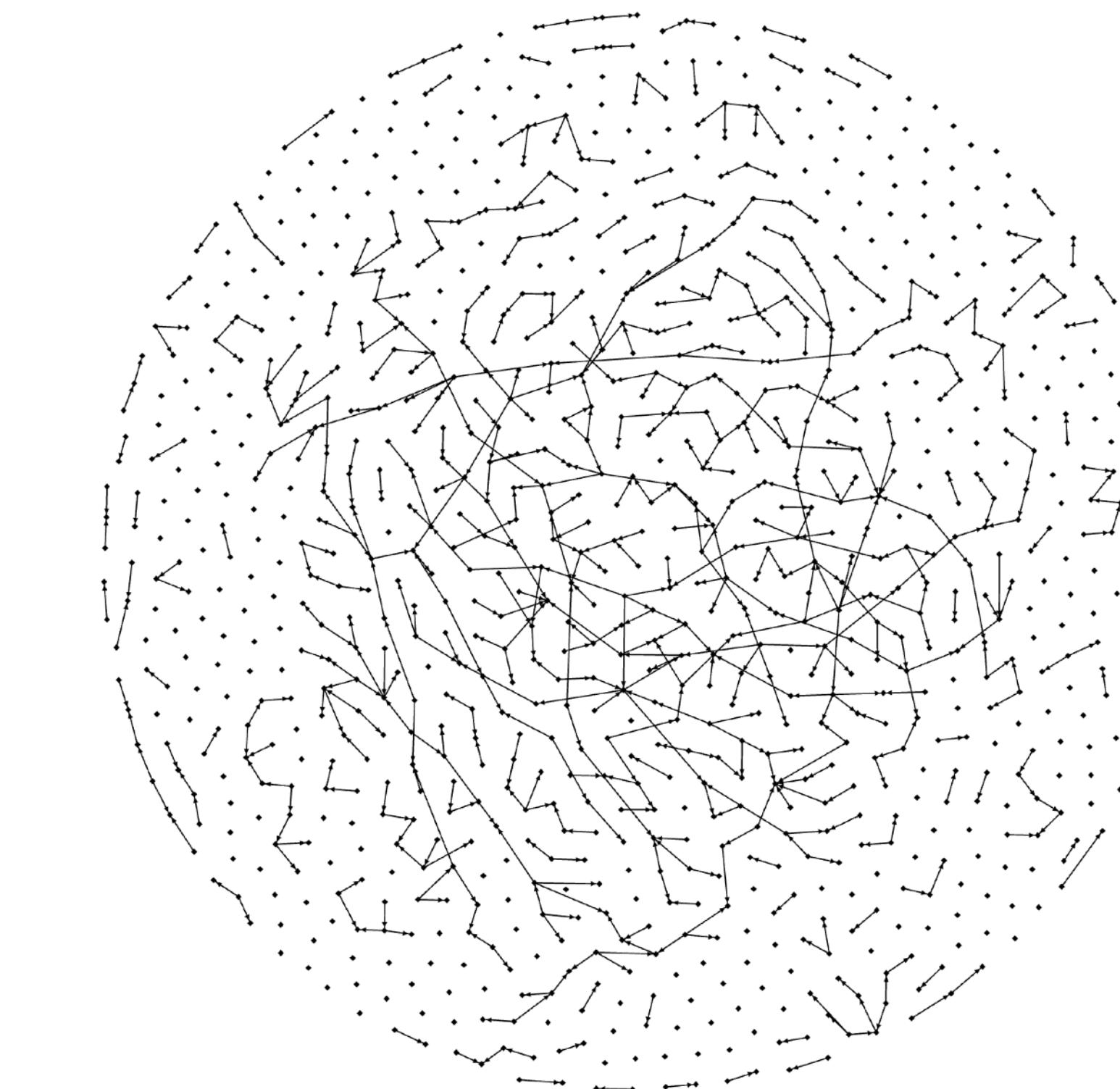


Phase transition in random networks : R is the mean number of contacts

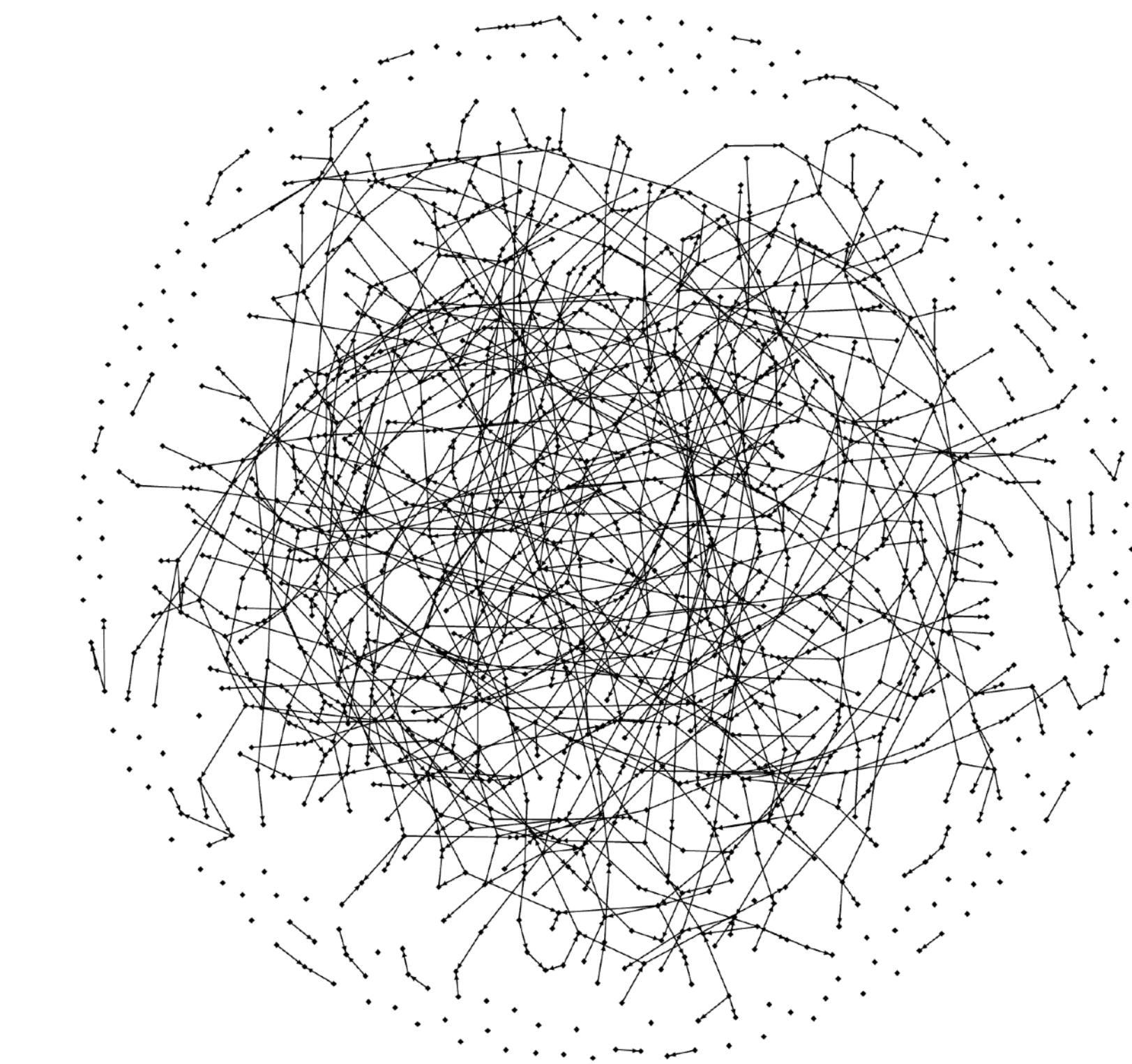
left $R=0.9$,



middle $R=1.2$

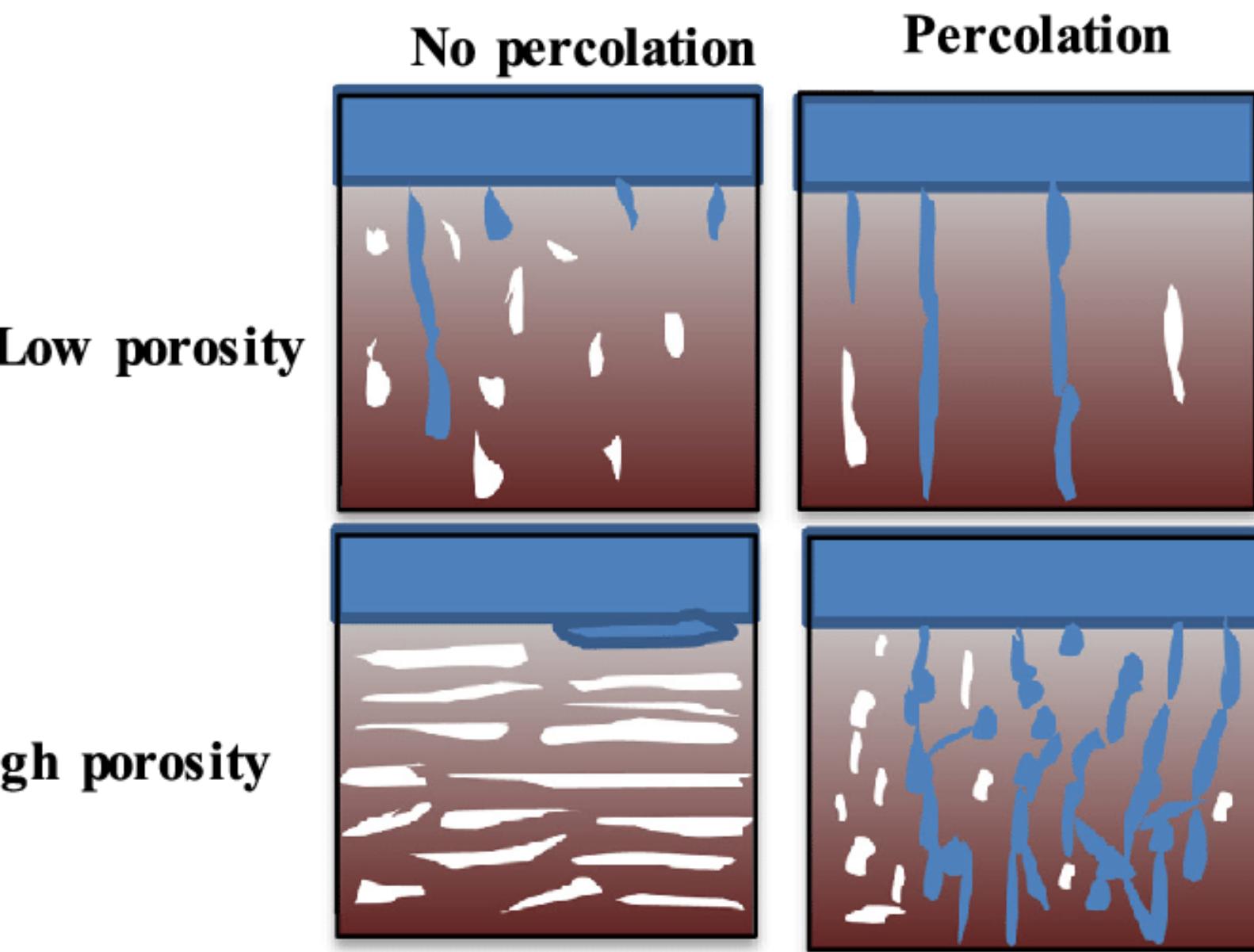
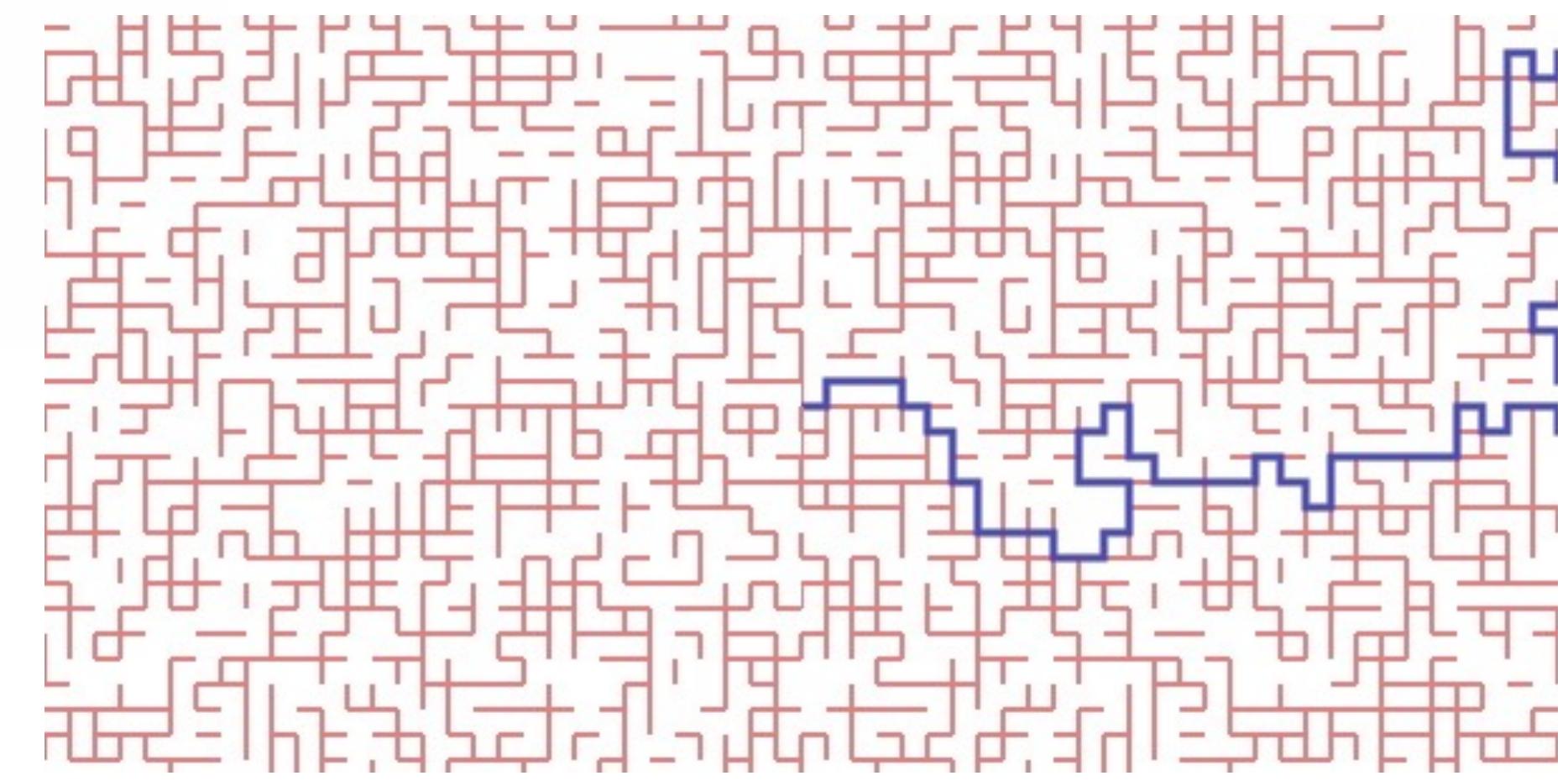


right $R=2$

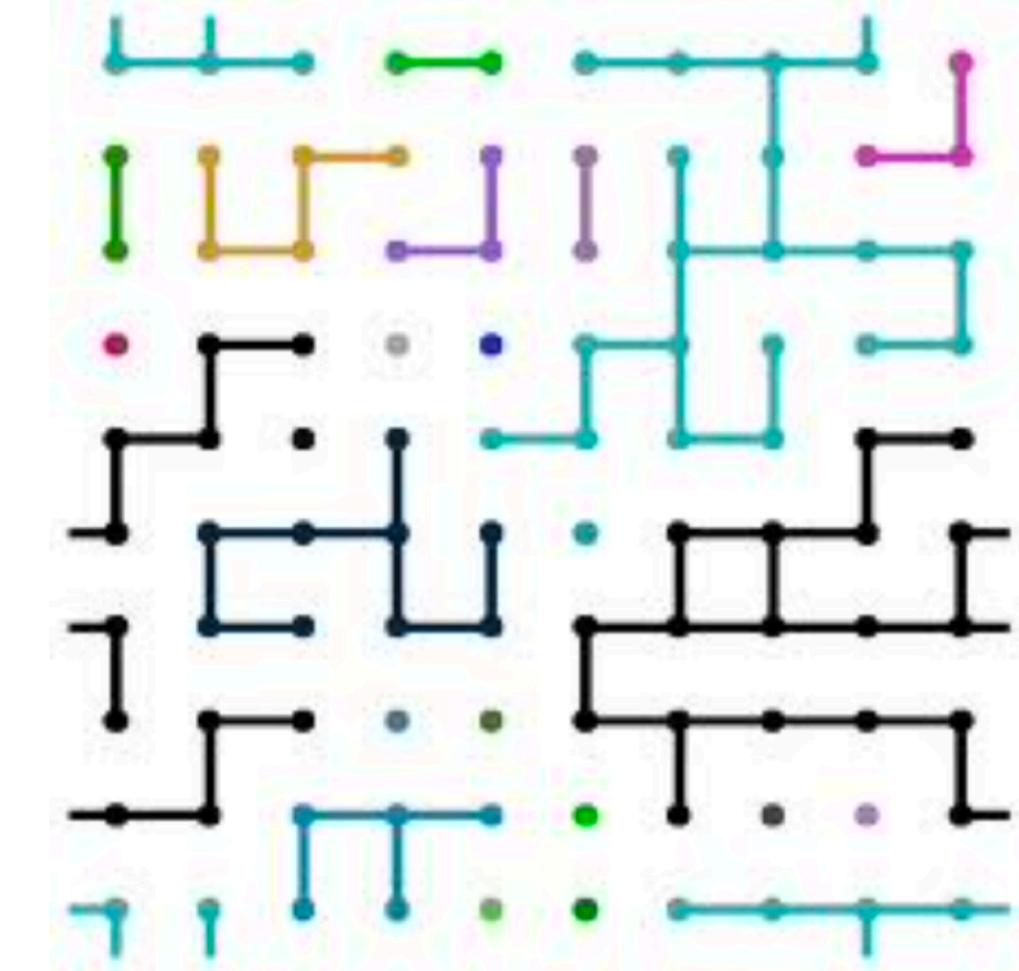


Aim of intervention measure : infection graph splits into many very small components (at most of $\log N$ size)

Percolation Theory



- Diffusion of a liquid through a porous media
- Broadbent&Hammersley 1957
- Mathematical breakthrough by Kesten in 1980
- Open edges between neighbours on a lattice are made with probability p
- For $p > 1/2$ an infinite cluster appears



Epidemic models on networks (agent based models)

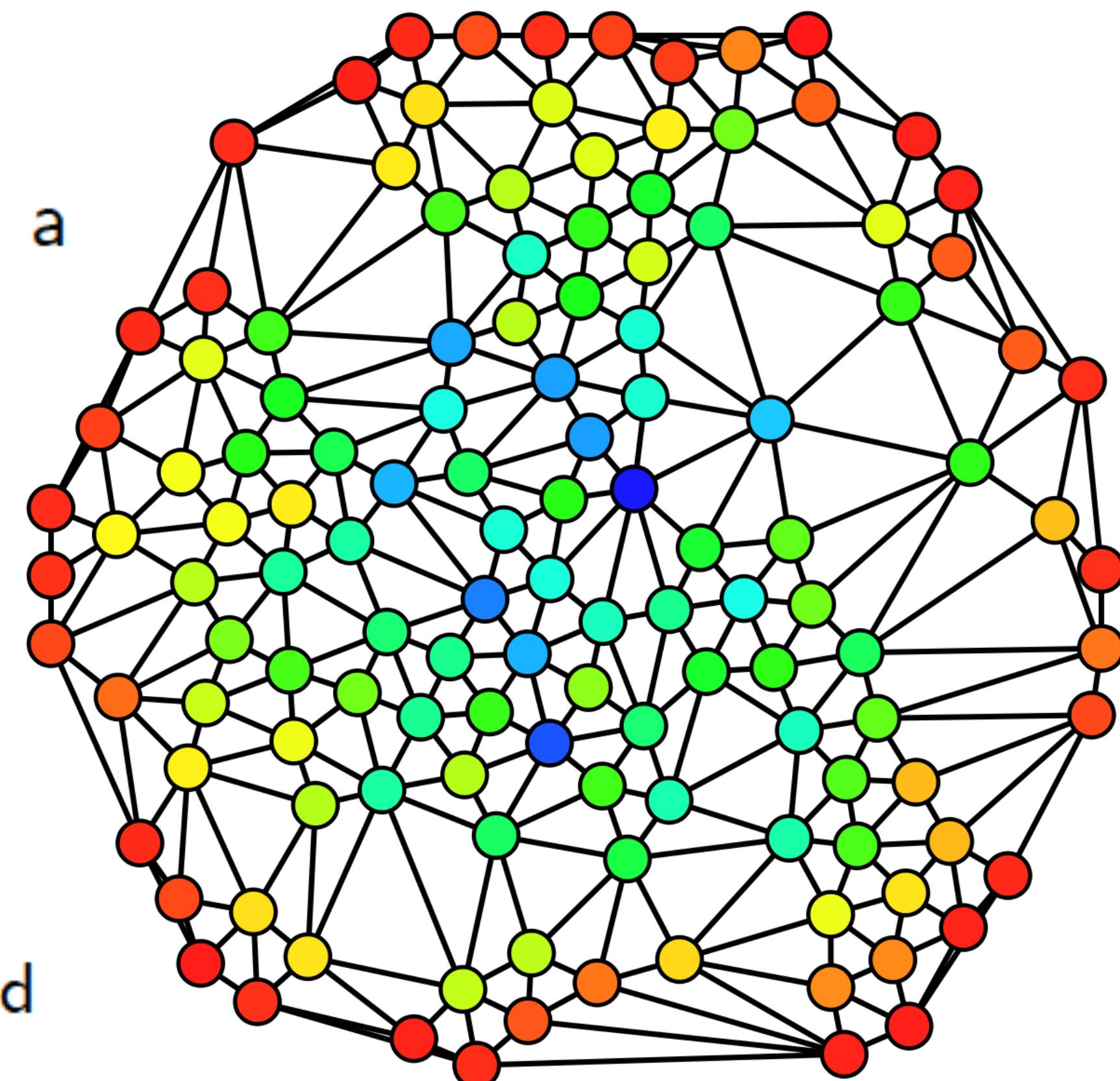
- types $x \in S$,
- infection states of individuals $\chi(i, t, x_i) \in A$
- contacts: $\kappa(x_i, y_j)$ defines probability of contact between individual i and j
- transmission probability (depending on infection states of individuals)
- both state and contacts could be time dependent

Bollobas Janson Riordan random graphs (2005) : beyond Erdos&Renyi

- the n vertices have properties $\{x_i\}$ with values in ground space S
- S could be \mathbb{R}^+ , \mathbb{R}^m , $[0, 1]$, or more complicated spaces
- the values $\{x_i\}$ are asymptotically μ distributed
- edges are independent and edge probabilities are defined via a kernel κ

$$\Pr(i \sim j) = \frac{\kappa(x_i, x_j)}{n}$$

- $\kappa(x, y) \geq 0$
- in epidemic applications
 - the kernel $\kappa(x, y)$ accounts for the social contact structure along which an infection might occur
 - the state space S accounts for a) socioeconomic features and b) for intrinsic medical features like incubation time or severeness of disease progression
 - typically the kernel is a sum of specific kernels for workplace, friendships, public places contacts



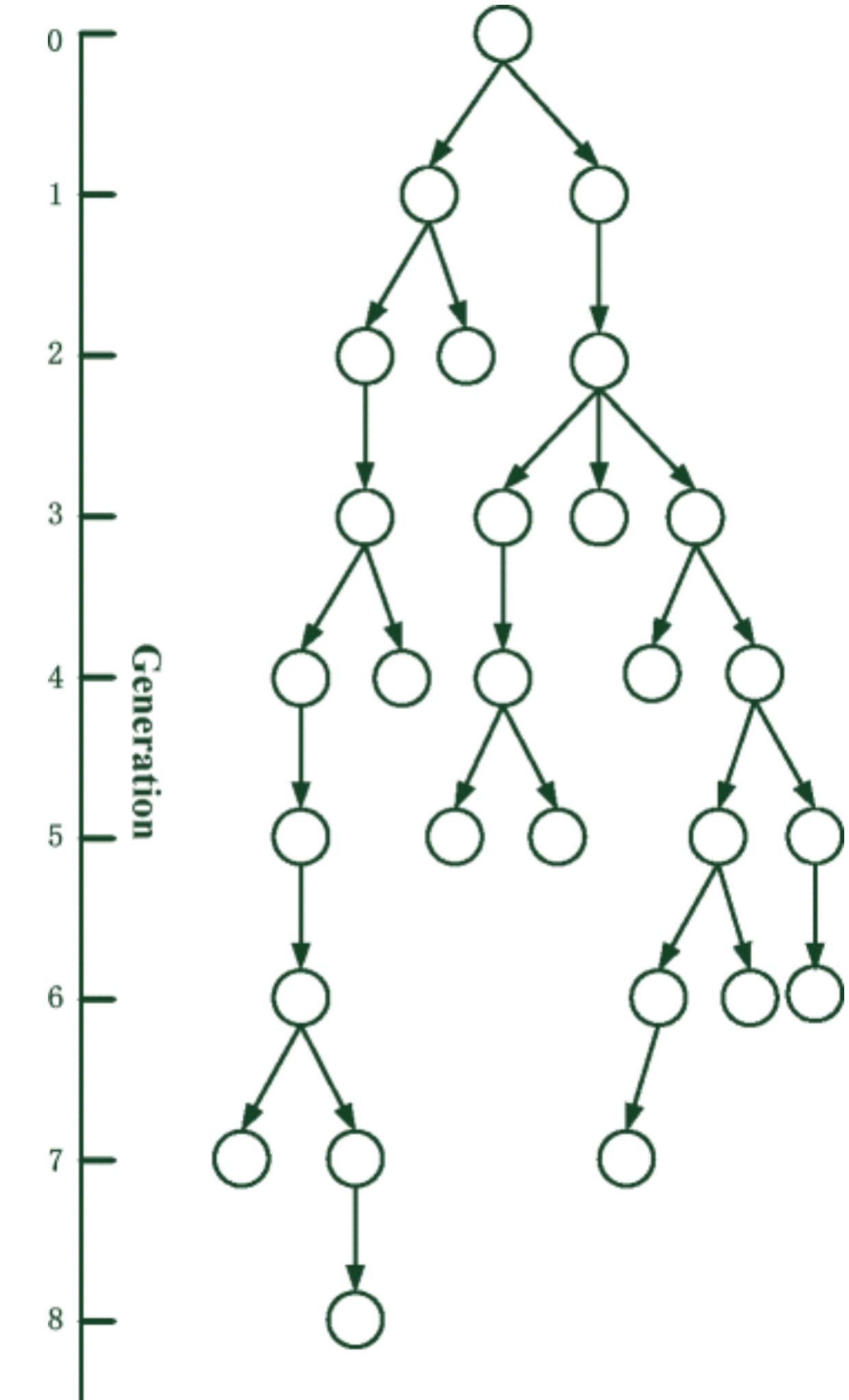
Features which can a priory assigned into BJR type graphs

- socio - economic properties like age, gender, profession, householdsize, workstatus
- time being infectious, incubation time , time being infectious outside household
- tested /not tested (independent of contact tracing)
- quarantined and when
- hospital stay, death
- discovered /not discovered in forward contact tracing
- *NOT* a priory decidable : timing when discovered in contact tracing , discover probability via backtracing



Local structure of BJR random graphs

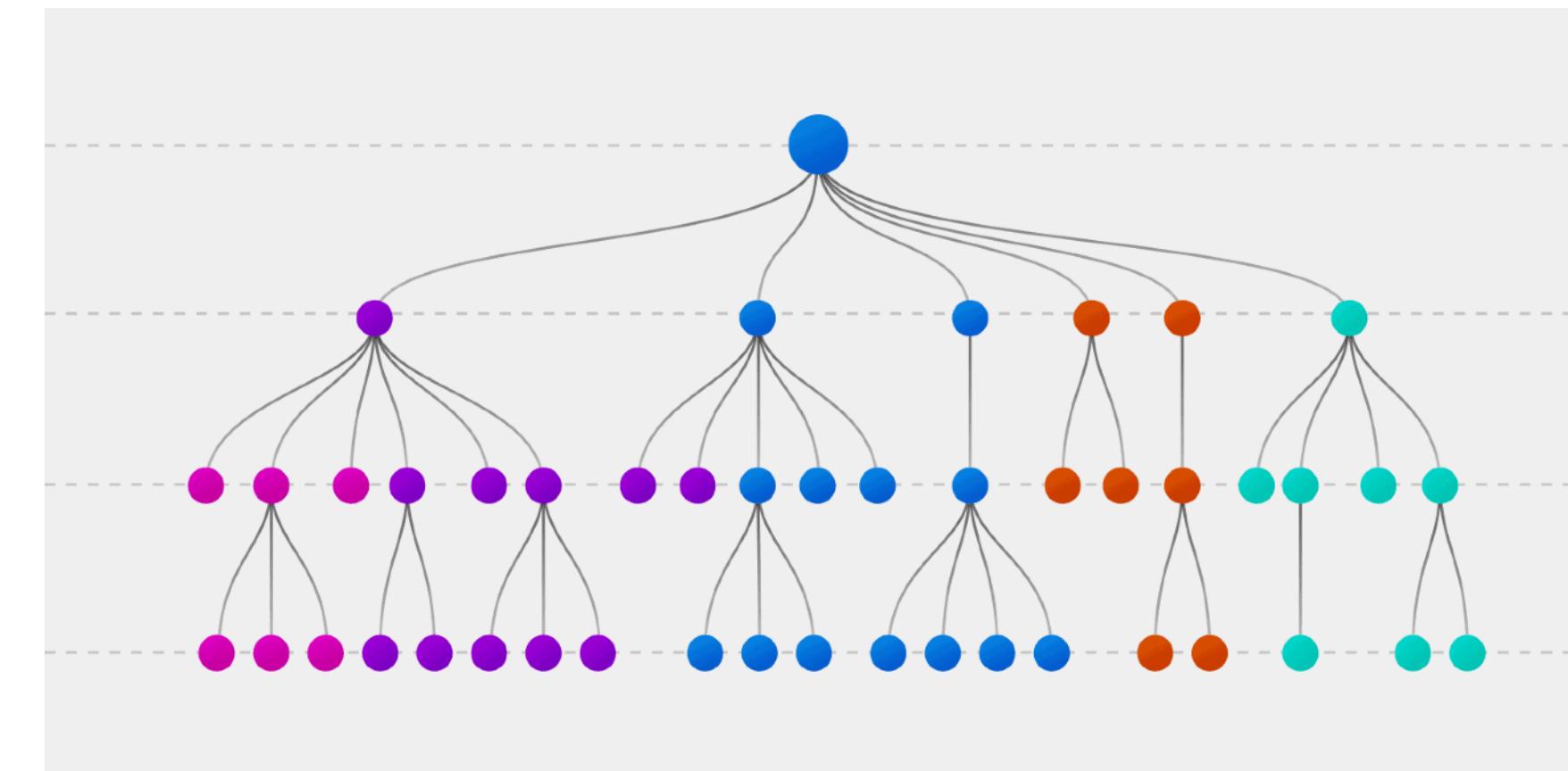
- tree like
- for n large: local structure given by multitype Poisson process with intensity $\kappa(x, y) d\mu(y)$
 - number of offsprings of vertex of type x with type values y in subset A is Poisson distributed with mean $\int_A \kappa(x, y) d\mu(y)$
 - Expected degree of type x node:
$$\mathbb{E}(d(x)) = \lambda(x) = \int_S \kappa(x, y) d\mu(y)$$
: expected total number of offsprings of type x vertex. Note: if μ has density $\varphi(x)$ than $d\mu(x) = \varphi(x) dx$
- Example: $\kappa(x, y) = cxy \Rightarrow \mathbb{E}(d(x)) = \lambda(x) = cx \int_S y d\mu(y) = cx \mathbb{E}_\mu(y)$



BJR random graphs and the transfer operator

- The role of connectivity constant c in Erdős&Renyi graphs is overtaken by the so-called transfer operator

$$Tf(x) =_{def} \int f(y) \kappa(x, y) d\mu(y)$$



- Tf can be seen as the expectation of the observable f over the neighbors of a type x individual
- The phase transition (epidemic threshold) is given by the norm-condition $\|T\|_2 = 1$ (L_2 - norm in a Hilbert space)
 - $\|T\|_2 > 1 \Rightarrow$ there exists a giant connected component of size $\rho \cdot n$
 - $\|T\|_2 < 1$ all components are small $o(n)$
- in SIR epidemics $\|T\|_2$ can be interpreted as the reproduction number R_o : the expected number of secondary cases an infected individual generates

BJR graphs : size of giant component

- The probability of an individual of type x to be in the giant component is given by the solution of

$$\rho(x) = 1 - e^{-T\rho(x)}$$

- $\rho(x)$ also gives the survival probabilities of a multitype - branching (infection process) starting at an individual of type x
- Solution can be easily obtained by iterating the dynamical system $\rho(x) \rightarrow 1 - e^{-T\rho(x)}$ starting with $\rho_0 \equiv 1$
- For $\|T\|_2 > 1 \Rightarrow \rho(x) \neq 0$ and $\lim_{n \rightarrow \infty} \frac{|\mathcal{C}_{\max}|}{n} = \rho$ with
$$\rho = \int_S \rho(x) d\mu(x)$$

Structural phase transitions in BJR graphs (directed and symmetric)

-

$\|T\|_{sp} > 1$: giant component \Rightarrow supercritical epidemic

$\|T\|_{sp} < 1$: many small disconnected component \Rightarrow subcritical

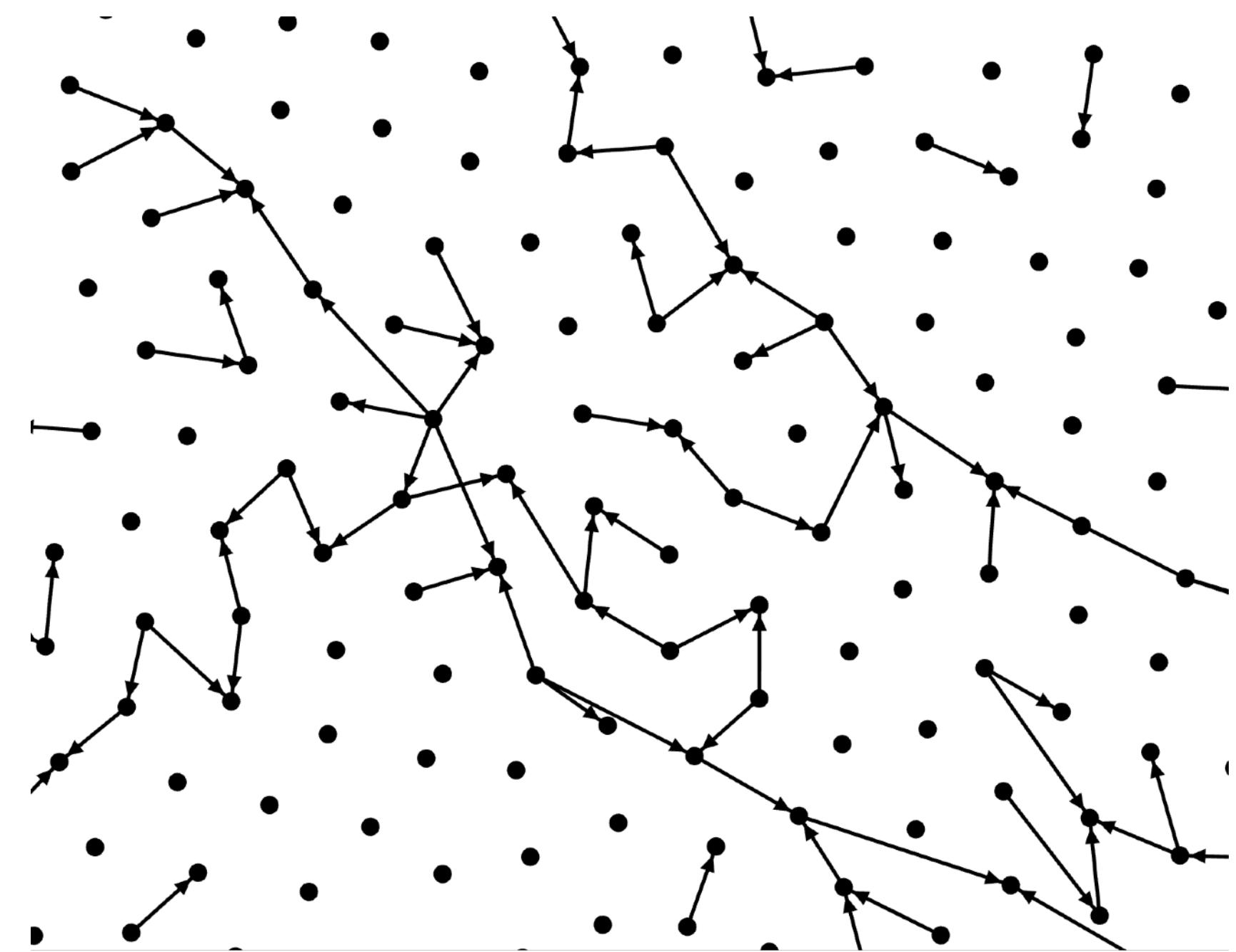
$\|T\|_{sp} = 1$: phase transition

- understanding the epidemics = understanding how $\|T\|_{sp}$ depends on parameters (medical + interventions+social)
- probability $\rho(x)$ of type x node to get infected is given by solution of

$$\rho(x) = 1 - \exp \left(- \int \kappa(y, x) \rho(y) d\mu(y) \right) \quad (12)$$

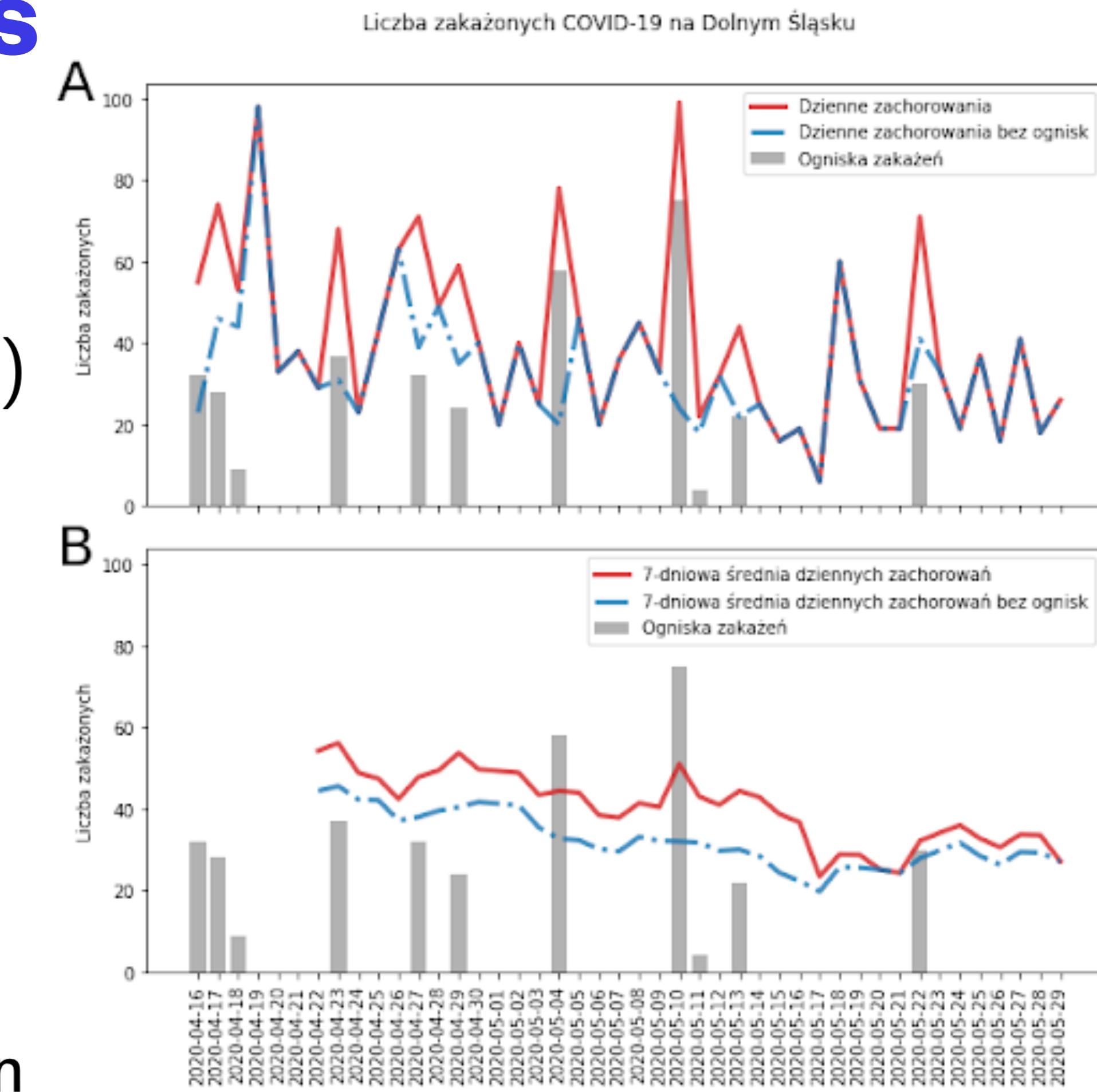
Directed BJR graphs

- $\kappa(x, y)$ not necessary symmetric :
$$\Pr \{i \rightarrow j\} = \min \left(1, \frac{\kappa(x,y)}{n} \right)$$
- there are now two branching processes corresponding to transfer operator:
 - $T^+ f(x) = \int \kappa(x, y) f(y) d\mu(y)$ and
 - $T^- f(y) = \int \kappa(x, y) f(x) d\mu(x)$
- T^+ describes the forward BP and T^- the backward BP, same norm because adjoint
- solution of $\rho^+(x) = 1 - \exp(-T^+ \rho^+(x))$ gives the survival/outbreak probability of BP starting with type x node
- but the probability to get infected is given by solution of backward process $\rho^-(x) = 1 - \exp(-T^- \rho^-(x))$



Application: super spreading events

- assume $x > 1$ heavy tailed distributed (e.g. $\text{const} \cdot x^{-\alpha}$)
- kernel : $\kappa(x, y) = cx$
- outbreak probability :
$$\rho^+(x) = 1 - \exp(-cx \int \rho^+(y) d\mu(y))$$
- infection probability :
$$\rho^-(y) = 1 - \exp(-c \int x \rho^-(x) d\mu(x)) \Rightarrow \rho^-(y) \text{ is constant } \rho^- \text{ and}$$
- $\rho^- = 1 - \exp(-\rho^- \cdot \mathbb{E}(x))$: the Erdos&Renyi equation
- conclusion : for the final prevalence of the epidemics we only need to know the expectation of x
- ρ^+ can be much smaller than ρ^-

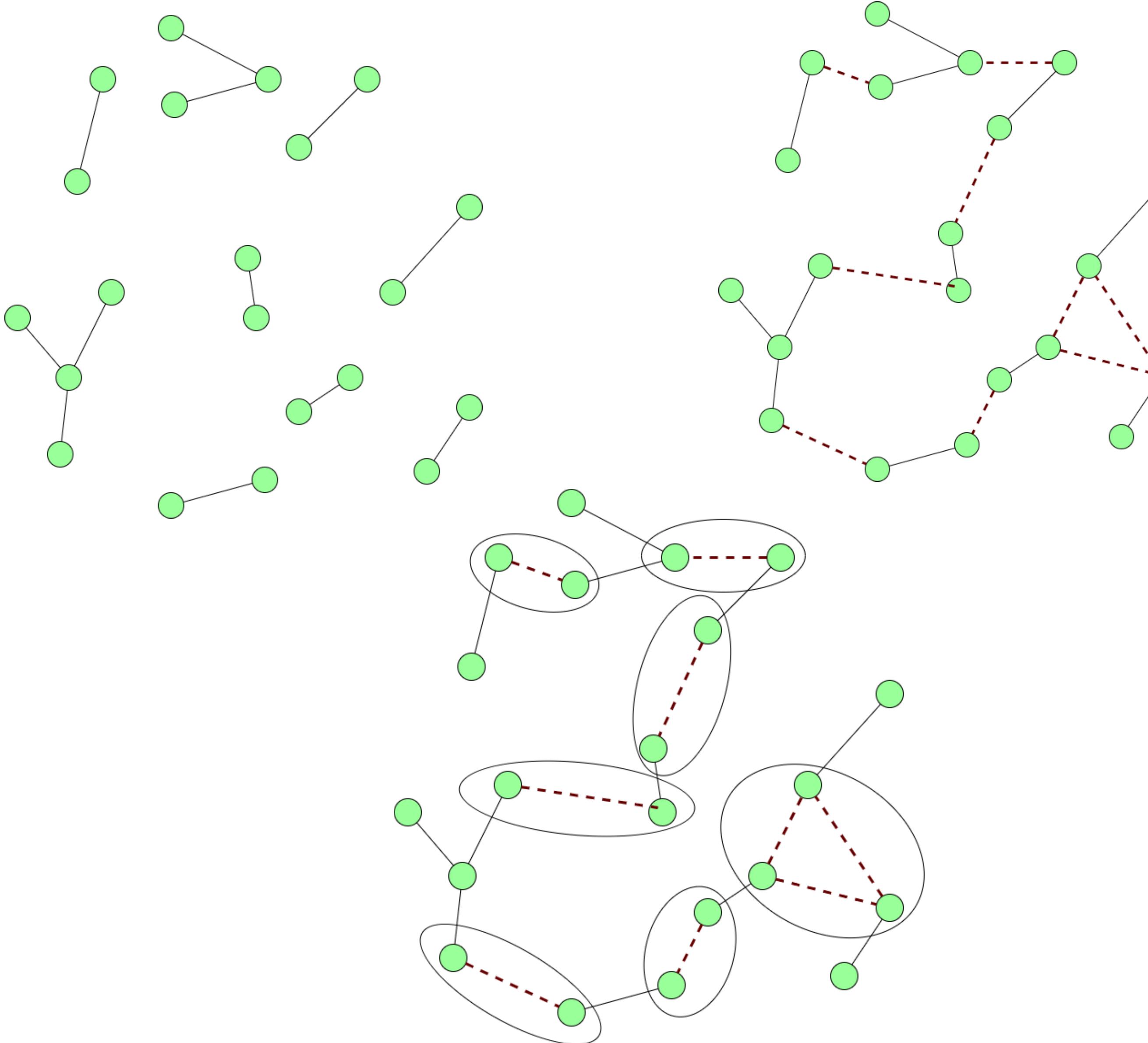


Household structure plays
significant role in epidemics spread



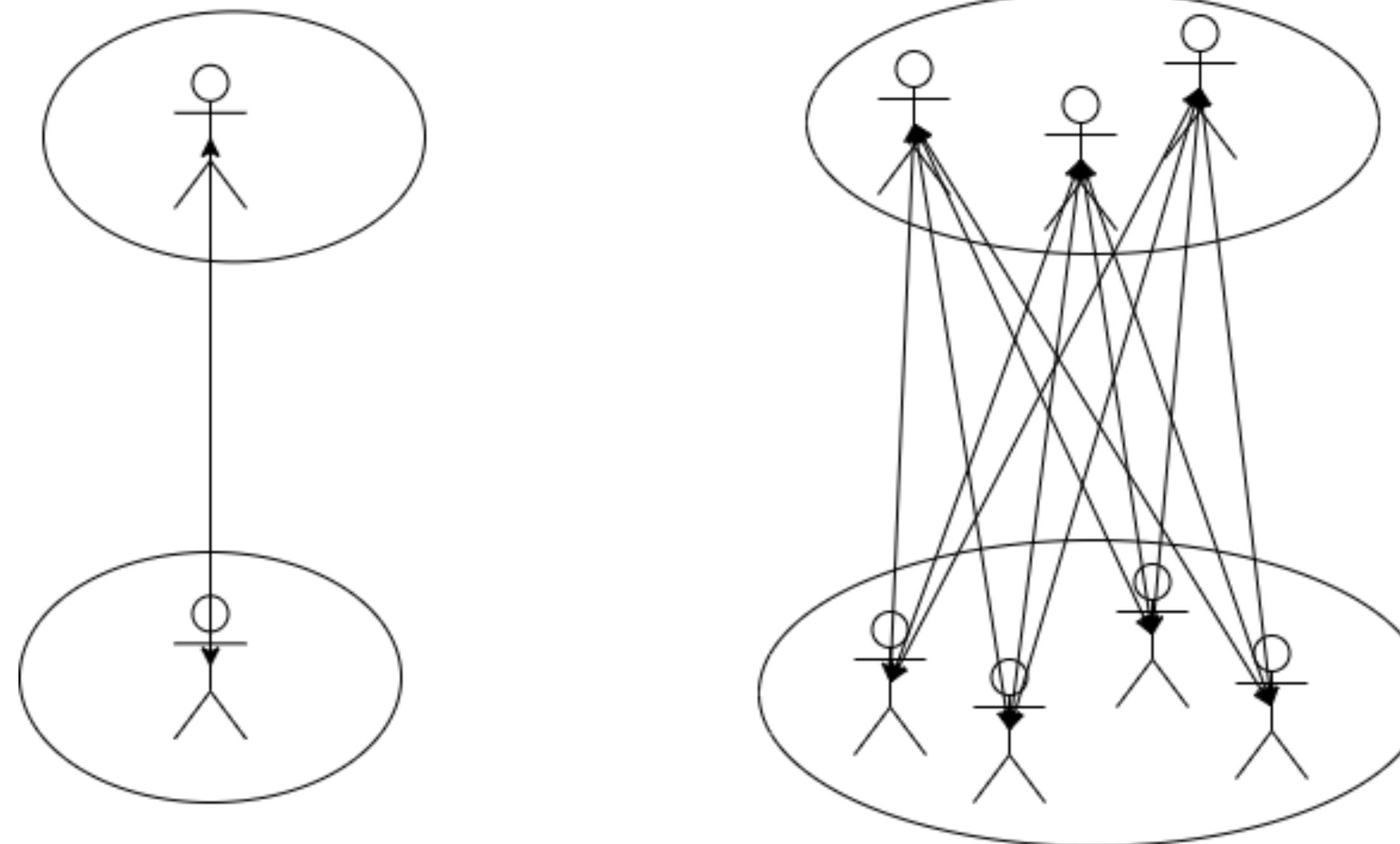
(joint work of MOCOS and Viola Priesemann team from MPI Goettingen)

Structural insights: Households are a key factor for epidemic dynamics



**From network between
individuals to network
between households**

Large households have catalytic effect on the spread because link probability between two households is proportional to the product of their sizes



Household graphs : simple example of constant attack rate a

- the out-household contact structure is an Erdös&Renyi graph with parameter c
 - in other words: out-household individual reproduction number is $Poiss(c)$
 - population is partitioned into households.
 - nodes are interpreted as households
 - node feature x_i is now the household size
 - E the expected household size
 - a_H the household attack rate: after a household of size k gets infected the expected number of remaining household members which get infected is $a_H \cdot (k - 1)$
 - $\eta(k)$ is the household size distribution (fraction of households of size k)
 - (setting can easily be extended to households with different age composition and age specific coupling between households)



Household graphs : simple example of constant attack rate a

- household kernel defining the link structure between households of size x and size y :

$$\kappa(x, y) = \frac{c}{E} (1 + a_H(x - 1)) \cdot y \quad (2)$$

- associated household transfer operator:

$$Tf(x) = \int \frac{c}{E} (1 + a_H(x - 1)) \cdot y \cdot f(y) d\mu(y) \quad (3)$$

- T is a rank one, non-symmetric linear operator
- spectral norm of the transfer operator is given by :

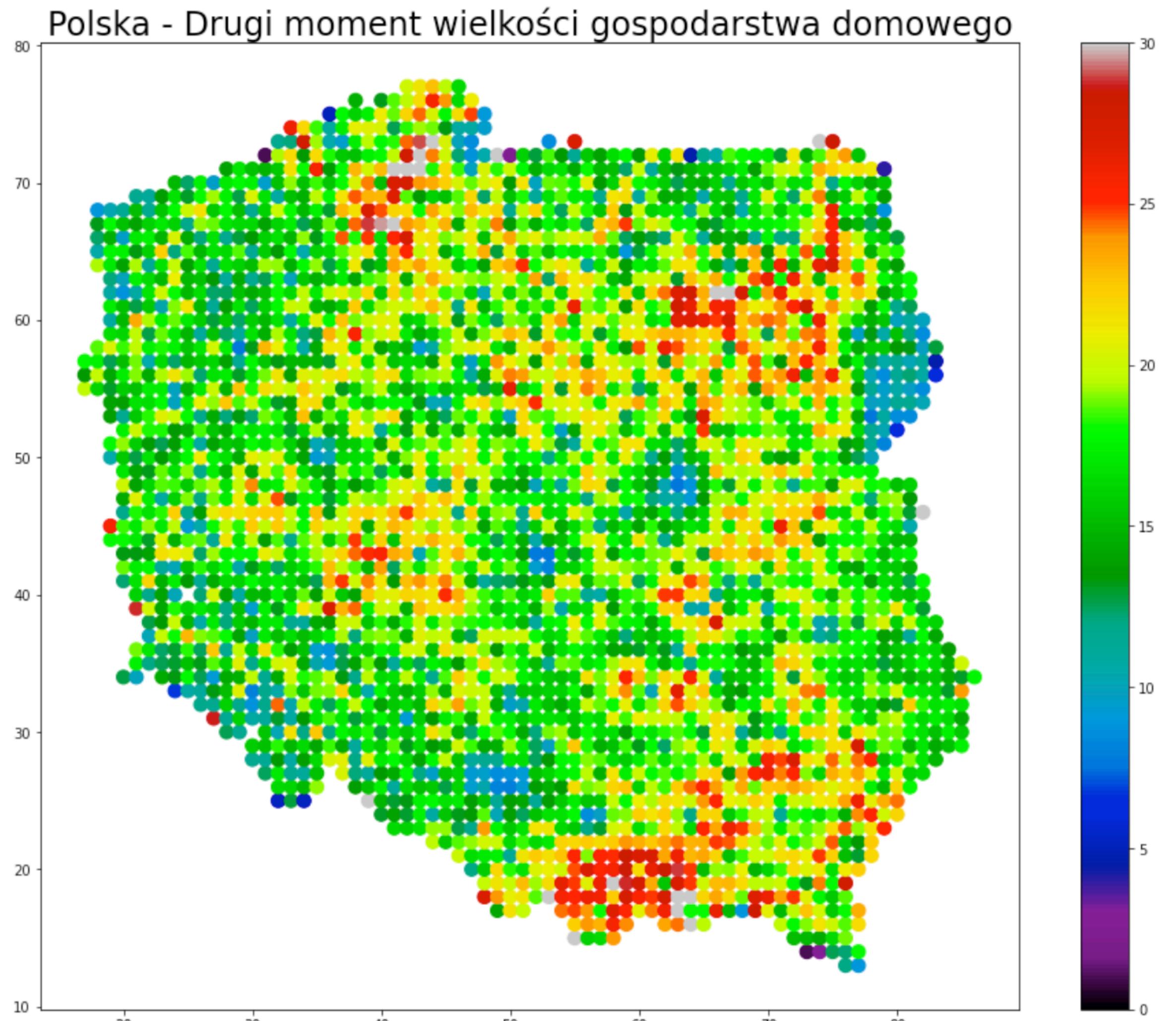
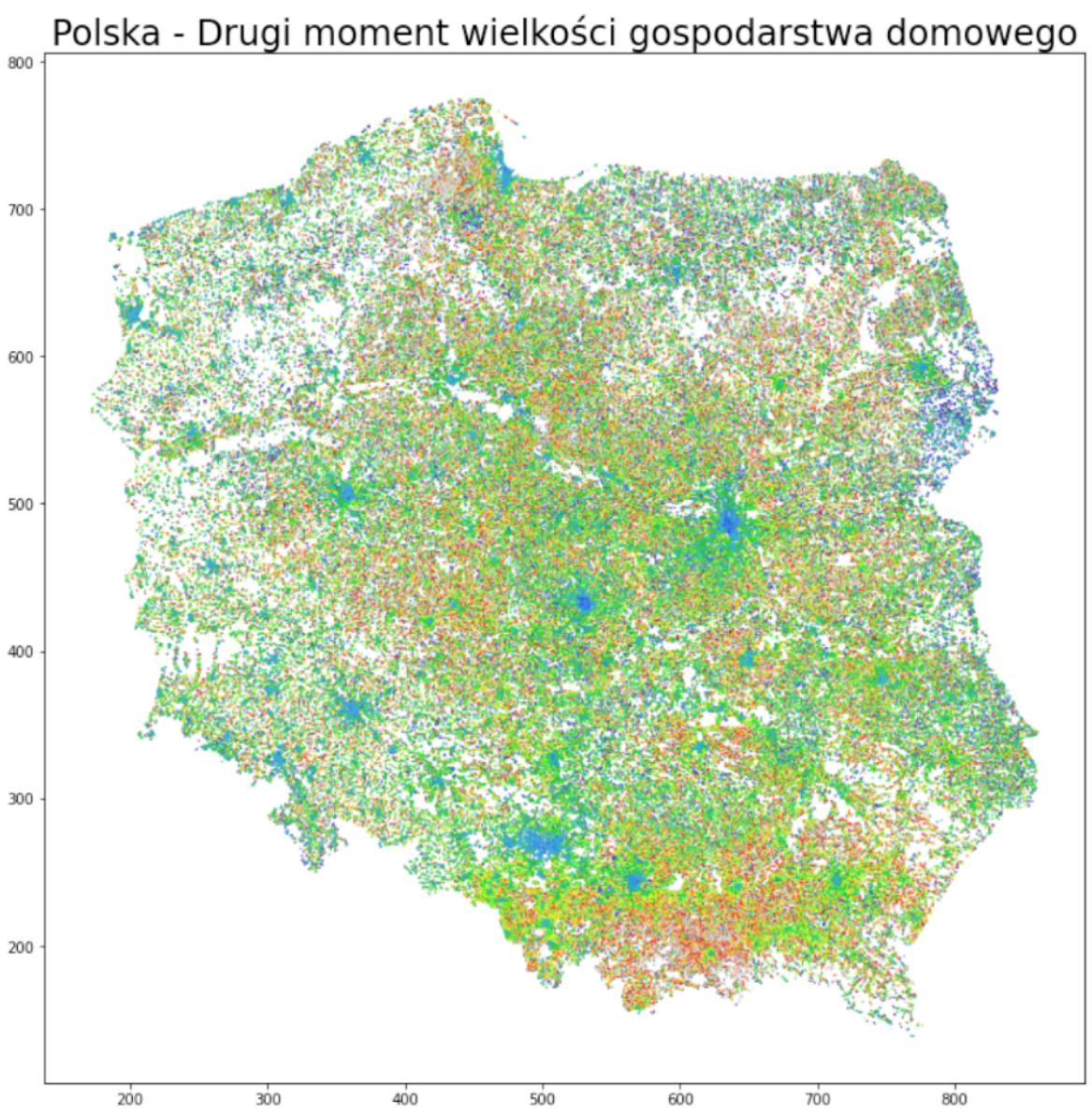
$$\frac{c}{E} \sum_{k \geq 1} (1 + a_H(k - 1)) \cdot k \cdot \eta(k) \quad (4)$$

$$= c \left(1 + a_H \left(\frac{m_2}{E} - 1 \right) \right); \text{ } m_2 \text{ second moment of } \eta \quad (5)$$

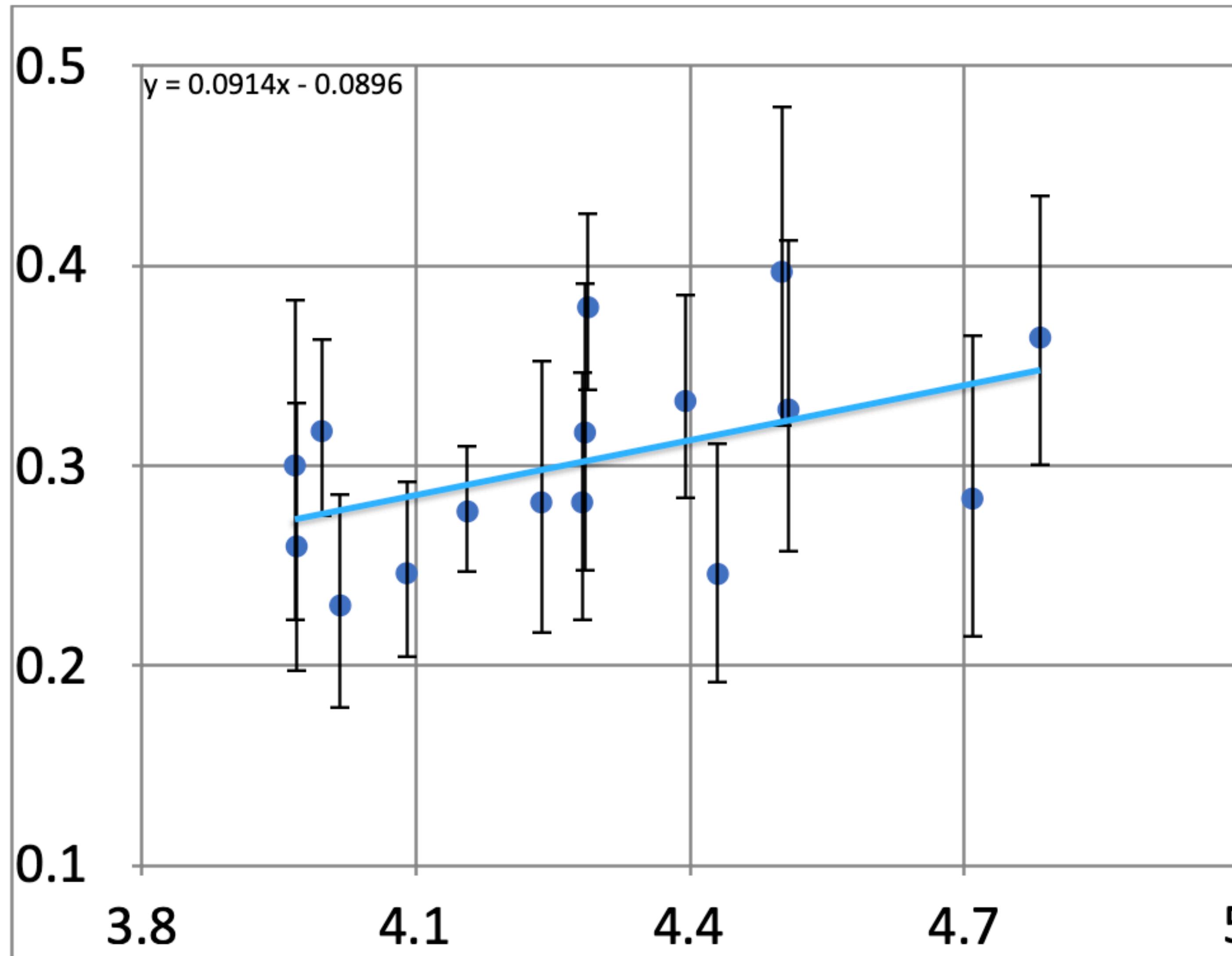
where $\eta(k)$ is the fraction of households of size k

Household graphs : simple example of constant attack rate a

- in case out-household individual reproduction number is not $Poiss(c)$ but from distribution φ replace c by expectation \bar{c} with respect to φ
- Poland: $\frac{m_2}{E} \simeq 4.26 ; (1 + 0.2(4.26 - 1)) \simeq 1.65 : a_H = 0.2$ (british variant)
- Germany: $\frac{m_2}{E} \simeq 2.75 ; (1 + 0.2(2.75 - 1)) = 1.35$
- age coupling kernel $\kappa(a, b)$ for age a with age b :
- household type not only given by size but also age composition and who is infected
- for spectral norm computation for realistic age kernels one has to work with empirical distribution of household composition (e.g. from microcensus data)



Empirical correlation: seroprevalence Poland for different Wojewodstwos (April/May 2021) versus moment-ratio. (OBSERCO study)

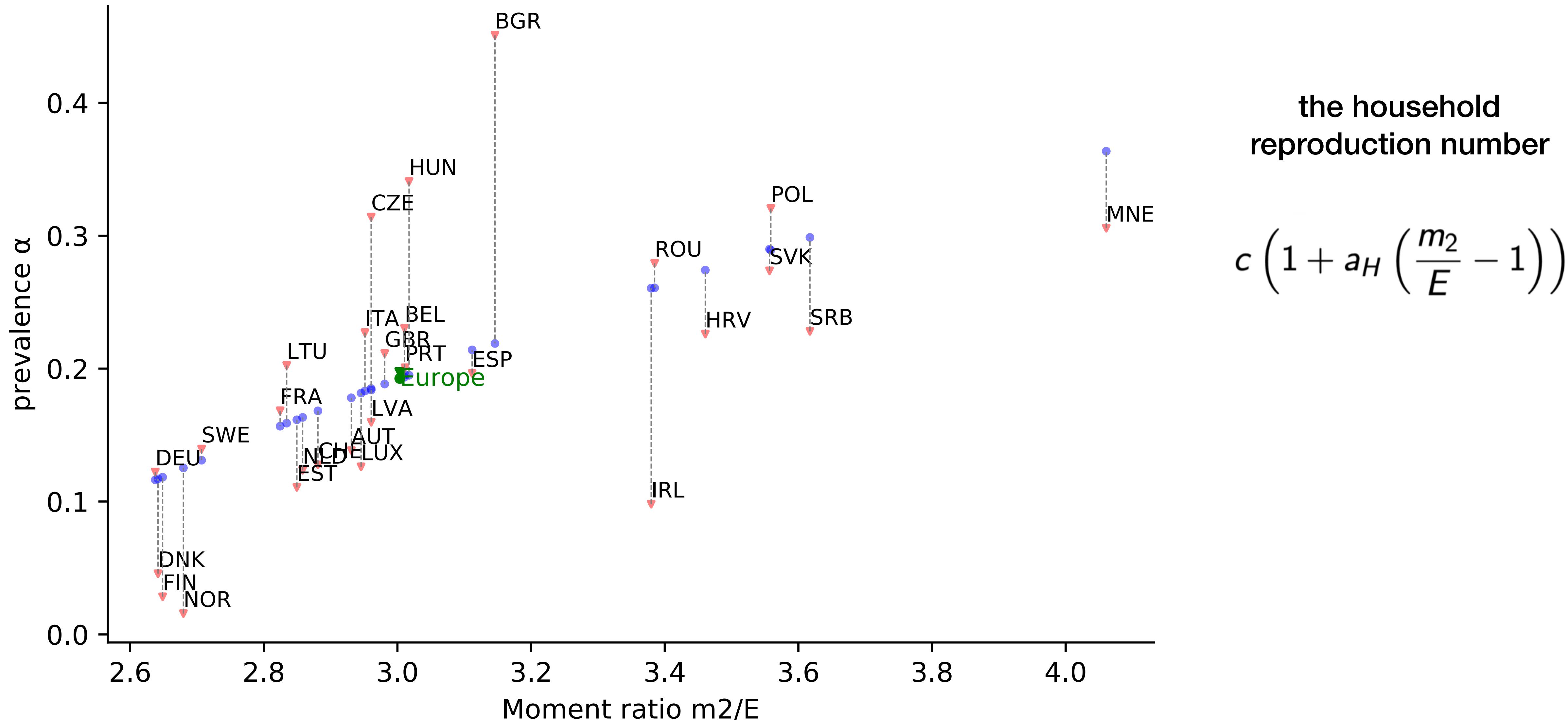


the household
reproduction number

$$c \left(1 + a_H \left(\frac{m_2}{E} - 1 \right) \right) :$$

Excess deaths versus moment-ratio

Blue: European average c predictions



Households and prevalence: looking backwards

- warm up : single households
- assume after some waves there are I infected in population of size n with $\alpha = I/n$
- each infected i would have "produced" c_i secondary cases if his contacts where all susceptible
- hence c_i is the number of infectious contacts of individual i
- let $c = \frac{1}{I} \sum c_i$ be the average
- what is the probability p that a (random chosen) individual got infected?
- clearly $p = \alpha$ but there is another way to compute p

Prevalence and reproduction number : population of singles

- what is the probability p that a (random chosen) individual got infected?
-

$$p = 1 - \prod_{i \in I} \left(1 - \frac{c_i}{n}\right) \quad (6)$$

$$\simeq 1 - \exp\left(-\frac{1}{n} \sum c_i\right) \quad (7)$$

$$= 1 - \exp\left(-\frac{\alpha}{I} \sum c_i\right) \quad (8)$$

$$= 1 - \exp(-\alpha \cdot c) \quad (9)$$

- but p also equals $\alpha \Rightarrow c$ (average) can be computed from α :

$$c = \frac{-\ln(1-\alpha)}{\alpha}$$

- the previous $p = 1 - \exp(-\alpha \cdot c)$ is the probability to get infected from an outhousehold contact
- Let μ_k be the expectation of infected individuals in households of size k
- we have

$$\sum_{k \geq 1} H_k \mu_k = \alpha \cdot n \quad (10)$$

$$\text{and hence} \quad (11)$$

$$\sum_{k \geq 1} \eta_k \mu_k = E \cdot \alpha \quad (12)$$

Computation of expectation number of infected in k-household

μ_k

- the previous $p = 1 - \exp(-\alpha \cdot c)$ is now the probability to get infected from an outhousehold contact
-

$$\begin{aligned}\mu_k &= \sum_{l=1}^k \binom{k}{l} p^l (1-p)^{k-l} \left(l + (k-l) \left(1 - (1-a_h)^l \right) \right) \\ &= pk + \sum_{l=1}^k \binom{k}{l} p^l (1-p)^{k-l} (k-l) \left(1 - q_H^l \right) \\ &= k \left(1 - (1-a_H p)^{k-1} (1-p) \right)\end{aligned}\tag{13}$$

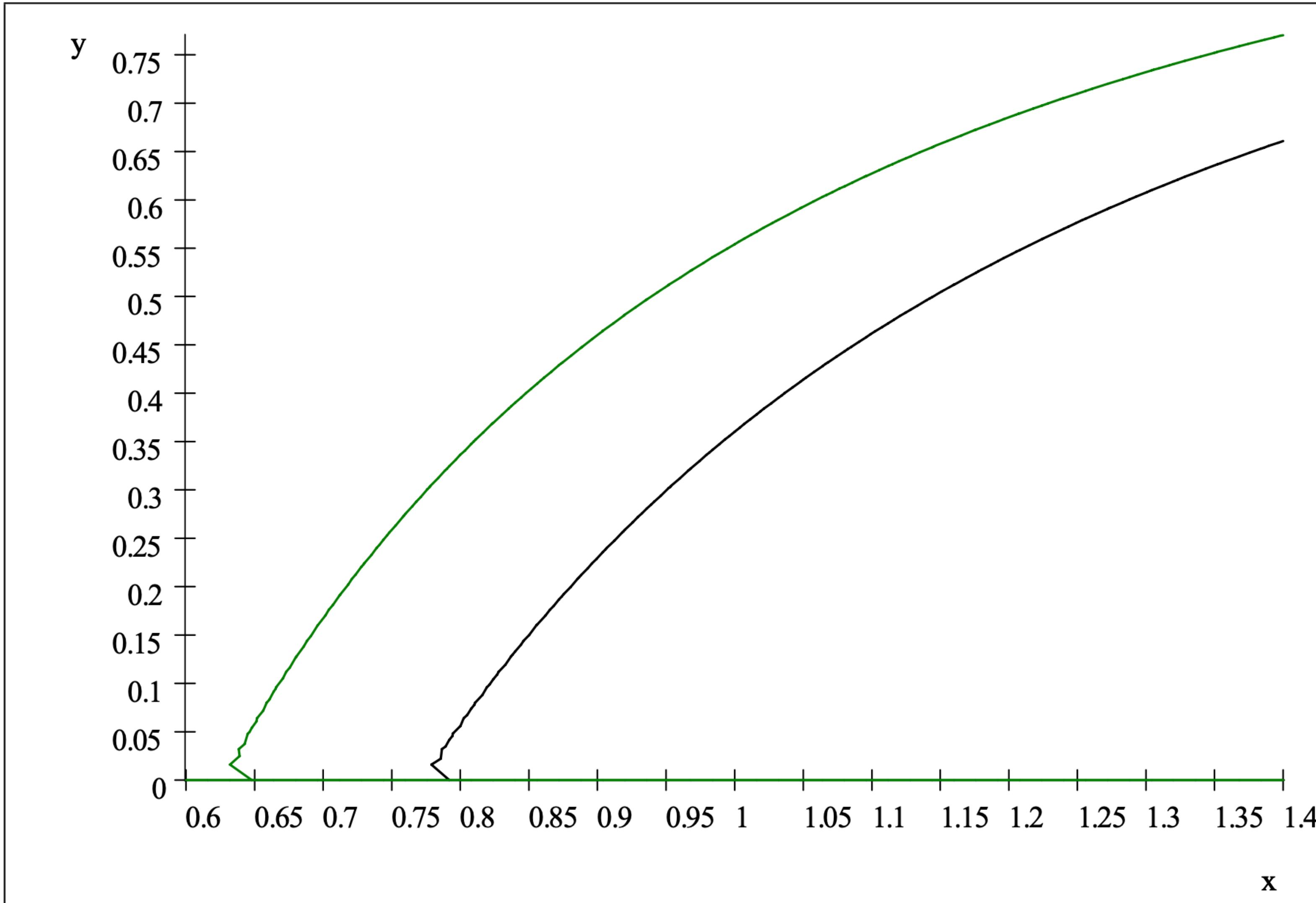
- to estimate c solve now : $\sum_{k \geq 1} \eta_k \mu_k = E \cdot \alpha$

Approximation for small attack rates and prevalences

For $\alpha < 0.5$ and a_H sufficiently small one can find a good approximate c or α value by solving:

$$(1 - e^{-\alpha c}) \left(1 + a_H \cdot e^{-\alpha c} \left(\frac{m_2}{E} - 1 \right) \right) = \alpha$$

Prevalence and out-household R : Poland versus Germany



$$(1 - e^{-x \cdot y}) \cdot 0.423 \cdot (1 + (1 - e^{-x \cdot y} \cdot (1 - 0.2(1 - e^{-x \cdot y})^1)) \cdot 0.322 \cdot 2 + (1 - e^{-x \cdot y} \cdot (1 - 0.2(1 - e^{-x \cdot y}))^2) \cdot 0.119 \cdot 3 + (1 - e^{-x \cdot y} \cdot (1 - 0.2(1 - e^{-x \cdot y}))^3) \cdot 4 \cdot 0.091 + (1 - e^{-x \cdot y} \cdot (1 - 0.2(1 - e^{-x \cdot y}))^4) \cdot 5 \cdot 0.035 = 2 \cdot y$$

Three countries compared

Country	Mean out-household reproduction factor	dark figure (calculated from the beginning of the pandemics)	Immunized population ratio due to past infection (pre-vaccine times)
Poland	0.89	~6	45%
Germany	0.83	~2.5	12%
Great Britain	0.88	~4.5	30%

Table 1. Characteristics of selected countries - the variables subject to this analysis were selected as having an impact on the number of deaths.

Prevalences in dependence of household size

Householdsize	Poland - 45 % prevalence: splits into household prevalences	Germany 12 % prevalence: splits into household prevalences
1	33%	9.5%
2	37%	11.2%
3	42%	12.9%
4	45%	14.6
5	49%	16.2
6	52%	-
7	56%	-

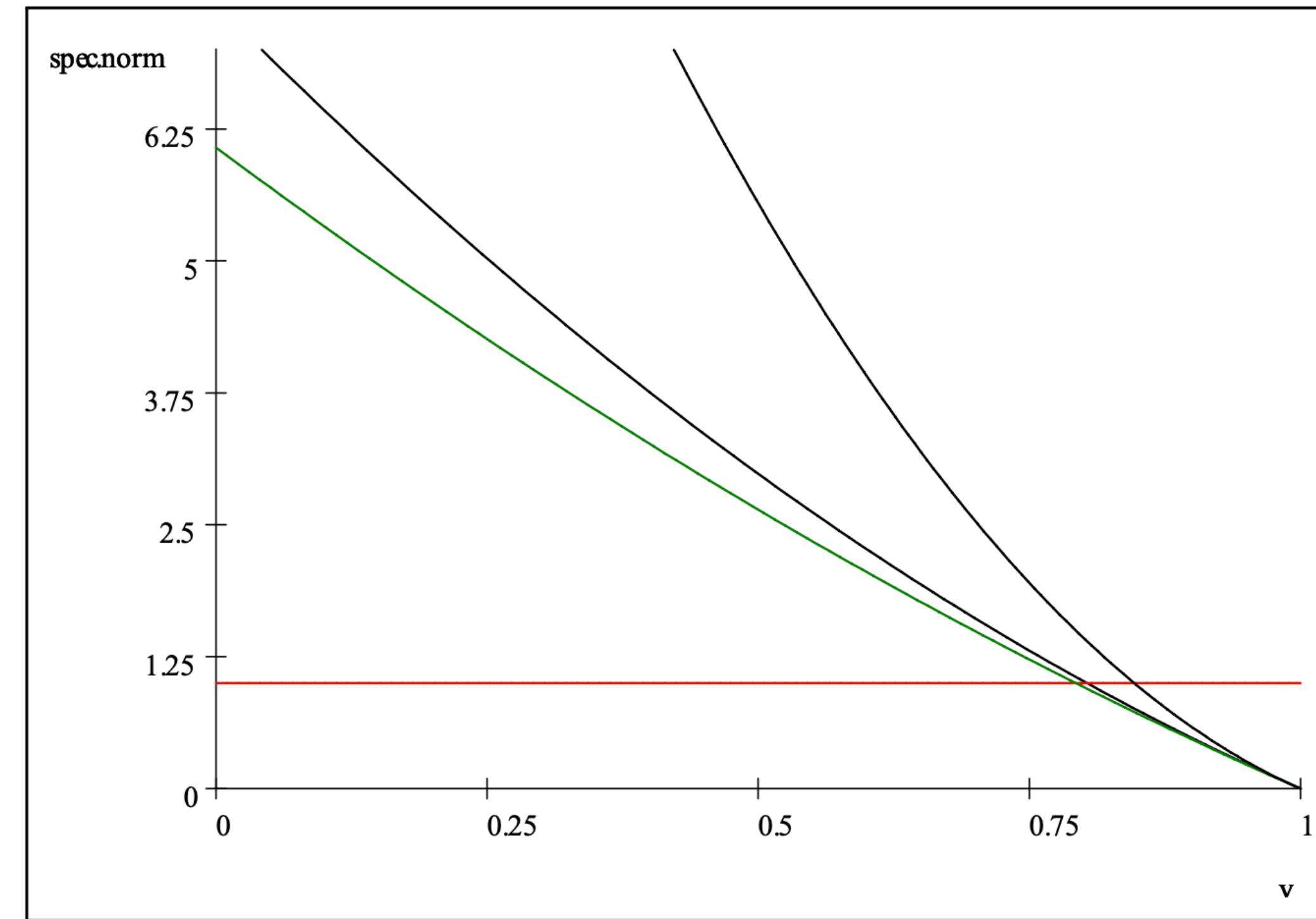
Virtual exchange of population and restrictions

Mean out-household reproduction factor (a measure for effectiveness of non-pharmaceutical intervention measures) calculated for ...	Modelled immunized population ratio due to past infection assuming household structure of...		
	...Poland	...Germany	...Great Britain
...Poland (0.89)	45%	21%	31%
...Germany (0.83)	38%	12%	22%
...Great Britain (0.88)	44%	20%	30%

Table 2. Illustration of the influence of the households structure on the immunized population ratio due to past infection (for a time just before starting distribution of vaccines on scale)

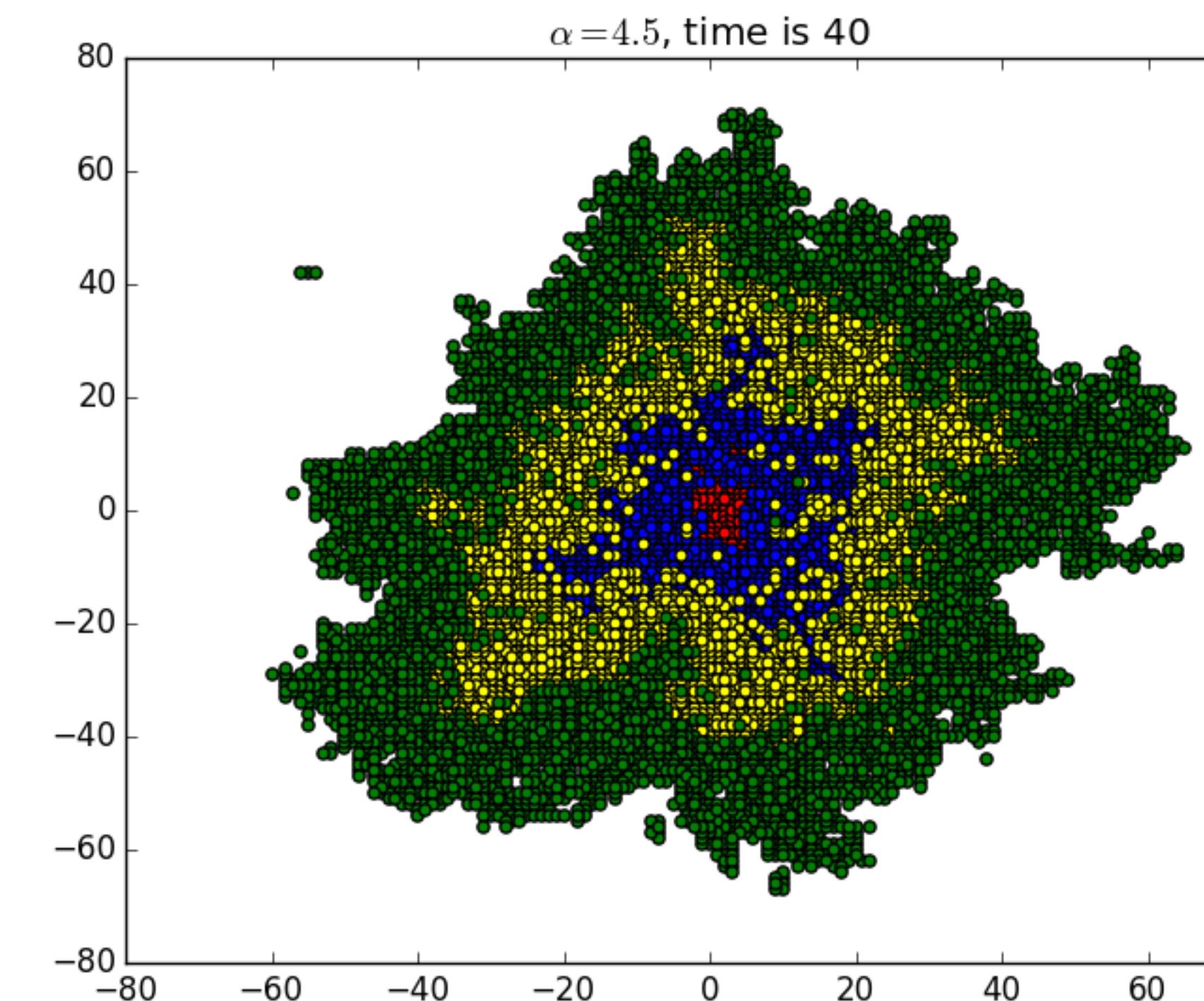
Application 1: vaccination

- fraction of population vaccinated : v
- vaccination uniform in population
- vaccination kernel
$$\kappa(x, y) = \frac{c}{E} (1 + a_H (1 - v)) (x - 1) (1 - v) y$$
- the spectral norm is given by
$$\|T_v\|_{sp} = c (1 - v) (1 + a_H (1 - v)) \left(\frac{m_2}{E} - 1 \right)$$



Timing and first passage percolation

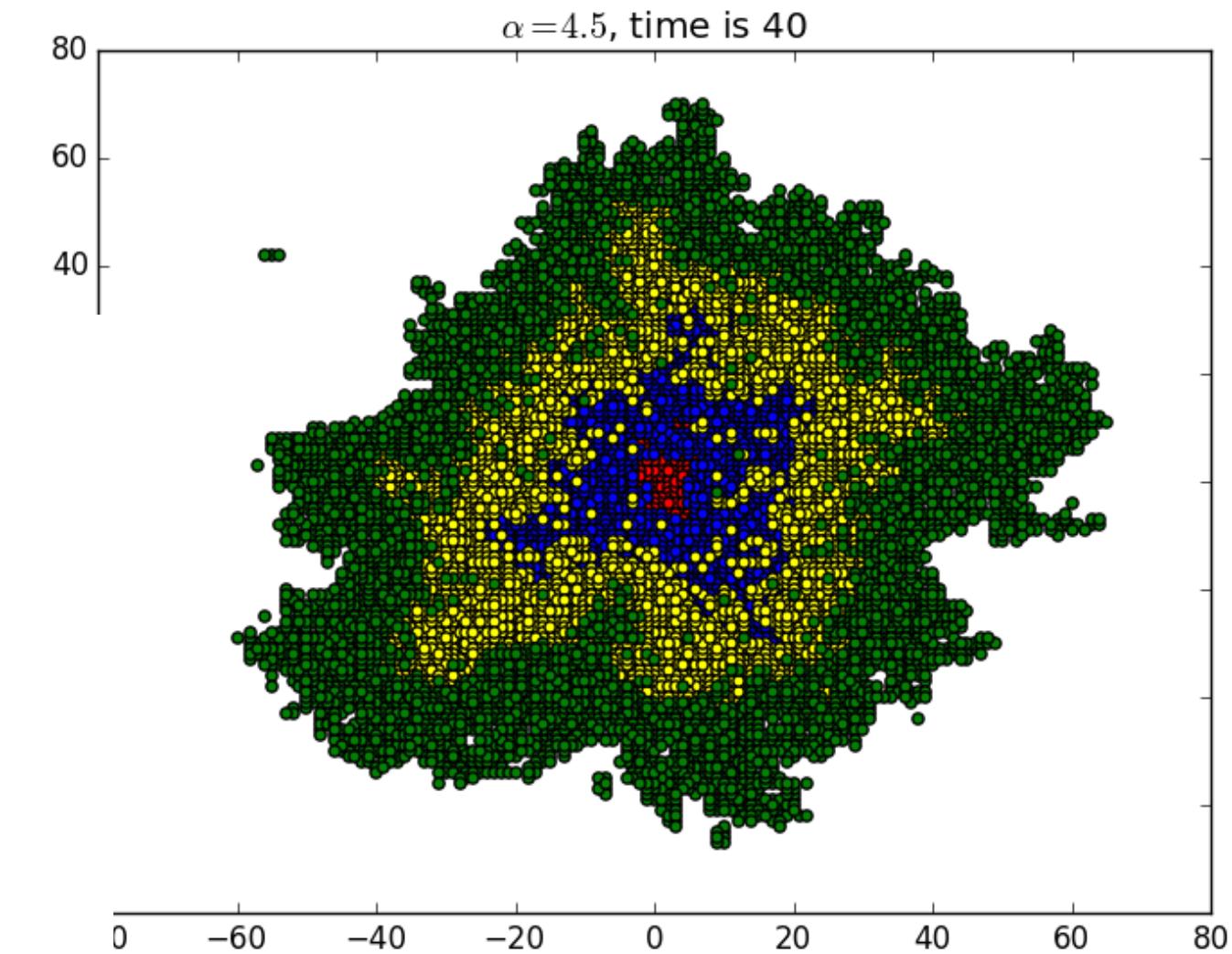
- assign "travel times" τ_{ij} to edges $i \sim j$, usually *iid* from a given distribution φ taking into account medical information of duration of incubation time and duration of infectious period (in epidemiology τ_{ij} are called serial intervals)
- interpretation: if node i gets infected at time t_i : node j (neighbor of i) gets infected at time $t_i + \tau_{ij}$
- analog to first passage percolation models
- $\{\tau_{ij}\}$ (or φ) together with a set of initial infected nodes I_0 defines then a time continuous stochastic infection process (typically non-Markov) on the random graph space
- the final size of an SIR infection is the size of the union of the connected components in which the initial infections I_0 are located.



Continuous time age dependent BP

- simplest case: single type Galton Watson process with "life time" distribution φ :
- Z_t : number of individuals alive at time t , and m expected offspring number
- Theorem :for $m > 1$: $Z_t \sim W \cdot e^{at}$ that is $\lim_{t \rightarrow \infty} Z_t e^{-at}$ converges to a limiting distribution W (pointwise) with

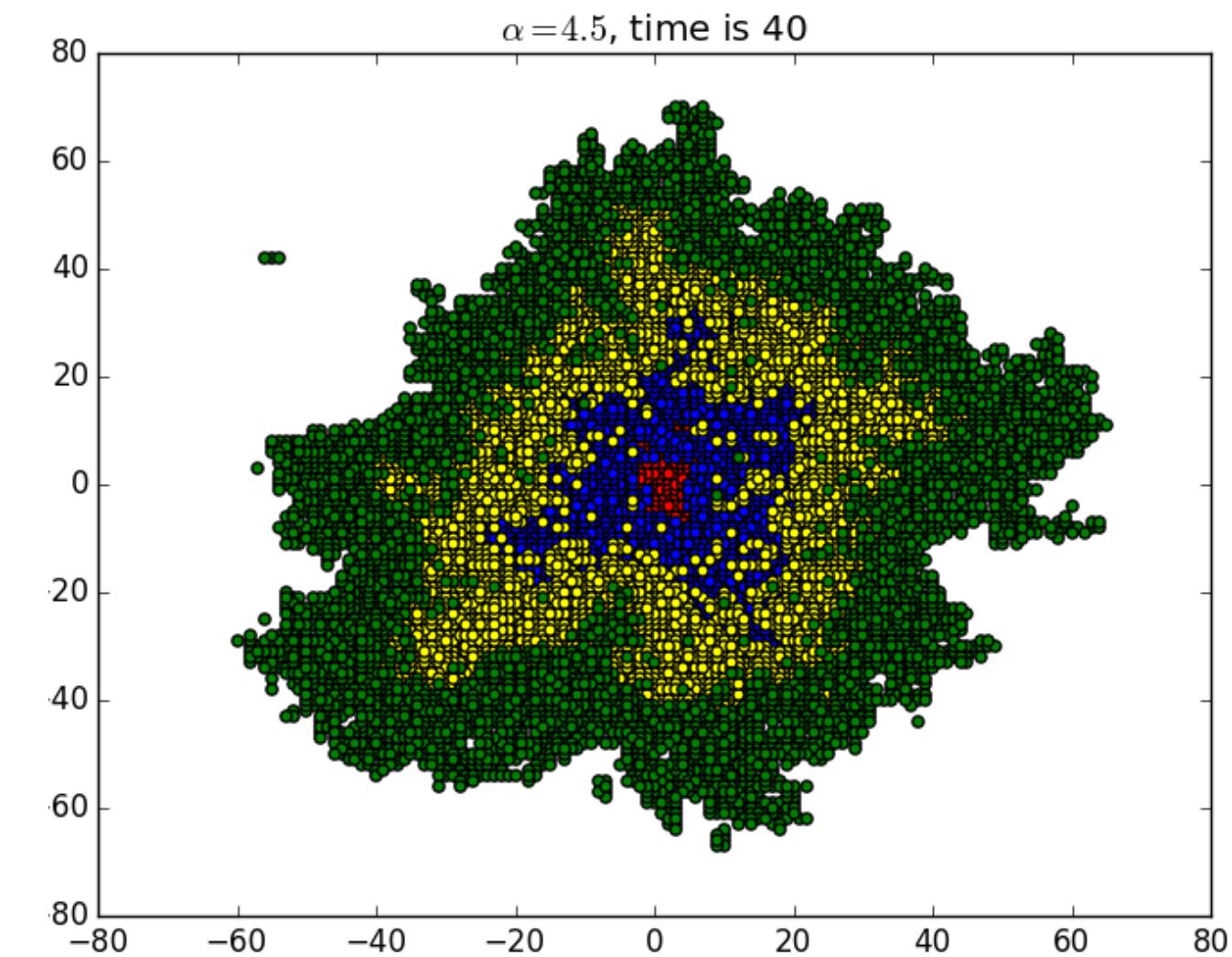
$$a \text{ solution of } m \int_0^{\infty} e^{-at} \varphi(t) dt = m \cdot \mathcal{L}_{\varphi}(a) = 1 \quad (9)$$



- Examples 1: $\varphi : Exp(\lambda) \Rightarrow a = \lambda(m - 1)$
- Example 2 : $\varphi : Gamma(k, \beta) \Rightarrow a = \beta(m^{1/k} - 1)$
- Example 3: $\varphi : U[0, \theta] \Rightarrow a = \frac{m\rho}{\theta}$ where ρ is the size of giant component in Erdos&Renyi graph with mean degree m
- Example 4: $\varphi : U[\theta, \theta + \epsilon] \Rightarrow a \sim \frac{1}{\theta} \log m$ for ϵ small

Continuous time multi type

- type dependent "life time" distribution $\varphi_{x,y}$ for directed links between type x and y individuals (serial intervals)
- operator $(T_a f)(x) = \int \kappa(x, y) \mathcal{L}_{\varphi_{x,y}}(a) f(y) d\mu(y) :$
 a solution of: $\|(T_a)\|_{sp} = 1$
- a is called the Maltusian parameter , $\mathcal{L}_{\varphi_{x,y}}(a)$: Laplace transform: $\int_0^\infty e^{-at} \varphi_{x,y}(t) dt$
- $\kappa(x, y)$ depends usually on time since contact structure might change in time
- Z_t : number of individuals "alive" at time t
- : $Z_t \sim W \cdot e^{at}$ that is $\lim_{t \rightarrow \infty} Z_t e^{-at}$ converges to a limiting distribution W (pointwise)



First passage percolation results for rank one kernels

Go to page 2

- iid travel times τ_{ij} to edges with finite second moments
- $L_n(i, j) =_{def}$ shortest travel time from i to j (if in same connected component) ; n number of nodes
- $H_n(i, j)$ = "Hopcount" = number of edges passed along the path which defines the shortest travel time $L(i, j)$
- Theorem (Bhamidi, Hofstad, Hooghiemstra 2018):

$$\left(\frac{H_n - \gamma \log n}{\sqrt{\beta \log n}}, L_n - \frac{\log n}{\alpha} \right) \xrightarrow{d} (Z, Q)$$

as $n \rightarrow \infty$ with $Z \sim \mathcal{N}(0, 1)$, Q continuos and Z and Q independent.

First passage percolation for directed BJR graphs

- independent but type dependent travel times $\tau_{ij} \sim \varphi_{x_i x_j}(\tau)$ to edges with finite second moments
- $L_n(i,j) =_{def}$ shortest travel time from i to j (if in same connected component) ; n number of nodes
- L_n travel time distribution for two random chosen nodes
- a maltusian parameter with respect to transfer operator (Tf) and $1 < \|Tf\| < \infty$
- Conjecture for directed BJR graphs (Bezborodov, Krueger 2021):

$$L_n - \frac{\log n}{a} \xrightarrow{d} Q$$

as $n \rightarrow \infty$ with Q :

$$Q = \frac{1}{a} (-\log W^+ - \log W^- - \Lambda + \text{const}) \quad (10)$$

- with W^+ limiting distribution of forward BP Z_t w.r to contact kernel $\kappa(x,y)$ and W^- backward BP limiting distribution w.r

Cumulative numbers

as a consequence one has $\lim_{n \rightarrow \infty} \mathbb{E} Z_{cum} \left(t + \frac{\log n}{a} \right) = \Pr(Q \leq t)$
conjecture : almost sure convergence of $Z_{cum} \left(t + \frac{\log n}{a} \right)$

End of Part I