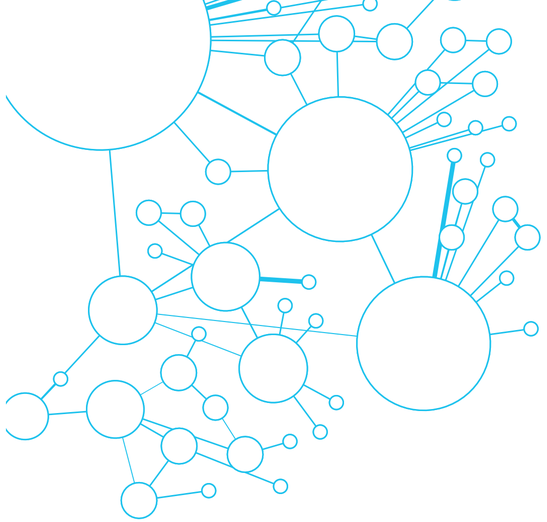
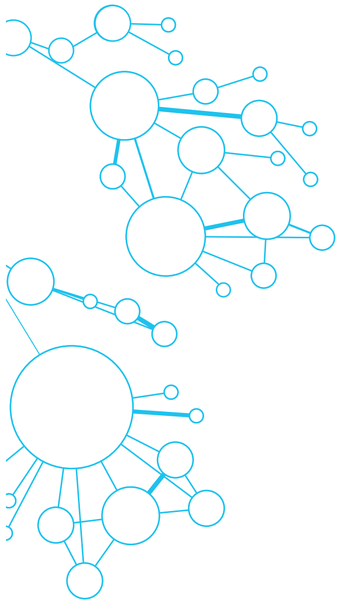




Northeastern

Epydemix:
open-source platform for stochastic epidemic modeling, simulation, and forecasting.

EPISTORM

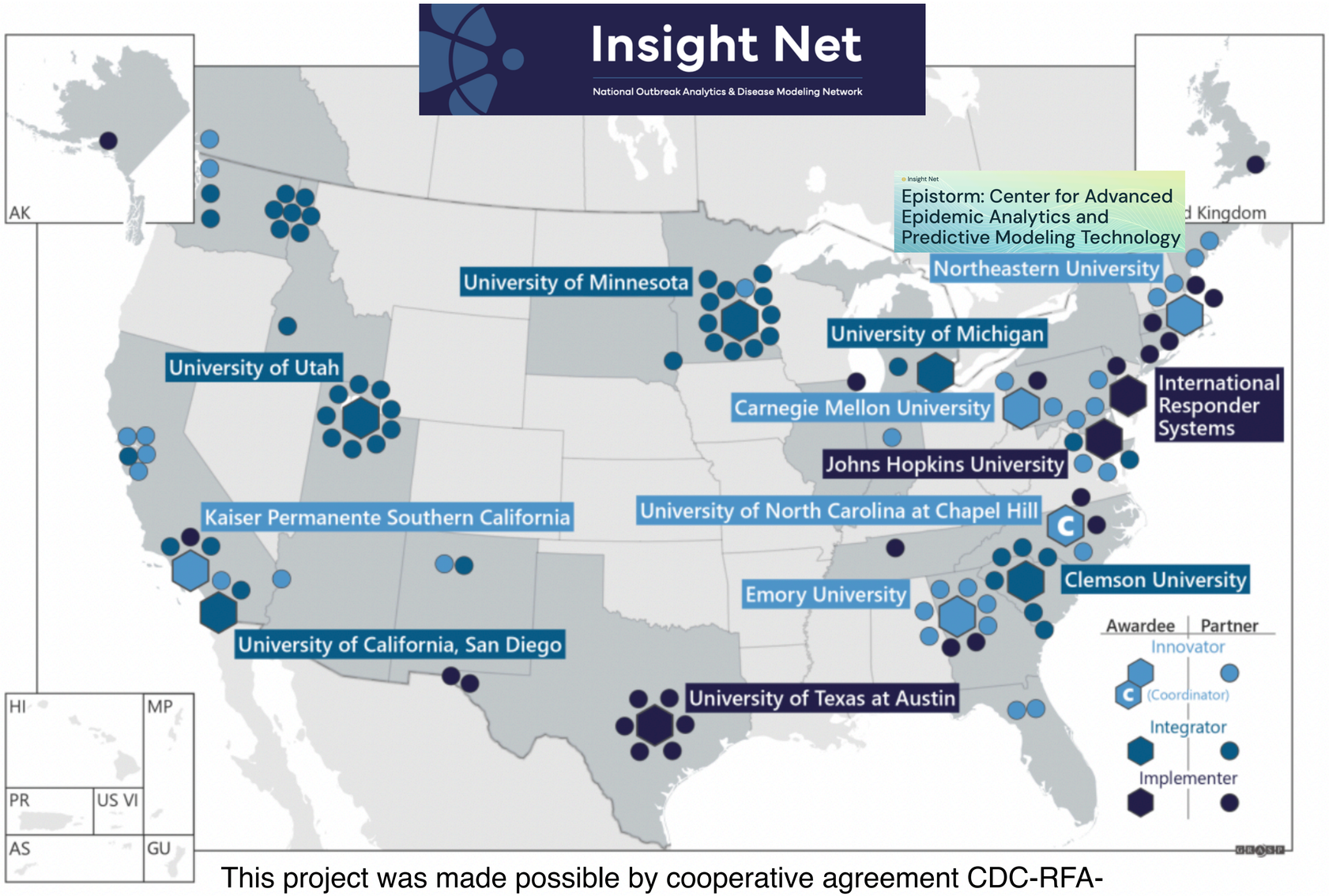
MOBS LAB

LABORATORY FOR THE MODELING OF BIOLOGICAL AND SOCIO-TECHNICAL SYSTEMS

@alexvespi

Introducing the team:

Jessica Davis,
Nicolo Gozzi,
Minami Ueda,
Alessandro Vespignani



This project was made possible by cooperative agreement CDC-RFA-FT-23-0069 from the CDC's Center for Forecasting and Outbreak Analytics.

- Establishment of National centers:
- Sustainability
- Develop, deploy, and maintain data/modeling technology
- Next-generation data pipeline
- Advance communications
- Engage/dialogue decision makers, government, private sector, and the public at large

Insight Net

Epistorm: Center for Advanced Epidemic Analytics and Predictive Modeling Technology

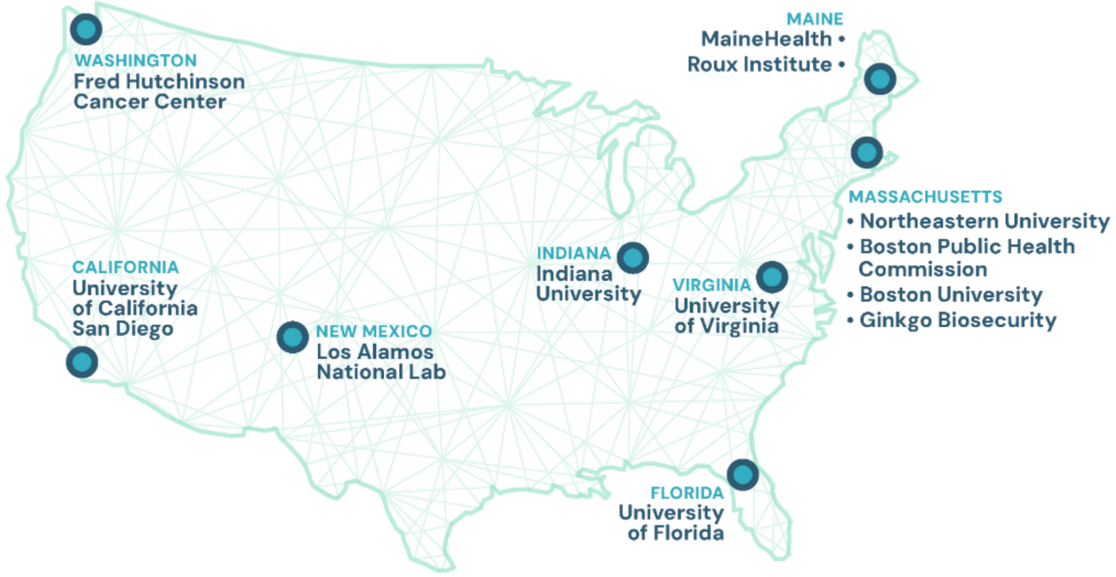
Epistorm is an innovation center of **InsightNet**, the national network for outbreak and disease modeling established by the CDC's Center for Forecasting & Outbreak Analytics (CFA) in 2023. As a consortium of research institutions, healthcare systems and private companies we develop innovative methodologies to integrate high-resolution mobility, airline travel, genomic and wastewater surveillance data—with mechanistic, statistical, and deep learning forecasting models to increase the accuracy of predictive analytics to help the U.S. make more informed decisions during future outbreaks of infectious diseases.

EPISTORM CONSORTIUM

- Northeastern University
- Boston Public Health Commission
- Boston University
- Fred Hutchinson Cancer Center
- Ginkgo Biosecurity
- Indiana University
- Los Alamos National Laboratory
- MaineHealth
- University of California San Diego
- University of Florida
- University of Virginia

PUBLIC HEALTH COLLABORATORS

- New Hampshire DHHS
- Maine DHHS
- Council of State and Territorial Epidemiologists



- **Epidemic modeling as public service:** Establish epidemic modeling infrastructure analogous to weather forecasting, providing continuous uncertainty-quantified predictions that inform public health decision-making before, during, and between outbreaks.
- **Collaborative modeling networks:** enable distributed teams of researchers and practitioners to contribute models, validate predictions against common benchmarks, and share methodological advances through standardized computational frameworks
- **Data-to-insight pipeline integration:** connect real-time surveillance data streams (emergency department visits, wastewater, mobility, clinical reporting) to mechanistic models with automated calibration frameworks like ABC-SMC for operational readiness.
- **Open-source ecosystem for epidemic intelligence:** build on the Epydemix platform to create modular, interoperable tools that enable rapid model development, calibration, and deployment across diverse disease systems and geographic scales

EPISTORM

Software

Epistorm-Mix

Dataset containing contact patterns for all ages of the US population collected during the post-pandemic era from a statistically representative sample.

epydemix

Open-source Python package designed for flexible, modular, and data-driven epidemic modeling.

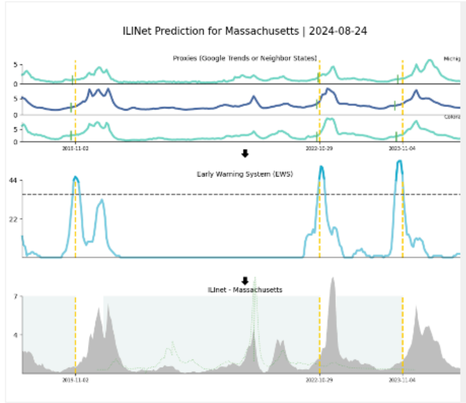
SummRt

R package and software to compare estimators of the time varying reproductive number; built in collaboration with the EpiForeSite center.

Reproductive number estimation R packages

Package to estimate the reproductive number at a spatially granular level with mobility data and also separately to incorporate reporting delays.

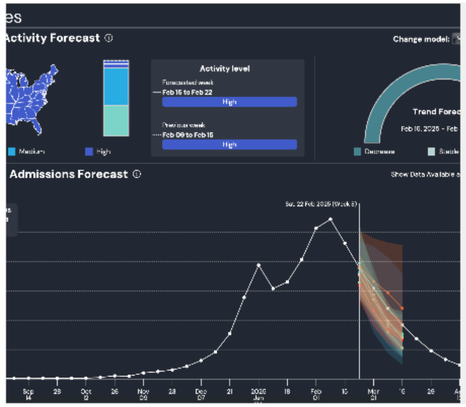
Platforms



Flu Early Warning System

Our real-time early warning systems have been able to identify future surges in hospitalizations 2 to 6 weeks before they are reported by traditional disease surveillance systems.

Launch site



Flu Forecast

One-stop dashboard for the visualization and analysis of the model contributed by Epistorm to the CDC Flusight forecasting program.

Launch site

Epistorm Mobility Dashboard

The dashboard provides public health officials, epidemic modelers, and researchers with continuous access to high-resolution mobility data critical for disease transmission analysis and outbreak forecasting.

Python Package

epydemix

An open-source Python package designed for flexible, modular, and data-driven epidemic modeling.

Epydemix supports the full modeling pipeline—from constructing stochastic compartmental models to running simulations, integrating real-world data, and calibrating parameters. Users can incorporate age-structured contact patterns, dynamic interventions, and population demographics with ease. Built-in Approximate Bayesian Computation (ABC) methods enable robust parameter estimation and model fitting, supporting forecasting, scenario exploration, and policy-relevant analyses. Epydemix bridges the gap between theoretical modeling and practical application, helping researchers and public health professionals translate models into actionable insights.

Get started ›

Tools

epyForecast

A desktop app to run short-term forecasts using real data, powered by Epydemix.

Details on how to download EpyForecast will be available soon.

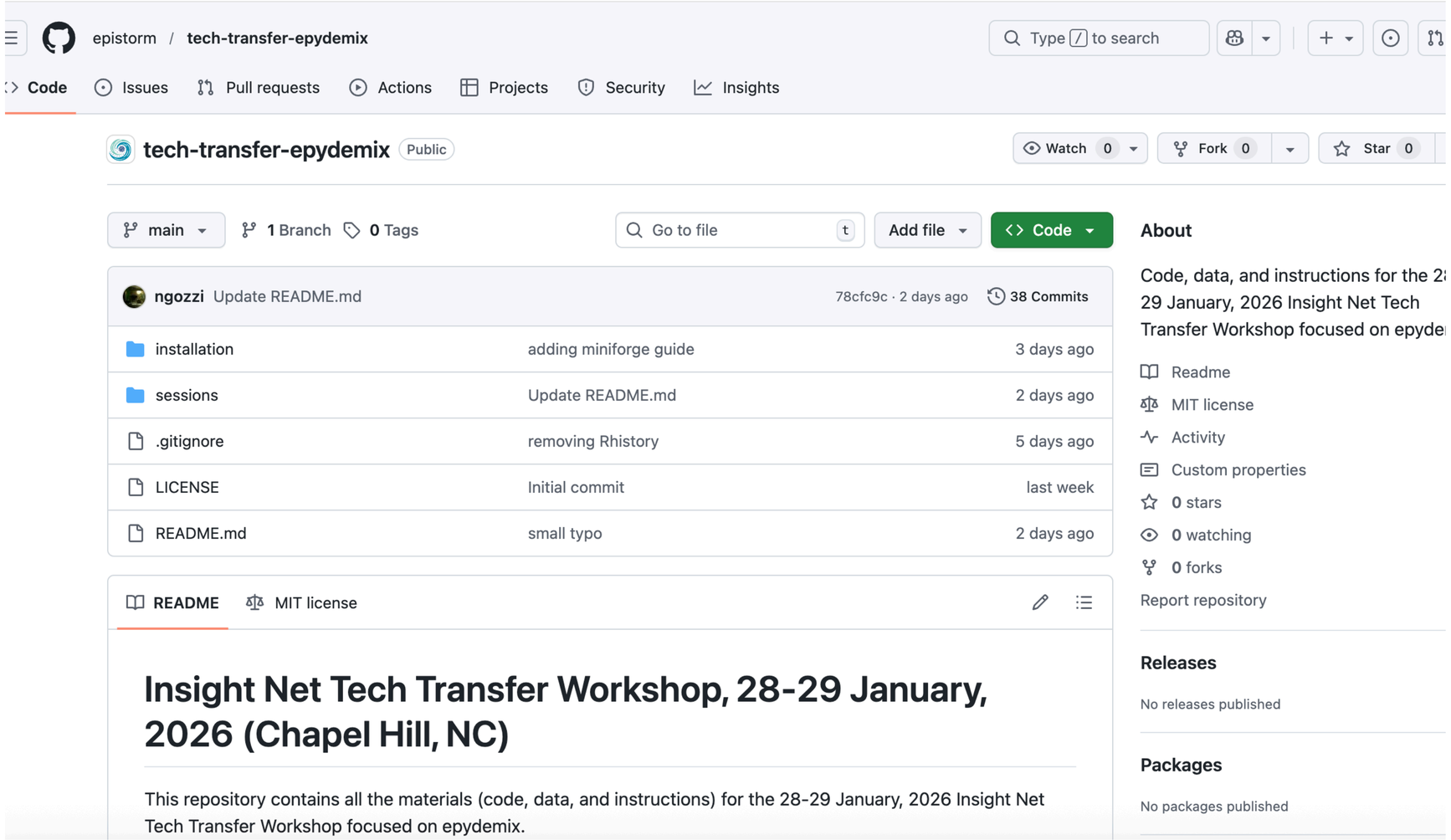
epyScenario

A web app to explore epidemic scenarios, powered by Epydemix.

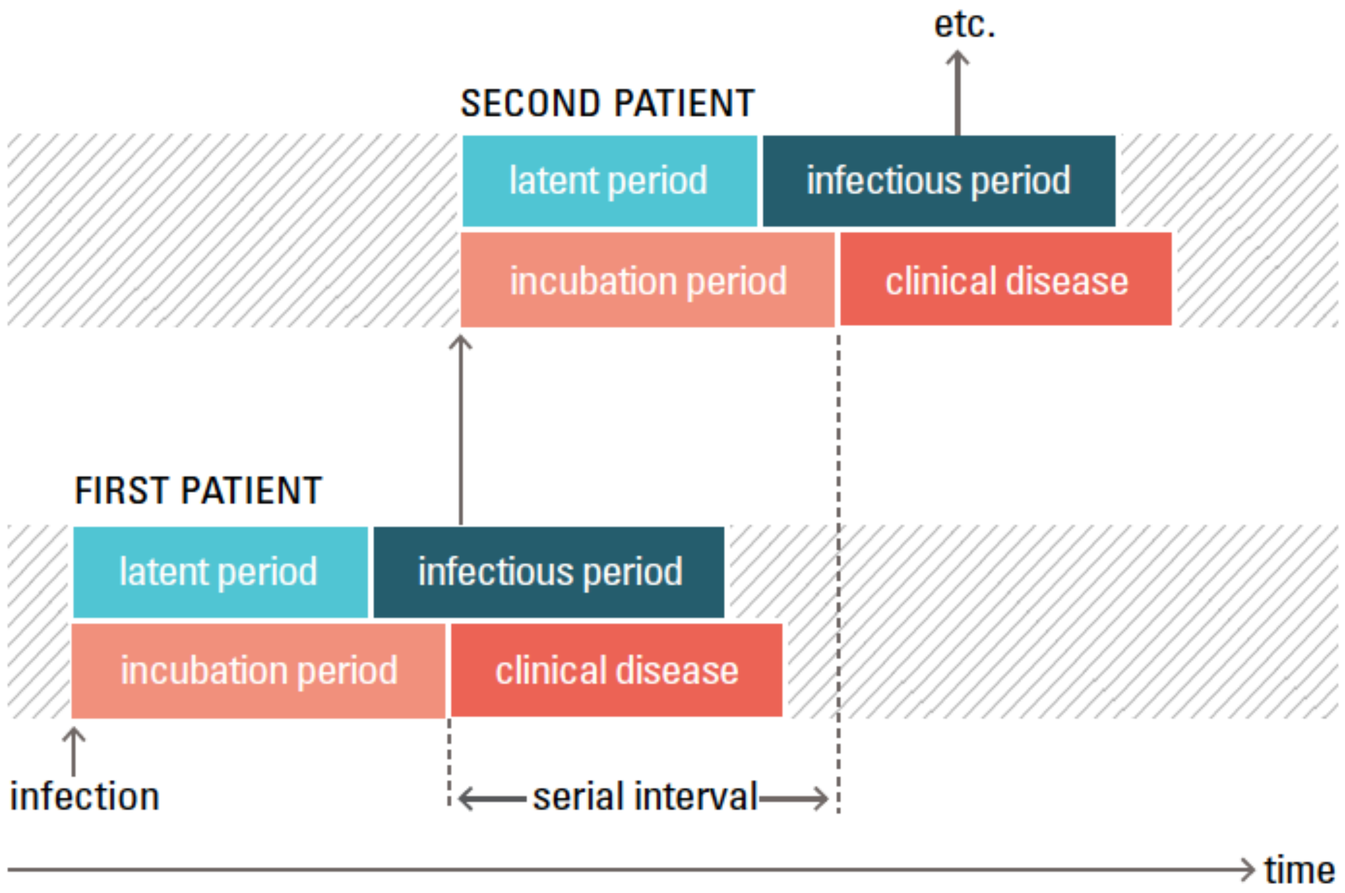
Launch ›

https://www.epydemix.org/

https://github.com/epistorm/tech-transfer-epydemix/tree/main



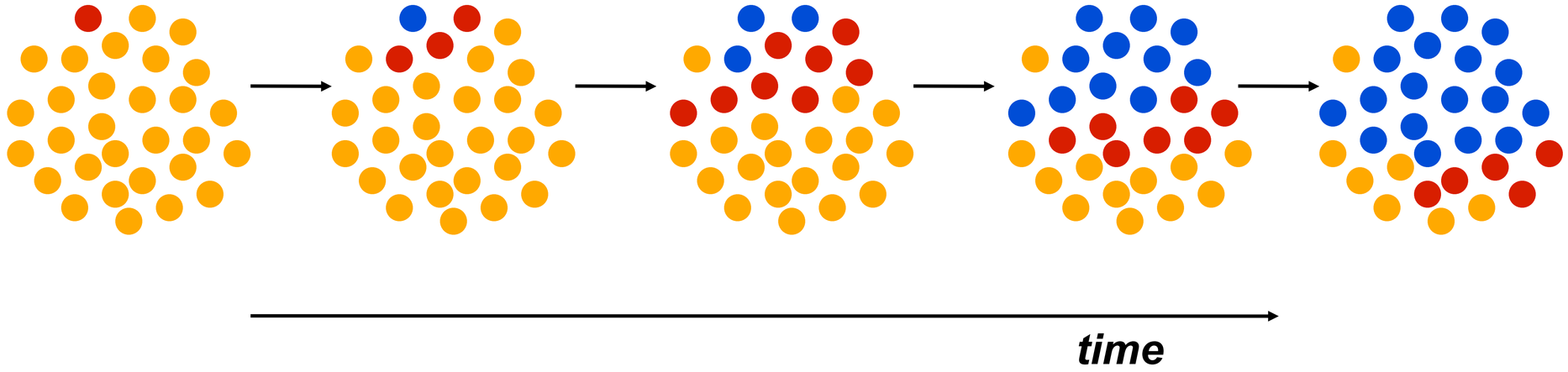
COMPARTMENTAL EPIDEMIOLOGICAL CHARACTERIZATION



Stochastic and discrete modeling of epidemics….

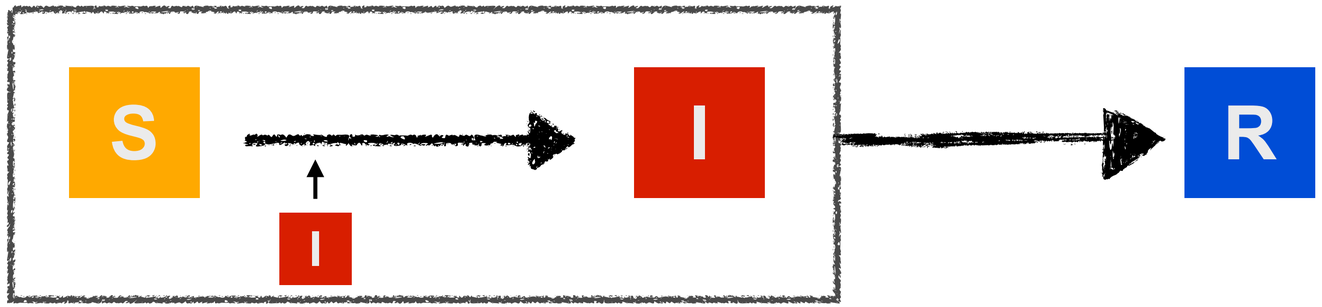
SIR model

dynamics



MATHEMATICAL EPIDEMIOLOGY CIRCA 1766

Based on the disease compartmental structure: Individuals are characterized by the disease stage: susceptible, infectious, recovered, exposed etc.etc



S = number of susceptible individuals
I = number of infectious individuals
R = number of recovered individuals

N = total population

$$N = S + I + R$$

Homogenous assumption

Transmission rate (force of infection) λ_t for each susceptible:

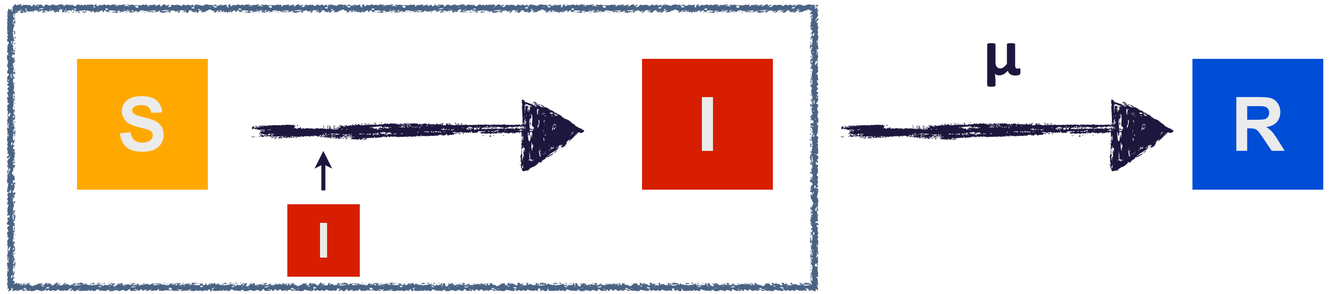
transmission rate x effective number of contacts per unit time

x proportion of contacts infectious

$$\lambda_t = \beta \, I / N$$

MOBS LAB

Disease dynamic transitions



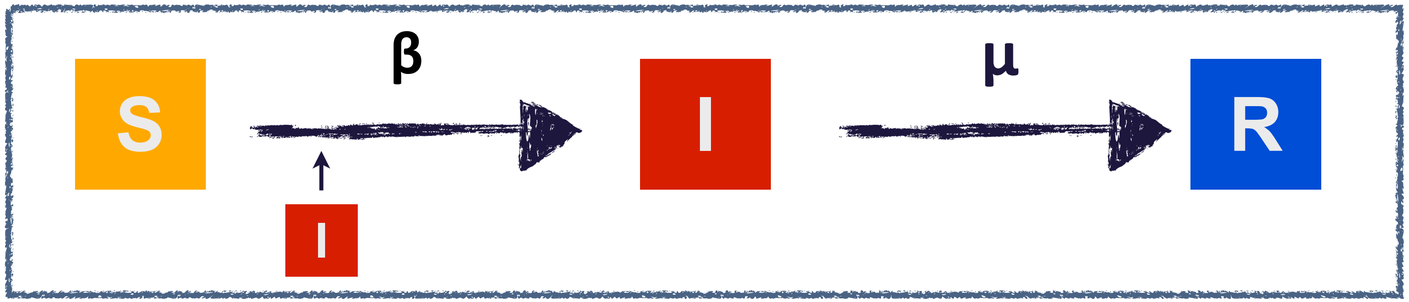
prob. of infection

$$P_{S \to I} = 1 - \exp(-\lambda_t \delta t) \simeq \lambda_t \delta t$$

prob. of recovery

$$P_{I \to R} = 1 - \exp(-\mu \delta t) \simeq \mu \delta t$$

Average dynamic



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

assuming $\delta t = 1$

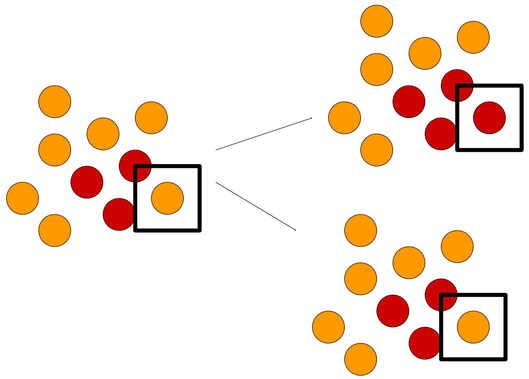
$$I_{t+1} = I_t + \beta \frac{I_t}{N} S_t - \mu I_t$$

$$R_{t+1} = R_t + \mu I_t$$

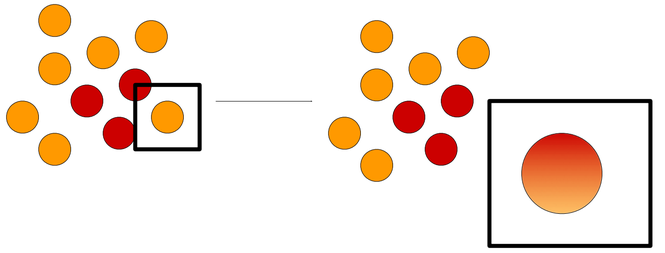
Average equations

- No stochastic effects => determinism

Different possibilities
Randomness



- Continuous variables => $s(t) = 1.5$

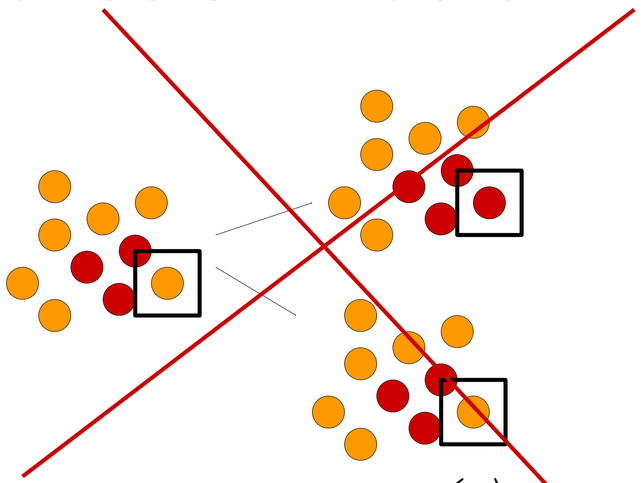


half / half

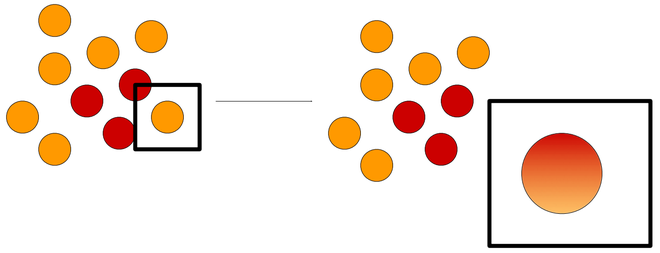
Average equations

- No stochastic effects => determinism

Different possibilities
Randomness

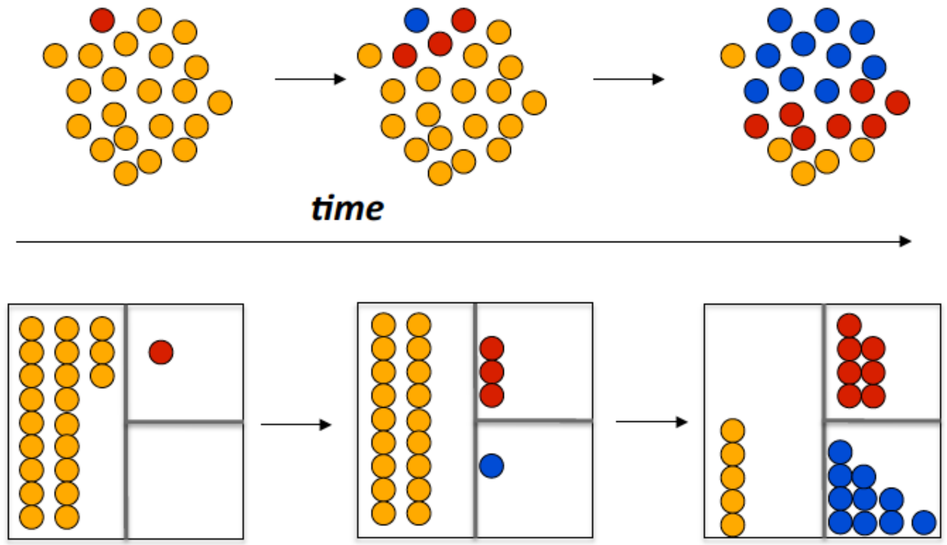


- Continuous variables => $s(t) = 1.5$



half / half

Stochastic dynamic



Binomial processes

Number of trials It , St
Probability of success:

$$\lambda_t = \beta I/N \quad ; \mu$$

$$\begin{aligned} S_{t+1} &= S_t - Bin(S_t, \lambda_t) \\ I_{t+1} &= I_t + Bin(S_t, \lambda_t) - Bin(I_t, \mu) \\ R_{t+1} &= R_t + Bin(I_t, \mu), \end{aligned}$$

Chain binomial processes

Basic reproductive number

R_0 is the average number of individuals infected directly by an infected individuals during his infectious period in a fully susceptible population.

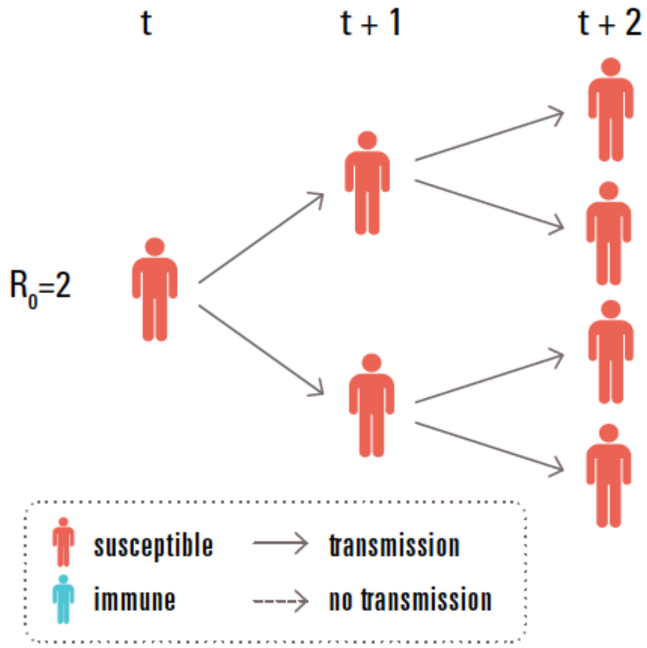
β is the average number of infections generated in a unitary time step

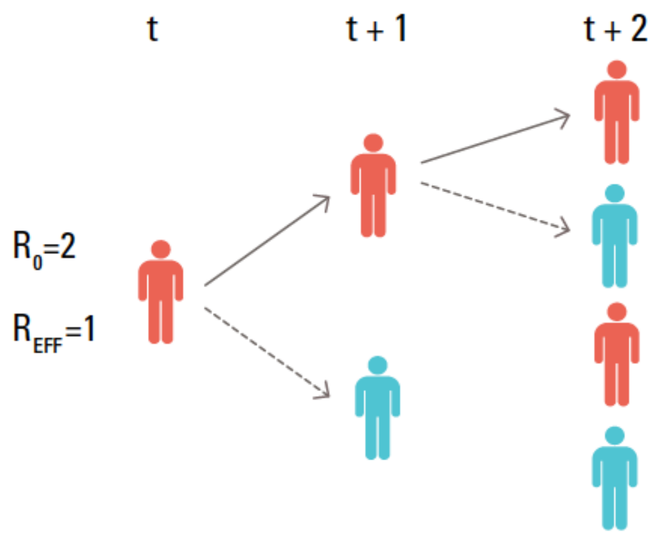
$\tau = 1/\mu$ is the average number an infectious individual stays infected

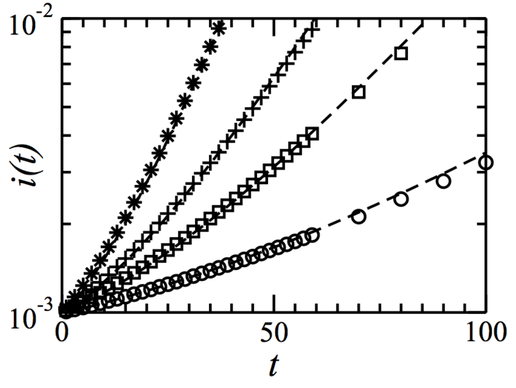
$$R_0 = \beta\tau = \frac{\beta}{\mu}$$

Basic reproductive number R_0

R_0 is the average number of individuals infected directly by an infected individuals during his infe
a fully susceptible population.







Exponential growth

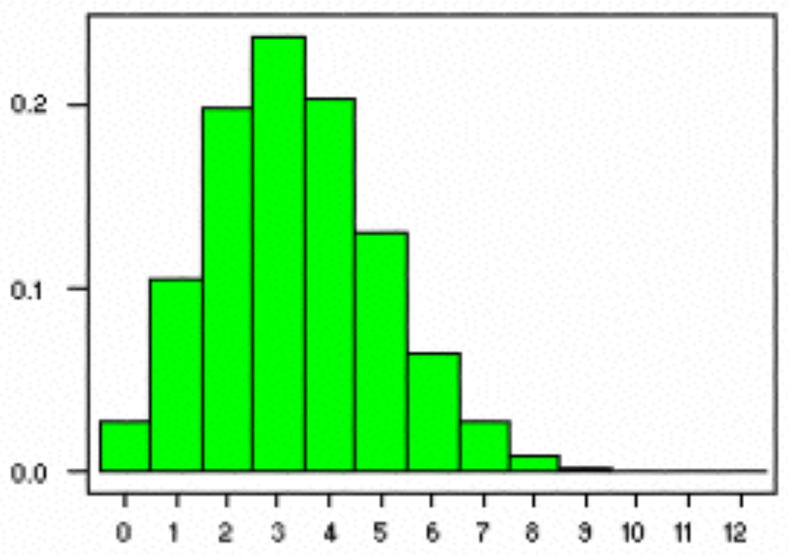
MOBS LAB

BINOMIAL DISTRIBUTION

$$f(x_i=k) = \binom{n}{k} p^k (1-p)^{n-k}$$

- Number of trials *n*
- Probability of success *p*

Binom(n,p)

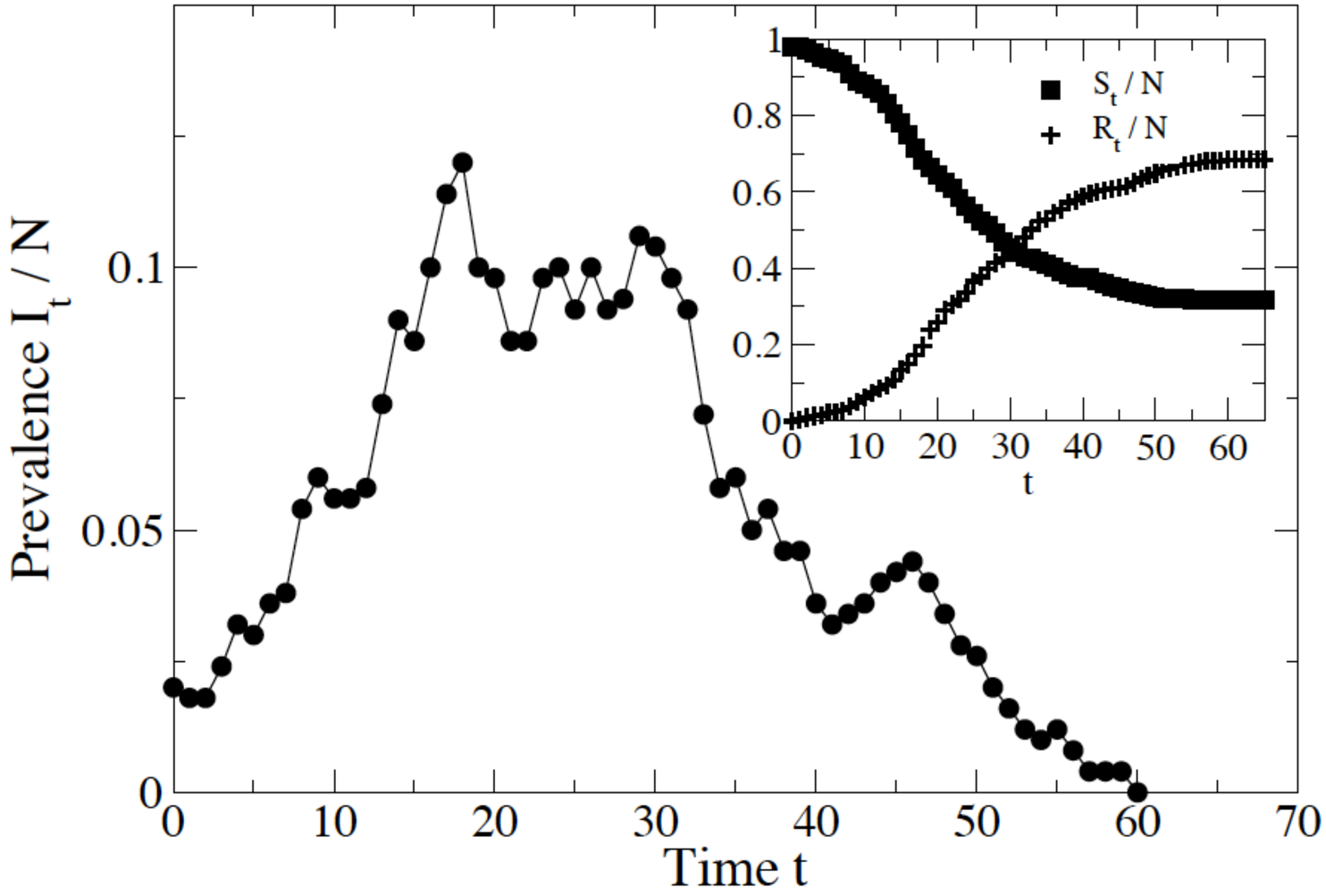


stochastic SIR model

```
time=0
   S=990   I=10   R=0

time=1
    S->I:    p=0.004000 S=990  -->  DS=5
    I->R:    p=0.200000 I=10   -->  DR=3
    S=985   I=12   R=3

time=2
    S->I:    p=0.004800 S=985  -->  DS=3
    I->R:    p=0.200000 I=12   -->  DR=2
    S=982   I=13   R=5

time=3
    S->I:    p=0.005200 S=982  -->  DS=2
    I->R:    p=0.200000 I=13   -->  DR=2
    S=980   I=13   R=7

time=4
    S->I:    p=0.005200 S=980  -->  DS=2
    I->R:    p=0.200000 I=13   -->  DR=3
    S=978   I=12   R=10

time=5
    S->I:    p=0.004800 S=978  -->  DS=4
    I->R:    p=0.200000 I=12   -->  DR=1
   S=974   I=15   R=11
```

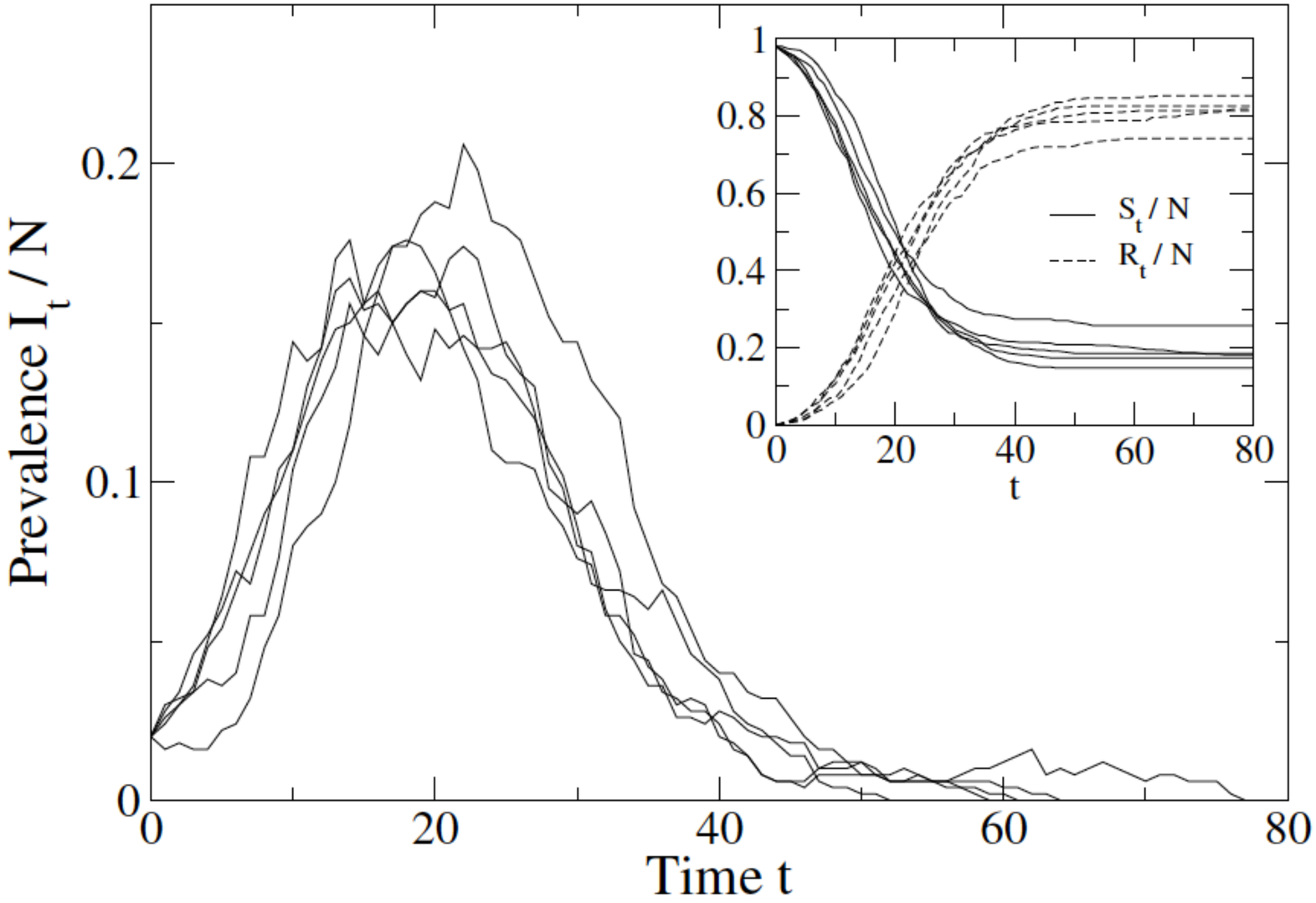
```
time=0
    S=990   I=10   R=0

time=1
    S->I:    p=0.004000 S=990  -->  DS=3
    I->R:    p=0.200000 I=10   -->  DR=2
    S=987   I=11   R=2

time=2
    S->I:    p=0.004400 S=987  -->  DS=7
    I->R:    p=0.200000 I=11   -->  DR=3
    S=980   I=15   R=5

time=3
    S->I:    p=0.006000 S=980  -->  DS=6
    I->R:    p=0.200000 I=15   -->  DR=3
    S=974   I=18   R=8

time=4
    S->I:    p=0.007200 S=974  -->  DS=7
    I->R:    p=0.200000 I=18   -->  DR=3
    S=967   I=22   R=11

time=5
    S->I:    p=0.008800 S=967  -->  DS=7
    I->R:    p=0.200000 I=22   -->  DR=6
    S=960   I=23   R=17
```

Single epidemic



MOBS LAB

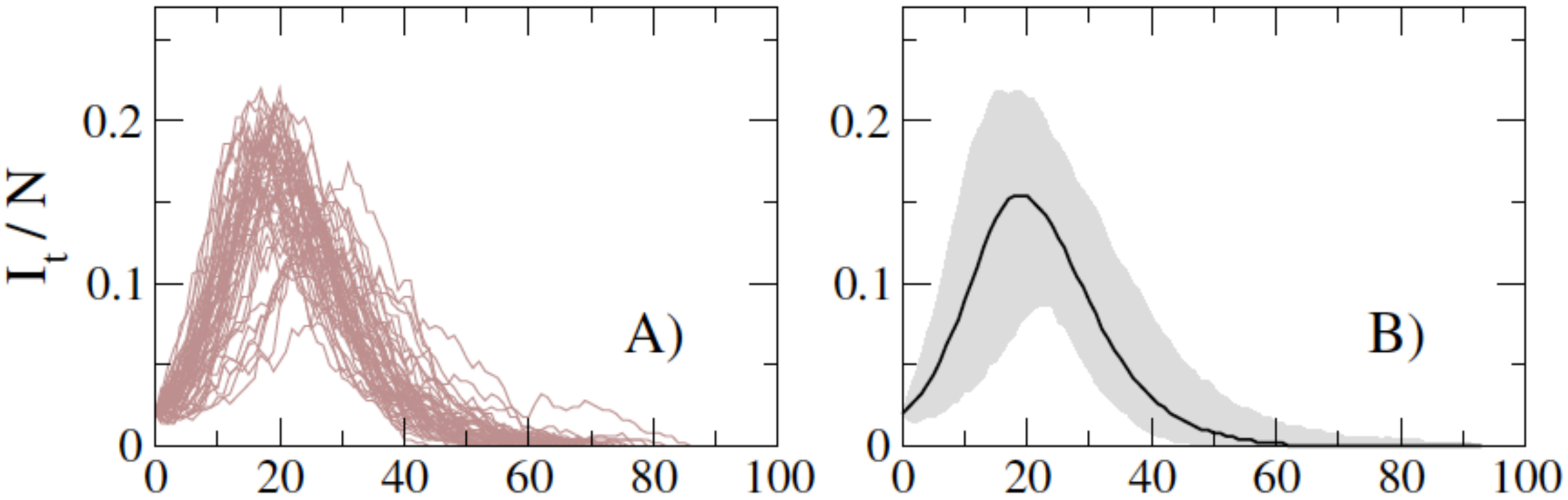
Multiple epidemic



MOBS LAB

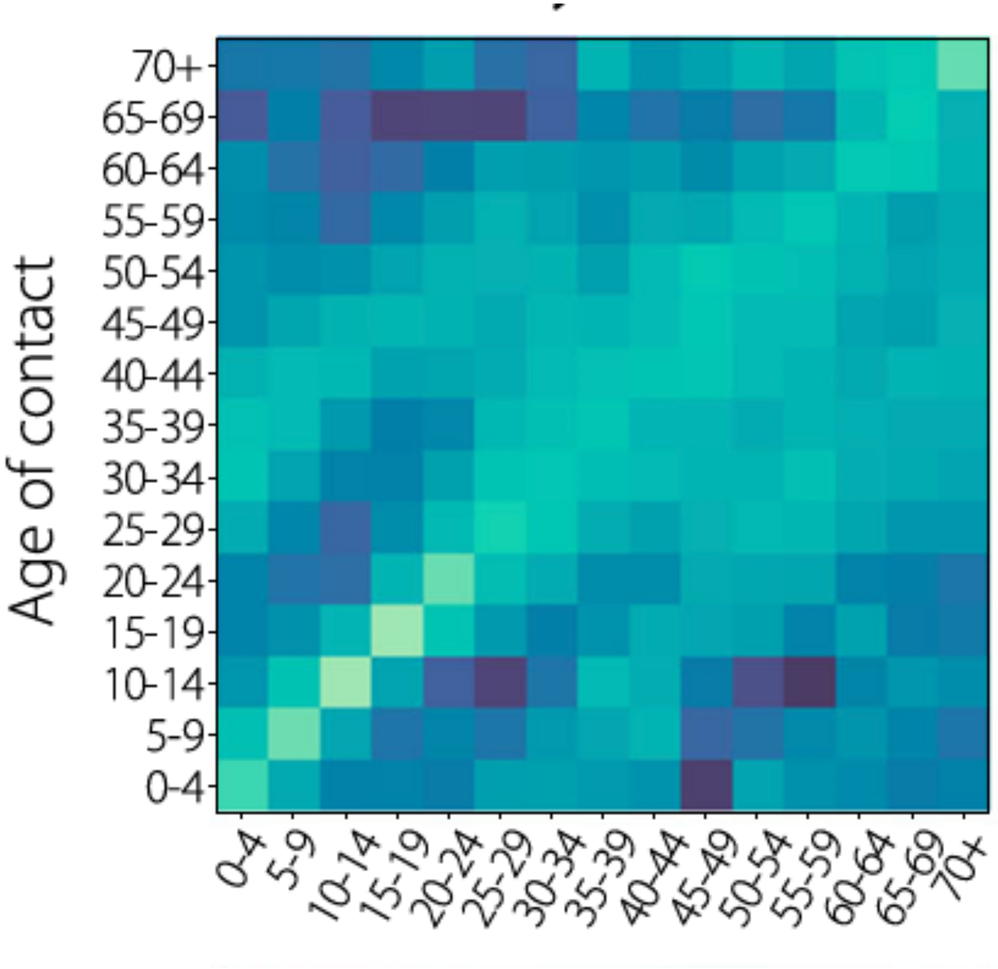
Multiple epidemic

Multiple epidemic



Networks: the invisible fabric of a population

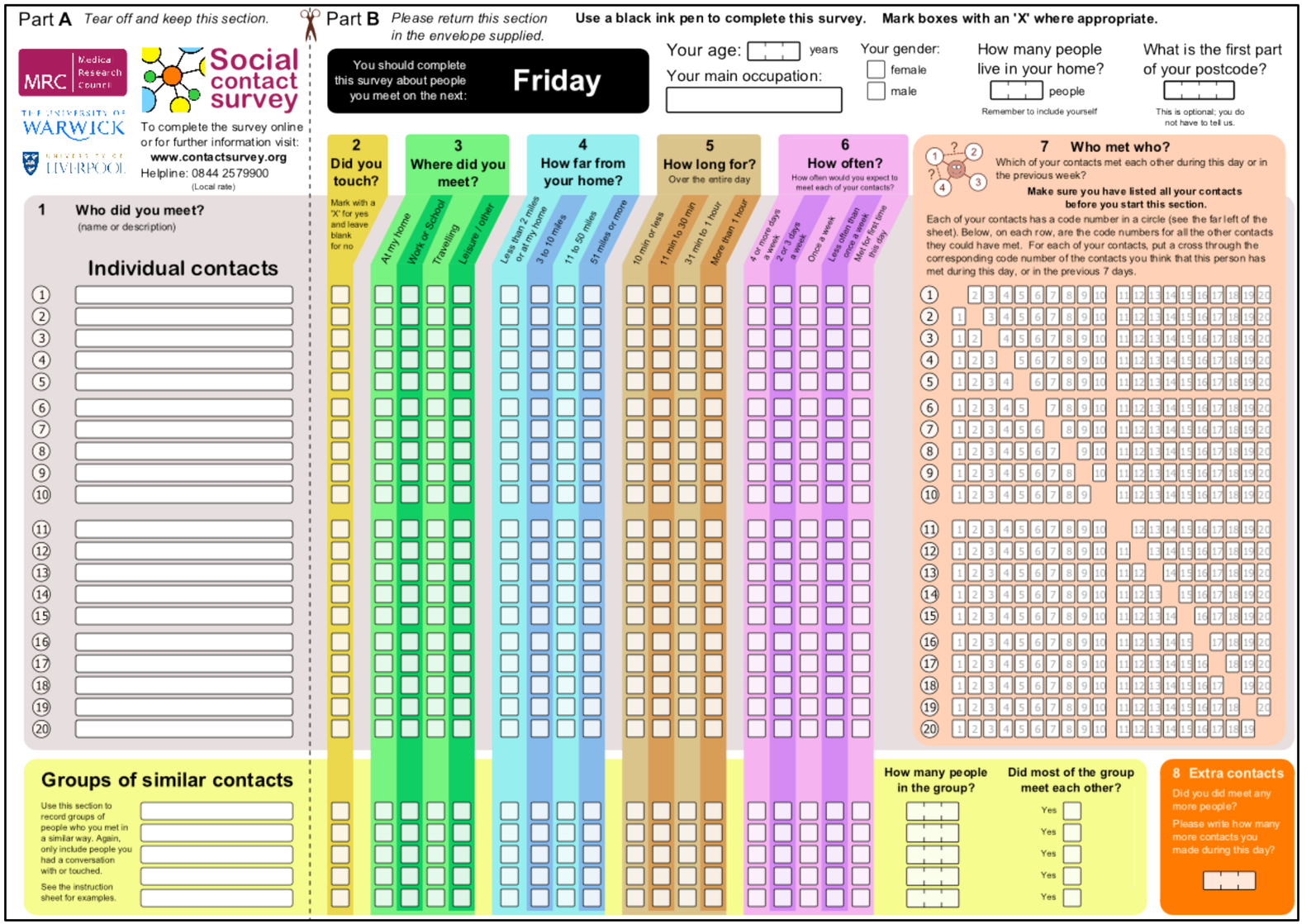
Effective Contact Matrices in Epidemic Modeling



- Quantify interactions between population groups (e.g., by age) that can lead to disease transmission.
- Incorporate heterogeneity in contact patterns, capturing who mixes with whom and the frequency of these interactions.
- Constructed using demographic data (e.g., age distribution, household composition, school enrollment, workplace participation) combined with behavioral assumptions about mixing patterns.
- Derived from empirical data, such as contact diaries, time-use surveys, wearable sensors, or mobile phone mobility data, which capture the frequency and context of interpersonal interactions across population subgroups (e.g., by age or setting).

MOBS LAB

manually



Part A *Tear off and keep this section.*

Part B *Please return this section in the envelope supplied.*

Use a black ink pen to complete this survey. Mark boxes with an 'X' where appropriate.

Medical Research Council

THE UNIVERSITY OF WARWICK

UNIVERSITY OF LIVERPOOL

Social contact survey

To complete the survey online or for further information visit:
www.contactsurvey.org
Helpline: 0844 2579900
(Local rate)

You should complete this survey about people you meet on the next: **Friday**

Your age: ☐☐ years

Your main occupation:

Your gender: ☐ female ☐ male

How many people live in your home? ☐☐ people
Remember to include yourself

What is the first part of your postcode?
This is optional; you do not have to tell us.

1 Who did you meet?
(name or description)

Individual contacts

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20

2 Did you touch?
Mark with a 'X' for yes and leave blank for no

3 Where did you meet?
At my home
Work or School
Travelling
Leisure / other

4 How far from your home?
Less than 2 miles or at my home
3 to 10 miles
11 to 50 miles
51 miles or more

5 How long for?
Over the entire day
10 min or less
11 min to 30 min
31 min to 1 hour
More than 1 hour

6 How often?
How often would you expect to meet each of your contacts?
4 or more days a week
2 or 3 days a week
Once a week
Less often than once a week
Met for first time this day

7 Who met who?
Which of your contacts met each other during this day or in the previous week?

Make sure you have listed all your contacts before you start this section.

Each of your contacts has a code number in a circle (see the far left of the sheet). Below, on each row, are the code numbers for all the other contacts they could have met. For each of your contacts, put a cross through the corresponding code number of the contacts you think that this person has met during this day, or in the previous 7 days.

Groups of similar contacts

Use this section to record groups of people who you met in a similar way. Again, only include people you had a conversation with or touched.

See the instruction sheet for examples.

	How many people in the group?	Did most of the group meet each other?
	☐☐	Yes ☐
	☐☐	Yes ☐
	☐☐	Yes ☐
	☐☐	Yes ☐
	☐☐	Yes ☐

8 Extra contacts

Did you did meet any more people?

Please write how many more contacts you made during this day?

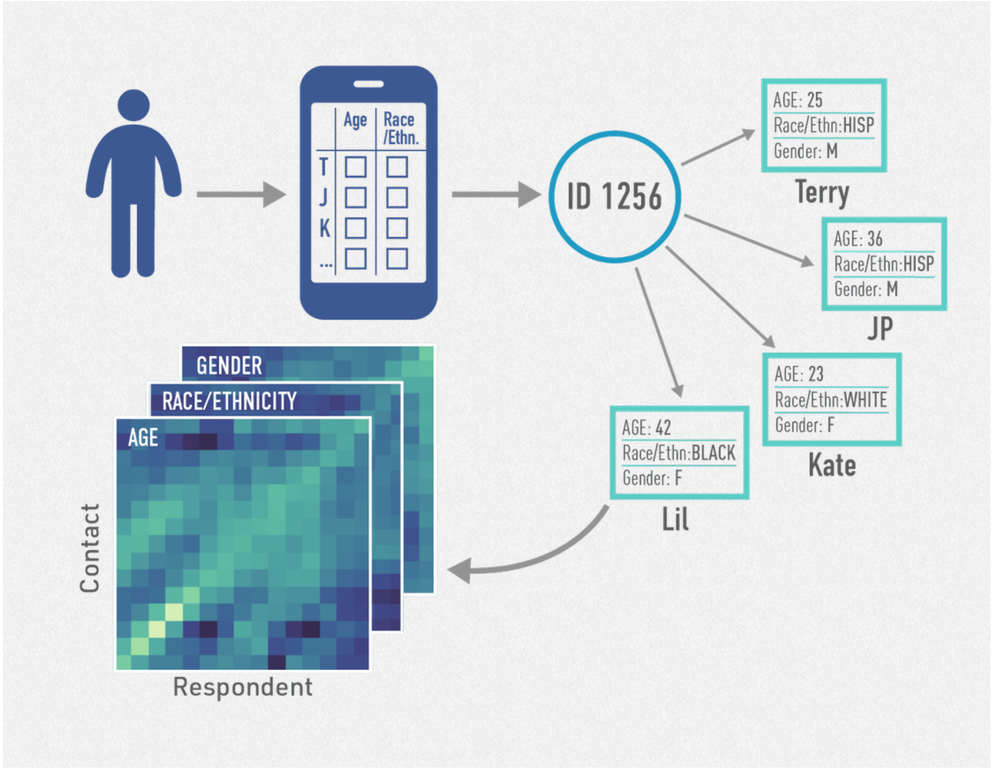
☐☐

Observatory of US Contact Patterns

What's new

Approach: Conducting multiple waves of a survey that provides regular snapshots of contact patterns in the US population. The survey is conducted on an established panel of individuals that is nationally representative by age, gender, and race/ethnicity. For each participant, the survey collects detailed information about the number and characteristics of their contacts.

Advancements: Quantifying US contact patterns in different socio-demographic and economic contexts for all ages and racial/ethnic groups.



Key takeaways

- Participants reported an average of 6.2 contacts per day, lower than pre-COVID estimates for other high-income countries. This reflects social changes brought by the COVID-19 pandemic s.
- The number of contacts varied with age, with school age individuals reporting the largest number of contacts and older adults reporting the lowest.
- No statistically significant difference in the number of contacts by gender was found.
- The number of contacts was different by race/ethnicity with NH Black reporting the lowest average number of contacts and Hispanic reporting the highest.

Epistorm-Mix

Mapping Social Contact Patterns in the Post-Pandemic United States

Epistorm-Mix provides privacy-preserving contact data and contact patterns characterization relevant for the spread of respiratory infectious diseases within the US population.

Built on a probability-based sample, Epistorm-Mix delivers population-level contact matrices and mixing indicators that reflect real-world heterogeneity across age, sex, race/ethnicity, income, and settings. These data enable epidemic models to better capture differential infection risks and to test targeted interventions in silico—offering actionable insight for seasonal preparedness and pandemic response.

The Repository

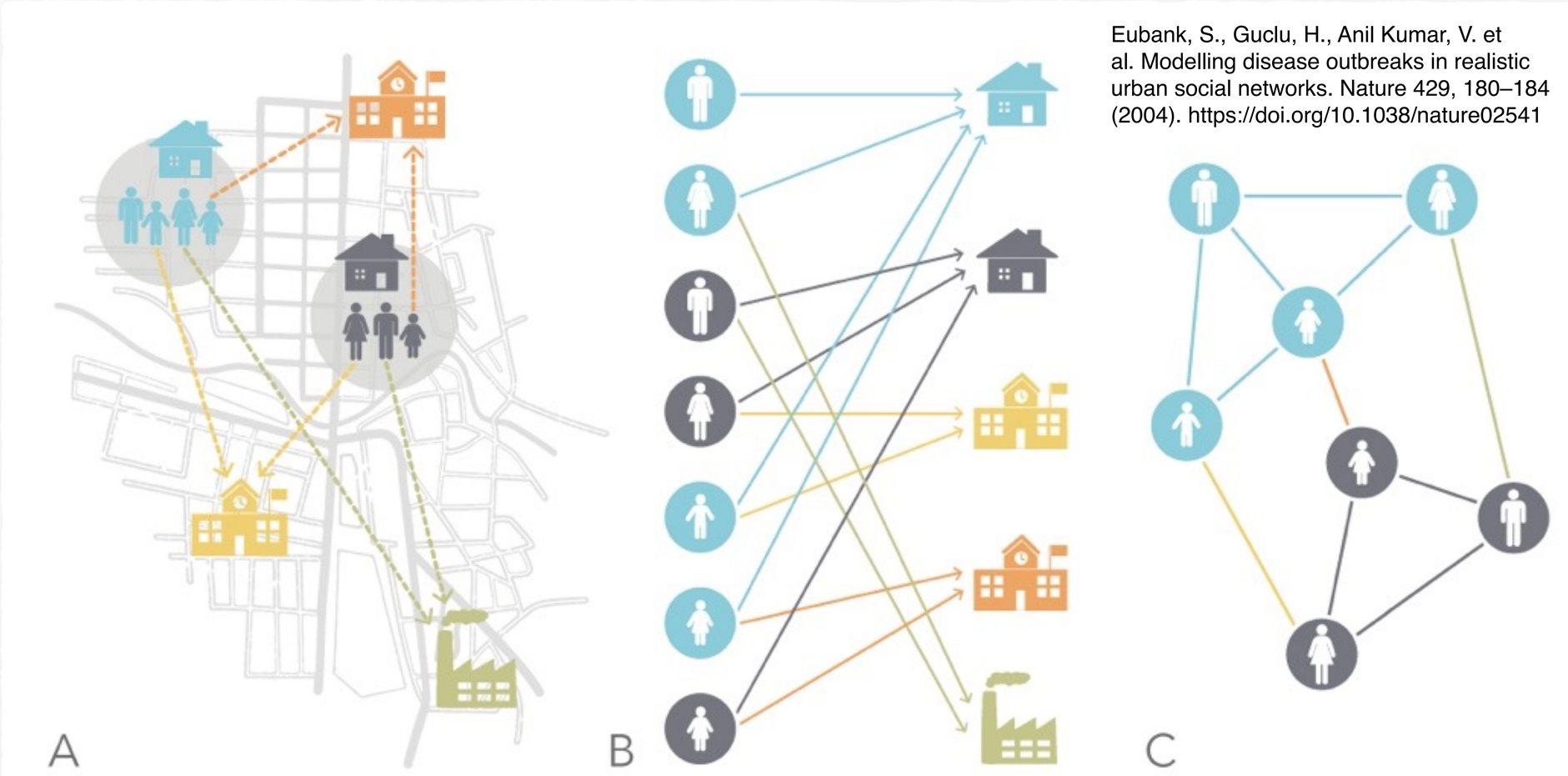
Use Epistorm-Mix to integrate realistic contact patterns into respiratory-disease models.

Repository with code, data, and tutorials:

Epistorm-Mix

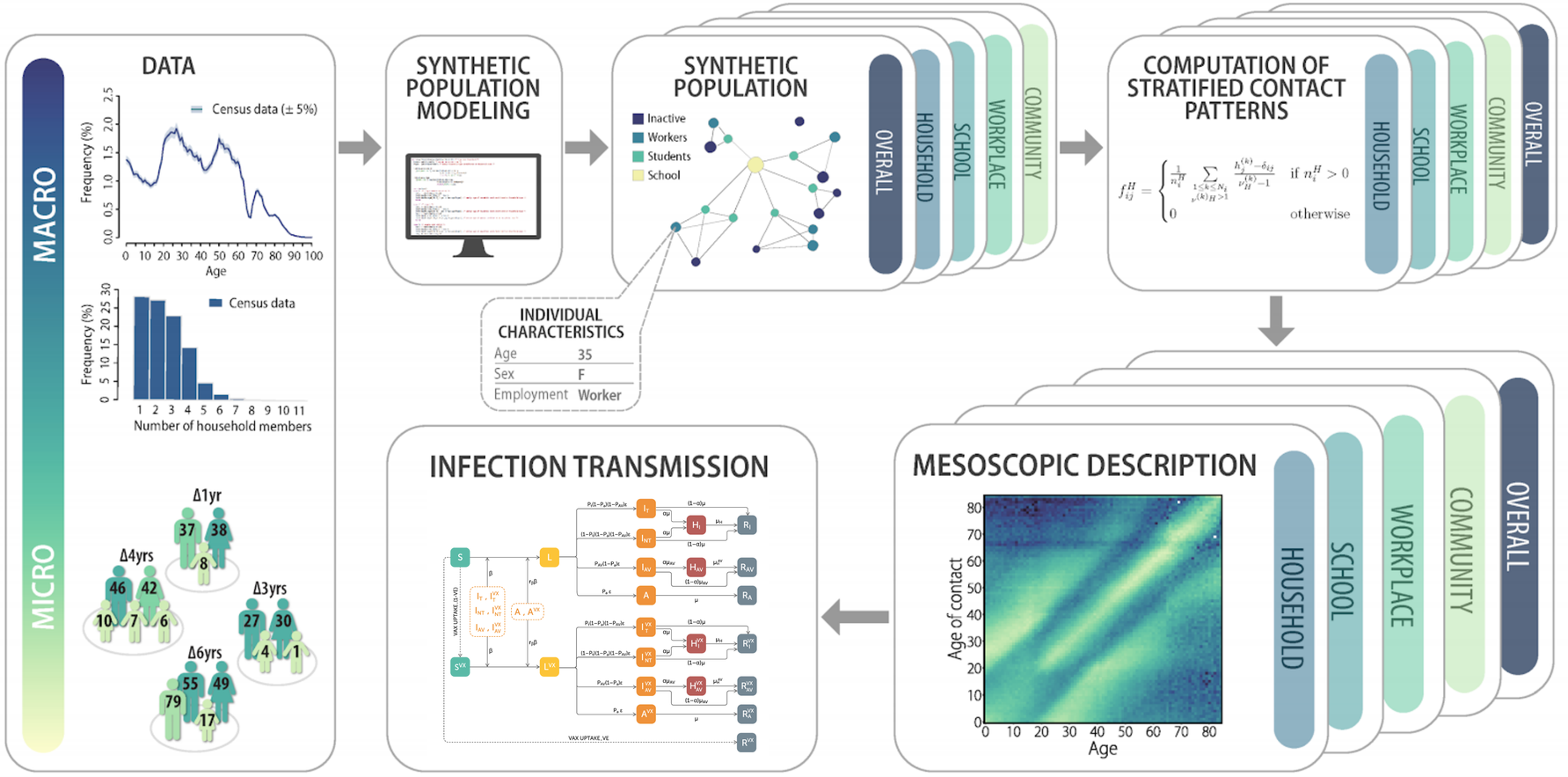
https://www.epistorm.org/data/epistorm-mix

EPISTORM



Eubank, S., Guclu, H., Anil Kumar, V. et al. Modelling disease outbreaks in realistic urban social networks. Nature 429, 180–184 (2004). https://doi.org/10.1038/nature02541

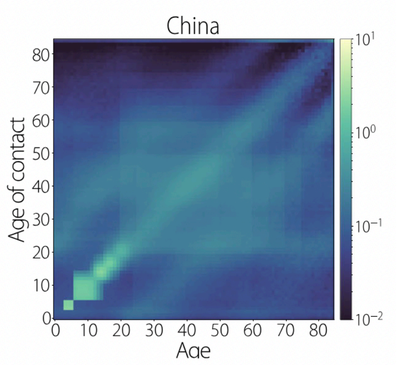
- Fumanelli, Laura, et al. "Inferring the structure of social contacts from demographic data in the analysis of infectious diseases spread." (2012): e1002673.
- Mistry, Dina, et al. "Inferring high-resolution human mixing patterns for disease modeling." Nature communications 12.1 (2021): 1-12.

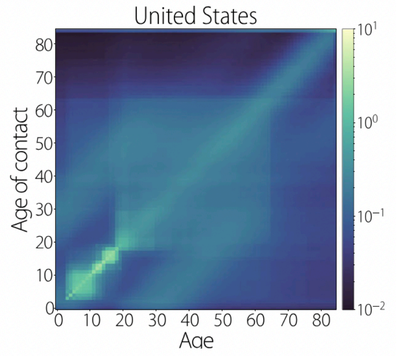


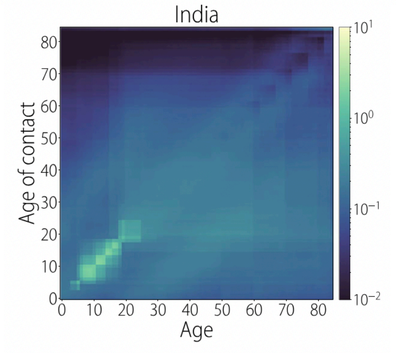
- Fumanelli, Laura, et al. "Inferring the structure of social contacts from demographic data in the analysis of infectious diseases spread." (2012): e1002673.
- Mistry, Dina, et al. "Inferring high-resolution human mixing patterns for disease modeling." Nature communications 12.1 (2021): 1-12.

$$M_{ij} = \sum_k \omega_k F_{ij}^k$$

Contact matrices can be stratified by setting (e.g., home, school, workplace, community), allowing the modulations of the number of contacts per settings.

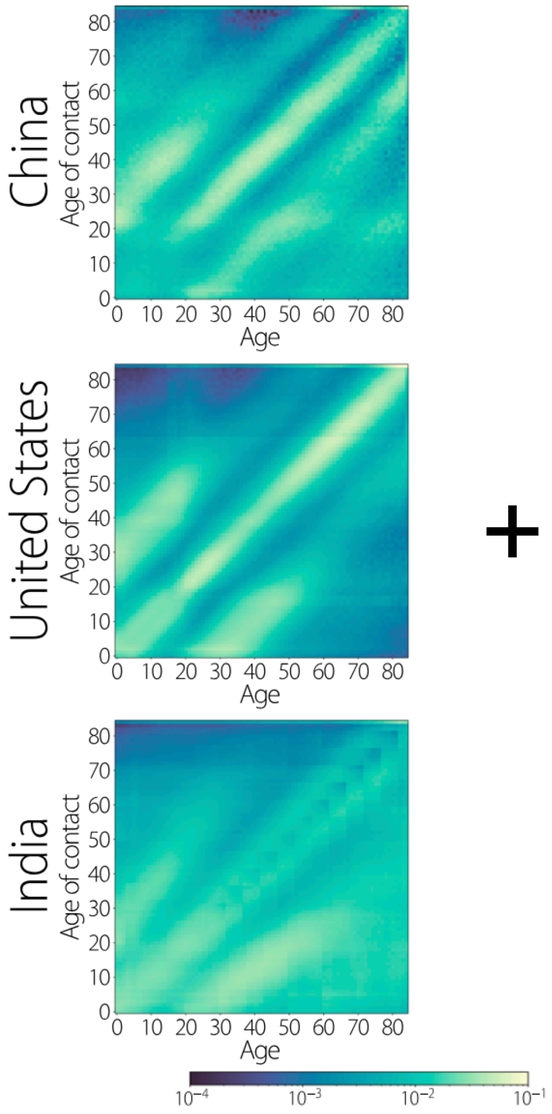




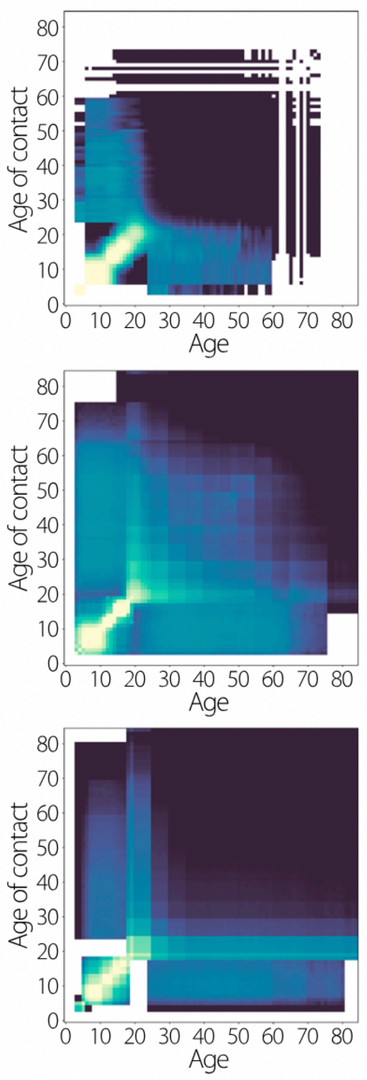


=

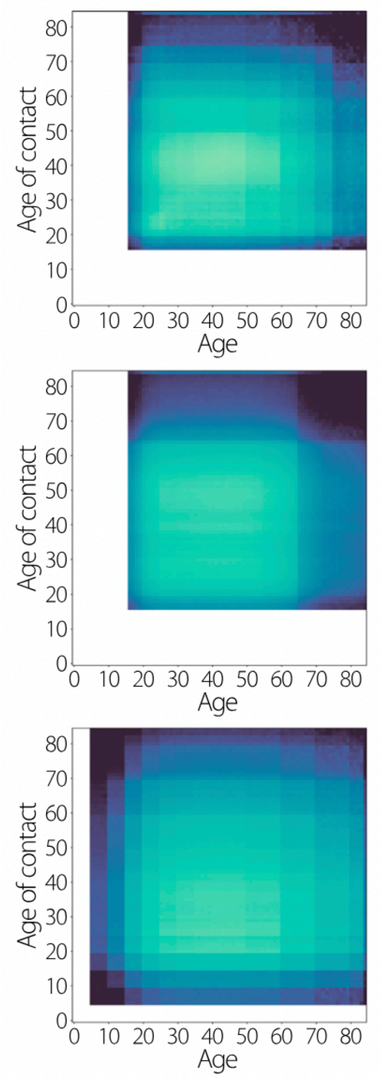
ω_1



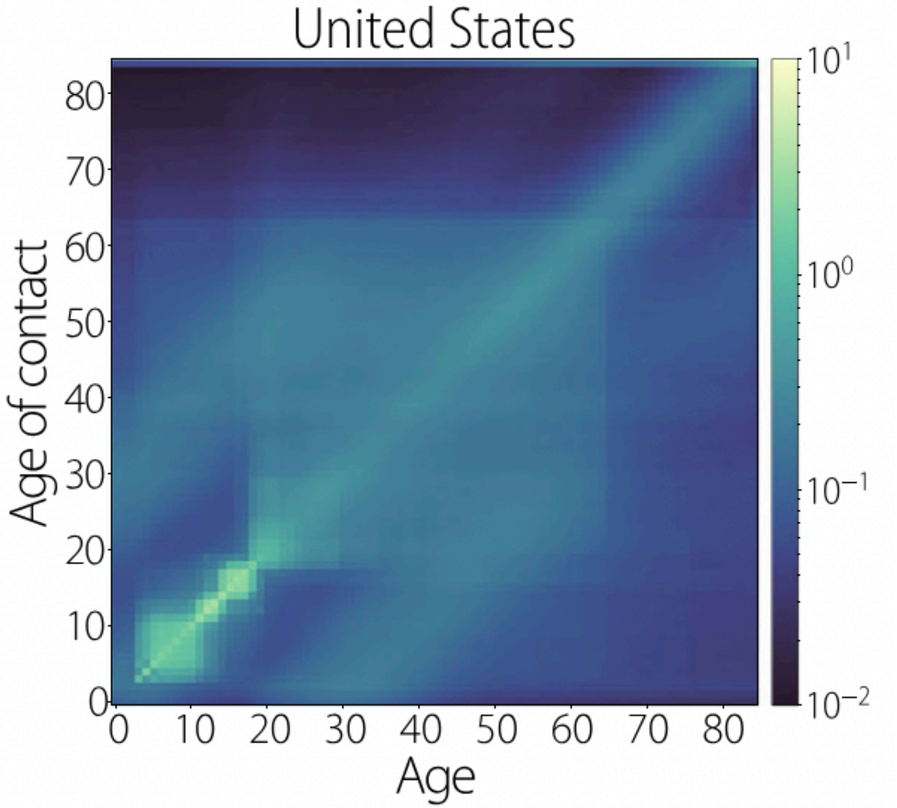
$+ \omega_2$



$+ \omega_3$



Core input to compartmental and agent-based models to simulate realistic epidemic dynamics



Chain Binomial processes

$$\lambda_t = \beta \sum_j M_{ij} \frac{I_j}{N_j}$$

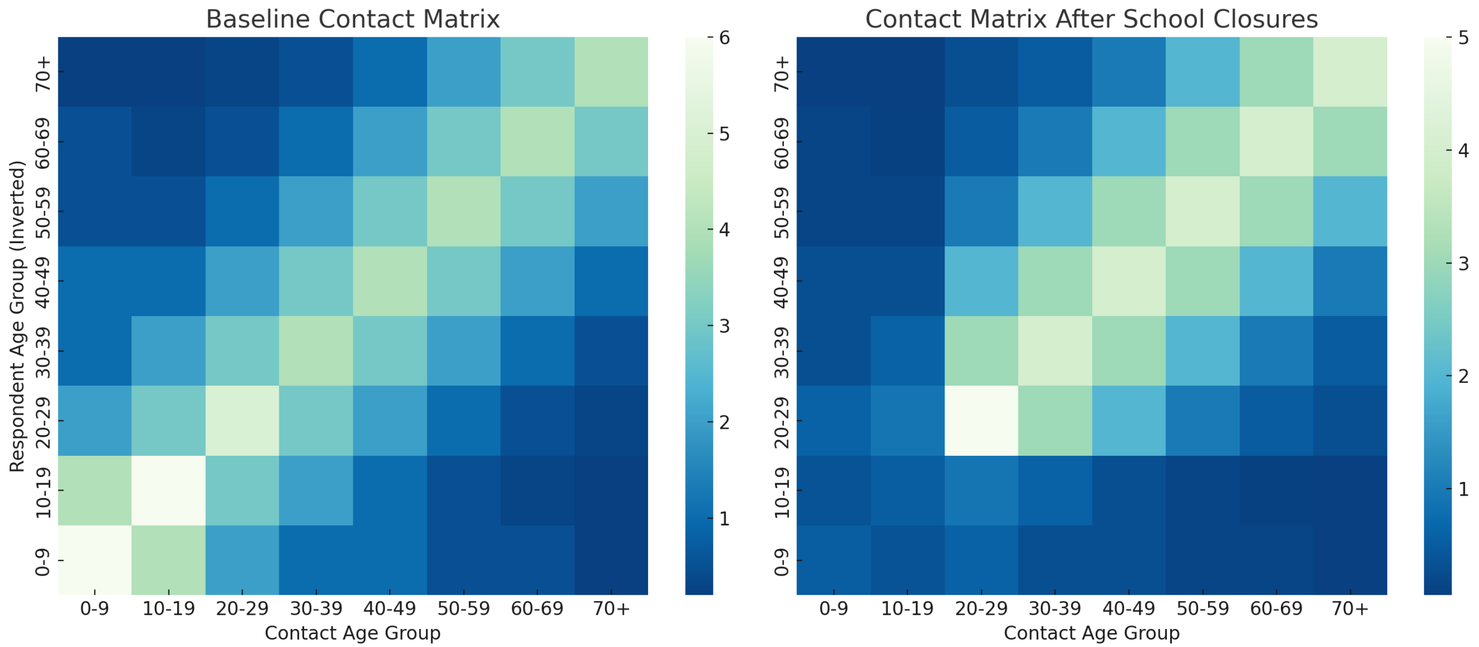
Age stratified force of infection

$$I_{i,t+1} = I_{i,t} - \mu I_{i,t} + Bin(S_{i,t}, \lambda_t)$$

- In transmission models (e.g., age-structured SIR/SEIR models), the force of infection is modulated by accounting for heterogeneous mixing across population groups.
- Enable calculation of next-generation matrices and the basic reproduction number (R_0), reflecting how contacts between different groups drive epidemic potential and spread.

MOBS LAB

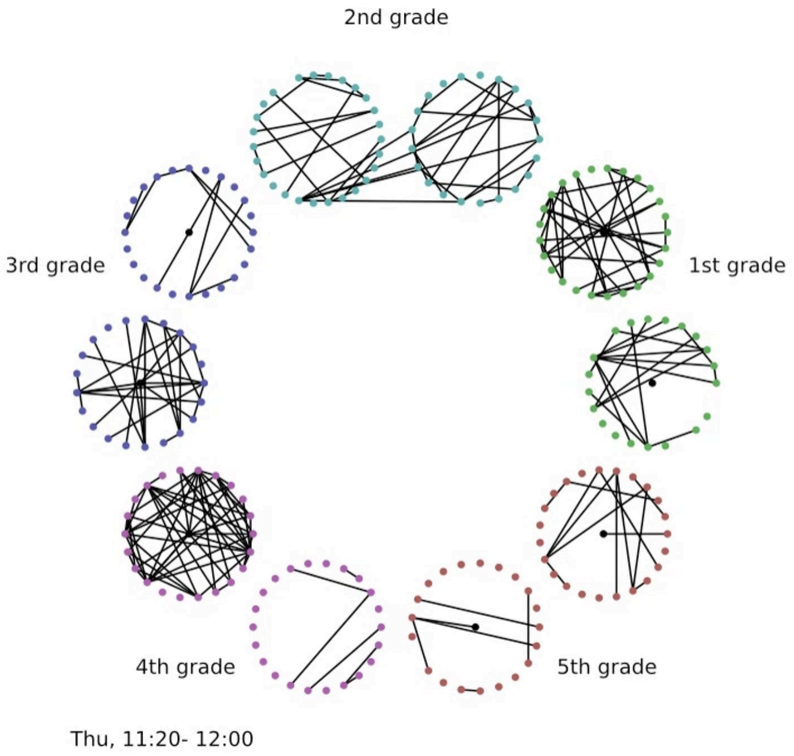
Scenario analysis and policy evaluation



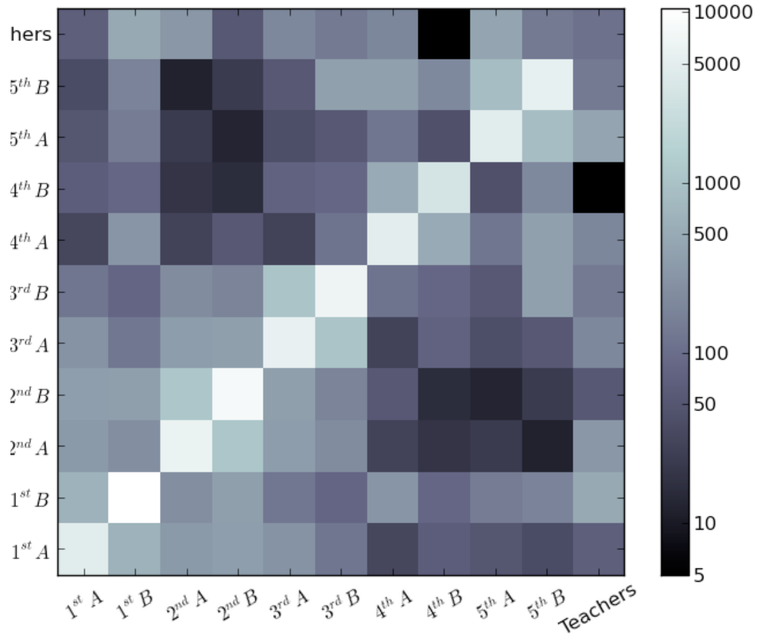
- **Contact patterns change under interventions**
- **Left:** The baseline contact matrix with high within-group and intergenerational contacts
- **Right:** The modified matrix simulates school closures, where interactions involving the 0–19 age groups are significantly reduced.

MOBS LAB

Flexible Stratification of Contact Matrices



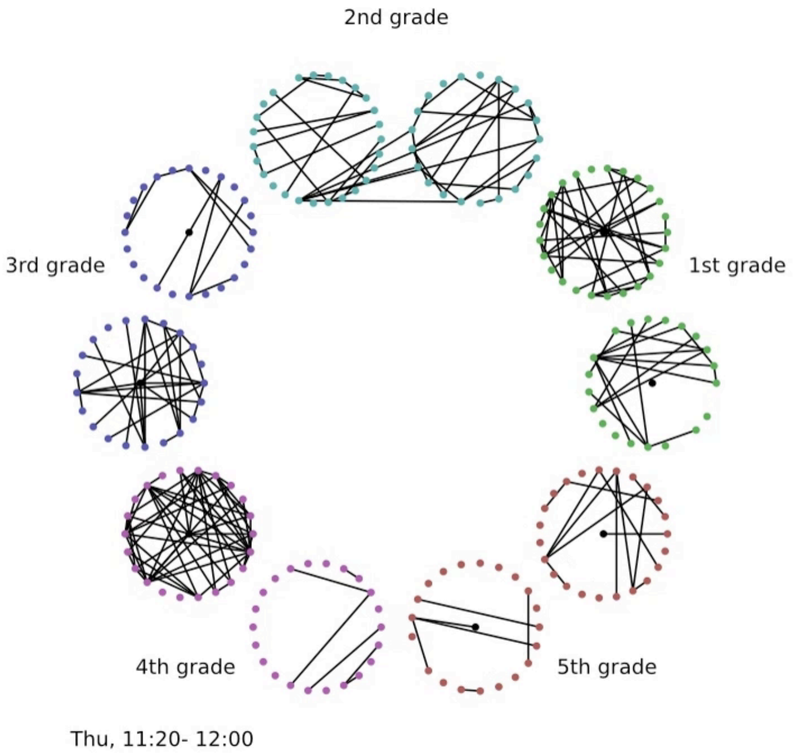
high-resolution contact network



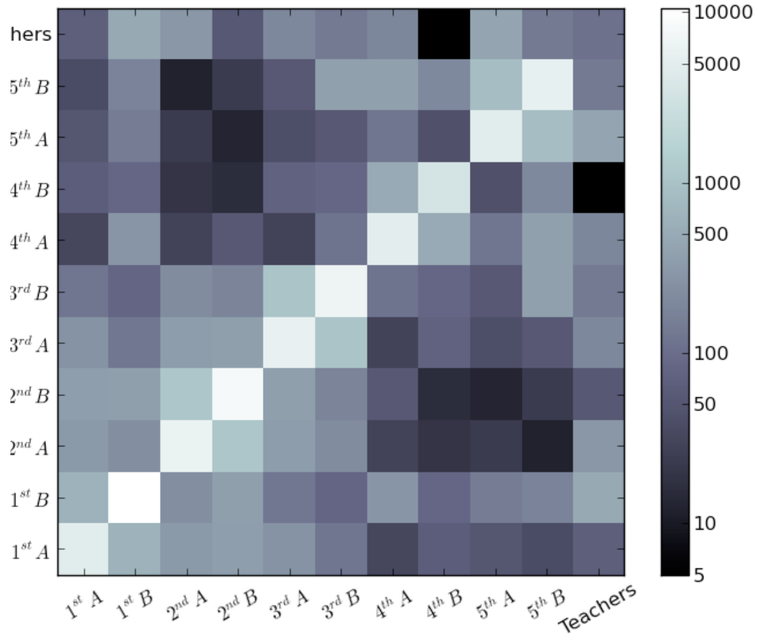
contact matrix

- Contact matrices can be defined across multiple dimensions, not just age — including geography, occupation, socioeconomic status, or risk groups (e.g., immunocompromised).
- Setting-specific matrices (home, school, workplace, community) can be combined with demographic or behavioral stratifications.
- Time-varying stratifications allow models to capture dynamic changes (e.g., during lockdowns, holidays, or vaccination rollouts).

Flexible Stratification of Contact Matrices



high-resolution contact network

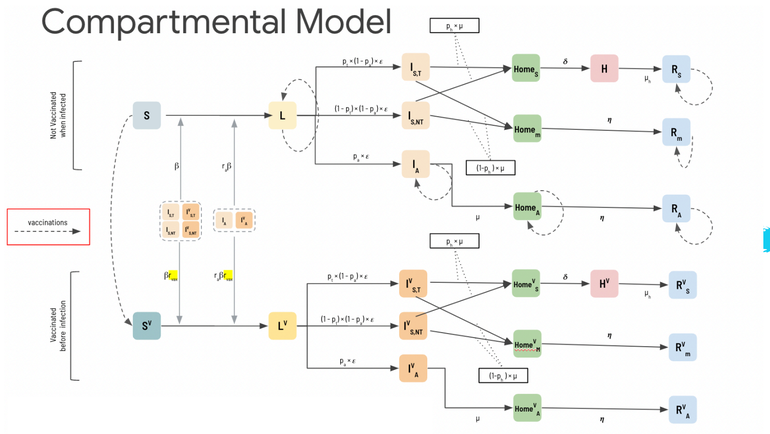


contact matrix

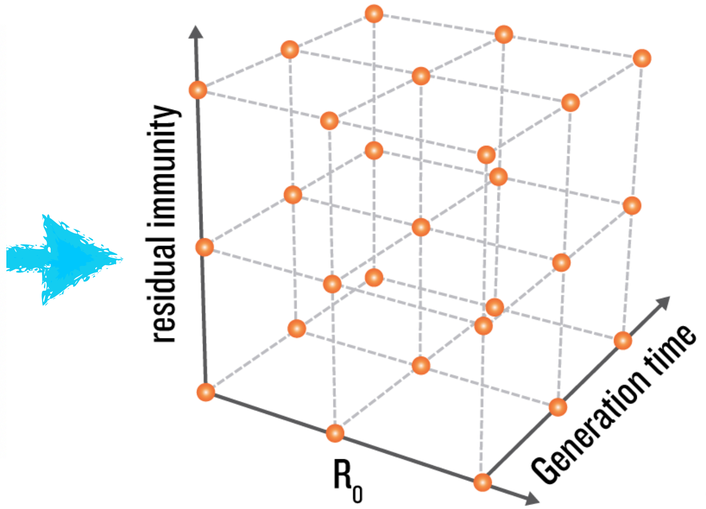
- Contact matrices can be defined across multiple dimensions, not just age — including geography, occupation, socioeconomic status, or risk groups (e.g., immunocompromised).
- Setting-specific matrices (home, school, workplace, community) can be combined with demographic or behavioral stratifications.
- Time-varying stratifications allow models to capture dynamic changes (e.g., during lockdowns, holidays, or vaccination rollouts).

MODELING/CALIBRATION

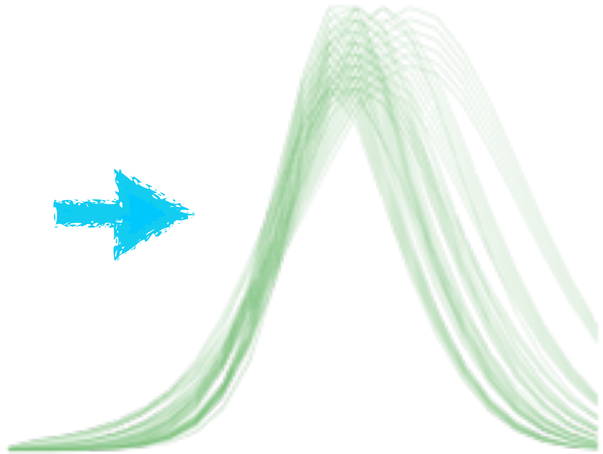
Disease Model



Phase Space



Epidemic curve

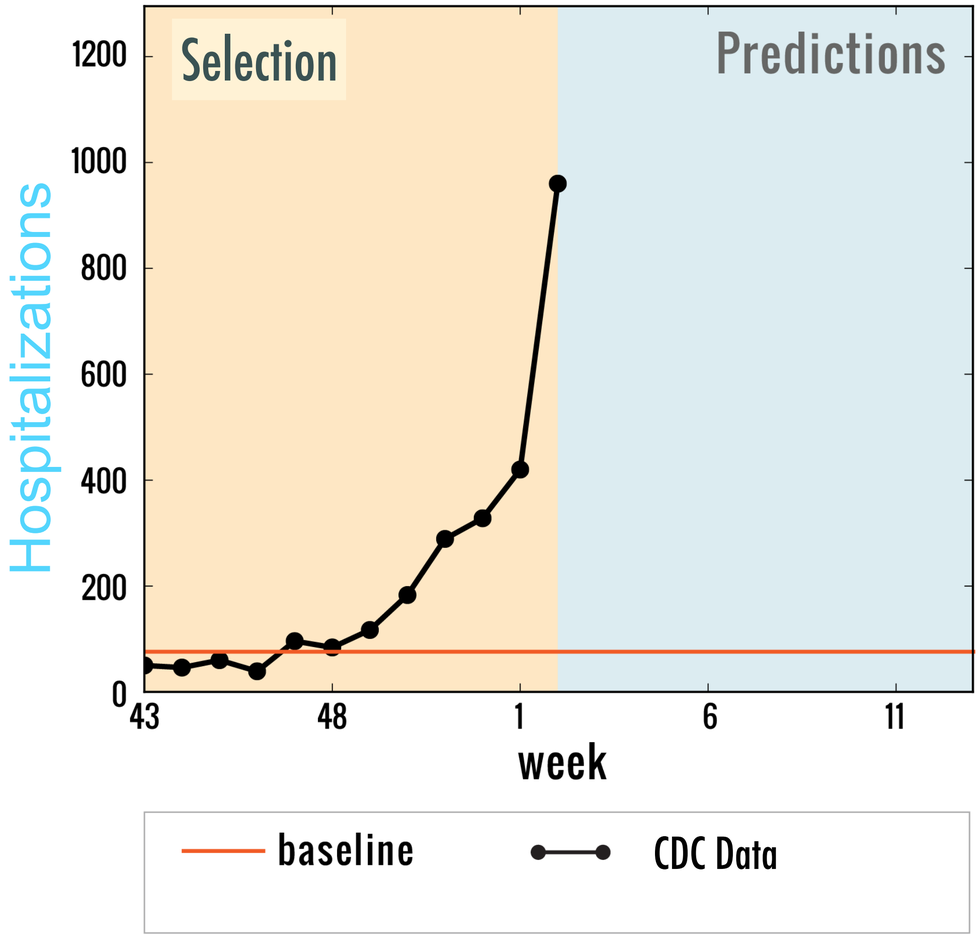


- parametrization

- Initial conditions
- Latin square sampling exploration

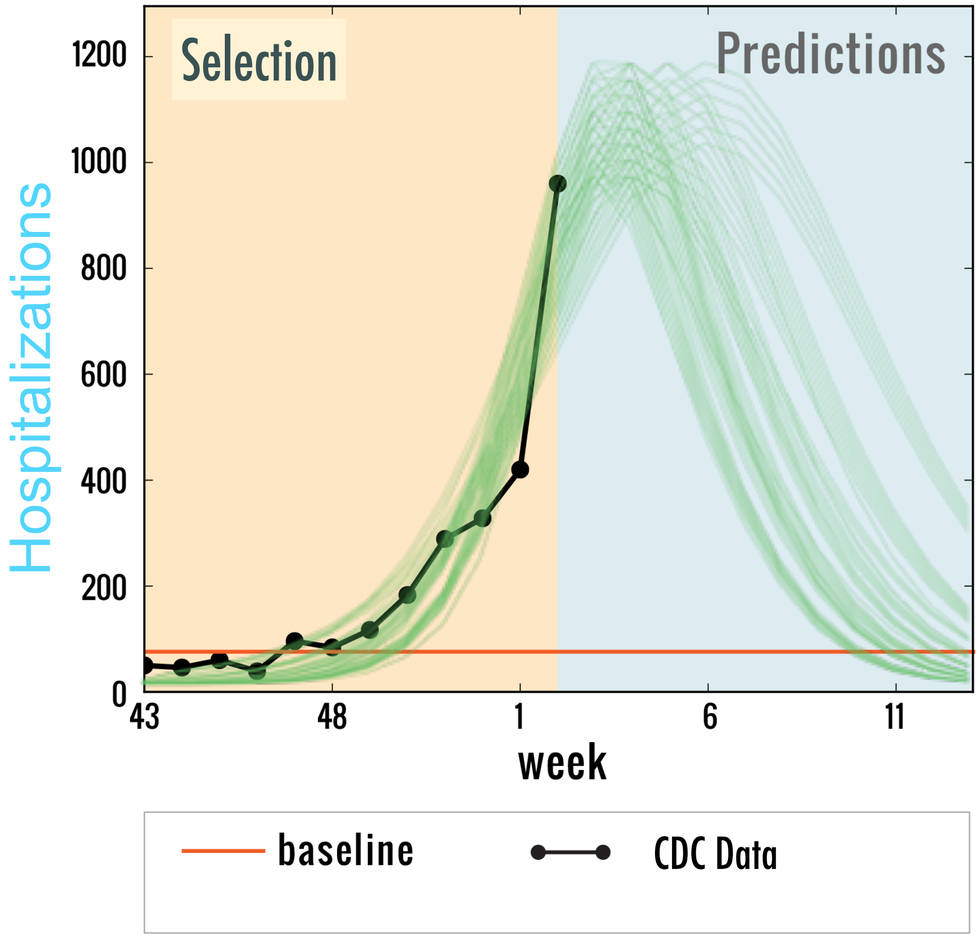
- Stochastic model generation

Model selection



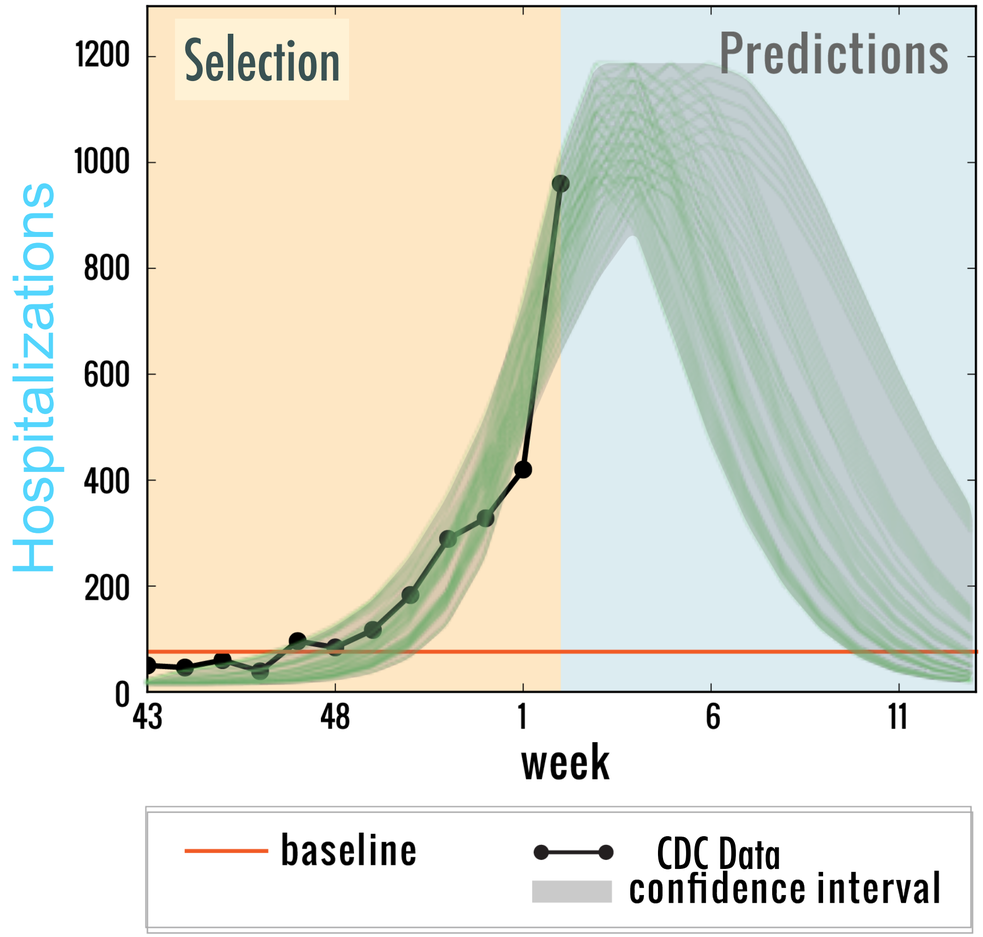
MOBS LAB

Model selection



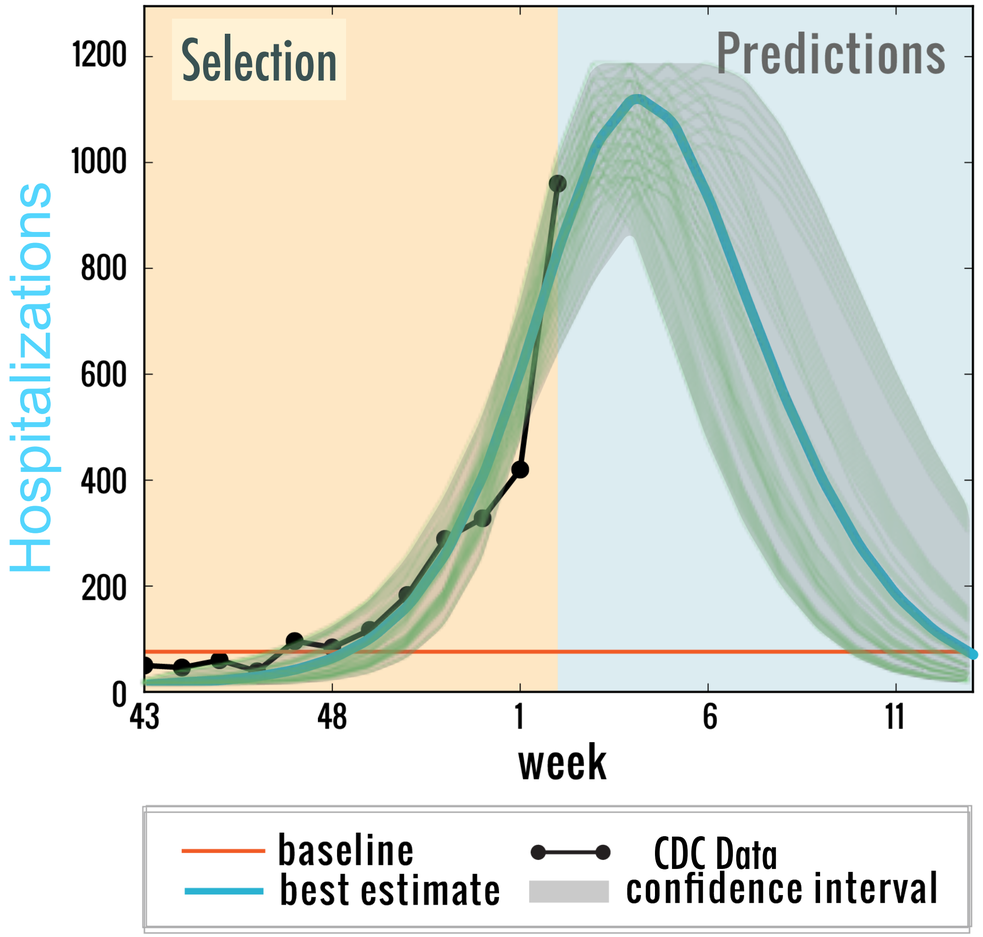
MOBS LAB

Model selection



MOBS LAB

Model selection



MOBS LAB

ABC approach

$D = observed\ data$

$\theta = parameter\ value$

$$p(\theta|D) = \frac{p(D|\theta)p(\theta)}{p(D)}$$

Rejection criterion

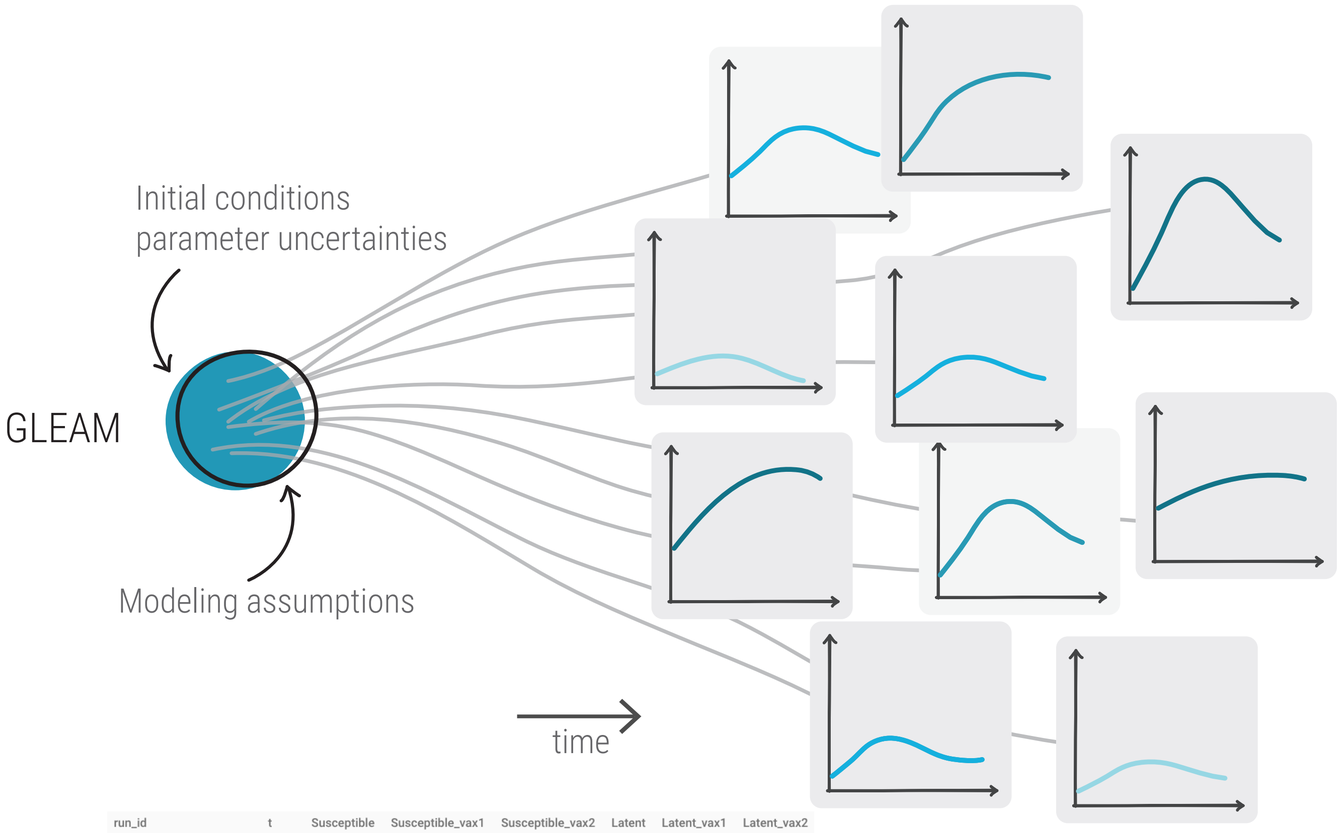
$$\rho(D, D^*) \leq \epsilon$$

1. Sample θ^* from the prior distribution $P(\theta)$.
2. Simulate data set from the model, using parameters θ^*, to get D^*.
3. If $\rho(D, D^*) \leq \epsilon$ accept θ^* otherwise reject.
4. Repeat until N particles (the parameter values or parameter sets) $\boldsymbol{\theta}^* = \{\theta_i^*;\ i = 1, ..., n\}$ are accepted.

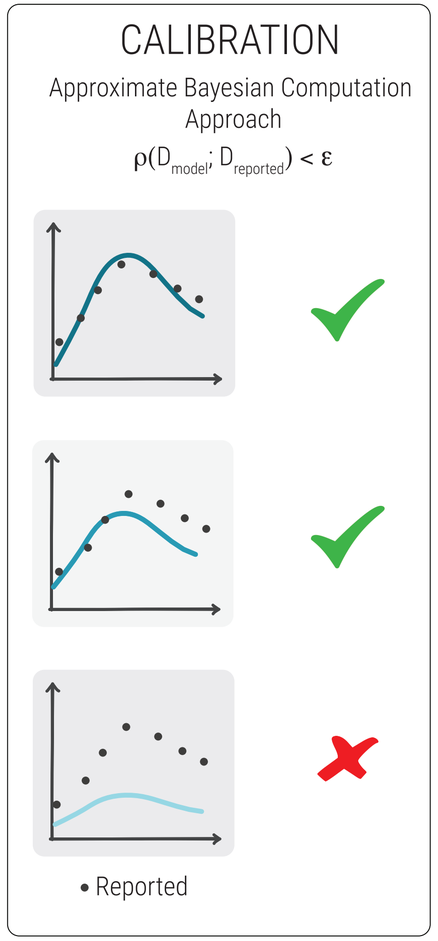
Sunnåker M, Busetto AG, Numminen E, Corander J, Foll M, Dessimoz C (2013) Approximate Bayesian Computation. PLoS Comput Biol 9(1): e1002803.
Minter &Rethkute, Approximate Bayesian Computation for infectious disease modelling, Epidemics, 29, December 2019, 100368

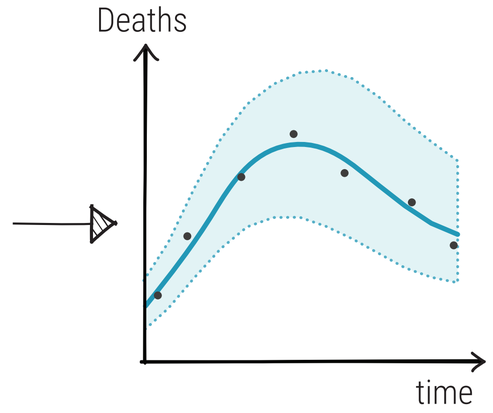
MOBS LAB

Model Calibration & analysis



run_id	t	Susceptible	Susceptible_vax1	Susceptible_vax2	Latent	Latent_vax1	Latent_vax2
70001000000000411	253	0	0	0	10	0	0
70001000000000910	253	0	0	0	4	0	0
70001000000000034	253	0	0	0	7	0	0
70001000000000297	253	0	0	0	14	0	0
70001000000000273	253	0	0	0	3	0	0
70001000000000175	253	0	0	0	10	0	0
70001000000000365	253	0	0	0	19	0	0
70001000000000695	253	0	0	0	10	0	0





MOBS LAB

epydemix is an open-source Python package designed for flexible, modular, and data-driven epidemic modeling.

It supports the full modeling pipeline—from constructing stochastic compartmental models to running simulations, integrating real-world data, and calibrating parameters. Users can incorporate age-structured contact patterns, dynamic interventions, and population demographics with ease. Built-in Approximate Bayesian Computation (ABC) methods enable robust parameter estimation and model fitting, supporting forecasting, scenario exploration, and policy-relevant analyses. Epydemix bridges the gap between theoretical modeling and practical application, helping researchers and public health professionals translate models into actionable insights.

GitHub Repository >

pip install epydemix

MOBS LAB