Understanding COVID-19

with non-markovian & agent-based models

George Wong

in collaboration with

Ahmed Elbanna, Nigel Goldenfeld, Sergei Maslov, Alexei Tkachenko,
Tong Wang, Zach Weiner, and Hantao Zhang

The problem

An unknown disease, SARS-COV-2, rapidly spreads across the planet. Its symptoms are unknown; its incubation and infectiousness periods are unknown; its severity is unknown.

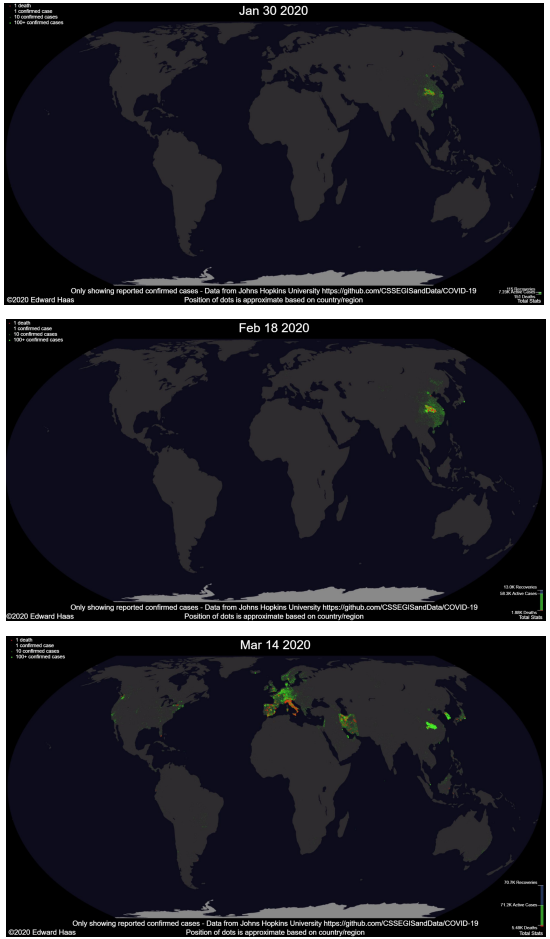
How do we determine when to close borders?

How do we determine whether to build new hospitals?

How do we predict different mitigation strategies' effectiveness?

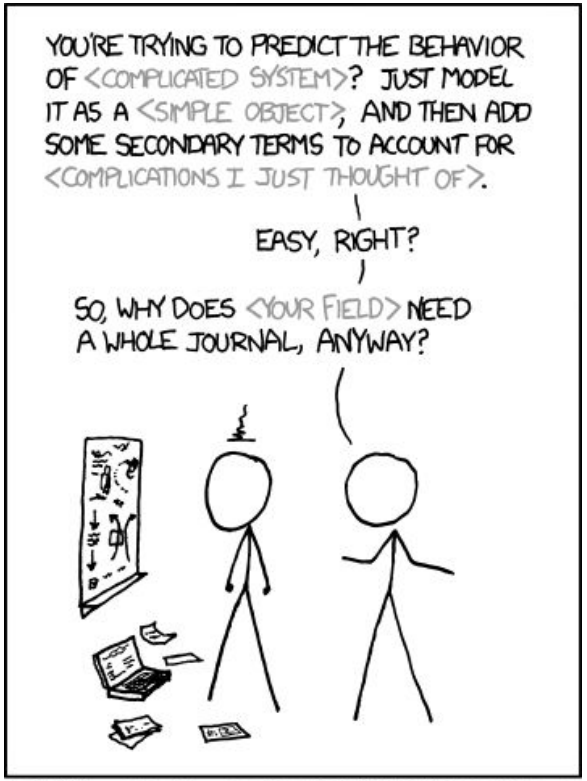
What is the ideal partition of a population to limit/quench spread?

...



Outline

- Subgrid picture of disease spread
- “Standard” compartmental models
- Extensions to “standard” models (c.f. 1927)
- Parameter inference
- Model shortfalls (percolation regime, heterogeneity)
- Extension to a campus

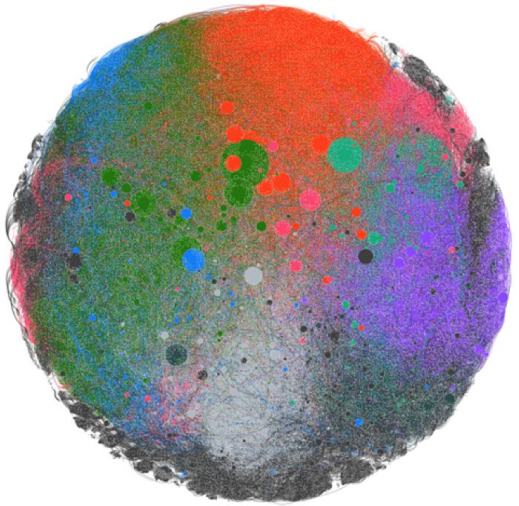


“Microscopic” system description

>**10,000,000 individuals** (Illinois) with different ages and pre-existing conditions

Individuals interact with each other according to **time-dependent social network**

Each node is a person, each edge is an interaction



“Microscopic” system description

>**10,000,000 individuals** (Illinois) with different ages and pre-existing conditions

Individuals interact with each other according to **time-dependent social network**

Infectious individuals **emit viral quanta** according to activity state

Table 3 – Quanta emission rates (ER_q) for a SARS-CoV-2 infected asymptomatic subject ($c_v=10^8$ copies mL^{-1}) as a function of the activity level and respiratory activity.

Activity level	Respiratory activity				
	Voiced counting	Whispered counting	Speaking	Breathing	Avg
Resting	49.9	12.1	320	10.5	98.1
Standing	74.8	18.1	480	15.7	147
Light exercise	161	39.1	1.03×10^3	33.9	317

Buonanno+ 2020

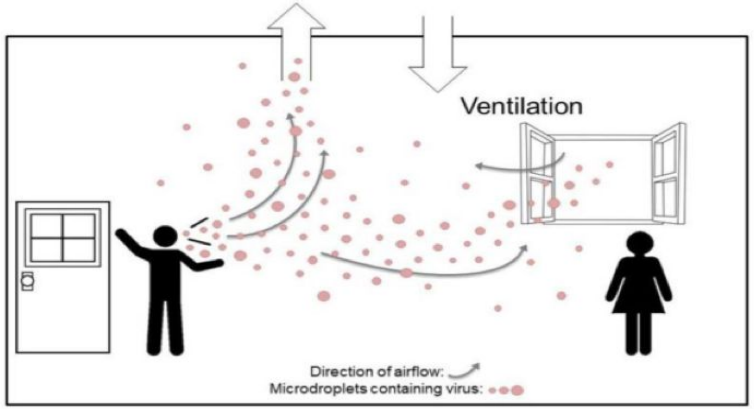
“Microscopic” system description

>**10,000,000 individuals** (Illinois) with different ages and pre-existing conditions

Individuals interact with each other according to **time-dependent social network**

Infectious individuals **emit viral quanta** according to activity state

Quanta spread according to **air flow patterns**



Morawska+ 2020

“Microscopic” system description

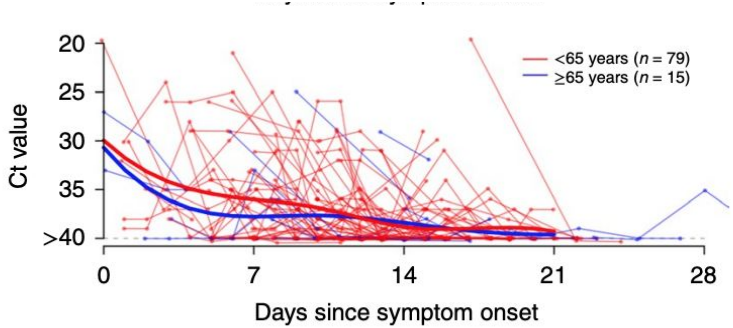
>**10,000,000 individuals** (Illinois) with different ages and pre-existing conditions

Individuals interact with each other according to **time-dependent social network**

Infectious individuals **emit viral quanta** according to activity state

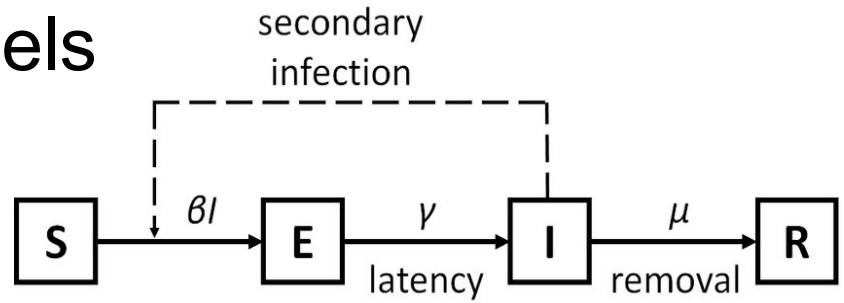
Quanta spread according to **air flow patterns**

Different disease progression per individual



He+ 2020

Standard compartmental models



Matthew Patrick+ 2016

In an **SEIR** model, every member of the population is assigned to a population subgroup:

Susceptible, **E**xposed, **I**nfectious, **R**emoved

Individuals transition through the network stages according to **reactions**:

S + I → E a susceptible person is infected

E → I an exposed person becomes infectious

I → R an infectious person recovers

$$\frac{dS}{dt} = \Lambda N - \mu S - \frac{\beta IS}{N}$$

$$\frac{dE}{dt} = \frac{\beta IS}{N} - (\mu + a)E$$

$$\frac{dI}{dt} = aE - (\gamma + \mu)I$$

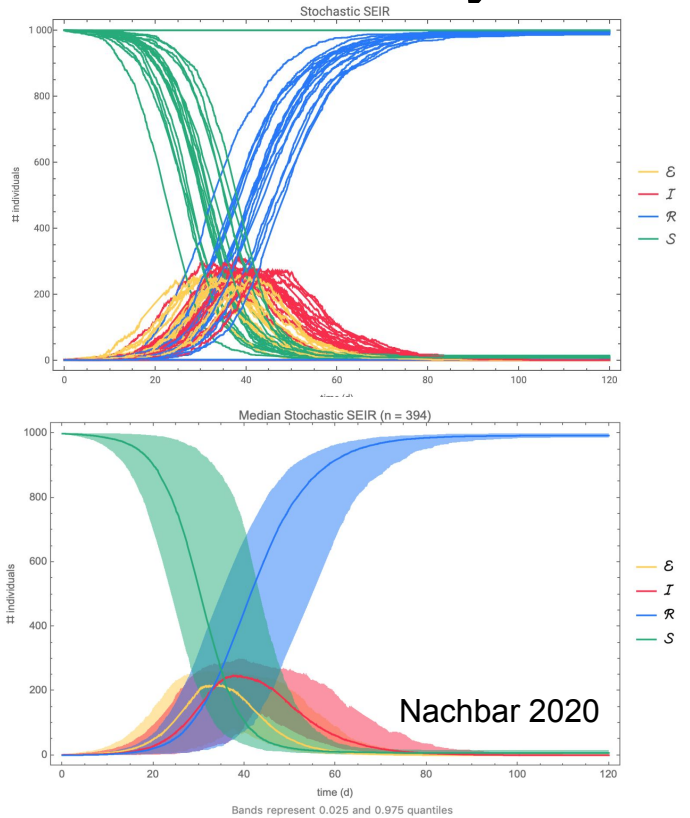
$$\frac{dR}{dt} = \gamma I - \mu R.$$

Standard compartmental models: +stochasticity

Real-world outbreaks are **not smooth**. Random noise is involved. Recast dynamical equations as **stochastic differential equations**, and use the **Gillespie algorithm** to produce trajectory.

1. Write reaction as rate = 1/time → timestep
2. Set dt = -log(1-X)/rate, X a R.U.V. in (0,1)

** extra details for systems with multiple reactions



Standard compartmental models: the problem

Question: Since the differential equations do not transition individuals from left state to the right state,
What is the distribution of "time spent" in a state?

Answer: exponential distribution!

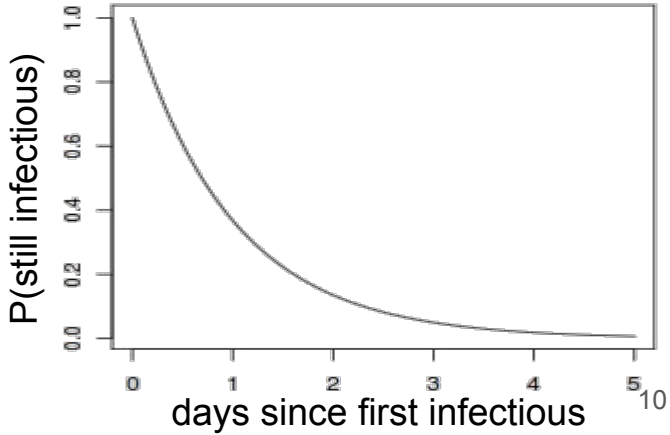
$$\frac{dS}{dt} = -\beta I S$$

$$\frac{dE}{dt} = +\beta I S - a E$$

$$\frac{dI}{dt} = +a E - \gamma I$$

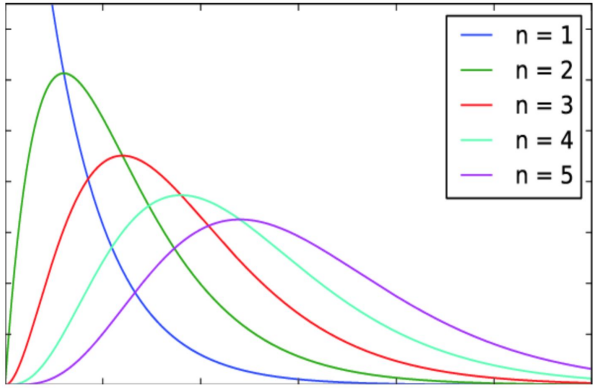
$$\frac{dR}{dt} = +\gamma I$$

Does not reflect the real world, which has reported latent/infectiousness profiles ~gamma distributions (e.g., Linton+ 2020)



Standard compartmental models: a solution

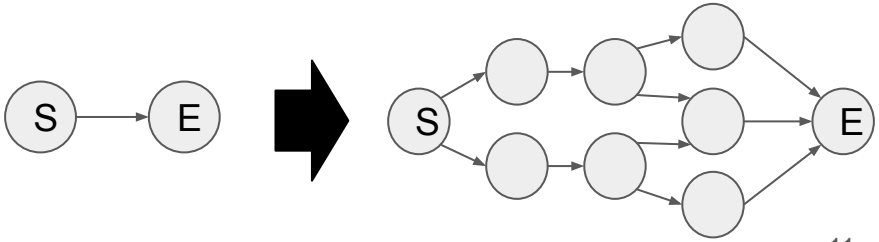
Just add compartments to the model! Rates between compartments will be exponential, but the **convolution of exponentials** will be an **Erlang distribution**.



Skottfelt+ 2014

Internal/parallel nodes can effectively produce **any distribution** you want (Hurtado+ 2019).

** related to the “exposed” compartment.



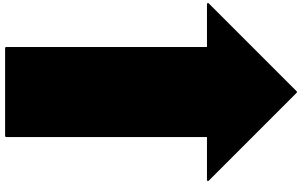
Standard compartmental models: a solution?

$$\frac{dS}{dt} = \Lambda N - \mu S - \frac{\beta I S}{N}$$

$$\frac{dE}{dt} = \frac{\beta I S}{N} - (\mu + a)E$$

$$\frac{dI}{dt} = aE - (\gamma + \mu)I$$

$$\frac{dR}{dt} = \gamma I - \mu R.$$



$$\frac{dS_i}{dt} = -\lambda_i S_i,$$
$$\frac{dE_{1i}}{dt} = \phi_i + \lambda_i S_i - \sigma E_{1i},$$
$$\frac{dE_{ki}}{dt} = \phi_i + \sigma E_{(k-1)i} - \sigma E_{ki} \qquad (k = 2, ..., m),$$
$$\frac{dP_{1i}}{dt} = \phi_i + (1 - \rho_i)\sigma E_{mi} - \zeta_{\text{symptomatic},i} P_{1i},$$
$$\frac{dP_{ki}}{dt} = \phi_i + \zeta_{\text{symptomatic},i} P(k-1)i - \zeta_{\text{symptomatic},i} P_{ki} \qquad (k = 2, ..., n),$$
$$\frac{dA_{1i}}{dt} = \phi_i + \rho_i \sigma E_{mi} - \eta_i A_{1i},$$
$$\frac{dA_{ki}}{dt} = \phi_i + \eta_i A_{(k-1)i} - \eta_i A_{ki} \qquad (k = 2, ..., p),$$
$$\frac{dI_{\text{M},1i}}{dt} = \phi_i + (1 - \kappa_i)\zeta_{\text{symptomatic},i} P_{ni} - \gamma_{\text{mild},i} I_{\text{M},1i},$$
$$\frac{dI_{\text{M},ki}}{dt} = \phi_i + \gamma_{\text{mild},i} I_{\text{M},(k-1)i} - \gamma_{\text{mild},i} I_{\text{M},ki} \qquad (k = 2, ..., q),$$
$$\frac{dI_{\text{S},1i}}{dt} = \phi_i + \kappa_i \zeta_{\text{symptomatic},i} P_{ni} - \zeta_{\text{hospitalized},i} I_{\text{S},1i},$$
$$\frac{dI_{\text{S},ki}}{dt} = \phi_i + \zeta_{\text{hospitalized},i} I_{\text{S},(k-1)i} - \zeta_{\text{hospitalized},i} I_{\text{S},ki} \qquad (k = 2, ..., r),$$
$$\frac{dI_{\text{H1},1i}}{dt} = \phi_i + \psi_1 \zeta_{\text{hospitalized},i} I_{\text{S},ri} - \gamma_{\text{hospitalized},i} I_{\text{H1},1i}$$
$$\frac{dI_{\text{H1},ki}}{dt} = \phi_i + \gamma_{\text{hospitalized},i} I_{\text{H1},(k-1)i} - \gamma_{\text{hospitalized},i} I_{\text{H1},ki} \qquad (k = 2, ..., r),$$

$$\frac{dI_{\text{H2},1i}}{dt} = \phi_i + \psi_2 \zeta_{\text{hospitalized},i} I_{\text{S},ri} - \zeta_{\text{critical},i} I_{\text{H2},1i}$$
$$\frac{dI_{\text{H2},ki}}{dt} = \phi_i + \zeta_{\text{critical},i} I_{\text{H2},(k-1)i} - \zeta_{\text{critical},i} I_{\text{H2},ki} \qquad (k = 2, ..., r),$$
$$\frac{dI_{\text{C2},1i}}{dt} = \phi_i + \zeta_{\text{critical},i} I_{\text{H2},ri} - \gamma_{\text{critical},i} I_{\text{C2},1i}$$
$$\frac{dI_{\text{C2},ki}}{dt} = \phi_i + \gamma_{\text{critical},i} I_{\text{C2},(k-1)i} - \gamma_{\text{critical},i} I_{\text{C2},ki} \qquad (k = 2, ..., r),$$
$$\frac{dI_{\text{H3},1i}}{dt} = \phi_i + (1 - \psi_1 - \psi_2)\zeta_{\text{hospitalized},i} I_{\text{S},ri} - \zeta_{\text{critical},i} I_{\text{H3},1i}$$
$$\frac{dI_{\text{H3},ki}}{dt} = \phi_i + \zeta_{\text{critical},i} I_{\text{H3},(k-1)i} - \zeta_{\text{critical},i} I_{\text{H3},ki} \qquad (k = 2, ..., r),$$
$$\frac{dI_{\text{C3},1i}}{dt} = \phi_i + \zeta_{\text{critical},i} I_{\text{H3},ri} - \mu_{\text{critical},i} I_{\text{C3},1i}$$
$$\frac{dI_{\text{C3},ki}}{dt} = \phi_i + \mu_{\text{critical},i} I_{\text{C3},(k-1)i} - \mu_{\text{critical},i} I_{\text{C3},ki} \qquad (k = 2, ..., r),$$
$$\frac{dR_i}{dt} = \gamma_{\text{mild},i} I_{\text{M},qi} + \gamma_{\text{hospitalized},i} I_{\text{H1},ri} + \gamma_{\text{critical},i} I_{\text{C2},ki} + \eta_i A_{pi},$$

Introducing Non-markovian models

The basic timing problem with compartment models comes from the fact that **individuals do not know how long they have been in a state**.

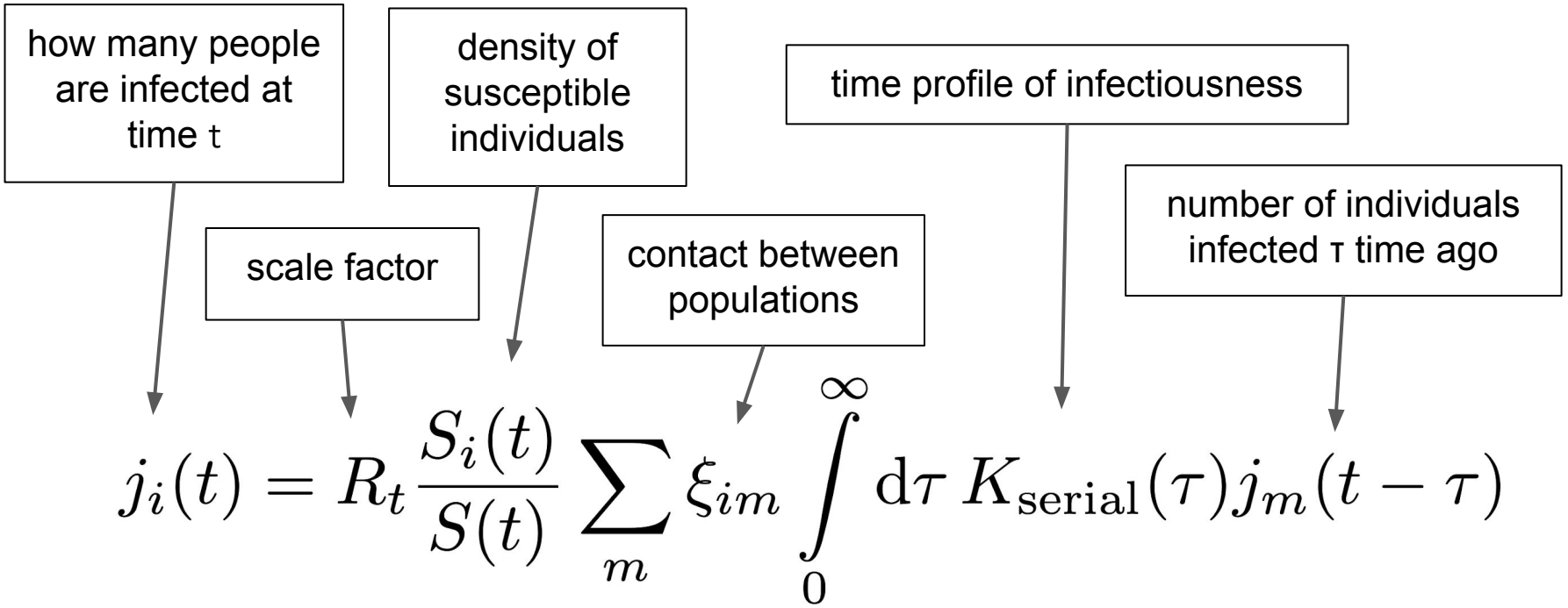
Fix: swap “single number” compartment populations for functions of time, i.e., swap differential equations for integro-differential equations.

** actually an integral equation shown here

$$j_i(t) = R_t \frac{S_i(t)}{S(t)} \sum_m \xi_{im} \int_0^\infty \mathrm{d}\tau \, K_{\text{serial}}(\tau) j_m(t-\tau)$$

Kermack–McKendrick theory (1927, 1932, 1933)

Introducing Non-markovian models

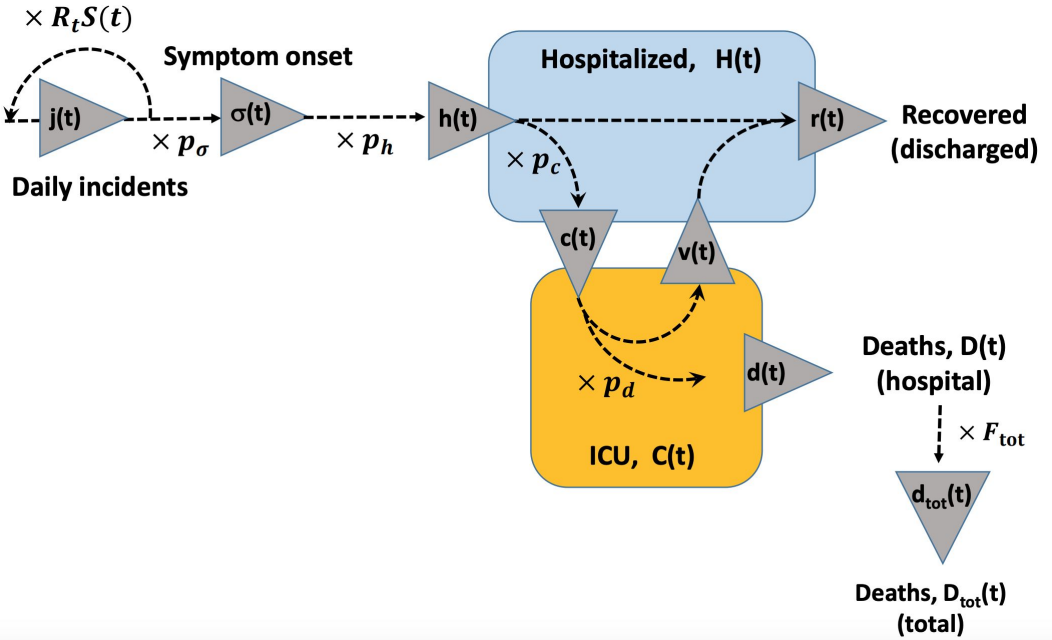


Calibrating the model to data

Data come primarily from the **healthcare system**, so must relate **infected** to **symptomatic**, to **hospitalized**, and so on

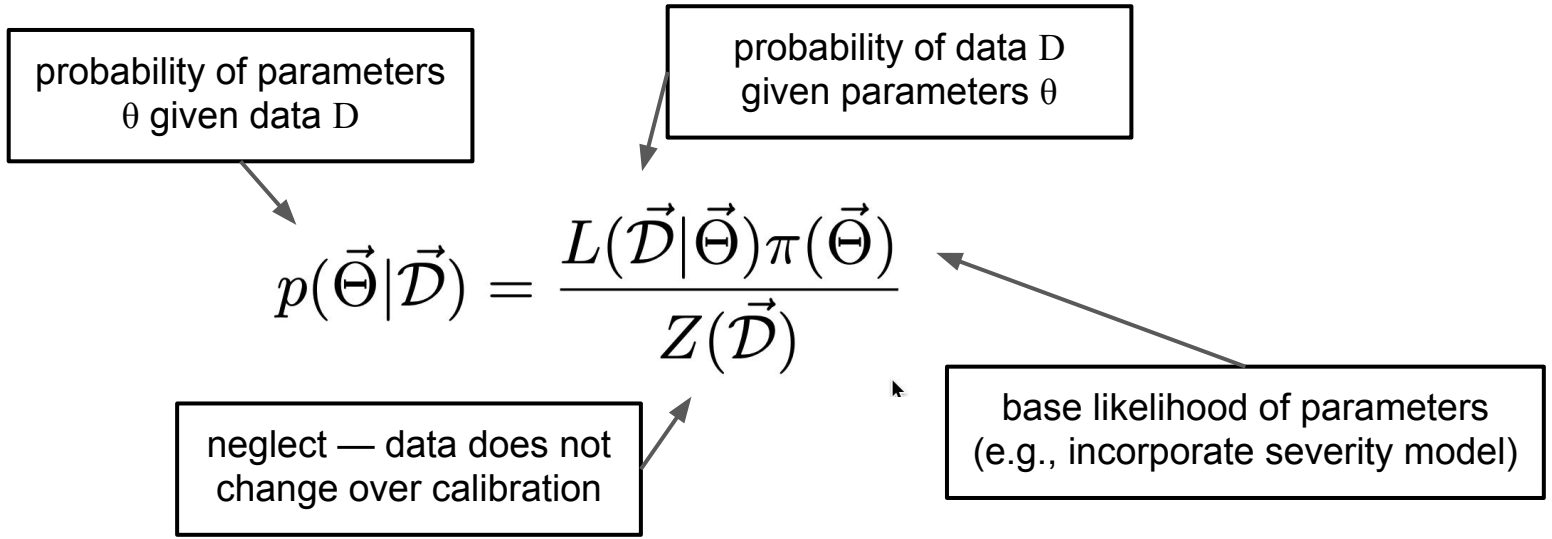
Model topology described by figure to the right

Dashed lines represent integral equations (as in previous slide)



Parameter inference

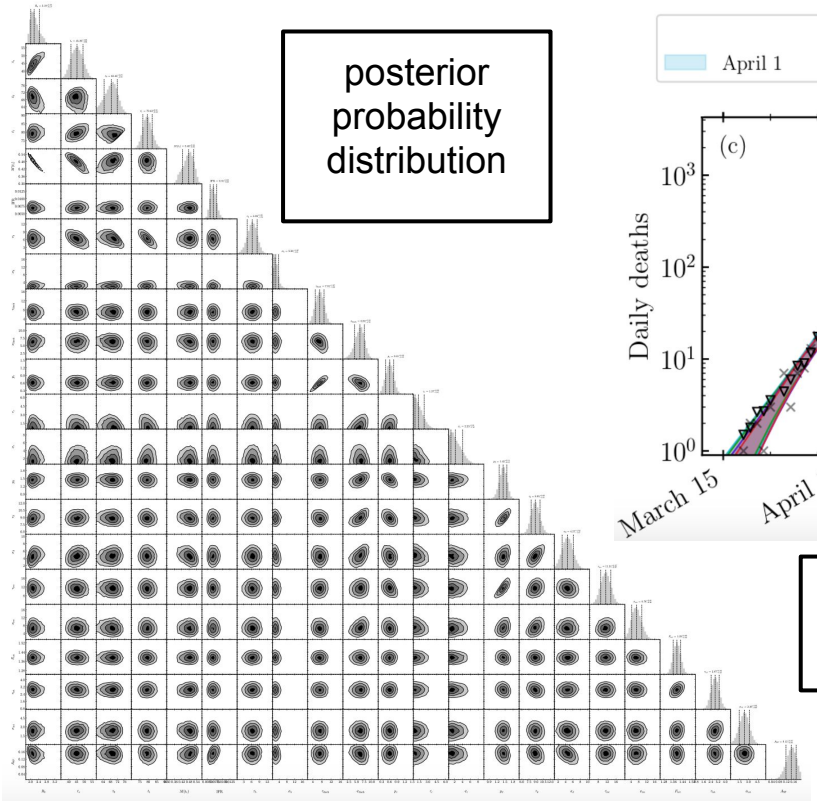
Find the model parameters that are **most likely to produced observed data**

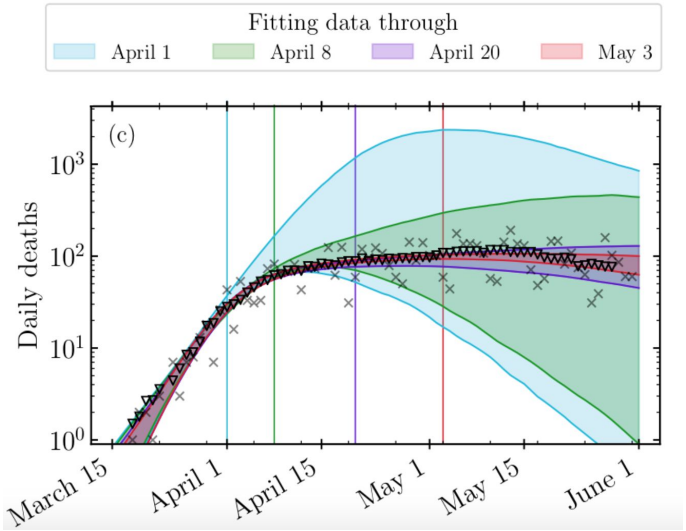


Use **Monte carlo Markov Chain** to maximize p over θ

Example calibration & correlation

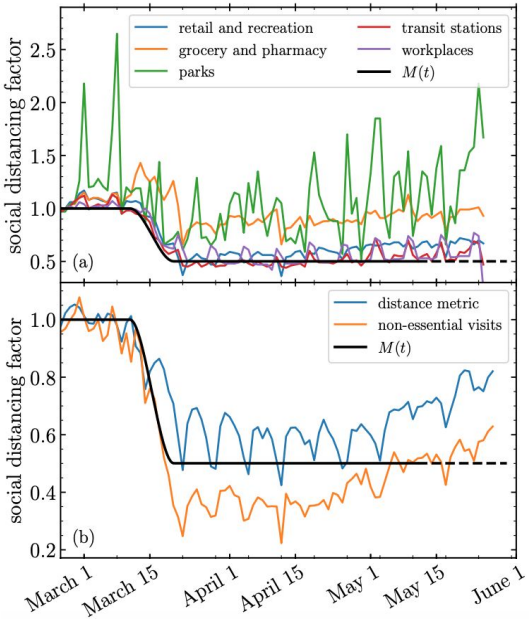
posterior probability distribution





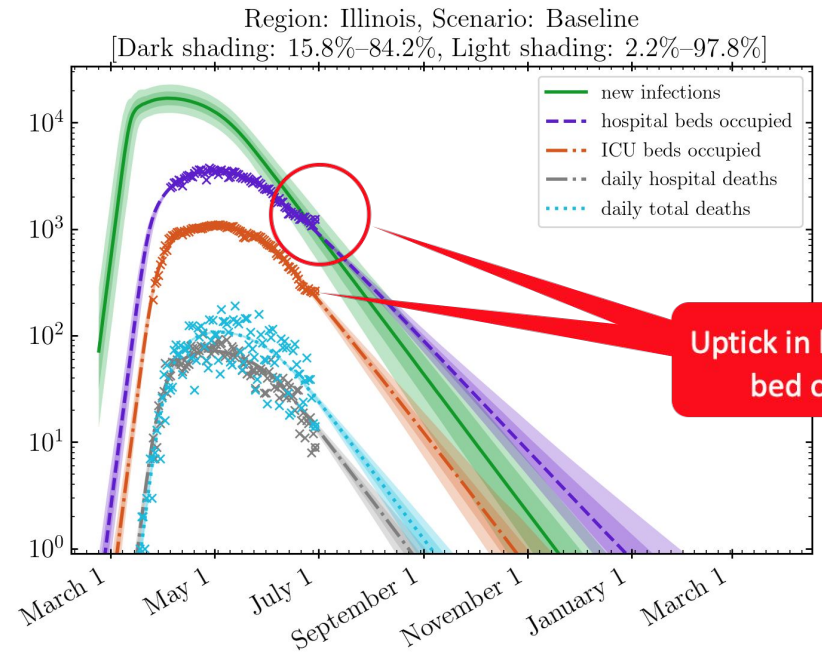
model robustness to new data

correlations with population mobility



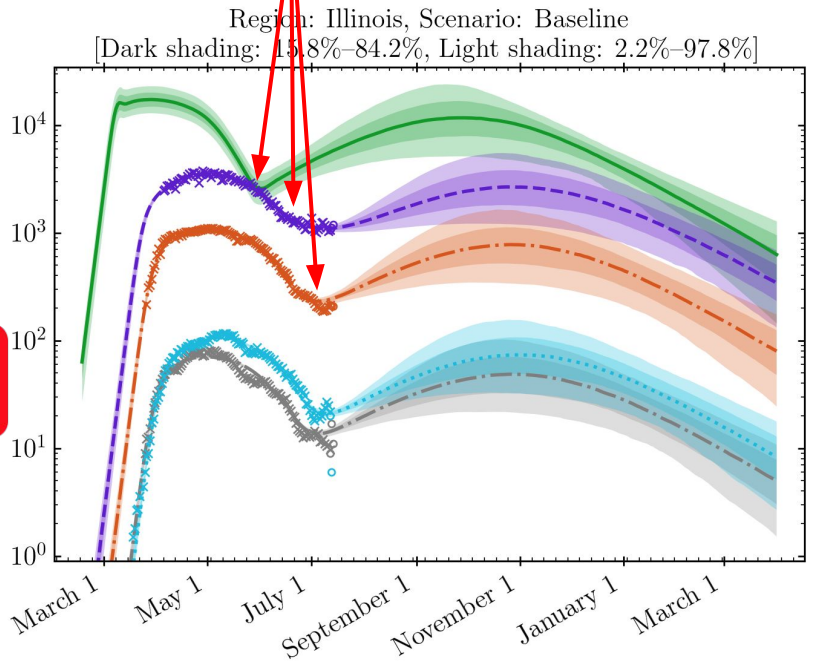
Early warning system

Restore Illinois:
phase 3 / 4 “no change” model tension



Uptick in hospital & ICU bed occupancy

Delay in signal from infections to hospitalizations, &c. allows for early-warning predictions



Learn more...

Model details (data sources, calibration procedures and comparisons, &c.) have been published.

Especially see references!

Production code is public

https://github.com/uiuc-covid19-modeling/pydemic

medRxiv preprint doi: https://doi.org/10.1101/2020.06.03.20120691.this version posted June 17, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Modeling COVID-19 dynamics in Illinois under non-pharmaceutical interventions

George N. Wong[1,†], Zachary J. Weiner[1,†], Alexei V. Tkachenko[2], Ahmed Elbanna[3], Sergei Maslov[1,4,5,*], and Nigel Goldenfeld[1,5,*]

[1]Department of Physics, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA
[2]Center for Functional Nanomaterials, Brookhaven National Laboratory, Upton, NY 11973, USA
[3]Department of Civil Engineering, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA
[4]Department of Bioengineering, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA
[5]Carl R. Woese Institute for Genomic Biology, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA

* Correspondence to: maslov@illinois.edu and nigel@illinois.edu.
† These authors contributed equally to this work.

(Dated: June 16, 2020)

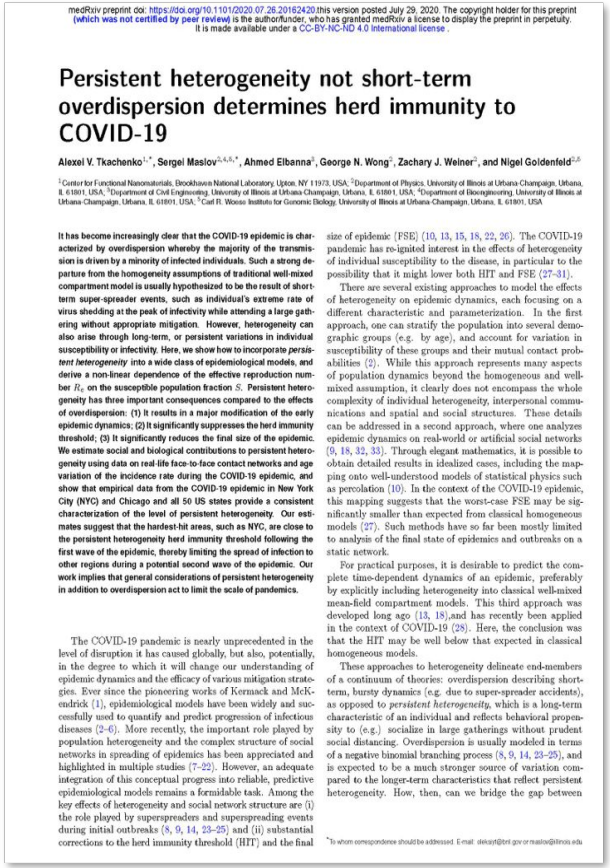
We present modeling of the COVID-19 epidemic in Illinois, USA, capturing the implementation of a Stay-at-Home order and scenarios for its eventual release. We use a non-Markovian age-of-infection model that is capable of handling long and variable time delays without changing its model topology. Bayesian estimation of model parameters is carried out using Markov Chain Monte Carlo (MCMC) methods. This framework allows us to treat all available input information, including both the previously published parameters of the epidemic and available local data, in a uniform manner. To accurately model deaths as well as demand on the healthcare system, we calibrate our predictions to total and in-hospital deaths as well as hospital and ICU bed occupancy by COVID-19 patients. We apply this model not only to the state as a whole but also its sub-regions in order to account for the wide disparities in population size and density. Without prior information on non-pharmaceutical interventions (NPIs), the model independently reproduces a mitigation trend closely matching mobility data reported by Google and Unacast. Forward predictions of the model provide robust estimates of the peak position and severity and also enable forecasting the regional-dependent results of releasing Stay-at-Home orders. The resulting highly constrained narrative of the epidemic is able to provide estimates of its unseen progression and inform scenarios for sustainable monitoring and control of the epidemic.

On January 24, 2020, the second known COVID-19 case to be diagnosed in the USA was reported in Chicago, Illinois. Community transmission of the disease was confirmed on March 8, 2020. During the subsequent ten days, the epidemic grew with a case doubling time of approximately 2.3 days, while testing capacity was essentially fixed. On March 21, 2020, a Stay-at-Home order was issued for the entire state of Illinois and subsequently extended on March 31, 2020 and again on April 23, 2020. The order was lifted on May 30, 2020 [1]. Responsible relaxation of the mitigation of COVID-19 must be informed by realistic and well-calibrated epidemiological modeling of the outcomes of any scenarios under consideration—not just of the resulting (increased) death toll but also of the stress placed upon the healthcare system. The purpose of this report is to present such an analysis.

A variety of modeling approaches are used by hospitals, public health officials, and state governments. These range between phenomenological models that use a curve-fitting procedure to match data, such as the daily death rate, and mechanistic methods that model the trajectory of the epidemic as individuals transition through several disease and healthcare-bound stages [2–5]. Only mechanistic models are able to make justifiable predictions while accounting for changes in the epidemic environment, such as the imposition or relaxation of community mitigation efforts. Of these, compartmental models like the Susceptible-Infectious-Recovered (SIR) models, and Susceptible-Exposed-Infectious-Recovered (SEIR) extensions, are widely used. Compartmental models describe how fractions of a homogeneous, well-mixed population progress through different states the disease, driven by interactions between infectious and susceptible individuals. In the simplest models, the dynamics is deterministic and the rates are constant in time, but many variants and extensions exist and are widely used.

In order to be practically useful, models must be calibrated to observed data [4, 6–8]. We calibrate the important dynamics of the model to several simultaneous streams of empirical data including total and in-hospital deaths, as well as hospital and ICU bed occupancy by COVID-19 patients. To avoid biases resulting from non-uniform and non-constant testing rates, which may be difficult to parameterize, we do not consider positive case data. The resulting model is a description of the epidemic as it progresses through the hospital system in Illinois; as it is clear that a non-negligible number of COVID-19 deaths occur outside the hospital environment (e.g., in homes and nursing homes especially), we augment our model with an effective description of the net incidence of deaths due to COVID-19.

There are many limitations to the types of models that we and others use to describe COVID-19, and these have been explored extensively in the literature, especially with regard to spatial structure [9], superspreader events and individuals [10–13], and the structure of contact networks [14–16]. A geographical region as large as the state of Illinois is not well-described as homogeneous, due to



medRxiv preprint doi: https://doi.org/10.1101/2020.07.26.20162420.this version posted July 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Persistent heterogeneity not short-term overdispersion determines herd immunity to COVID-19

Alexei V. Tkachenko[1,*], Sergei Maslov[2,4,5,*], Ahmed Elbanna[3], George N. Wong[2], Zachary J. Weiner[2], and Nigel Goldenfeld[2,5]

[1]Center for Functional Nanomaterials, Brookhaven National Laboratory, Upton, NY 11973, USA. [2]Department of Physics, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA. [3]Department of Civil Engineering, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA. [4]Department of Bioengineering, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA. [5]Carl R. Woese Institute for Genomic Biology, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA

It has become increasingly clear that the COVID-19 epidemic is characterized by overdispersion whereby the majority of the transmission is driven by a minority of infected individuals. Such a strong departure from the homogeneity assumptions of traditional well-mixed compartment model is usually hypothesized to be the result of short-term super-spreader events, such as individual's extreme rate of virus shedding at the peak of infectivity while attending a large gathering without appropriate mitigation. However, heterogeneity can also arise through long-term, or persistent variations in individual susceptibility or infectivity. Here, we show how to incorporate *persistent heterogeneity* into a wide class of epidemiological models, and derive a non-linear dependence of the effective reproduction number R_e on the susceptible population fraction S. Persistent heterogeneity has three important consequences compared to the effects of overdispersion: (1) It results in a major modification of the early epidemic dynamics; (2) It significantly suppresses the herd immunity threshold; (3) It significantly reduces the final size of the epidemic. We estimate social and biological contributions to persistent heterogeneity using data on real-life face-to-face contact networks and age variation of the incidence rate during the COVID-19 epidemic, and show that empirical data from the COVID-19 epidemic in New York City (NYC) and Chicago and all 50 US states provide a consistent characterization of the level of persistent heterogeneity. Our estimates suggest that the hardest-hit areas, such as NYC, are close to the persistent heterogeneity herd immunity threshold following the first wave of the epidemic, thereby limiting the spread of infection to other regions during a potential second wave of the epidemic. Our work implies that general considerations of persistent heterogeneity in addition to overdispersion act to limit the scale of pandemics.

The COVID-19 pandemic is nearly unprecedented in the level of disruption it has caused globally, but also, potentially, in the degree to which it will change our understanding of epidemic dynamics and the efficacy of various mitigation strategies. Ever since the pioneering works of Kermack and McKendrick (1), epidemiological models have been widely and successfully used to quantify and predict progression of infectious diseases (2–6). More recently, the important role played by population heterogeneity and the complex structure of social networks in spreading of epidemics has been appreciated and highlighted in multiple studies (7–22). However, an adequate integration of this conceptual progress into reliable, predictive epidemiological models remains a formidable task. Among the key effects of heterogeneity and social network structure are (i) the role played by superspreaders and superspreading events during initial outbreaks (8, 9, 14, 23–25) and (ii) substantial corrections to the herd immunity threshold (HIT) and the final size of epidemic (FSE) (10, 13, 15, 18, 22, 26). The COVID-19 pandemic has re-ignited interest in the effects of heterogeneity of individual susceptibility to the disease, in particular to the possibility that it might lower both HIT and FSE (27–31).

There are several existing approaches to model the effects of heterogeneity on epidemic dynamics, each focusing on a different characteristic and parameterization. In the first approach, one can stratify the population into several demographic groups (e.g. by age), and account for variation in susceptibility of these groups and their mutual contact probabilities (2). While this approach represents many aspects of population dynamics beyond the homogeneous and well-mixed assumption, it clearly does not encompass the whole complexity of individual heterogeneity, interpersonal communications and spatial and social structures. These details can be addressed in a second approach, where one analyzes epidemic dynamics on real-world or artificial social networks (9, 18, 32, 33). Through elegant mathematics, it is possible to obtain detailed results in idealized cases, including the mapping onto well-understood models of statistical physics such as percolation (10). In the context of the COVID-19 epidemic, this mapping suggests that the worst-case FSE may be significantly smaller than expected from classical homogeneous models (27). Such methods have so far been mostly limited to analysis of the final state of epidemics and outbreaks on a static network.

For practical purposes, it is desirable to predict the complete time-dependent dynamics of an epidemic, preferably by explicitly including heterogeneity into classical well-mixed mean-field compartment models. This third approach was developed long ago (13, 18),and has recently been applied in the context of COVID-19 (28). Here, the conclusion was that the HIT may be well below that expected in classical homogeneous models.

These approaches to heterogeneity delineate end-members of a continuum of theories: overdispersion describing short-term, bursty dynamics (e.g. due to super-spreader accidents), as opposed to *persistent heterogeneity*, which is a long-term characteristic of an individual and reflects behavioral propensity to (e.g.) socialize in large gatherings without prudent social distancing. Overdispersion is usually modeled in terms of a negative binomial branching process (8, 9, 14, 23–25), and is expected to be a much stronger source of variation compared to the longer-term characteristics that reflect persistent heterogeneity. How, then, can we bridge the gap between

*To whom correspondence should be addressed. E-mail: alexjyt@bnl.gov or maslov@illinois.edu

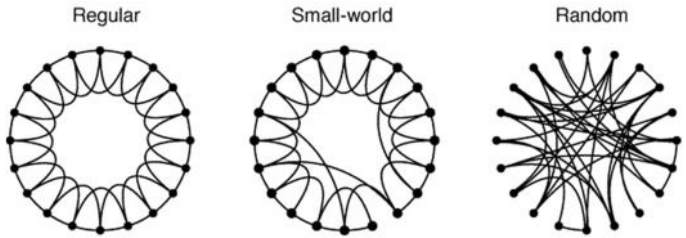
Modeling a university population

The mean-field model deals in **effective parameters** that approximate network heterogeneity, mitigation/intervention measurements, changing timetables, …

Unless the relationships between real world details and the effective parameters are well understood, guessing parameter values begs the question.

Idea: **explicitly treat known network structure** (class schedules, number of restaurants, room volumes, …) and marginalize over uncertainty.

⇒ use agent-based models



Watts+ 1998

Agent-based model overview

Independently track location & infection state of (40k) campus-bound students, faculty, staff

Include complete course schedule, estimate out-of-class schedule

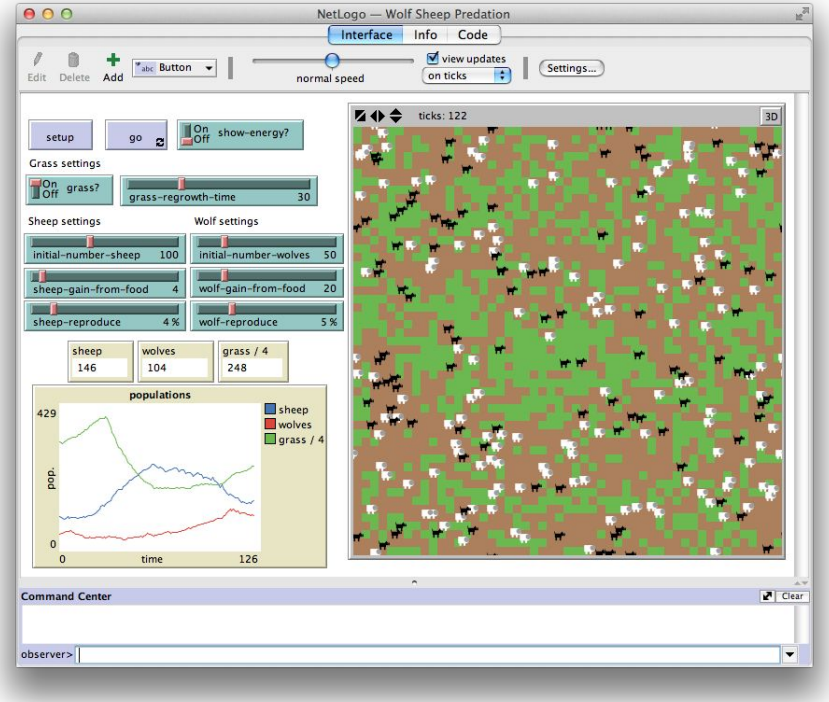
Compute ingested viral quanta based on proximity

Set disease profiles based on literature

Simulate contact tracing by proximity

Simulate effects of quarantine and isolation

Netlogo, an off-the-shelf ABM simulator



Agent-based modeling is hard

- Don’t know details of disease infectivity
- Don’t know (e.g.) airflow patterns in classrooms, bars, libraries, dorms...
- Don’t know effects of interventions
- Under-constrained model for “return to campus”
- Making sense of contact tracing data requires understanding infection
- Student social life (before & after COVID) under-constrained
- Hard to estimate compliance / failures of contact tracing

Agent-based modeling is hard

… but it is necessary

- Produce multiple scenarios, marginalize over uncertainty, update model as time goes by and more data is available
- Identify general warning trends
- Estimate effects of different mitigation strategies
- Exploration → understanding

Better is good.

Agent-based model: infection detail

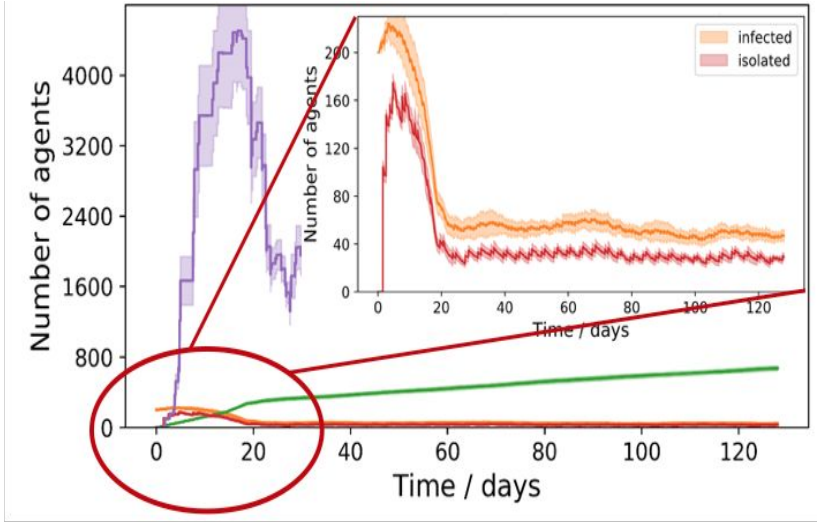
- The world comprises **zones**, physical locations with volumes and airflow rates
- Each agent has a **schedule** that defines when to be in which zones
- Each agent has **internal infection timers**, that track disease progression
- If an agent is infected, they **deposit viral quanta** into zones as they move
- Viral quanta are **localized** and **decay with time** according to ventilation
- A viral quantum is an **infection probability**
- Individuals are **infected according to ingested viral quanta** when leaving a zone

Agent-based model: mitigation detail

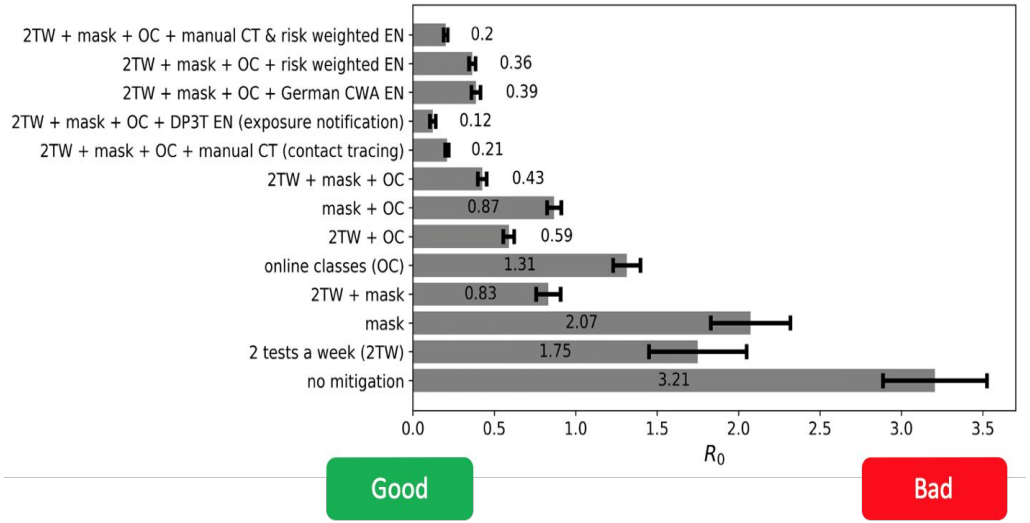
- Explore size **threshold for shift to online** classes (remove classes from schedule)
- Vary **testing frequency** per demographic
- Limit **indoor population density** (e.g., restaurants, bars, …)
- Vary the contact tracing app **adoption rate**
- Effects of **quarantine/isolation compliance**, threshold for sustainability
- Investigate effect of **mask ordinances** (in classrooms, libraries, buses, outside)

Agent-based model: data products

simulated epidemic trajectories



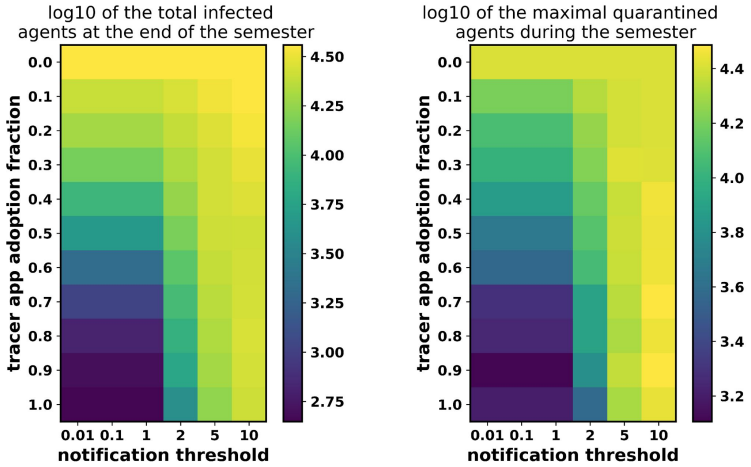
estimated mitigation effectiveness



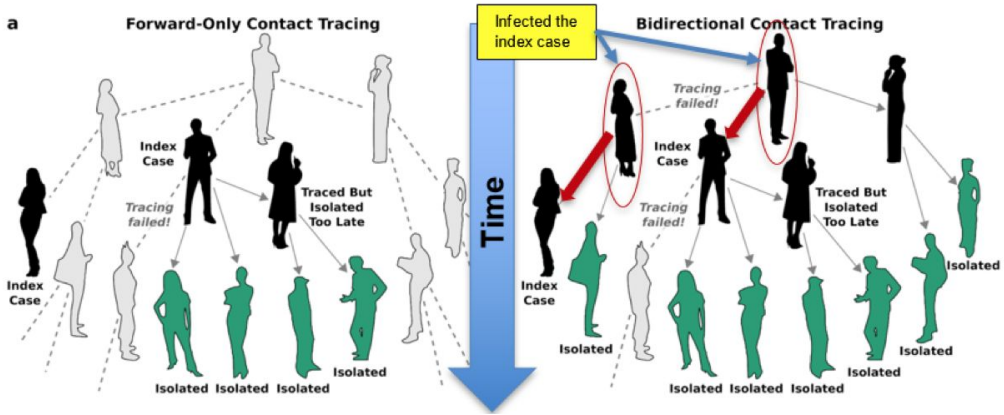
Agent-based model: contact tracing methods

Cost/benefit tradeoff between too many notifications (→ ignored) and too few notifications (→ insufficient containment).

Explore effectiveness of forward- versus bidirectional contact tracing.



Wang+

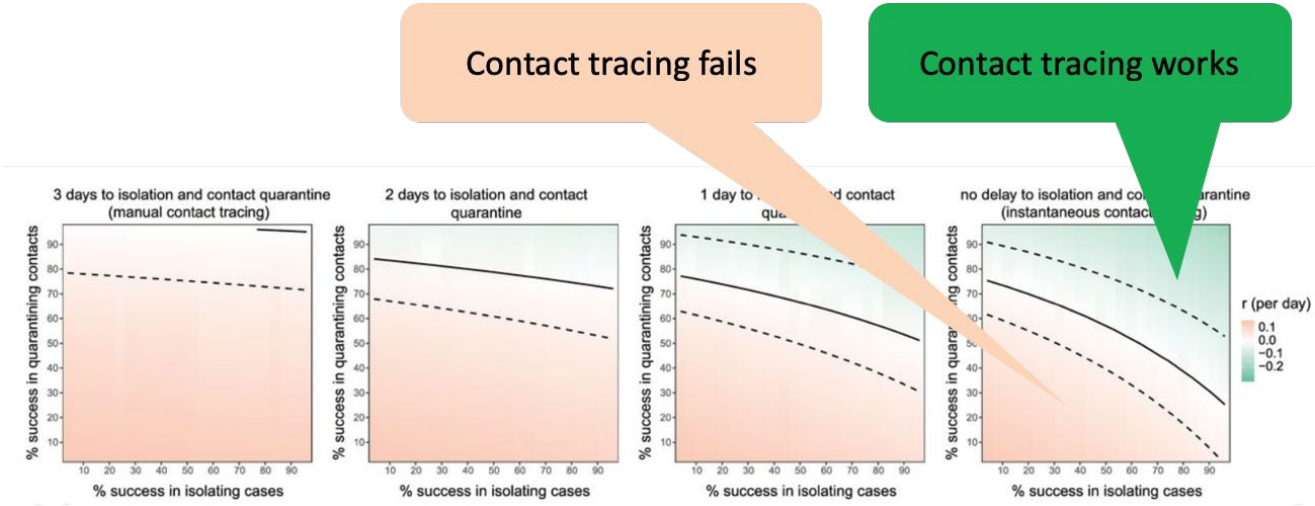


Bradshaw+

Agent-based model: contact tracing methods

Minimizing the delay between identification and quarantine/isolation **is crucial**!

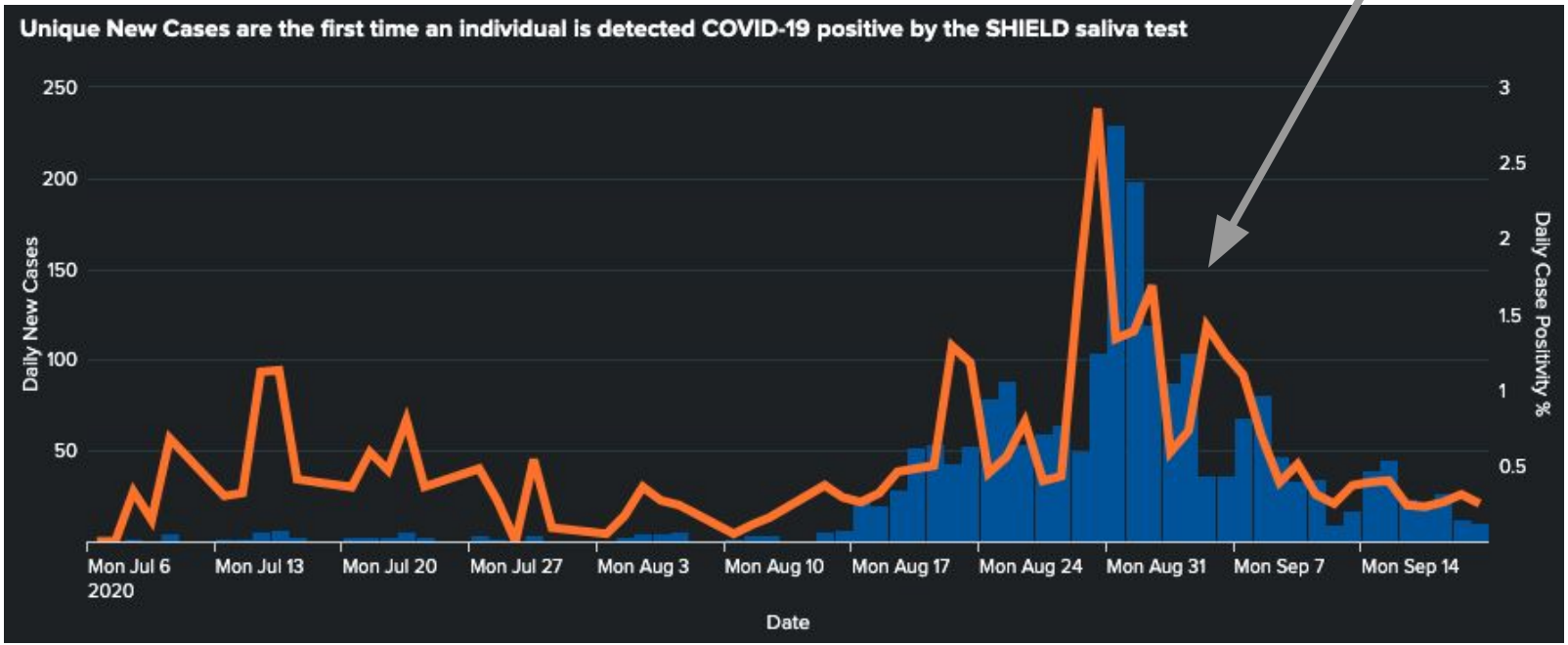
If delay > 2 days, contact tracing **will not work**.



Ferretti+ 2020

Agent-based model: results

Definitive plan is effective!



Thank you!