# Biocomplexity and AI: A symbiotic relationship

Madhav Marathe (marathe@virginia.edu)
Biocomplexity Institute &
Department of Computer Science
University of Virginia

https://biocomplexity.virginia.edu/person/madhavmarathe
https://nssac.github.io/sociotechnicaldigitaltwins
https://net.science/home

**TR BI-2026-7**

**Support:**

NIH R01GM109718, NSF CINES OAC-1916805, NSF DIBBS ACI-1443054, NSF BIG DATA IIS-1633028, NSF Expeditions in Computing CCF-1918656, NSF PREPARE CNS-2041952, NSF Eager CCF-1917819, DTRA, US Centers for Disease Control and Prevention 75D30119C05935, CDC PGCOE, CDC CSTA, VDH UVA-BIO610-GY23, UVA Contagion Science program

Thanks to Siva, Rajesh, Narahari, Vijay and others at ICTS and IISc and to members of the research team, our students, postdocs, and collaborators at SUNY Albany, CDC, NEU, NIH, DTRA, Georgia Tech, JHU, UMD, Princeton, Stanford, VT, Yale, Argonne, IISc, Pittsburgh and Purdue Supercomputing Centers, Persistent Systems, and others.

**43 administrative & research staff**

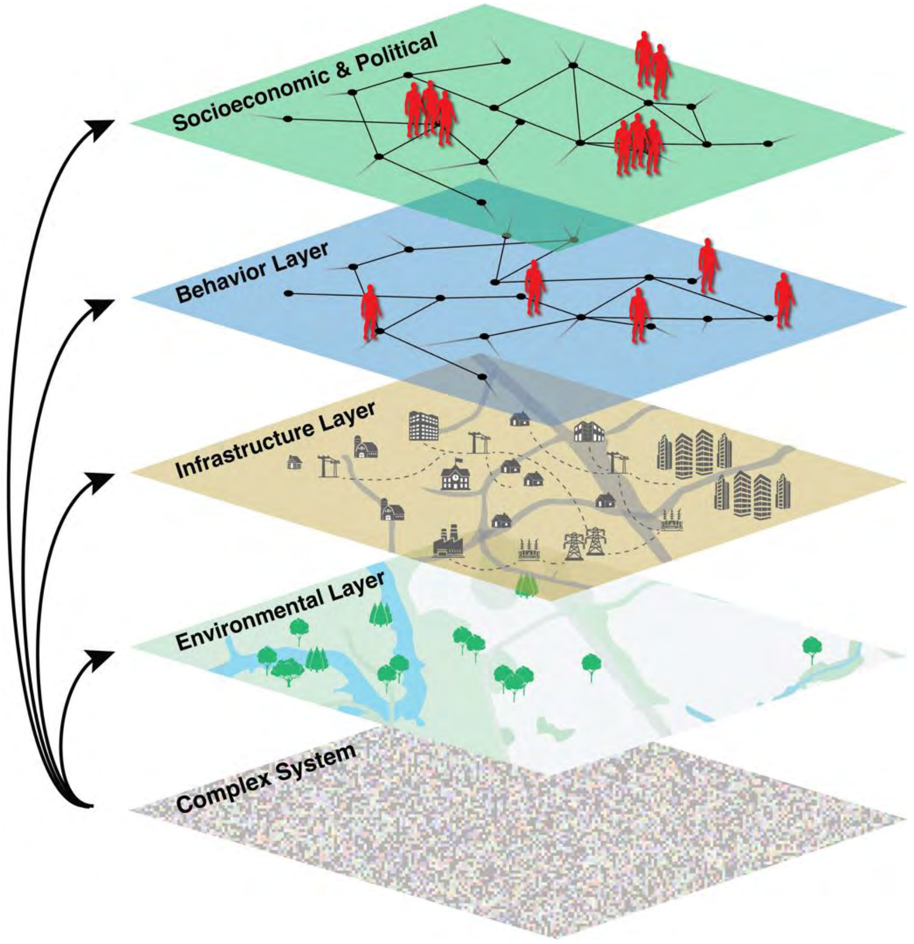
**17–20 graduate students**

**6–25 undergraduate students**

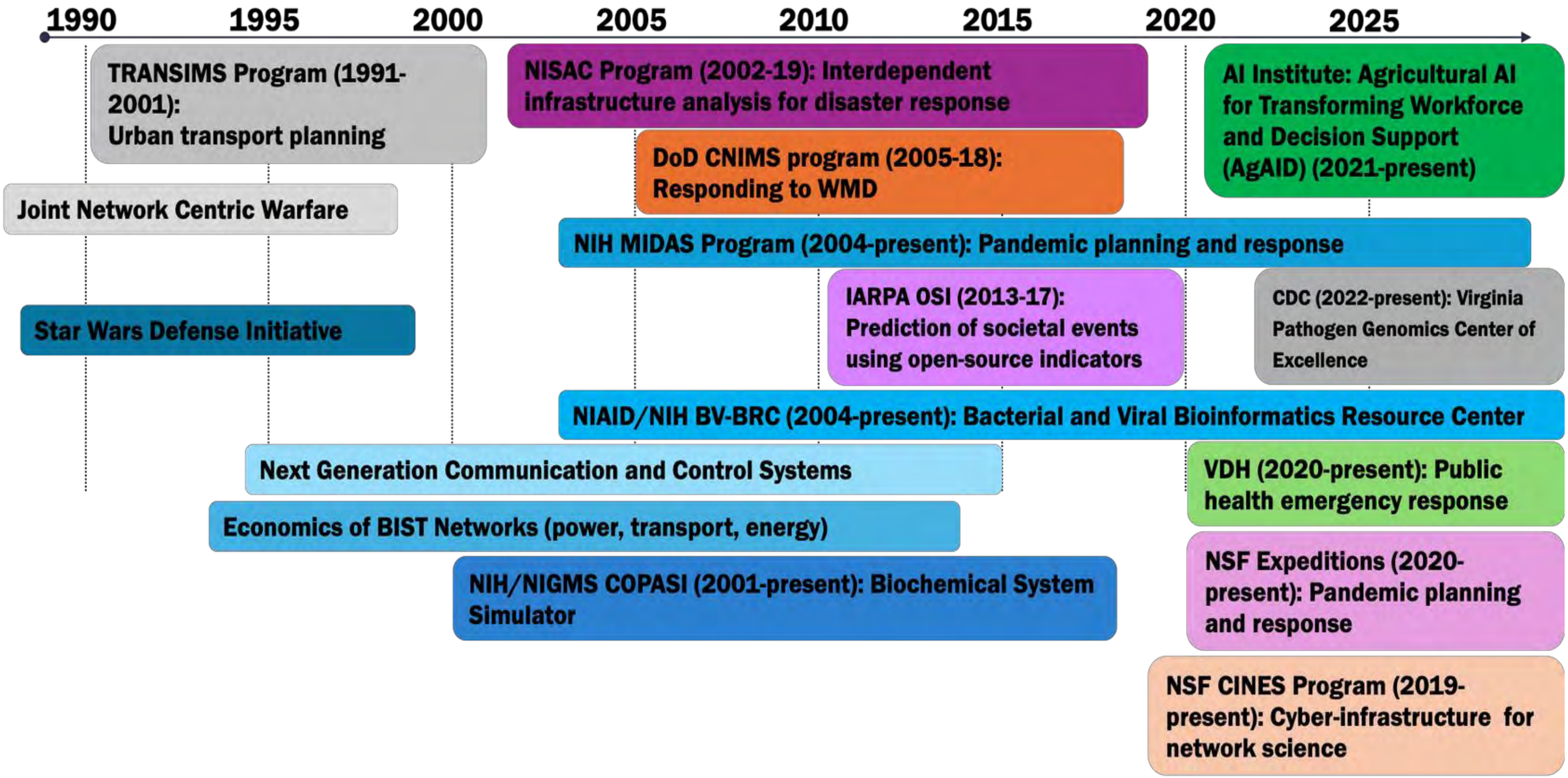
# Biocomplexity Institute

**Biocomplexity refers to the interactions of living and non-living agents and ecologies, including humans and their enabling technologies.**

- Mission-driven projects that lead to fundamental advances in science and technology
- Problems that cannot be solved by a single faculty member or within the narrow boundaries of a discipline
- Problems that require *sophisticated tools and diverse intellectual resources*
- Tangible societal and scholarly impact (over 350 publications)



# A Brief History



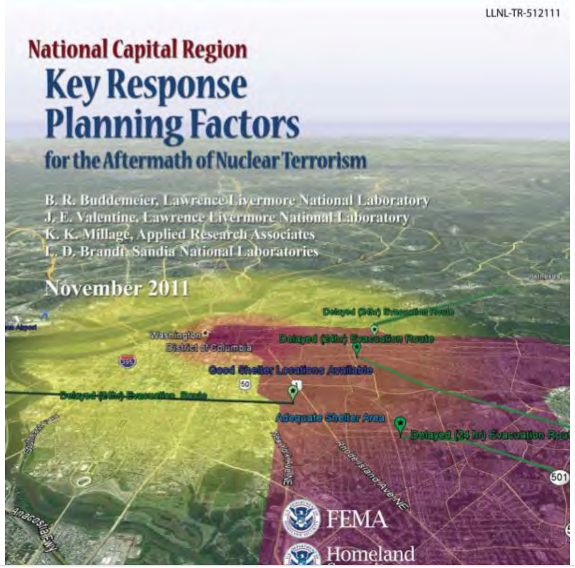
# TRANSIMS:
# Urban Transport Planning Modeling Environment

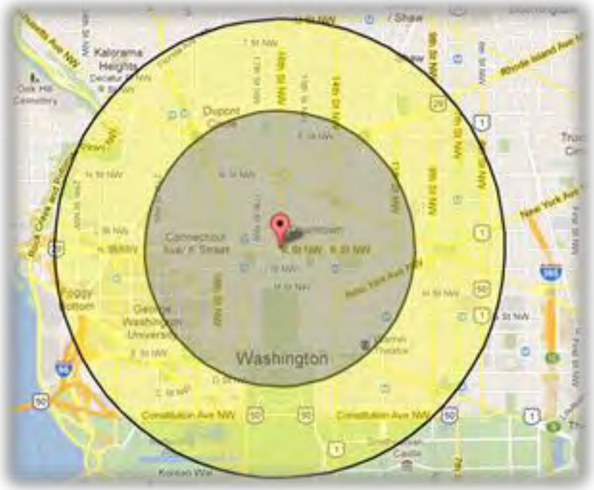
- TRANSIMS: (1991-2001)
  - Set up to address important national problems
  - Assessing traffic flows during congestion
  - Predicting household behavior when the infrastructure changes
  - Assessing social justice issues in proposed infrastructure changes
  - Predicting environmental impact
- Requirements set by:
  - Government (DOT), regional planning organizations, university researchers & transportation consultants
- Technology developed in collaborations with MPO
  - Developed by team at LANL
  - Case studies with Albuquerque, Dallas, Portland MPOs
- TRANSIMS is currently an open-source system managed by Argonne National Laboratory

*Figure 1 – White House Area Street Closures*

https://www.ncpc.gov/docs/WHATS_Final_Report.pdf

# National Planning Scenario 1




## Hypothetical Scenario:

- Unannounced 10 kt detonation of an Improvised Nuclear Device (IND)
- 11:15am May 15$^{th}$, 2006
- 16$^{th}$ and K Street, Washington DC


## Traditional Focus

- Prompt impact on overall health
- Responders and their safety
- Victims as essentially passive
- Cold war scenarios or low-level exposure to radiation

*WSC 2015, AAMAS JAAMAS, Mitch Waldrop's articles in Science and PNAS*

# What did we learn

Even a *partially restored* communication system has disproportionately positive impact on the overall behavior:

- Fewer deaths, better health outcomes & reduction in anxiety
- Policy question: how do we build a self forming comm network?

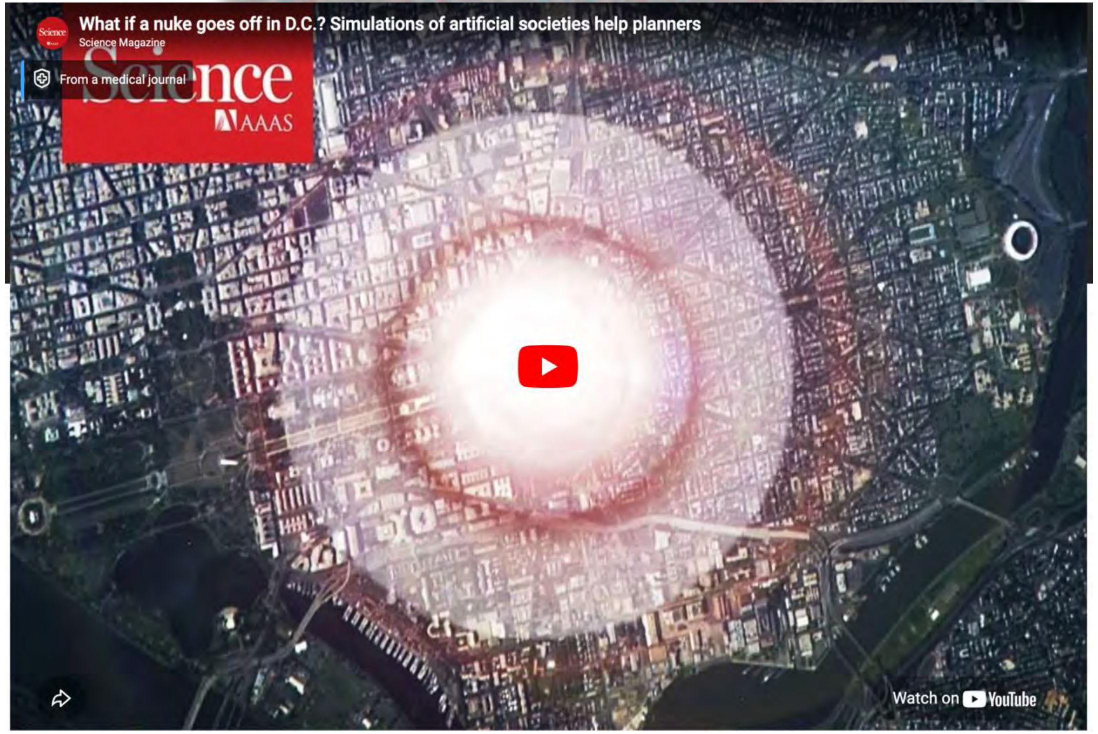
Large part of power network unlikely to be operational for >2 years

- NO significant cascading failures beyond a small area likely
- Important implications on how the city and its surroundings will be reconstituted

<u>*Individual emergent behaviors*</u>: personal choices + contextualized dynamic decisions + institutionalized processes and policies

Models: understandable, explainable & computationally efficient

- Individual and collective behavior and behavioral adaptation
- Sensitive to contextual information



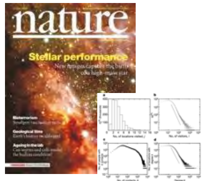

UVA | BIOCOMPLEXITY INSTITUTE

# Pandemic Modeling and Response: 2000–Present

**COVID-19 resource page:** https://covid19.biocomplexity.virginia.edu/
**COVID work done by NSSAC:** https://nssac.github.io/covid-19/index
**NSF Expeditions Project:** https://computational-epidemiology.org
**NSF PREPARE: Virtual Organization Project:** https://prepare-vo.org

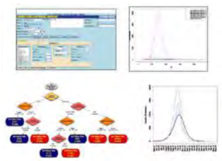
# Since 2004 BI has supported federal agencies during all major outbreak response efforts

2004
Smallpox

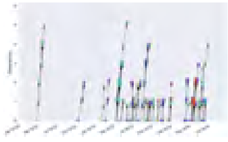



2005
H5N1

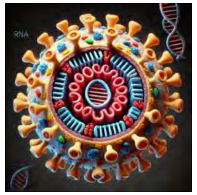

2009
H1N1



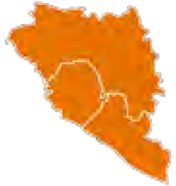
2012
MERS



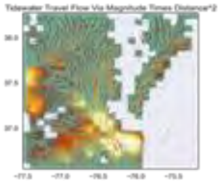
2013
Influenza



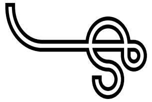
2014
Ebola



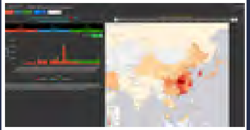
2016
Zika



2018
Ebola



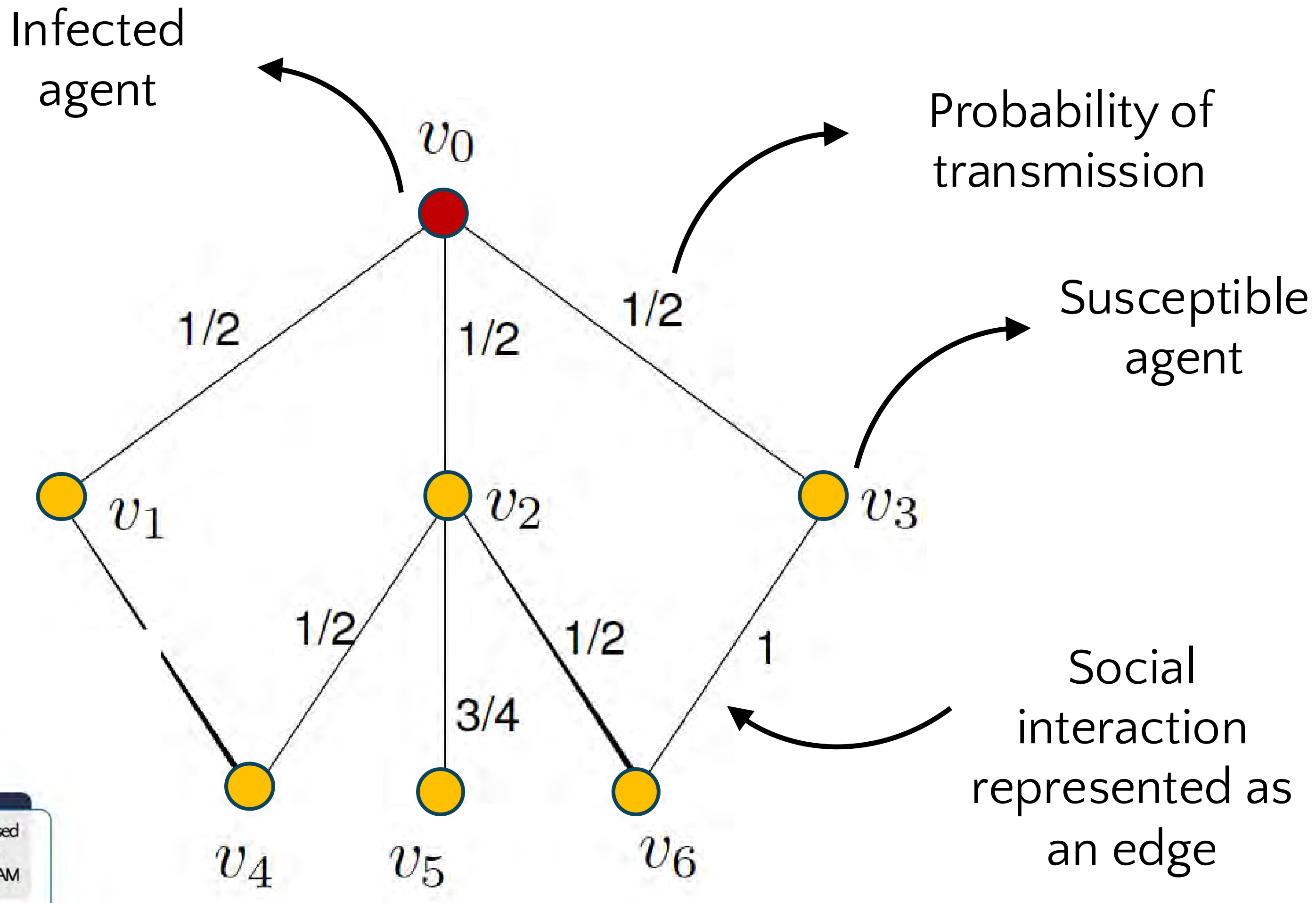
2020
COVID-19

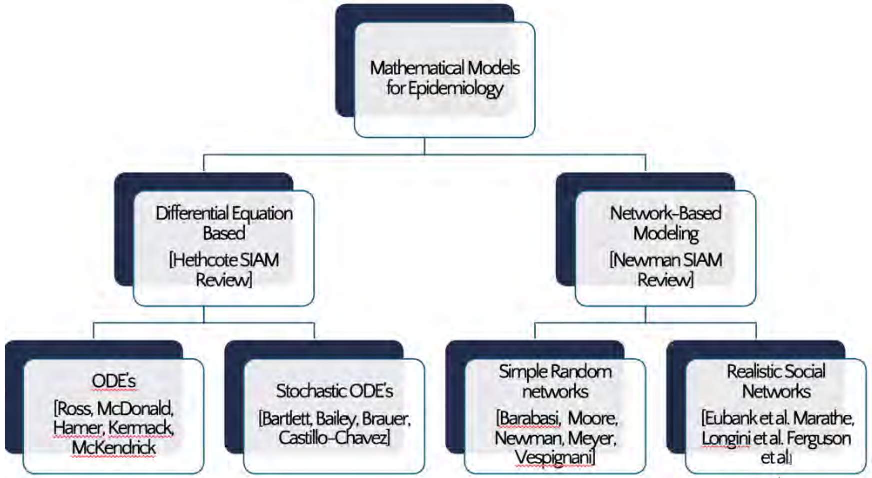


2024
H5N1

# Models based on networks

| | |
|---|---|
| 𝕊 | 𝕊usceptible |
| 𝕀 | 𝕀nfected |
| ℝ | ℝecovered |



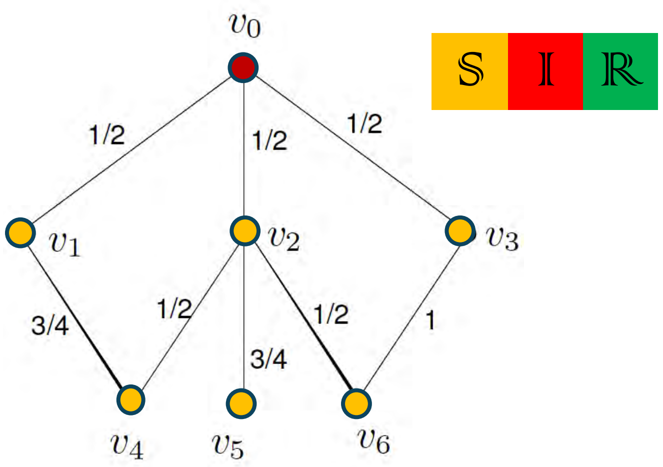




Pandemic Informatics: Preparation, Robustness, and Resilience A Computing Community Consortium (CCC) Quadrennial Paper, E. Bradley, M. Marathe, M. Moses, W. Gropp, and D. Lopresti, 2020.

UVA | BIOCOMPLEXITY INSTITUTE

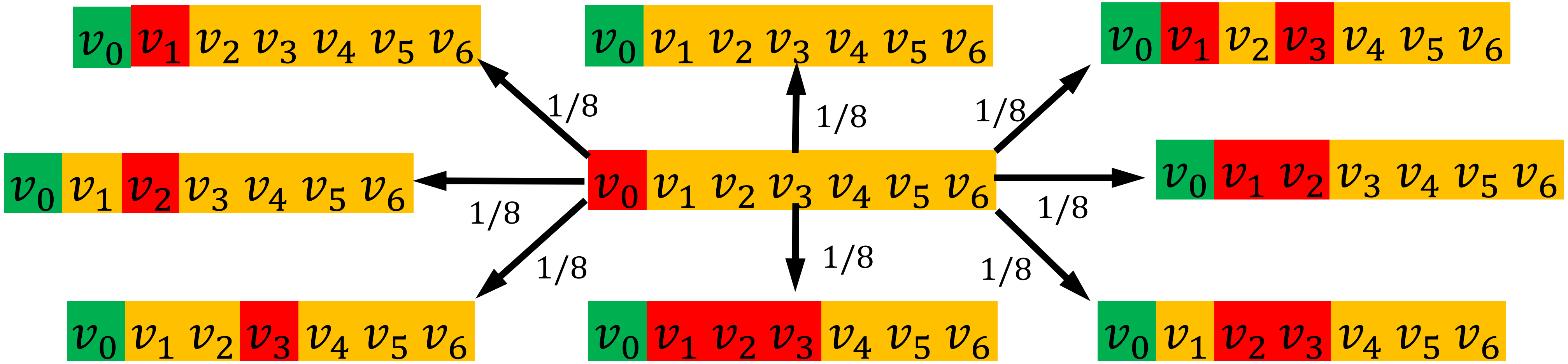
# Understanding the dynamics requires the study of exponentially large system

Temporal Dynamics represented by Markov Chain (just a fragment is depicted)




Markov chain (partial) exponentially larger than the network

A social network with $\boldsymbol{n}$ nodes & **3** disease states per nodes yields a chain of size $\boldsymbol{3^n}$




UVA | BIOCOMPLEXITY INSTITUTE

# The reachability problem

- Given, time $t = 0$,
  - A fixed network $G(V, E)$ with $n$ vertices and $m$ edges
  - A single infected individual
  - SIR dynamics with infectious and incubation period (more generally we are given a within host disease manifestation function)
  - The transmission probability on edge $\{u, v\}$ is $p_{\{u,v\}}$
- Question:
  - (Reachability) Find the state of the system after $t$ time units
  - **State:** we can consider various functional forms, e.g. the dendrogram, the total number of infections until $t$, number of active infections at time $t' \leq t$
  - Given that the system is stochastic, each state estimation usually comes with an estimate of the probability associated with that state

*Rosenkrantz, Levin, Marathe et al. PNAS*

# Computational cost of analyzing SIR dynamics

- The state assessment problem is **#P**-hard
  - *Implies that simulations are needed for understanding the dynamics*
  - Even short-term state assessment is hard
  - Result due to stochastic nature of the dynamical process
  - Similar results hold for inference and intervention analysis problems
- *Associated Markov chain size – astronomical in the size of the description*
  - *Network with $10^6$ nodes & (3 states: S,I,R), the chain has $3^{10^6}$ nodes; impossible to do naïve search*

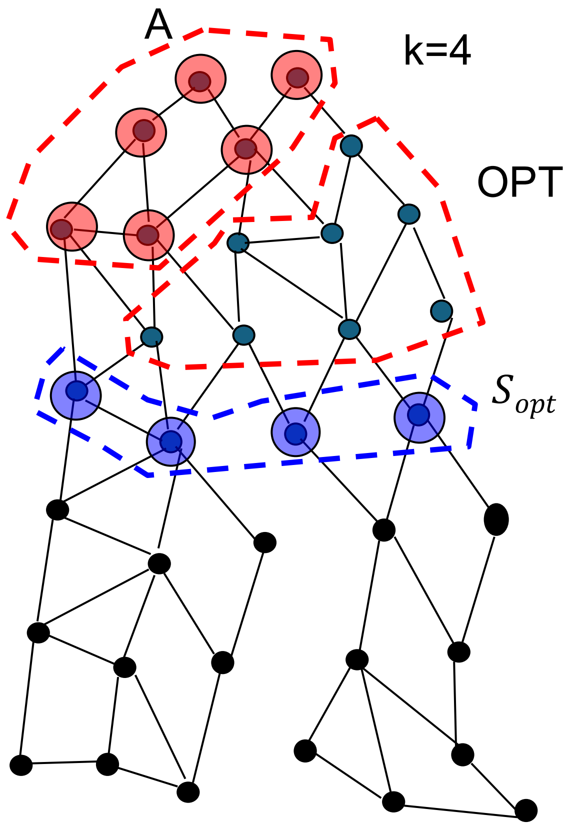
Simulations are therefore necessary to solve this problem

Rosenkrantz DJ, Vullikanti A, Ravi SS, Stearns RE, Levin S, Poor HV, Marathe MV. Fundamental limitations on efficiently forecasting certain epidemic measures in network models. *Proceedings of the National Academy of Sciences*. 2022 Jan 25;119(4):e2109228119.

UVA | BIOCOMPLEXITY INSTITUTE

# Control problems: Vaccine allocation

**Given**: A fixed network $G(V, E)$ with $n$ vertices and $m$ edges. SIR dynamics with infectious and incubation period = 1. The transmission probability on edge $\{u, v\}$ is $p_{\{u,v\}}$. Initial infected set $A$ and budget $k$ for #vaccines
**Objective**: Select set $S$ (of size $k$) to vaccinate so that expected #infections is minimized




- **Theorem:** $\forall\, \mu \geq 1, and\ p_{\{u,v\}}$ = 1, a polynomial time algorithm that vaccinates $(1 + \mu)k$ nodes & results in $\leq (1 + \frac{1}{\mu})$OPT infected people. Here OPT denotes the optimal number of infections that when $k$ individuals are vaccinated

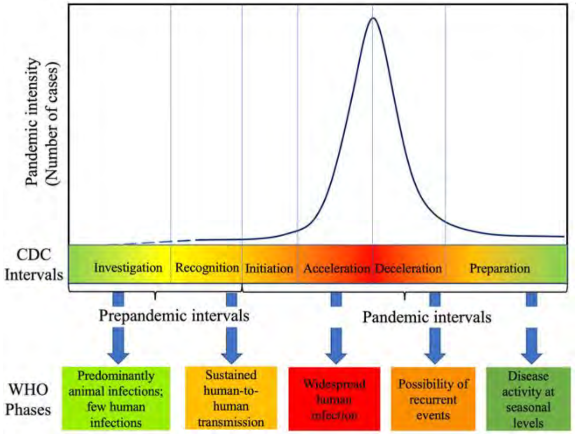
*Eubank, Vullkanti et al: DIMACS, SODA*

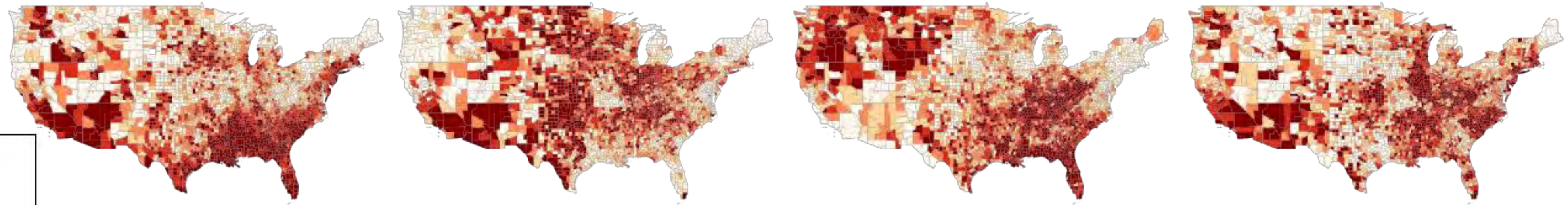
UVA | BIOCOMPLEXITY INSTITUTE

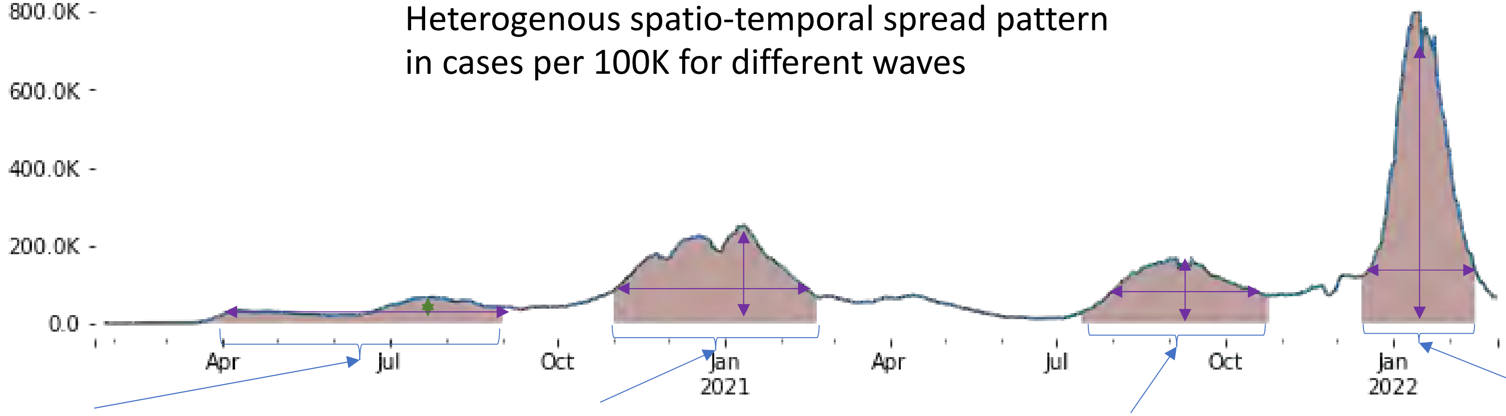
# COVID-Response

**COVID-19 resource page:** https://covid19.biocomplexity.virginia.edu/
**COVID work done by NSSAC:** https://nssac.github.io/covid-19/index
**NSF Expeditions Project:** https://computational-epidemiology.org
**NSF PREPARE: Virtual Organization Project:** https://prepare-vo.org

# Evolution of COVID-19 pandemic was complicated






Heterogenous spatio-temporal spread pattern in cases per 100K for different waves

Varying duration, peak Intensity, and growth rates across waves

Reasons: Varying interventions, testing capacity, transmissibility characteristics of variants, etc.

First wave (wild type):
- Lack of immunity
- Mixing in the population during the initial phase

Measures:
- **Extreme restrictions** (lockdown)
- Mask wearing

Second wave:
- **Relaxed social interventions, holidays, university reopening**

Measures:
- **Behavioral adaptations**
- Mask wearing
- Partially vaccinated population

Third wave (Delta):
- Relaxed social interventions
- **Unvaccinated individuals**

Measures:
- **Vaccinations**
- Behavioral adaptations
- Mask wearing

Fourth wave (Omicron):
- **Immune escape**
- Relaxed social interventions

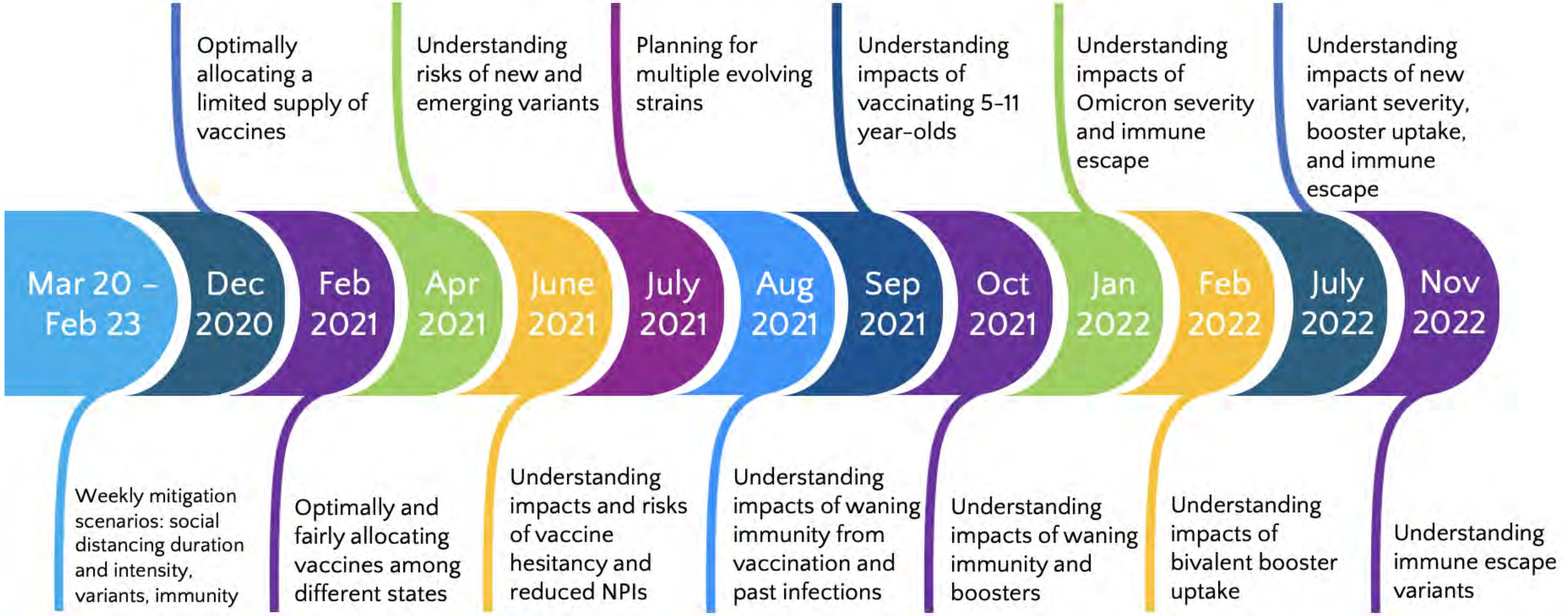
Measures:
- **Booster vaccinations**
- Behavioral adaptations

UVA BIOCOMPLEXITY INSTITUTE

# Real-time operational response during a pandemic

- **Real-time**: Strict deadlines drive the time to build, test & complete tasks
- **Operational**: Problems derived from current needs; resultin analysis should be useful to decision makers and need not be perfect every time
- **Evolving**: The pandemic evolves constantly -- pipeline needs constant adaptation
  - Acquire, process, organize and ingest diverse data
    - Vaccine uptake, mobility patterns, interventions, …
  - Modify, enhance, test and deploy models under strict deadline
    - New situations require new abstractions and models
  - Prepare wide ranging documents, analyses, and presentations for communicating the results (to analysts, decision makers, broad public, press,…)

# Supporting COVID Response: 2020–2025




## International & Federal agencies

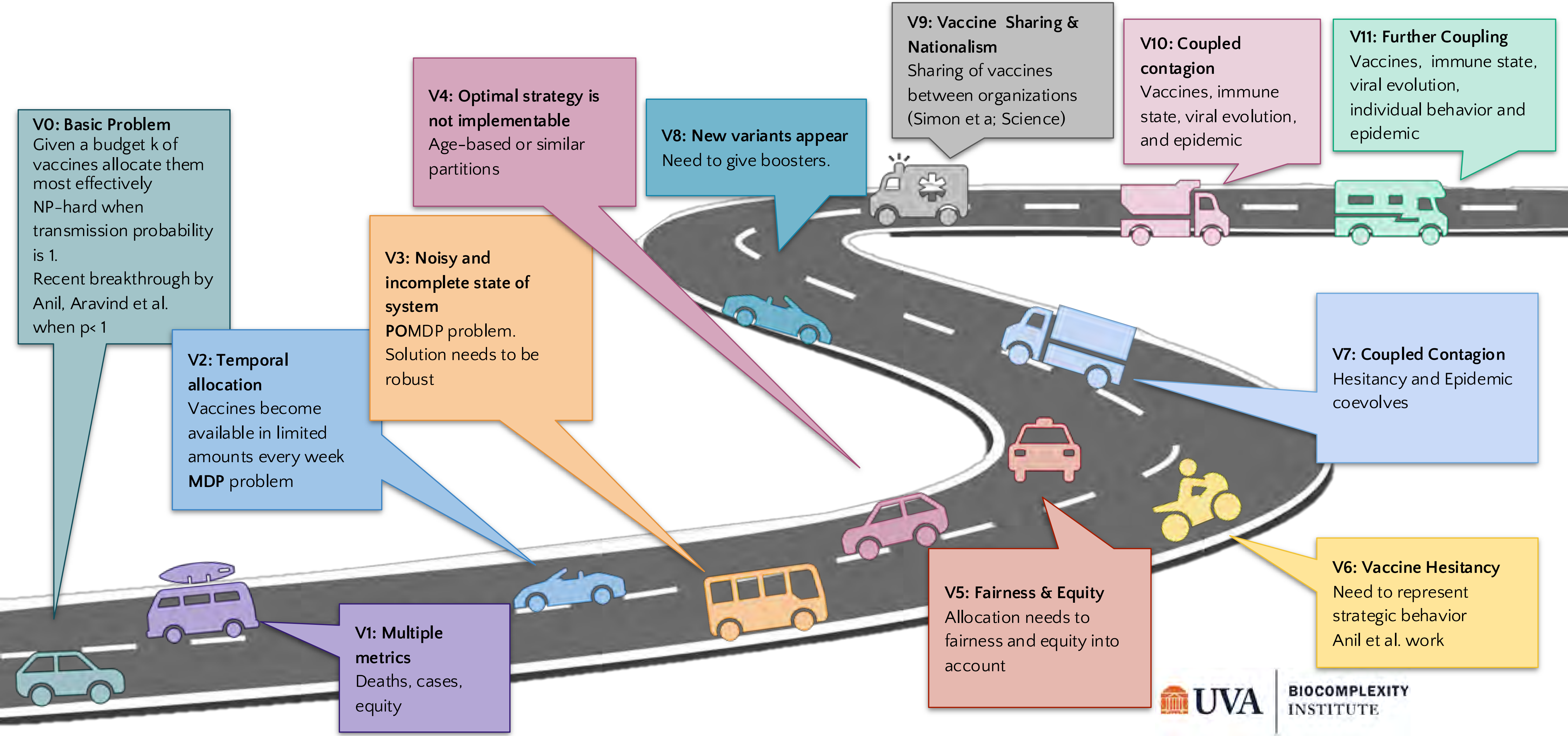
## Commonwealth of Virginia

Primary modeling for planning and response efforts

## Local agencies

Virginia Hospital and Healthcare Association (comprised of 27-member health systems and 110 community and specialty hospitals)

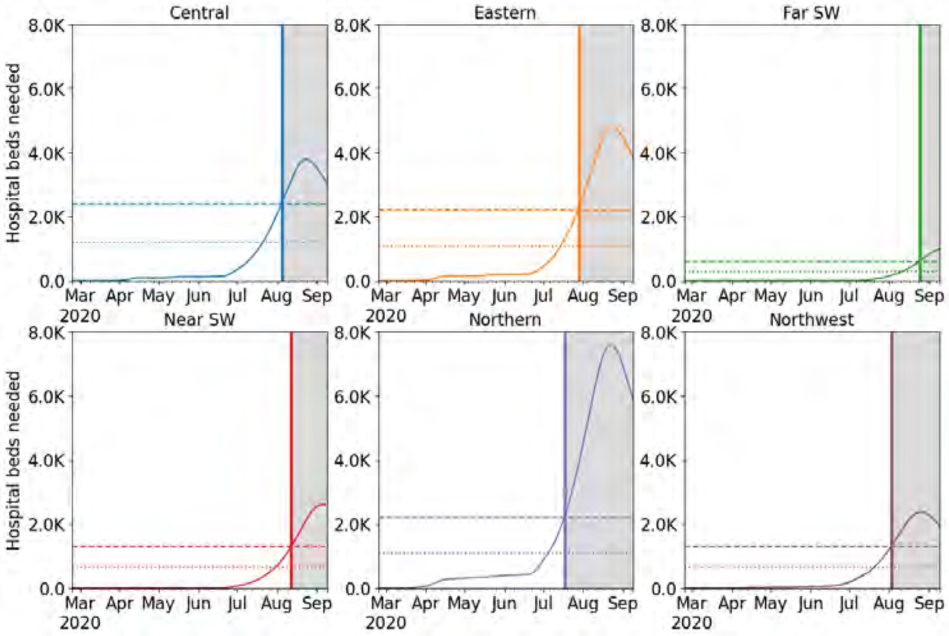
# Motivated new fundamental questions – Vaccine allocation: from simple to complex




UVA | BIOCOMPLEXITY INSTITUTE

# Guidance on field hospitals

As of April 13, 2020, model estimated at least 3 months of lag before hitting hospital capacity under the "Pause" scenario




**April 13, 2020**

HEALTH & MEDICINE

**U.Va.'s coronavirus model projects Virginia will have enough hospital capacity in next few months**

By Peter Coutu and Gary A. Harki

Expand

Gov. Ralph Northam, left, speaks, accompanied by state Secretary of Health and Human Resources Dr. Daniel Carey, right, during a news conference on the state's preparedness for the coronavirus at the Capitol in Richmond, Va., on Wednesday, Mar. 4, 2020. (AP Photo/Steve Helber) (Steve Helber/AP)

**April 17, 2020**

**Virginia is 'pivoting away' from field hospitals**

**Dr. Danny Avula of Richmond/Henrico health departments says hospitals likely can handle coming surge.**

PUBLISHED APRIL 17, 2020 BY KATE ANDREWS

virginia BUSINESS

Virginia is "pivoting away" from needing field hospitals to handle patient overflow because health systems believe they'll be able to manage any surge in COVID-19 cases, said Dr. Danny Avula, the director of Richmond and Henrico County's health departments, Friday evening.

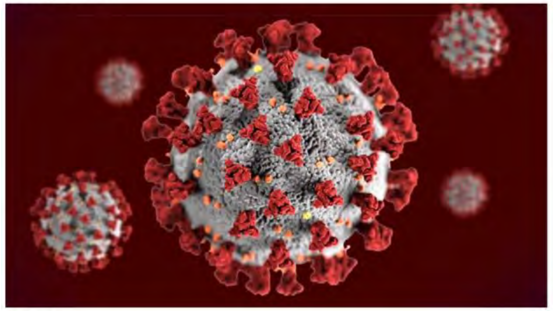
"We have not seen any significant increase in hospitalization [and] ventilator use is under 25%," Avula said during a news briefing. "It feels like the calm before the storm, and hopefully the storm will never come."

Dr. Danny Avula, director of Richmond and Henrico County health departments, at his April 17 news conference

Examining the University of Virginia's COVID-19 forecasting model, Avula said, Virginia seems

**April 21, 2020**

**Virginia hits pause on plans for 3 large field hospitals**



COVID-19 (AP)

*Updated: Apr. 21, 2020 at 5:52 PM EDT*

**May 7, 2020**

U.S. Field Hospitals Stand Down, Most Without Treating Any COVID-19 Patients

May 7, 2020 · 1:15 PM ET
Heard on All Things Considered

JOEL ROSE

Construction at the COVID-19 field hospital at McCormick Place in Chicago on April 10. The city pared back plans for a 3,000-bed temporary hospital at the nation's largest convention center as infection numbers decreased.

UVA | BIOCOMPLEXITY INSTITUTE

# Tools to support decision makers in real-time

Best Paper Award for SIG KDD 2021

## Mobility-based policy analysis tool

Interactive tool to analyze the impact on future cases based on altering mobility to to different types of locations

https://nssac.bii.virginia.edu/covid-19/dashboard/

## Global COVID-19 cases

Provides interactive resource for monitoring and analyzing COVID-19 health statistics from around the world

https://nssac.bii.virginia.edu/covid-19/vmrddash/

## Medical Resource Dashboard

Provides interactive resource for monitoring current and future projected hospitalization burden

https://nssac.bii.virginia.edu/covid-19/vaxstatdashboard/

## Vaccination Statistics Dashboard

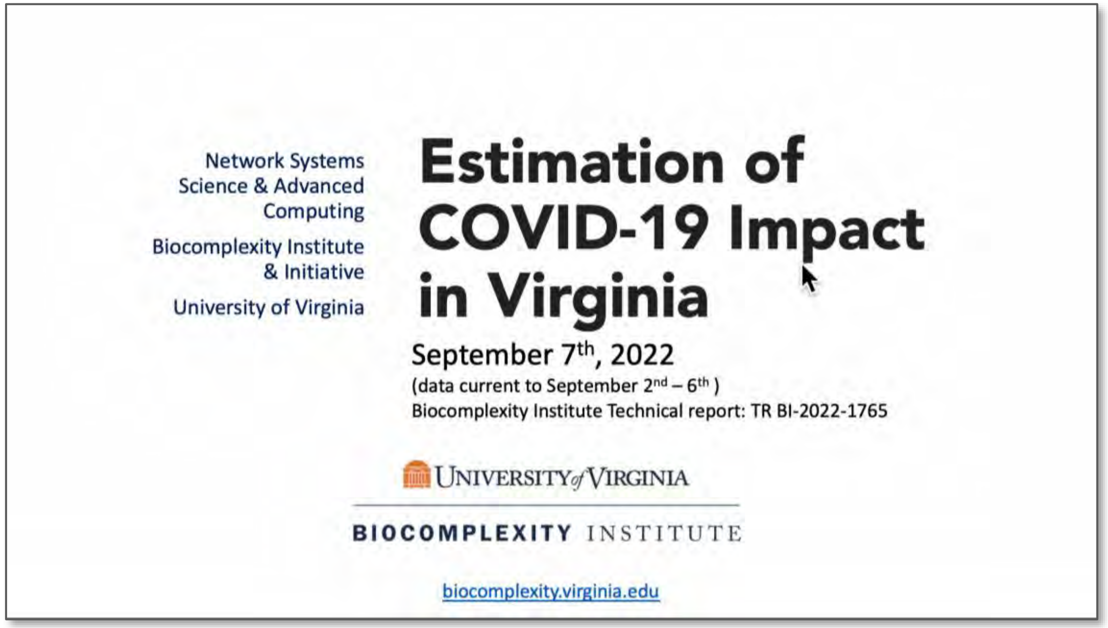
Provides predictive modeling and multiple reports assessing vaccination coverage

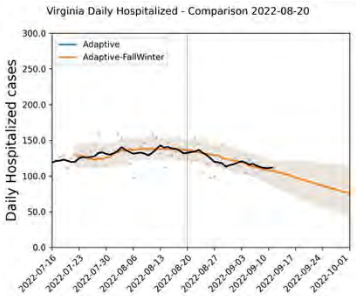
Chang S, Wilson M, Lewis B, Mehrab Z, Dudakiya KK, Pierson E, Koh PW, Gerardin J, Redbird B, Grusky D, Marathe M. Supporting COVID-19 policy response with large-scale mobility-based modeling. *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, 2021 August (2632-2642).

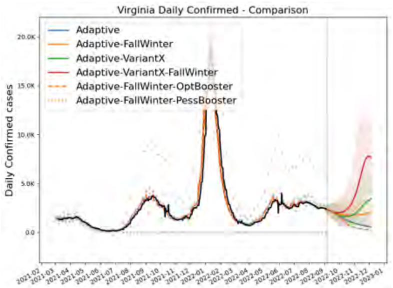
UVA | BIOCOMPLEXITY INSTITUTE

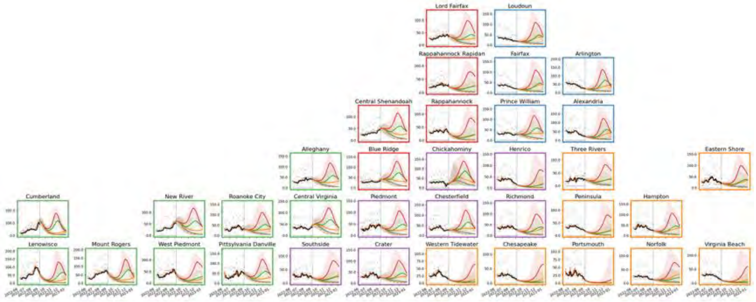
# Projections and analyses throughout pandemic

- **Over 250+ weekly updates of model projections and analyses**
  - Regular briefings to VDH, Sec of Health, other cabinet members, local health districts and hospital systems
  - Ad-hoc meetings with special teams at VDH, universities, etc.
- **Situational Assessment**
  - Cases, hospitalizations, variants, associated metrics
  - Literature review
- **Model Description**
  - Structure and scenarios
- **Model Results**
  - Interpretation and evaluation



https://www.vdh.virginia.gov/coronavirus/category/covid-19/model/

UVA | BIOCOMPLEXITY INSTITUTE

# Contributing to CDC supported scenario modeling hub (SMH) for public health policy and decision making

https://covid19scenariomodelinghub.org/index.html

NPI decisions after primary vaccination (**MMWR May 2021**)

Centers for Disease Control and Prevention

MMWR

Morbidity and Mortality Weekly Report

Early Release / Vol. 70 May 5, 2021

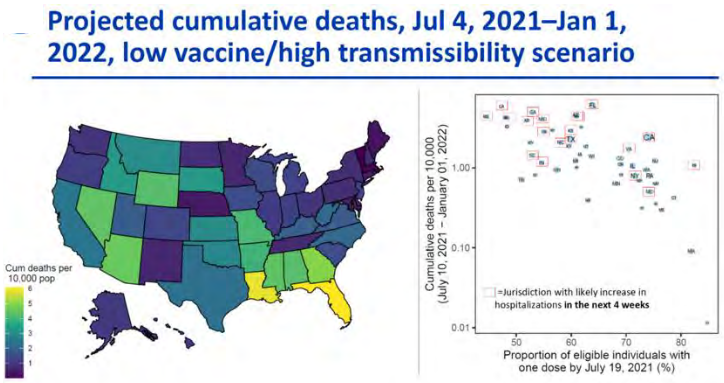
**Modeling of Future COVID-19 Cases, Hospitalizations, and Deaths, by Vaccination Rates and Nonpharmaceutical Intervention Scenarios — United States, April–September 2021**

Rebecca K. Borchering, PhD[1,*]; Cécile Viboud, PhD[2,*]; Emily Howerton[1]; Claire P. Smith[3]; Shaun Truelove, PhD[3]; Michael C. Runge, PhD[4]; Nicholas G. Reich, PhD[5]; Lucie Contamin, MS[6]; John Levander[6]; Jessica Salerno, MPH[6]; Wilbert van Panhuis, PhD[6]; Matt Kinsey, PhD[7]; Kate Tallaksen, MS[7]; R. Freddy Obrecht, PhD[7]; Laura Asher, MPS[7]; Cash Costello, MS[7]; Michael Kelbaugh[7]; Shelby Wilson, PhD[7]; Lauren Shin[7]; Molly E. Gallagher, PhD[7]; Luke C. Mullany, PhD[7]; Kaitlin Rainwater-Lovett, PhD[7]; Joseph C. Lemaitre, MS[8]; Juan Dent, ScM[3]; Kyra H. Grantz[3]; Joshua Kaminsky, MS[3]; Stephen A. Lauer, PhD[3]; Elizabeth C. Lee, PhD[3]; Hannah R. Meredith, PhD[3]; Javier Perez-Saez, PhD[3]; Lindsay T. Keegan, PhD[9]; Dean Karlen, PhD[10]; Matteo Chinazzi, PhD[11]; Jessica T. Davis[11]; Kunpeng Mu[11]; Xinyue Xiong, MSc[11]; Ana Pastore y Piontti, PhD[11]; Alessandro Vespignani, PhD[11]; Ajitesh Srivastava, PhD[12]; Przemyslaw Porebski, PhD[13]; Srinivasan Venkatramanan, PhD[13]; Aniruddha Adiga, PhD[13]; Bryan Lewis, PhD[13]; Brian Klahn, MS[13]; Joseph Outten[13]; James Schlitt, PhD[13]; Patrick Corbett[13]; Pyrros Alexander Telionis, PhD[13]; Lijing Wang, MS[13]; Akhil Sai Peddireddy[13]; Benjamin Hurt, MS[13]; Jiangzhuo Chen, PhD[13]; Anil Vullikanti, PhD[13]; Madhav Marathe, PhD[13]; Jessica M. Healy, PhD[14]; Rachel B. Slayton, PhD[14]; Matthew Biggerstaff, ScD[14]; Michael A. Johansson, PhD[14]; Katriona Shea, PhD[1,†]; Justin Lessler, PhD[3,†]

[1]The Pennsylvania State University, State College, Pennsylvania; [2]Fogarty International Center, National Institutes of Health, Bethesda, Maryland; [3]Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland; [4]U.S. Geological Survey, Laurel, Maryland; [5]University of Massachusetts Amherst, Amherst, Massachusetts; [6]University of Pittsburgh, Pittsburgh, Pennsylvania; [7]Johns Hopkins University Applied Physics Laboratories, Laurel, Maryland; [8]École polytechnique fédérale de Lausanne, Lausanne, Switzerland; [9]University of Utah, Salt Lake City, Utah; [10]University of Victoria, Victoria, British Columbia,

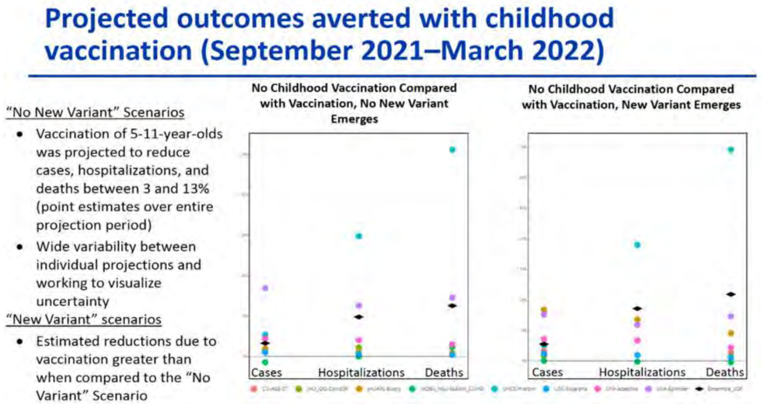
Expansion of vaccine eligibility to 5–11-year-olds (Sep 2021)

**Lancet Regional Health**



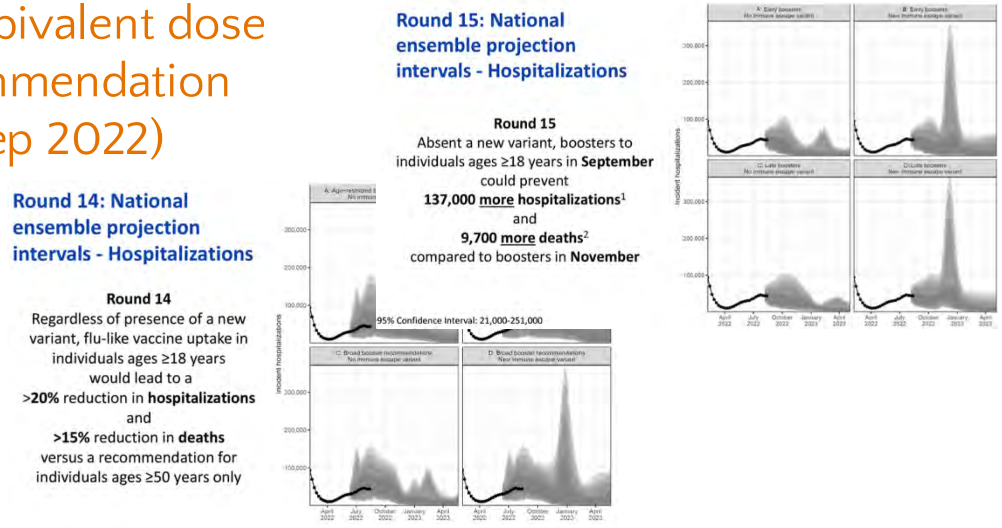

Slide courtesy: Matt Biggerstaff, CDC

Potential impact of Delta wave across states **Elife**, **Jul 2022**)




Slide courtesy: Matt Biggerstaff, CDC

Role in bivalent dose recommendation (Sep 2022)

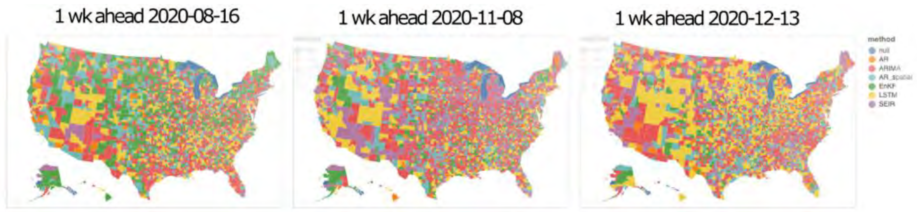


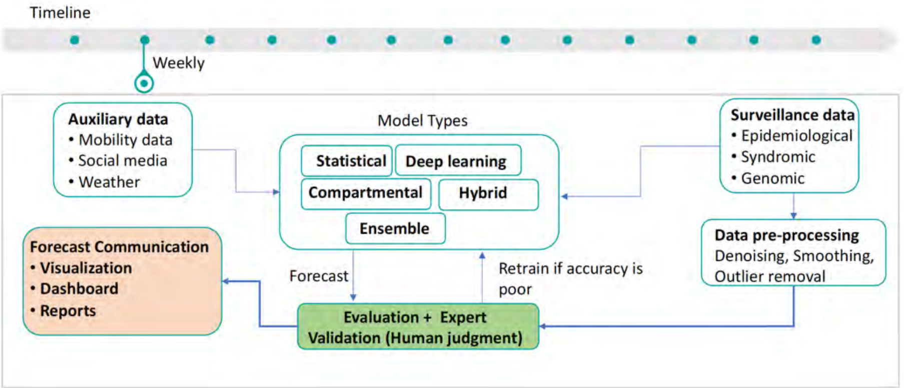

UVA | BIOCOMPLEXITY INSTITUTE

# Supporting CDC Forecasting Hub

**Challenge:** Support multiple agencies with weekly forecasts and analysis – US Forecast Hub, EU Forecast Hub, VDH, and DTRA; work closely with IISc for Indian forecasts

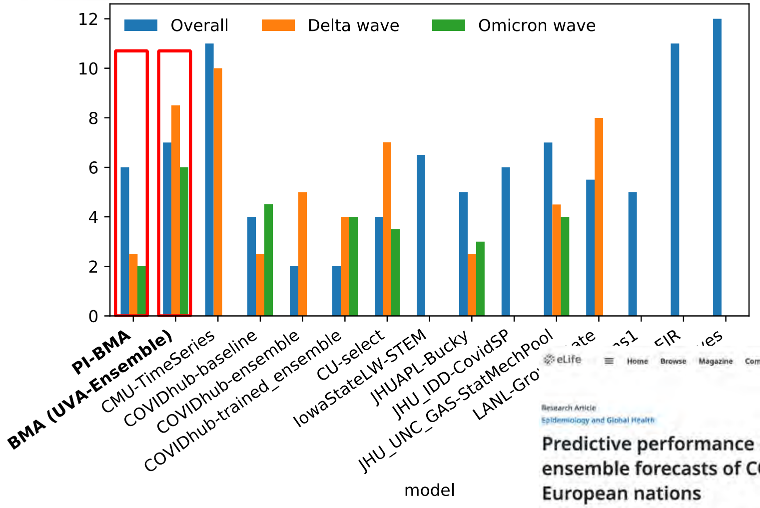
Multiple forecasts targets and resolution – 1—4 weeks ahead cases for 3000 US counties, 50 states, and EU countries and daily hospitalizations 1—28 days ahead for US states




**Computational advances:**

- Multi-model data driven for weekly national forecasting
- Phase driven forecasting
- Bayesian ensembling methods




**Social and Policy Impacts:** Forecasting Hub forecasts are provided on the CDC data portal and are incorporated in federal government decision making. Provided the CDC with weekly updates for this county-level ensemble of statistical, machine learning, and mechanistic models since August 2020.


eLife Home Browse Magazine Community About

Research Article

Epidemiology and Global Health

**Predictive performance of multi-model ensemble forecasts of COVID-19 across European nations**

Katharine Sherratt, Hugo Gruson, Rok Grah, Helen Johnson, Rene Niehus, Bastian Prasse, Frank Sandmann, Jannik Deuschel, Daniel Wolffram ... Sebastian Funk see all

London School of Hygiene & Tropical Medicine, United Kingdom; European Centre for Disease Prevention and Control (ECDC), Sweden; Karlsruhe Institute of Technology, Germany; Robert Koch Institute, Germany; University of Massachusetts Amherst, United States; Boston Children's Hospital and Harvard Medical School, United States; Third Faculty of Medicine, Charles University, Czech Republic; Ecole Polytechnique Federale de Lausanne, Switzerland; Éducation nationale, France see all

*KDD 2021, IAAI 2022, Nature SD, Elife 2022*

# How do we approach these problems?

UVA | BIOCOMPLEXITY INSTITUTE

# Three simple steps that combine network science, AI and high-performance computing

Synthesize a digital similar of a city/country

- Combines data and ML to create digital similar
- Yields a multi-scale, relational social contact network
- Has desirable statistical properties

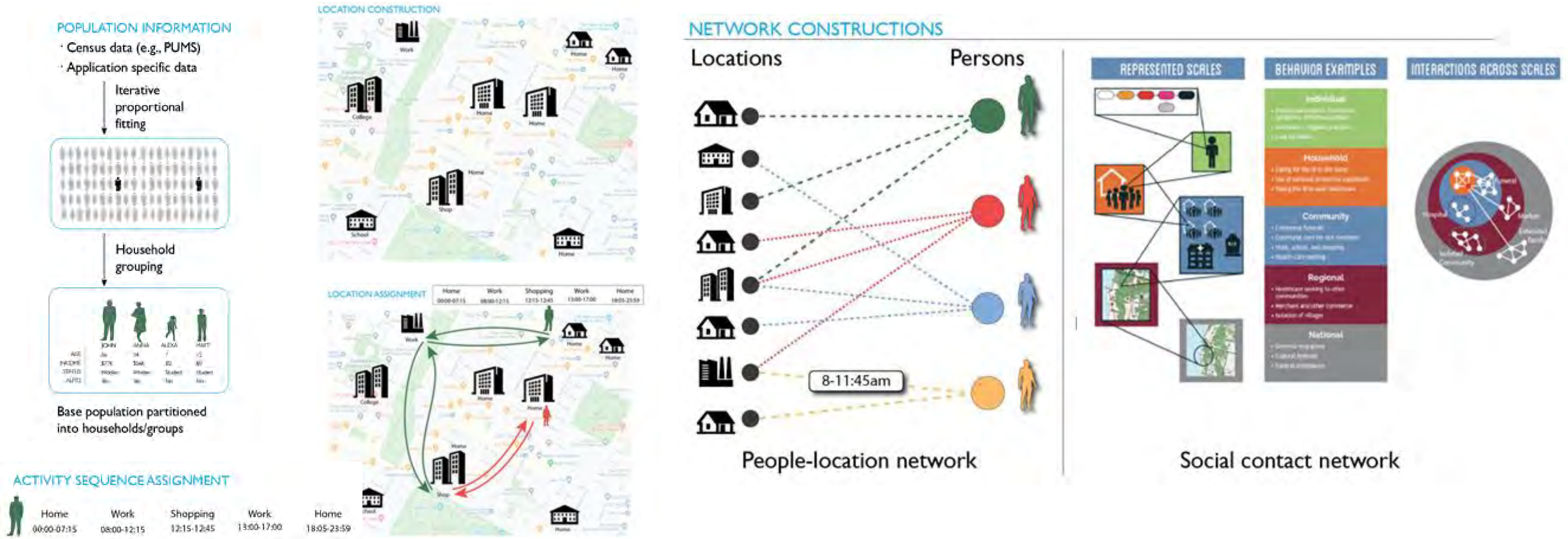
Build AI-enabled agent-based simulations

- Scalable to high-performance computing systems
- Rich representation of policies
- Works with irregular and time-varying networks

Develop AI + Simulation-driven workflows and apps

- Calibration
- Forecasting
- Explanation
- Planning and what-if scenario analysis
- Resource allocation
- Inference

# Digital Similars (Twins) of US social contact network




**Digital twin captures heterogeneous spatially explicit, multi-scale, multi-theory, multi-level social contact network representations**

A Tutorial on Generating Synthetic Populations for Social Modeling, IJCAI 2016, New York City, July 10, 2016 & AAMAS 2016, Singapore, May 9, 2016 and AAMAS 2017, Brazil.
http://people.virginia.edu/~ss7rs/synthetic_population_tutorial/IJCAI_2016_generating_synthetic_populations_for_social_modeling_tutorial.pdf

UVA | BIOCOMPLEXITY INSTITUTE

# Why Digital Similars?

**Provides a spatially and individually resolved data structure**

- Fusing of diverse information in a statistically consistent manner, e.g. (demographic + energy + cell phone) data
- Agents have nominal (age, income), declarative (activities they take) and procedural (e.g. how to drive) data
- Enables information privacy and attribution

**A social coordinate system for incorporating new (intra- and inter) agentic information**

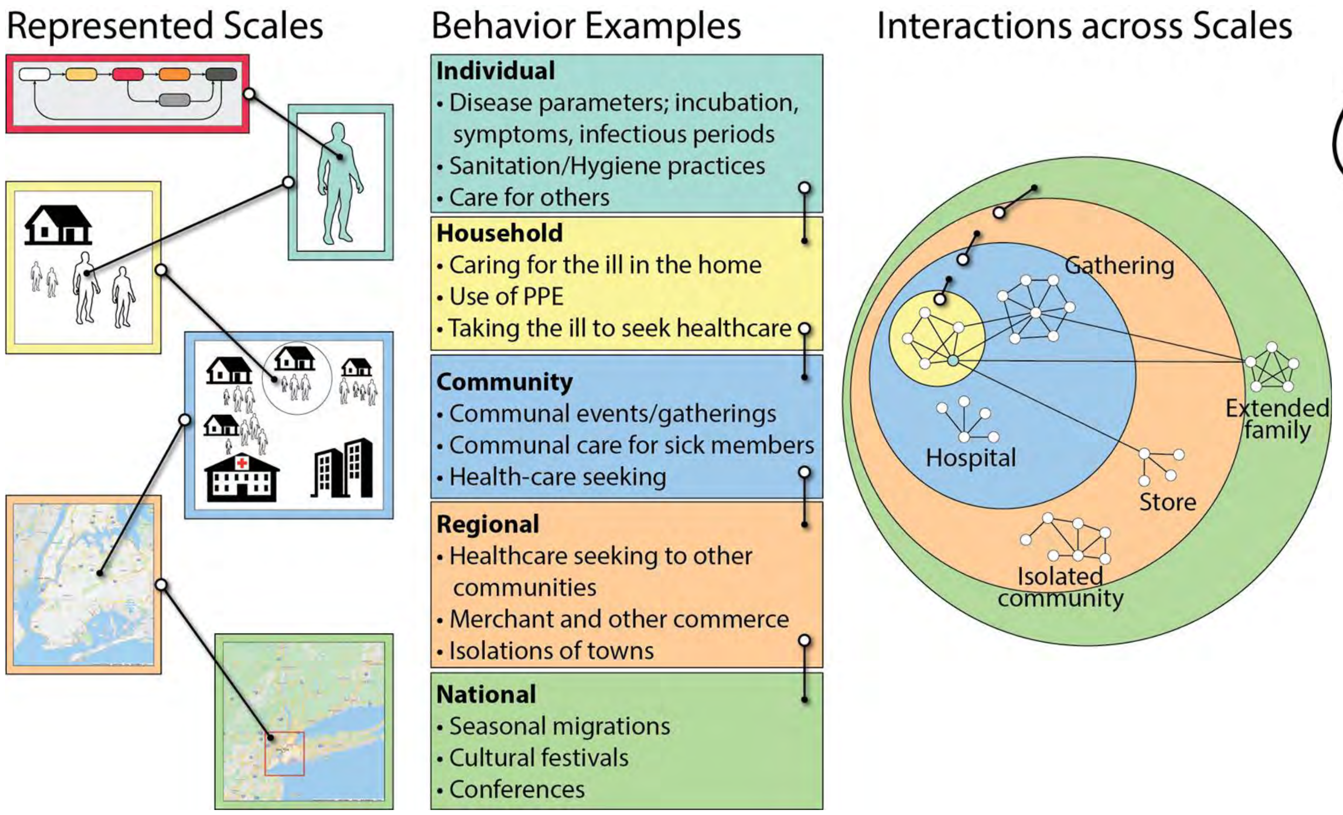
**Agentic representations enable causal interaction-based models & data integration**

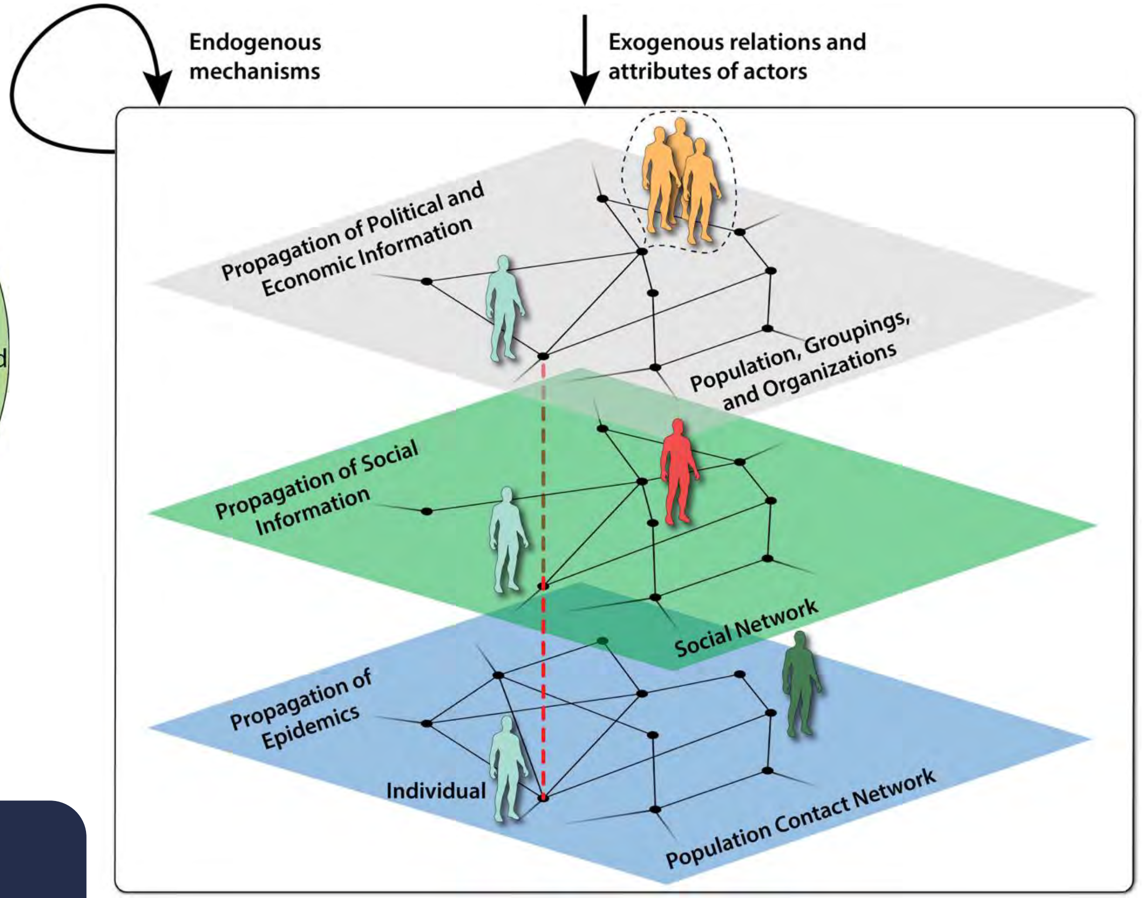
- Putting procedural, social and behavioral theories in motion, e.g. contagion theories of Granovetter, Macy, Schelling, et al.
- Explicitly represent interactions as networks
- Model outputs can then be aggregated to any desired level

**Supports next generation modeling and analytics at-scale**

UVA | BIOCOMPLEXITY INSTITUTE

# What it Captures: Multiple Scales and Theories



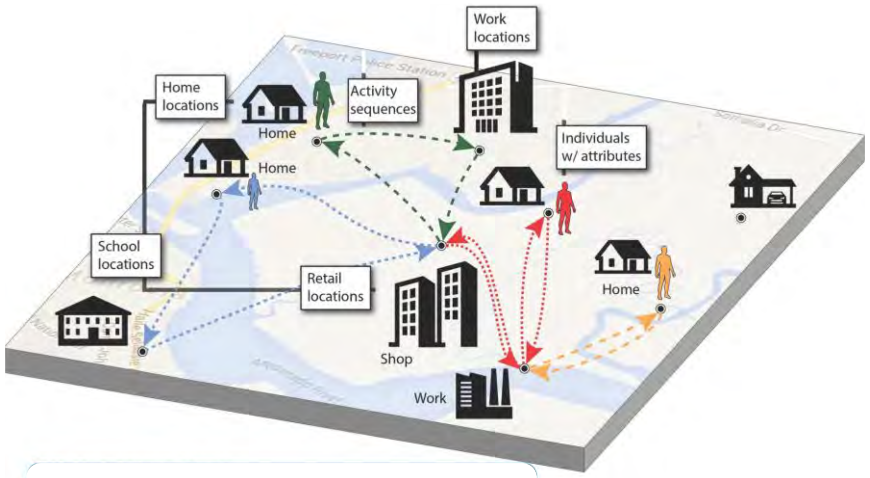




Digital twins capture heterogeneous spatially explicit, multi-scale, multi-theory, multi-level social contact network representations

*A Tutorial on Generating Synthetic Populations for Social Modeling, IJCAI 2016, New York City, July 10, 2016 & AAMAS 2016, Singapore, May 9, 2016 and AAMAS 2017, Brazil. https://sites.google.com/view/samarthswarup/teaching*

UVA | BIOCOMPLEXITY INSTITUTE

# Scale & Size




## US Population Digital Twin

| People & households | Activity sequence | Activity locations | Individual locations | Contact network |
|---|---|---|---|---|

**AIIA Digital Twin Model: (Only one of its kind)**

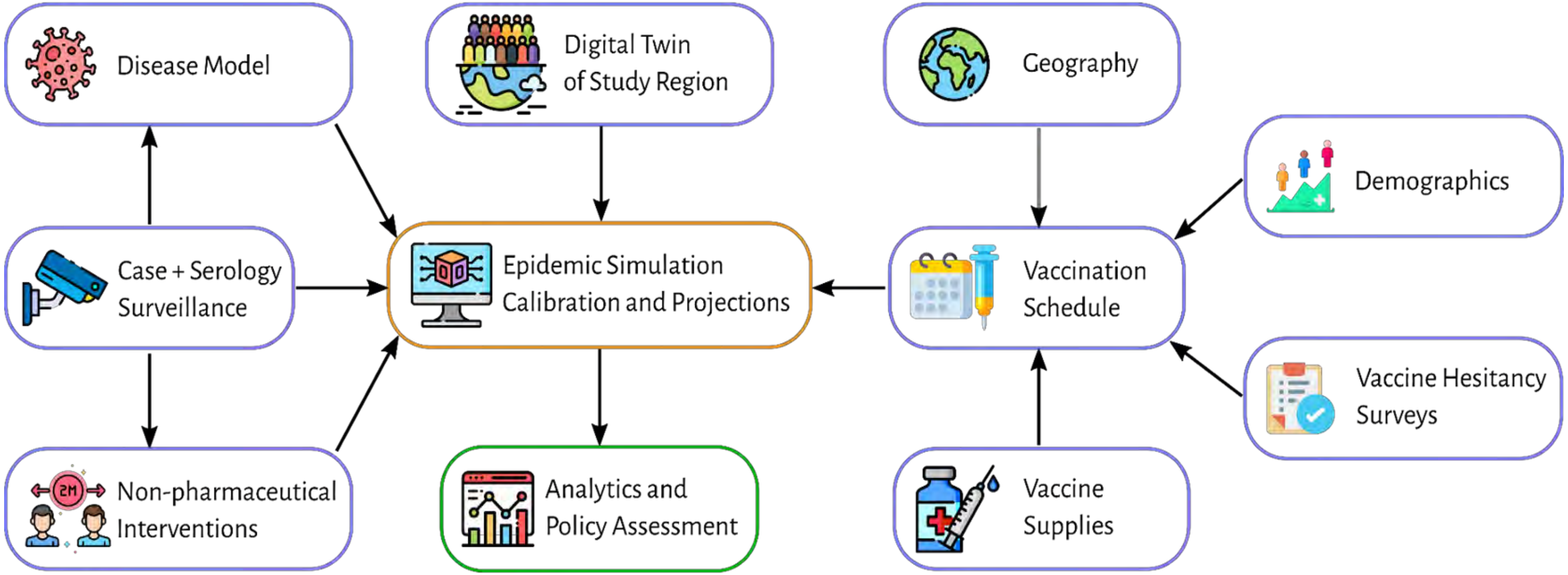
- Covers 48 contiguous US states
- 2.5 Terabytes of data! 12Hours on Rivanna
- Protects privacy – no personally identifiable data

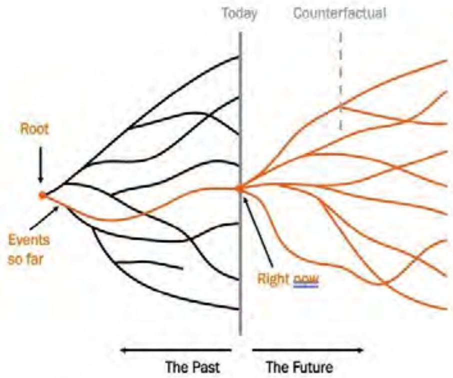
## Data Sets

- US Census demographic distributions at block group resolution
- American Community Survey (ACS) commuting flow data
- ACS Public Use Micro Sample data 2013-2017
- US Census National, State, and County Housing Unit Totals
- National Household Travel Survey 2017 and 2001
- Bureau of Labor Statistics employment data
- National Center for Education Statistics data
- Microsoft Building Footprint data
- Building Footprint USA (now Lightbox)
- HERE location and Points of Interest data
- Black Knight geographical parcel data

# Live Digital Twins for real-time response

Incorporates multiple new datasets updated as the pandemic evolves

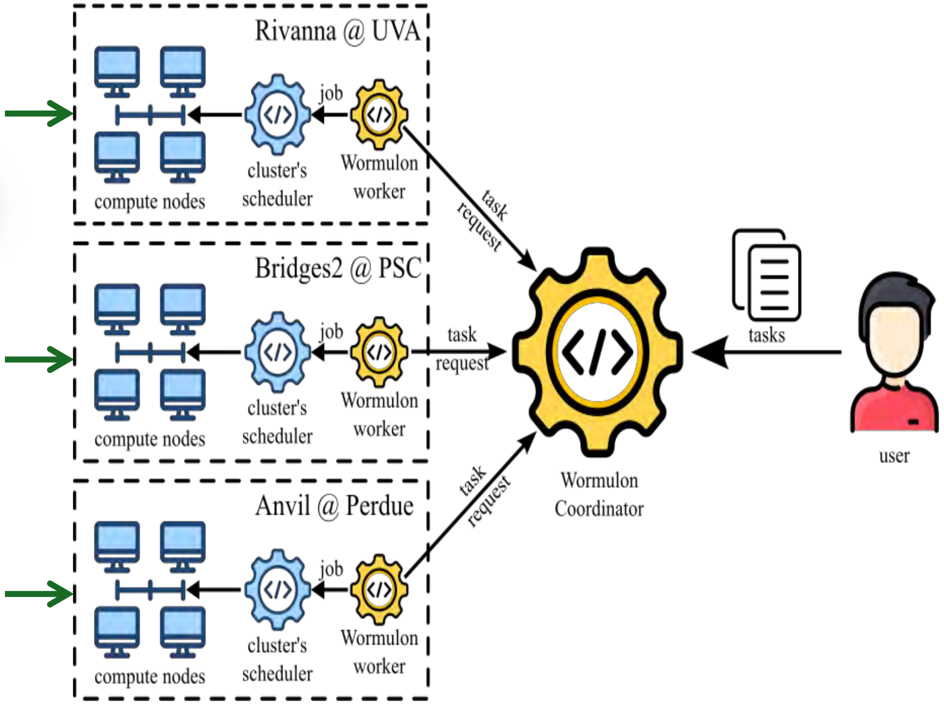



Bringing the Digital Twin to life:
contextualize the DT with current ground conditions

# Parallel socio-epidemic simulators and analytics engine




**2550 simulation instances**
*calibration*

**~10200 simulation instances**
*projections*

**3100 US counties 500 days**
*confirmed case count data*

**~20GB synthetic population**
*person, household, location, and activities data*

**~1.5B entries, ~4GB aggregate output data**
*250 days, 200 health states, 3 counts*

**~4TB individual-level output data**
*multi--million state transitions = multi-billion entries*

The Hybrid HPC pipeline for national pandemic planning and response can run: 400 replicates of national (US) individualized multi-agent socio-epidemic simulations (300M nodes & 12B edge networks yielding over 1Trillion interactions over 200 simulated days) in 33 hours on 9K cores spread over two (and more recently three) machines

ACM Gordon Bell Finalist – Special Prize for HPC-Based COVID-19 Research ( in IJHPC 2022) IPDPS, eLife, IEEE Big Data, PNAS Nexus

UVA | BIOCOMPLEXITY INSTITUTE

# Examples of scientific questions we address

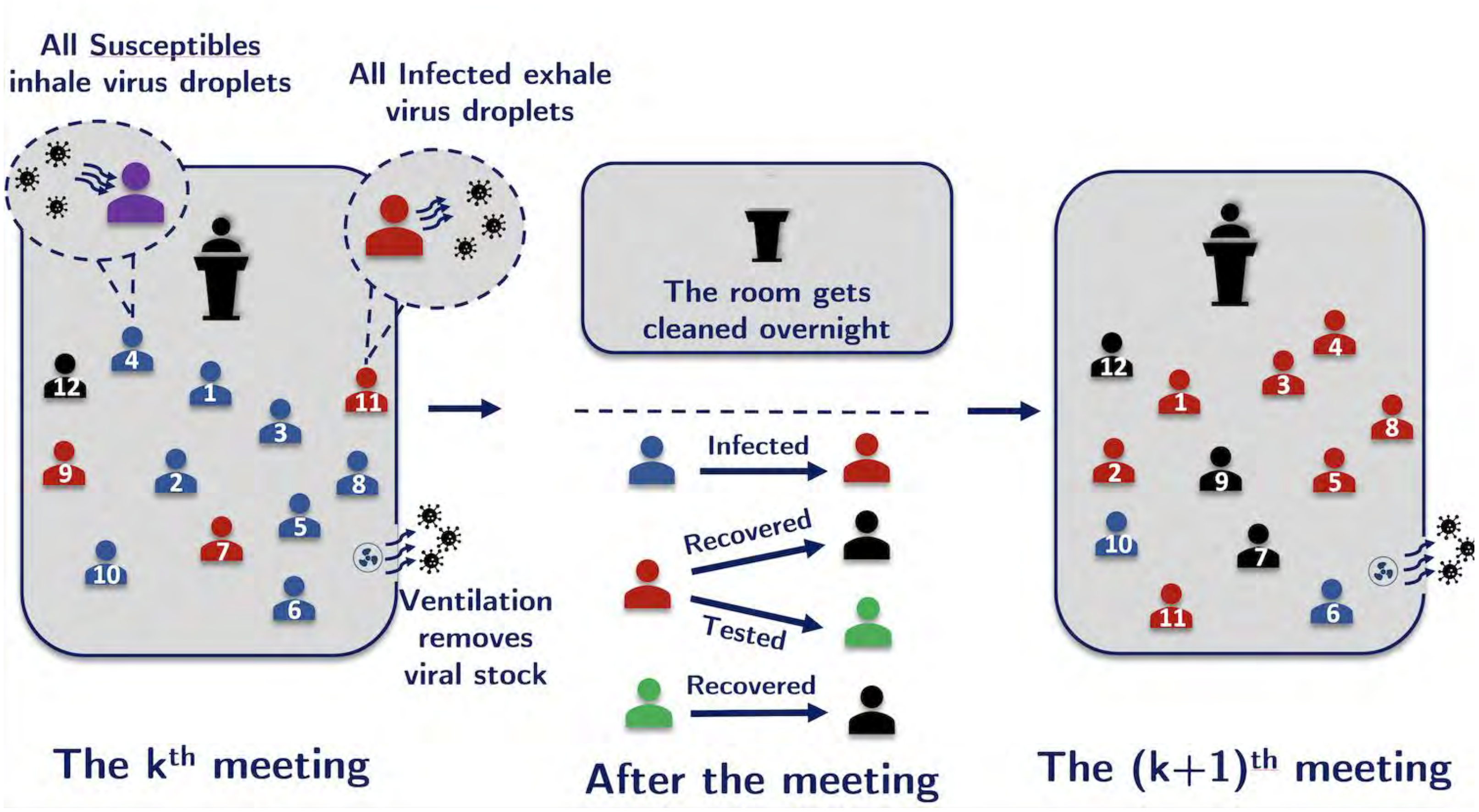
UVA | BIOCOMPLEXITY INSTITUTE

# Example 1: Tradeoffs across temporal scales

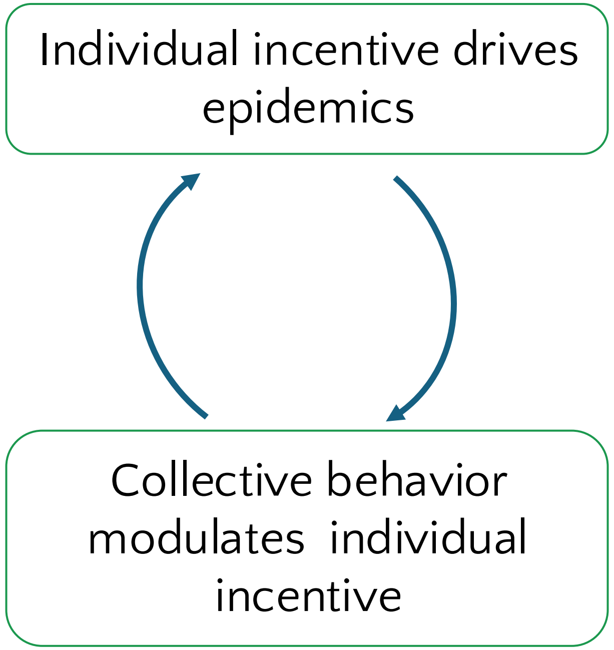
PNAS RESEARCH ARTICLE | BIOPHYSICS AND COMPUTATIONAL BIOLOGY
OPEN ACCESS

**Airborne disease transmission during indoor gatherings over multiple time scales: Modeling framework and policy implications**

Avinash K. Dixit[a,1], Baltazar Espinoza[b], Zirou Qiu[b,c], Anil Vullikanti[b,c], and Madhav V. Marathe[b,c,1]




Robert H. MacArthur award lecture. *Ecology, 1992.*
**The Problem of Pattern and Scale in Ecology**
*Simon A. Levin*




EPIDEMIOLOGY

**Social Factors in Epidemiology**

Chris T. Bauch[1] and Alison P. Galvani[1,2]

The coupling of social and biological contagion in human populations can have positive or negative outcomes.

UVA | BIOCOMPLEXITY INSTITUTE

# Scenarios, control variables and policy insights

| Scenario | Description | Control variables | Policy insights | Dynamic components |
|---|---|---|---|---|
| Short time-scale | Single meeting at a single day. | *a, b, c, d* | For short meetings, ventilation and masking have similar effects.<br>For long meetings, ventilation has a greater impact than masking. | Viral load |
| Medium short time-scale | Multiple meetings at a single day. | *a, b, c, d, e* | For short meetings, break times of 10–15 min have a similar impact than 50% mask compliance.<br>For long meetings, break times of 20–25 min have a similar impact than 50% mask compliance. | Viral load |
| Medium-long time-scale | Single meeting at multiple days. | *a, b, c, d, f* | For short meetings, masking and testing shows that a trade-off follows a 2:1<br>For long meetings, masking and testing shows a trade-off of 1:1. | Viral load<br>Infectious individuals |
| Long time scale | multiple meetings at multiple days. | *a, b, c, d, e, f* | For short meetings, a 60% testing reduction is balanced by around 20-min breaks.<br>For long meetings, a 40% testing reduction is balanced by around 20-min breaks. | Viral load<br>Infectious individuals |

a. Group size b. Air mass (room size) c. Mask compliance d. Ventilation e. Break times f. Testing

UVA | BIOCOMPLEXITY INSTITUTE

# BRAVE: Revolutionizing Indoor Air Safety

- BRAVE: $40M ARPA-H grant to develop a smart building system to detect and respond to airborne health threats
- BRAVE: Led by Linsey Marr at Virginia Tech with collaborators from BI@UVA, WashU, UC Davis, Michigan, Yale, Penn State, Emory, Signature Sc. and industry partners
- <u>Goal:</u>
  - Will develop enhanced biosensors + modeling to detect pathogens and trigger HVAC interventions (ventilation, filtration, UV)
  - Reduce respiratory illnesses by 25% and improve public health in communal spaces
- Biocomplexity Institute (UVA) leads modeling and simulation efforts and develop multi-scale AI-enabled decision support systems

**BREATHE research teams kick off efforts to enhance indoor air quality to improve health, advance the future of smart buildings**

*Teams will seek to develop systems that reduce harmful pathogens and allergens in buildings across the country, ensuring indoor air quality is always safe and healthy*

The Advanced Research Projects Agency for Health (ARPA-H), an agency within the U.S. Department of Health and Human Services (HHS), today announced the research and development teams receiving awards from the Building Resilient Environments for Air and Total Health (BREATHE) program. BREATHE intends to advance the next generation of smart and healthy buildings by developing integrated systems that will provide continual measurement and risk assessment of indoor air quality and deploy real-time interventions, like extra ventilation or disinfection, to reduce airborne threats to human health.

Poor indoor air quality exacerbates chronic diseases like asthma and allergies and is a leading cause of preventable respiratory illnesses like the flu. Airborne pathogens and allergens can have a significant effect on people's health, especially children, the elderly, and those living with chronic illnesses. Even though Americans spend 90% of their lives indoors, we do more to monitor and reduce outdoor air threats than those inside our homes, schools, and workplaces. Rather than only treating illness after it occurs, BREATHE aims to create healthier indoor environments that prevent disease, improve productivity, and enhance quality of life for everyone.

"BREATHE will revolutionize public health by greatly advancing our ability to detect and address indoor air quality threats like never before," said Jason Roos, Ph.D., **ARPA-H Acting Director**. "ARPA-H's investment has the potential to bring about the next-generation of smart buildings, making sure indoor air is always clean and healthy."

The agency's initial total commitment to these teams is up to $156 million over 5 years, to establish this groundbreaking approach. Other Transactions Agreements (not procurement contracts, grants, or cooperative agreements) vary in funding amount per awardee and are contingent upon each team meeting aggressive and accelerated milestones.

# Example 2: Dueling dynamics over multiple networks

PNAS | RESEARCH ARTICLE | ENVIRONMENTAL SCIENCES

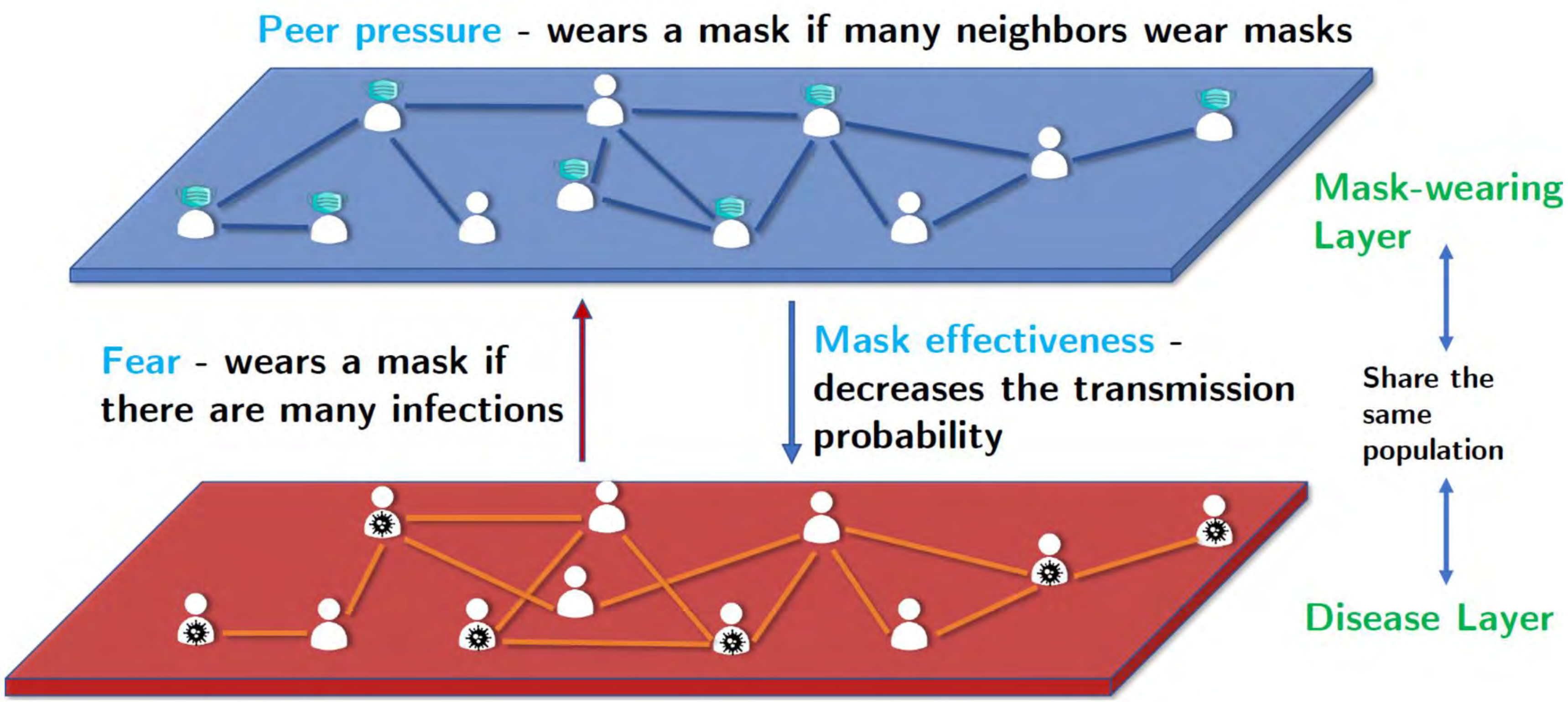
**Understanding the coevolution of mask wearing and epidemics: A network perspective**

Zirou Qiu[a,b,1], Baltazar Espinoza[b], Vitor V. Vasconcelos[c,d,e,f], Chen Chen[b], Sara M. Constantino[f,g,h,i], Stefani A. Crabtree[j,k], Luojun Yang[l], Anil Vullikanti[a,b], Jiangzhuo Chen[b], Jörgen Weibull[m], Kaushik Basu[n,o], Avinash Dixit[p], Simon A. Levin[l,1], and Madhav V. Marathe[a,b,1]

**Behavior is modeled by a set of threshold conditions**

Mechanisms: Peer pressure, fear, pro-sociality
Response: Use or not face mask

Social Interaction layer

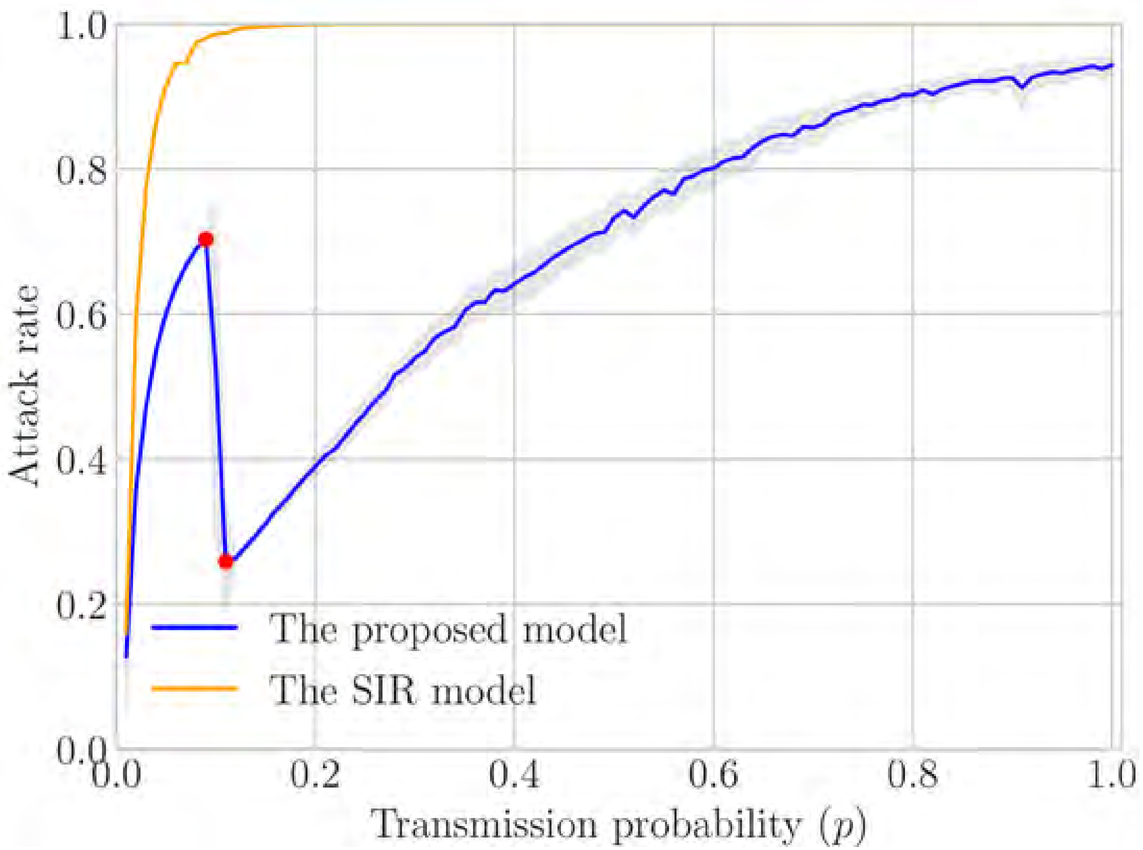



Proximity Contacts layer

Collective human behavior drives epidemics

Epidemics drive collective human behavior

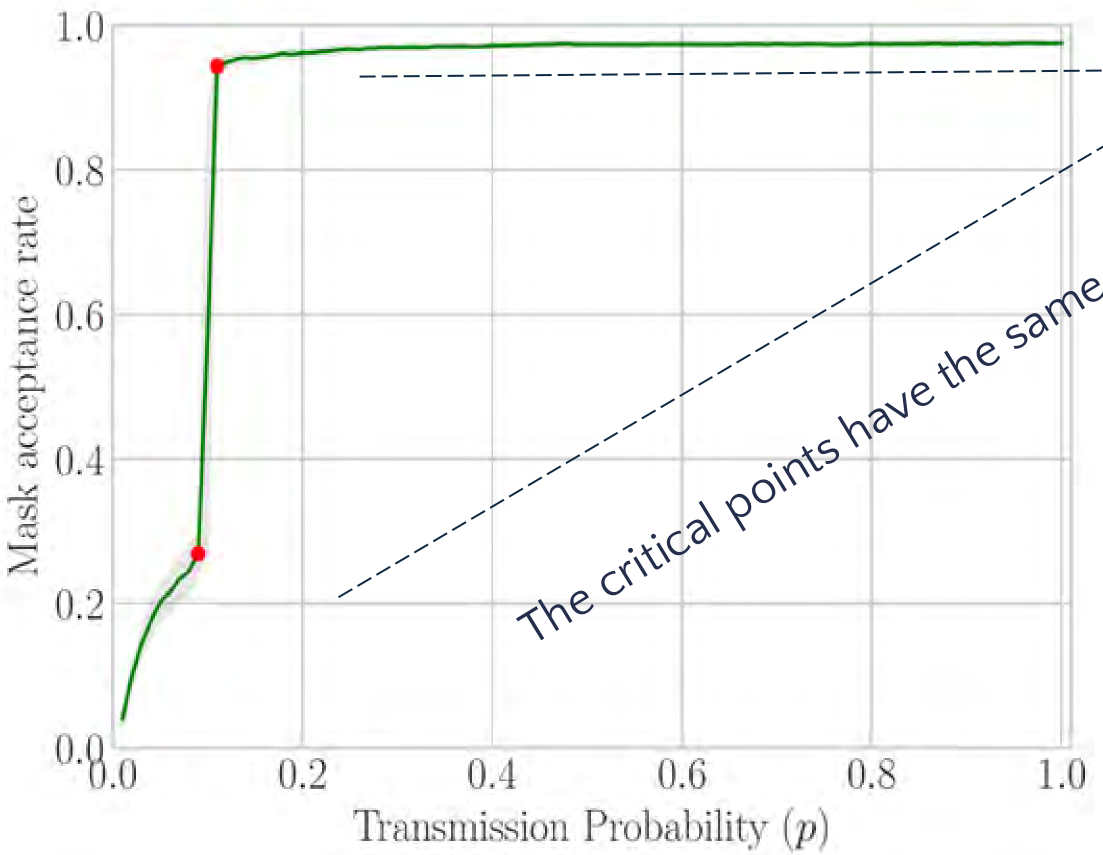
# A phase transition of dueling dynamics

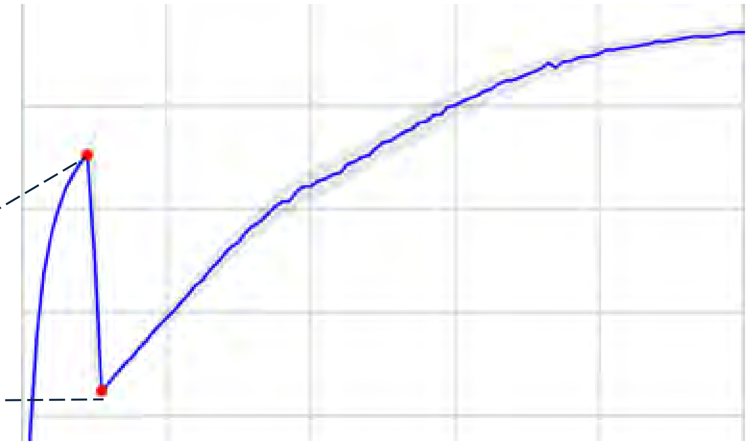



- A **monotonic increase** of the attack rate under the SIR model as the disease becomes more infectious
- The disease dynamic for the behavior model exhibits a **phase transition**, characterized by two **critical points**
- Studied sensitivity of the model w.r.t:
  - Initial condition
  - Randomness of the network
- Observed a low variance

UVA | BIOCOMPLEXITY INSTITUTE

# A phase transition of the social dynamic



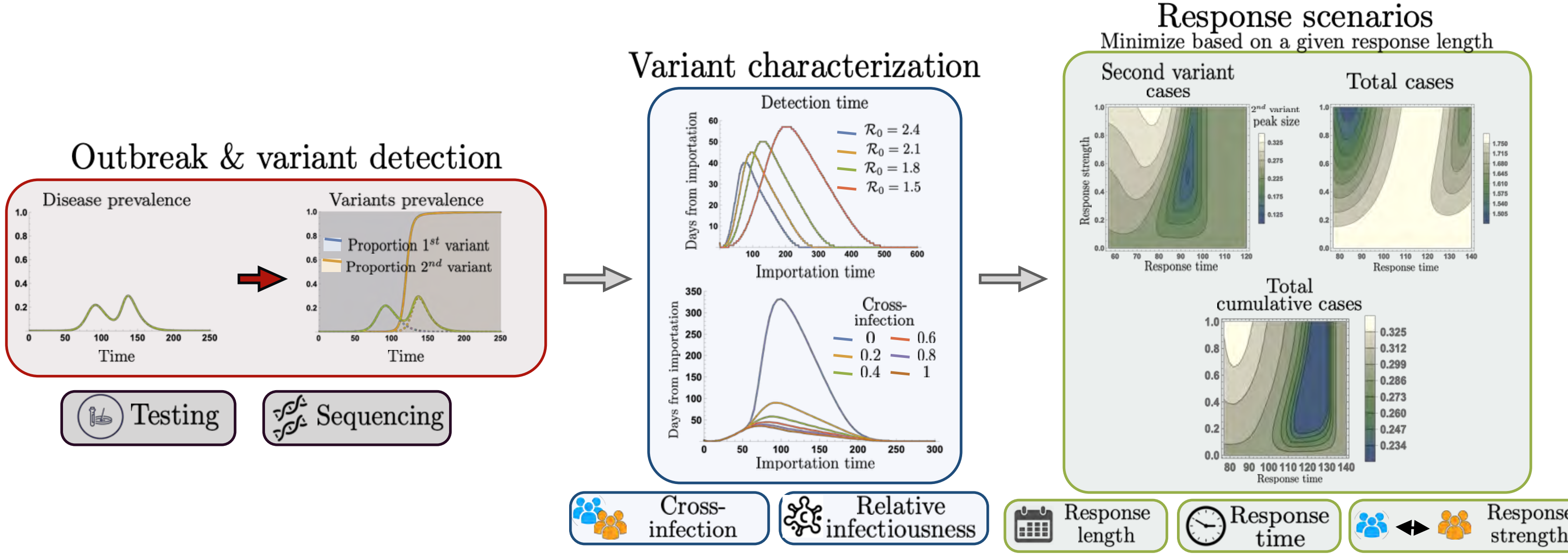
- The mask acceptance rate increases gradually as *p* increases, up to the first critical point
- Then mask acceptance rate rises drastically as *p* continues to increase to the second tipping point
- The social dynamic enters a saturation stage after the second critical point

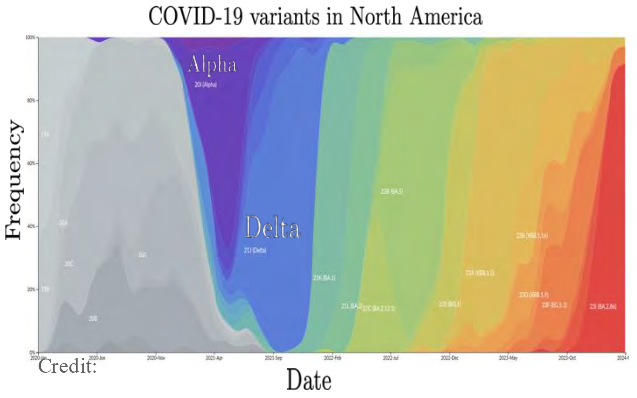
The two critical points can be captured with a finer granularity where the system exhibits a **sharp phase transition**.

UVA | BIOCOMPLEXITY INSTITUTE

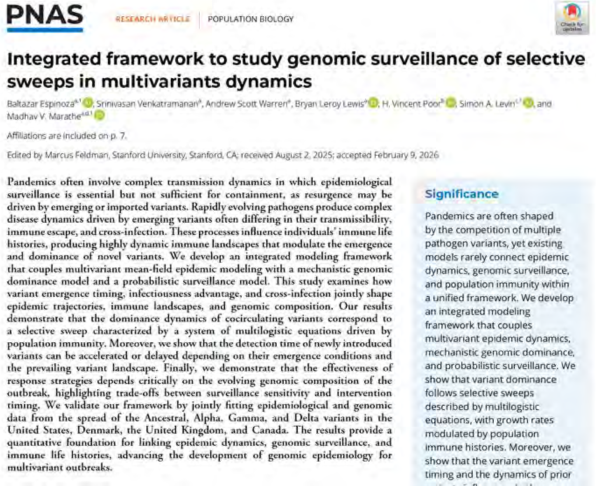
# Example 3: Genomic epidemiology and biosurveillance

- Variant dominance time in the US*
  - Delta – within 7-13 weeks after detection
  - Omicron – within 4-6 weeks after detection
- How do different variant's importation conditions shape epi-genomic dynamics?
- How fast can we detect an outbreak? A novel variant?
- How do intervention effects depend on the variant's competing dynamics?




nextstrain.org



PNAS RESEARCH ARTICLE | POPULATION BIOLOGY

**Integrated framework to study genomic surveillance of selective sweeps in multivariants dynamics**

Baltazar Espinoza, Srinivasan Venkatramanan, Andrew Scott Warren, Bryan Leroy Lewis, H. Vincent Poor, Simon A. Levin, and Madhav V. Marathe

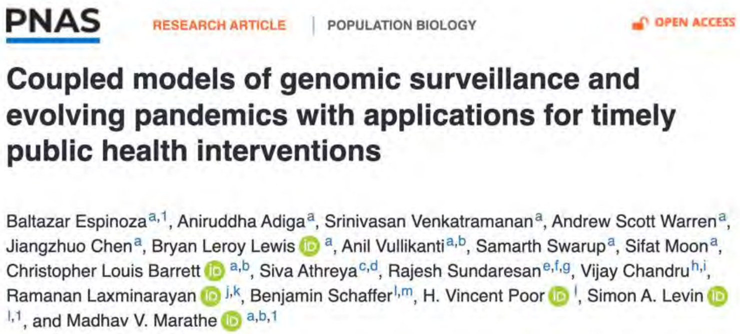
Affiliations are included on p. 7.

Edited by Marcus Feldman, Stanford University, Stanford, CA; received August 2, 2025; accepted February 9, 2026

Pandemics often involve complex transmission dynamics in which epidemiological surveillance is essential but not sufficient for containment, as resurgence may be driven by emerging or imported variants. Rapidly evolving pathogens produce complex disease dynamics driven by emerging variants often differing in their transmissibility, immune escape, and cross-infection. These processes influence individuals' immune life histories, producing highly dynamic immune landscapes that modulate the emergence and dominance of novel variants. We develop an integrated modeling framework that couples multivariant mean-field epidemic modeling with a mechanistic genomic dominance model and a probabilistic surveillance model. This study examines how variant emergence timing, infectiousness advantage, and cross-infection jointly shape epidemic trajectories, immune landscapes, and genomic composition. Our results demonstrate that the dominance dynamics of cocirculating variants correspond to a selective sweep characterized by a system of multilogistic equations driven by population immunity. Moreover, we show that the detection time of newly introduced variants can be accelerated or delayed depending on their emergence conditions and the prevailing variant landscape. Finally, we demonstrate that the effectiveness of response strategies depends critically on the evolving genomic composition of the outbreak, highlighting trade-offs between surveillance sensitivity and intervention timing. We validate our framework by jointly fitting epidemiological and genomic data from the spread of the Ancestral, Alpha, Gamma, and Delta variants in the United States, Denmark, the United Kingdom, and Canada. The results provide a quantitative foundation for linking epidemic dynamics, genomic surveillance, and immune life histories, advancing the development of genomic epidemiology for multivariant outbreaks.

**Significance**

Pandemics are often shaped by the competition of multiple pathogen variants, yet existing models rarely connect epidemic dynamics, genomic surveillance, and population immunity within a unified framework. We develop an integrated modeling framework that couples multivariant epidemic dynamics, mechanistic genomic dominance, and probabilistic surveillance. We show that variant dominance follows selective sweeps described by multilogistic equations, with growth rates modulated by population immune histories. Moreover, we show that the variant emergence timing and the dynamics of prior



PNAS RESEARCH ARTICLE | POPULATION BIOLOGY OPEN ACCESS

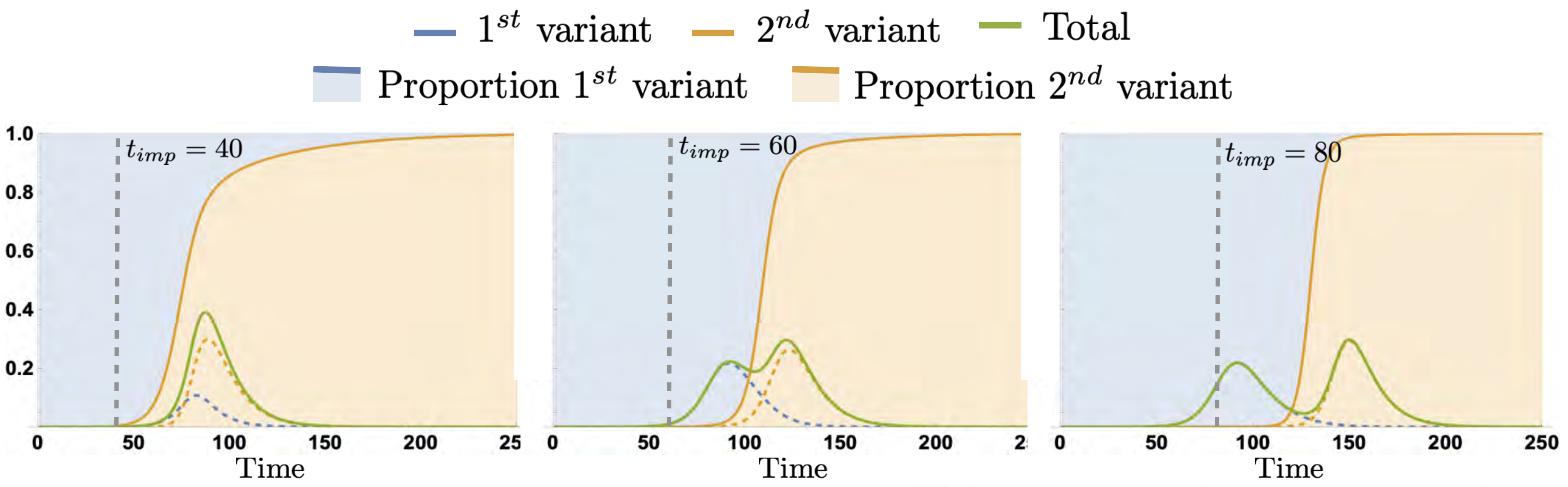
**Coupled models of genomic surveillance and evolving pandemics with applications for timely public health interventions**

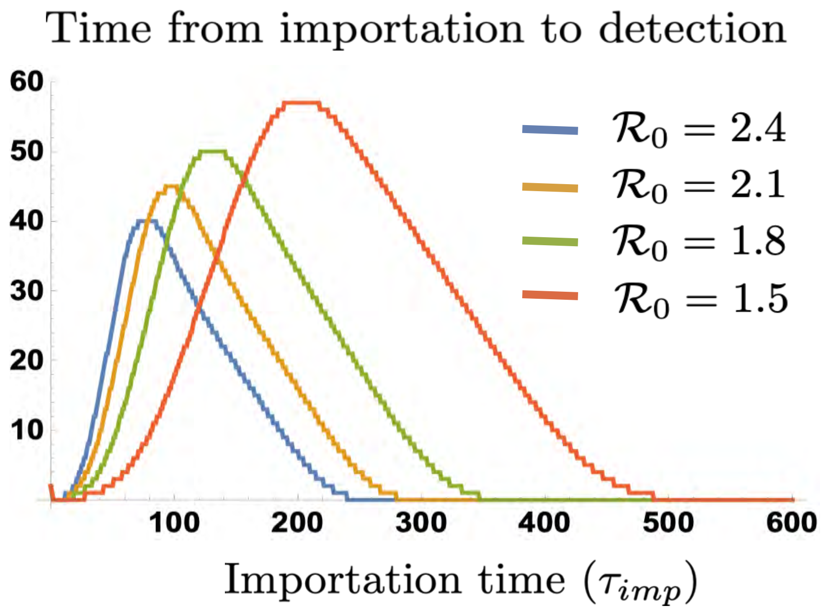
Baltazar Espinoza[a,1], Aniruddha Adiga[a], Srinivasan Venkatramanan[a], Andrew Scott Warren[a], Jiangzhuo Chen[a], Bryan Leroy Lewis[a], Anil Vullikanti[a,b], Samarth Swarup[a], Sifat Moon[a], Christopher Louis Barrett[a,b], Siva Athreya[c,d], Rajesh Sundaresan[e,f,g], Vijay Chandru[h,i], Ramanan Laxminarayan[j,k], Benjamin Schaffer[l,m], H. Vincent Poor[l], Simon A. Levin[l,1], and Madhav V. Marathe[a,b,1]

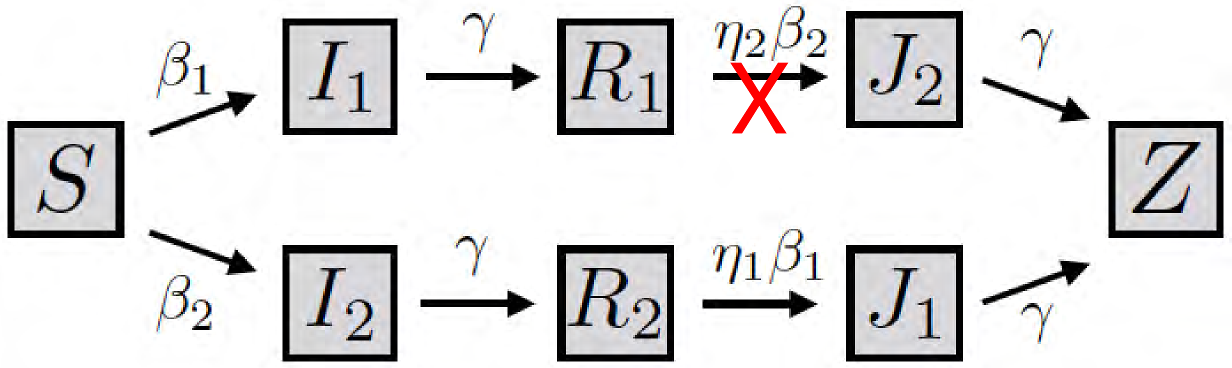
*Chan Y, Irvine MA, Prystajecky N, Sbihi H, Taylor M, Joffres Y, et al. Emergence of SARS-CoV-2 Delta Variant and Effect of Nonpharmaceutical Interventions, British Columbia, Canada. Emerg Infect Dis. 2023;29(10):1999–2007. https://doi.org/10.3201/eid2910.230055

UVA BIOCOMPLEXITY INSTITUTE

# Variant detection and intervention implementation



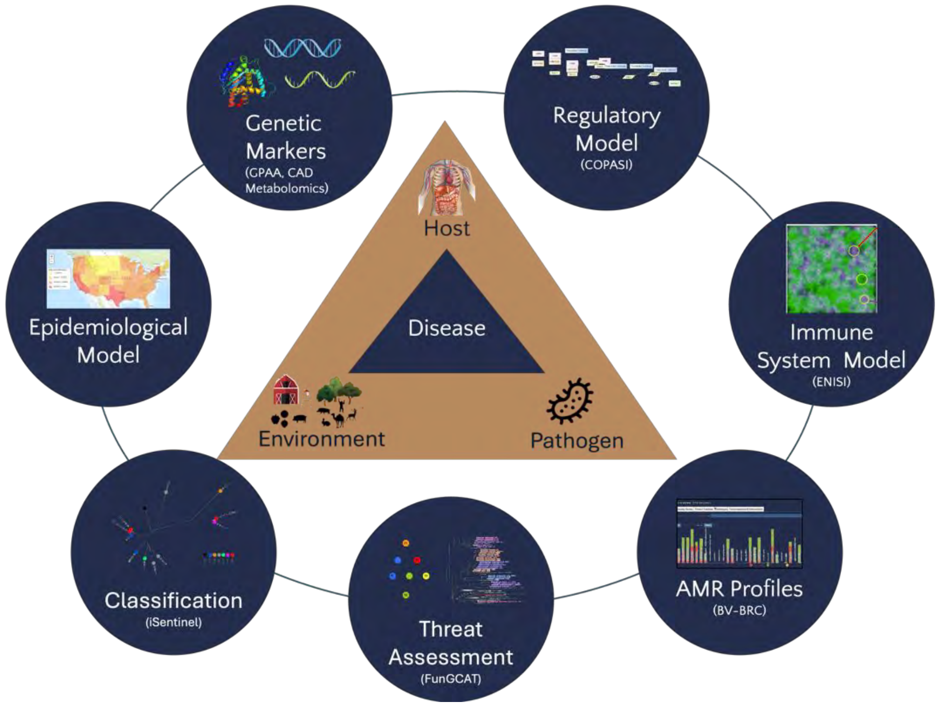

Competing variant dynamics impact the detection time of novel variants

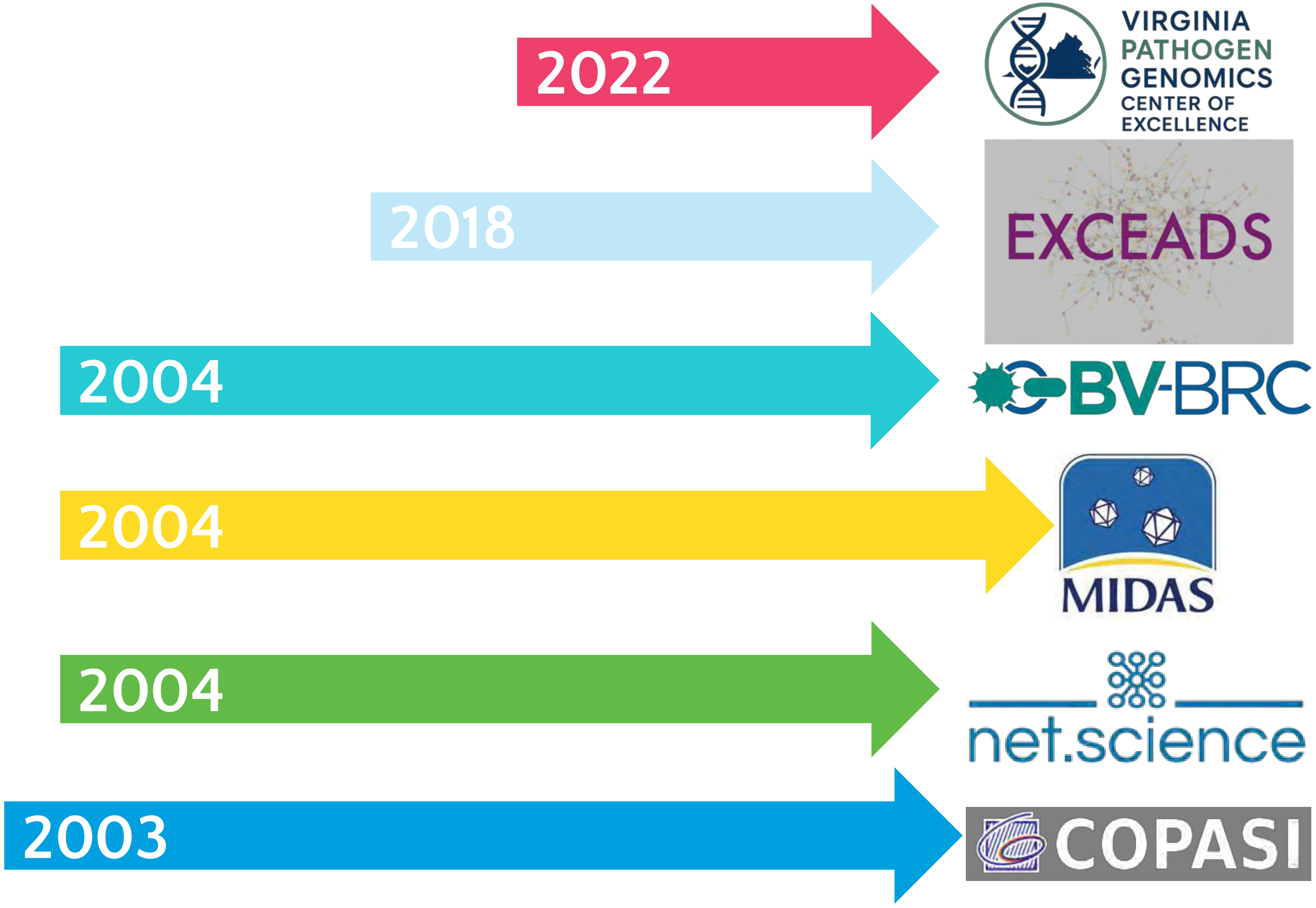
UVA | BIOCOMPLEXITY INSTITUTE

# Developing computational and data infrastructures for health sciences for over 25 years

UVA | BIOCOMPLEXITY INSTITUTE

# Long standing expertise in building Cyber-infrastructures and Cores



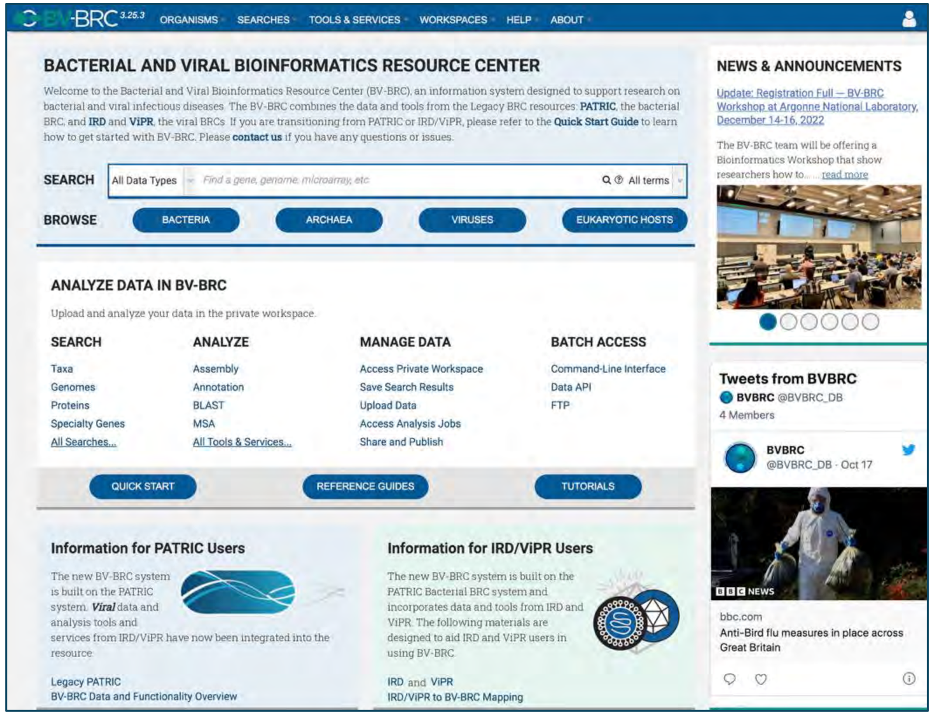

https://va-pgcoe.org/

More about this in preprint: https://arxiv.org/abs/2411.01386
Data visualization and downloads at https://ditto.bii.virginia.edu

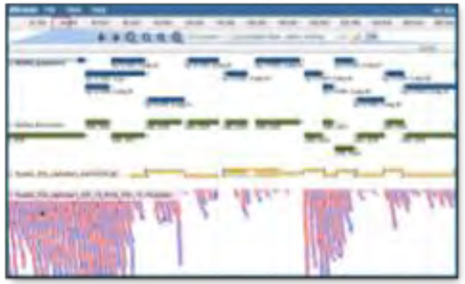
UVA | BIOCOMPLEXITY INSTITUTE

# Bacterial & Viral Bioinformatics Resource Center (BV-BRC)

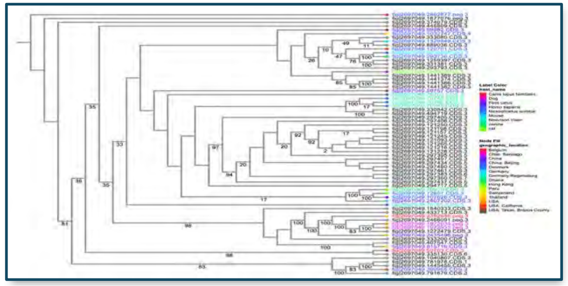
- 22+ years continuous funding ($130M total, $65M BI)
- 77K+ registered users
- 3M+ pageviews/year
- 30K citations
- 1.3M+ bacterial genomes; 15M+ viral genomes
- 1M+ user analysis jobs; 820TB user workspace
- 100+ workshops/webinars with over 4K participants
- [https://www.bv-brc.org](https://www.bv-brc.org)



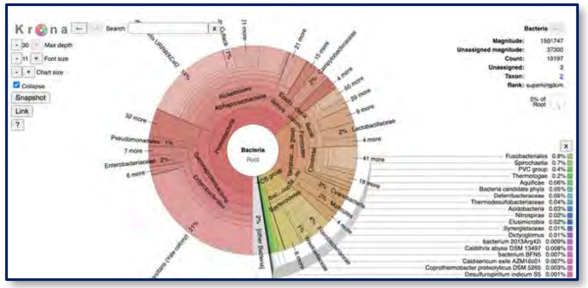

Genome Browser



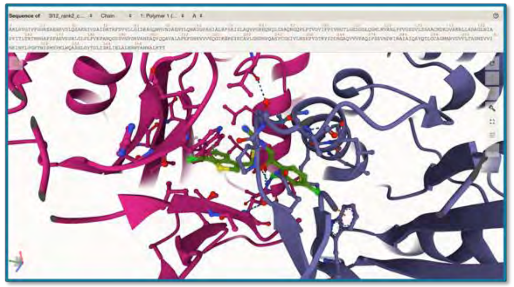
Phylogenetic Tree Analysis



Metagenomic Analysis



Protein Structure Viewer

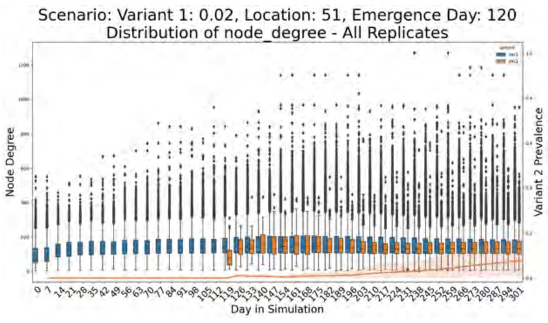


Olson et al., Introducing the Bacterial and Viral Bioinformatics Resource Center (BV-BRC): a resource combining PATRIC, IRD and ViPR, Nucleic Acids Res 2022

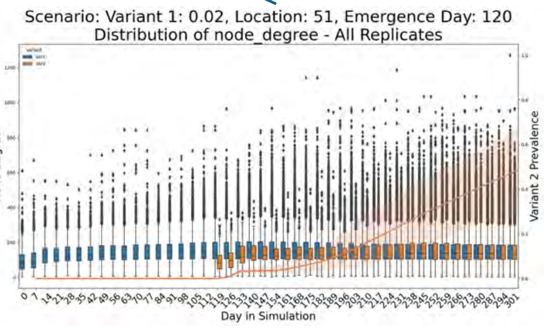
UVA | BIOCOMPLEXITY INSTITUTE

# Pathogen Genomic Center of Excellence – VA

Impact of importation scenarios on variant sampling




Household Importation




Distributed Importation

## Data Products

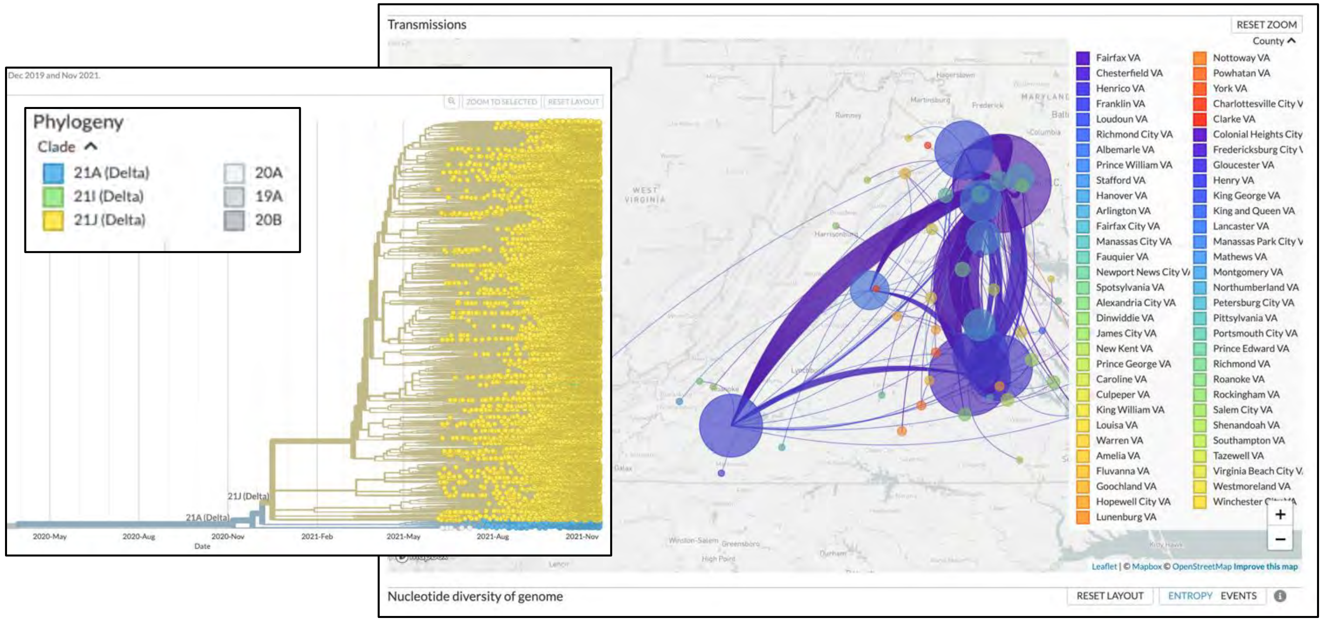
- Multi-layer digital similars
- Integrated genomic information to enhance situational awareness
- Sampling frame based on population characteristics

## Scalable Software Infrastructure

- Create tools for PH
- Enable new functionality
- Lower barrier to entry

## Analysis Techniques & Publications

- Quantify impact of strategies
- Understand spread based on genomic information
- Benchmarking bioinformatics and phylodynamics




Detailed geographic and demographic case movement and resource allocation

https://va-pgcoe.org/

UVA | BIOCOMPLEXITY INSTITUTE

# MIDAS

- **Global Modeling Network**: MIDAS unites experts advancing computational, statistical, and mathematical models of infectious disease dynamics.
- **Modeling & Data Infrastructure**: MCC develops and maintains US Scenario Modeling Hub infrastructure, including standardized GitHub repositories, pipelines, and interactive dissemination portals.
- **Methodological Innovation**: Active R&D in ensembling, data harmonization, benchmarking, and epidemiological reasoning, leveraging ML/AI to modernize outbreak modeling.
- **Technical Community Leadership**: Coordinates sustained collaboration through seminars, shared resources, and the annual MIDAS meeting to accelerate rigor, reuse, and impact.
- https://midasnetwork.us/covid-19/


UVA | BIOCOMPLEXITY INSTITUTE

# Reducing antibiotic resistant infections using hospital digital twins

Surveillance and control of the spread of hospital acquired infections (e.g., MRSA)

- Prediction of MRSA can help clinicians in choosing antibiotic treatment
  - Testing is limited and incurs delays
- Optimize infection control policies, especially for seniors and at nursing homes


Representative results

- Effective prediction tools for clinicians using Electronic Health Record (EHR) data from UVA hospital
  - Supports antimicrobial stewardship
- Optimal policies for discontinuation of isolation precautions at UVA hospital (currently in use)
- Estimating source of infection during outbreak investigations
- Ongoing: estimation of MRSA risk at Long Term Care Facilities

Cui et al., NPJ Digital Medicine 2025
Kamruzzaman et al., Improving Risk Prediction of Methicillin-Resistant Staphylococcus aureus Using Machine Learning Methods With Network Features: Retrospective Development Study, JMIR AI 2024
Cui et al., Modeling relaxed policies for discontinuation of methicillin-resistant Staphylococcus aureus contact precautions, ICHE 2024
Jang et al., Detecting Sources of Healthcare Associated Infections, AAAI 2023

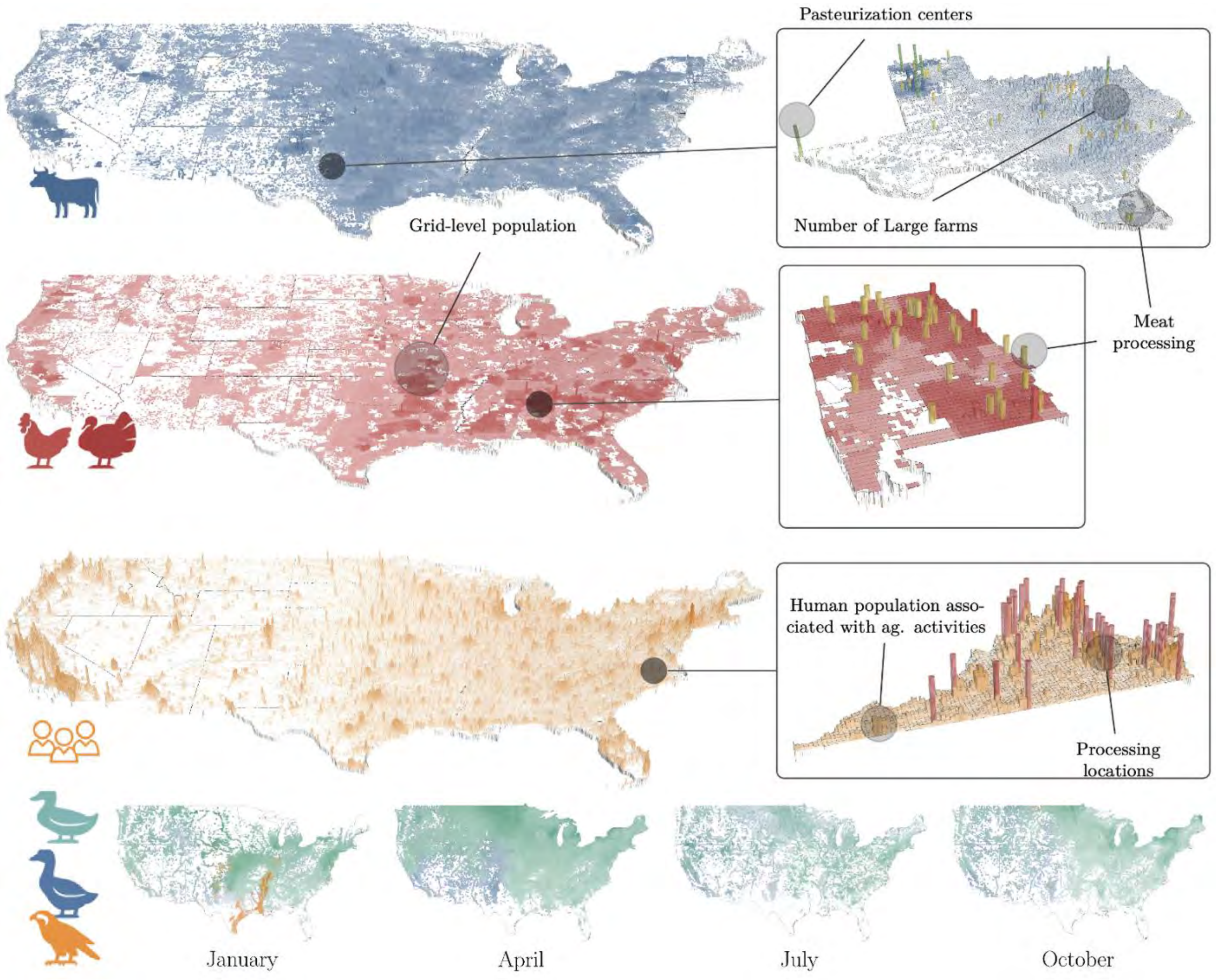
UVA BIOCOMPLEXITY INSTITUTE

# Ongoing work

# Livestock Digital Similar

A US-scale multi-layered high-resolution digital representation of:

- **Livestock**: Multiple farm-level livestock populations modelled from Census of Agriculture and Gridded Livestock of the World datasets
- **Processing centers**: Milk, meat, eggs and poultry
- **Human population**: By size and occupation type based on our US v2.4.0 digital similar
- **Wild birds**: Multiple species of migratory birds from eBirds Avian Abundance data



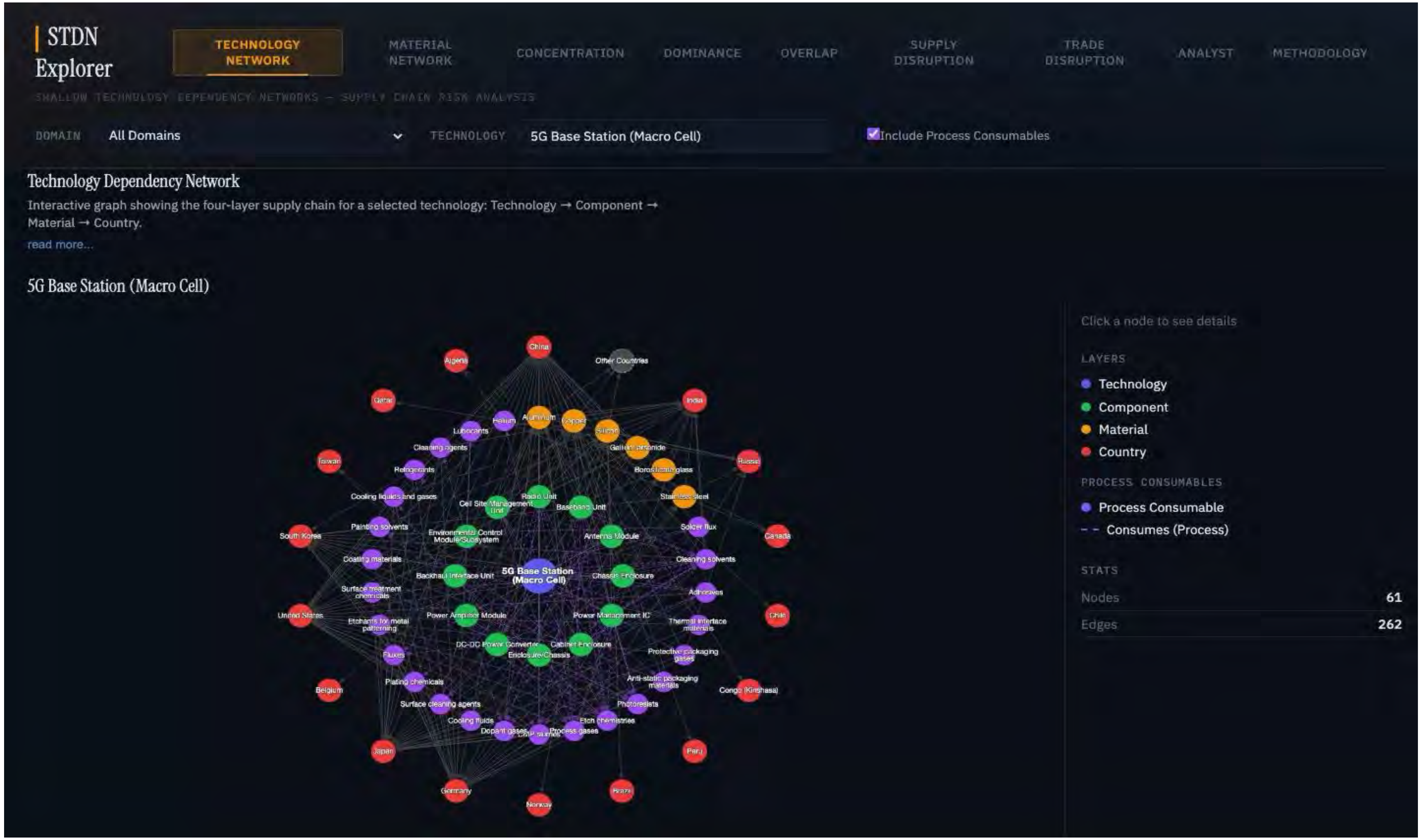

- More about this in preprint: https://arxiv.org/abs/2411.01386
- Data visualization and downloads at https://ditto.bii.virginia.edu

https://va-pgcoe.org/; **to appear in PNAS May 2026**

UVA BIOCOMPLEXITY INSTITUTE

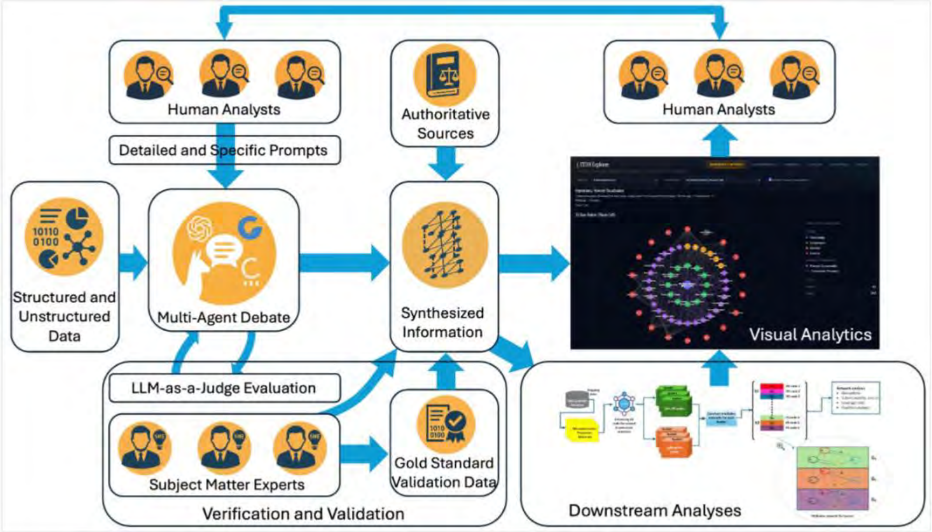
# CMAP: Critical Materials Supply Chain Analytical Platform

Supply chain disruptions and foreign dependencies are no longer theoretical risks, but active concerns shaping federal investment, industrial policy,
and defense readiness planning



# Helium Case: A 40% Cut to Qatar's Production

What the analytical stack reveals — and what it changes




*Multi-method pipeline: agent debate, validation, and visual analytics produce evidence-backed dependency data.*

**15.5%**

GLOBAL SUPPLY REMOVED

at a 40% Qatar production cut

**152/180**

SYSTEM TECHNOLOGIES EXPOSED

across microelectronics, biotech, pharma

## WHAT THE DATA SHOWS

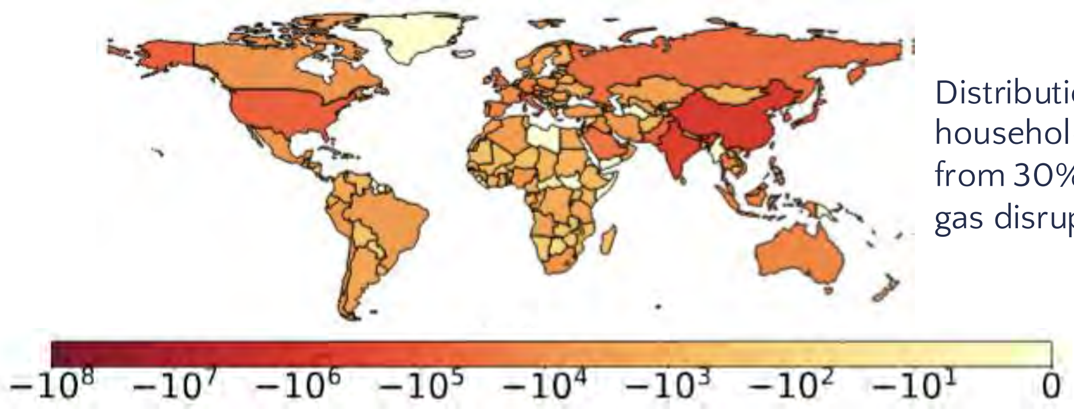
- Concentration: HHI 2,815; Qatar 38.8% + US 35.3% = 95% of global production from just 6 countries
- Qatar's profile: "narrow but deep" — leads in only 1 material, yet hits 60 of 60 microelectronics technologies (36.2% avg share)
- US direct exposure has pivoted: Qatar dominated imports through 2021; Canada has supplied ~97% since 2022
- Locked in: only 2 countries have held the top US supplier position across 9 years — no meaningful substitution

## Policy Implications

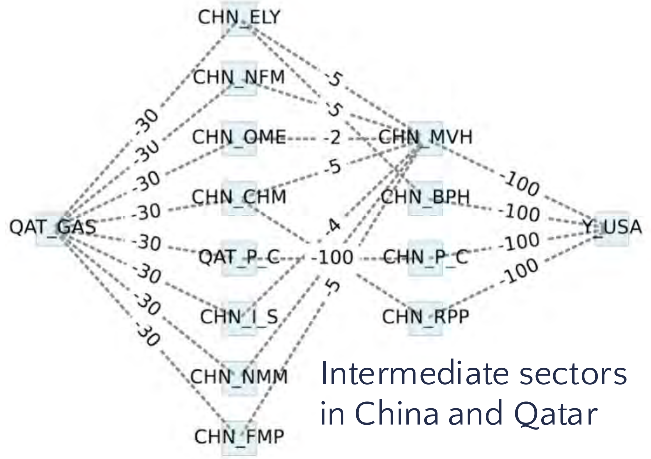
- Price-shock risk (global supply) and US import-exposure risk are now distinct questions with distinct answers
- Process consumables like helium sit outside existing critical-minerals frameworks
- The 2024 Federal Helium Reserve wind-down left no stockpile buffer for a commodity with HHI 2,815 and no substitution channel

UVA | BIOCOMPLEXITY INSTITUTE

# Cascading impact of Qatar gas disruptions




Distribution of household losses from 30% Qatar gas disruption




Intermediate sectors in China and Qatar

Disruption propagation network showing how US households are affected by Qatar disruption

## WHAT THE DATA AND MODELS SHOW

- Qatar gas disruptions disruption leads to significant losses for multiple sectors in Asia and Europe, with the largest aggregate impacts observed in India, China, and South Korea
- Expansion of production capacity among major gas-producing countries improves supply conditions and leads to broader output gains
- However, benefits remain concentrated in a few large economies

## WHAT IT MEANS FOR POLICY

- US households are also impacted, even though US is a big gas producer, and doesn't import directly from Qatar
- Adaptive mechanisms such as trade reallocation and production expansion can alleviate the effects of supply shocks but their effectiveness is limited and heterogeneous

**17%**
QATAR LNG SUPPLY REMOVED
Likely for over 3 years

**24%**
HELIUM DISRUPTION
From Qatar Rass Laffan facility

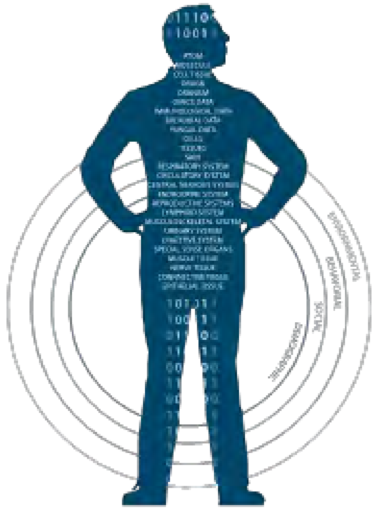
**6%**
SULPHUR DISRUPTION
From Qatar Rass Laffan facility

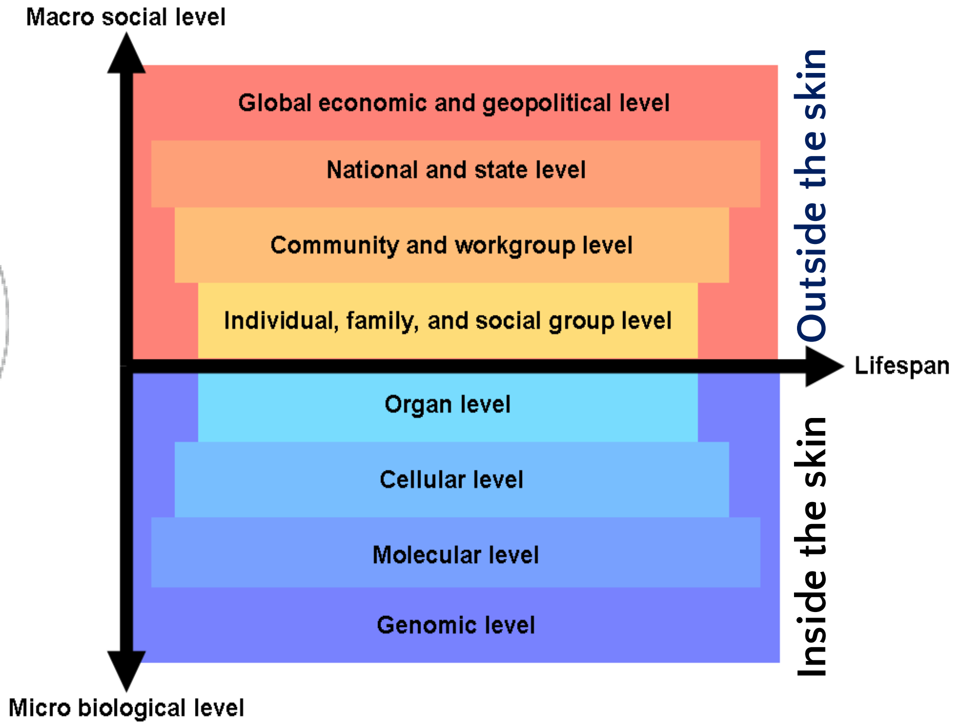
UVA | BIOCOMPLEXITY INSTITUTE

# Where we are heading

UVA | BIOCOMPLEXITY INSTITUTE

# A Grand challenge: From Genomics to Populomics to "Sociomics": Digital Twins for Human Ecology






Develop the computing and data infrastructure to represent synthetic humans, synthetic populations and ultimately synthetic societies to facilitate advancing translation of cellular and molecular data into clinical decision-making

*__Unifying scientific foundation__*: Science of Networks and Interactions

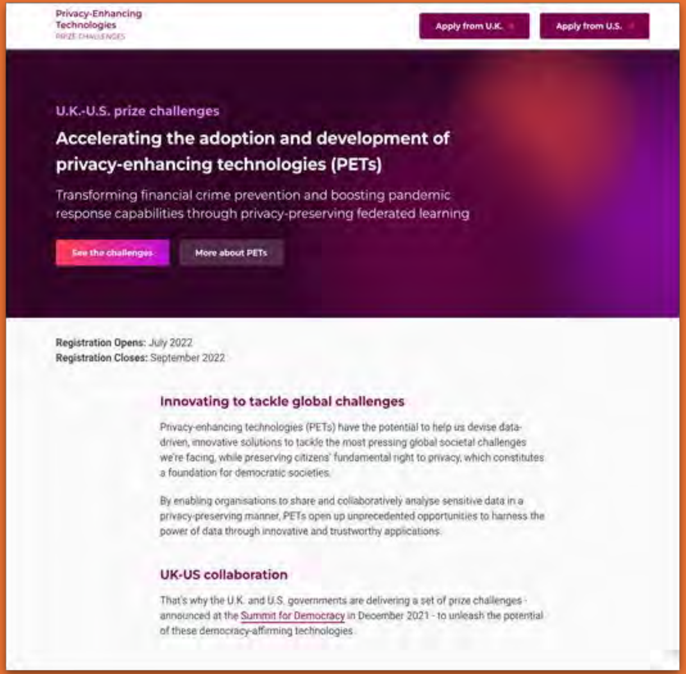
Glass, McAtee (2006). *Soc. Sci. Medicine*, 62: 1650-1671 and *Abrams, Behavior Bridges Biology and Society: Toward Systems Integration. A Transdisciplinary and Translational Challenge*

UVA BIOCOMPLEXITY INSTITUTE

# Pervasive, Personalized and Precision ($P^3$) decision-making & analytics for biocomplex networks

- **Pervasive:** Enable decision maker to make decisions at any place, anytime, and on any device
- **Personalized:** Enable decision maker to get personalized information that reflects her context
- **Precise:** Provide decision maker with precise information in space, time, and context

US-UK Prize Challenge: Privacy Enhancing Technologies




https://petsprizechallenges.com/

UVA | BIOCOMPLEXITY INSTITUTE

# Key Takeaways

- **The notion of agents and interaction are two central concepts in AI and Biocomplexity**
- **Advances in AI are proving to be force multipliers in the study of Biocomplex systems**
  - Neurosymbolic methods for developing surrogate models for planning and control problems in pandemic science
  - Agentic methods for creating more efficient workflows and decision support systems
  - Digital twins of human body for personalized medicine and care
- **Concepts from biocomplex systems can prove useful in understanding next generation AI systems**
  - Ideas in cognition, social cognition for designing future generation AI systems
  - Ideas from Immunology to develop secure AI systems
  - Ideas from Ecology