EpiSimdemics: An Efficient Algorithm for Simulating the Spread of Infectious Disease over Large Realistic Social Networks

KEITH R. BISSET

Network Dynamics and Simulation Science Laboratory
Virginia Bioinformatics Institute
Virginia Tech

Supported by Grants from NIH MIDAS, NSF HSD, CDC COE, and DoD.

NDSSL

Acknowledgments

Virginia Tech: Members, Network Dynamics & Simulation Science Laboratory, VBI

Karla Atkins, Chris Barrett, Keith Bisset, Richard Beckman, Deepti Chafekar, JiangZhou Chen, Lisa Durbeck, Steve Harris, Stephen Eubank, Annette Feng, Xizhou Feng, V.S. Anil Kumar, Bryan Lewis, Achla Marathe, Madhav Marathe, Gabriel Mateescu, Henning Mortveit, Paula Stretz

External Collaborators:

Aravind Srinivasan (U. Maryland), Doug Roberts (RTI)

NDSSL

Epidemiology is the study of patterns of health in a population and the factors that contribute to these patterns.

Computational Epidemiology is the development and use of computer models to understand the spatio-temporal diffusion of disease through populations.

Modeling the spread of infectious disease

1. Size, scale, and importance of the problem

Controlling the spread of infectious disease is one of the most important global health problems today.

2. Interaction based modeling

The only way to answer some important questions.

3. The EpiSimdemics Model

EpiSimdemics is the largest, most detailed model for evaluating the spread of contagious disease across realistic social contact networks.

VIRGINIA BIOINFORMATICS INSTITUTE AT VIRGINIA TECH VBI

NDSSL

Burden of Infectious Disease

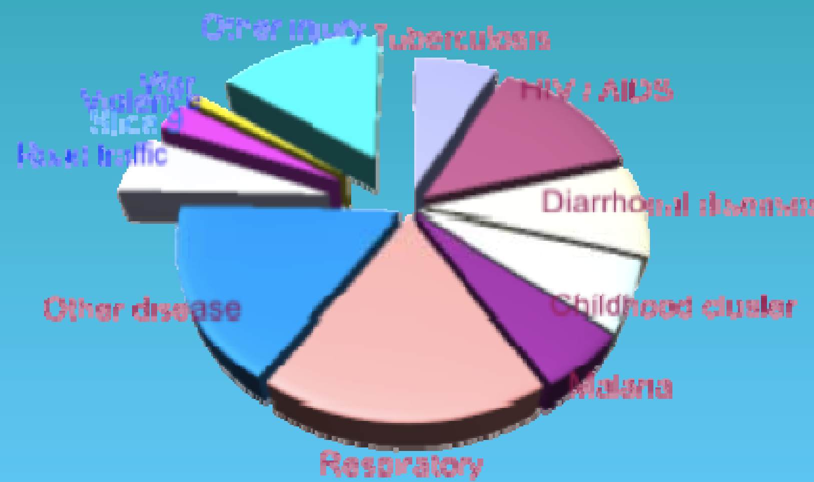
1918 Influenza Pandemic

- Worst Infectious Disease epidemic experienced by humankind
- Inside of 2 years, 50-100 million dead (3-6% world population)
- Every country was effected, very few communities were unscathed

Influenza

- Every year a new family of virus strains circulate the globe
- Seasonal pattern occurs in all places in world, though season shifts by location
- Infects ~30 million and kills 36,000 Americans each year
- Causes an estimate 200,000 hospitalizations in the US a year
- Annual cost of ~$10 billion

Deaths from Injury
5.2 million



Deaths from Communicable Disease
18.4 million

VIRGINIA BIOINFORMATICS INSTITUTE AT VIRGINIA TECH

www.who.int/topics/global_burden_of_disease/en

NDSSL

Why an HPC solution is needed

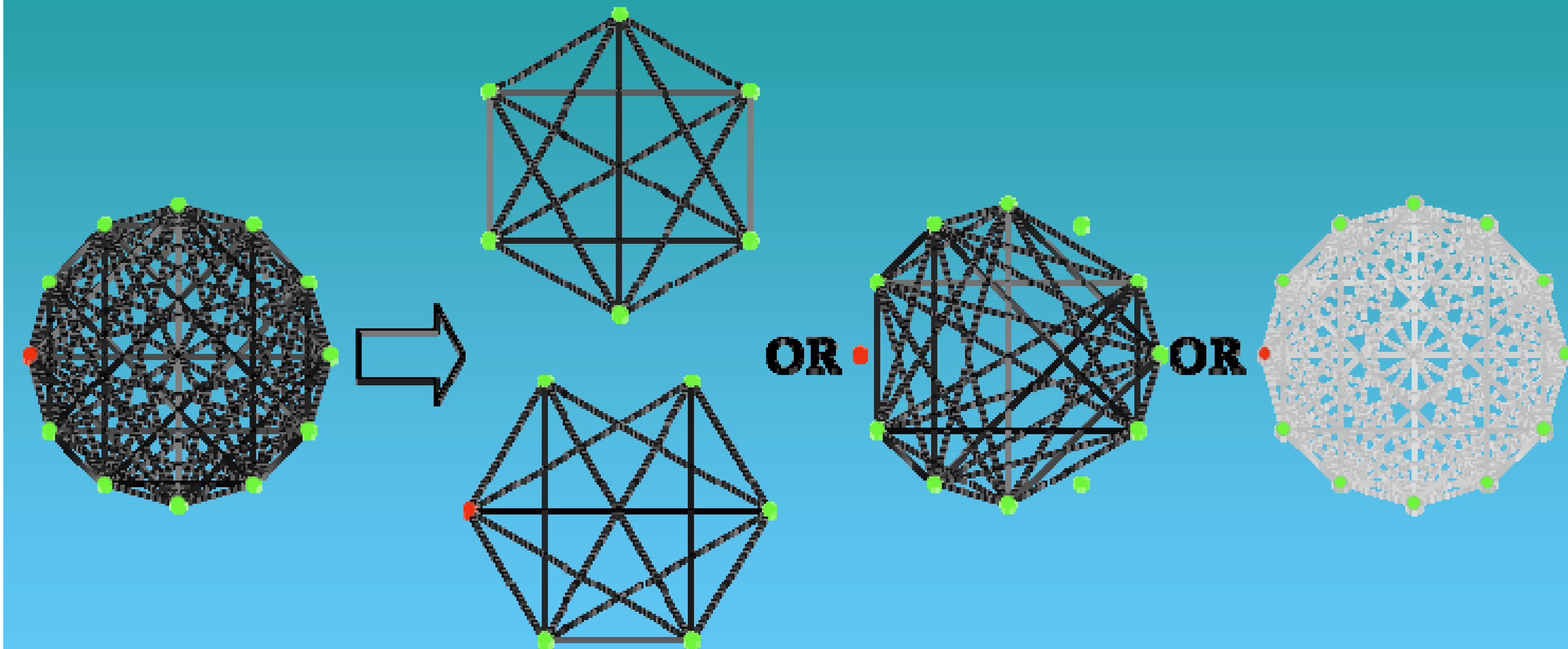
- 300 million people in United States, performing 1.5 billion daily activities
  - Input data at least 100GB
- Each person has a unique set of demographics, daily activities, and behaviors
- Complex interactions between spread of disease, government interventions, and individual reactions
- Models need to be run many times
  - Stochastic model ⇒ multiple replicates
  - Large experimental design ⇒ multiple configurations
  - 5000 runs for a study not unusual

Why Interaction-Based Models

- Certain questions need an individual model to answer correctly
- Some interventions have “unintended consequences” that need to be modeled
- Effects of complex interactions are hard to model
- Creates a co-evolving system that modifies the social network structure in response to the locally perceived progression of the disease and certain policies based on global considerations
- Models shape the questions that are asked

NDSSL

Models shape thinking



Reduce work contact by 50%?

NDSSL

Simdemics:
High resolution network based modeling

1. Create a synthetic population
   Sampling Contingency Tables, Assignment Problems
2. Derive a social network
   Assign activities (CART Trees), locations (Gravity models), Construction and analysis of large networks
3. Create a model of disease transmission
   Design probabilistic timed finite state automata based on data
4. Study how the disease spreads over this network
   Simulation of a diffusion process
5. Study effects of different policies and behavioral change
   Optimal Selection of critical individuals, whom do we isolate, vaccinate, implementable policies.
   Investigate how individual decisions affect disease spread

VIRGINIA BIOINFORMATICS INSTITUTE AT VIRGINIA TECH VBI

NDSSL

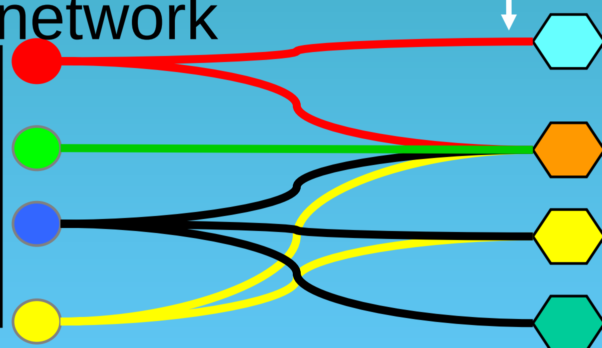
Step 1: Synthetic populations

- **Who**: *People*
  - Individuals
  - Household structure
  - Statistically identical to U.S. Census
  - Assigned to Home Locations

NDSSL

Step 2: Dynamic social contact network

- Demographically match schedules
- Assign appropriate locations by activity and distance
- Determine duration of interaction
- Generate social network

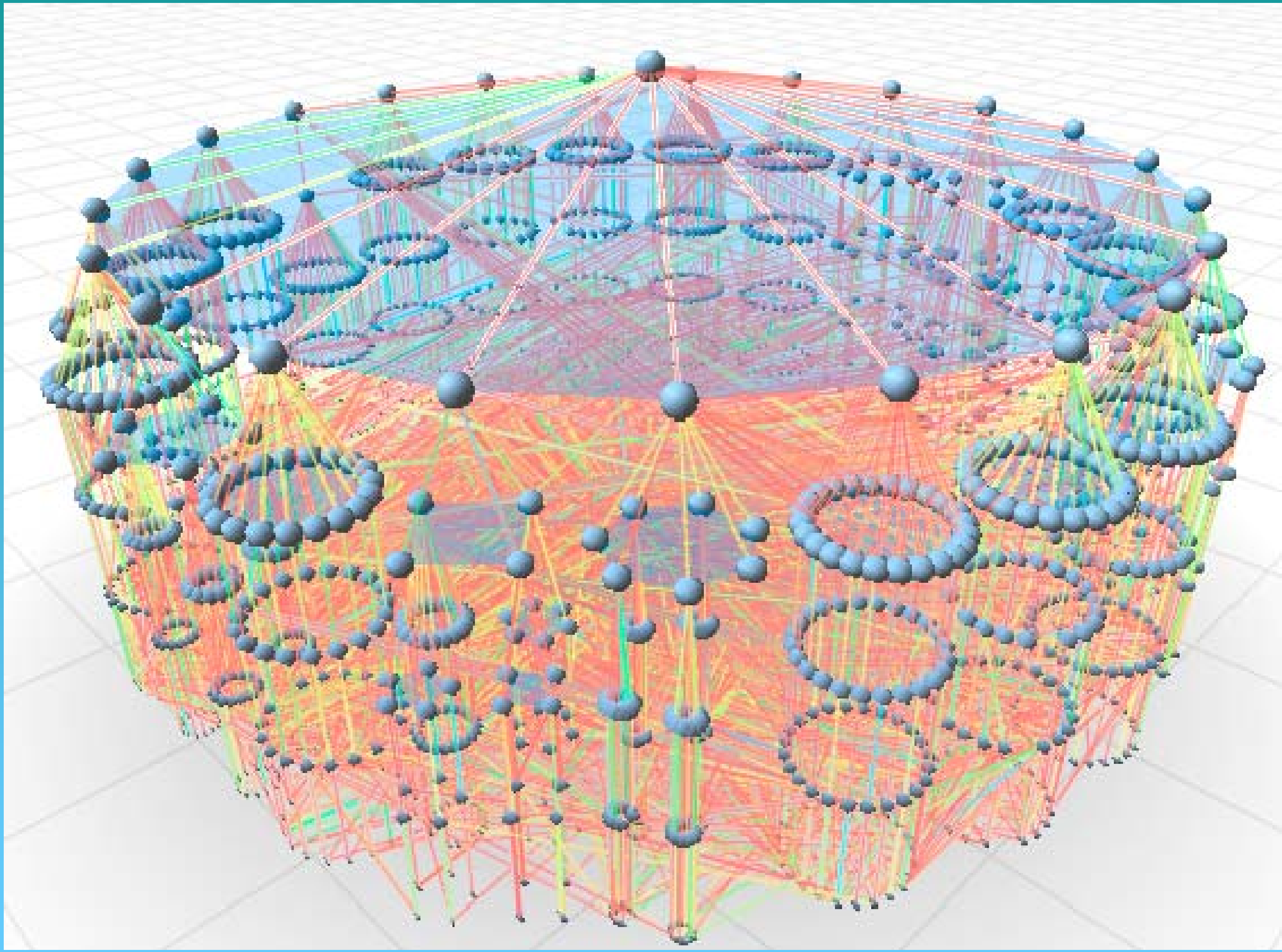


People Vertex:
- age
- household size
- gender
- income ..

Location Vertex:
- (x,y,z)
- land use .
- Business type

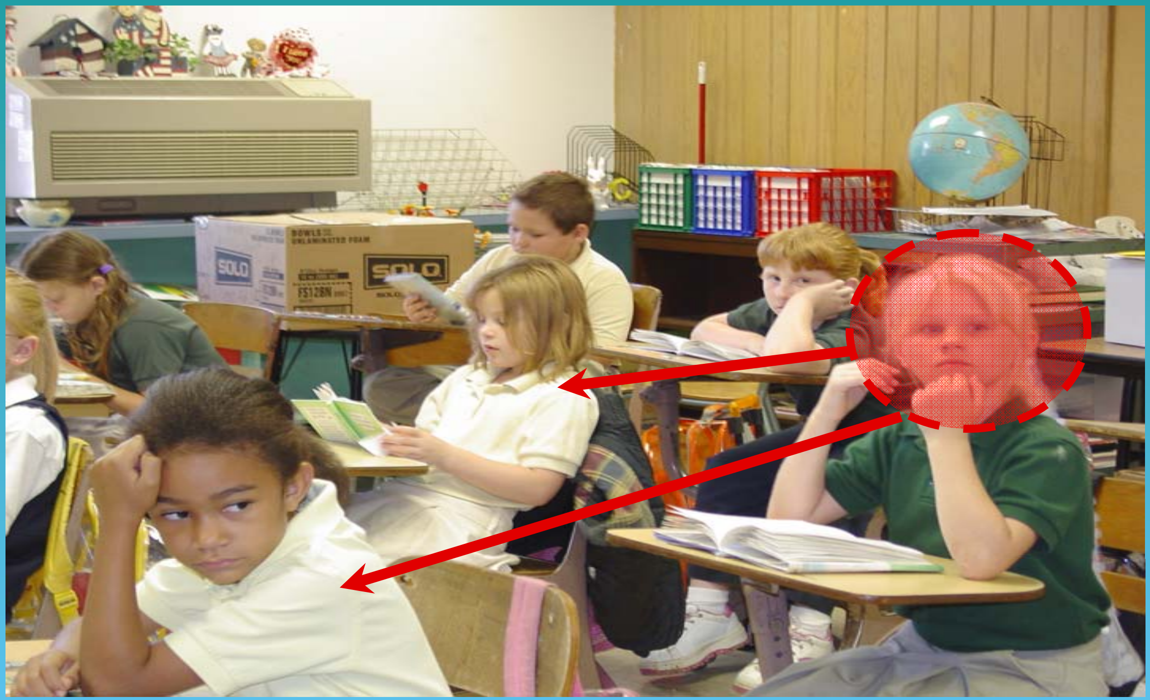
Edge labels
- activity type: shop, work, school
- (start time 1, end time 1)
- (start time 2, end time 2)

Person - Person Contact Network



NDSSL

Step 3: Model disease propagation



$$p_i = 1 - exp(\tau \sum_{r \in R} N_r ln(1 - r s_i \rho))$$

Disease can be spread from one person to another.

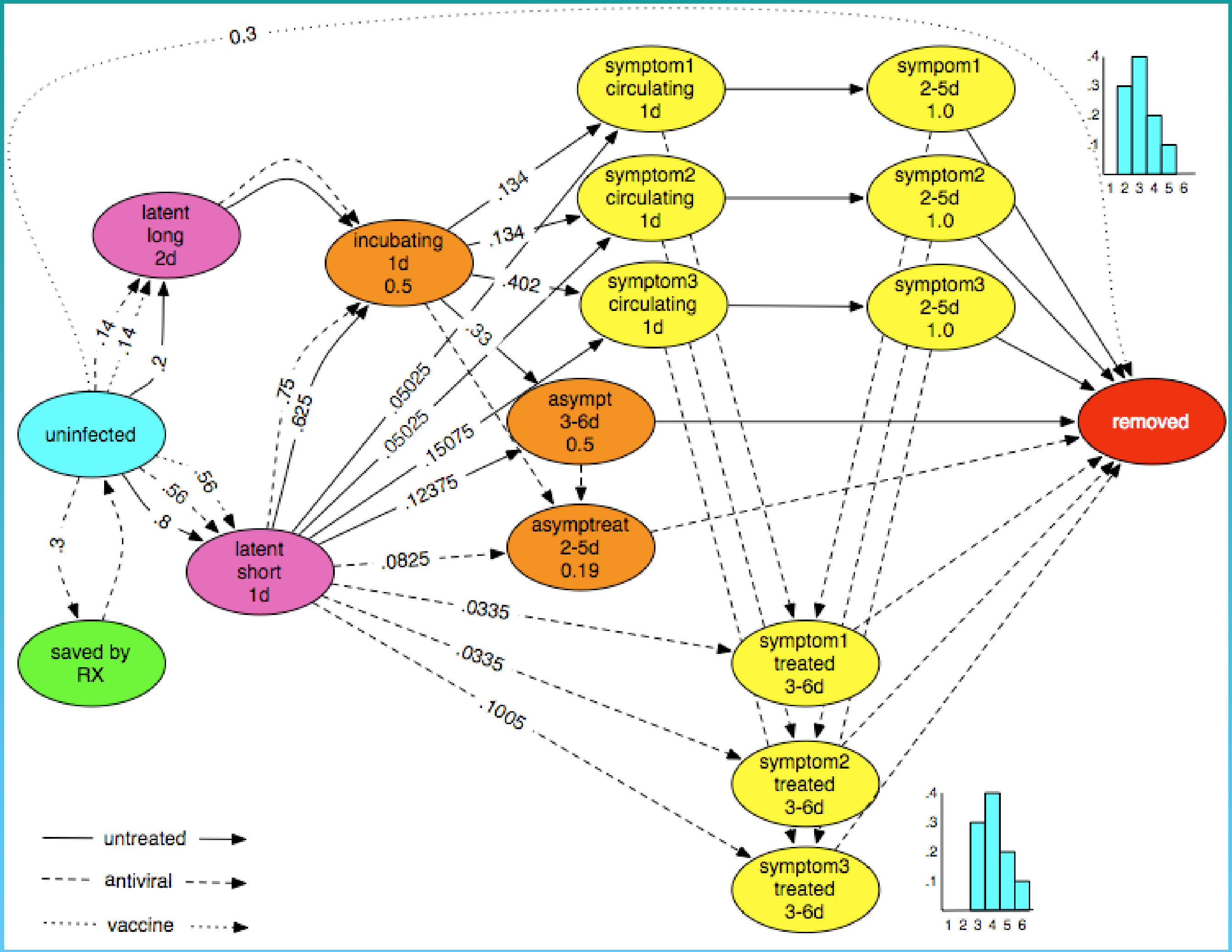
The probability of transmission can depend on:

- type of disease
- duration and type of contact
- person's characteristics
    - age, health state, etc.

VIRGINIA BIOINFORMATICS INSTITUTE AT VIRGINIA TECH

NDSSL

Step 3: Model disease progression



Probabilistic Timed Transition System

NDSSL

Step 5: Study Policy and Behavior

Targeted Layered Containment

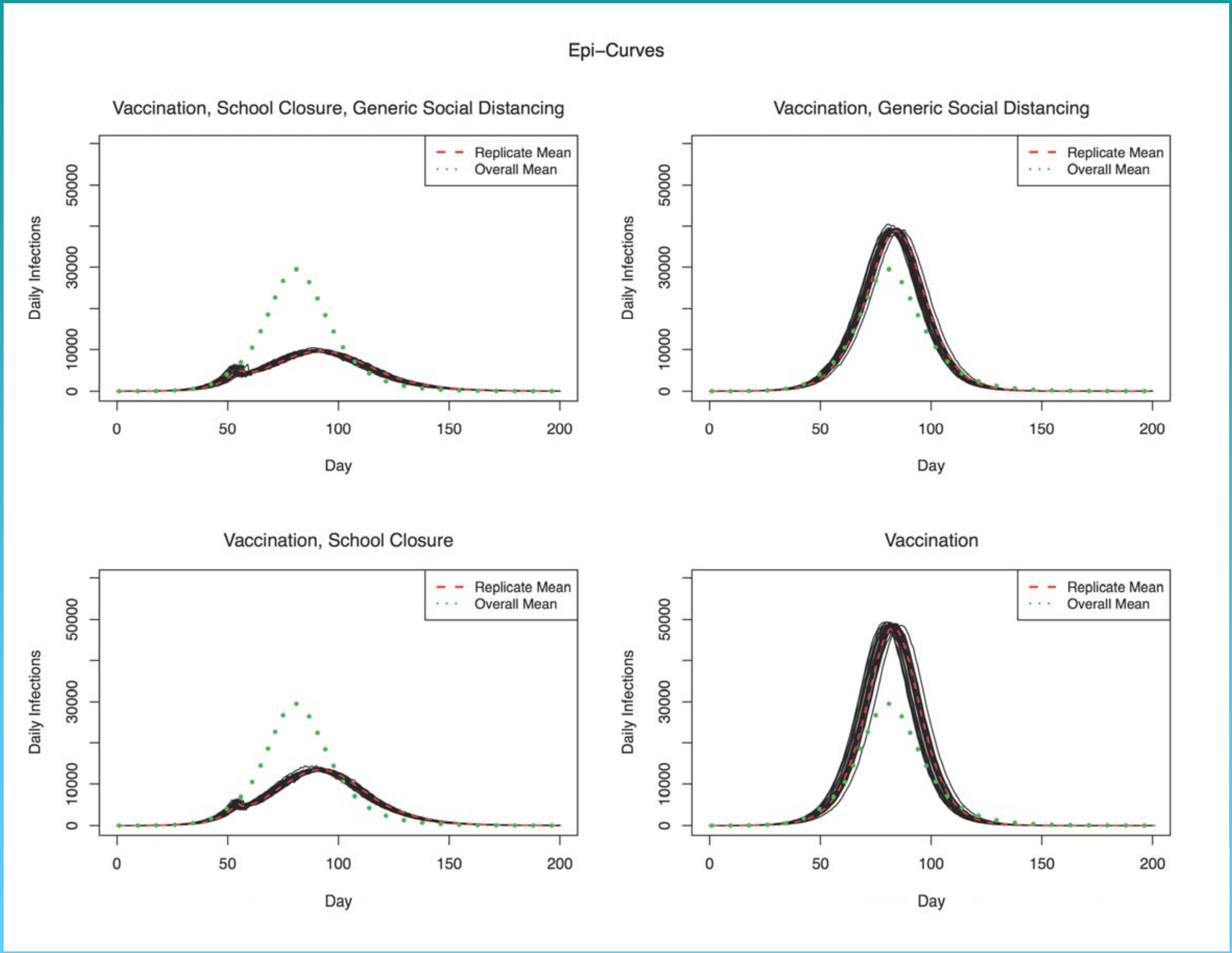
- Treat household contacts
  - Prophylax, isolate, and/or quarantine
- Close Schools
  - Keep children isolated at home
- Workplace Social Distancing
- Community Social Distancing

Each layer has its own trigger point and compliance rate.

Compliance can depend on individual demographics.

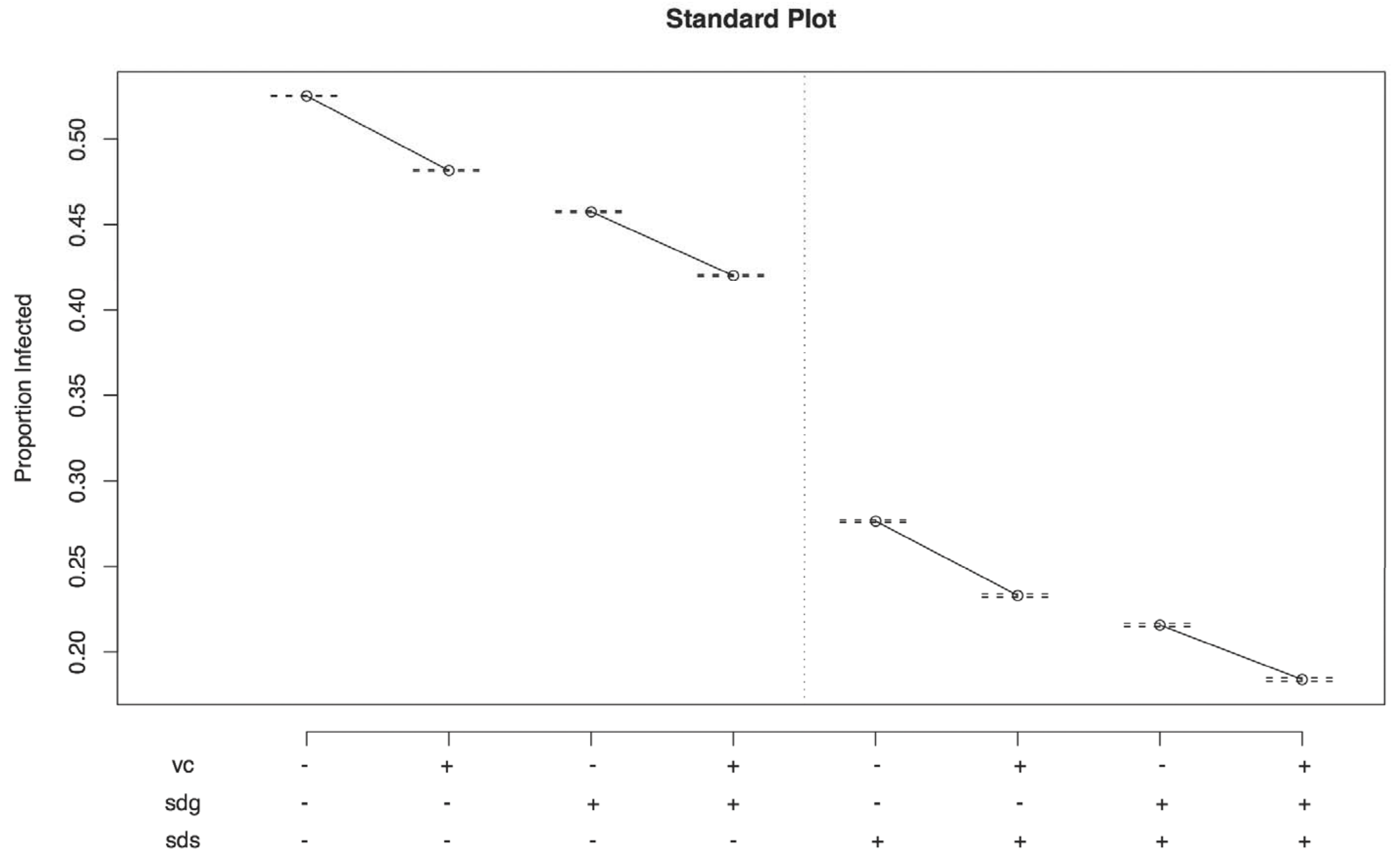
The scenario is specified externally from the model.

Sample Results



NDSSL

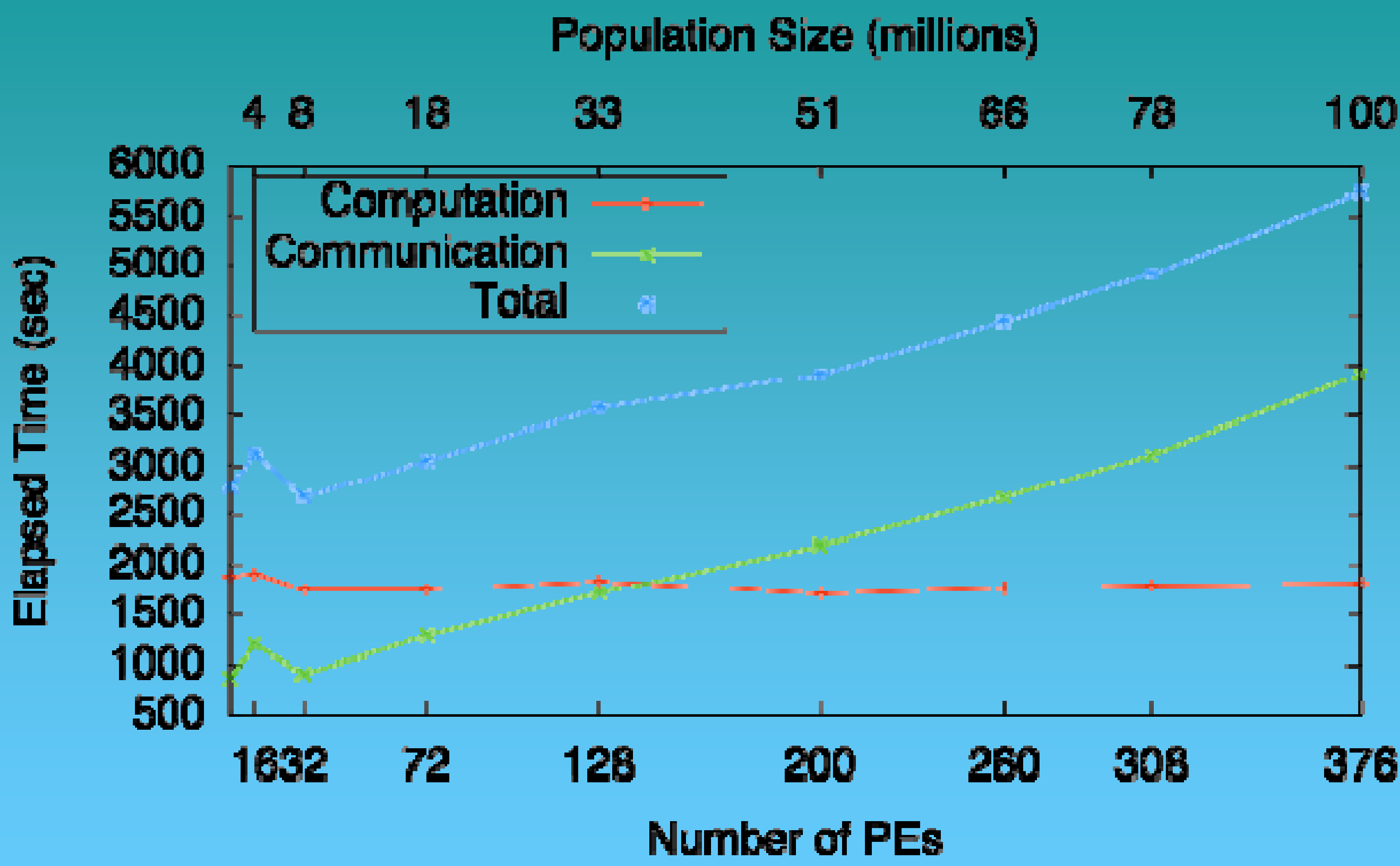
Sample Results



VIRGINIA BIOINFORMATICS INSTITUTE AT VIRGINIA TECH

NDSSL

Performace - Weak Scaling



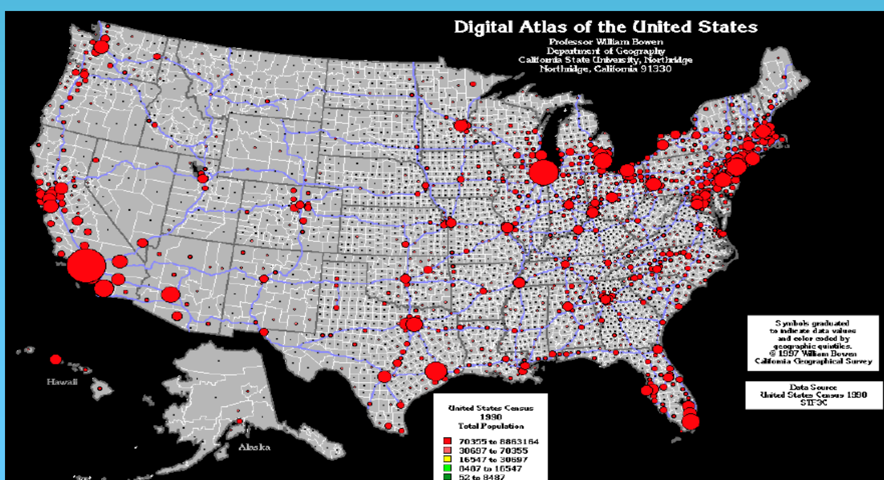
VIRGINIA BIOINFORMATICS INSTITUTE AT VIRGINIA TECH

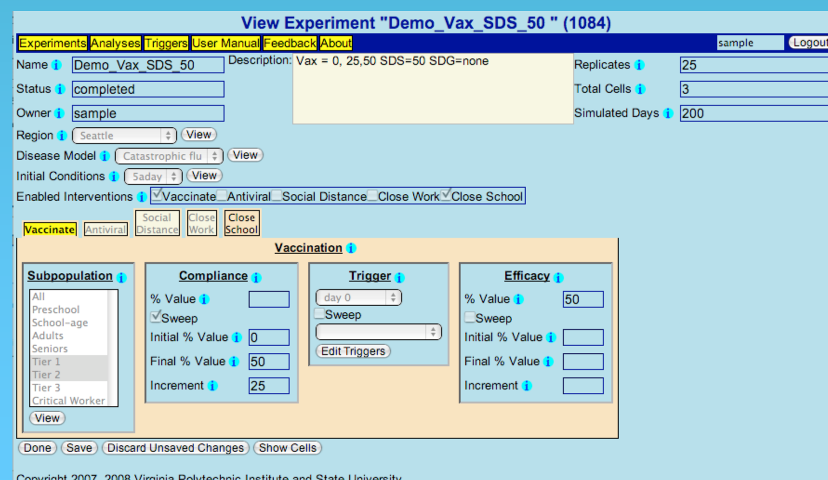
NDSSL

Conclusions

- Detailed modeling of infectious disease is an important problem
- EpiSimdemics
  - Flexible
  - Scalable
  - Highly detailed
  - Interaction based
- Used for multiple sponsor-defined studies
  - NIH, CDC, HHS, DoD, DHS

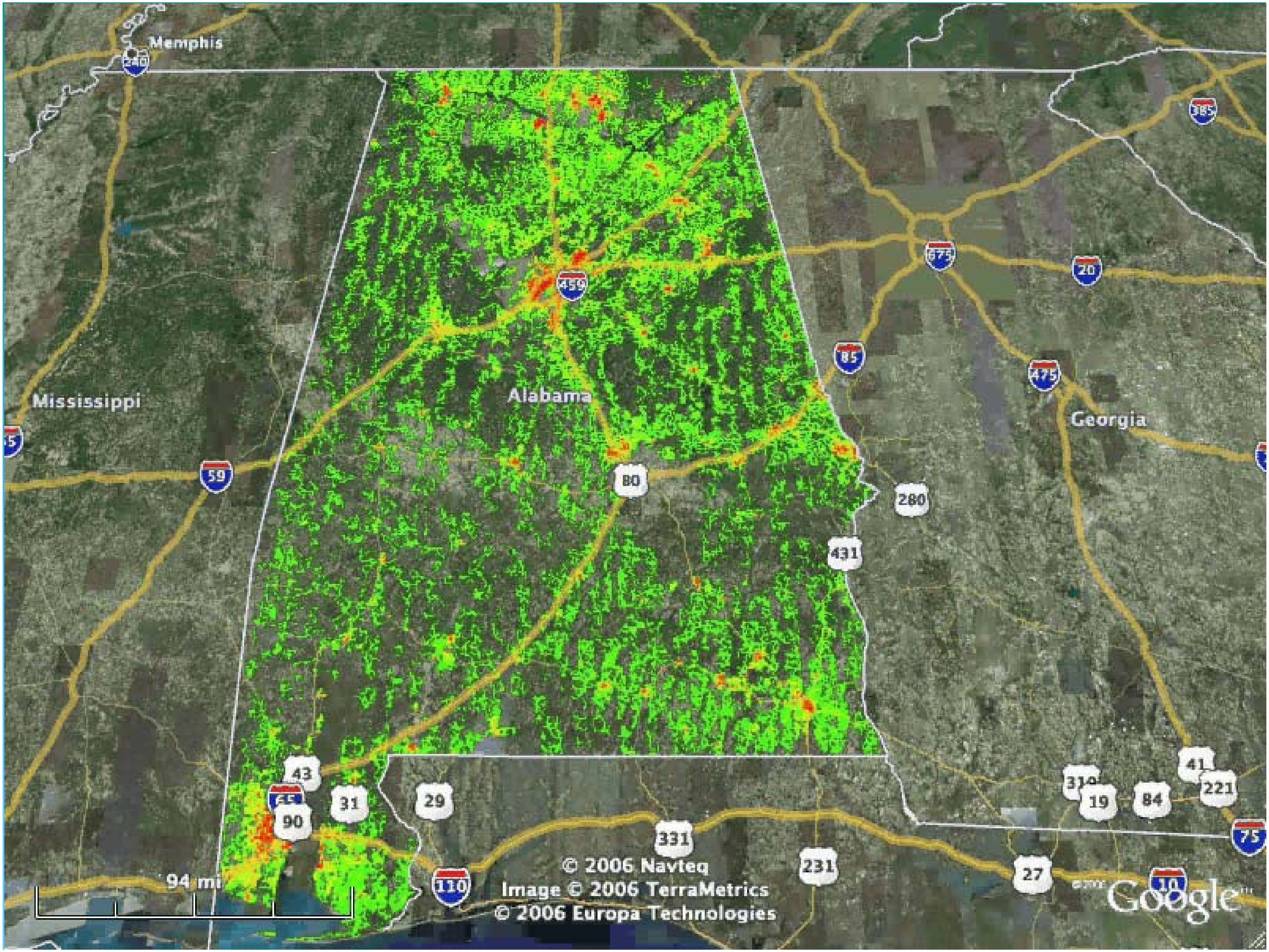
Where are we headed

- National Scale Modeling - terascale to petascale
- General Contagion Model
  - vector borne disease, norms, fads, influence
- Multiple co-evolving diseases
- General behavior models
- Simfrastructure: a Grid Based, SOA, Distributed Modeling Framework
- 

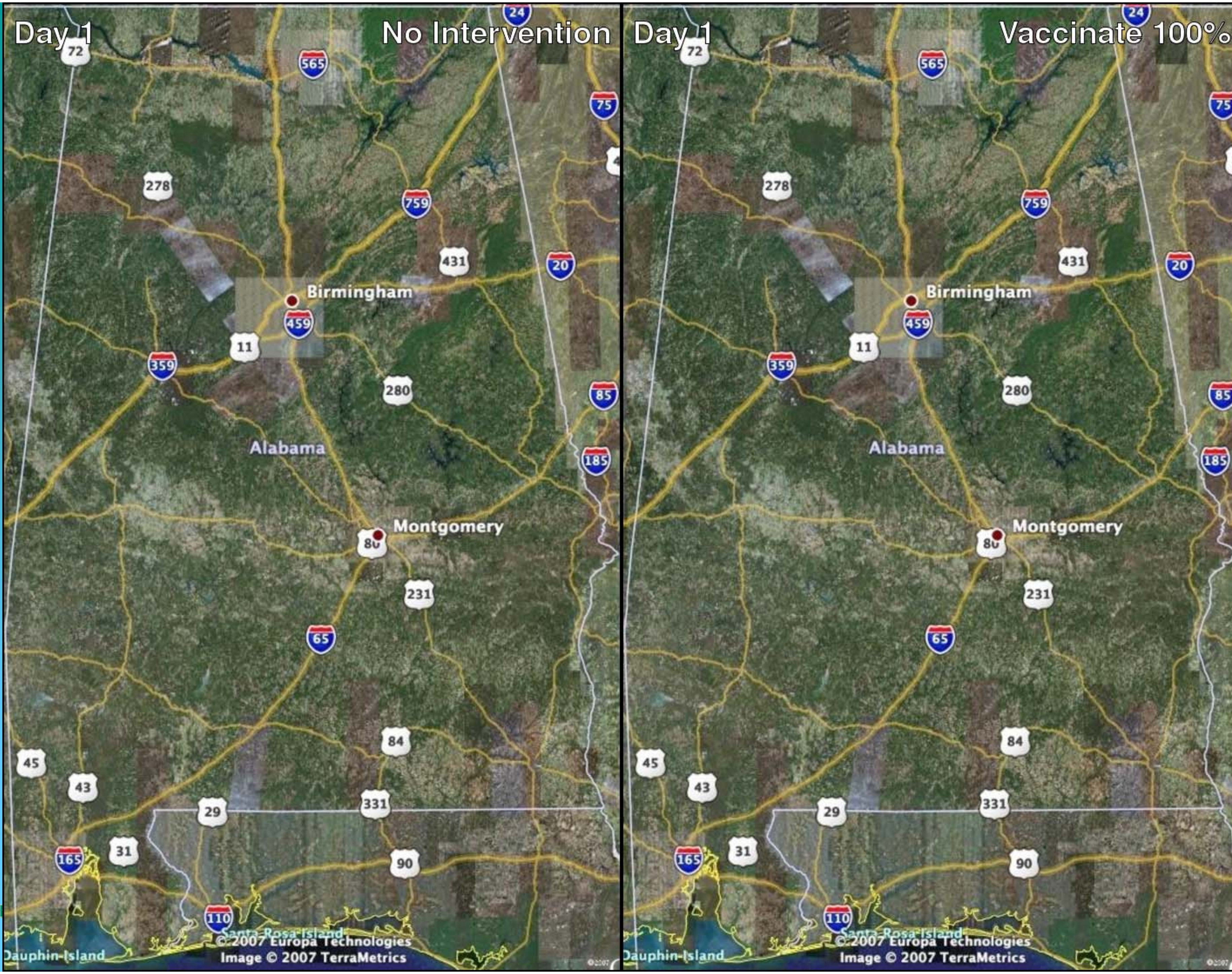


VIRGINIA BIOINFORMATICS INSTITUTE AT VIRGINIA TECH

NDSSL



Memphis
Mississippi
Alabama
Georgia
94 mi
© 2006 Navteq
Image © 2006 TerraMetrics
© 2006 Europa Technologies
Google


Day 1
No Intervention
Day 1
Vaccinate 100%
Birmingham
Alabama
Montgomery
Dauphin Island
Santa Rosa Island
© 2007 Europa Technologies
Image © 2007 TerraMetrics