# Synthetic social habitats for public policy and decision making

**Madhav Marathe**
*Network Dynamics and Simulation Science Laboratory(NDSSL), Biocomplexity Institute & Dept. of Computer Science*
*Virginia Tech*

TR-2017-042
*http://www.bi.vt.edu/ndssl*

VirginiaTech
*Invent the Future®*

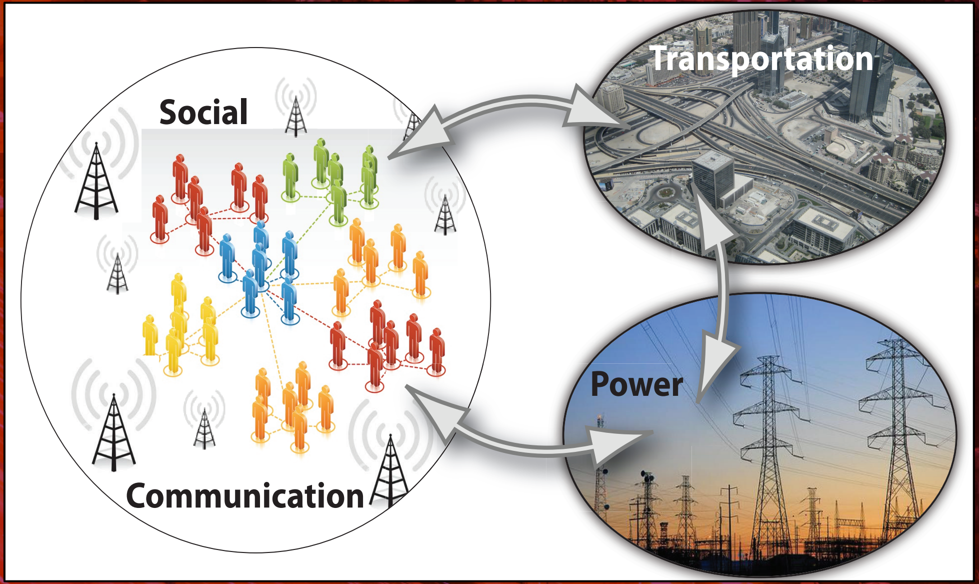
BIO COM PLEX

# Acknowledgements


- Thanks to members of the Network Dynamics and Simulation Science Laboratory, BI and Discovery Analytics Center (DAC), both at Virginia Tech and our collaborators at SUNY Albany, UMD, Yale, Harvard, Northeastern and others.
- Support
  - National Science Foundation: HSD grant SES-0729441, NSF PetaApps grant OCI-0904844, NSF NetSE grant CNS-1011769, NSF SDCI grant OCI-1032677, NSF Big Data and NSF DIBBS projects.
  - Defense Threat Reduction Agency grant HDTRA1-11-1-0016, DTRA CNIMS contract HDTRA1-11-D-0016-0001, HDTRA1-11-D-0016-0005
  - National Institute of Health Midas grant 2U01GM070694-09,
  - Intelligence Advanced Research Projects Activity (IARPA) via the US Department of Interior (DoI) National Business Center (NBC): D12PC000337.

# Increasing Interdependence and Coupling Between Socio-Technical Systems

Multi-scale, multi-network modeling is critical




ICT advances **imply** qualitative change for supporting public policy

Unprecedented opportunities and potential risks to the global environment, social stability

# Data & computing are pervasive

Everyone wants to make pervasive real-time and personalized decisions

VirginiaTech
Biocomplexity Institute

# Modeling for integrated reasoning about situations and actions

- **<u>Vision:</u>** Real time policy support for ST systems
- **<u>Goal:</u>** Build a flexible suite of informatics tools that
  - *Synthesize*: available data to produce consistent and meaningful representation of the underlying system
  - *Provide*: range of interpretations of incoming measurements
  - *Evaluate*: range of response actions and behaviors
  - *Monitor*: Effect of policy responses
  - Support coordination among diverse stakeholders
- Want to go beyond prediction
- Systems should be useable by analysts and not just by computing experts

EDITORIAL

Harvey V. Fineberg is president of the Institute of Medicine.

Mary Elizabeth Wilson is associate professor of Global Health and Population at the Harvard School of Public Health and associate clinical professor at Harvard Medical School, Boston, MA.

## Epidemic Science in Real Time

FEW SITUATIONS MORE DRAMATICALLY ILLUSTRATE THE SALIENCE OF SCIENCE TO POLICY THAN AN epidemic. The relevant science takes place rapidly and continually, in the laboratory, clinic, and community. In facing the current swine flu (H1N1 influenza) outbreak, the world has benefited from research investment over many years, as well as from preparedness exercises and planning in many countries. The global public health enterprise has been tempered by the outbreak of severe acute respiratory syndrome (SARS) in 2002–2003, the ongoing threat of highly pathogenic avian flu, and concerns over bioterrorism. Researchers and other experts are now able to make vital contributions in real time. By conducting the right science and communicating expert judgment, scientists can enable policies to be adjusted appropriately as an epidemic scenario unfolds.

In the past, scientists and policy-makers have often failed to take advantage of the opportunity to learn and adjust policy in real time. In 1976, for example, in response to a swine flu outbreak at Fort Dix, New Jersey, a decision was made to mount a nationwide immunization program against this virus because it was deemed similar to that responsible for the 1918–1919 flu pandemic. Immunizations were initiated months later despite the fact that not a single related case of infection had appeared by that time elsewhere in the United States or the world (www.iom.edu/swinefluaffair). Decision-makers failed to take seriously a key question: What additional information could lead to a different course of action? The answer is precisely what should drive a research agenda in real time today.

In the face of a threatened pandemic, policy-makers will want real-time answers in at least five areas where science can help: pandemic risk, vulnerable populations, available interventions, implementation possibilities and pitfalls, and public understanding. Pandemic risk, for example, entails both spread and severity. In the current H1N1 influenza outbreak, the causative virus and its genetic sequence were identified in a matter of days. Within a couple of weeks, an international consortium of investigators developed preliminary assessments of cases and mortality based on epidemic modeling.*

Specific genetic markers on flu viruses have been associated with more severe outbreaks. But virulence is an incompletely understood function of host-pathogen interaction, and the absence of a known marker in the current H1N1 virus does not mean it will remain relatively benign. It may mutate or acquire new genetic material. Thus, ongoing, refined estimates of its pandemic potential will benefit from tracking epidemiological patterns in the field and viral mutations in the laboratory. If epidemic models suggest that more precise estimates on specific elements such as attack rate, case fatality rate, or duration of viral shedding will be pivotal for projecting pandemic potential, then these measurements deserve special attention. Even when more is learned, a degree of uncertainty will persist, and scientists have the responsibility to accurately convey the extent of and change in scientific uncertainty as new information emerges.

A range of laboratory, epidemiologic, and social science research will similarly be required to provide answers about vulnerable populations; interventions to prevent, treat, and mitigate disease and other consequences of a pandemic; and ways of achieving public understanding that avoid both over- and underreaction. Also, we know from past experience that planning for the implementation of such projects has often been inadequate. For example, if the United States decides to immunize twice the number of people in half the usual time, are the existing channels of vaccine distribution and administration up to the task? On a global scale, making the rapid availability and administration of vaccine possible is an order of magnitude more daunting.

Scientists and other flu experts in the United States and around the world have much to occupy their attention. Time and resources are limited, however, and leaders in government agencies will need to ensure that the most consequential scientific questions are answered. In the meantime, scientists can discourage irrational policies, such as the banning of pork imports, and in the face of a threatened pandemic, energetically pursue science in real time.

**– Harvey V. Fineberg and Mary Elizabeth Wilson**

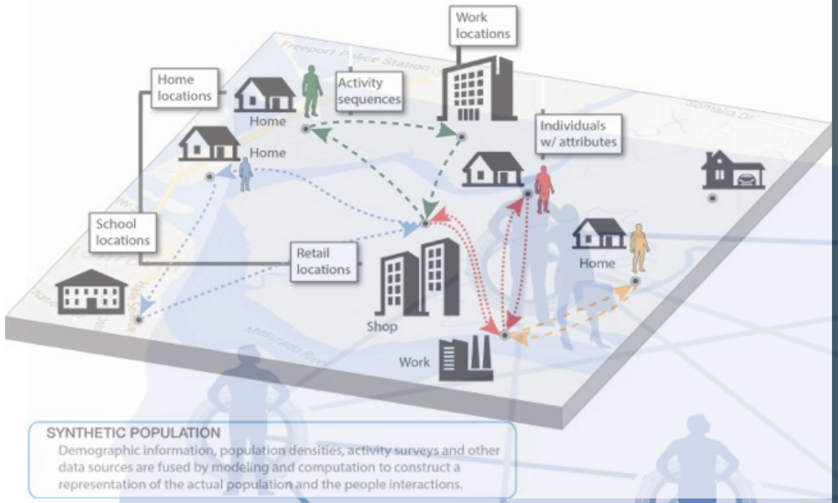
10.1126/science.1176297

*C. Fraser *et al.*, *Science* 11 May 2009 (10.1126/science.1176062).

CREDITS: (LEFT) ED LAUSCH; (RIGHT) JUPITERIMAGES

# A Natural Approach:
Create living social habitats

VirginiaTech
Biocomplexity Institute




**Step 1:** Create a spatially explicit, highly resolved synthetic social coordinate system

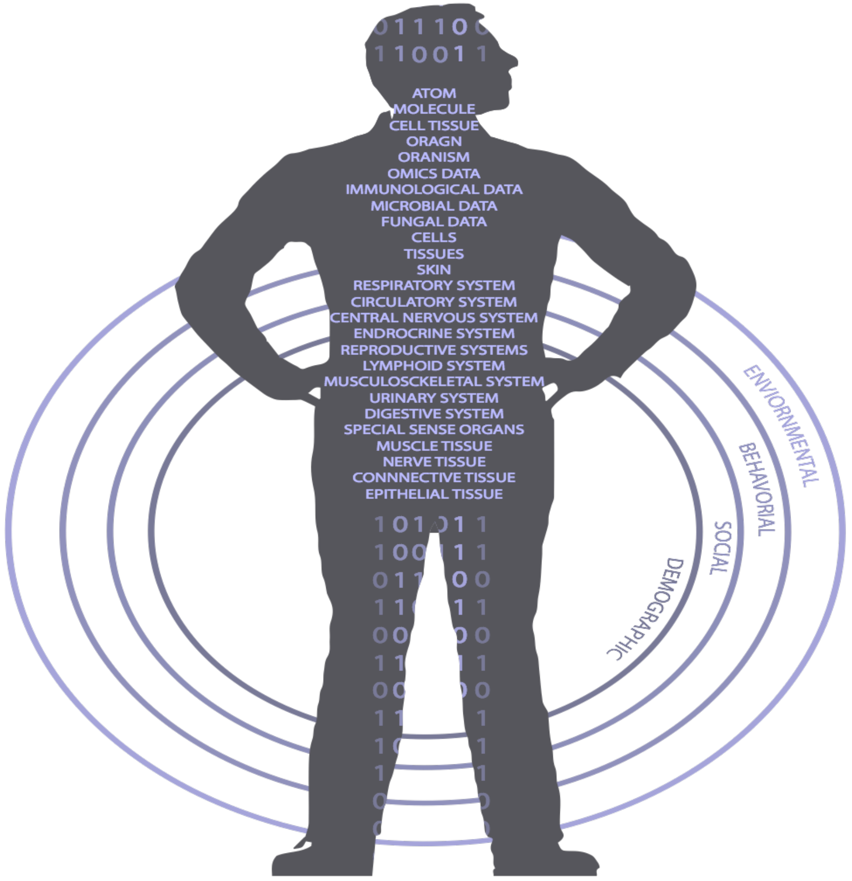
A natural structure for multi-scale modeling and analytics

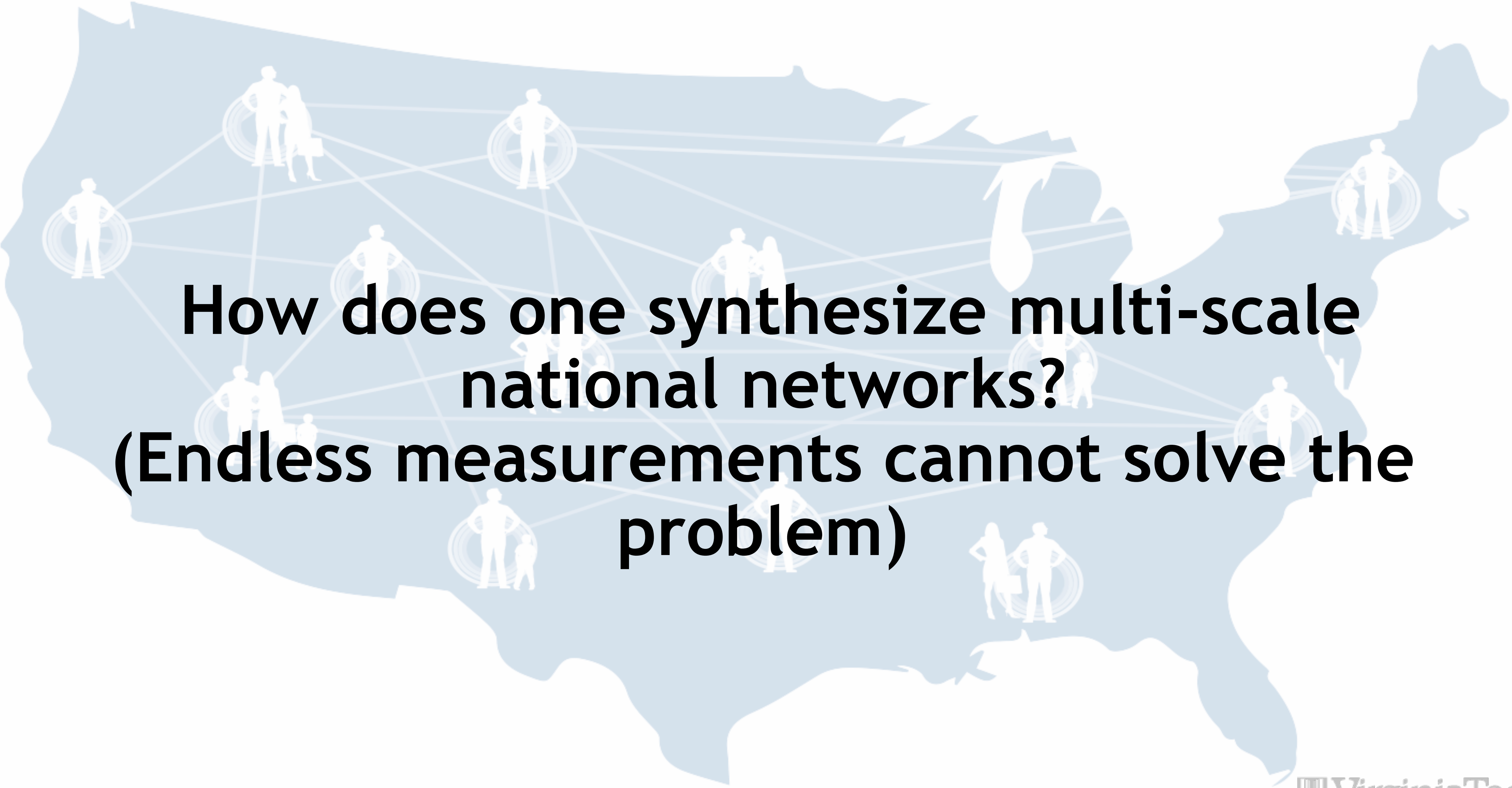
- **A Tutorial on Generating Synthetic Populations for Social Modeling**
- **IJCAI 2016, 2016 & AAMAS 2016, 2016**
- **http://staff.vbi.vt.edu/swarup/synthetic_population_tutorial/**


VirginiaTech
Biocomplexity Institute

# What is a synthetic agent & network ?

- A *representation* of elements' and states that is not intended to precisely match any snapshot of the system, but to provide a statistically accurate overall picture:
  - people, places, things
  - cells, cytokines, organs
- A *synthesis* of incommensurate data
- E.g.: A synthetic human agent
  - Can have demographic, social, health, cognitive, cultural attributes
  - These attributes need to be statistically accurate to attributes of humans

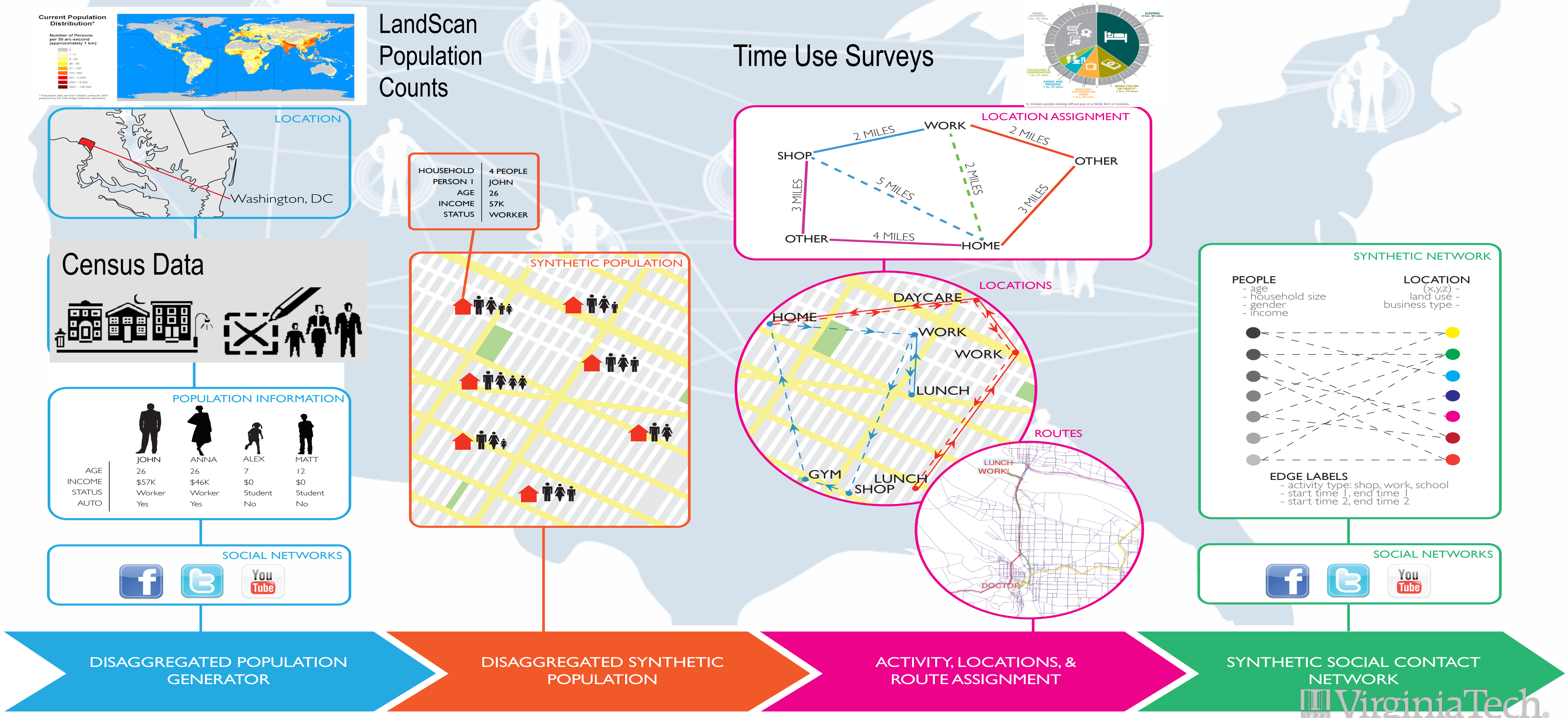




# How does one synthesize multi-scale national networks? (Endless measurements cannot solve the problem)

Virginia Tech
Biocomplexity Institute

# Constructing synthetic multi-scale synthetic networks at scale




VirginiaTech
Biocomplexity Institute

# Multi-scale models of mobility and social interactions

**Within homes and small communities**

- Sensors or surveys
- Homes, hospitals, work place, funeral homes
- Random graph models
- 1 m to 1 km

**Daily movement in a city**

- Activity surveys, phones, GPS traces
- Home to work, school recreation
- CART methods
- 1-50 km

**Intercity across states**

- Cell phone data, satellite data, surveys from UN
- Migration or business
- CART/gravity models
- 50-1000 km

**Between Countries**

- Cell phone, flight data, satellite data
- Mass migration, business
- CART models
- 100 to 1000s of km

# Why is synthetic information useful?

- Provides a spatially and individually resolved data structure
  - Fusion of diverse information in a statistically consistent manner, e.g. (demographic + health + 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 natural data structure for multi-scale models

A socio-technical multi-scale *coordinate system* for incorporating new (intra- and inter) agentic information

# Global synthetic information

2GB/M people
Storage

7 Billion
Synthetic individuals

28+ Billion
Interactions

40+
Databases

220 countries
synthetic populations and networks constructed

50K+
Files in which data is stored

5 Days
Compute time

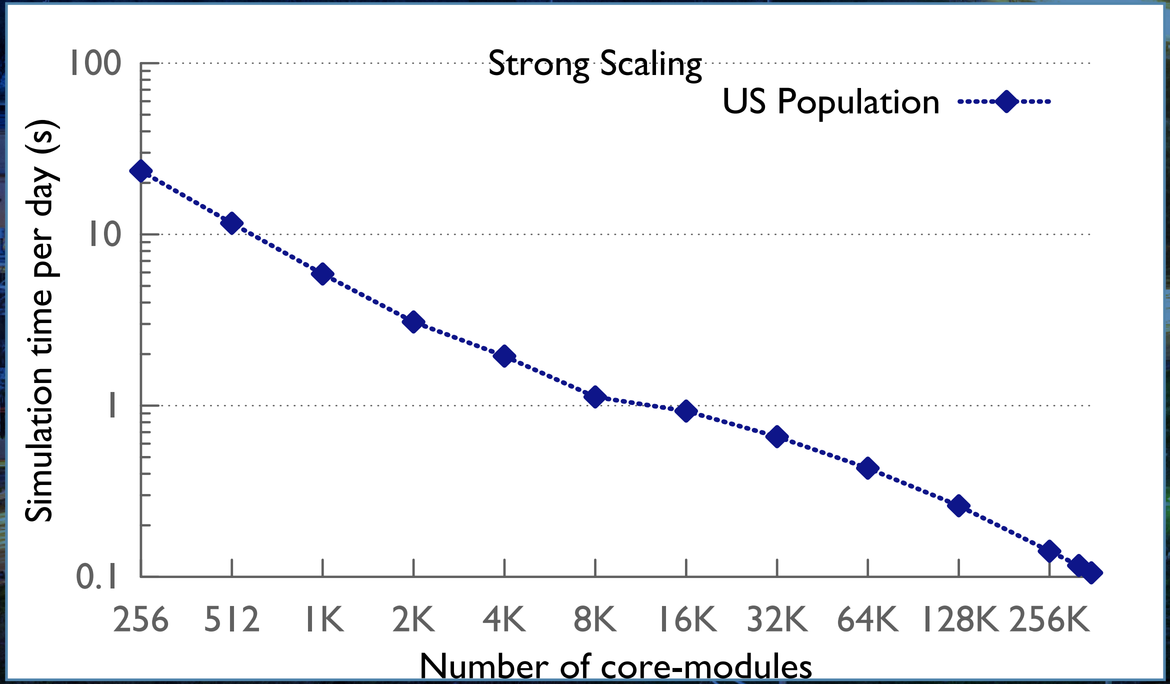
8TB
Storage

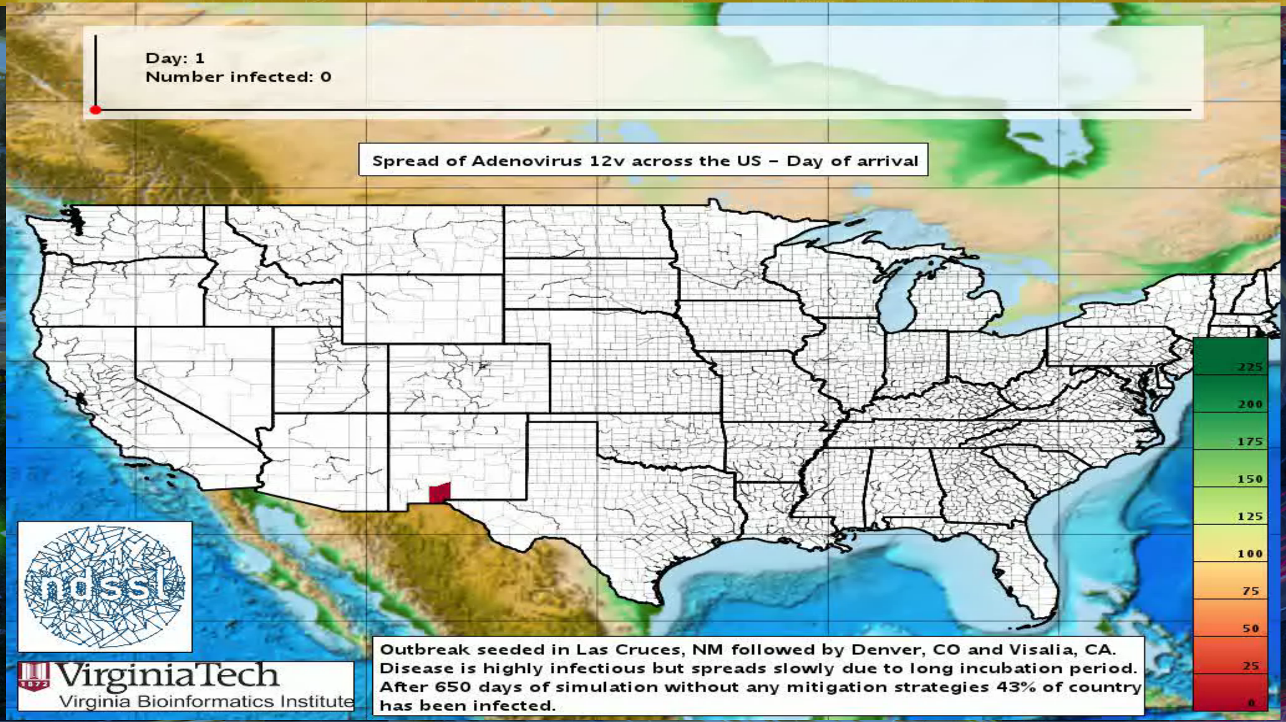
**First data driven global synthetic populations and proximity networks**

# Step 2:
# Use HPC-oriented causal models for interpolation, consistency & extrapolation

VirginiaTech
Biocomplexity Institute

# Extreme-scale (spatial, temporal, network, machines, data) causal multi-scale models*



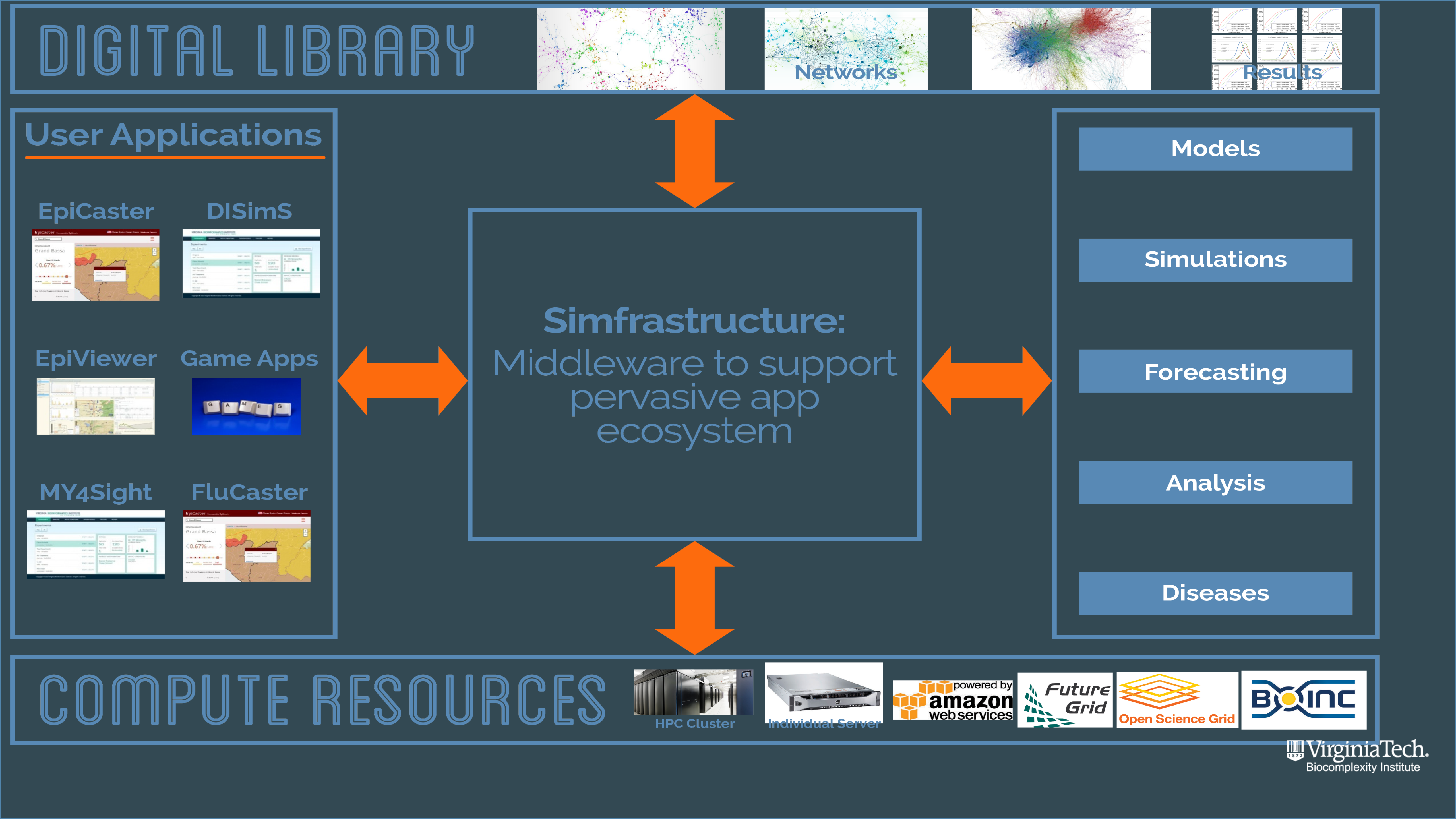

**National simulations of epidemic dynamics over social-proximity networks:**
**300 Million nodes, 15 Billion edges, 1 Trillion interactions,**
**750,000 Cores 10 seconds**

* SC, IPDPS, CCCGRID, WSC,

VirginiaTech
Biocomplexity Institute

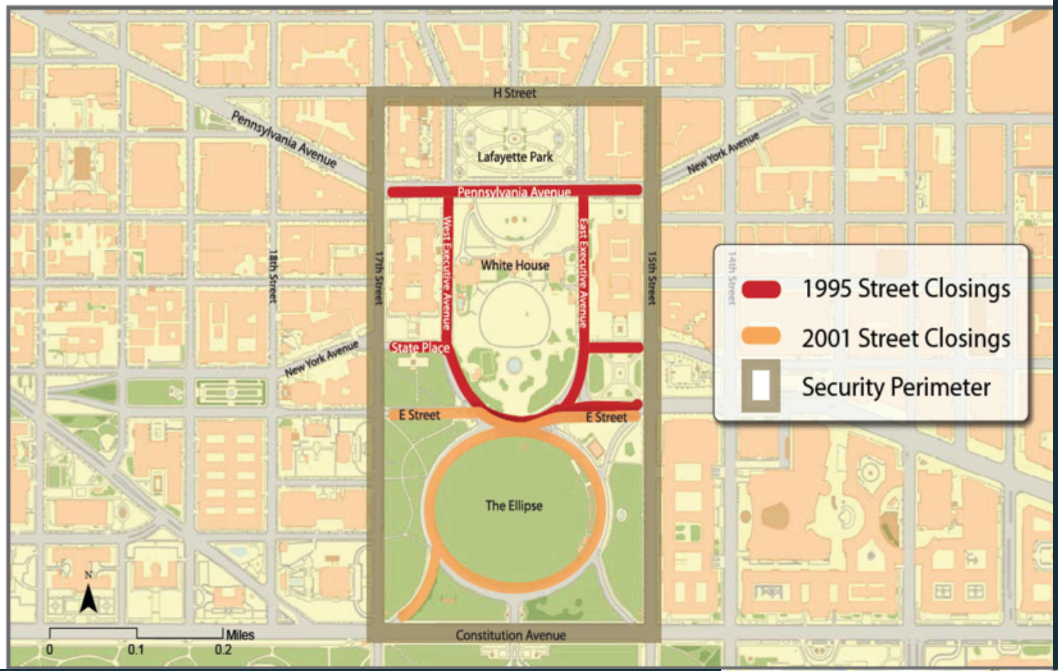

VirginiaTech
Biocomplexity Institute

**Step 3:**
Develop advanced technologies that support *pervasive real-time decision making*

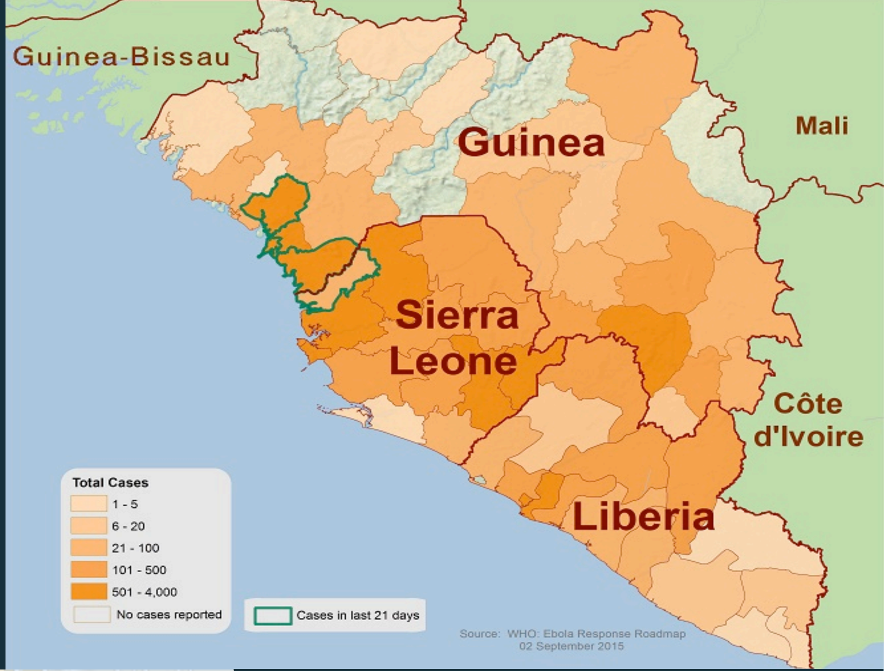

DIGITAL LIBRARY
Networks
Results
User Applications
EpiCaster
DISimS
EpiViewer
Game Apps
MY4Sight
FluCaster
Simfrastructure:
Middleware to support pervasive app ecosystem
Models
Simulations
Forecasting
Analysis
Diseases
COMPUTE RESOURCES
HPC Cluster
Individual Server
powered by amazon web services
Future Grid
Open Science Grid
BOINC
VirginiaTech
Biocomplexity Institute

# Examples of a Large Program Over 25 Years

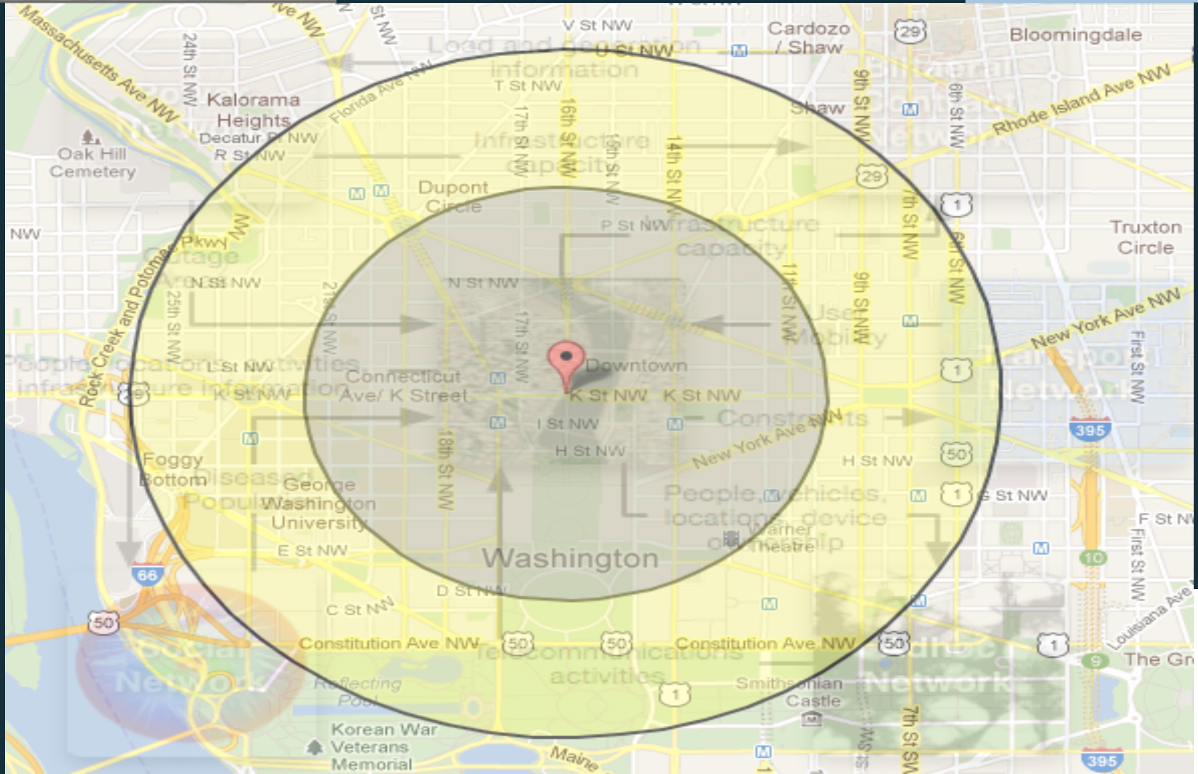
Biocomplexity Institute




# National Planning Scenario 1

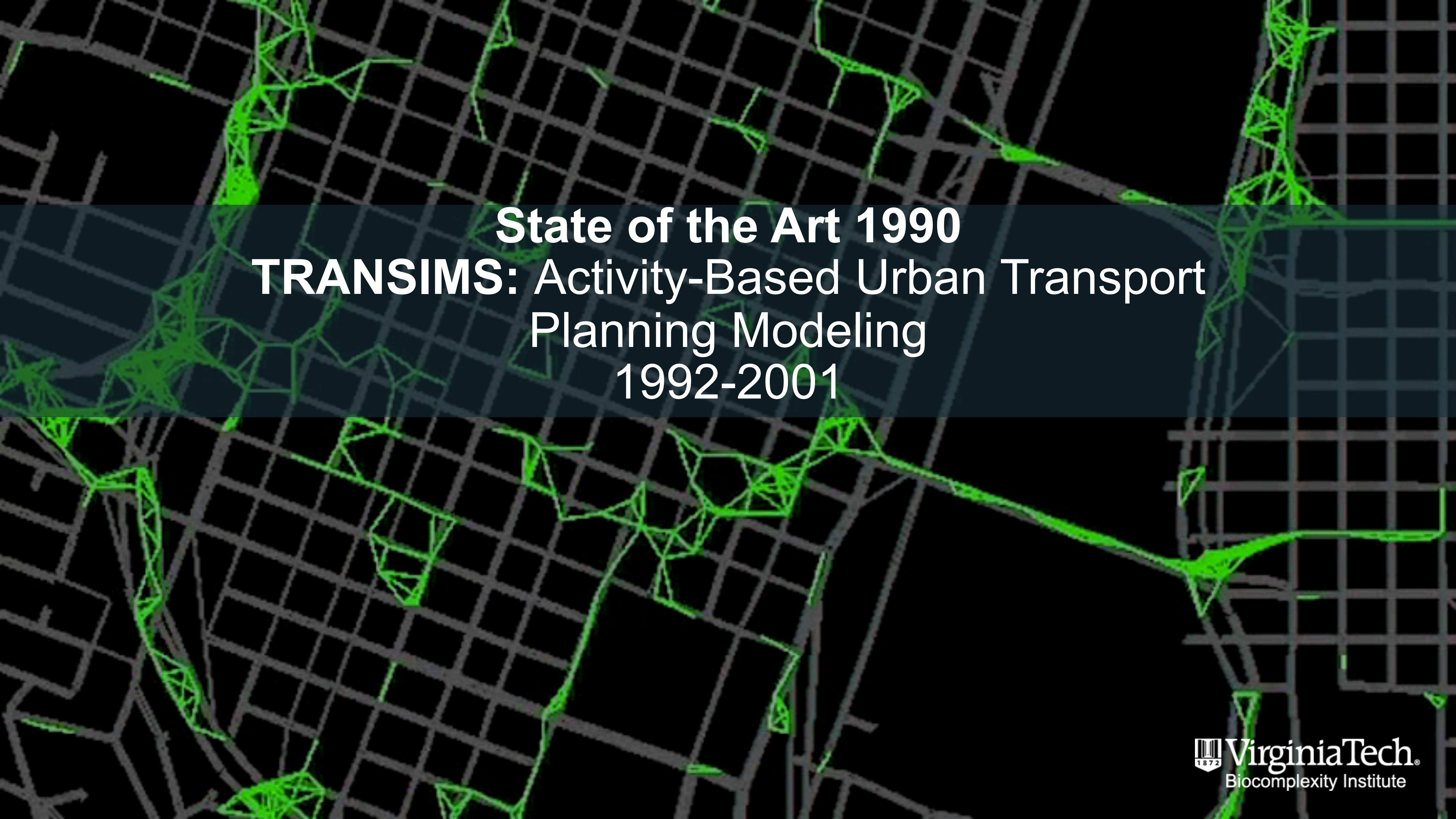



# National Capitol Planning Commission




# Real-time Ebola Outbreak Response

VirginiaTech
Biocomplexity Institute



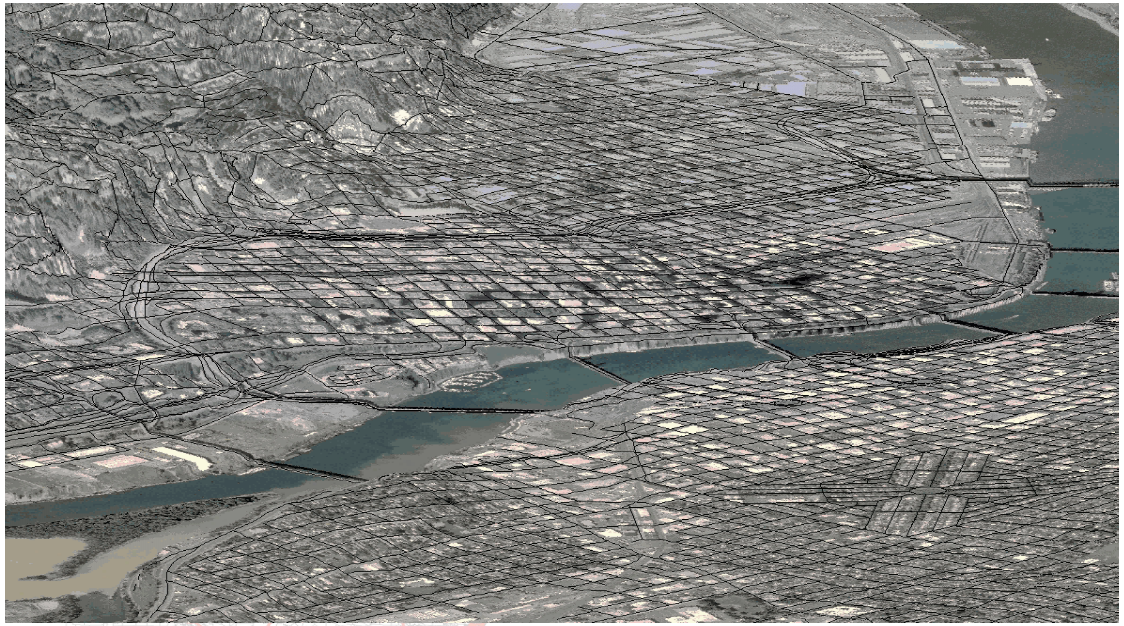
# State of the Art 1990

## **TRANSIMS:** Activity-Based Urban Transport Planning Modeling

## 1992-2001

# TRANSIMS: Activity based urban transport planning environment

- TRANSIMS: (1991-2001)
  - Set up to address important national problems
  - Assessing traffic flows during congestion
  - Predicting household behavior when the infrastructure changes
  - Assessing the social justice issues in proposed infrastructure changes
  - Predicting the environmental impact



- Requirements set by:
  - Government (DOT), Regional planning organizations
  - University researchers, Transportation consultants
- Technology developed, and demonstrated in close collaboration with Metropolitan planning offices
  - Case studies with Albuquerque, Dallas, Portland MPOs
  - Open source system managed by Argonne National Laboratory



VirginiaTech
Biocomplexity Institute

# White House Area Transportation Study (National Capitol Planning Commission)

- Purpose: Examine traffic problems around White House due to street closures
  - Objective: Mitigate the impacts of closures for travelers in the downtown area
  - Study by USDOT: 2005-2010
  - AECOM
- Metrics examined include:
  - Travelers experiences
  - Travelers interactions with each other
  - Amount of time to travel
  - Cost to travel
- Compare solution alternatives:

Times & cost between scenarios

- New tunnels construction – Infrastructure
- Open or close streets
- Transit improvements – bus
- Traffic management and traffic operations

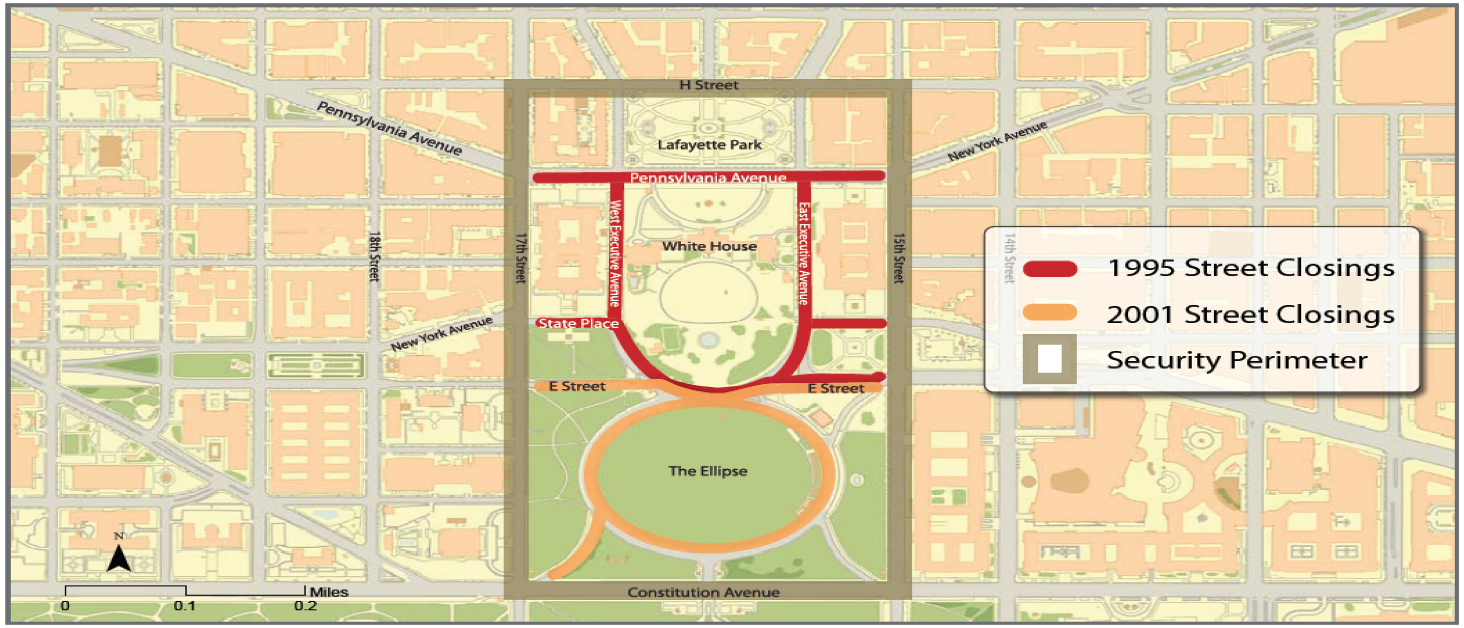



*Figure 1 – White House Area Street Closures*

http://www.ncpc.gov/ncpc/Main%28T2%29/Planning%28Tr2%29/PlanningStudies%28Tr3%29/Transportation.html

VirginiaTech Biocomplexity Institute

**CNIMS:** Comprehensive National Incident Management System (2005-present)

## National Planning Scenario 1: Studying the social and behavioral effects of large-scale WMD incidents.

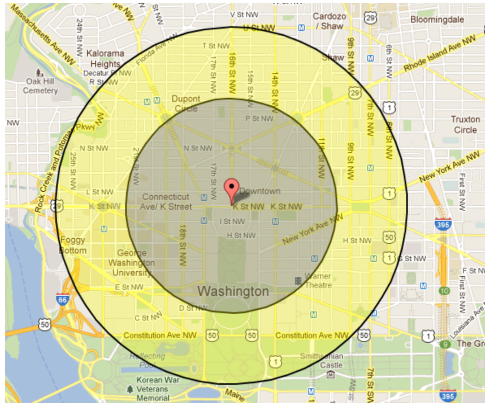
Deploying pervasive computing applications that support analysts and responders in their efforts to combat and respond to large human and naturally initiated crises.

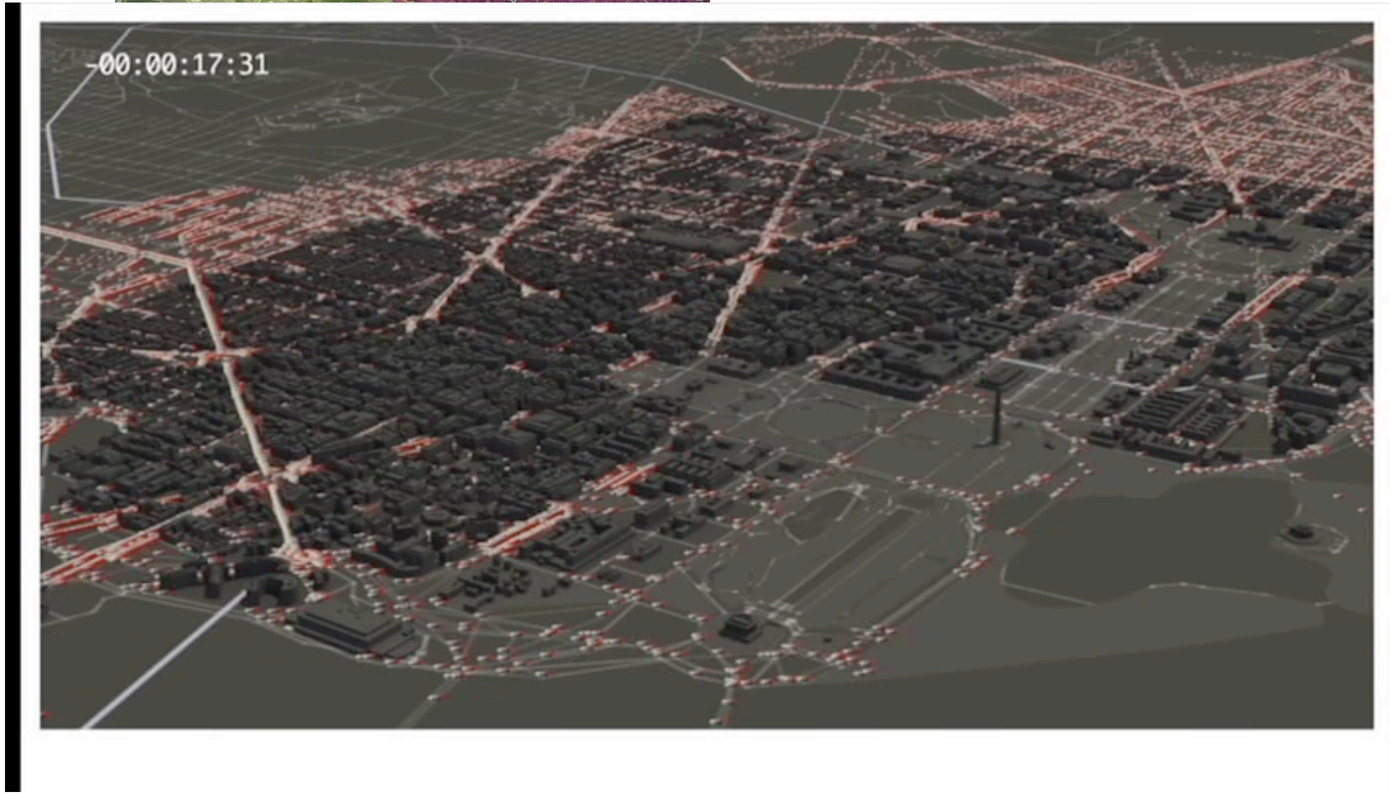
VirginiaTech
Biocomplexity Institute

# National planning scenario 1

## Hypothetical Scenario:

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

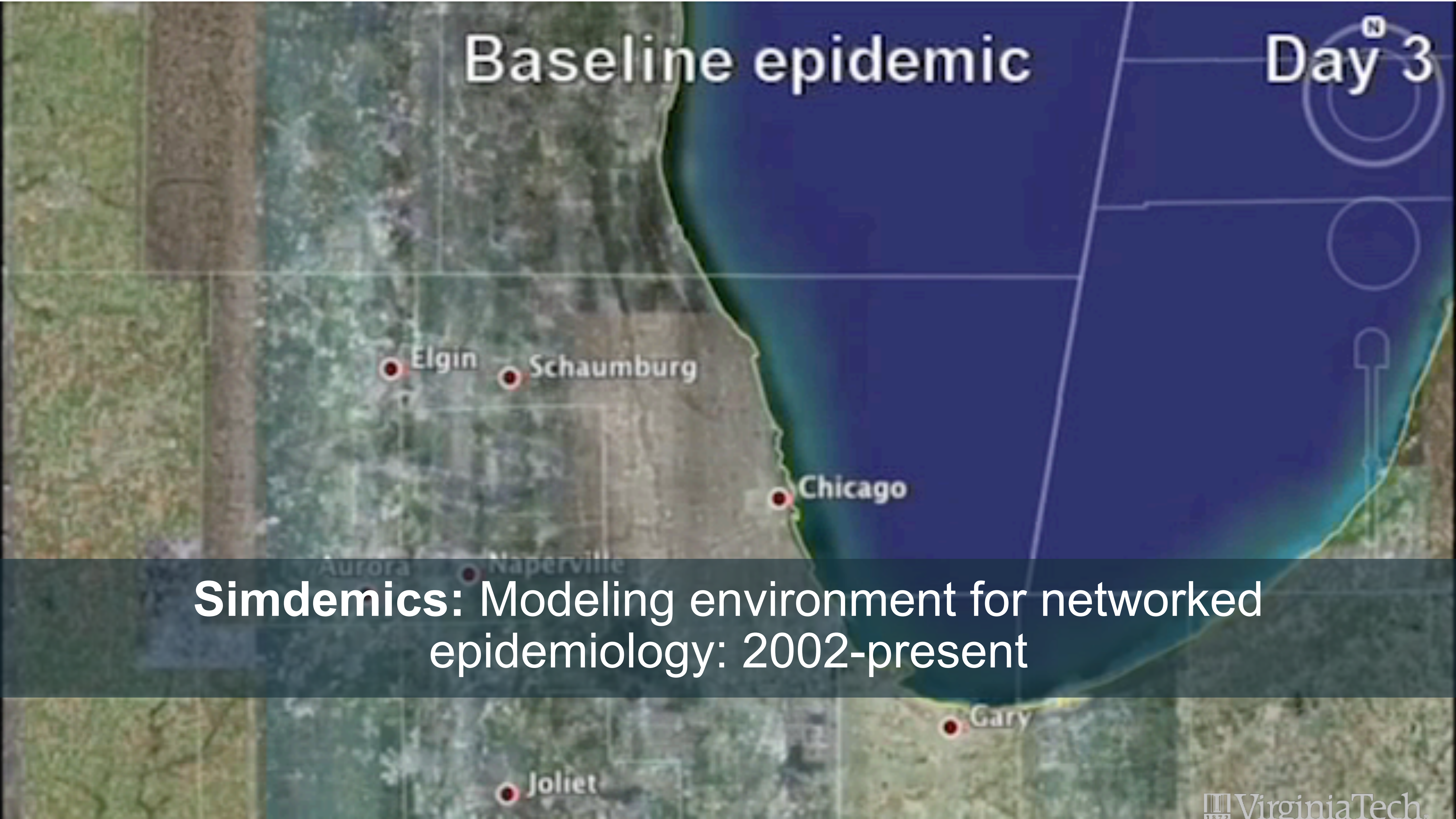




## Traditional Focus

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

# Two Illustrative conclusions

- Even a *partially restored* communication system has disproportionately positive impact on the overall behavior,
  - Leads to fewer casualties, better health outcomes & reduction in anxiety.
  - Policy question: how do we build a self forming communication network using components that belong to diverse stakeholders with few strategic nodes
- The power network suffers a huge loss and large portions of the network will unlikely be operational for at least  year or two.
  - IF the protection devices work as planned, NO significant cascading failures beyond a small area
  - Important policy implications: how the city and its surroundings will be reconstituted.

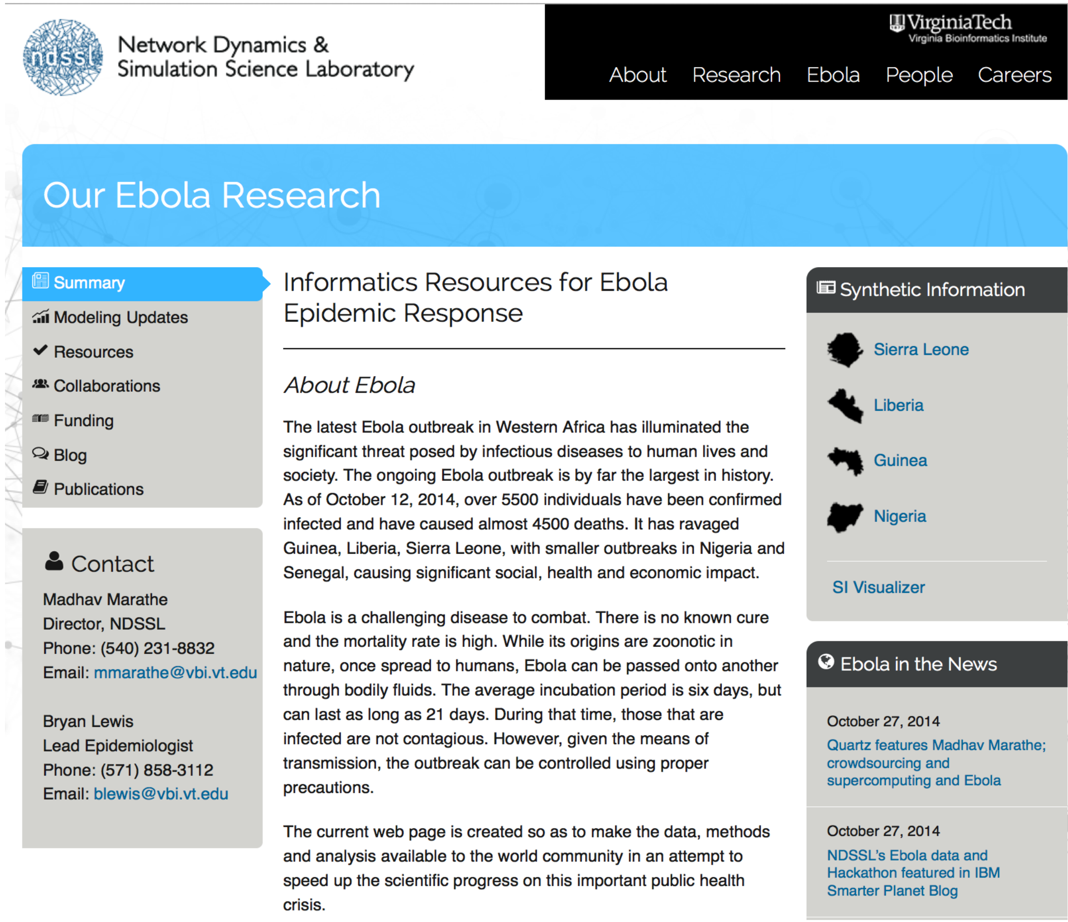



**Simdemics:** Modeling environment for networked epidemiology: 2002-present

# Informing policy during Ebola outbreak

**URL: https://www.vbi.vt.edu/ndssl/featured-projects/ebola**

- *Estimating* basic epidemiological parameters for the outbreak
- *Forecasting* the ongoing epidemic with & without control
- *Assessing* the threat of imported cases in the US causing secondary infections
- *Efficiently allocating* potential pharmaceutical treatments
- *Locating Emergency treatment centers* and assessing their impact
- *Estimating* the need for supplies such as personal protective equipment
- Analyzing social media for public mood & sentiments
- Assessing the potential spread of Ebola to Latin American countries




VirginiaTech
Biocomplexity Institute

# Example of work by other groups

- Galvani and her group at Yale (CIDMA)
  - A combination of case isolation and hygienic burial practices could reduce transmission to an extent that disease elimination became a realistic goal (Pandey et al. Science 2014);
  - role of non-survivors is significant (Yamin et al. Ann. of Int. med, 2015)
  - Studied utility of ring vaccinations; mobile app for collecting ground data
- Vespignani and his colleagues
  - Developed early agent based models and used it for forecasting and assessing interventions (PLOS Current Outbreaks, 2014 & Lancet'15)
  - Recently conducted an Ebola challenge to understand our ability to forecast in controlled settings
- Donnelly, Dye and the WHO team
  - Analyzed case data to estimate epidemiological parameters (NEJM'14)

# Other examples of supporting public policies

- White House Homeland Security Council  for smallpox mass vaccination
  - Do we need mass vaccination?  How do we protect critical workers? [*Nature'04*]
- Federal Influenza Plan: OHS & DHHS --  NIH MIDAS project
  - TLC: Targeted Layered Containment, Importance of Social Distancing [*PNAS'08*]
- Pandemic Planning for Military Preparedness: DoD
  - Impact of layered interventions for force projection: Public versus military health epidemiology [WSC'09,IHI'12]
- Zika Response (Work done by CIDMA)
  - Evaluate the effectiveness of pregnancy delay policies on incidence and prevalence of prenatal zika infection in Colombia: Mass delay for 9 months or more is the most effective strategy
  - ***Cost-effectiveness decision tool***:  Evaluate the cost-effectiveness of expenditures towards Zika control intervention. The maximum expenditure that a country should invest to avert 1 Zika infection depends on the how wealthy the country is (GDP per capita) and how many women are at risk (determined by the birth rate).

*Ndeffo-Mbah ML et al. 2016) Mitigating prenatal zika virus infections in the Americas. Ann Inter Med 165(8):551-559. Alfaro-Murillo JA, et al. (2016) A Cost-Effectiveness Tool for Informing Policies on Zika Virus Control. PLoS Negl Trop Dis*

# Summary

# What did we learn: a short list

- Delicate balance between rigorous science and real-time response
  - Modified Tukey's theorem: "An approximate answer to the right question within a given time is worth a great deal more than a precise answer to the right (wrong) question much later"
- Data sharing, access, and making it computer readable is a  challenge
  - Nothing new, but becomes even more critical during real-time decision making
- Embedded interactions with policy makers
  - Working closely with policy makers as the questions are being discussed and getting their feedback as models are formulated and analysis is being done is often critical: e.g. H5N1 TLC study, H1N1 response, Ebola response, ….
- V&V remains a vexing question: traditional approaches in physical sciences are not always appropriate: not just predictive validity
  - Models are used for variety of purposes: reasoning, forecasting, counterfactuals, ….
- Derive requirements for refining and improving models during studies
  - Tools were developed in this manner; allows us to learn from past experience and support future studies.

Thank you

Questions:

Madhav Marathe, 540 808 3292 (mmarathe@vt.edu)
https://www.bi.vt.edu/ndssl