turning knowledge into practice

Modeling and Simulation in Social Sciences

Georgiy Bobashev, Ph.D.

bobashev@rti.org

Dec. 12th, 2016

RTI INTERNATIONAL

3040 Cornwallis Road ■ P.O. Box 12194 ■ Research Triangle Park, North Carolina, USA 27709

Phone 919-541-6167 Fax 919-541-6722 e-mail bobashev@rti.org

RTI International

Headquartered in North Carolina, with offices in the US and around the world

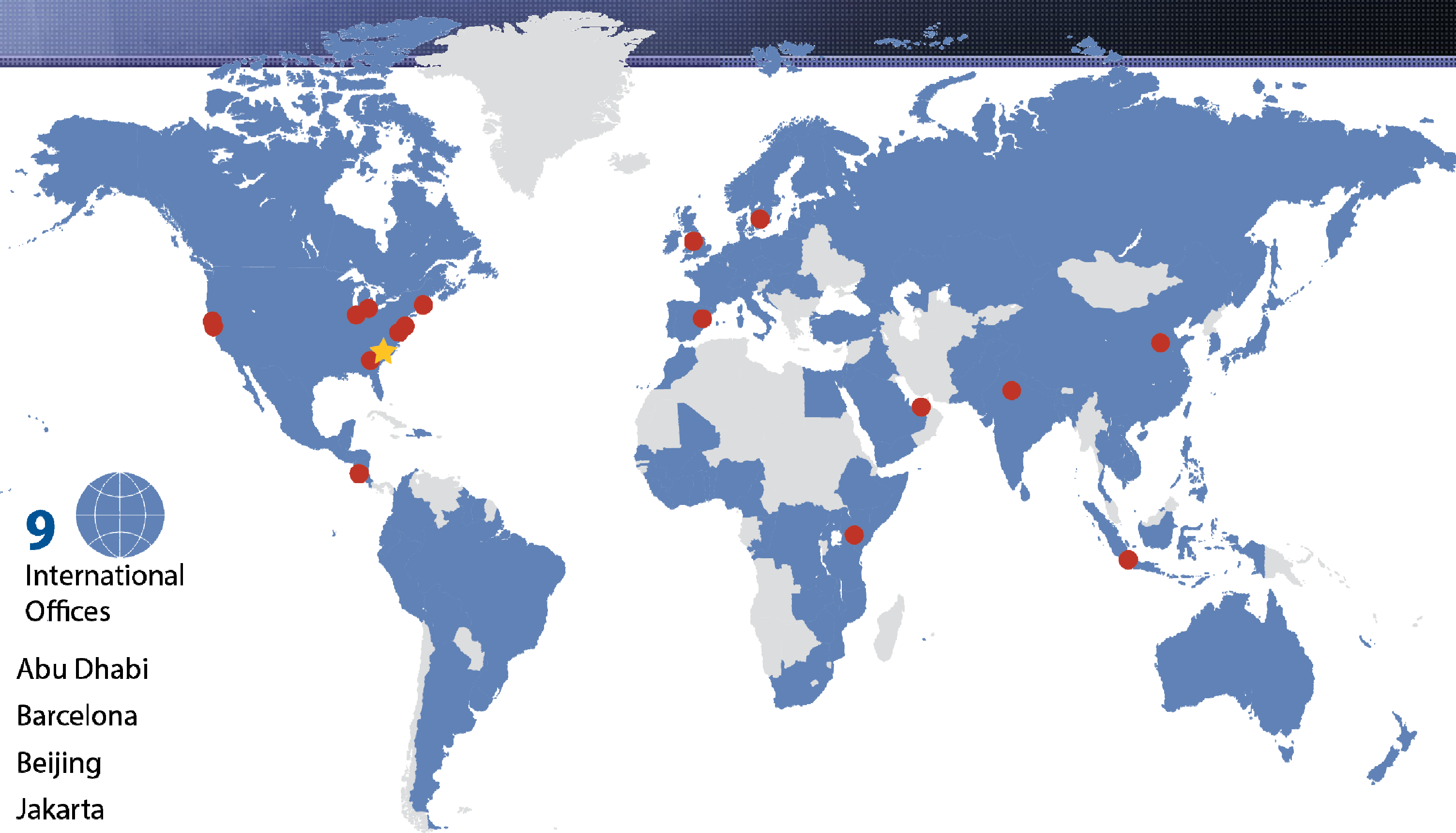
RTI is an independent, nonprofit institute that provides research, development, and technical services to government and commercial clients worldwide. Our mission is to improve the human condition by turning knowledge into practice.

What Is RTI International

RTI is an independent, nonprofit institute that provides research, development, and technical services to government and commercial clients worldwide.

Our mission is to improve the human condition by turning knowledge into practice.

Global Presence – Office Locations



9 U.S. Offices

RTP, NC
Washington, DC
Rockville, MD
Atlanta, GA
Chicago, IL
Waltham, MA
Ann Arbor, MI
Berkeley, CA
San Francisco, CA

9 International Offices

Abu Dhabi
Barcelona
Beijing
Jakarta
Ljungskile
Nairobi
New Delhi
Manchester
San Salvador

Global Presence – Workforce

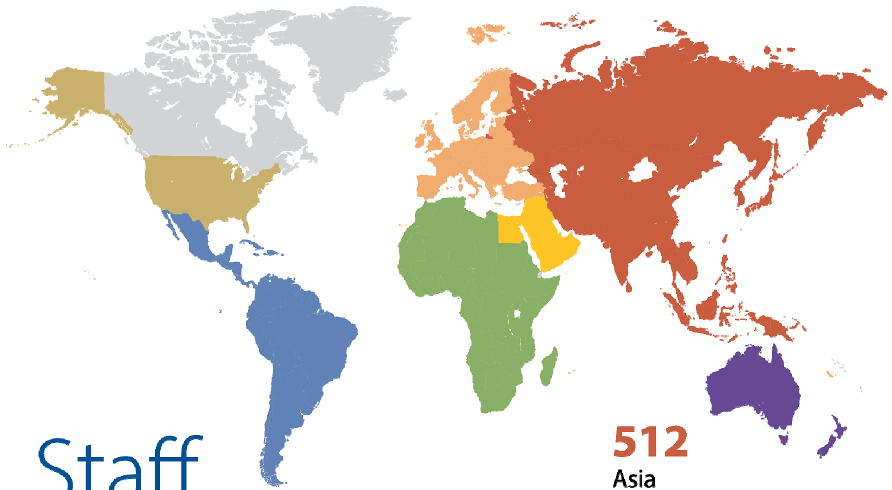
29
Fellows

250
Degree Fields

90
Languages

105
Nationalities

4,150 Staff Members
Worldwide



Staff By Region

2,874
United States

45
Latin America and the Caribbean

594
Africa

512
Asia

61
Europe

62
Middle East

2
Australia

Course Outline

Day1. Overview of the course and Agent-Based Models

Model types by modeling objective and modeling techniques

Day2. NetLogo laboratory. Examples of ABM. Build your first model

Day3. ODD protocol. Develop an ABM following ODD

Day4. Experiments using ABMs

Day5. Discussions of class homework projects

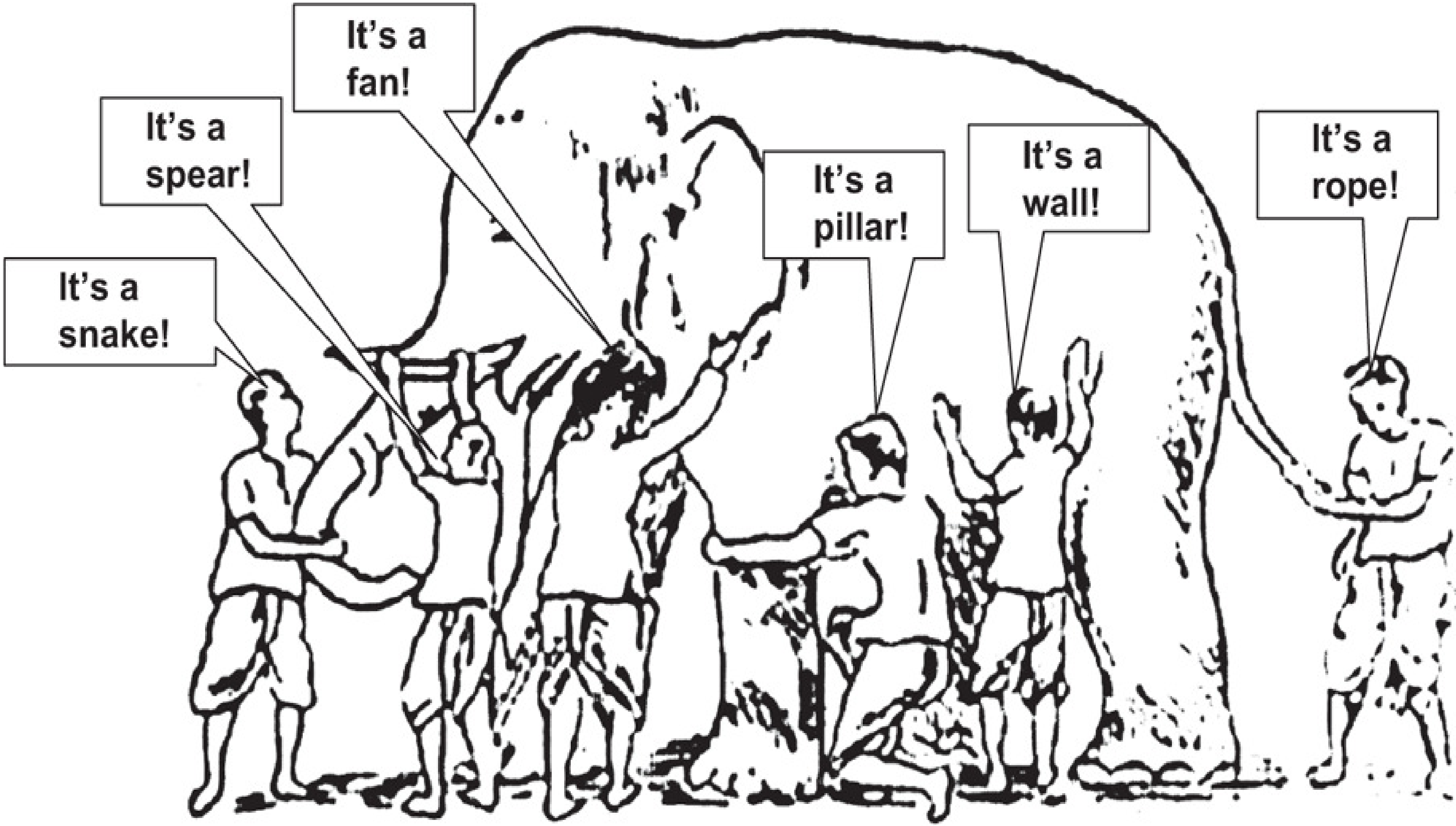
Examples of Models

$$dS/dt = m - \beta SI - \delta S$$

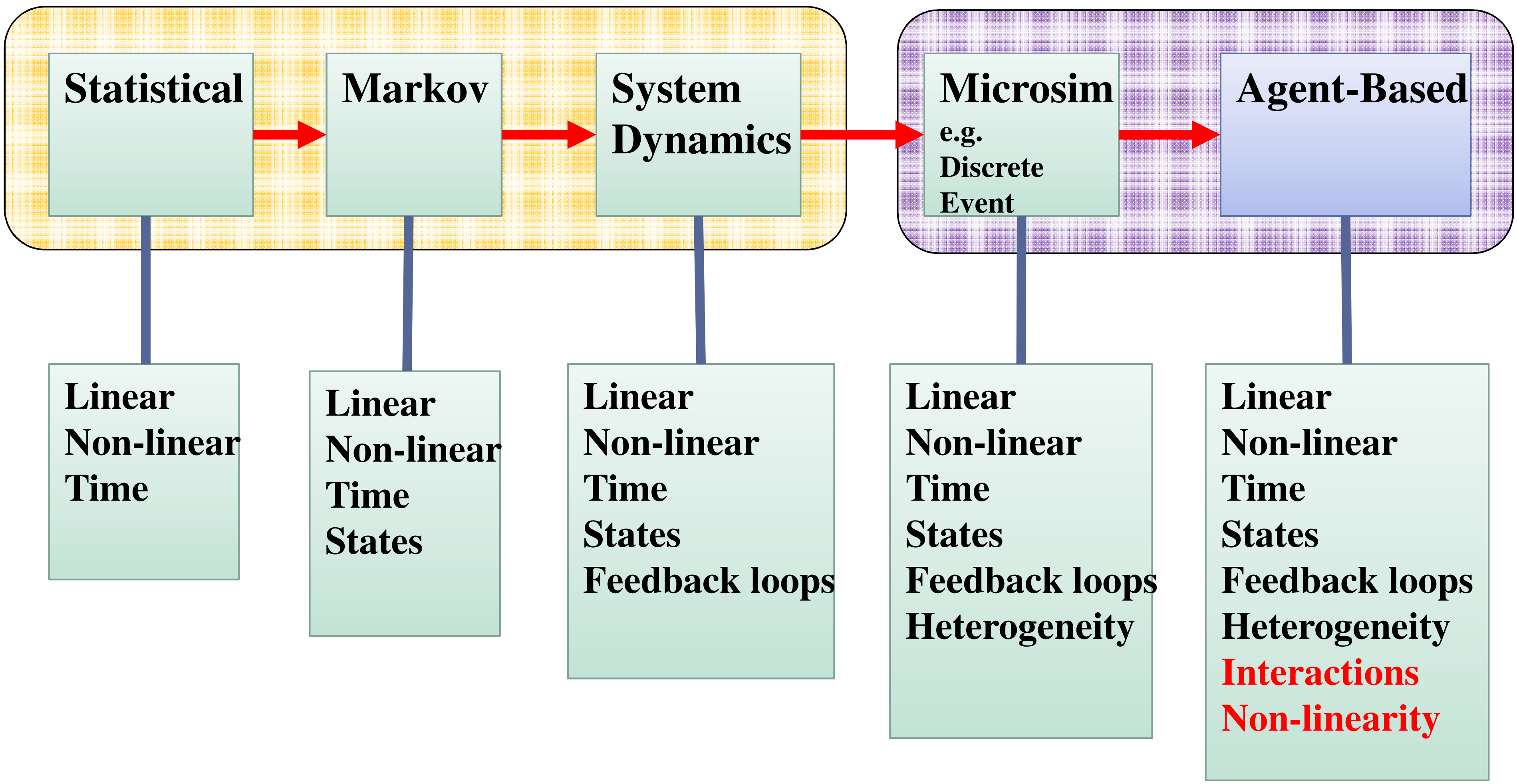
$$dE/dt = \beta SI - \lambda E - \delta E$$

$$dI/dt = \lambda E - \gamma I - \delta I$$

Reductionist and Systems Approach



Hierarchy of Simulation Models (non-consistent terminology)



Global vs. Local Rules

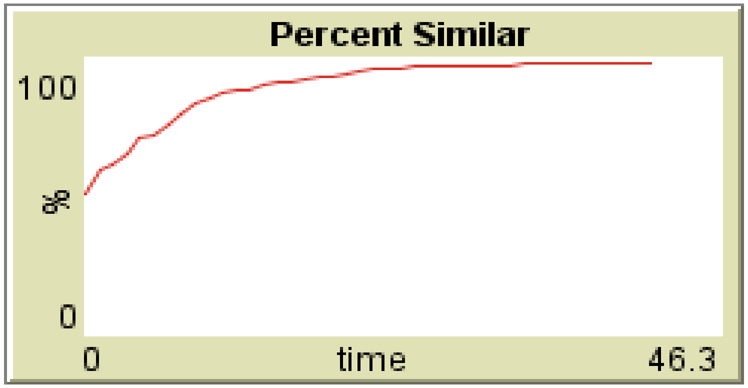
Classic orchestra | **Jazz band**

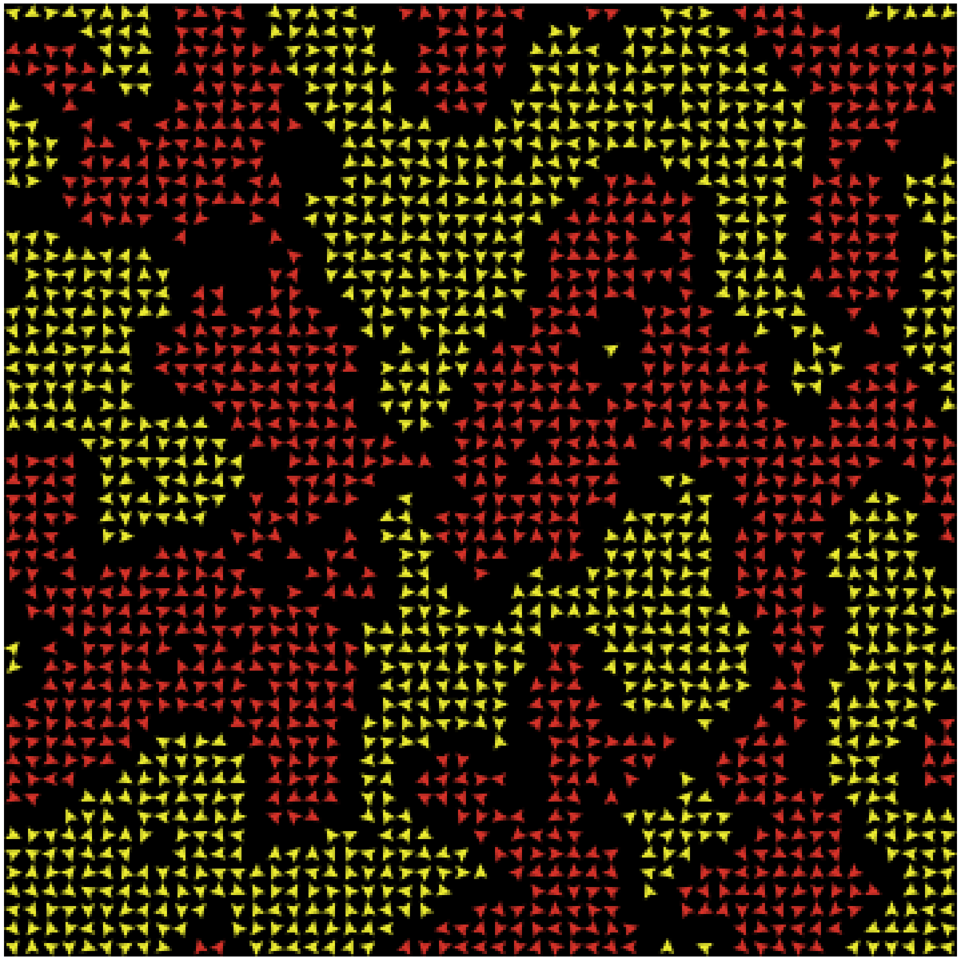
Examples of Theoretical ABMs

Easy start with NetLogo examples

http://ccl.northwestern.edu/netlogo/

Agent-based Model of Segregation Shelling (1971)

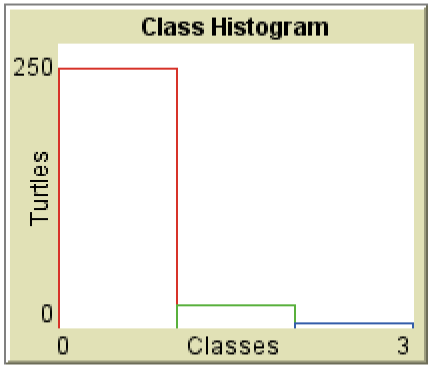


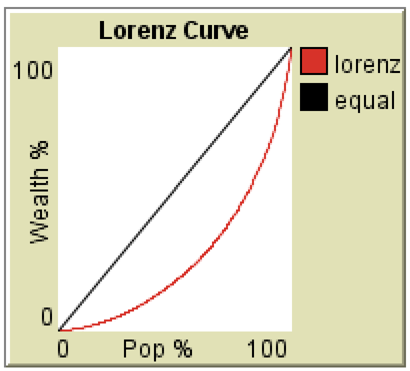


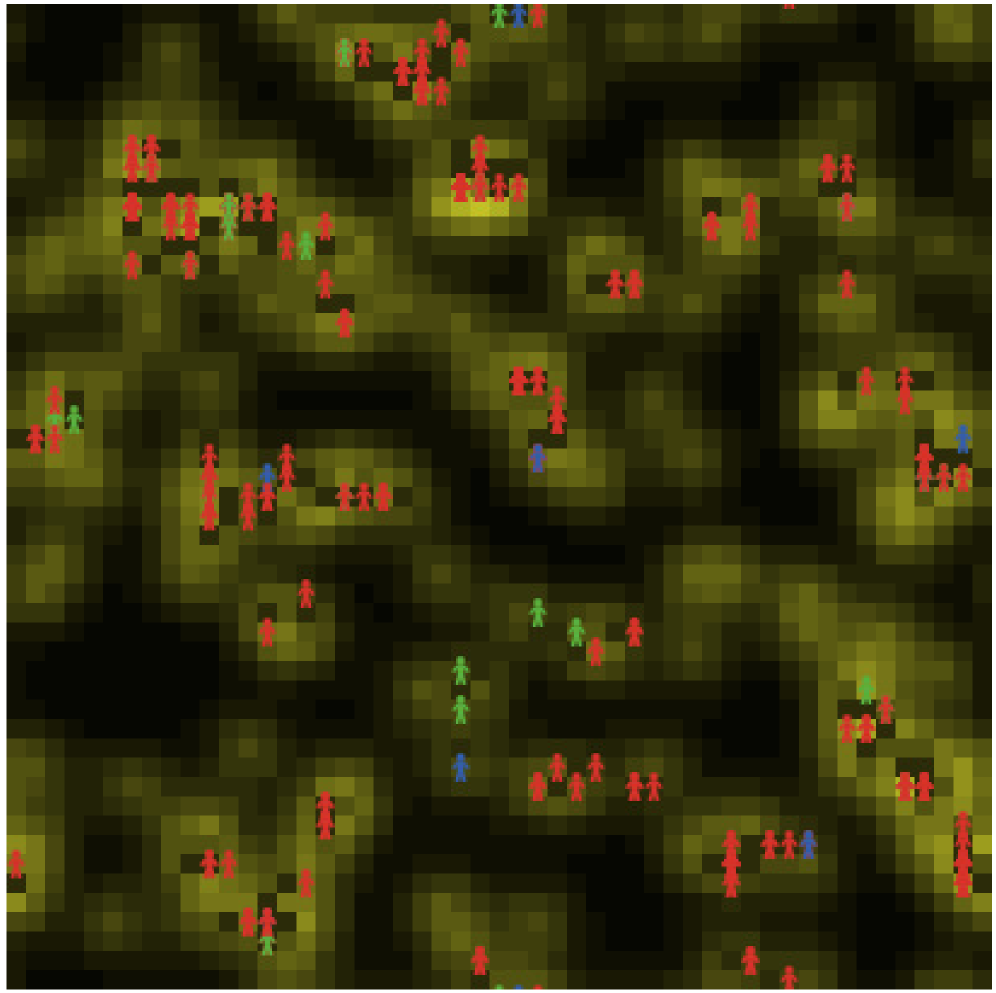
Demand 47%, result 98%

Demand 77%, no solution

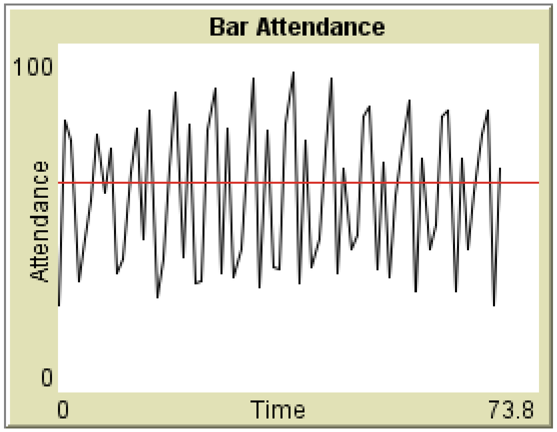
Agent-based Model of Wealth Distribution (simple rules lead to a Pareto distribution)

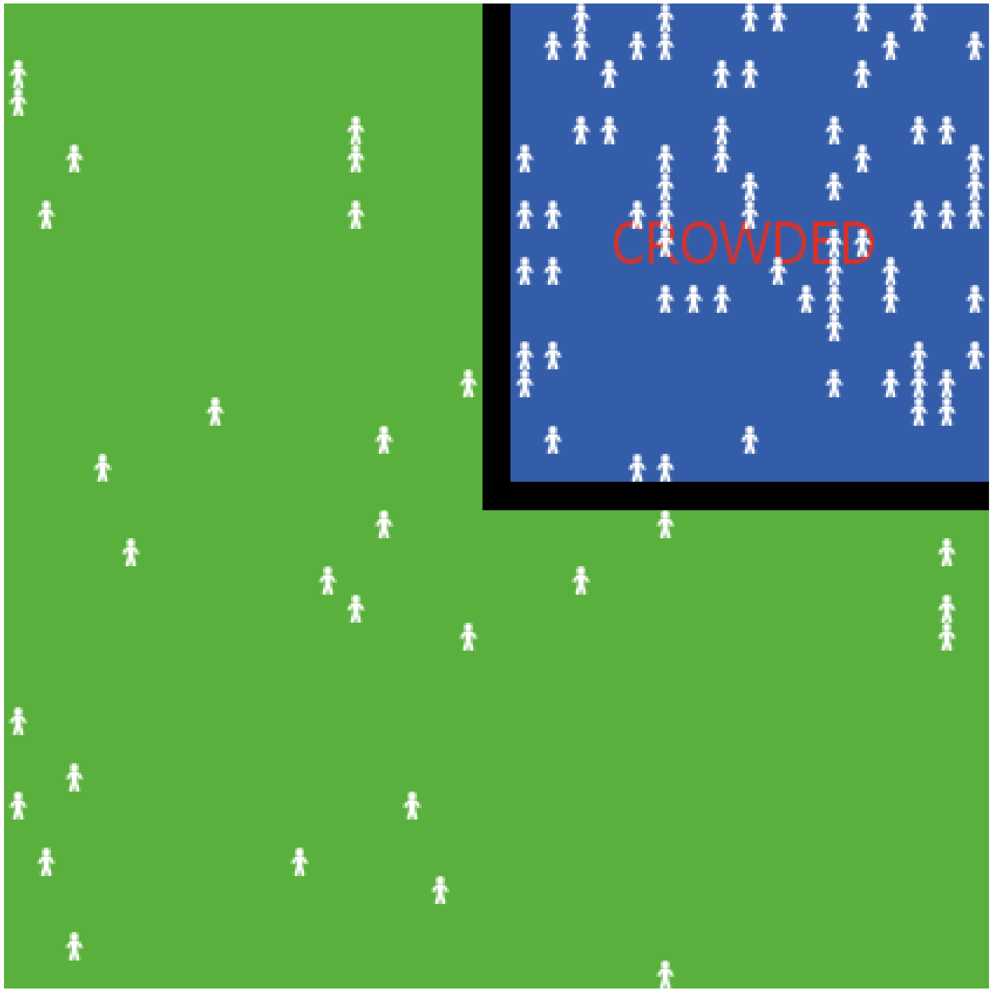






Agent-based Model of El Farol Bar

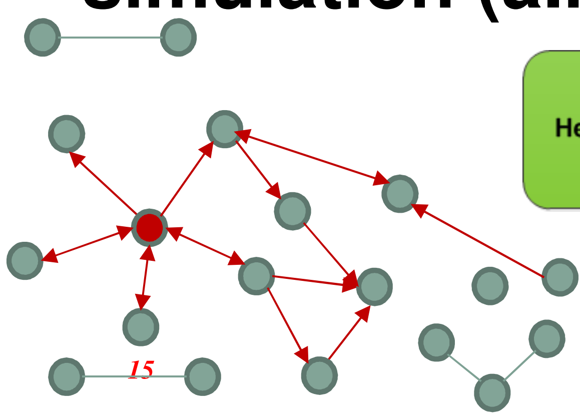


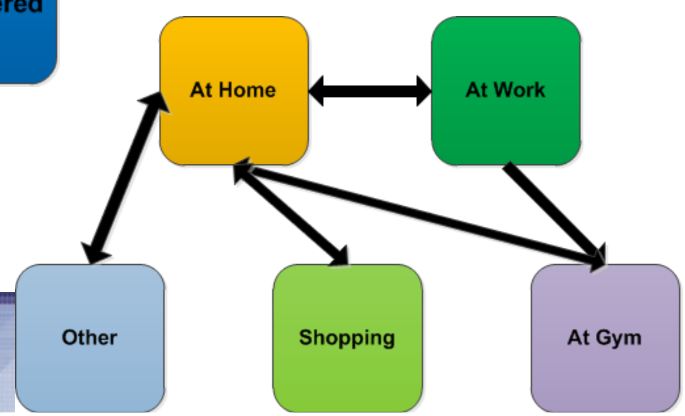


Agents in Agent-Based Models

- **Agents are computer objects that are defined by states, transitions between the states and rules of interaction between each other and environment**

- **If agents are passive the model is just a micro-simulation (a.k.a. discrete event)**

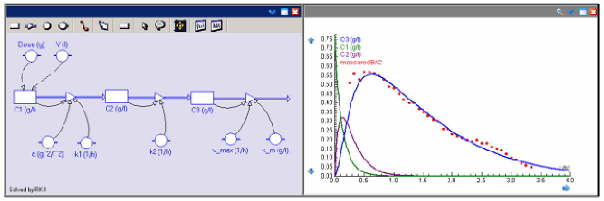




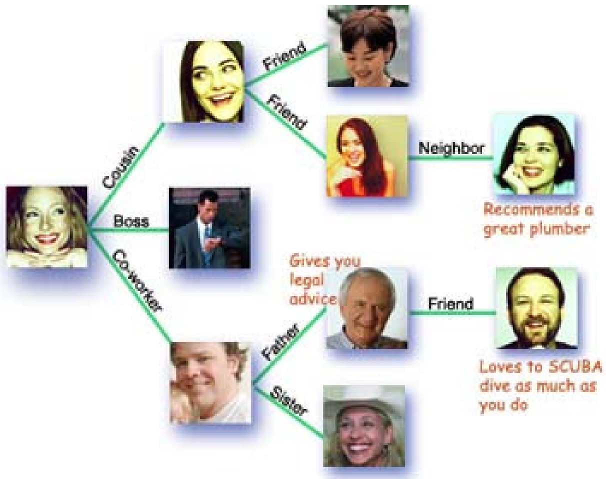
Agent-based Model

- Agents can make decisions based on rules
- Agents can be adaptive
- Agents can have several state charts and internal dynamics

Agent



- Agents can have social networks
- Networks can have life of their own, (i.e. be agents)



Agent-based Models

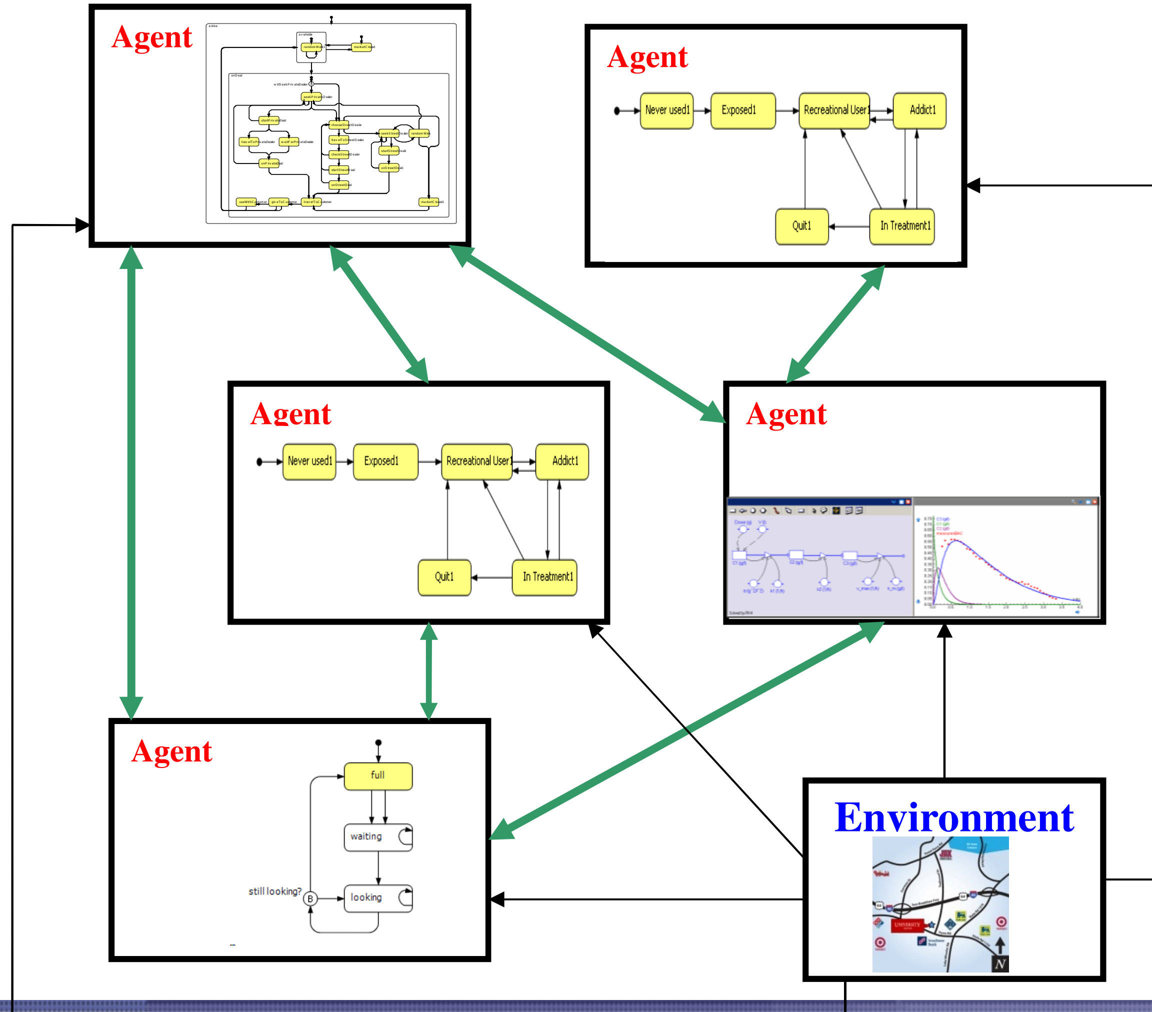
Social Networks

Behavior

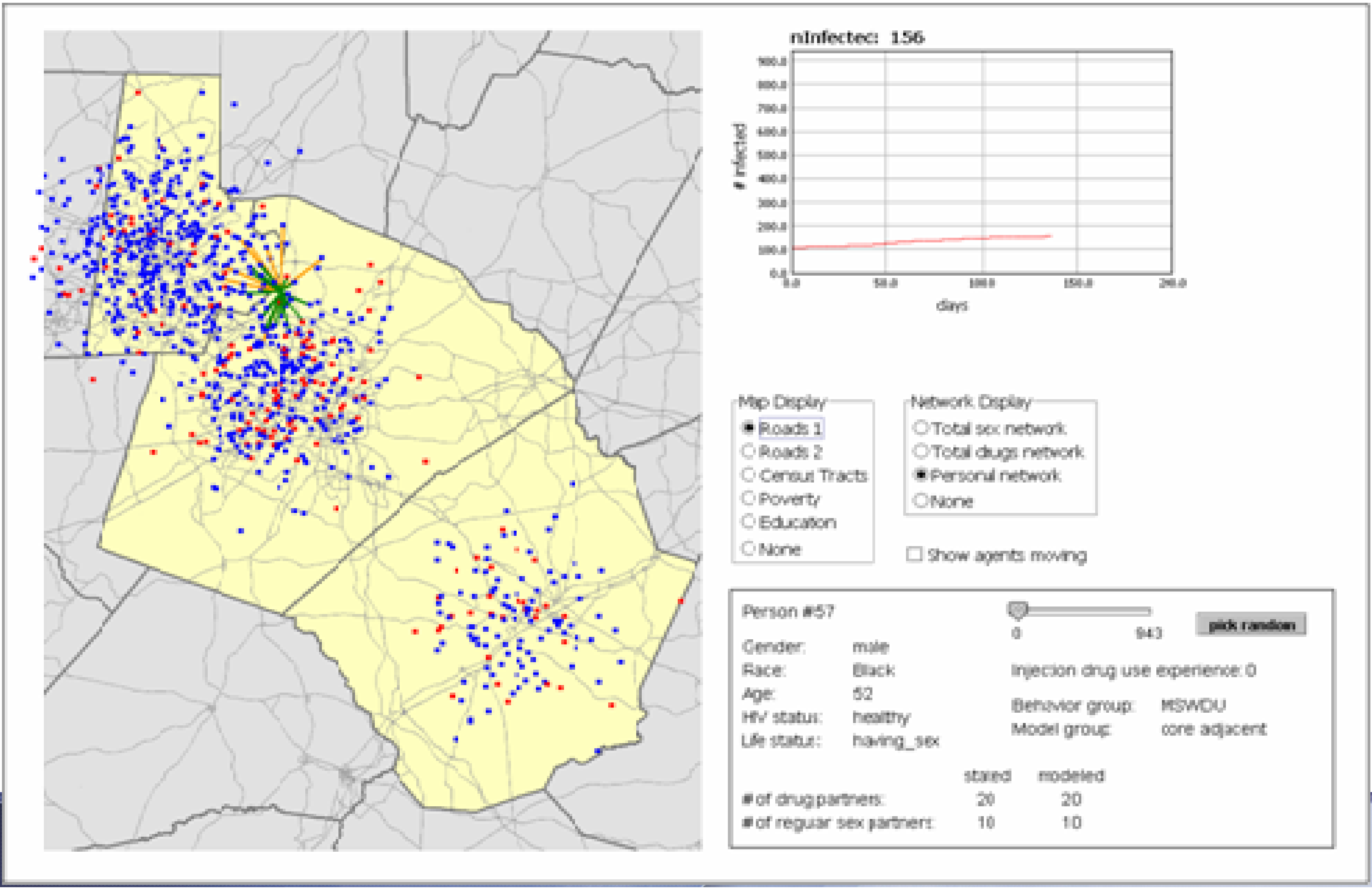
Interaction

Self-organization

Environment can be added as a system dynamics model



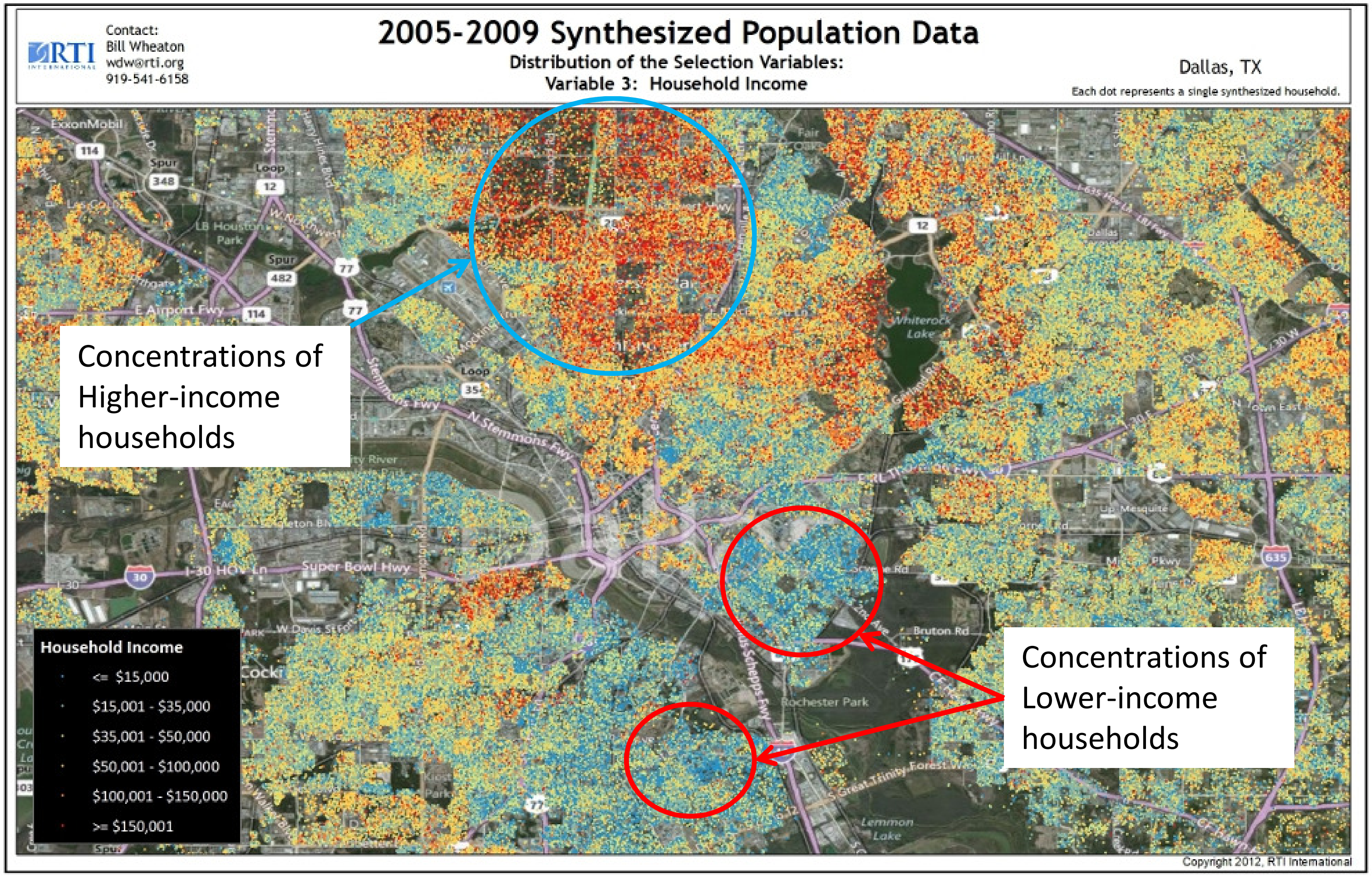
GIS and Synthetic Population



Practical Applications of ABMs

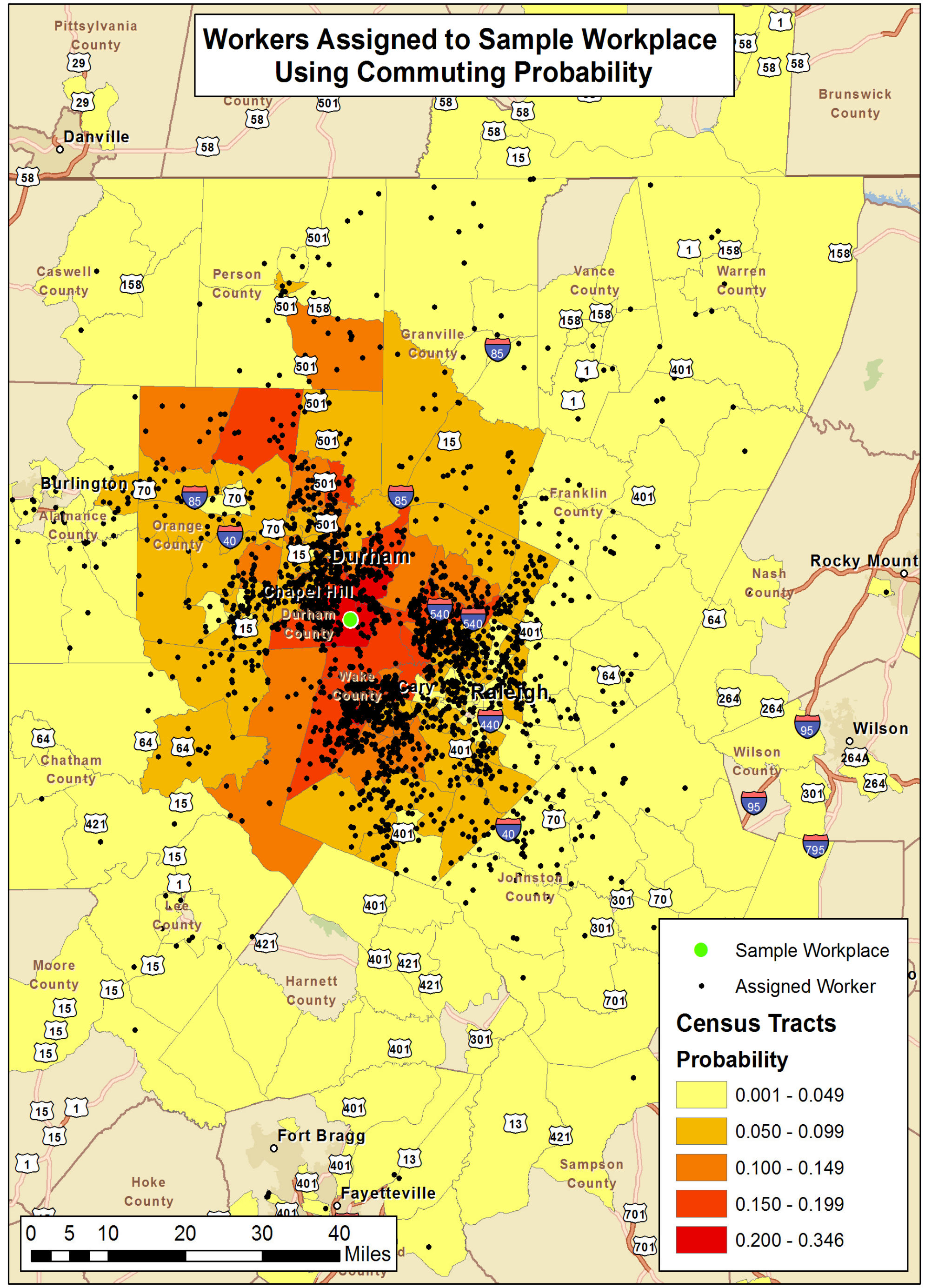
- Infectious diseases
- Chronic diseases
- Social Organizations
- Security/Criminal Justice
- Intervention Optimization
- Risk Assessment
- Cluster Identification/ Community Detection

Synthetic Populations
Dallas, TX Household Income



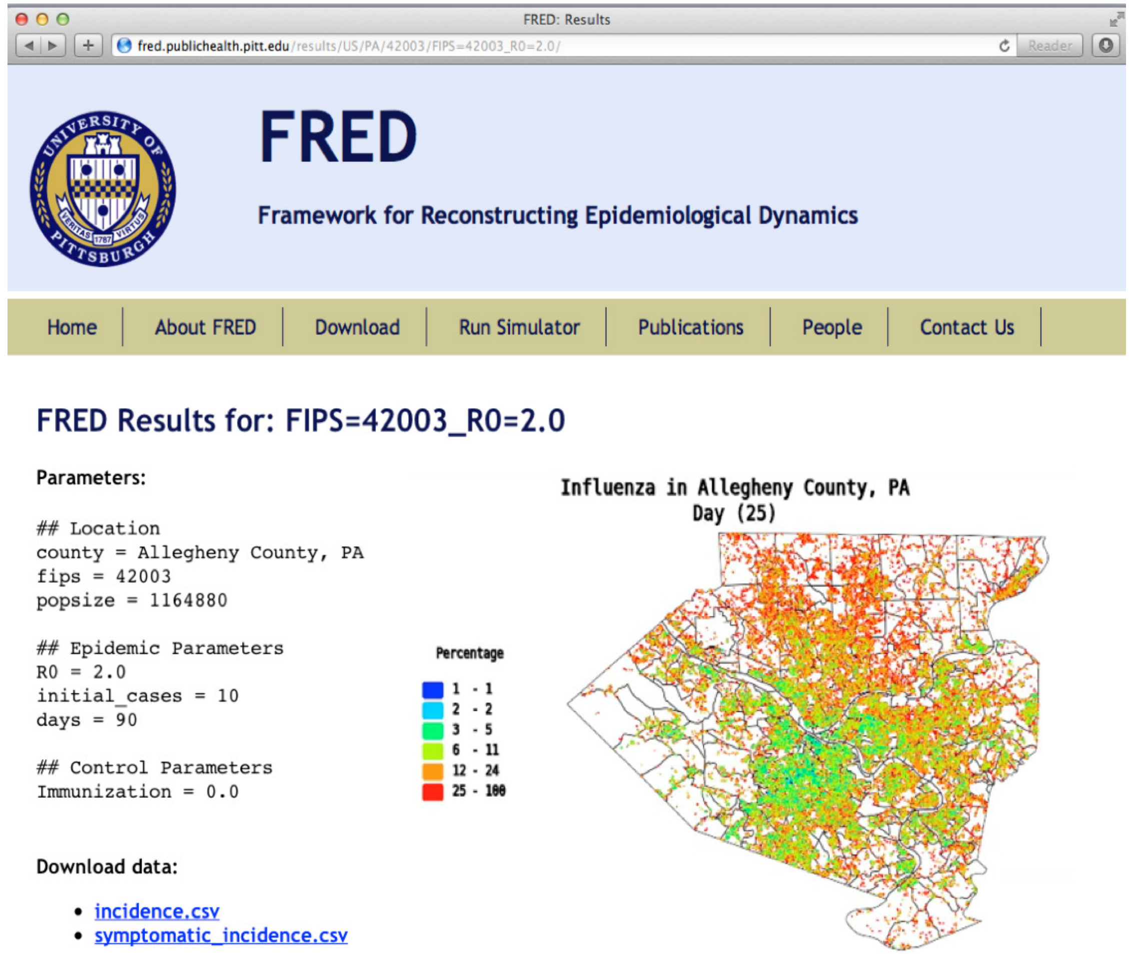
Schools and Workplaces: Encoded Social Networks

- U.S.:
  - Students assigned to schools based on location and capacity by grade
  - Workers assigned to workplaces based on commuting patterns, locations and sizes of workplaces
- International:
  - Given counts of schools by size or business by size
  - Estimate locations based on population distribution
  - Assign agents as in U.S.



Example: FRED (Framework for Reconstructing Epidemic Dynamics)

- The MIDAS ABM community is moving towards models that are standardized and can be set up and run quickly
- Any county or arbitrary set of counties in the U.S.

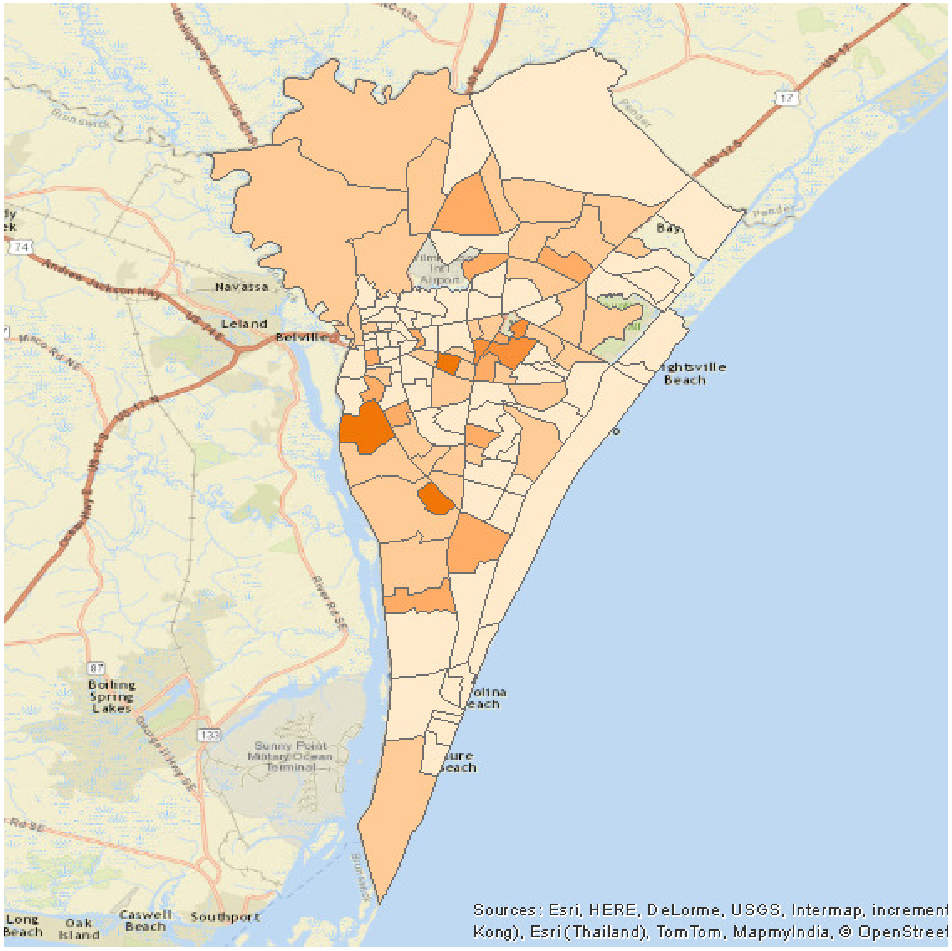


RTI INTERNATIONAL

Community detection and evacuation optimization

Tammy Armstrong, Greg Dimock, John Gauf,
Josh Knippel, Vanessa Moglin, Alice Toms

Community Detection - Results



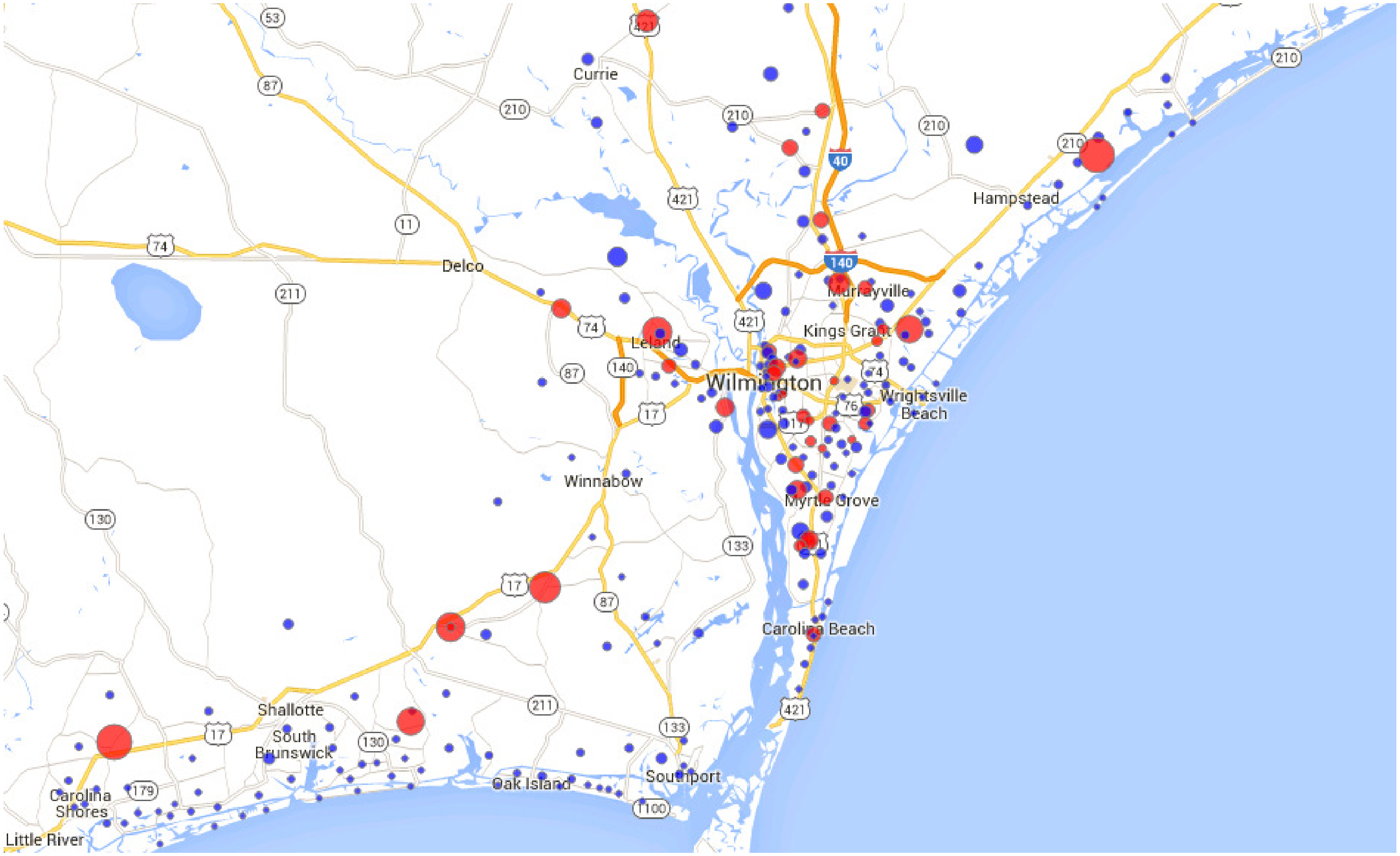
Legend:

Percent of block group

- 0.0 - 5.2%
- 5.2 - 10.4%
- 10.4 - 15.6%
- 15.6 - 20.8%
- 20.8 - 26.0%

Lower Income Immigrants (5%)

Shelter Optimization - Results

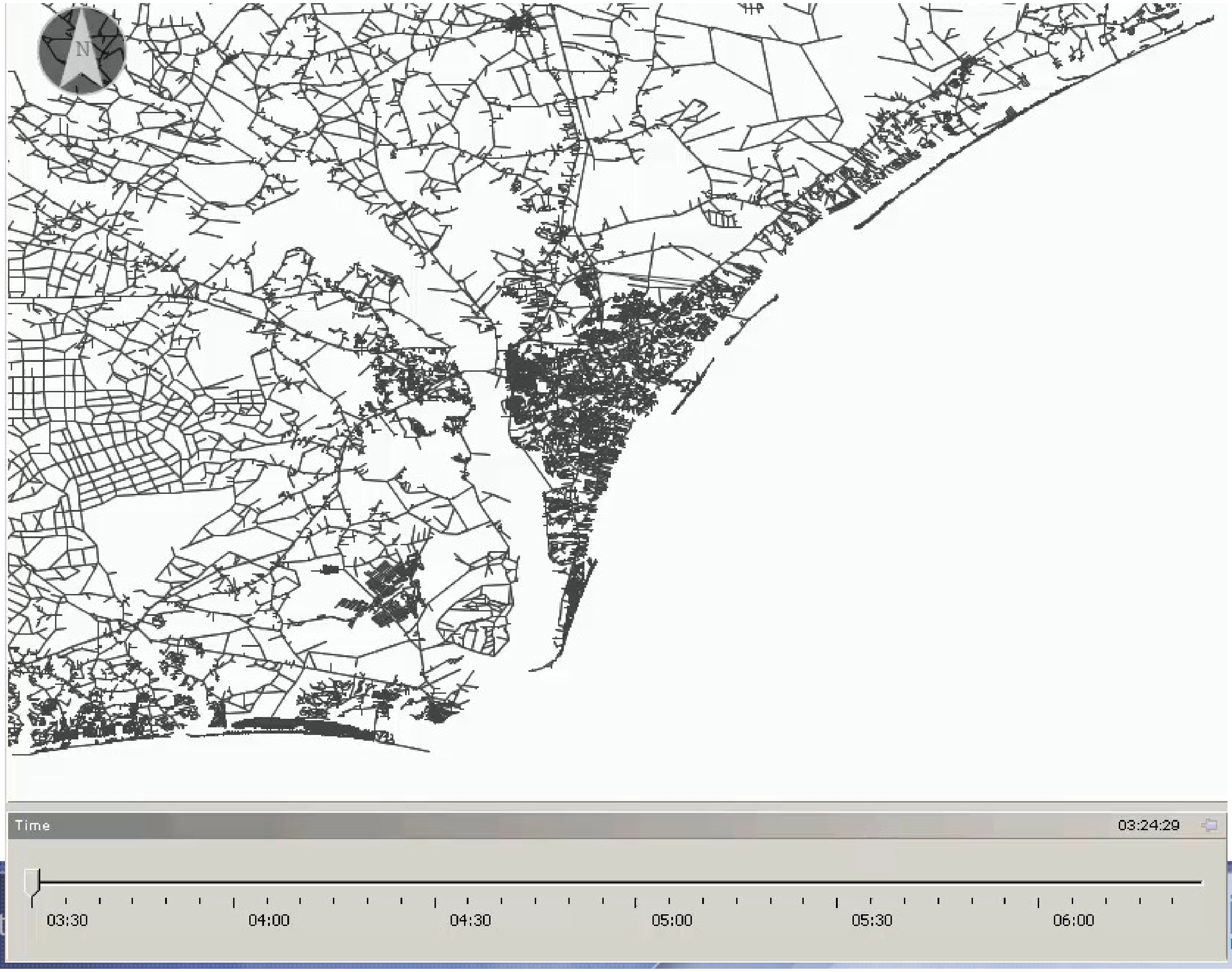


Version 2

Evacuation Optimization

Evacuation Optimization - Results

Evacuation Optimization - Results



Advantages of ABMs

Local level description (Policies are global, behavior is local. Use of synthetic populations, e.g., RTI's populations that match census at block level)

Multi-level modeling (below the skin – above the skin) (Models for tumor development, physiology, treatment, as well as behavior, environment, etc.)

Mean-field models might not be accurate because of Jensen's inequality

Allow modelers to create "virtual" societies (Individuals and institutions can be directly represented and the effects of their actions and interactions observed)

Conversion of cohort studies into population-level studies (Transition data is often collected through prospective studies, but the parameters are needed at the national level)

"Natural" setup for complex behavior (When heterogeneity is high no need to create artificial categories)

Major ABM's challenges

- **Individual-level data collection**
- **Validation is defined differently than for compartmental models**
  - Component validation rather than results validation
  - Pattern validation
- **Added uncertainty due to the propagation of error and stochasticity**

ODD "Standard" Protocol to Develop and Report an ABM

(Grimm et al. 2006, Railsback and Grimm, 2011)

	Purpose
Overview	Entities, state variables and scales
	Process overview and scheduling
	Emergence
	Objectives
	Adaptation
Design	Learning
concepts	Prediction
	Sensing
	Interaction
	Stochasticity
	Collectives
	Observation
	Initialization
Detail	Input data
	Submodels

Model Purposes

Model is a simplification of the reality intended to promote understanding

(SystemsWiki)

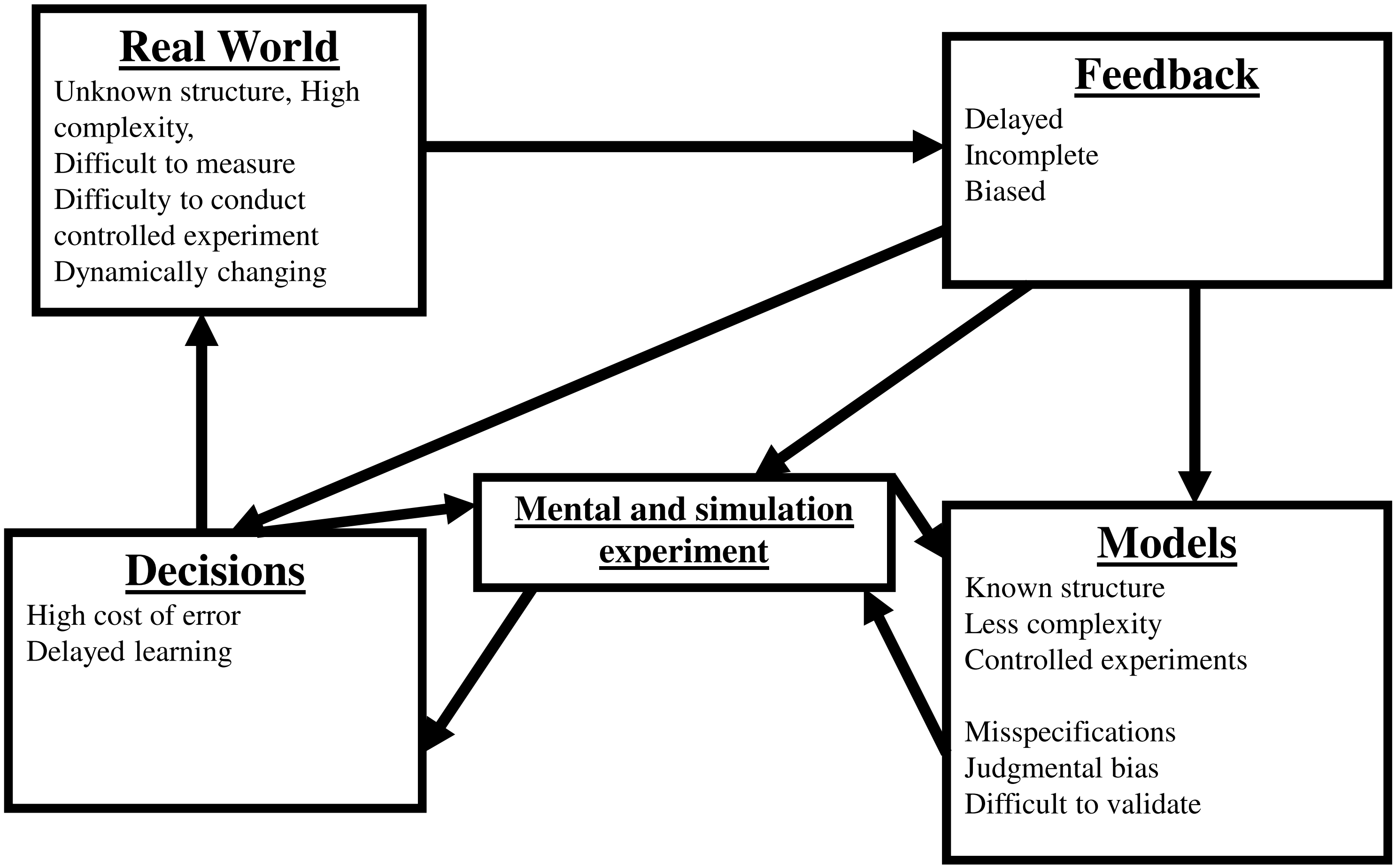
How Mathematical and Simulation Modeling Can be Used?

- Models represent the essence of the studied system in formal terms in order to
  - Understand (components and causality)
  - Describe (understanding in formal terms and compare to benchmark)
  - Predict (future relationships or numbers)

All Models are Wrong But Some are Useful (G. Box)

- **To answer a specific question (Will the intervention work? What will be its effect?)**
- **To identify data gaps (Social network data is rare available to understand initiation of drug use)**
- **To identify knowledge gaps (Do we have all necessary components to avoid unexpected consequences?)**

The Ideal Use of Simulation Modeling in Science



Predictive Models Predict a Number

It is difficult to make predictions, especially about the future *- Yogi Berra*

Predictive Models (predict a number)

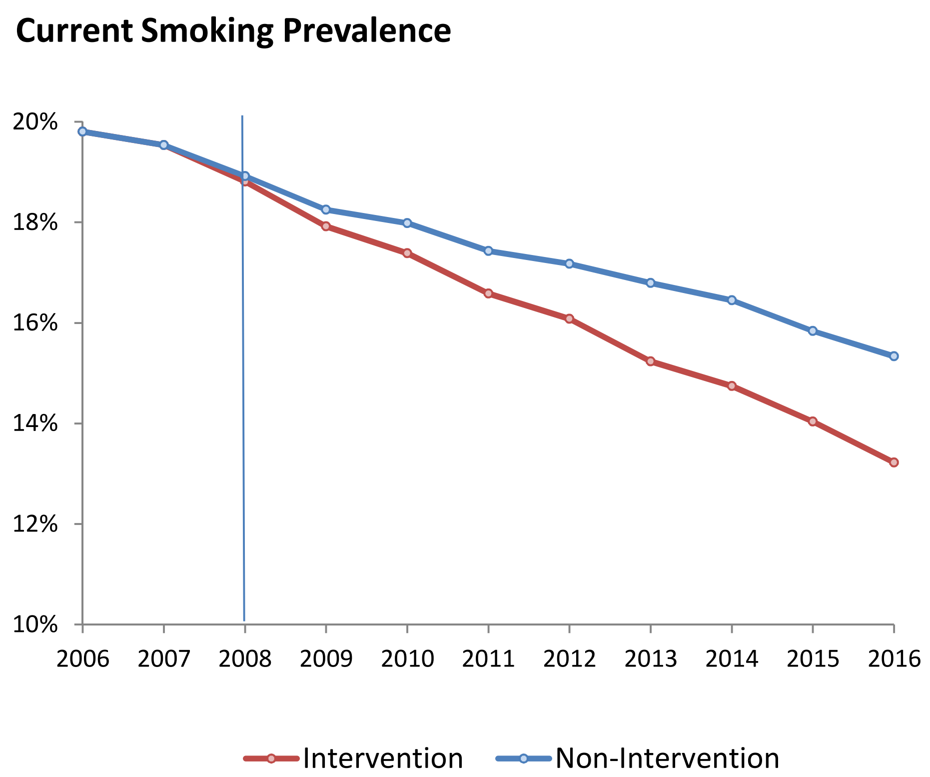
- **We know how many people smoke now and know the trends. Predict how many people will be smoking in 5 years**
- **Often the best predictor of tomorrow is today**

Can we do better than that?

Future=current + change

- **All predictions are based on "what if" assumptions**

Example of Predicting a Number
States Increase Funding to CDC Recommended Levels
(RTI, CDC tobacco report, 2008)



% Difference Between Intervention and Non-Intervention		
Year	Average Cumulative Funding	Current Smoking
2008	34.9%	-0.5%
2009	64.8%	-1.8%
2010	90.7%	-3.3%
2011	113.0%	-4.9%
2012	132.5%	-6.4%
2013	149.8%	-9.3%
2014	164.7%	-10.4%
2015	177.7%	-11.4%
2016	188.9%	-13.8%

Predictive Models Based on Association

- **Extending a regression**

Step 1. Fit a regression

$y=\beta_0+\beta_1 x+\varepsilon=N(\beta_0+\beta_1 x,\ \sigma^2)$

Step 2. Predict a new number

$$\hat{y}=\hat{\beta}_0+\hat{\beta}_1 x_1^{\,*}=\hat{\beta}_0+\hat{\beta}_1\bar{x}_1+\hat{\beta}_1\Delta x_1=y+\hat{\beta}_1\Delta x_1$$

Step 3. Estimate the variance of the prediction

$$Var(\hat{y}+\varepsilon)=Var(\hat{\beta}_0+\hat{\beta}_1 x_1^{\,*}+\varepsilon)=\sigma^2\left(\frac{1}{n}+\frac{(x^*-\bar{x})^2}{S_{xx}}\right)+\sigma^2$$

Decision Models
Choose Between the Actions

I don't want to make the wrong mistake *- Yogi Berra*

Example

What to Do in Case of Fire in a City?

(Drury et al. *A Principled Method of Scenario Design for Testing Emergency Response Decision-Making.* 2007)

Parameter Spaces

(Gary Klein ~ 2006)

- Question If there a wild fire what can we do put it down?
- **Background space (What is given?)**
- **Situation space (What are the variable factors?)**
- **Decision space (What can we potentially do?)**
- Develop a predictive model, an outcome, a cost metric
- Identify possible "what if" scenarios
- Compare the outcomes on the cost (regret) metrics

What to Do in Case of a Fire?

- Background space: Terrain, Season, Fixed resources, Time and location of the fire
- Situation space: Time and location of the fire, Variable resources, Weather forecast
- Decision space: How many trucks to send to put down the fires?
- Model: Simulation model of fire transmission, Outcome is regret ($)
- Scenarios: Consult the experts
- Conclusions: Have we considered politics?

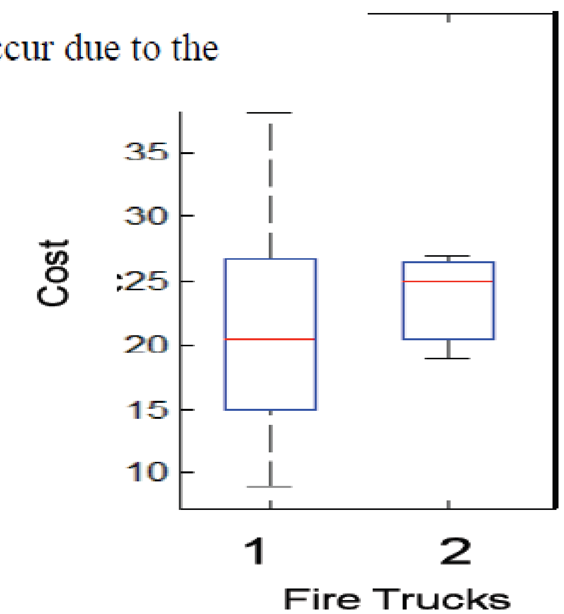
What to Do in Case of a Wild Fire?

(Drury et al. 2007)

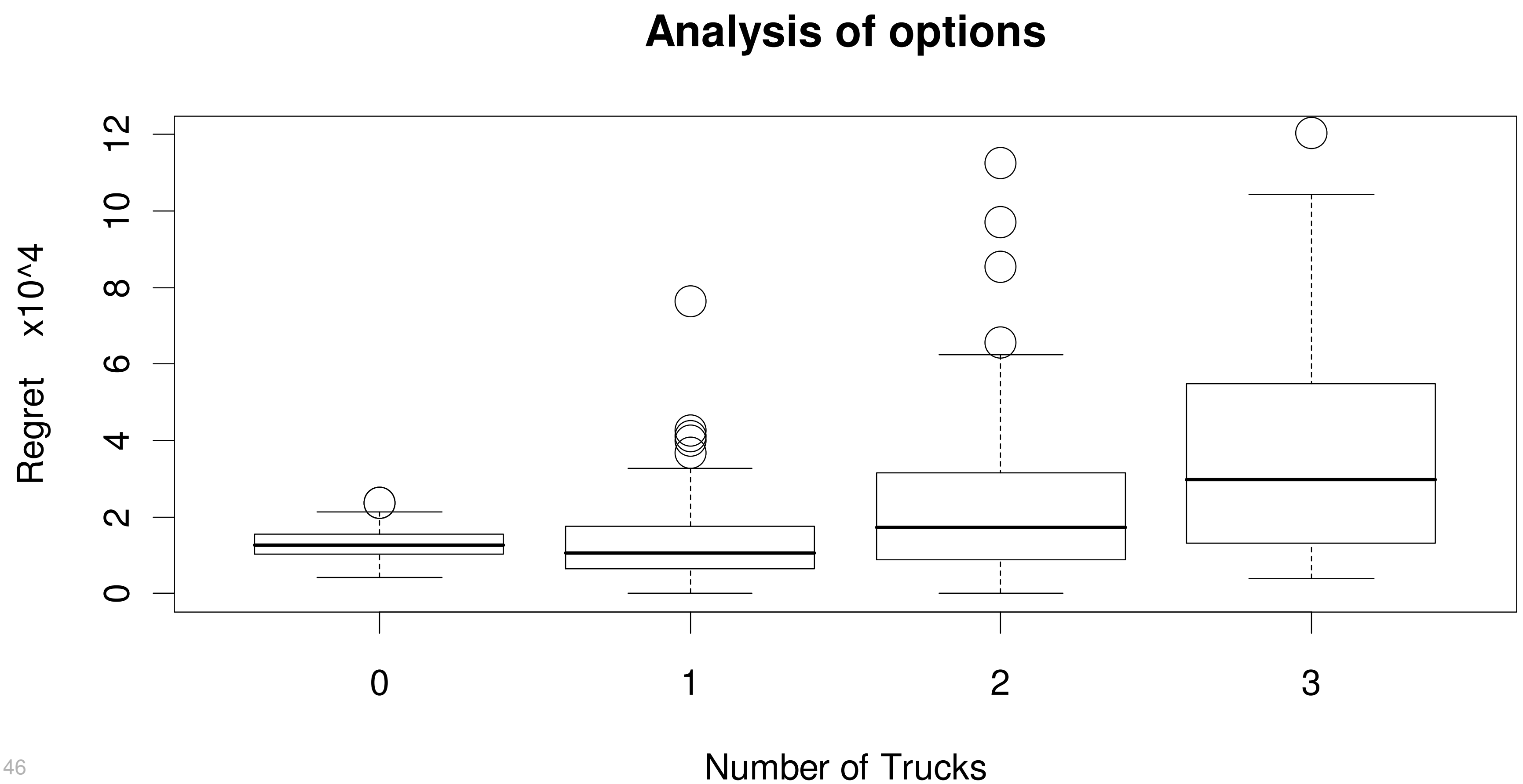
- **Regret (cost) measure**

Cost = $f\{M_i, R, PD_p, PD_f, I_p, I_f\}$, where

M_i	Initial magnitude of the incident
R	Cost of sending resources, scaled based on the number of resources allocated
PD_p	Property damage costs for the current incident
PD_f	Any additional property damage costs for future incidents that occur due to the response made for the current incident
I_p	Cost of injuries and/or deaths for the current incident
I_f	Any additional costs of injuries and/or deaths for future incidents that occur due to the response made for the current incident



Risk vs. Uncertainty



How Decisions are Made in Real Life?

Table 1. Example Challenging Scenario Components Based on Cost-function Conflicts

No.	Value/Parameter 1	Value/Parameter 2	Example Scenario Summary
1	Low M_i	High PD_p	A small fire has occurred in a jewelry shop
2	Low M_i	High I_p	A small fire has occurred in a crowded classroom
3	High M_i	Low PD_p	A large brushfire is burning well away from structures
4	High M_i	Low I_p	A large fire is burning in an uninhabited warehouse
5	High PD_p and High I_p	Possible High PD_f	A fire is burning in a crowded jewelry shop, but the city has been alerted that terrorists are threatening to bomb a (closed) art gallery
6	High PD_p and High I_p	Possible High I_f	A fire is burning in a crowded jewelry shop, but elsewhere in the city there is a packed stadium and home-town fans may riot if their team loses the game

Action Uncertainty

- **A. Unambiguous. There was clearly one course of action COA that was better than the others in all dimensions. This set of cases functioned as the control condition.**
- **B. Ambiguous: magnitude (Mi) versus current cost (PDp and/or Ip). These scenarios were of types 1 - 4 in Table 1.**
- **C. Ambiguous: current costs (PDp and/or Ip) versus future costs (PDf and/or If). These scenarios were of types 5 and 6 from Table 1.**
- **D. Ambiguous: best case costs (at the 0-percentile cost and 25-percentile levels) versus worst case costs (at the 75-percentile cost and 100-percentile cost levels) versus median costs. These scenarios reflected the robustness conflicts.**

Risk vs. Uncertainty

- 2X2 table by risk and uncertainty
- What is more important?
- We usually know how to deal with risk: reduce it
- Uncertainty can be irreducible. How to prepare?
- Game theory and preparedness

Understand a Theoretical Relationship

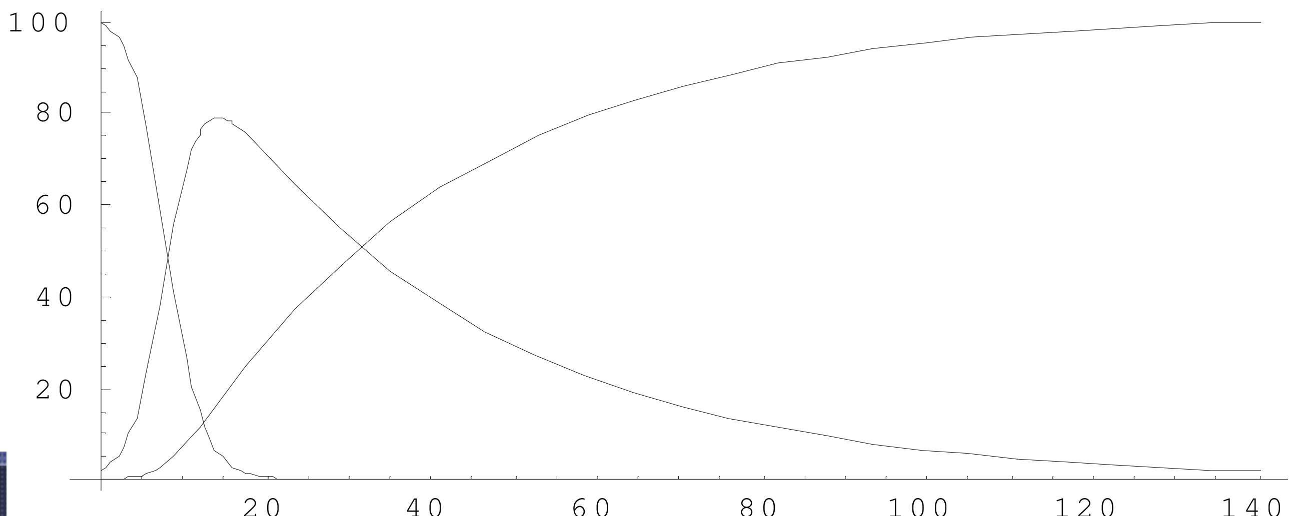
In theory there is no difference between theory and practice. In practice there is *–Yogi Berra*

Understand a Relationship

- Complex question that doesn't have neither a clear answer, nor supporting data, nor a conceptual model
- Develop a conceptual model
- Show that the model can represent reality
- Unexpected/Counterintuitive solution

Kermack-McKendrick SIR Model (1927)

- Susceptible Change Rate: $dS/dt = - \beta SI$
- Infective Change Rate: $dI/dt = \beta SI - \gamma I$
- Removed Change Rate: $dR/dt = \gamma I$



Important Theoretical Insights

- Existence of an epidemic threshold
- Necessary (but not stiochastically sufficient) condition for an epidemic to start, i.e.

$$dI/dt > 0 \Rightarrow \beta SI - \gamma I > 0$$

- Which translates into a critical proportion for susceptible population

$$S_{critical} > \gamma / \beta \Rightarrow \gamma / (\beta\, S_{critical}) < 1 \Rightarrow$$

- Concept of “Herd Immunity” Don’t need to vaccinate everyone!!

Basic Reproductive Rate (Ratio) R_0

Basic Reproductive Ratio $R_0 \equiv \dfrac{\beta S_0}{\gamma}$

A number of secondary infections per one infected individual

$R_0 > 1 \Rightarrow$ epidemic is possible

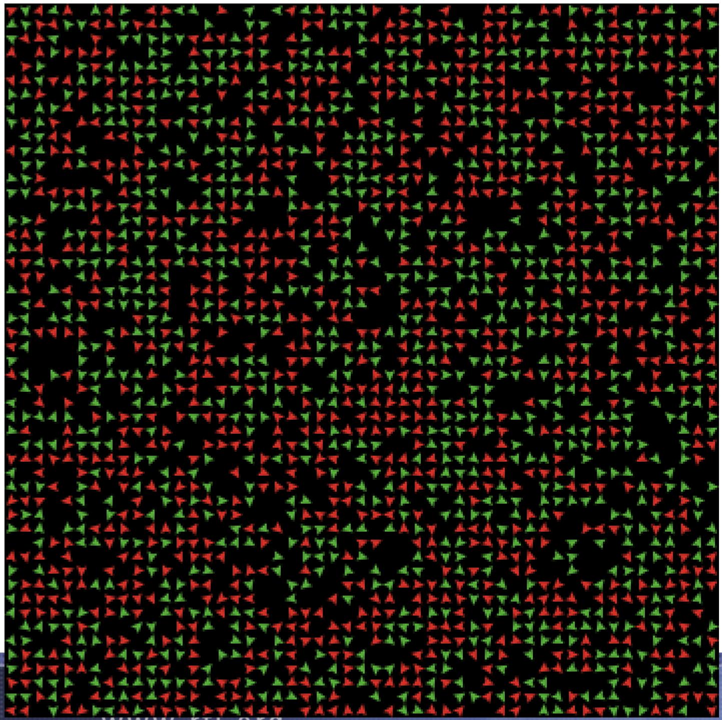
$R_0 < 1 \Rightarrow$ epidemic will not occur

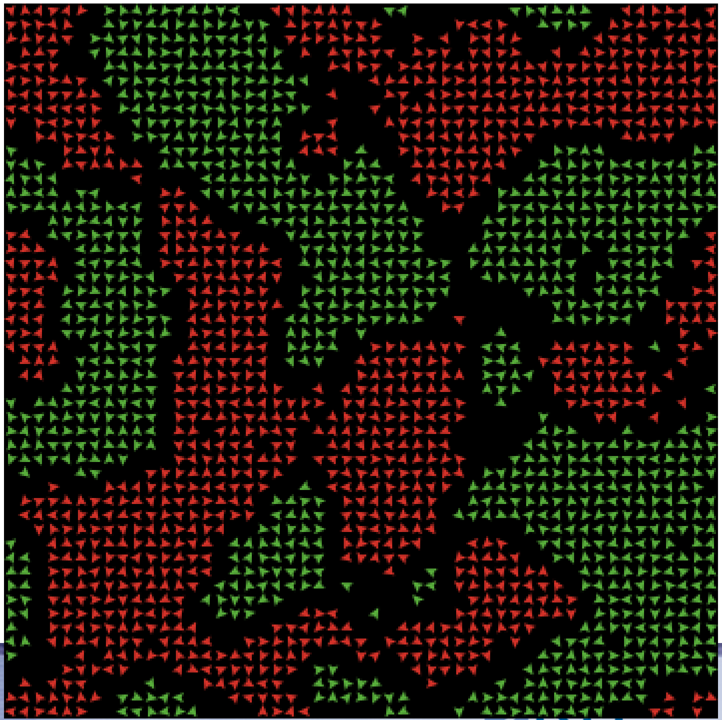
Need to vaccinate $1 - 1/R_0$ proportion of the population

This model is wrong, but so is Malthusian, Lotka-Volterra's, Shelling's, and many others

Shelling Segregation Models

In 1971 Shelling showed that a small preference for one's neighbors to be of the same color could lead to total segregation. He used coins on graph paper to demonstrate his theory by placing pennies and nickels in different patterns on the "board" and then moving them one by one if they were in an "unhappy" situation.





Risk Estimation

The future ain't what it used to be *- Yogi Berra*

Estimation of Risk

Understanding of risks affects our daily lives.

- **Go through the screening frame or get patted down?**
- **What is the risk getting cancer from low radiation?**
- **How to get an accurate estimate?**
- **Decision made on a single dose (42’ of regular life or 2’ flying)**
- **Exact studies are often unethical, infeasible and unnecessary**

Example

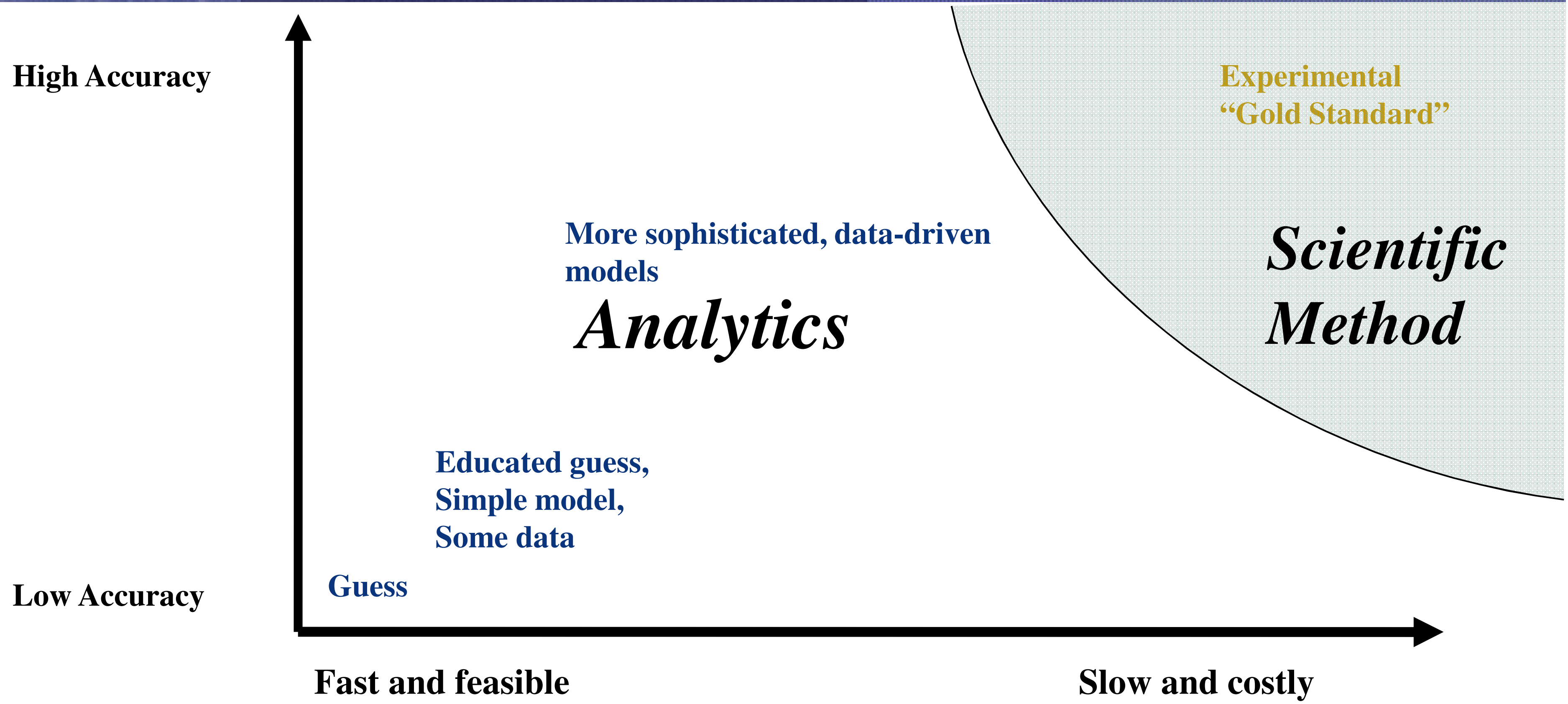
Estimate HIV Risk for a Specific Person

Estimation of Risk

What is the increase in risk of getting HIV form different behaviors?

- **Similarity with radiation but there is a difference**
- **Categorical vs. continuous**
- **Multifactor contribution, feedback loops, non-linearities**
- **How much accuracy do we need and how should a study look to answer this question?**

Compromise of Precision and Speed



Guess

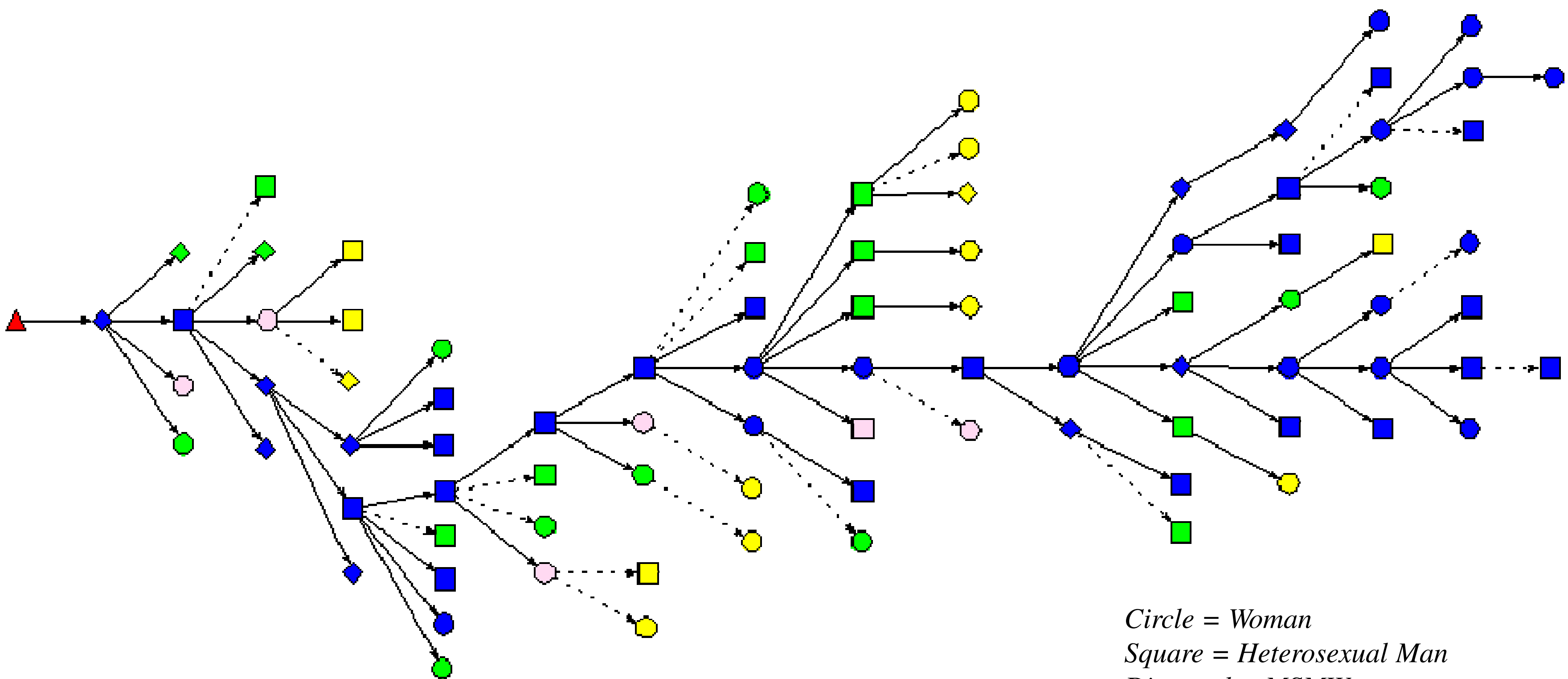
- The more sex partners the bigger the risk
- The more sharing syringes the higher the risk
- If someone has many sex partners, injecting drugs and sharing syringes then very high risk
- Difficult to quantify risk without numbers

A Cross-sectional Sample

- Sexual Acquisition and Transmission of HIV Cooperative Agreement Program (SATH-CAP), NIDA, William Zule PI

- About 2,000 members of at-risk group:
  - Men who have sex with men (MSM)
  - Drug users (DU)
- Respondent-driven sample (RDS) for data collection

RDS Sample



Circle = Woman
Square = Heterosexual Man
Diamond = MSMW
Triangle = MSM
Blue – Hard drug user
Green, yellow & pink – Recruited as different types of sex partners
Dotted line – Ineligible

Poorly Educated Guess.

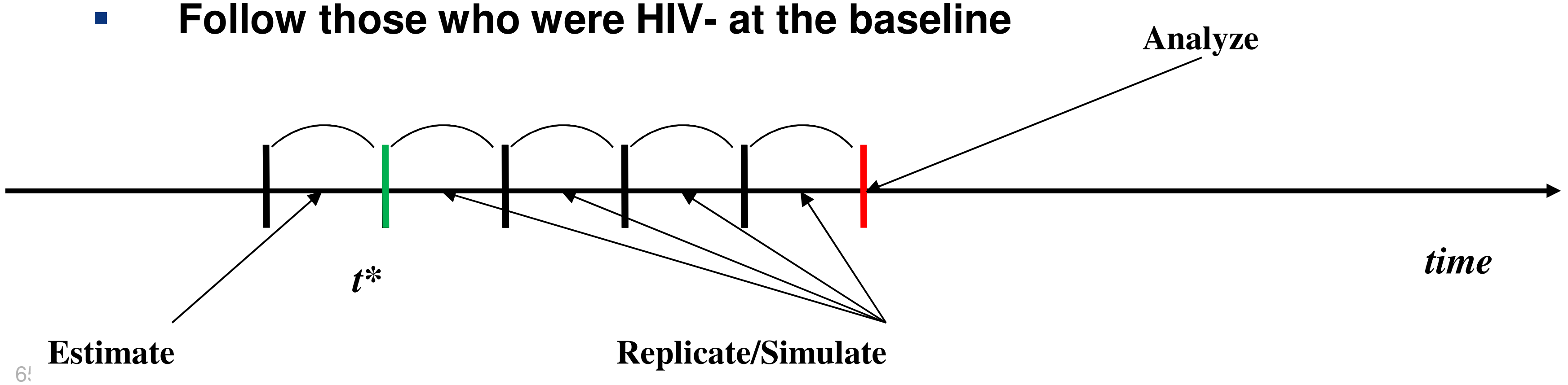
Absence of Causal Model Applied to Cross-sectional Data
Prevalence vs. Incidence?

Number of sex partners in past 6 months among MsM and MsMW

Level	Rel_risk of HIV	Odds	Odds Ratio
1	ref	0.21	ref
2-5	1.65	0.40	1.91
6-9	0.83	0.17	0.81
10+	0.80	0.16	0.76

Main Concept

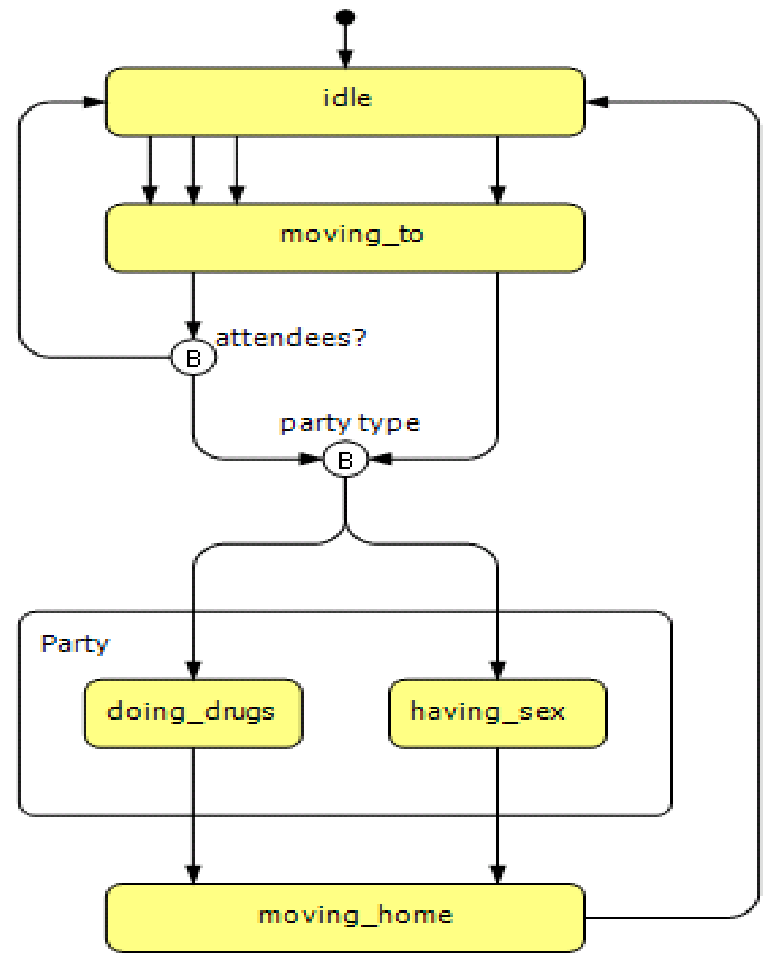
- Future=Current + Change
- Estimate the rate of behaviors, for example:
  - The number of unprotected sex acts in the past month
  - The number of equipment sharing in the past month
- Repeat many times in the community of HIV + and HIV –
  - Follow those who were HIV- at the baseline



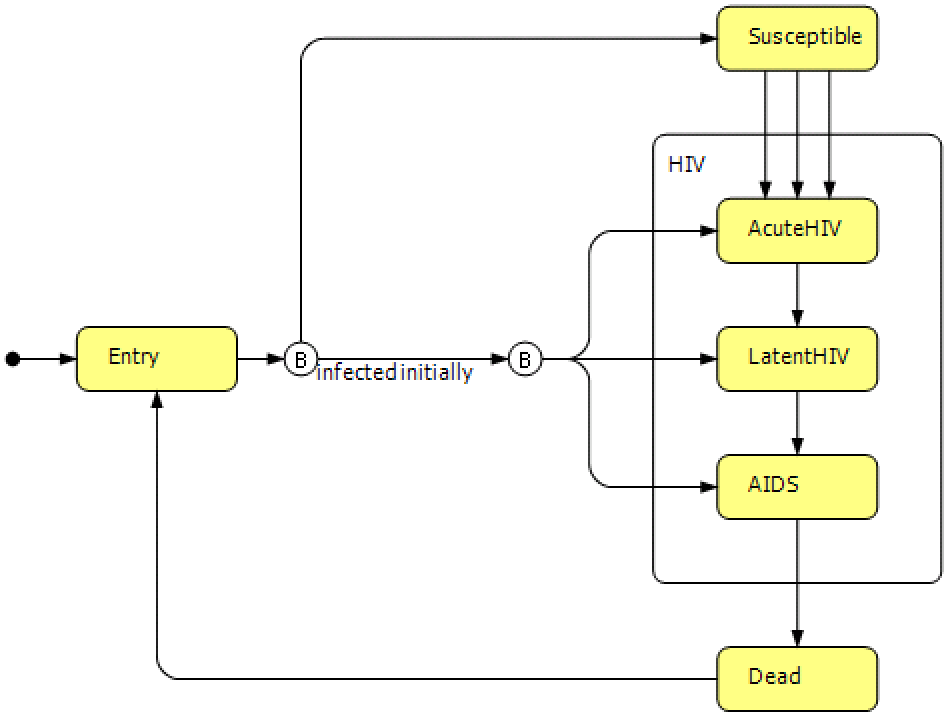
Components of the Model

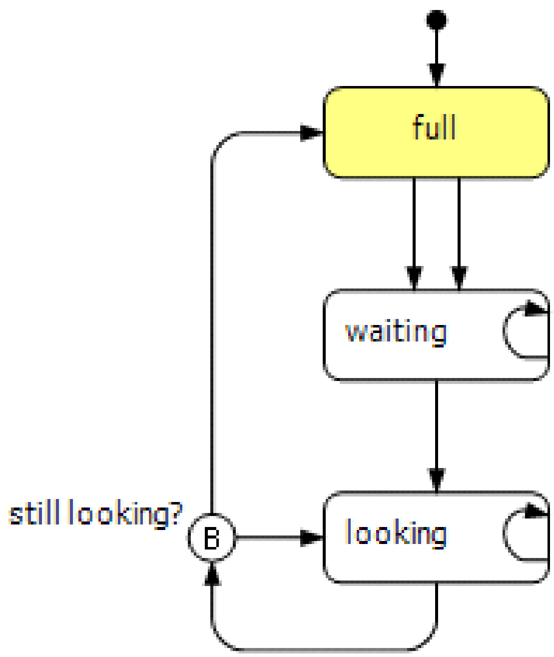
- Viral load and HIV progression
- Sexual behavior
- Drug-using behavior
- Syringe types
- Structure of sexual and equipment-sharing networks
- Sexual and drug use mixing matrices (who has sex with whom)

Individual State Diagram



Model of HIV Spread on Sexual and Drug-Using Networks

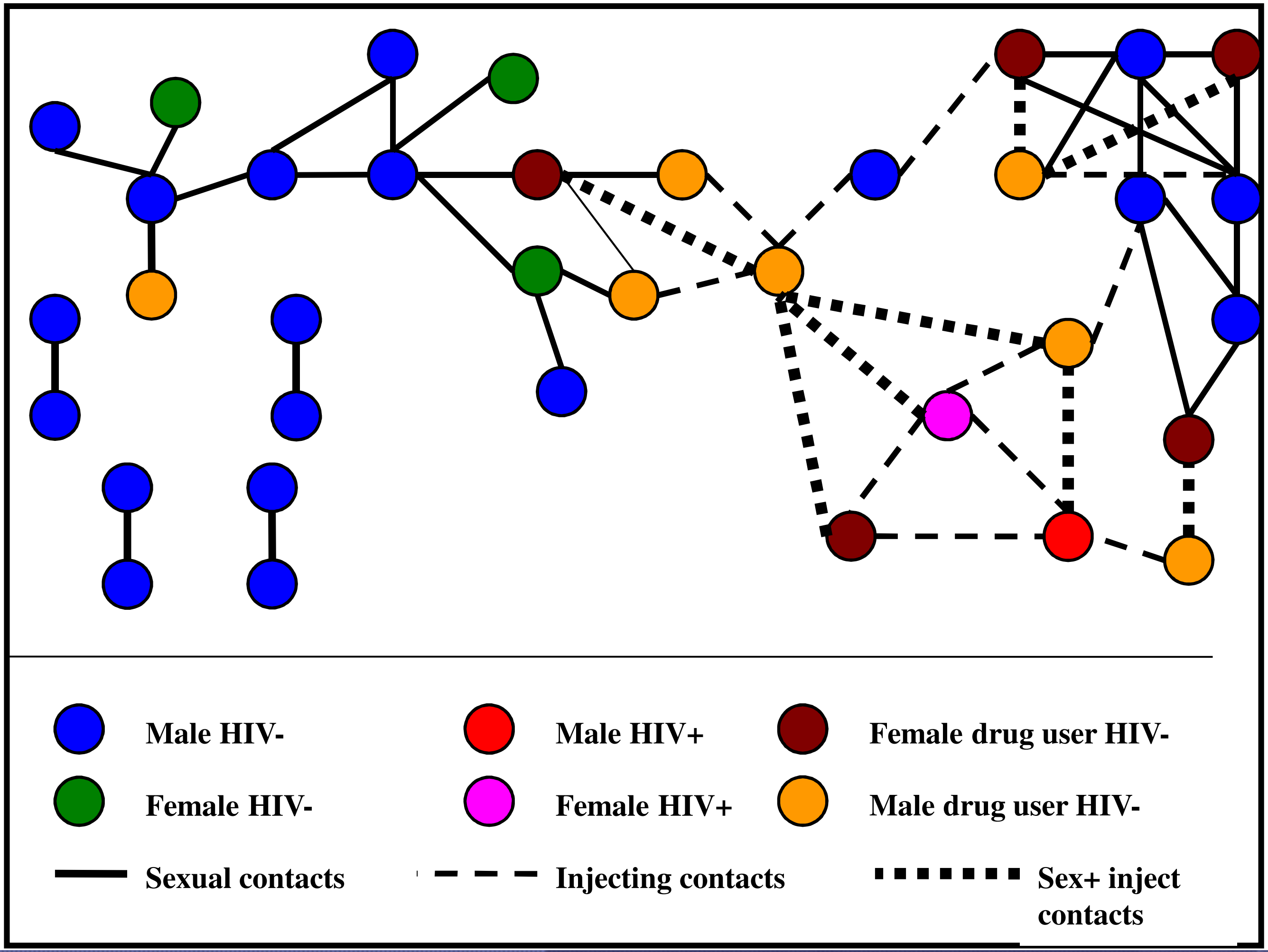




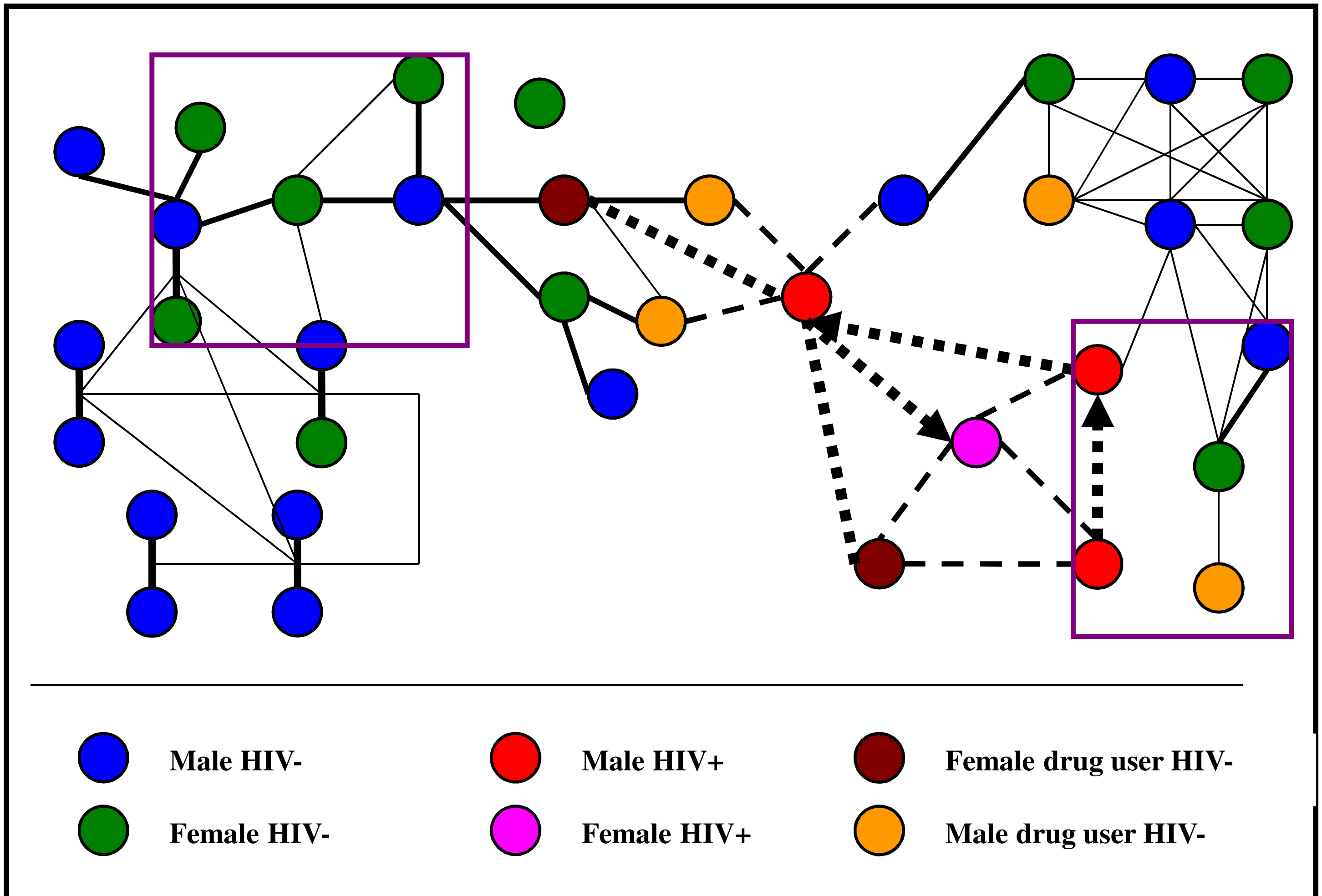
Population and Network

- Increase the sampled group by factor of 10 based on independent estimates.
- Estimate a mixing matrix (who has with whom) and (who injects and with whom).
- Connect agents based on the link's distribution to assure approximate balance of the egocentric link reports.
- Networks are functional and evolving.

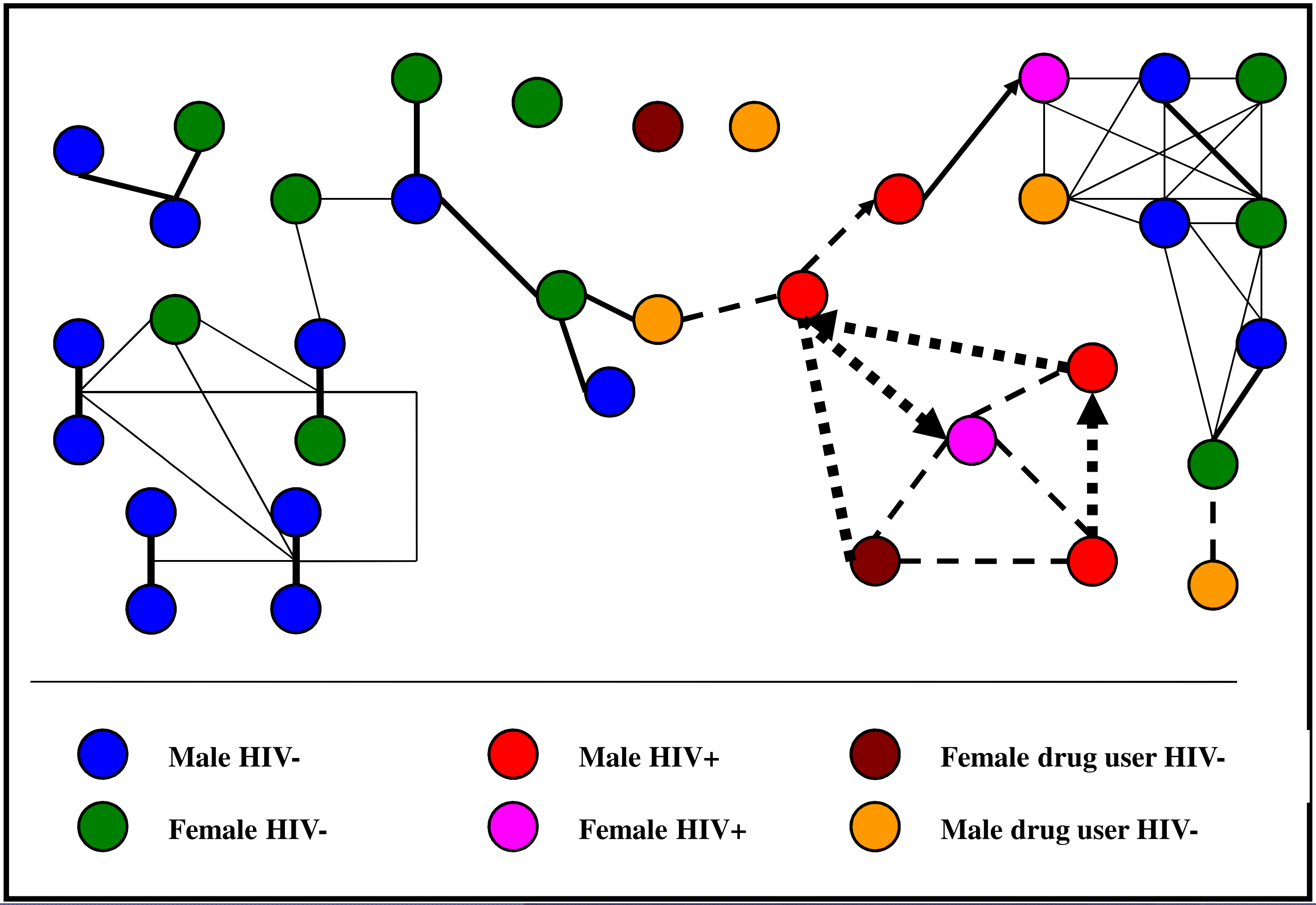
Sexual and Drug-Using Network



HIV Spread on Networks (1 year later)



HIV Spread on Networks (2 years later)



Model Parameters

- Survey data
  - Demographics, Sexual behaviors, Drug-use behaviors, Mixing matrix
- Peer-Reviewed Publications
  - Partner change dynamics, HIV transmission probabilities by sexual behaviors, sex of partner, type of sex (oral, anal, vaginal), condom use; by drug-use behaviors, using safe syringe, sharing syringes
- Educated guess
  - Network structure and contacts, Dynamics of links

Estimates Based on Prevalence vs. Model-based Incidence

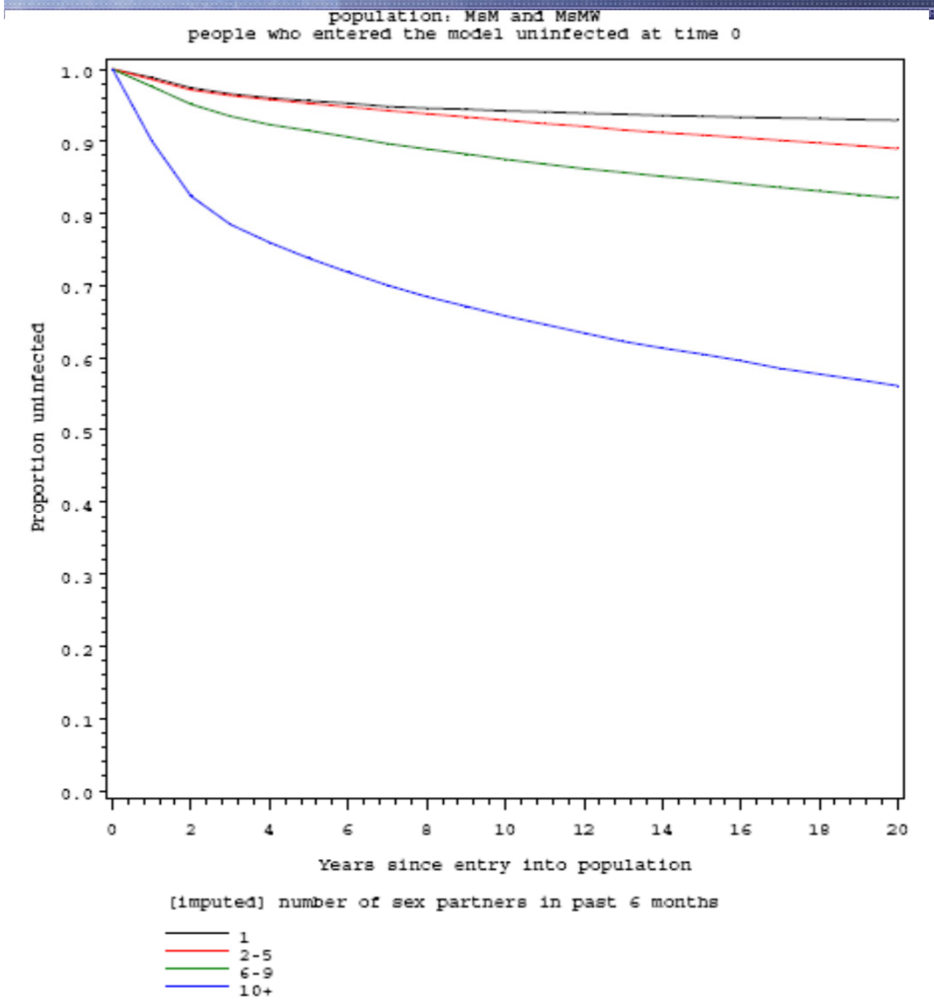
Number of sex partners in past 6 months. MsM and MsMW who entered the model at time 0.

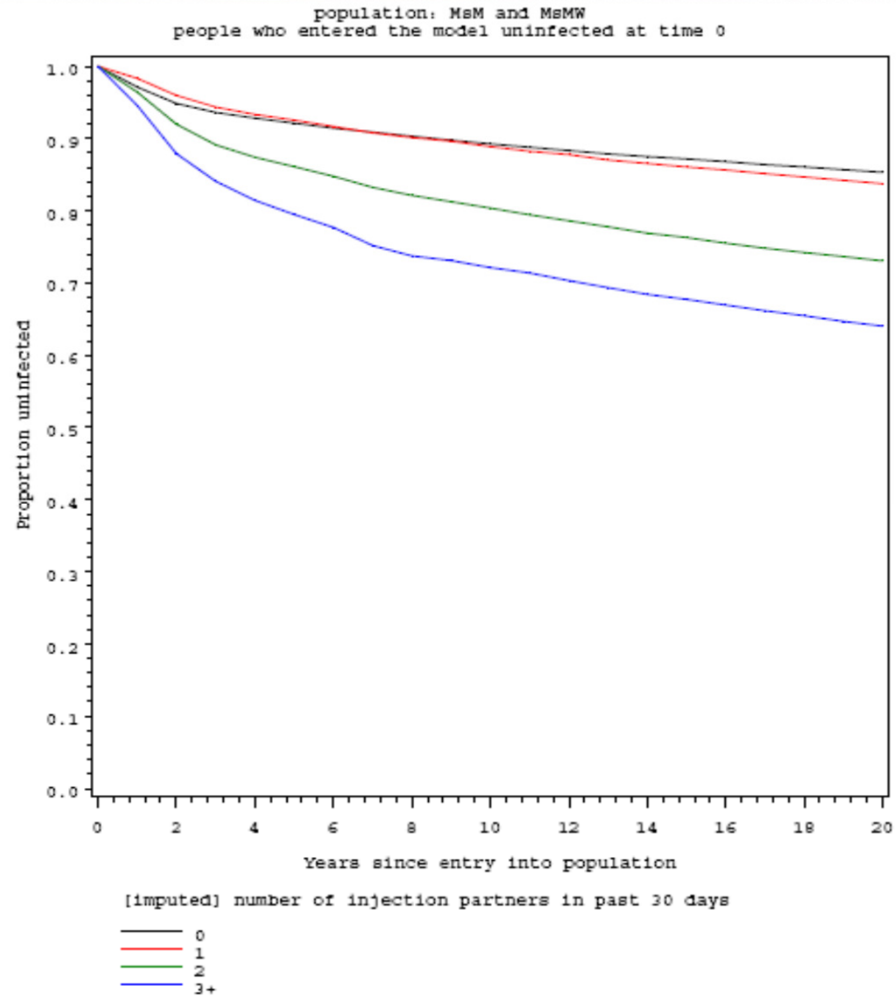
Level	Rel_risk of HIV	Odds	Odds Ratio
1	ref	0.21	ref
2-5	1.65	0.40	1.91
6-9	0.83	0.17	0.81
10+	0.80	0.16	0.76

Number of sex partners in past 6 months. MsM and MsMW who were not infected at the baseline. Assessed after 5 years.

Level	Rel_risk of HIV	Odds	Odds Ratio
1	ref	0.06	ref
2-5	1.02	0.06	1.03
6-9	1.88	0.12	1.98
10+	5.29	0.44	7.17

Surviving HIV is Related to the Number of Sexual and Injecting Partners





Who Are At the Highest Risk?

Number of sex partners past 6 months	Number of injection partners	Average rate of partner change per year	Number of people with whom shared a syringe in 30 days
1	3	0.6	13
500	0	23.2	0
80	0	32.4	0
27	2	16.9	2
22	0	15.1	0
3	10	3.7	3
5	10	4.5	11
6	2	4.3	2
2	2	3.3	4

Who Are At the Highest Risk?

Characteristic	P(infected after 3 years) >= 10%	P(infected after 3 years) < 10%	P(infected after 3 years) < 3%
Median age	43.0	41.0	41.0
Race distribution			
Black	87.0	77.6	77.5
White	10.7	14.0	13.5
Hispanic	0.3	2.4	2.2
Gender distribution			
Male	52.9	52.6	51.6
Female	47.1	47.4	48.4
Sex behavior distribution - men			
MSM	23.5	10.1	8.9
MSW	37.2	68.6	69.7
MSMW	37.2	8.8	5.3
not active	2.2	12.5	16.1

Who Are At the Highest Risk?

Characteristic	P(infected after 3 years) >= 10%	P(infected after 3 years) < 10%	P(infected after 3 years) < 3%
Median number of sex partners (sexually active only)	9	2	1
Median number of yearly sex partner changes (sexually active only)	5.1	3.4	0.8
Percentage stimulant users	84.5	64.3	62.5
Percentage injection drug users	31.2	6.0	5.8
Median number of injection partners (injection drug users only)	3	1	1