Individual-Based Models: Introduction, Tradeoffs & Tools

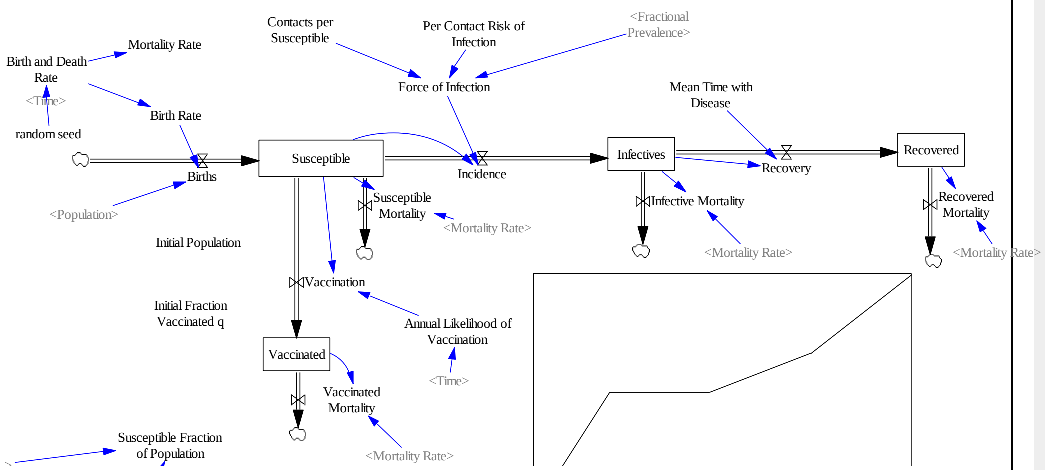
Nathaniel Osgood

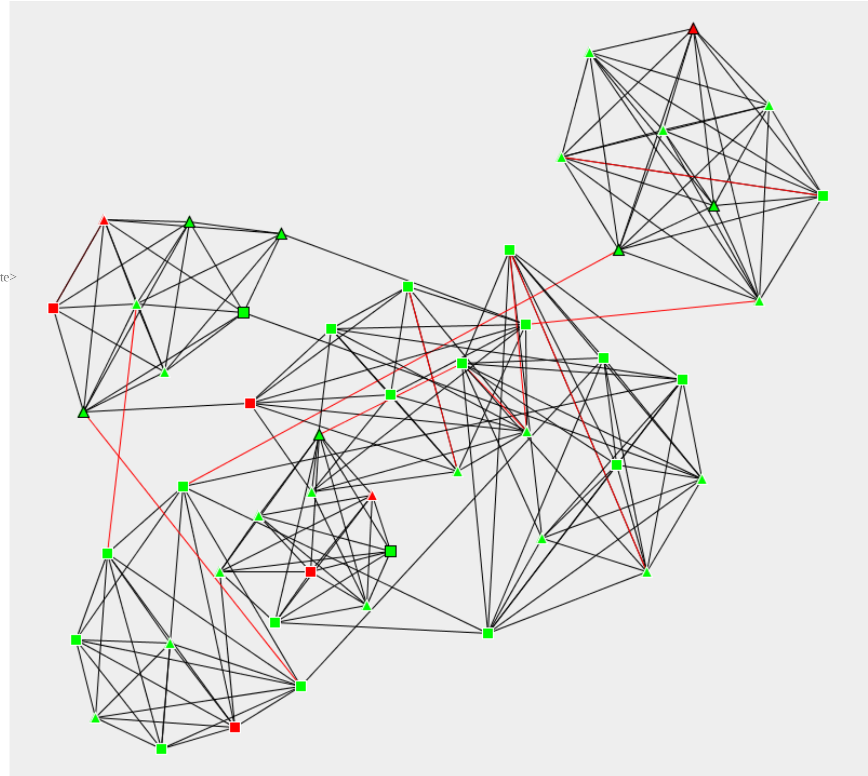
April 6, 2010

Dynamic Models for Health

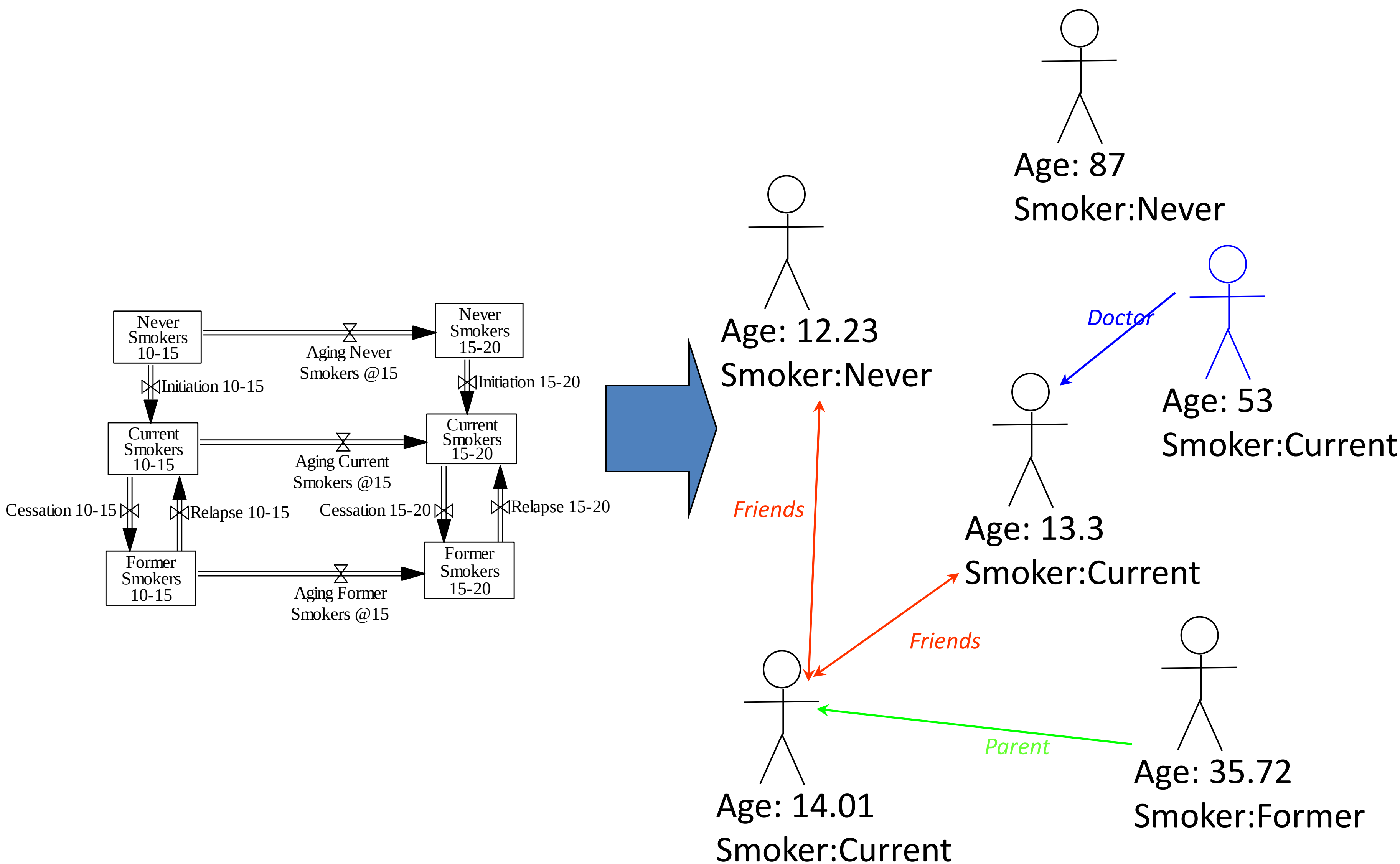
- Classic: Aggregate Models
  - Differential equations
  - Population classified into 2 or more state variables according to attributes
  - |State Variables|,|Parameters| << |Population|
- Recent: Individual-Based Models
  - Governing equations approach varies
  - Each individual evolves
  - |State Variables|,|Parameters| $\propto$ |Population|

Contrasting Model Granularity

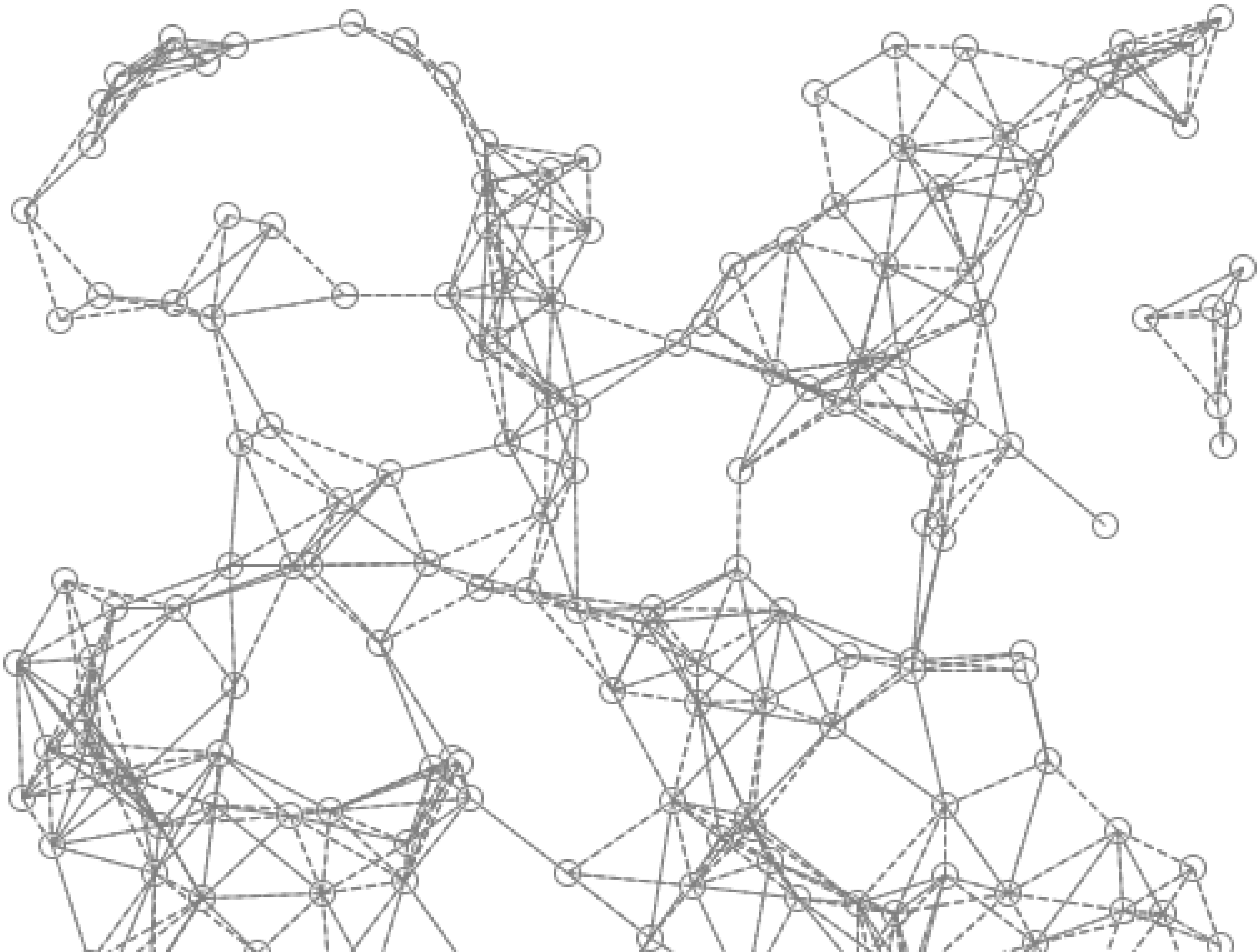




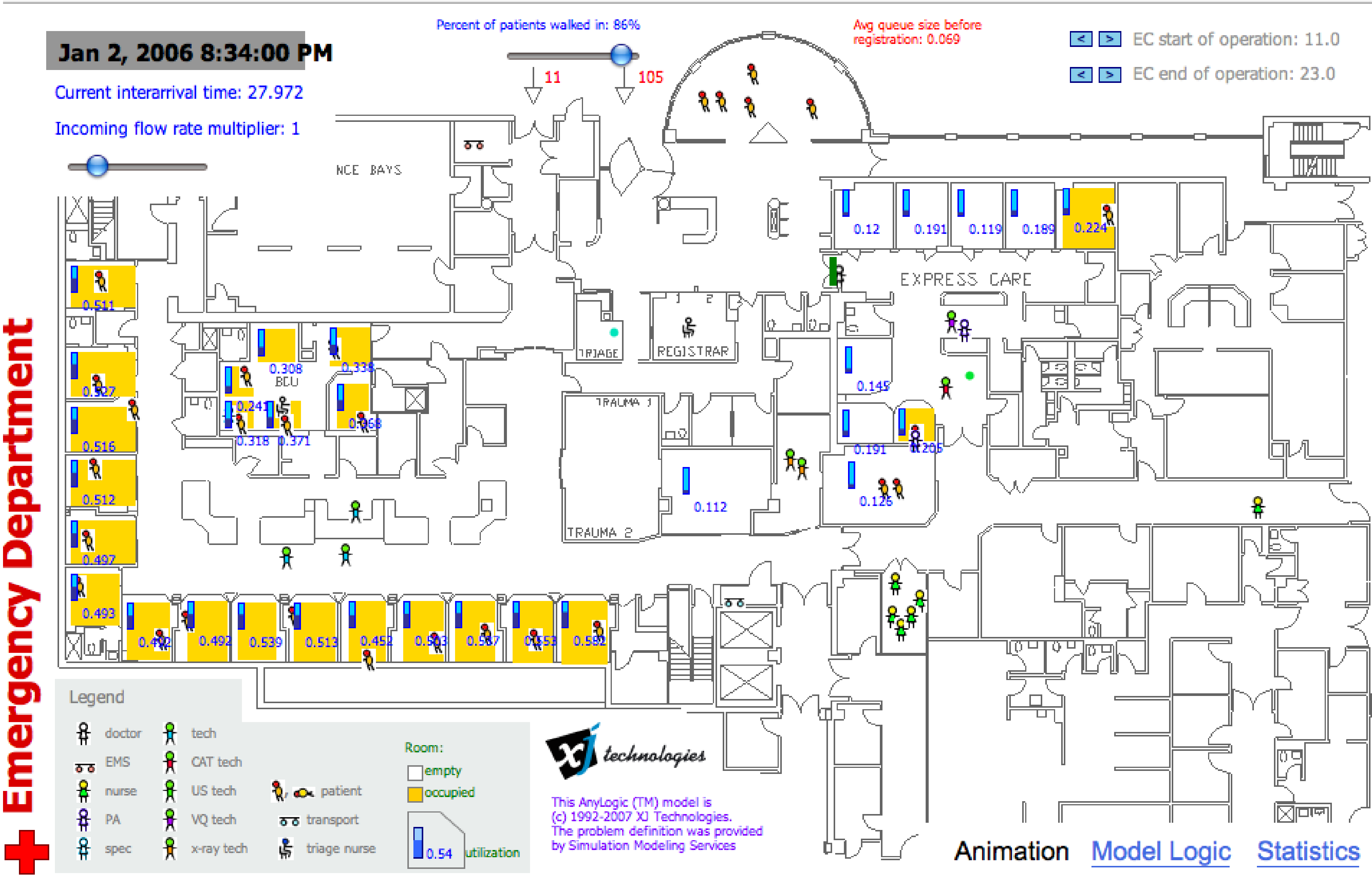
Interacting Individuals



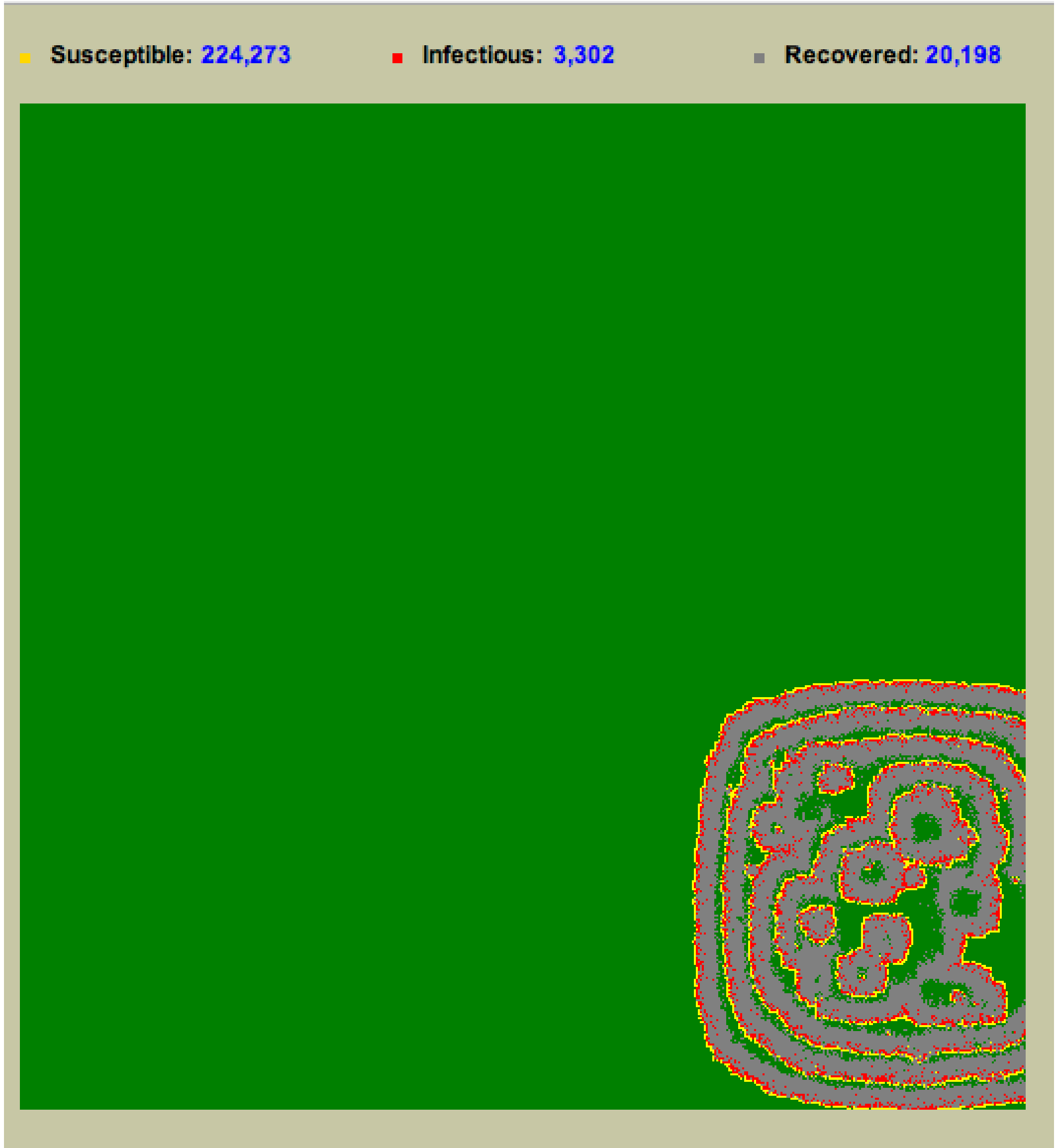
Network Embedded Individuals



Irregular Spatial Embedding



Regular Spatial Embedding



Agenda

- Motivations & Context
- Comparing Aggregate & Individual Based Models
- Granularity Tradeoffs
- Tools for individual-based modeling
  - Individual-Based Modelers in SD
  - Individual-based models in Agent-Based tools
- Other tradeoffs
- Looking forward

Importance of Heterogeneity

- Heterogeneity often significantly impacts policy effectiveness
  - Policies preferentially affect certain subgroups
    - Infection may be maintained within certain subgroups even though would tend to go extinct with random mixing in the entire population
  - Policies alter balance of heterogeneity in population
    - Shifts in the underlying heterogeneity can change aggregate population statistics
  - Given a non-linear relationship, inaccurate to use the mean as a proxy for whole distribution
- **Assessing policy effectiveness often requires representing heterogeneity**
- ***Flexibility* in representating heterogeneity is hard to achieve in aggregate (coarse-grained) models**

Longitudinal Heterogeneity

- There can be great heterogeneity not only cross-sectionally, but also longitudinally
  - Particularly in a path-dependent system, trajectories that are originally close may diverge dramatically
- Capturing this longitudinal disparity can be important for understanding intervention effects

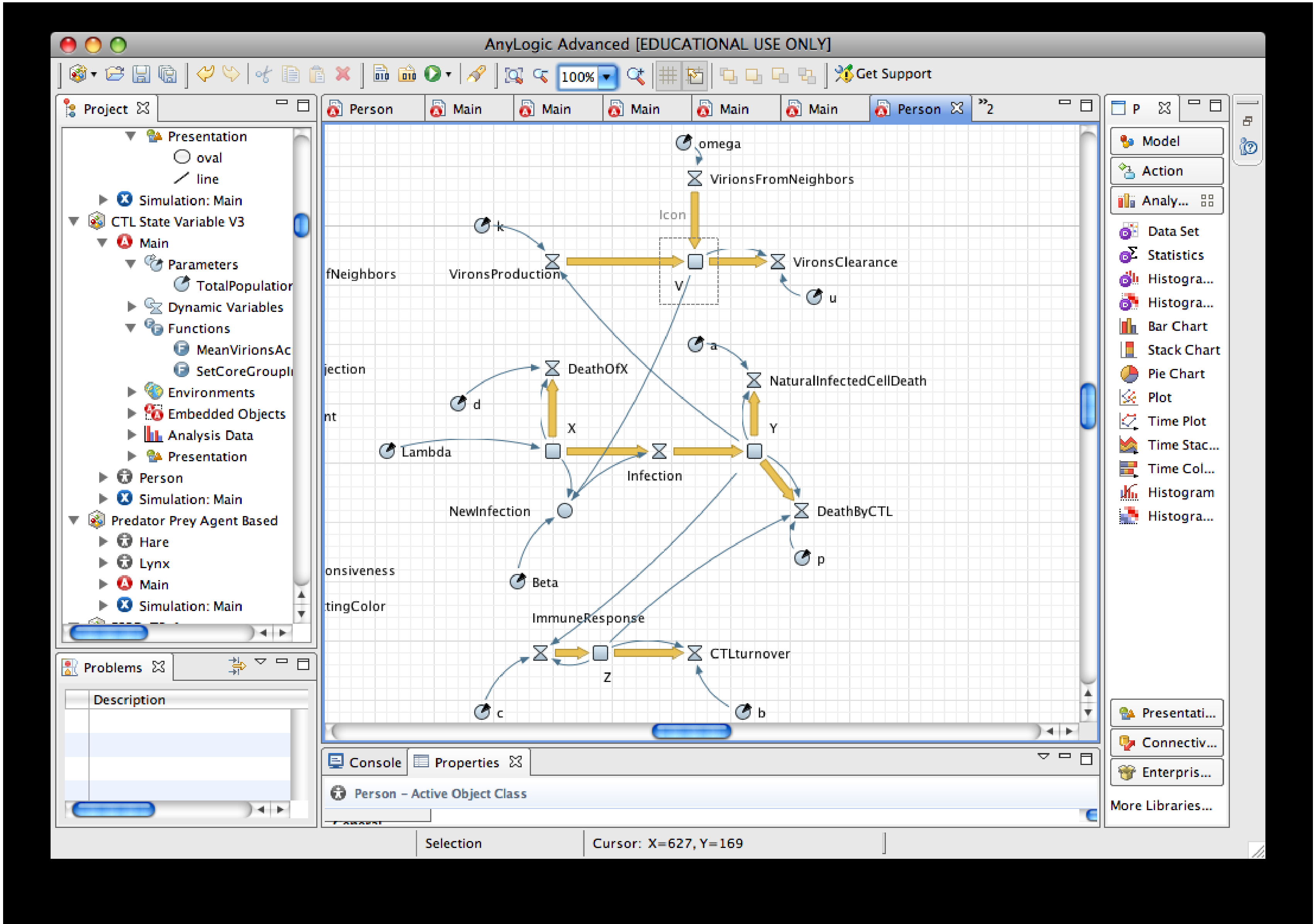
Agenda

✓ Motivations & Context
- Comparing Aggregate & Individual Based Models
- Granularity Tradeoffs
- Tools for individual-based modeling
  – Individual-Based Modelers in SD
  – Individual-based models in Agent-Based tools
- Other tradeoffs
- Looking forward

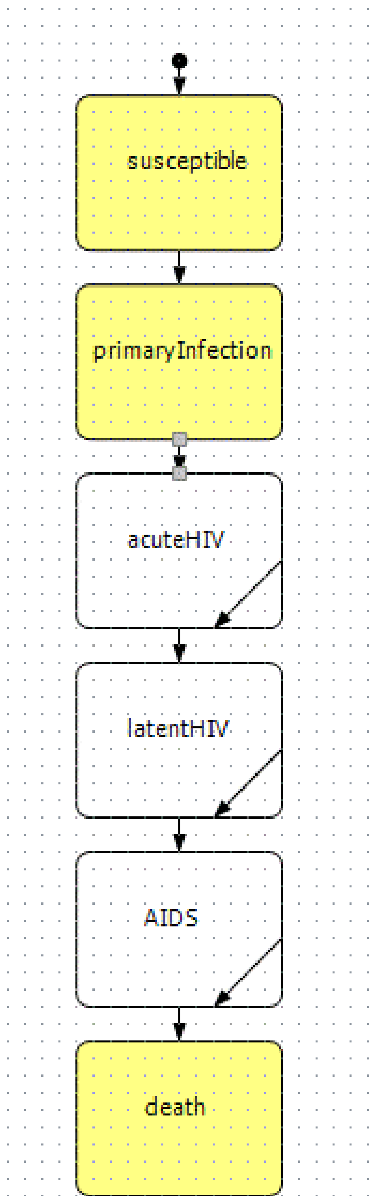
Elements of Individual State

- Example Discrete
  - Ethnicity
  - Gender
  - Categorical infection status
- Continuous
  - Age
  - Elements of body composition
  - Metabolic rate
  - Past exposure to environmental factors
  - Glycemic Level

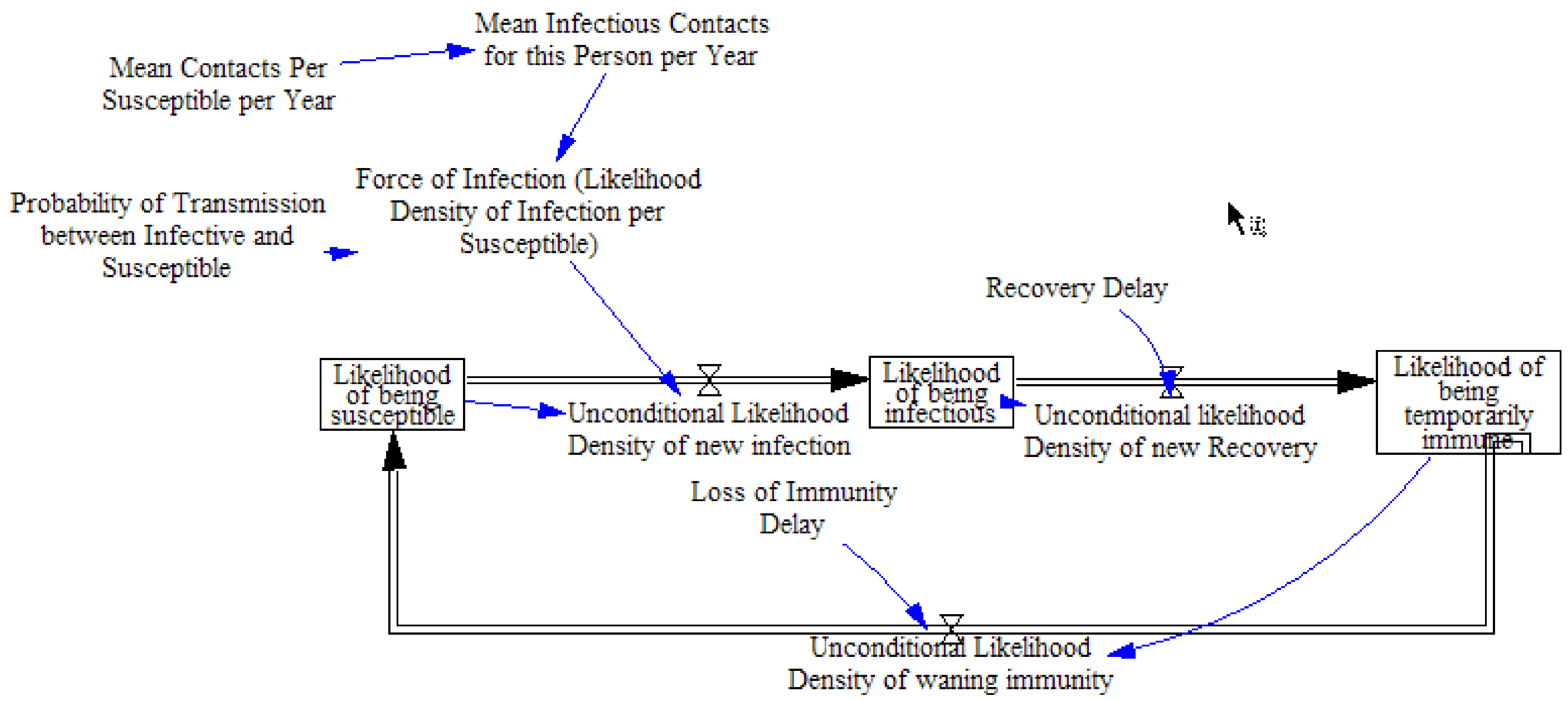
Example of Continuous Individual State



Example of Discrete States
Binary Presence in Discrete State

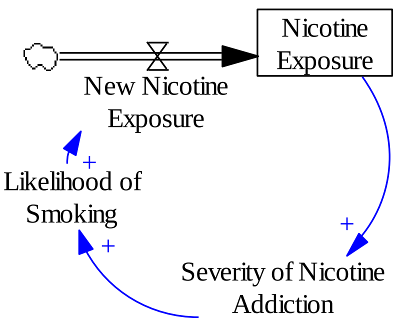


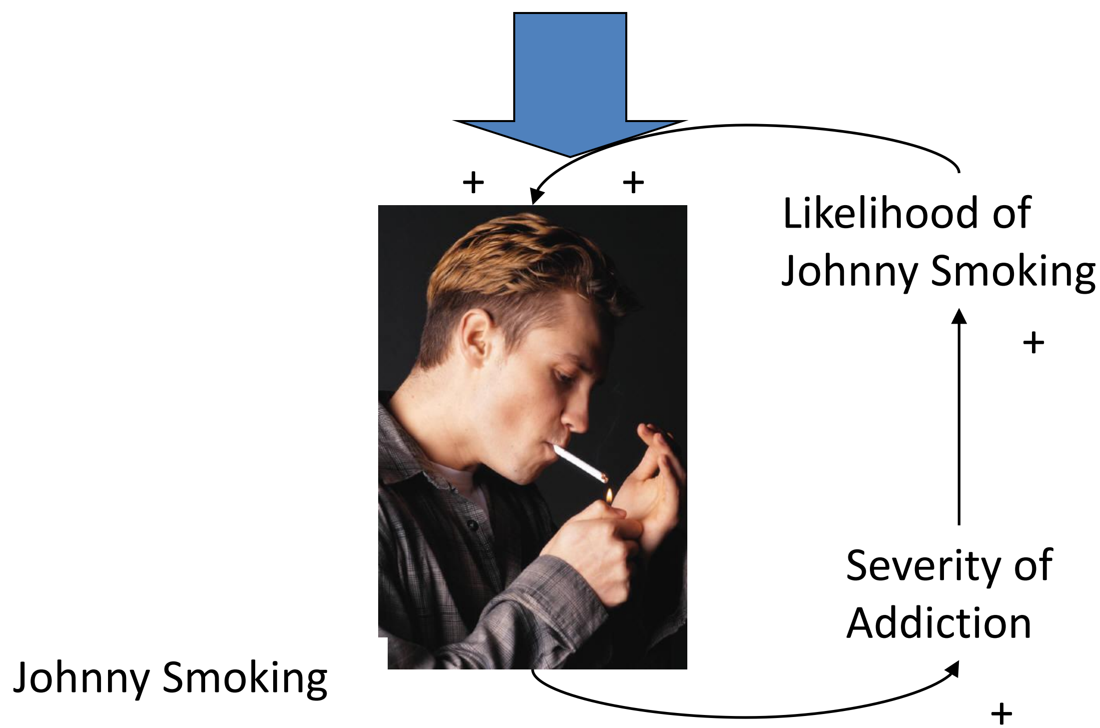
Example of Likelihood of Presence in Discrete State



Feedbacks

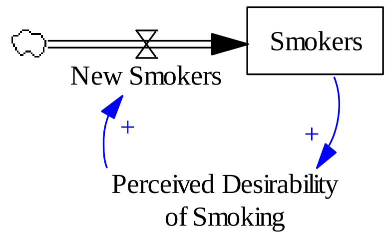
- Some aggregate feedbacks lie within individual agent





Feedbacks

- Many aggregate feedbacks are *between* agents



Aggregate Model

Johnny's Father's Smoking

+

Johnny's Perception Of Desirability Of Smoking

+

Johnny Smoking

+

+

Timmy's Perception Of Desirability Of Smoking

Agent-Based Model

Agenda

- ✓ Motivations & Context
- ✓ Comparing Aggregate & Individual Based Models
- Granularity Tradeoffs
- Tools for individual-based modeling
  - Individual-Based Modelers in SD
  - Individual-based models in Agent-Based tools
- Other tradeoffs
- Looking forward

Granularity Selection: Problem Specific

- Selection of granularity is a function of question that are asking – not of the "*true nature of the system*"
- Quanta of most obvious system components may not align with needs for insight
  - May gain benefits from higher-level representation
    - Many high-level behavior of complex systems can be explained with very simple models
    - Often gain greater insight from simpler model: Cf Gas laws vs. lattice gas model
  - May wish to seek lower level model
    - Small infection spread model : Characterization at level of immune response rather than monolithic person

Myth of Individual-Based Models as “Modeling from the Bottom Up”

- A single person is a natural locus of description
  - Presents for care
  - Lives
  - Dies
  - Coupled internal systems
- But the world has no natural “bottom”
  - It is frequently desirable to include within a person a great deal of “within the skin” detail
- The issues of model depth & breath are just as pressing in individual-based models as in aggregate modeling

Contrasting Benefits

Aggregate Models

- Easier
  - Construction
  - Calibration
  - Parameterization
  - Formal analysis (Control theoretic & Eigenspace techniques)
  - Understanding
- Performance
  - Lower baseline cost
  - Population size invariance
- Less pronounced stochastics
  - Less frequent need for Monte Carlo ensembles
- Quicker construction, runtime $\Rightarrow$More time for understanding,refinement

Individual-Based Models

- Better fidelity to many dynamics
- Stronger support for highly targeted policy planning
- Ability to calibrate to & validate off of longitudinal data
- Greater heterogeneity flexibility
- Better for examining finer-grained consequences
  - e.g. transfer effects w/i pop.
  - Network spread
- Simpler description of some causal mechanisms

Key Needs Motivating Individual-Based Modeling

- Need to calibrate against information on **agent history**
- Need to capture **progression** of agents **along multiple pathways** (e.g. co-morbidities)
- Wish to characterize **learning by and/or memory** of agents based on experience, or **strong history dependence** in agents
- Need to capture distinct **localized perception** among agents
- Seeking to intervene at points in, change behavior on, explain phenomena over or explain dynamics **across networks**
- Seek **distinct interventions for** many **heterogenous categories**
- Need to **capture impact** of intervention across **many categories**
- When it is much simpler to **describe behavior at indiv. level**
- Seek **flexibility** in exploring different **heterogeneity dimensions**
- **Needs of stakeholders** to engage with individual-based models

Key Needs Motivating Aggregate-Based Modeling

- Need to **execute quickly** (e.g. for user interaction)
- Describe/understand system behaviour across **all possible values for parameters**
- Seeking to mathematically **analyze the model** (e.g. to determine location or stability of equilibria)
- Need to **calibrate to match lots of data**
- Want to **use mathematical tools**, such as control theory or proofs
- **Desire of stakeholders** to work at higher level
- Behavior for different subgroups **differs only in degree**
- No recourse to **software engineering knowledge**
- **Lack of detailed knowledge** of network structure/ individual-level behaviour/Individual-level data

Simpler Causal Description

- Understanding of *individual* behavior sometimes exceeds that of collective behavior
  - Response to locally visible incentives
  - Company's response to competition
  - Young person's response to peer pressure
  - Individual's response to scarcity of good
- Sometimes it is very difficult to derive *a priori* the aggregate dynamics resulting from individual behavior
- Individual model can be simpler, more transparent

Fidelity to Dynamics

- Adequate characterization of system's causal processes may require fine-grain representation
  - Rich heterogeneity
  - Learning and adaptation
  - Response to local incentives
  - Memoryful processes
  - Behavior over persistent networks
- Aggregate behavior is *not* necessarily the same as |Population|*(Behavior of "average" individual)
- May be able to calibrate an aggregate model to results of individual-level model *post hoc*

Example of Concern: History Information

- Heterogeneity with respect to individual history can be highly important for future health
  - Whether vaccinated
  - *in utero* exposure
  - Degree of glycemic control over the past decade
  - Exposure to adiposity
  - Previous exposure to a pathogen
- In some areas of health, we have access to longitudinal data that provides information on individual historical trajectories.

Capturing History Information

- Individual based model
  - Both discrete & continuous history information can be readily captured
    - Categorical/discrete: State (in statechart) or variable
    - Continuous: Variable
  - Readily able to capture records of trajectories
- Aggregate model
  - Categorical/discrete: Limited discrete history information can be captured by disaggregating stocks
    - Curse of dimensionality provides tight limits on # of aspects of history can be recorded
  - Continuous: Almost always infeasible
  - Very complex to provide distributions of trajectories (via convolution of potentially changing PSFs of stocks)

Longitudinal Fidelity: Individual-Based Models

- An individual-based model provides easily accessible *cross-sectional* and longitudinal descrip. of system state
  - The system state at a particular moment in time is cross-sectional
  - By following & recording the trajectories of particular individuals, we can obtain longitudinal description
- In Calibration & validation, we can do rich comparison of both longitudinal and cross-sectional descriptions against available point or time-series data
  - It is in principle possible to have a model that accords with cross-sectional data, but which is at odds longitudinally

Longitudinal Fidelity: Aggregate Models

- An aggregate model provides an ongoing series of *cross-sectional* descriptions of system state
  - In Calibration & validation, we can do rich comparison of these cross-sectional descriptions against available point or time-series data
  - Because the model does not track individuals, we generally cannot explicitly extract model longitudinal trajectories from the model for comparison with historical data we have longitudinal trajectories

Aggregate Models & Trajectories

- While they may not be easy to study explicitly, aggregate models do impose some assumptions about the trajectories of individuals
- This reflects the assumption of a Markovian system: An aggregate model will assume that the placement of an individual at a particular stock in the model adequately summarizes all the historical information needed to describe future dynamics
- While it is somewhat awkward to do, we can test the longitudinal data at different particular components to see how well it holds up to Markov

Example of Markovian Concern

- For example, such a model assumes that the route of entry to a stock is independent of the route of exit
- E.g. If in longitudinal data we don't see independence between routes of entry to a model stock & routes of leaving that stock, that feature of the system may be poorly approximated by that model
  - In some cases, this could be of concern

Shortcomings of Aggregate Comparisons

- If we find that aspects of the data are Markovian with respect to model stocks we can be hopeful about our structure
- Common problems
  - Due to attribute-based disaggregation, a model that incorporates all necessary historical information is too big
  - We may not have data on transitions through particular model stocks – and thus cannot test if it adheres to Markovian assumptions with respect to those stocks
  - We cannot easily compare longitudinal model predictions vs. historic data (see next slide)

Comparisons of Model & History that are Difficult in an Aggregate Model

- Proportions of people with certain history characteristics (e.g. fraction of women who develop T2DM who have had 2 or more bouts of gestational diabetes, those with a certain duration of time separating TB infection sand active TB)
  - Can be very valuable for calibration
  - This is critical for assessing model accord with observed effect size (Relative Risk/Odds ratio)
- Model vs. historic trajectories (e.g. for timing of some transitions) for people with certain history characteristics

Example of Additional Information from Longitudinal Data

- Consider trying to distinguish pairs of situations
- e.g.: Smoking
  - Situation 1: One set of people quit & stay quit as former smokers, another set remain as current smokers
  - Situation 2: The entire set of people cycle through situations where they quit, relapse & repeat
- These two situations have very different health consequences
- We'd probably choose vary different sets of interventions for these two situations
- Similar examples are easy to imagine for obesity, STIs, TB, glycemic control & diabetes, etc.

Trajectories Summary

- If either or both of the following is true….
  - You have significant longitudinal information you'd strongly like the model to match
  - You have good reason to think that trajectory history has important consequences for health
- Then you should build a model that captures this history information
  - By disaggregating stocks, you can capture limited *discrete* history information in an aggregate model (e.g. whether a person was exposed *in utero*, Time Since Quit for FS, whether a woman has had a history of gestational diabetes)
  - There is significantly greater flexibility for collecting continuous or discrete history information for guiding individual dynamics & for calibration/validation comparison to historic longitudinal data

Calibration & Validation Comparisons

- We can compare statistics from histories in an individual-based model to statistics from actual histories
  - See if matches non-markovian nature
  - See how matches distribution of times

Recall: Importance of Heterogeneity

- Heterogeneity often significantly impacts policy effectiveness
  - Policies preferentially affect certain subgroups
    - Infection may be maintained within certain subgroups even though would tend to go extinct with random mixing in the entire population
  - Policies alter balance of heterogeneity in population
    - Shifts in the underlying heterogeneity can change aggregate population statistics
  - Given a non-linear relationship, inaccurate to use the mean as a proxy for whole distribution
- **Assessing policy effectiveness often requires representing heterogeneity**
- ***Flexibility* in representating heterogeneity is hard to achieve in aggregate (coarse-grained) models**

Impacts of Heterogeneity on Policy Effectiveness

- Value of breast cancer detection (Park & Lees)
- Impact of airbags on deaths (Shepherd&Zeckhauser)
- Value of hernia operations (Neuhauser)
- Impact of cardiovascular disease interventions (Chiang)
- Controlling blood pressure (Shepherd&Zeckhauser)
- Effectiveness of mobile cardiac care unit (Shepherd&Zeckhauser)
- Value of breast cancer treatment (Fox)
- Taeuber paradox (Keyfitz)

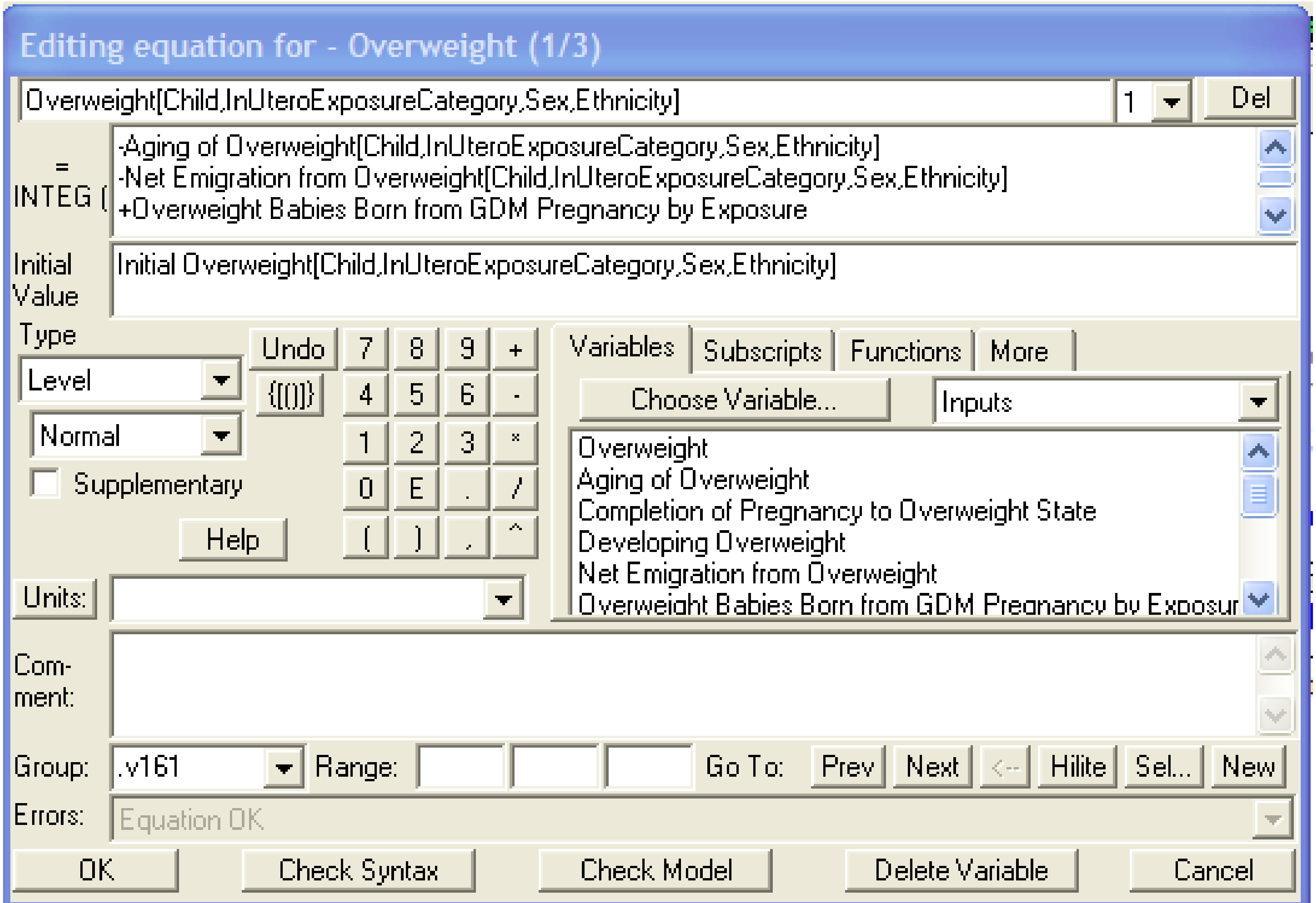
Frequent Heterogeneity Concerns

- No clear boundaries at which to divide people up into discrete categories
- Many dimensions of heterogeneity simultaneously
  - Capturing state with respect n factors requires n dimensions of heterogeneity!
- Need to consider progression along many dimensions simultaneously

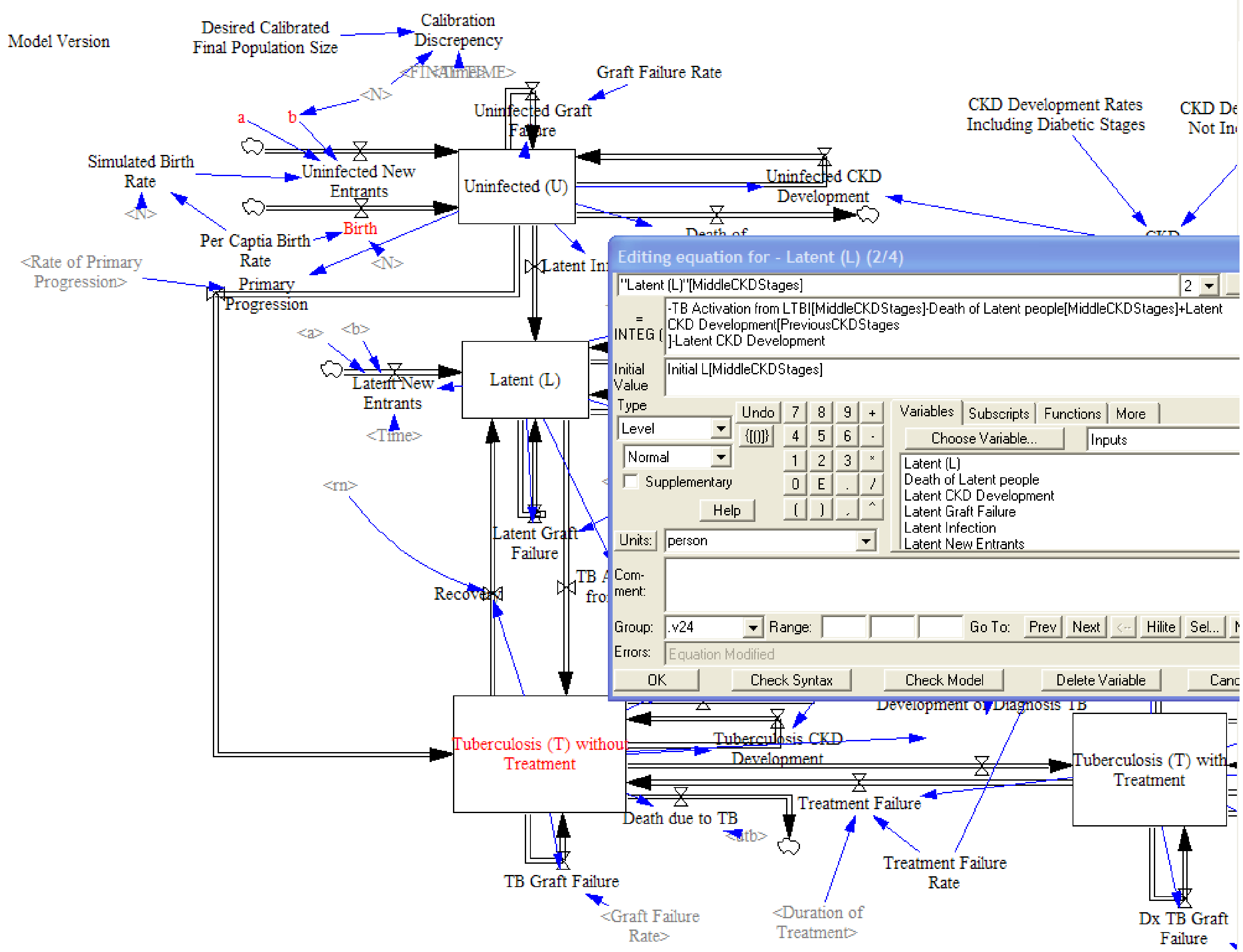
Challenges for Aggregate Model Formulation: Heterogeneity

- Two aggregate means for representing heterogeneity are limited:
  - Attribute-based disaggregation
    - Need *n* dimensions to capture individual state with respect to *n* factors
    - **Poor (geometric) scaling to large # dimensions of heterogeneity**
    - Global structural, equation changes required to incorporate new heterogeneity dimensions
    - Awkwardness in stratifying
  - Co-flows
    - Efficient and precise but highly specialized

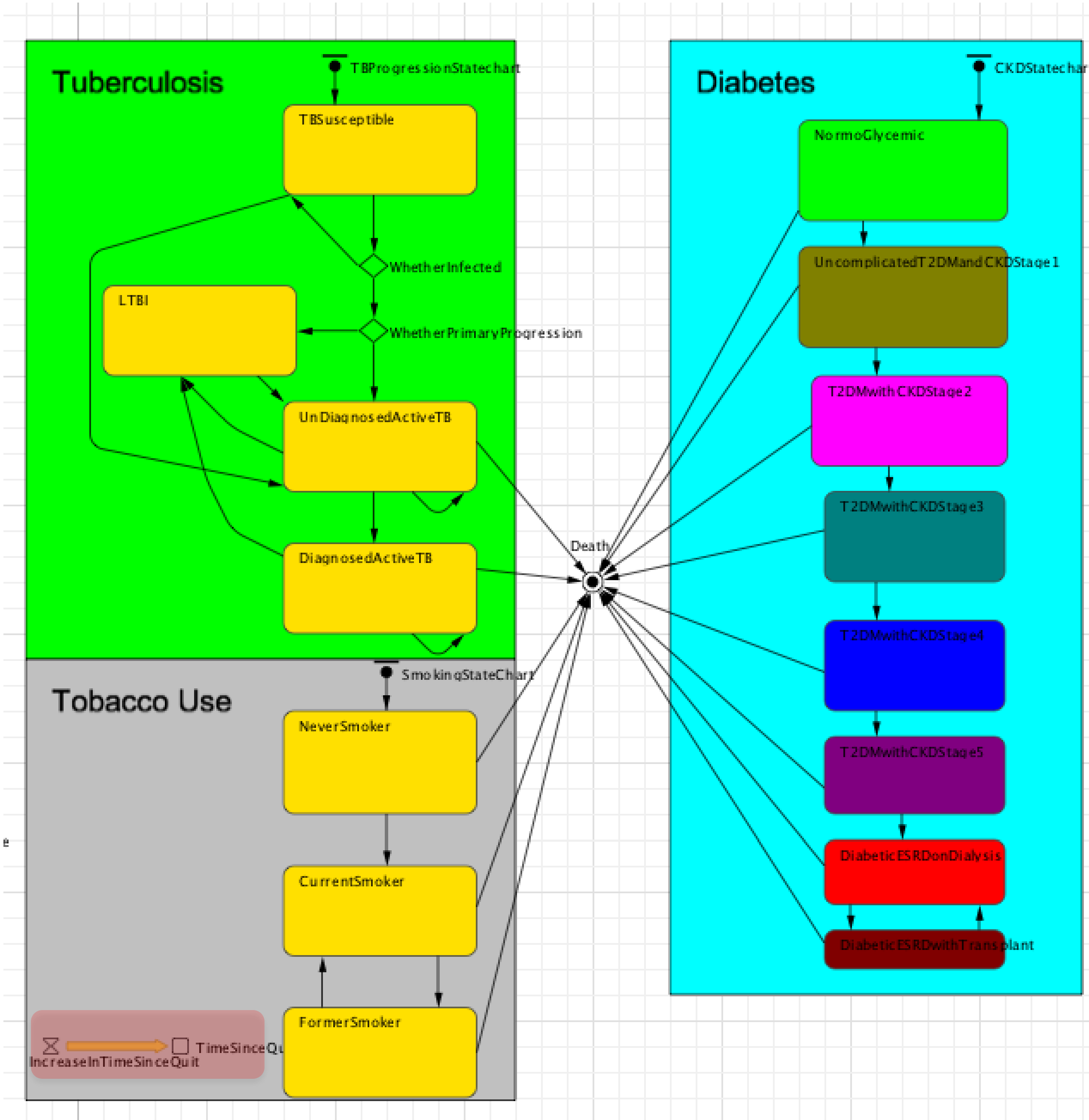
Fragility of Multi-Dimensional Subscripting



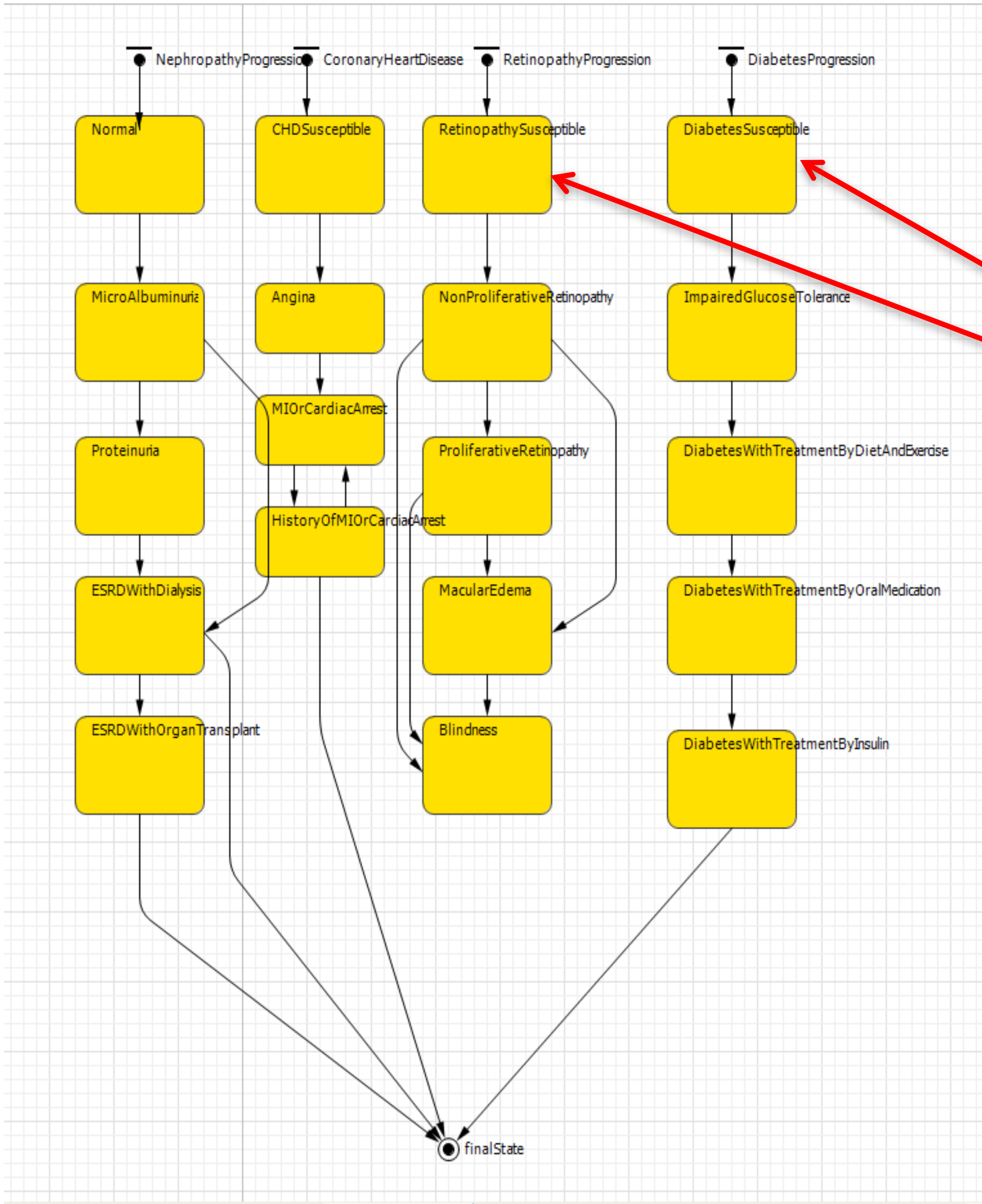
Combinatorial Subscripting: Multi-Dimensional Progression



Parallel Transitions



Parallel State Transition Diagrams



A person is in some particular state with respect to each of these (condition specific) state transition diagrams

This requires representing combinations of possibilities in an aggregate model

Capturing Heterogeneity in Individual-Based vs. Aggregate Models

- Consider the need to keeping track a new piece of information for each person (with d possible values)
  - E.g. age, sex, ethnicity, education level, strain type, city of residence, etc.
- Aggregate Model: Add a subscript
  - This multiplies the model size (number of state variables into which we divide individuals) by *d*!
- Individual based model: Add field (variable/param)
  - If model already has c fields, this will increase model size by a fraction 1/c.

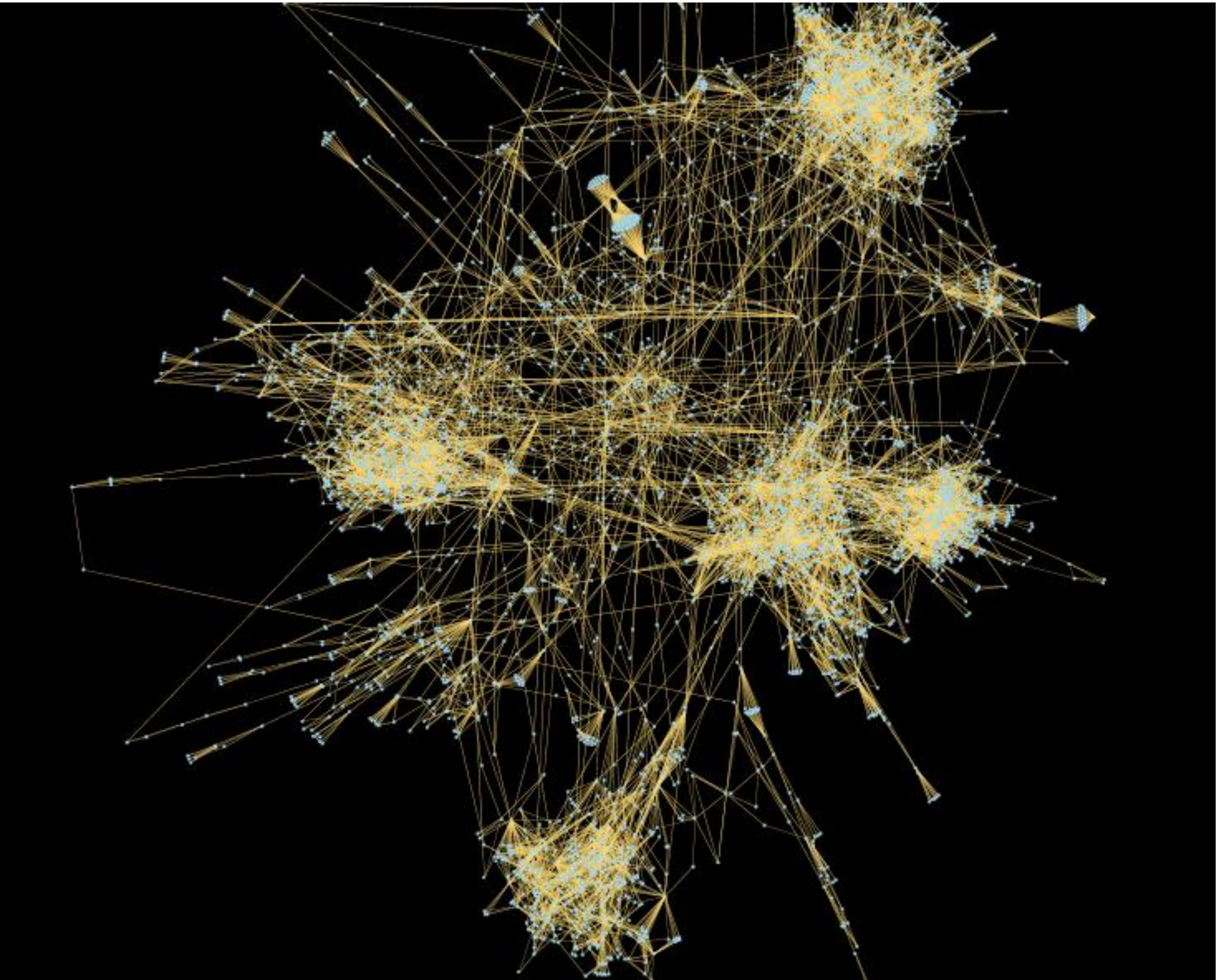
Desired: Flexibility in Representing Heterogeneity

- It is desirable to capture heterogeneity in a flexible fashion.
- More judicious exploration of whether to represent heterogeneity
  - Examine whether some observed covariation might simply be due to colinearity
    - Represent added heterogenity dimensions with no causal interaction, see if model covariations matches what is seen in external world
      - e.g. represent age in a TB model, see if rates of LTBI by age in the model match age-specific infection rate observations
  - Try adding in new dimension of heterogeneity & effects, and see if has impact that is both substantive & plausible

Challenges for Model Formulation: Persistent Interaction

- Network topologies can affect qualitative behavior
- Aggregate representations of network structure are expensive and awkward
- IBM permit expressive, efficient characterization of both dense & sparse networks
- While percolation over many topologies can be simulated in aggregate models, parameter calibration often requires finer-grained simulation

Network Clustering



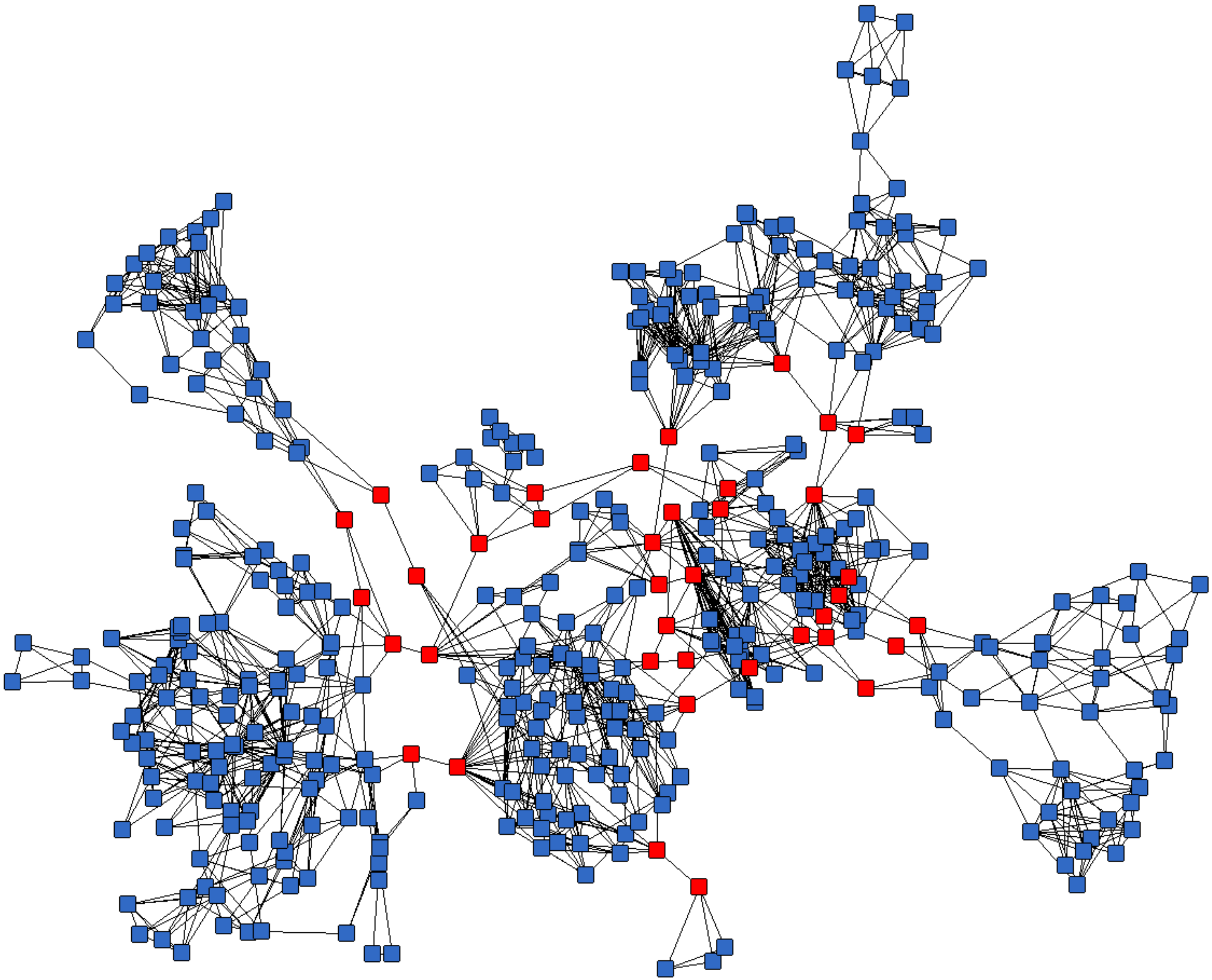
Preliminary case contact Network

Restricted to nodes of degree 2+

Clusters distinctive by
- Geography
- Ethnicity

- Data extraction: A. Al-Azem

Identifying Bridging Individuals



- Preliminary case contact network
- Restricted to nodes of degree 2+
- Data analysis & image: A. Al-Azem

Slides Adapted from External Source

Redacted from Public PDF for Copyright Reasons

Multi-scale Phenomena

- Frequently we are concerned about phenomena on a variety of scales
  - Aggregate societal & policy level
  - Institutional level
  - Individual level
  - Intra-institutional level

Slides Adapted from External Source

Redacted from Public PDF for Copyright Reasons

Finer Grained Policy Planning

- In the presence of networks or non-well-mixed populations, big difference in effects of targeted interventions
- e.g.
  - Targeted intervention within scale-free network
  - Impact of incentives on competition and cooperation
  - Impact of road structure on traffic jams

Parameterization & Calibration

- Individual-based models have many parameters
  - Estimating all of the parameters can require much effort
  - Calibration generally underdetermined (large # of possible sets of parameter values that could calibrate well)
  - May need to make simplifying assumptions
- Pronounced individual-level stochastics frequently require Monte-Carlo calibration

Individual-Based Model Performance Scaling

- Performance varies with population size
  - Large populations impose high computational resource demands
  - Scaling can be superlinear (e.g. $O(n^2)$ connections to consider)
  - This can frequently lead to simulations taking minutes at the least, commonly hours or even days
- Desire to characterize stochastic nature of individual-level behavior typically requires Monte Carlo approaches
  - This can lead to days or weeks to complete

Memoryless vs. Memoryful Processes

- ODE models can adequately capture only *memoryless transition processes out of a stock*
  - Stocks treated as "well-mixed": Transition probability does not depend on residence time
  - Memoryful processes can be approximated, but requires changing model structure to reflect a simple functional relationship (n^{th} order delays)
- IBM can record residence time in state & allow probability of transitions to depend on this

Slides Adapted from External Source

Redacted from Public PDF for Copyright Reasons

Individual vs. Aggregate Models: Necessary Tradeoffs

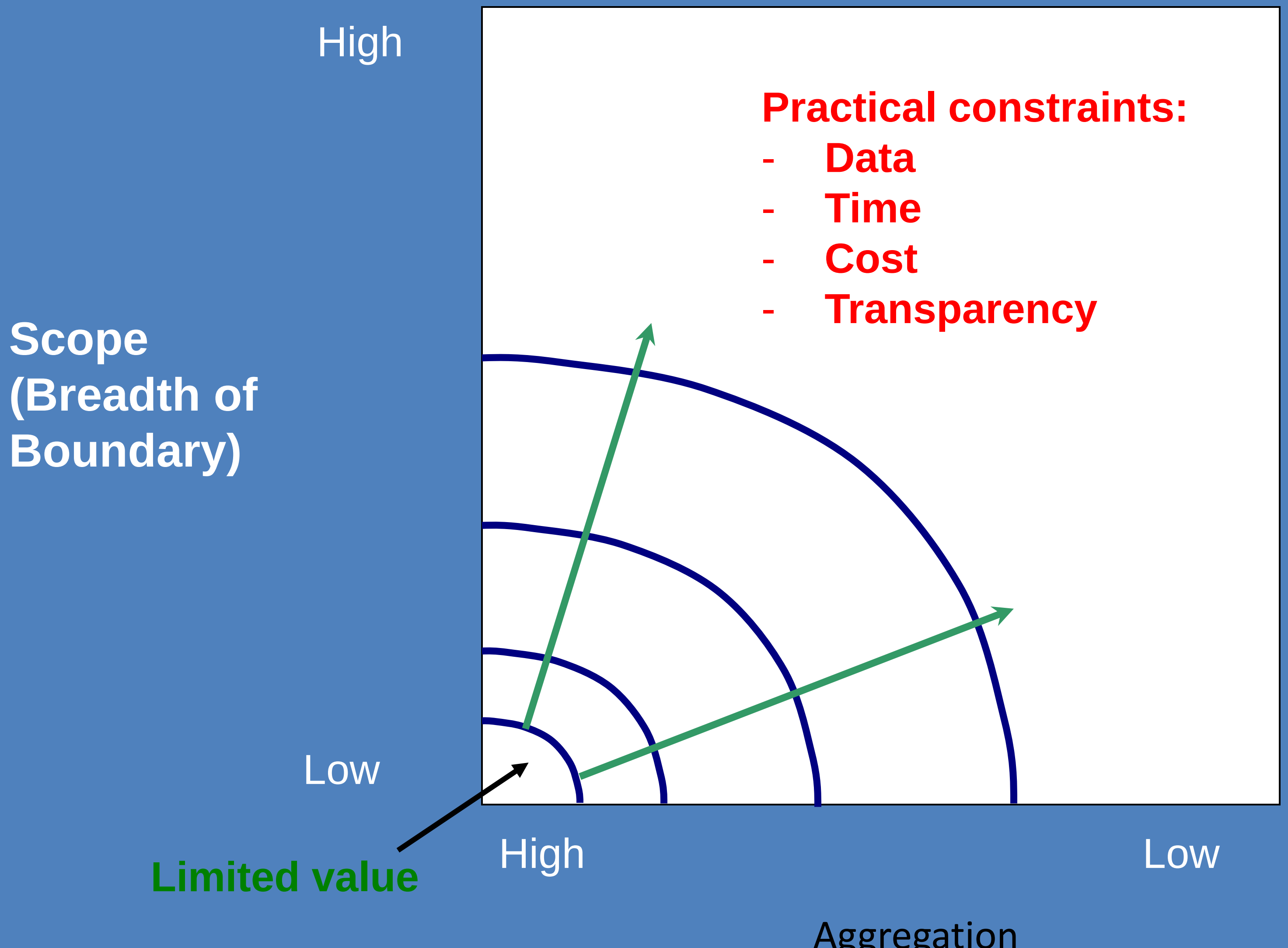
	Transition Generality	Network Representation	Calibration	Performance Issues				Capturing Learning/ Adapation
				Basal	Scaling with Population	Scaling with Heterogeneity	Need for Stochastics /Monte Carlo	
Individual Models	++	++				++		++
Aggregate Models		+	++	++	++		+	

- Both individual-level and aggregate modeling have *inherent* and non-trivial *tradeoffs*
- Both approaches likely to retain strong appeal in systems modeling

Areas of Advantage of Individual-Based Modeling

- Examining finer-grained consequences
  - Network spread
  - Transfer effects within population
  - Detailed spatial dynamics
  - Effects of population heterogeneity
  - Effects of highly targeted policies
  - Effects of individual-level synergies (e.g. multiple risk factors)
- Simple individual-based description of causal mechanisms
- Sufficient individual-level (distributional) data are available for policy modeling beyond exploratory models

Inevitable Tradeoffs



Agenda

✓ Motivations & Context
✓ Comparing Aggregate & Individual Based Models
✓ Granularity Tradeoffs
- Tools for individual-based modeling
  - Individual-Based Modelers in SD
  - Individual-based models in Agent-Based tools
- Other tradeoffs
- Looking forward

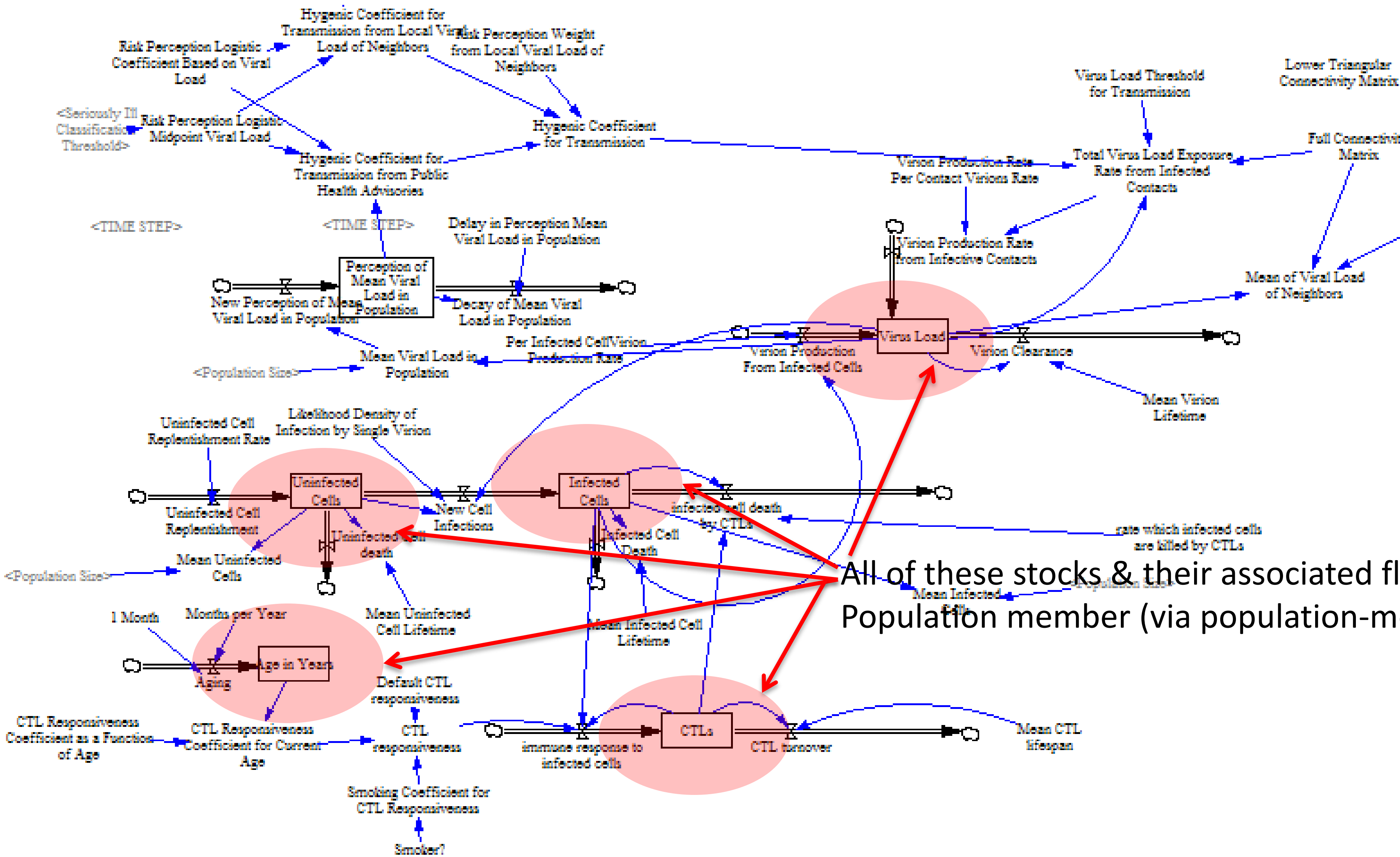
Hallmarks of Complex Systems

- Delays
  - Represented at individual/environment level
  - Generality of rules allows for memoryful stochastic processes
- Nonlinearities
  - Rules expressed with arbitrary algorithms – can encode arbitrary functions of model state
- Stochastics or Uncertainties
  - Many agents’ behaviors will be stochastic

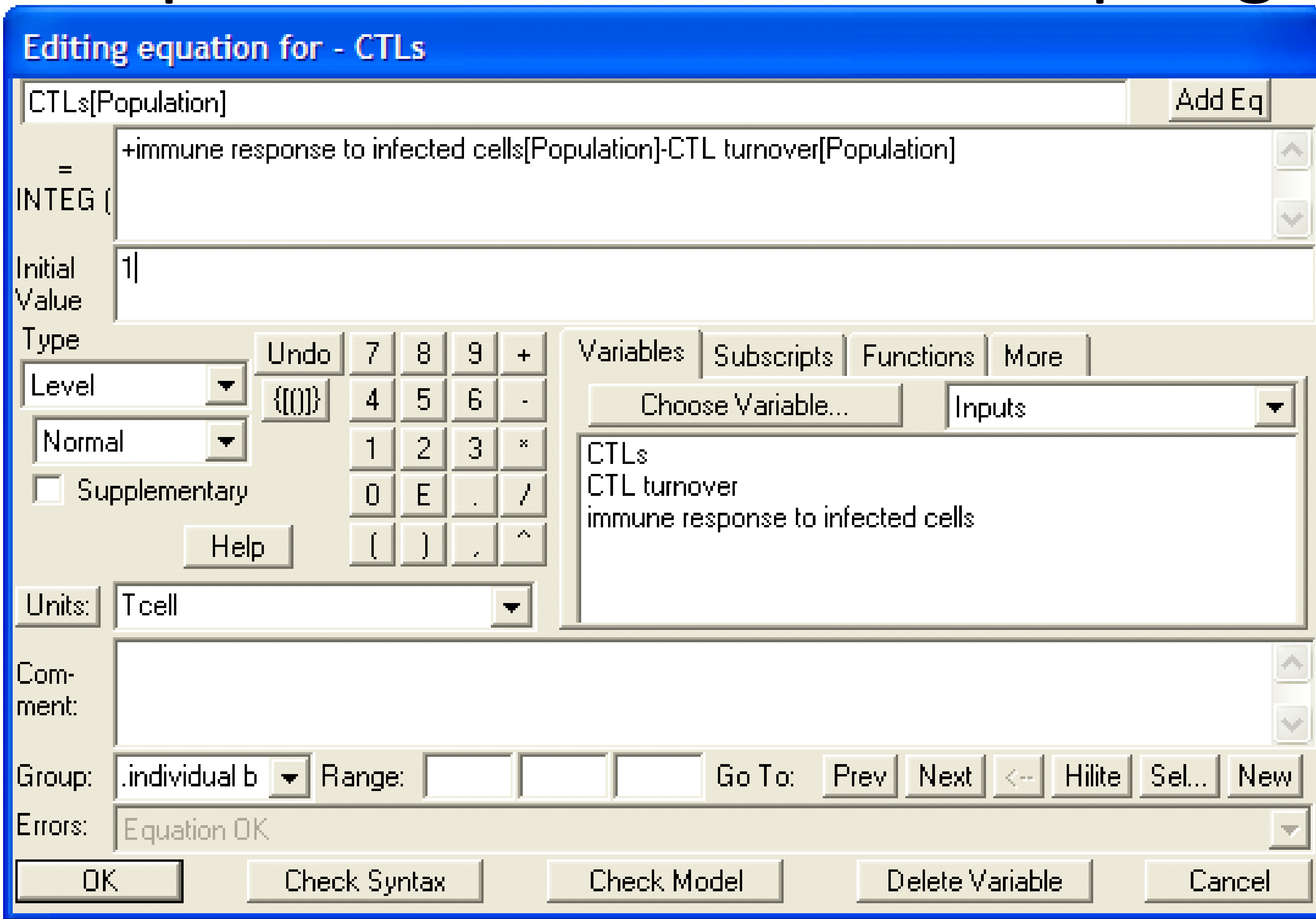
System Dynamics & Individual-Based Modeling

- Individual-based models can be created using
  - Traditional System Dynamics software
    - Small populations:
      - Separate stocks for each individual
      - Hand-drawn connections
    - Larger Populations
      - Subscripting stocks by population member
      - Binary network matrices
  - Stock & flows in other dynamic modeling software
    - e.g. in AnyLogic
  - System Dynamics methodology
    - Feedback-centric reasoning
    - Process-based work

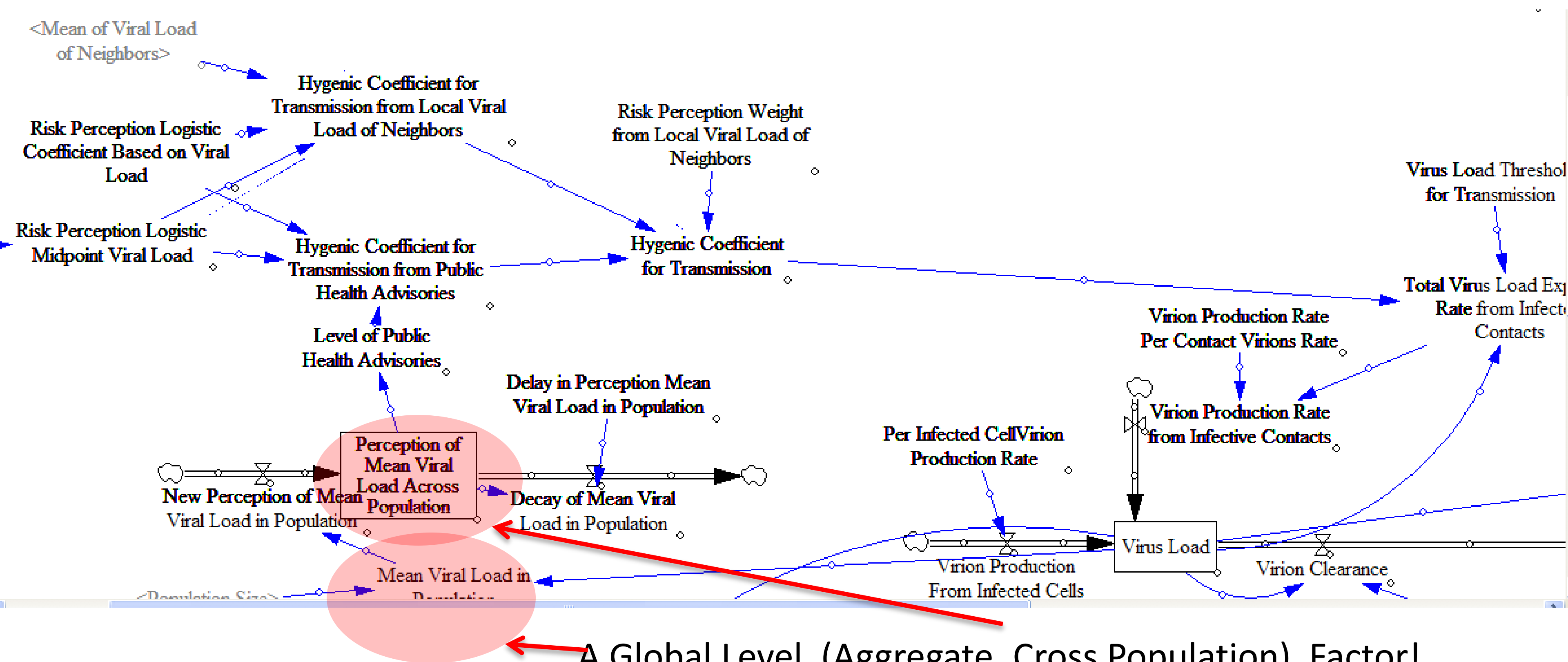
Individual-Based Model in Vensim



Population-Member Subscripting



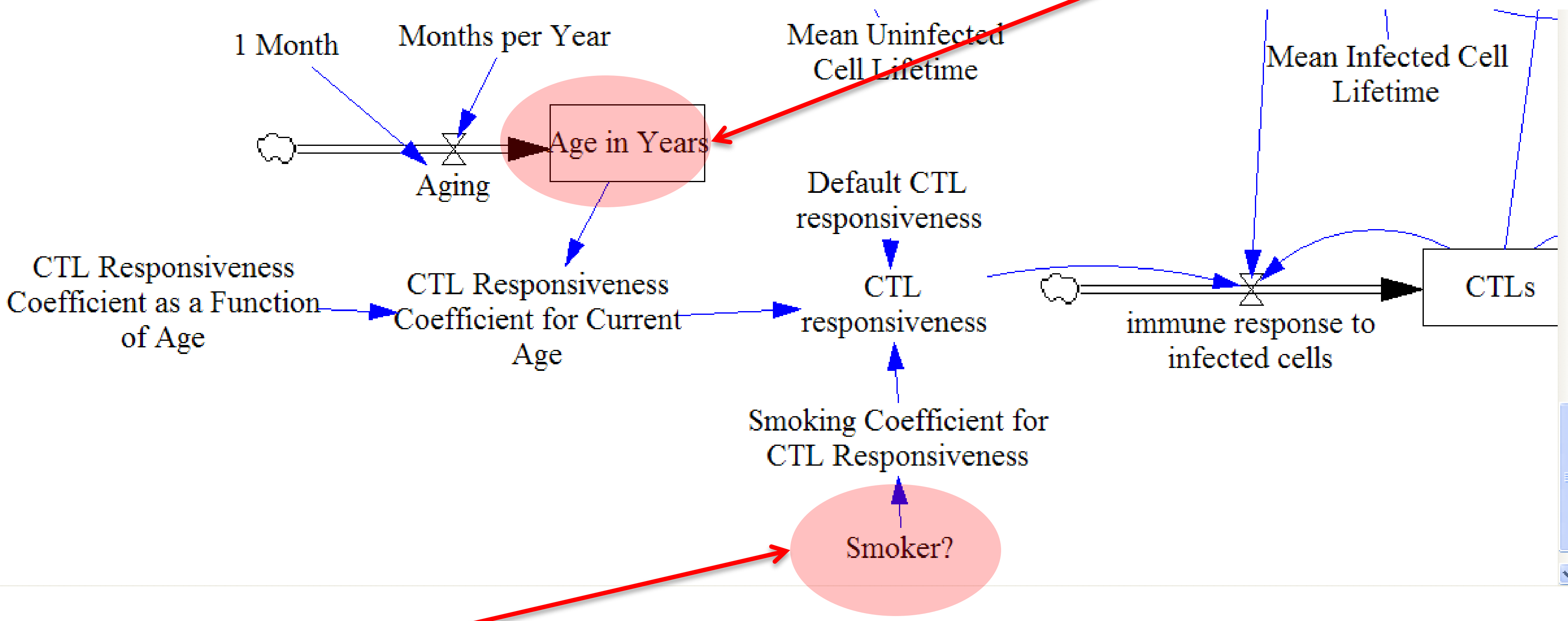
Example Interactions between Global & Local Levels



A Global Level (Aggregate, Cross Population) Factor!

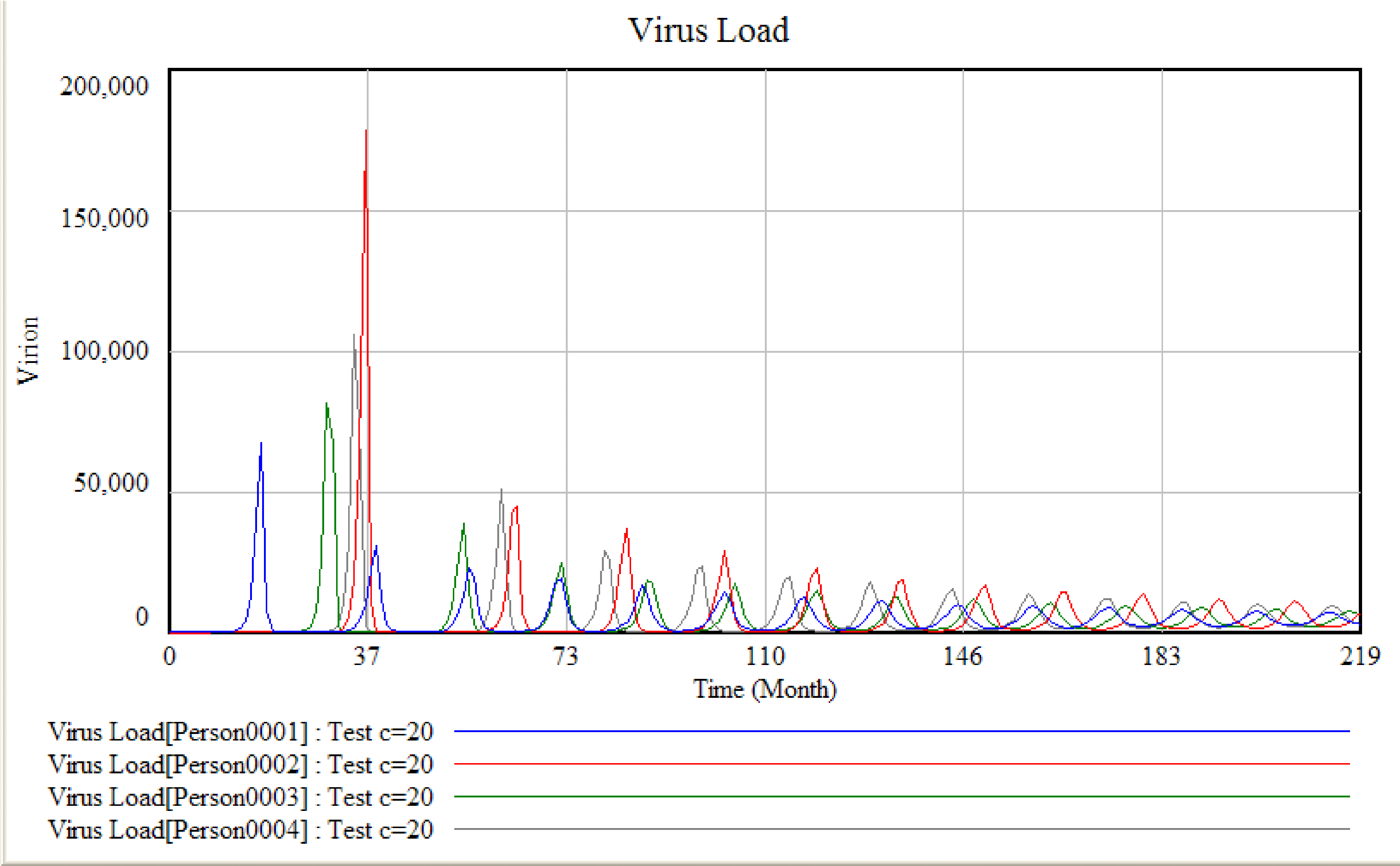
Example Individual-Level Risk Factors

An Individual-Level Risk Factor



Another Individual-Level Risk Factor (here, represented categorically, but we could Represent it as a continuous variable – e.g. cumulative smoke exposure, some estimate of cumulative physiologic damage from smoke, a moving average of smoke exposure, etc.)

Impact of Risk Factors on Individual Dynamics



Population Subscripting Tradeoffs

Advantages

- Conceptually simple
- Can SD tools
  - State trajectory file recording
  - Easy construction, structure visualization
    - No programming
  - Sensitivity analysis
  - Easy to aggregate

Disadvantages

- Difficult to visualize network structure & spread or spatial embedding
- Awkward to realize changing population size

Agent-Based Systems: A Glimpse

- (Current) agent-based model characteristics
  - One or more populations composed of individual agents
    - Each agent is associated with some of the following
      - State (continuous or discrete e.g. age, health, smoking status, networks, beliefs)
      - Parameters (e.g. Gender, genetic composition, preference fn.)
      - Rules for interaction (traditionally specified in general purpose programming language)
    - Embedded in an environment
    - Communicate via messaging
  - Environment
- Emergent aggregate behavior

Array output format

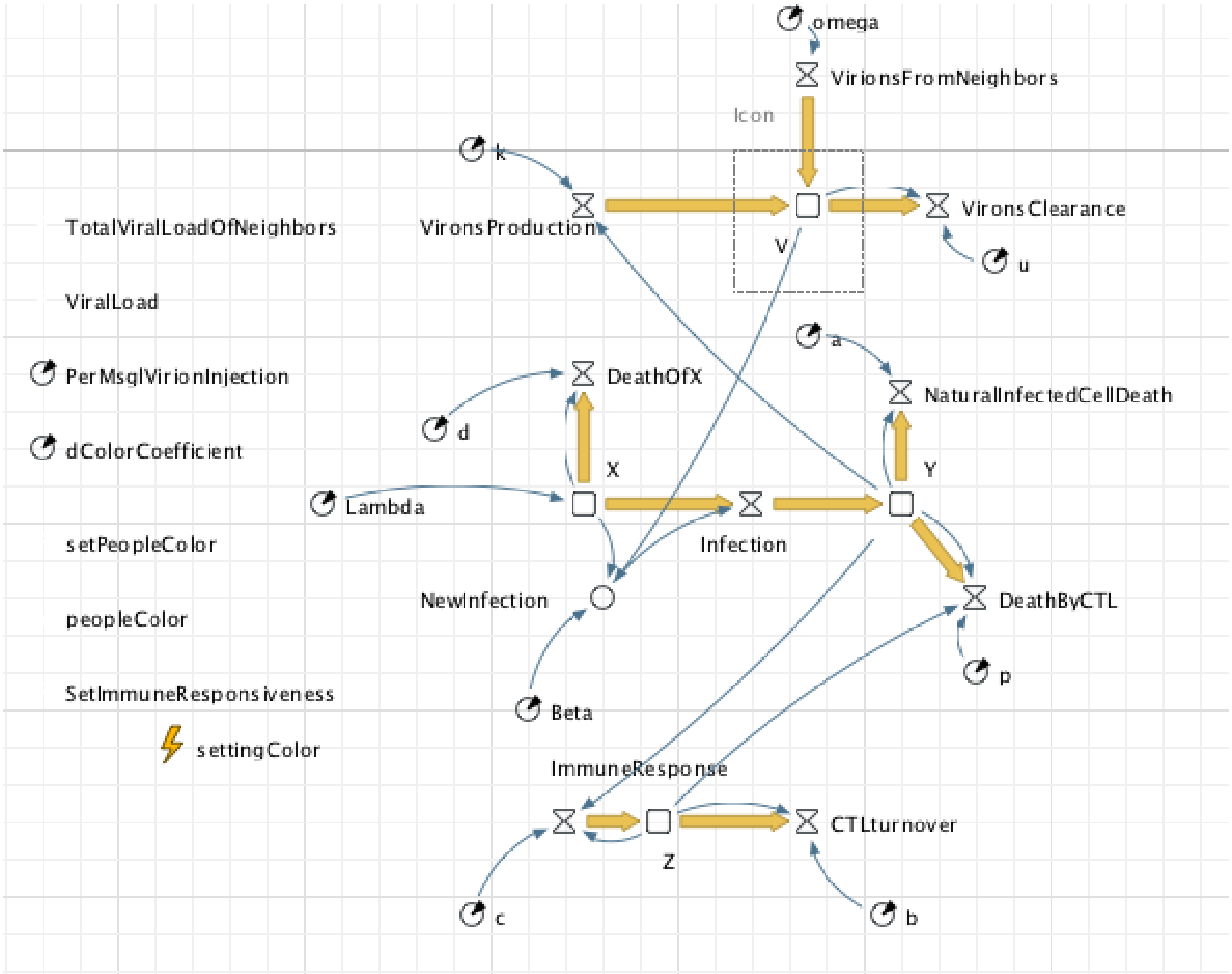
"Full Connectivity Matrix[Population,Population]" at time 263 Runs: Test

Full Connectivity Matrix[Population,Po	Person0001	Person0002	Person0003	Person0004	Person0005	Person0006
Person0001	0	0	0	0	0	0
Person0002	0	0	0	0	0	0
Person0003	0	0	0	0	0	0
Person0004	0	0	0	0	0	0
Person0005	0	0	0	0	0	0
Person0006	0	0	0	0	0	0
Person0007	0	0	0	0	0	0
Person0008	0	0	0	0	0	0

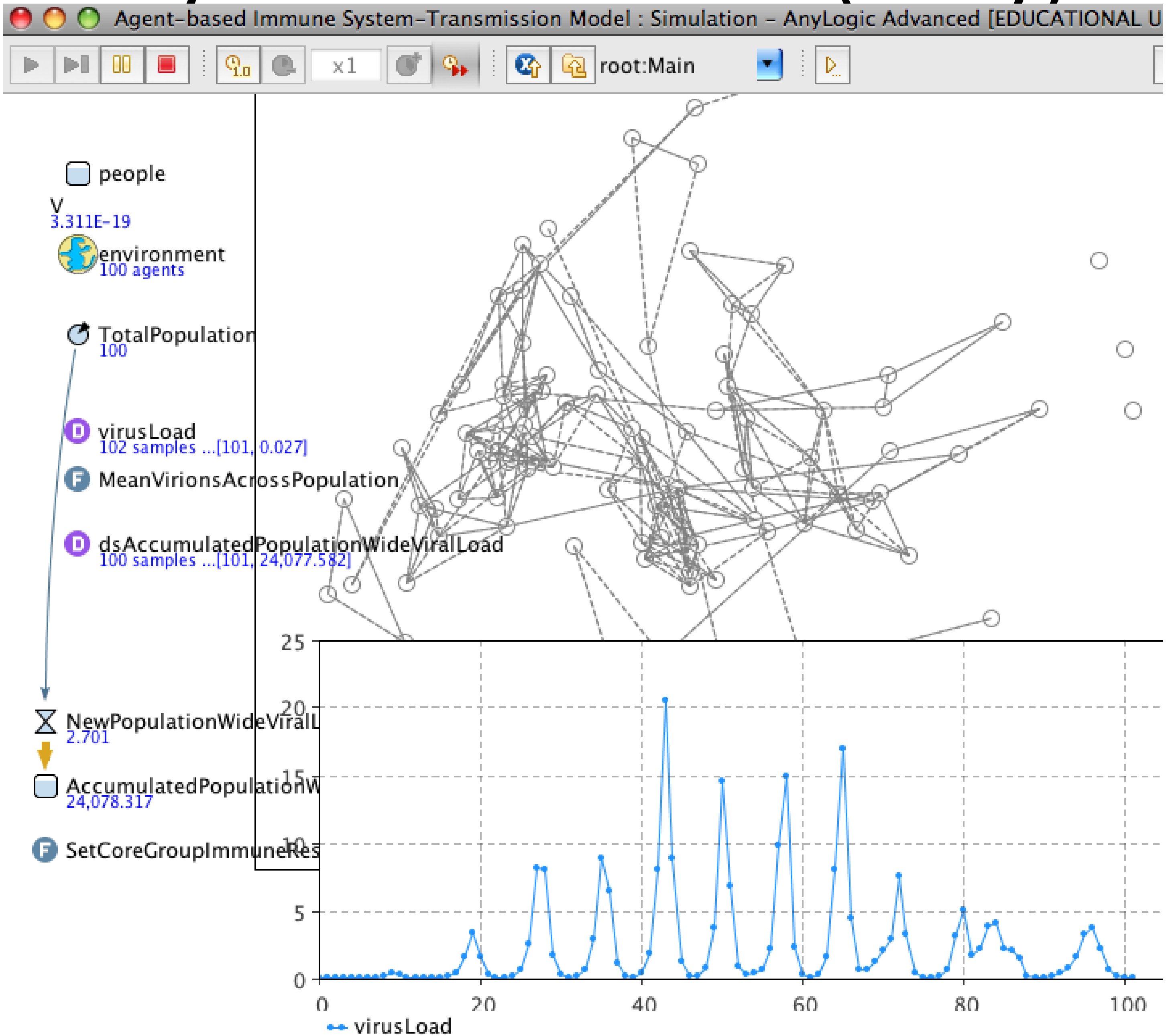
Agent-Based Modeling

- We can capture individuals in many ways
- I view Agent based models (ABM) as a type of individual-based modeling that encapsulates a given individual as a *software object* with
  - Methods
  - Properties
- Objects provide a convenient abstraction for individuals
- Agent-based models currently require writing at least some code in programming languages
- We can formulate SD models w/i agent-based tools
  - I view such models as simultaneously SD & ABM
- We can follow an SD process to build & use agent-based models

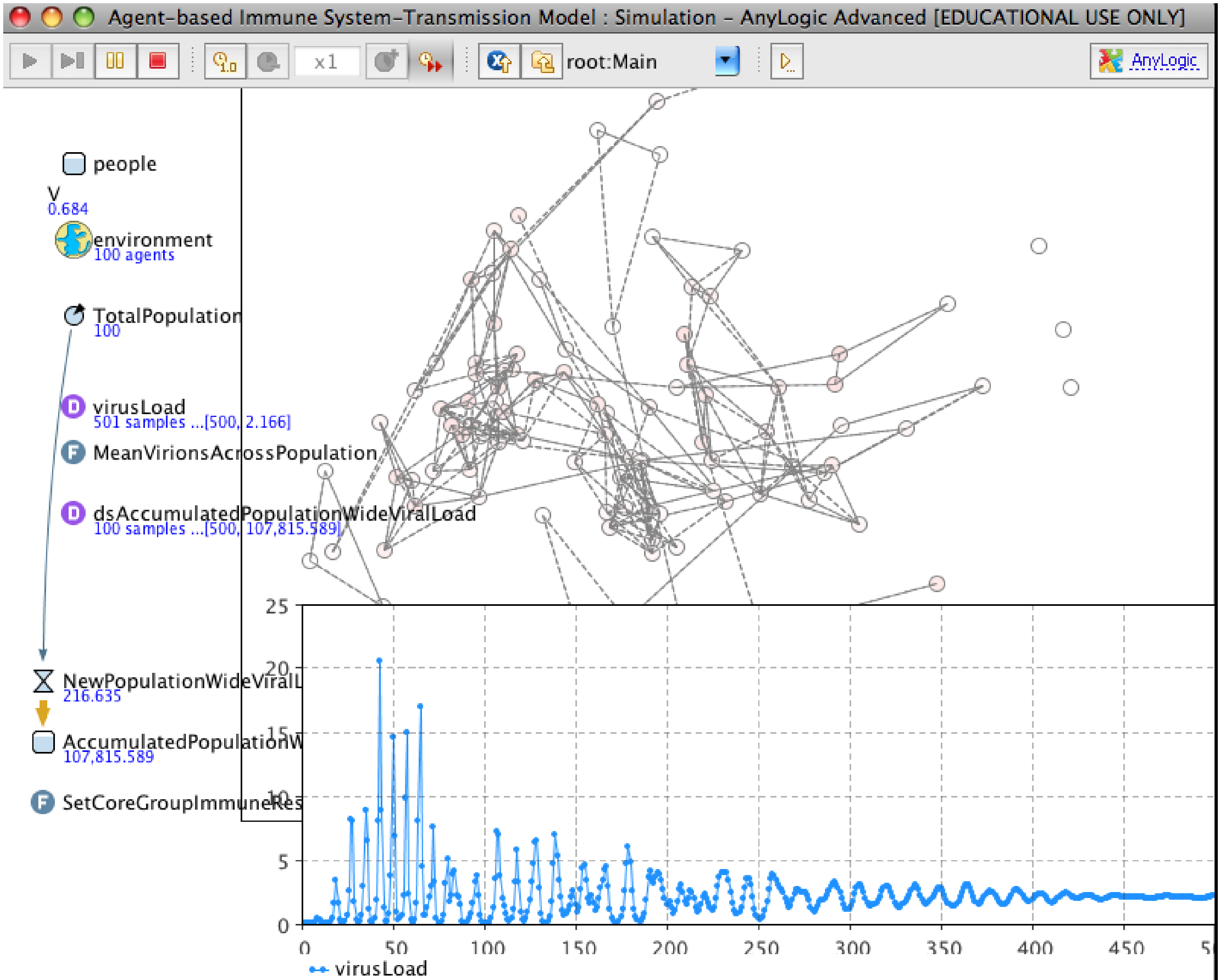
A Model in AnyLogic



Dynamic Behavior (Early)



Steady-State Behavior



Agent-Based Modeling

- We can capture individuals in many ways
- I view Agent based models (ABM) as a type of individual-based modeling that encapsulates a given individual as a *software object* with
  - Methods
  - Properties
- Objects provide a convenient abstraction for individuals
- Agent-based models currently require writing at least some code in programming languages
- We can formulate SD models w/i agent-based tools
  - I view such models as simultaneously SD & ABM
- We can follow an SD process to build & use agent-based models

The (Current) Package Deal

- **ABM (AnyLogic)**
  - Supports individual-based or aggregate
  - No trajectory files
  - Both discrete & continuous rules & states
  - Primarily imperative specification
  - Algorithmic (imperative)
  - Little/No explicit mathematical semantics
  - Modularity mechanisms
  - No metadata
- **Traditional system dynamics packages**
  - Supports individual-based or aggregate
  - Trajectory files well supported
  - Poor discrete rule support
  - Declarative specification
  - Equational notation & reasoning
  - Explicit mathematical semantics
  - Monolithic
  - Limited metadata (unit checks)

Current Package Deal: Modeling Implications (From my Perspective)

	Transparency	Performance	Ease of Creation	Generality	Analyzability/ Understanding	Ease of Parameterization	Ease of Calibration	Accuracy	Model Breadth	Scalability (Population)	Scalability (Heterogeneity)	Modifiability
TSD	++	++	++	+	++	++	++	+	++	++	-	+
ABM	+		+	++		+	+	++			++	+

Current ABM and TSD packages both have important advantages

Central Points: Looking Forward

- Most current differences reflect important but non-essential methodological choices / tool characteristics
- In the long run, these differences will likely lessen and the choice that will remain is that of model granularity
- Both individual-based models and aggregate models will play important roles in system dynamics
- There are good reasons to use all of individual-based models, aggregate models, and hybrid systems