

Individual-Based Models: Introduction, Tradeoffs & Tools

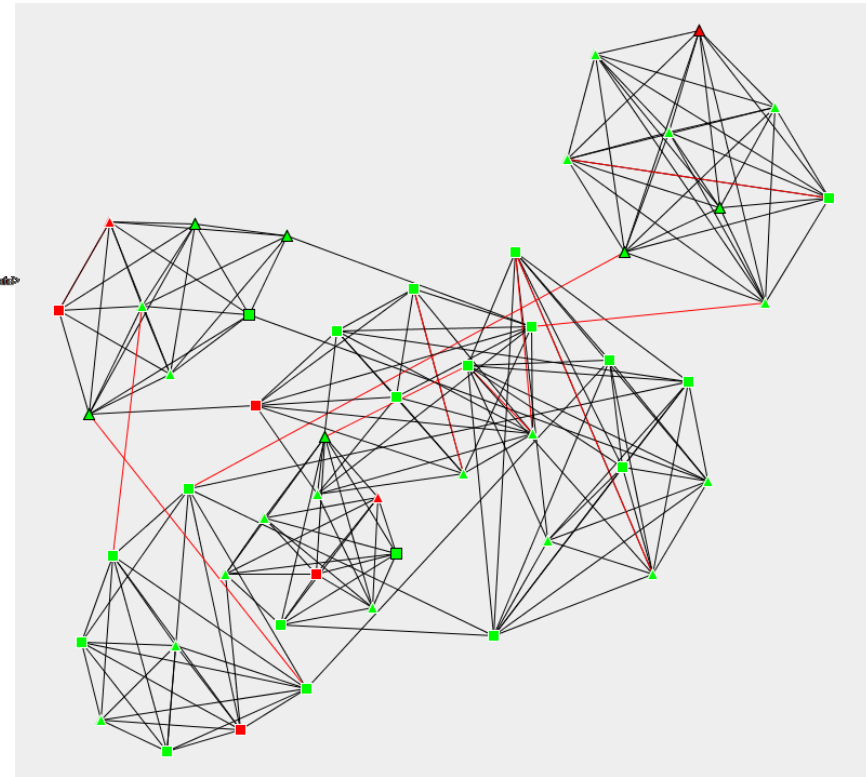
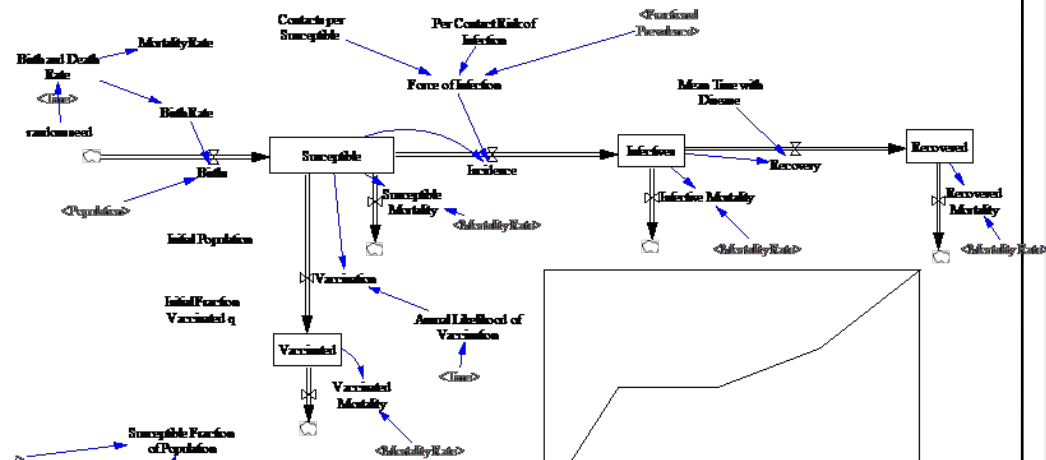
Nathaniel Osgood

October 29, 2009

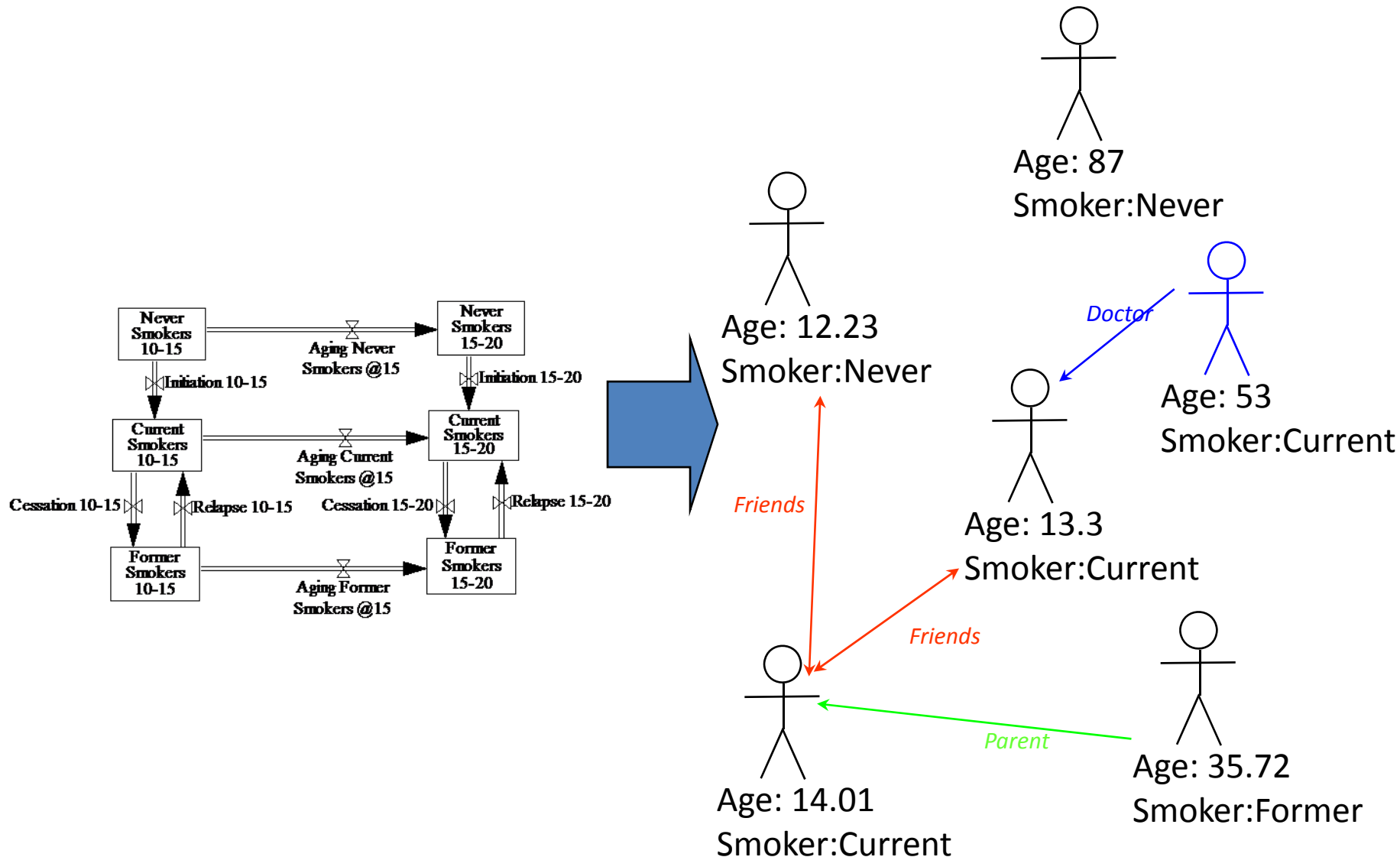
Dynamic Models for Health

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

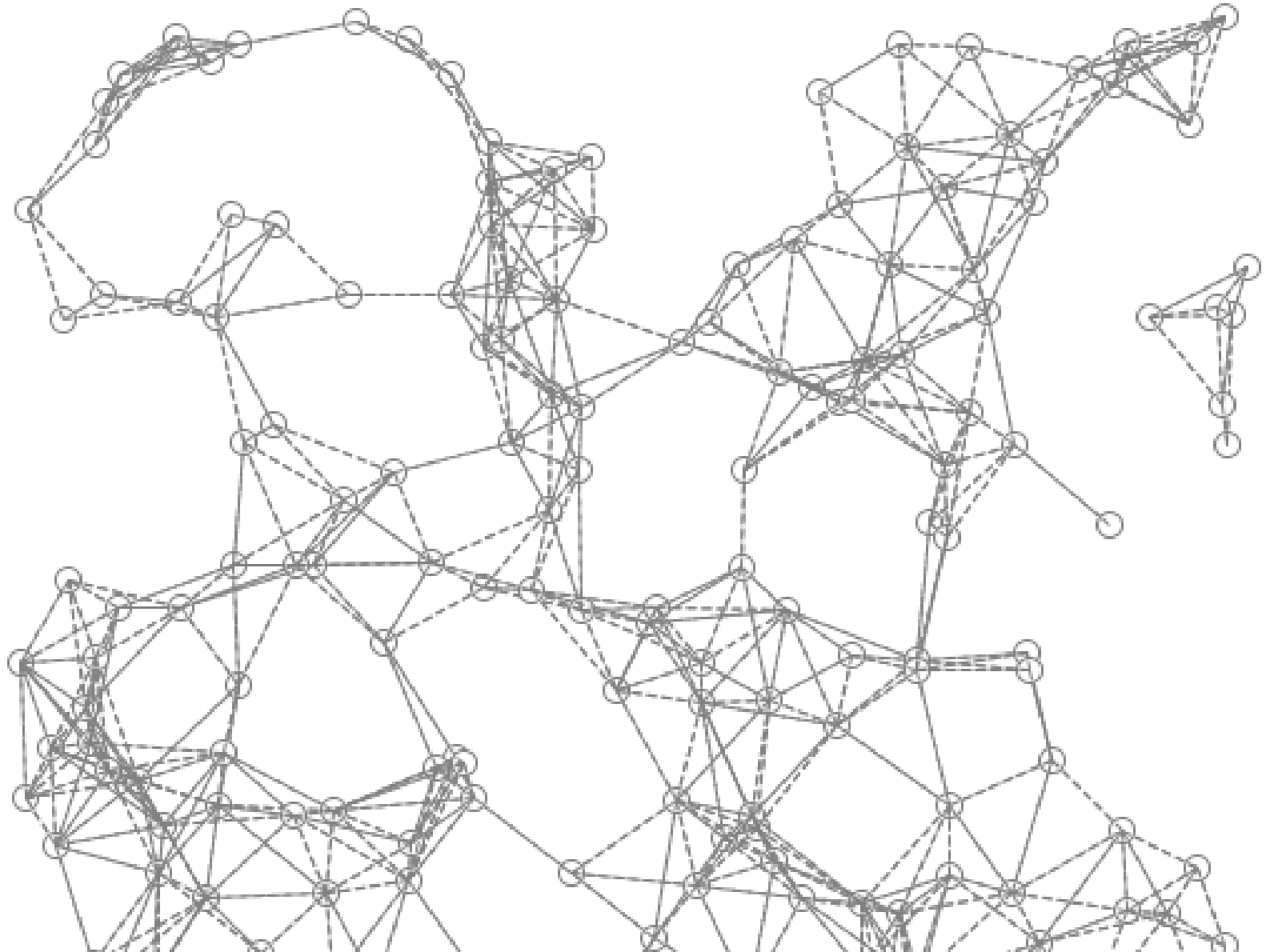
Contrasting Model Granularity



Interacting Individuals



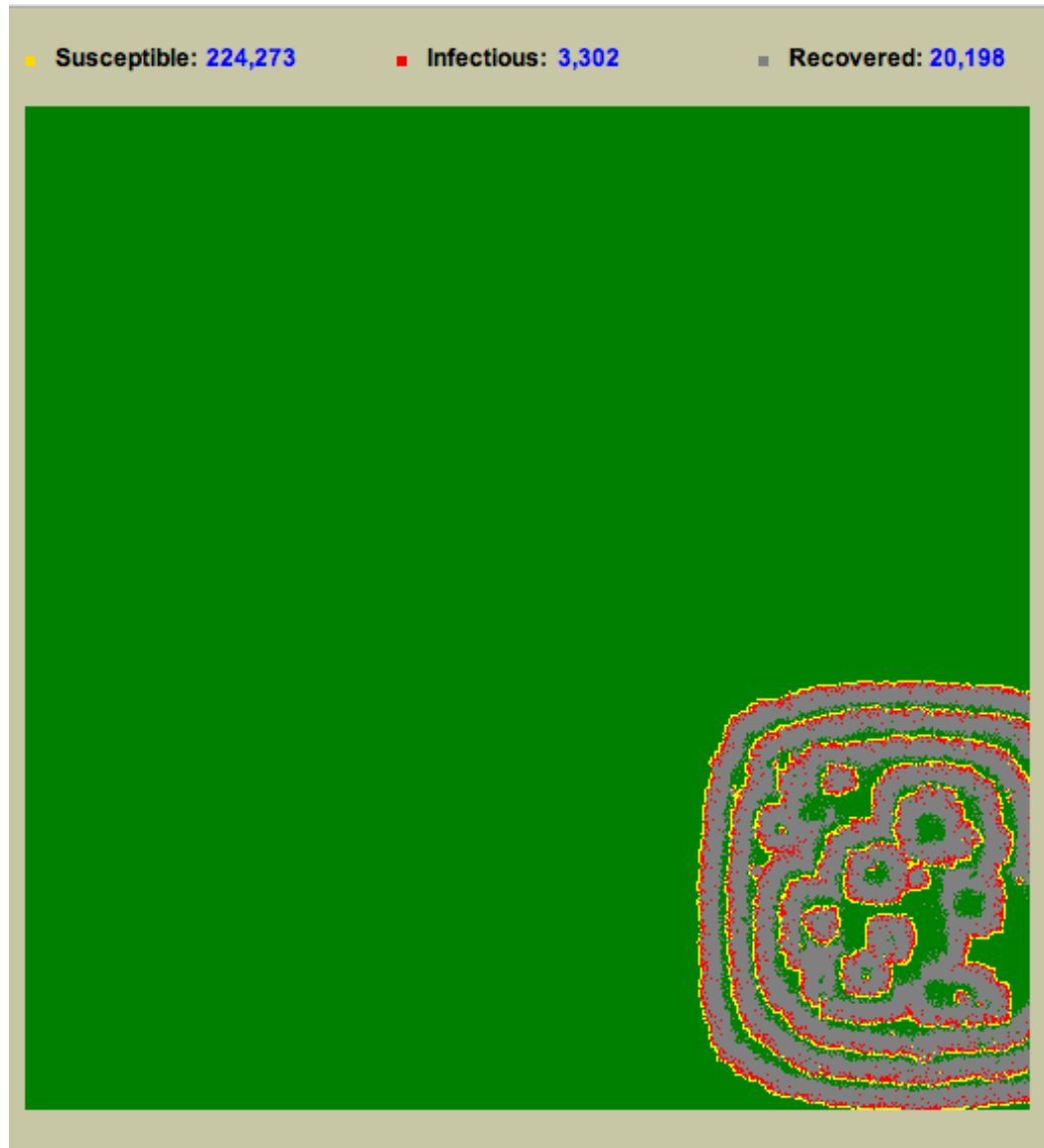
Network Embedded Individuals



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 representing 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

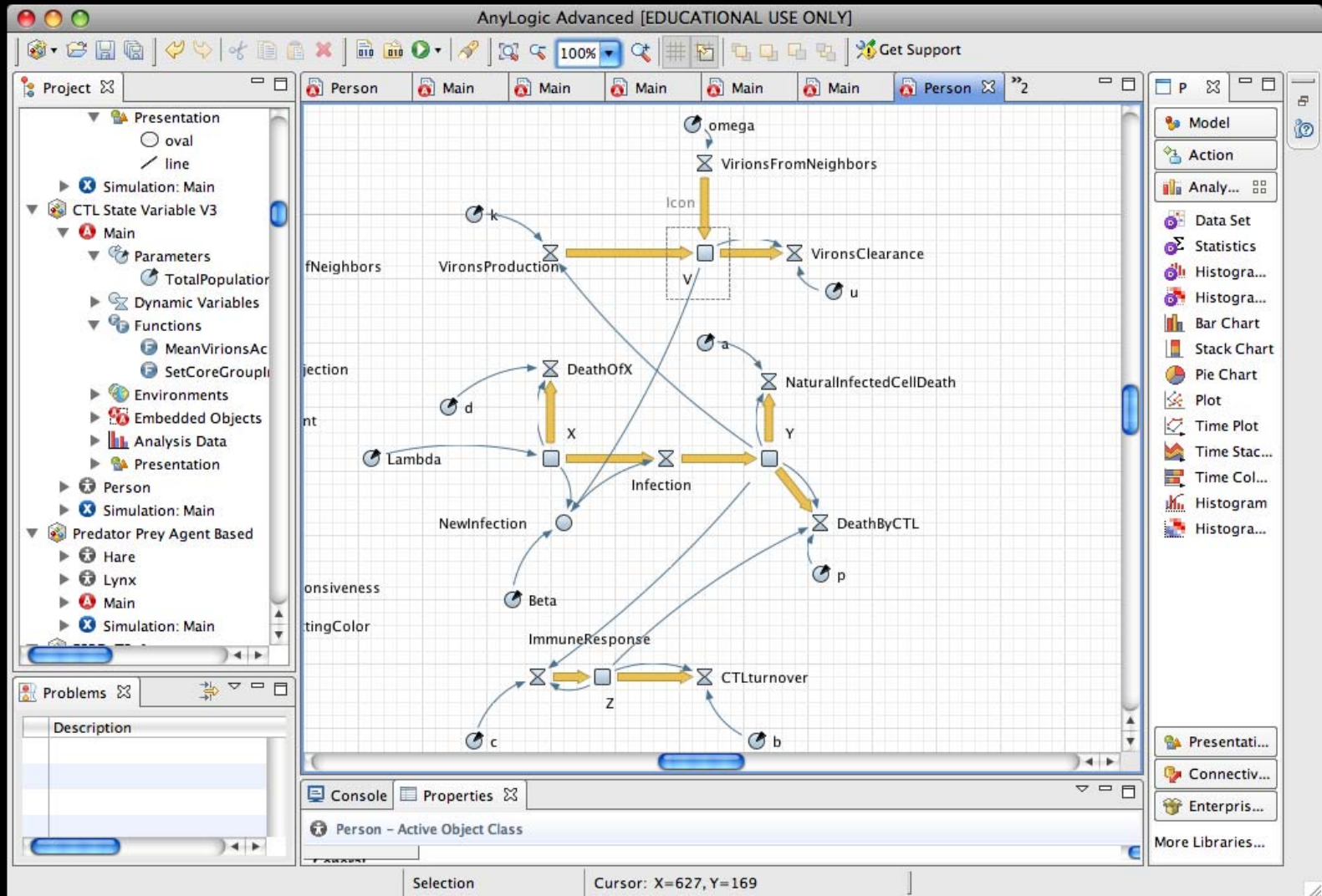
Agenda

- ✓ 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

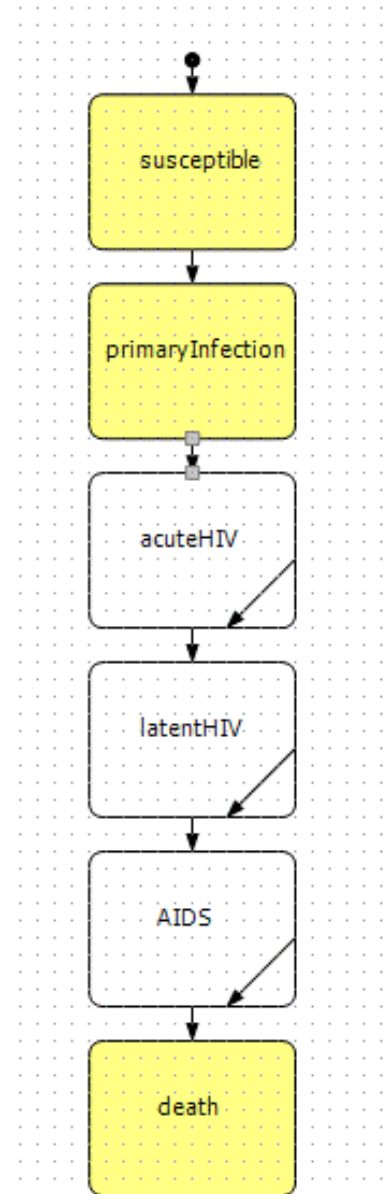
- Example Discrete
 - Ethnicity
 - Gender
 - Categorical infection status
- Continuous
 - Age
 - Elements of body composition
 - Metabolic rate
 - Past exposure to environmental factors
 - Glycemic Level

Example of Continuous Individual State

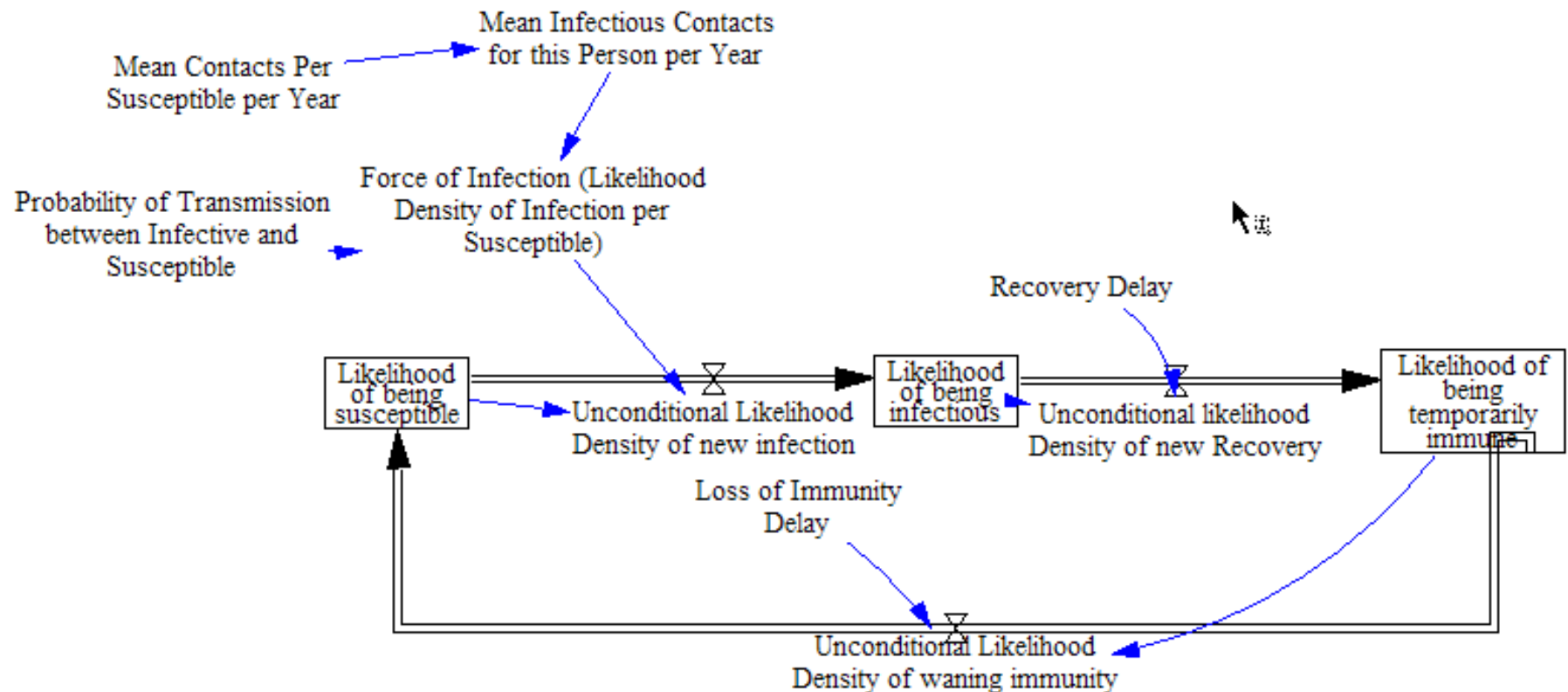


Example of Discrete States

Binary Presence in Discrete State

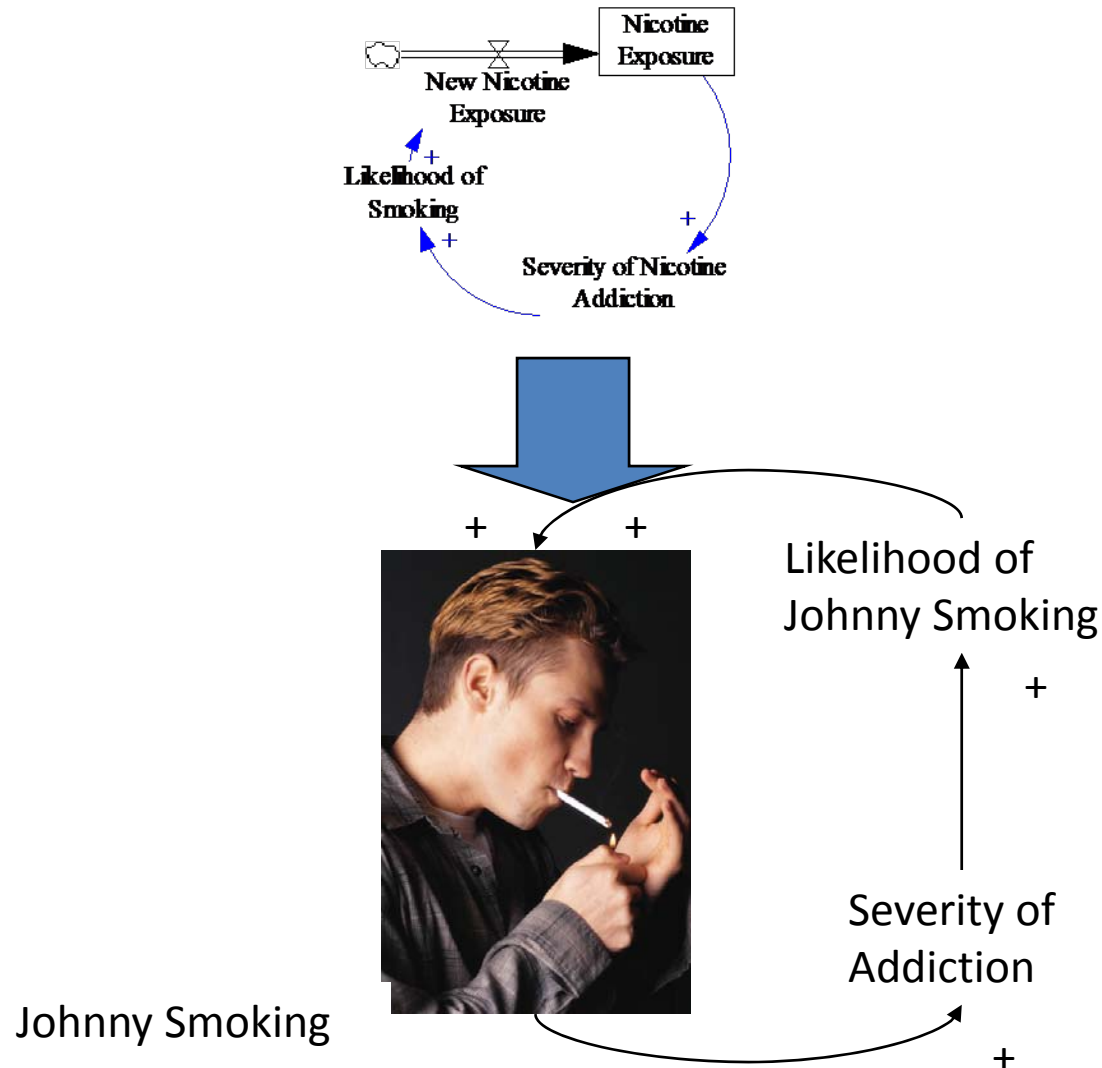


Example of Likelihood of Presence in Discrete State



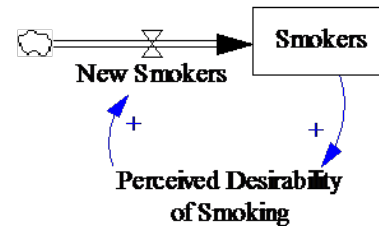
Feedbacks

- Some aggregate feedbacks lie within individual agent

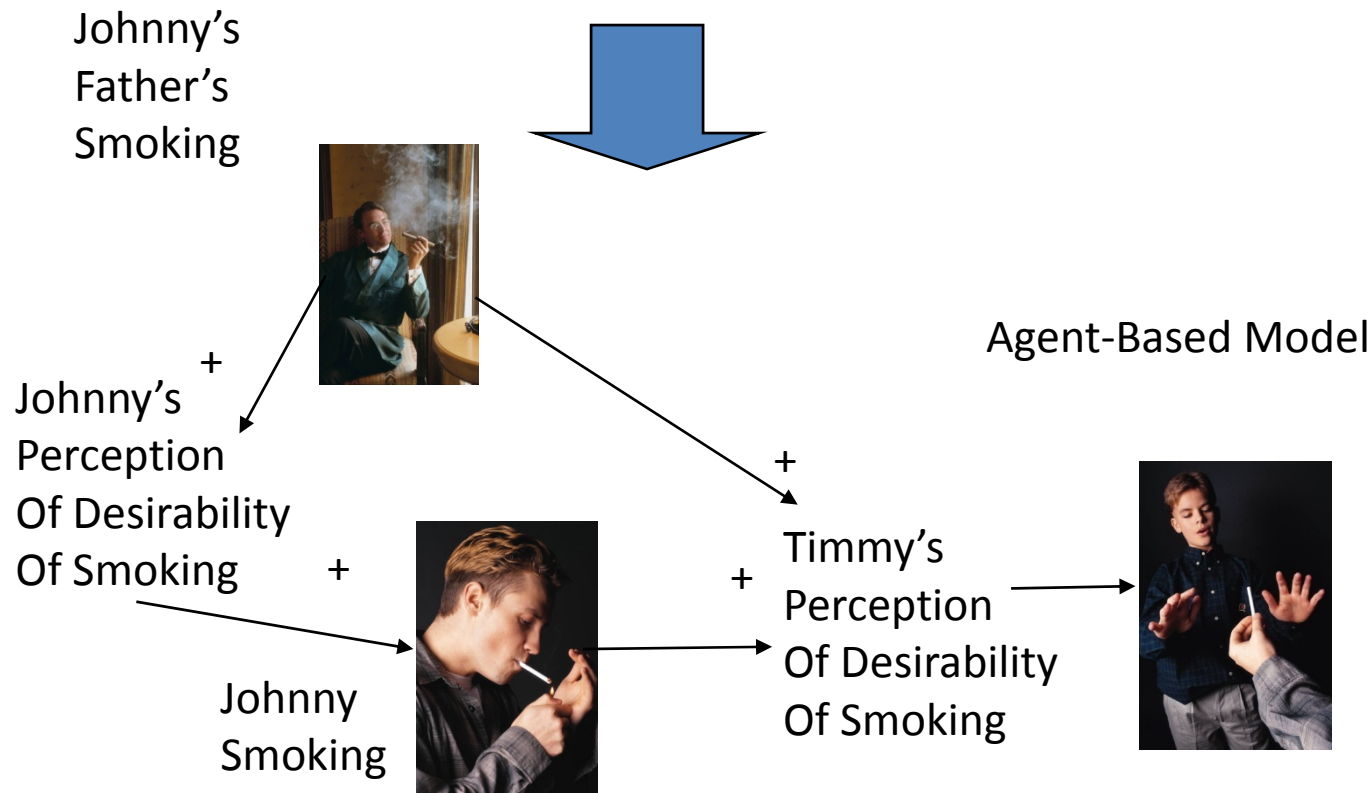


Feedbacks

- Many aggregate feedbacks are *between* agents



Aggregate Model



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
 - Analysis & 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

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

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 $|\text{Population}| * (\text{Behavior of "average" individual})$
- 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)
 - 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 representing 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 factor
 - **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

Editing equation for - Overweight (1/3)

Overweight[Child,InUteroExposureCategory,Sex,Ethnicity] 1 Del

=
INTEG (
-Aging of Overweight[Child,InUteroExposureCategory,Sex,Ethnicity]
-Net Emigration from Overweight[Child,InUteroExposureCategory,Sex,Ethnicity]
+Overweight Babies Born from GDM Pregnancy by Exposure

Initial Value Initial Overweight[Child,InUteroExposureCategory,Sex,Ethnicity]

Type
Level
Normal
☐ Supplementary
Help

Undo {({})} 7 8 9 + 4 5 6 - 1 2 3 * 0 E . / () , ^

Variables Subscripts Functions More
Choose Variable... Inputs
Overweight
Aging of Overweight
Completion of Pregnancy to Overweight State
Developing Overweight
Net Emigration from Overweight
Overweight Babies Born from GDM Pregnancy by Exposure

Units:

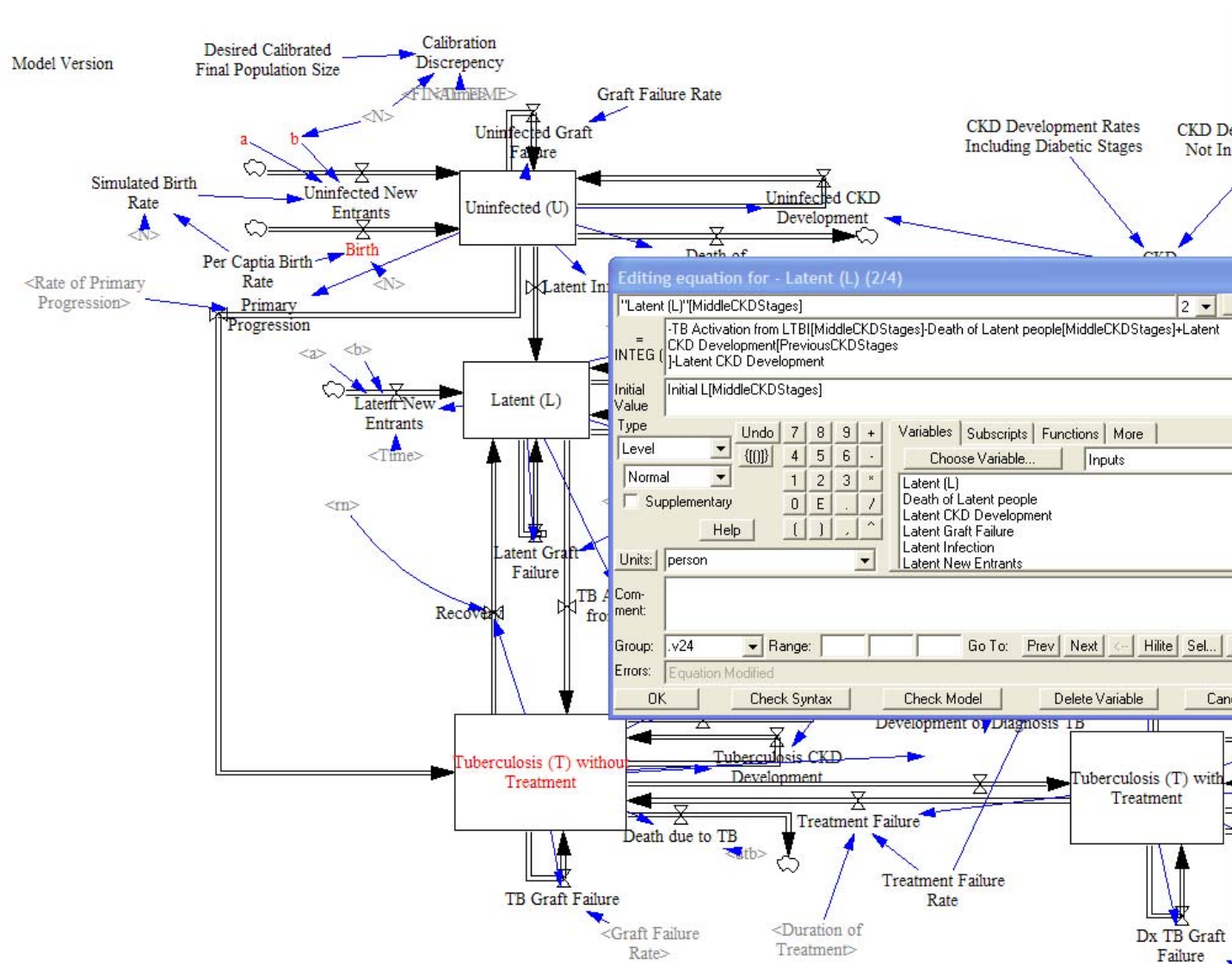
Comment:

Group: .v161 Range: Go To: Prev Next Hilite Sel... New

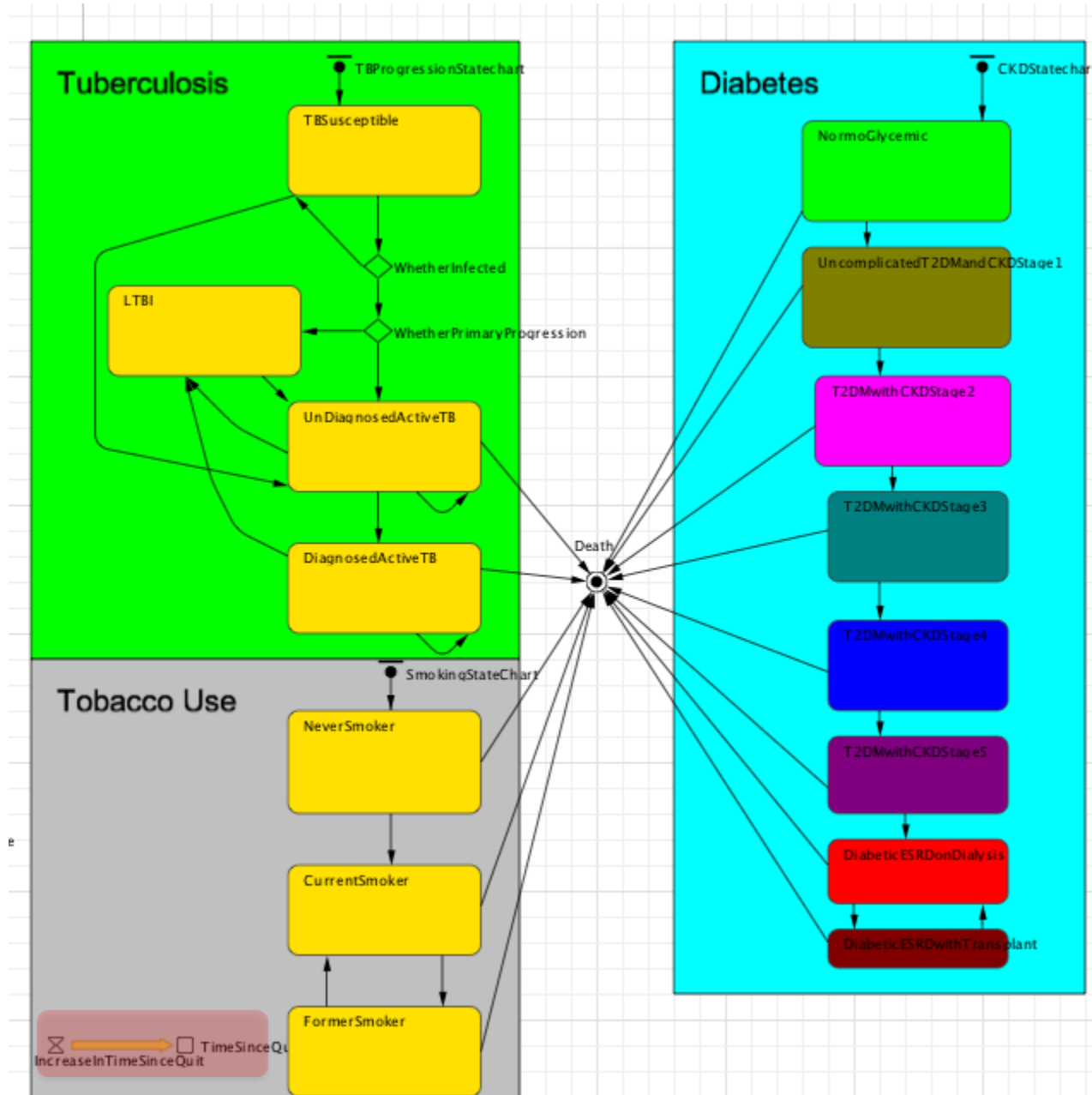
Errors: Equation OK

OK Check Syntax Check Model Delete Variable Cancel

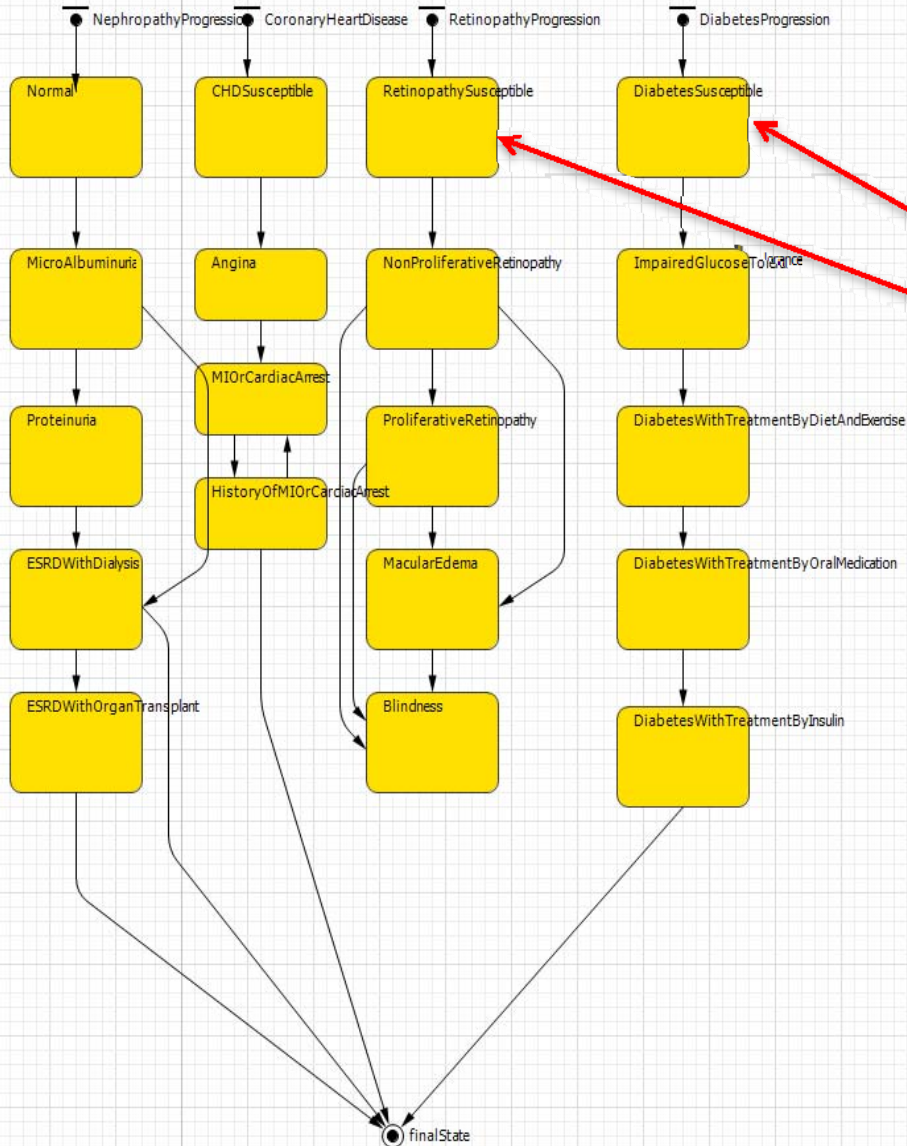
Combinatorial Subscripting: Multi-Dimensional Progression



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

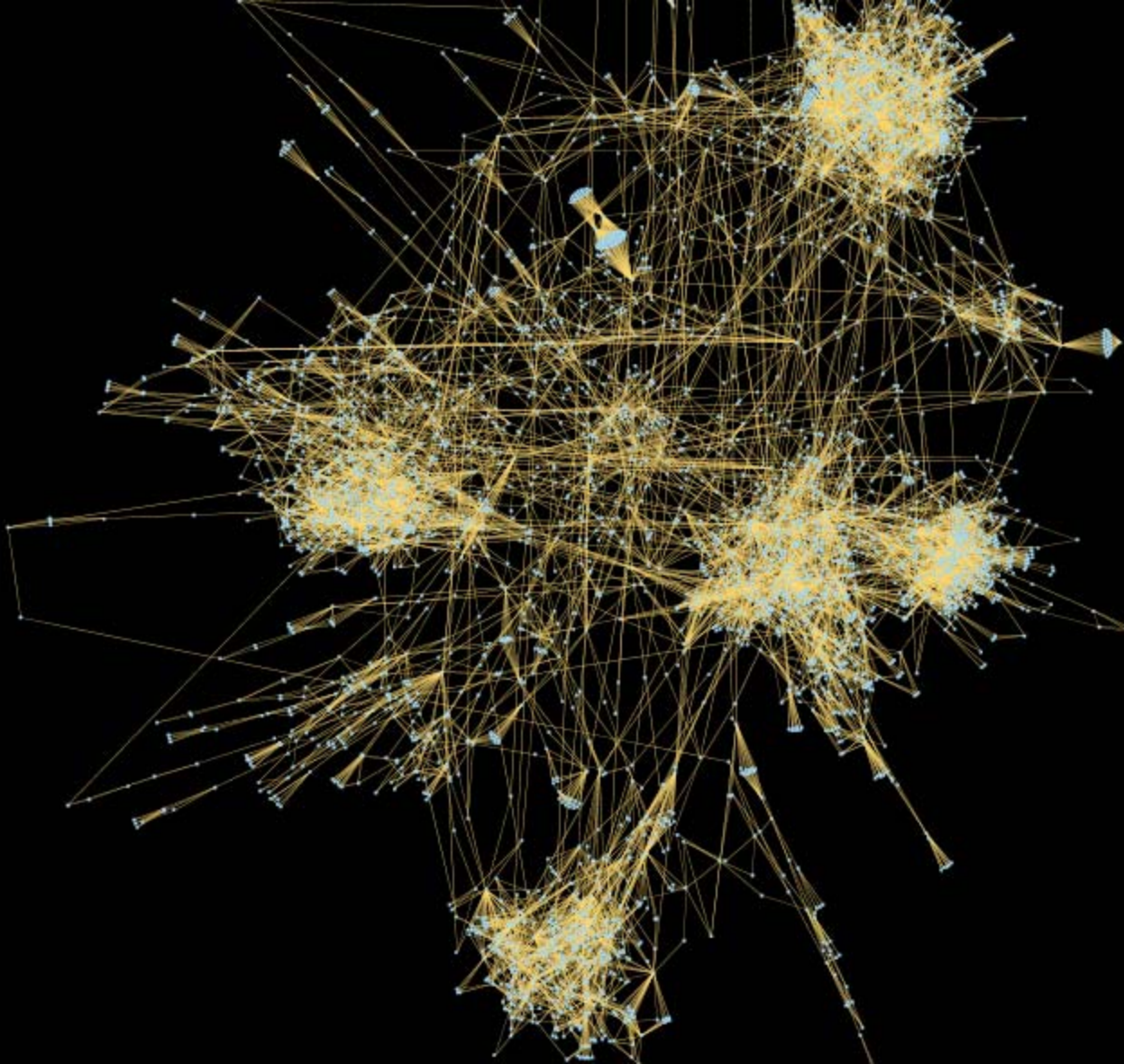
Capturing Heterogeneity in Individual-Based vs. Aggregate Models

- 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 $1/c$.

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

Social Network Analysis: Preliminary Analysis Suggests Pronounced Clustering



Preliminary
case contact
Network

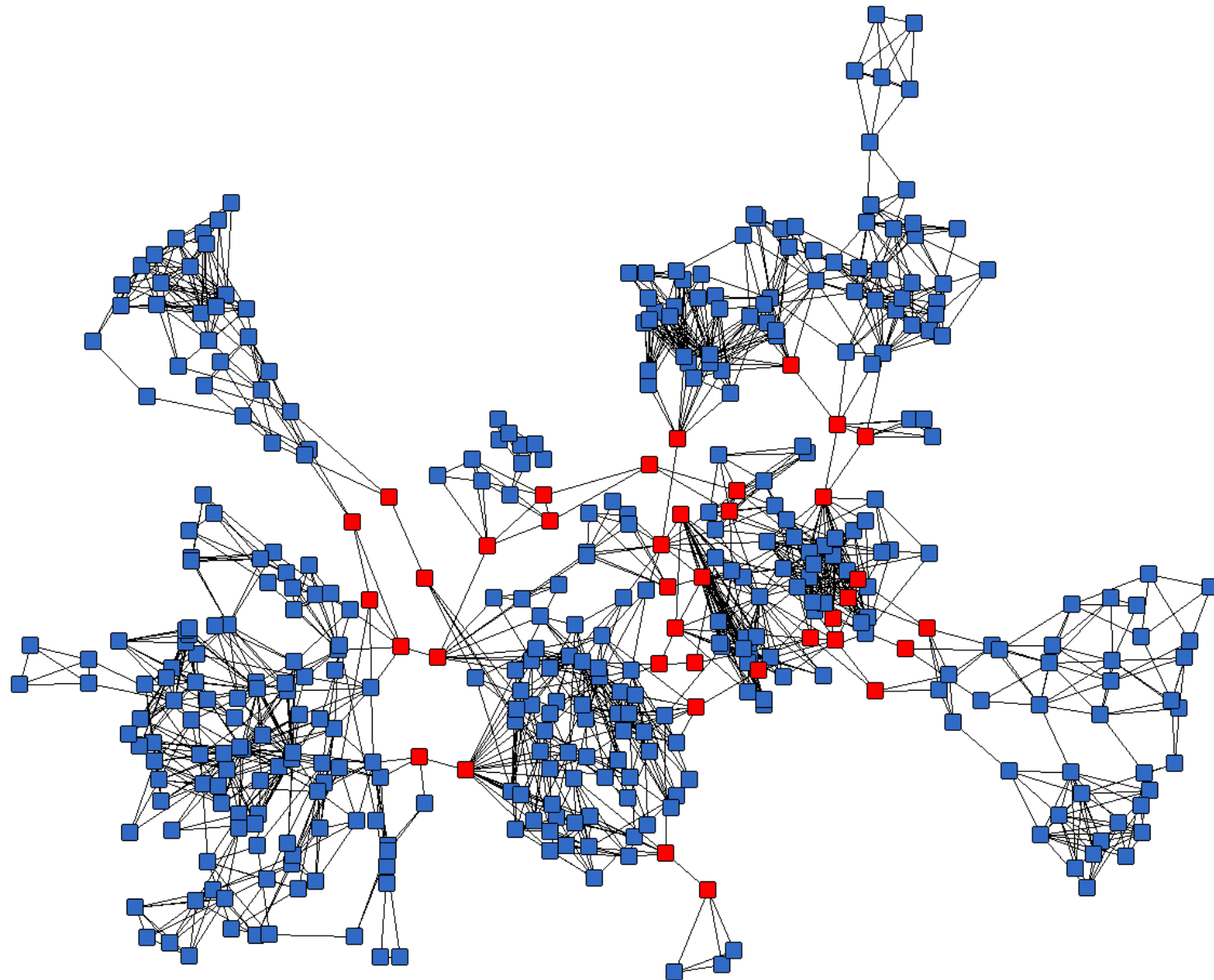
Restricted to
nodes of
degree 2+

Clusters
distinctive by

- Geography
- Ethnicity

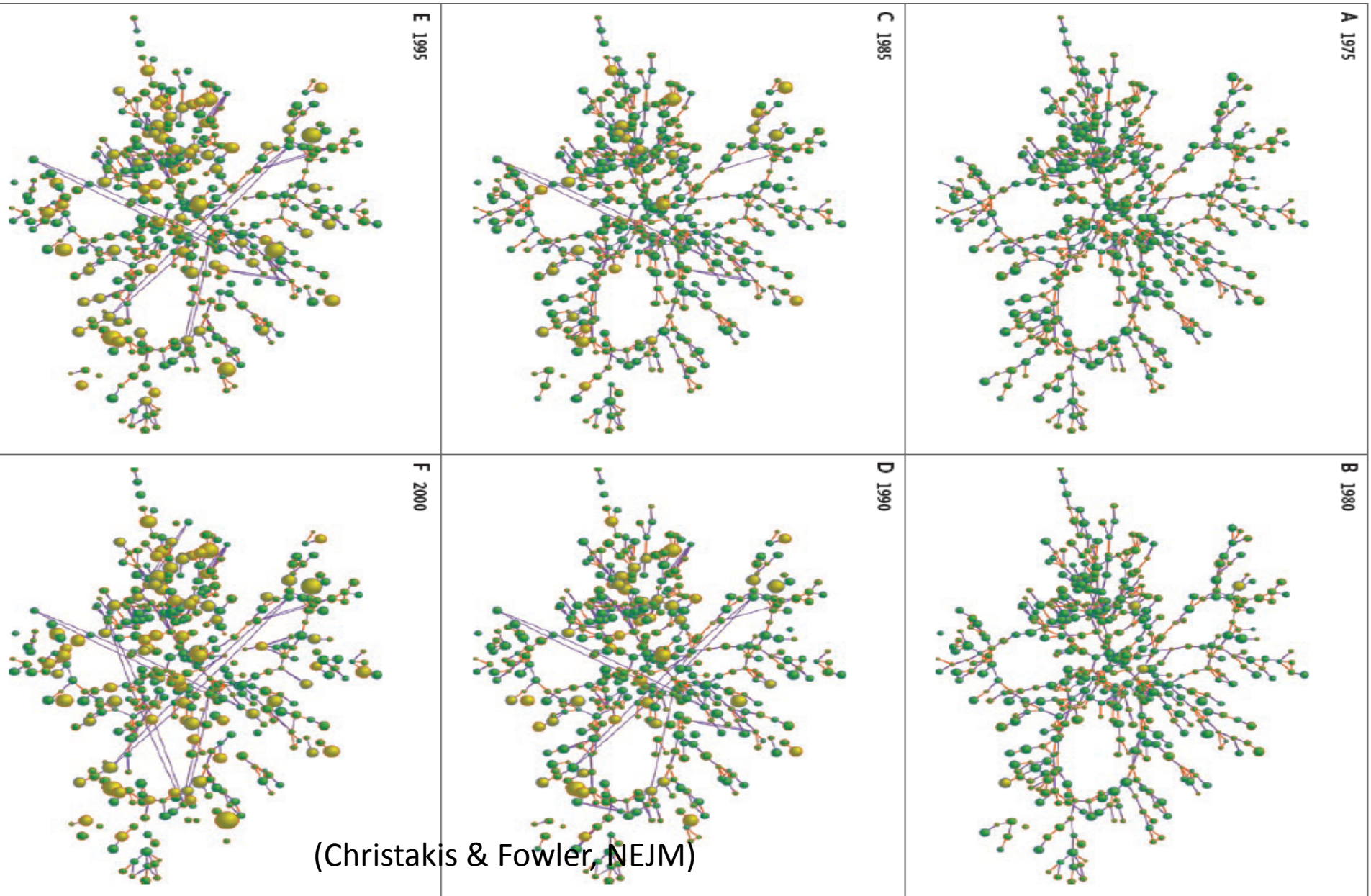
• Data extraction: A.
Al-Azem

Identifying Bridging Individuals



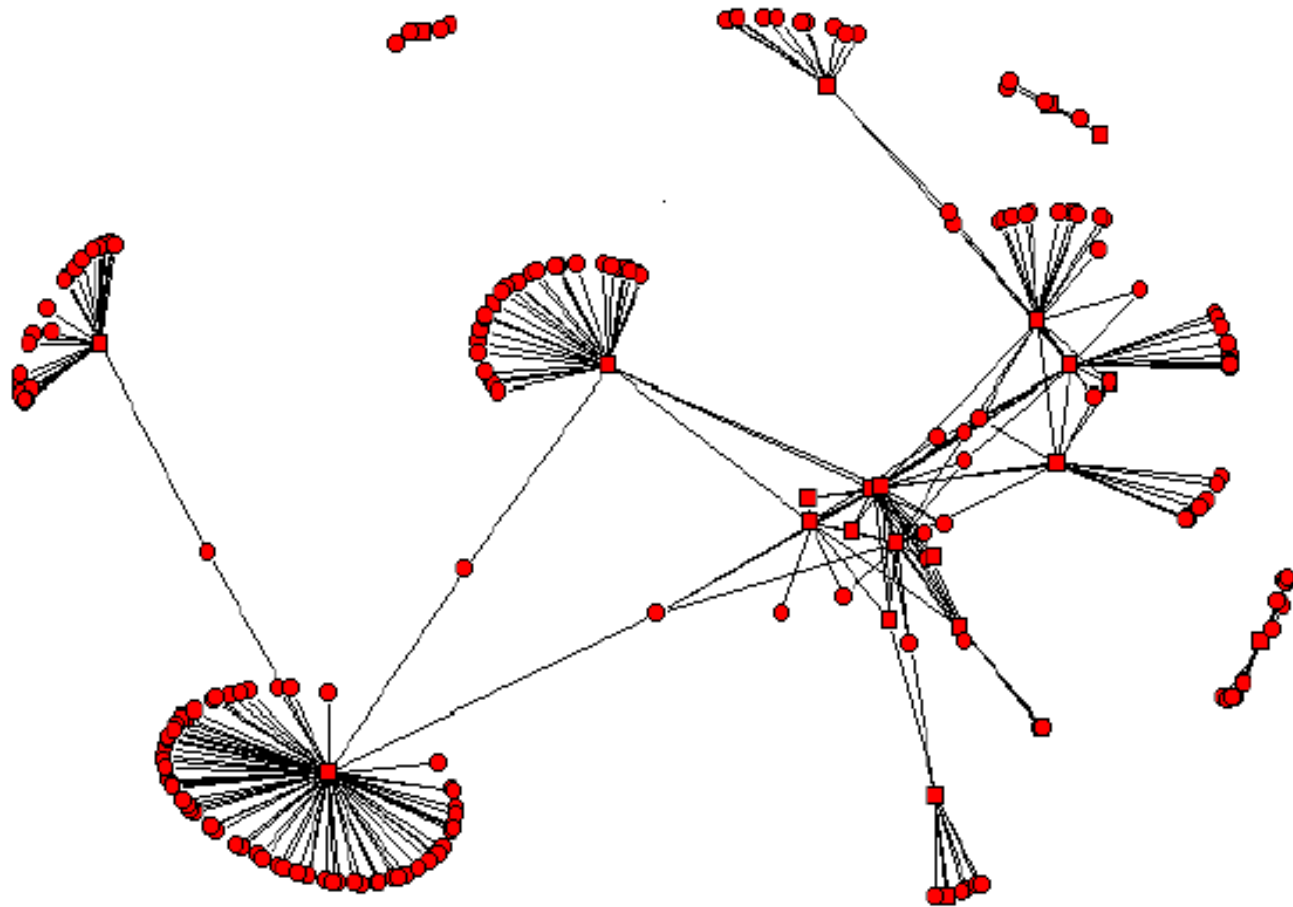
- Preliminary case contact network
- Restricted to nodes of degree 2+
- Data analysis & image: A. Al-Azem

Network Spread of Obesity



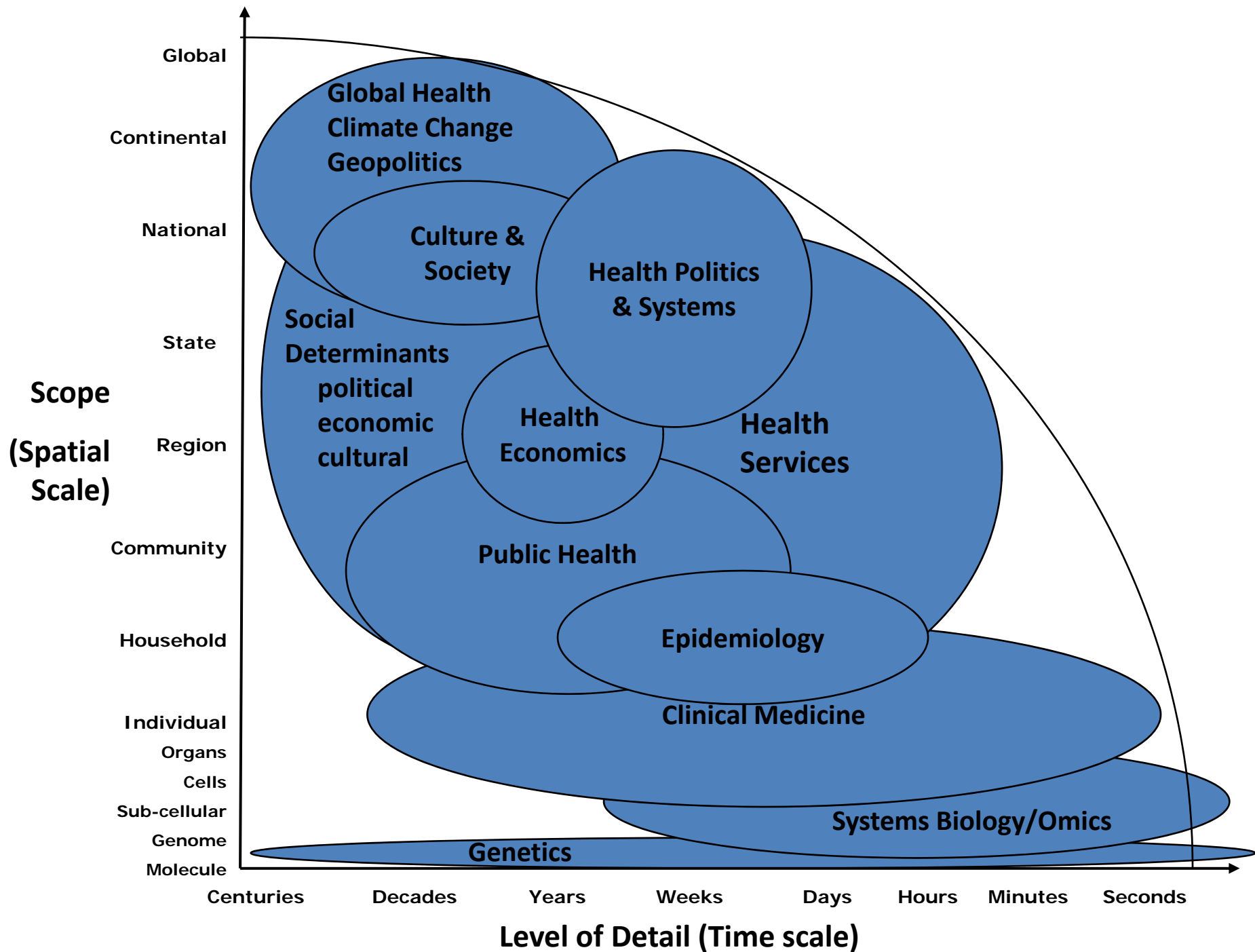
(Christakis & Fowler, NEJM)

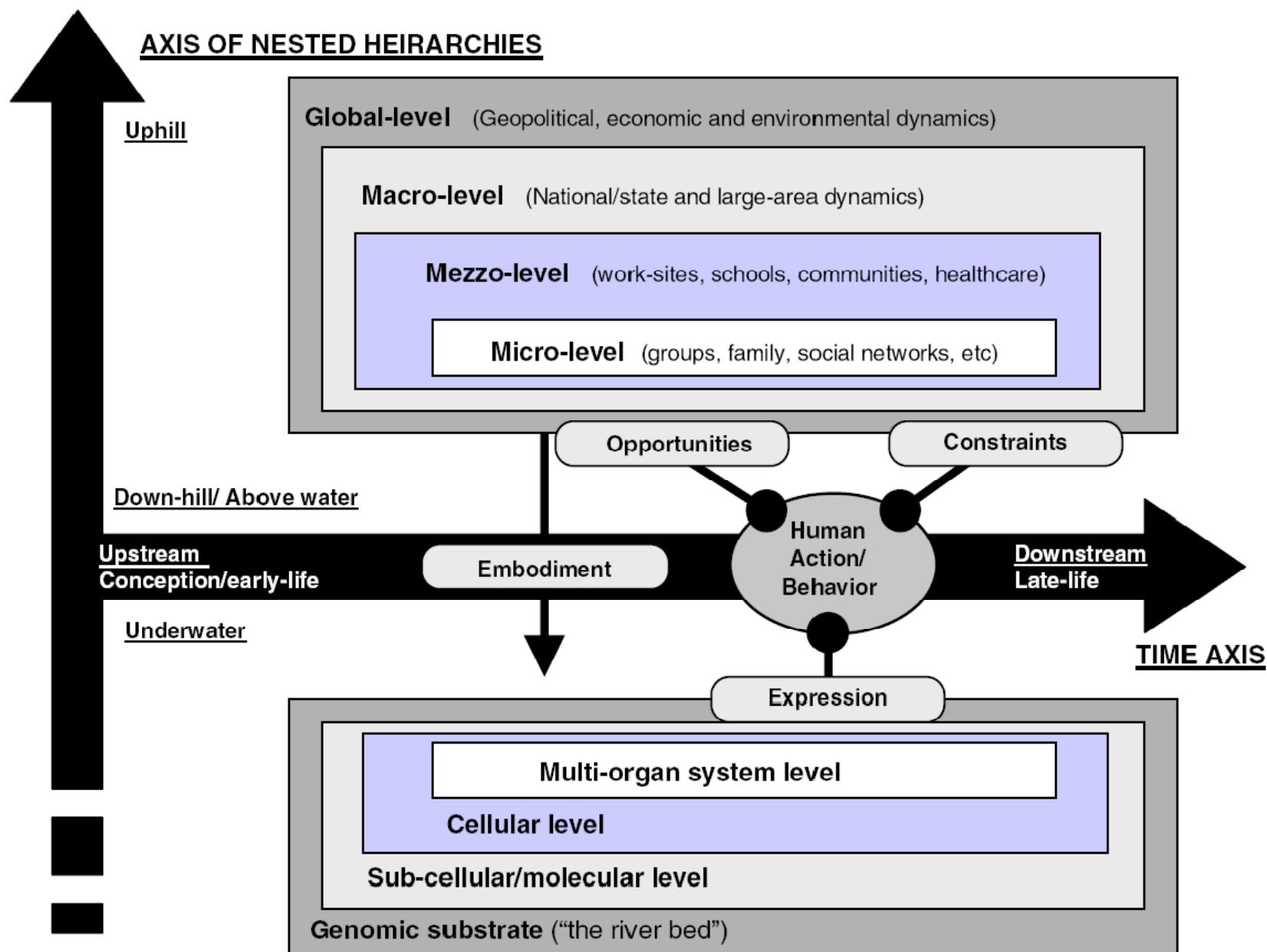
TB Infection and Contact Network



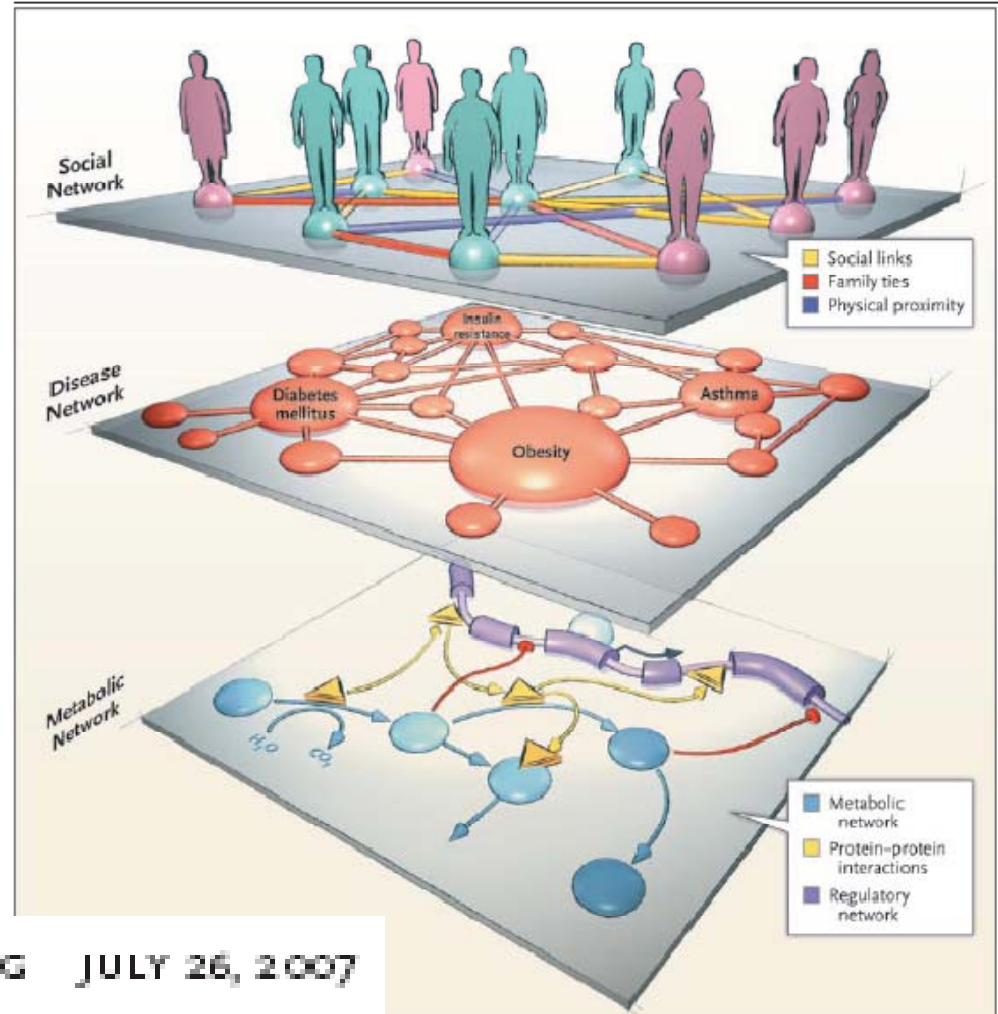
Multi-scale Phenomena

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





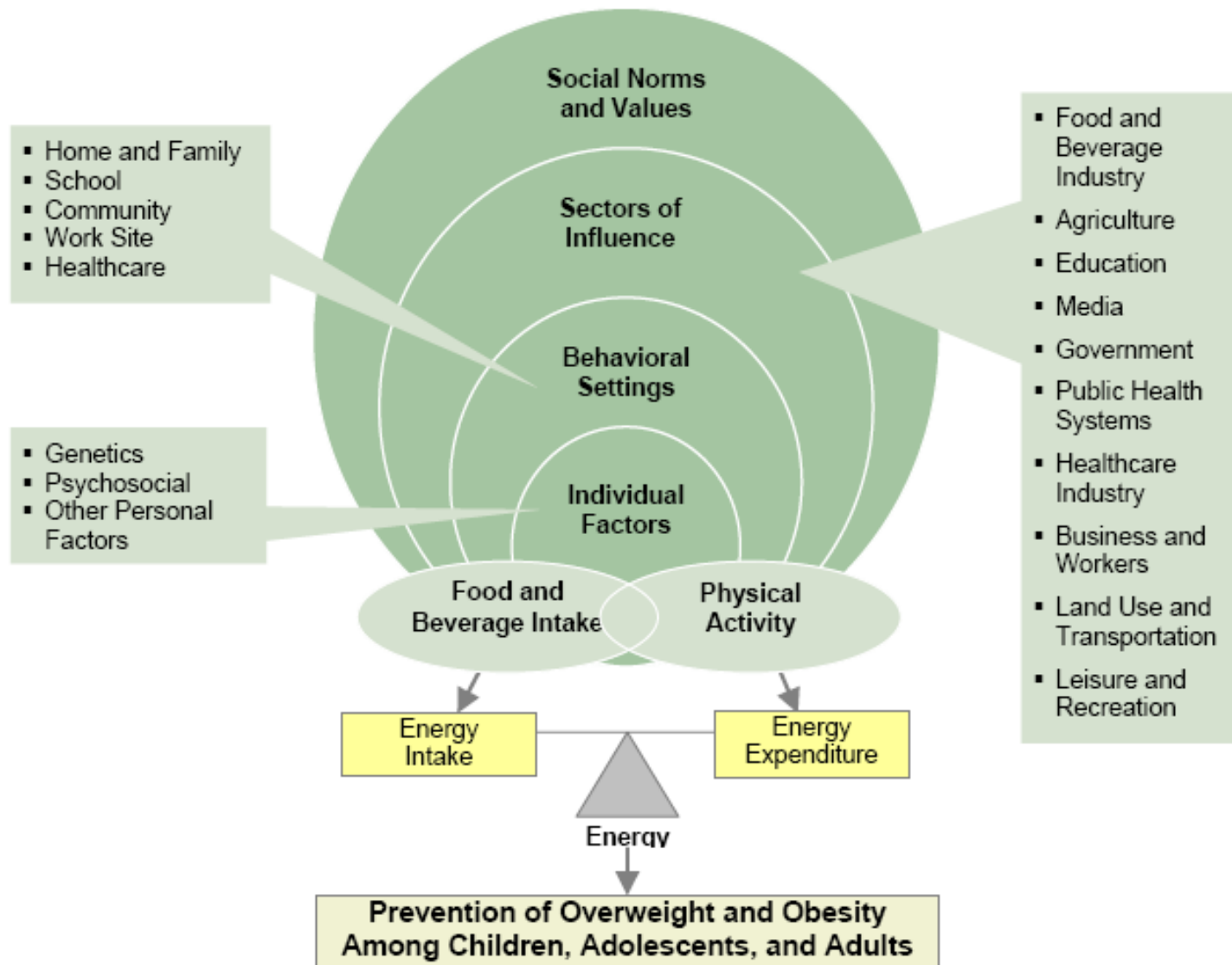
Network Medicine — From Obesity to the “Diseasome” Albert-László Barabási, Ph.D.



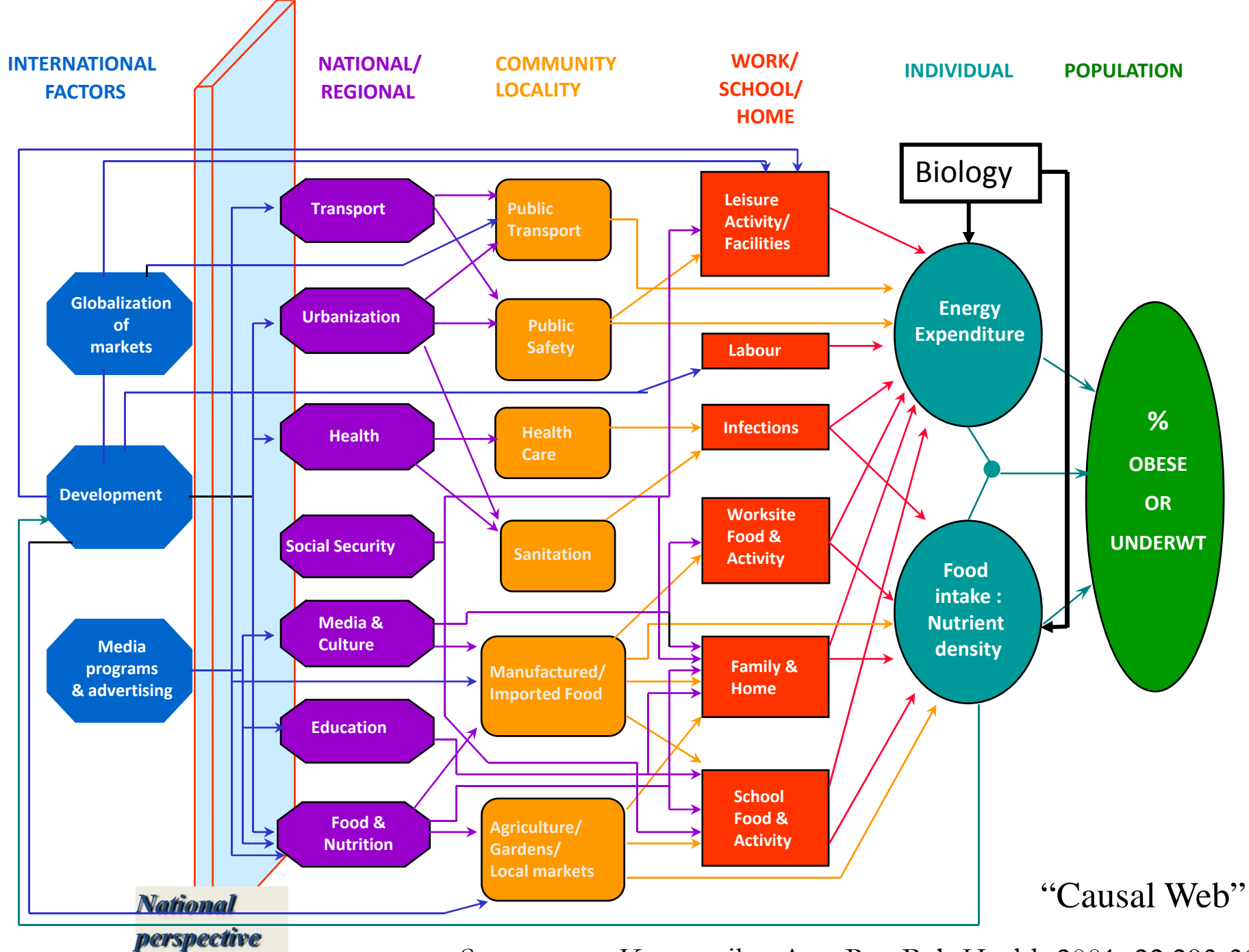
N ENGL J MED 357:4 WWW.NEJM.ORG JULY 26, 2007

Figure 1. Complex Networks of Direct Relevance to Network Medicine.

Although they are often treated separately, most human diseases are not independent of each other. Many diseases are associated with the breakdown of functional modules that are best described as subnetworks of a complex network connecting many cellular components. Therefore, an understanding of the functionally relevant genetic, regulatory, metabolic, and protein–protein interactions in a cellular network will play an important role in understanding the pathophysiology of human diseases (bottom layer). One way to visualize the ensuing potential interrelationships among human diseases is to construct a disease network (middle layer) in which two diseases are connected if they have a common genetic or functional origin. For example, on the basis of our current knowledge of disease genes, obesity is connected to at least seven other diseases such as diabetes, asthma, and insulin resistance, since genes associated with these diseases are known to affect obesity as well. The third network of key importance to human disease is the social network, which encompasses all human-to-human interactions (e.g., familial, friendship, sexual, and proximity-based contacts) that play a role in the spread of pathogens (top layer). These networks also have an important role in the spread of obesity. Efforts to understand the interactions between the cellular, disease, and social networks are part of network medicine, which aims to quantify the complex interlinked factors that may contribute to individual diseases.



Note: Adapted from IOM 2005 (Figure 3-2, p.85)



Source: see Kumanyika Ann Rev Pub Health 2001; 22:293-308

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

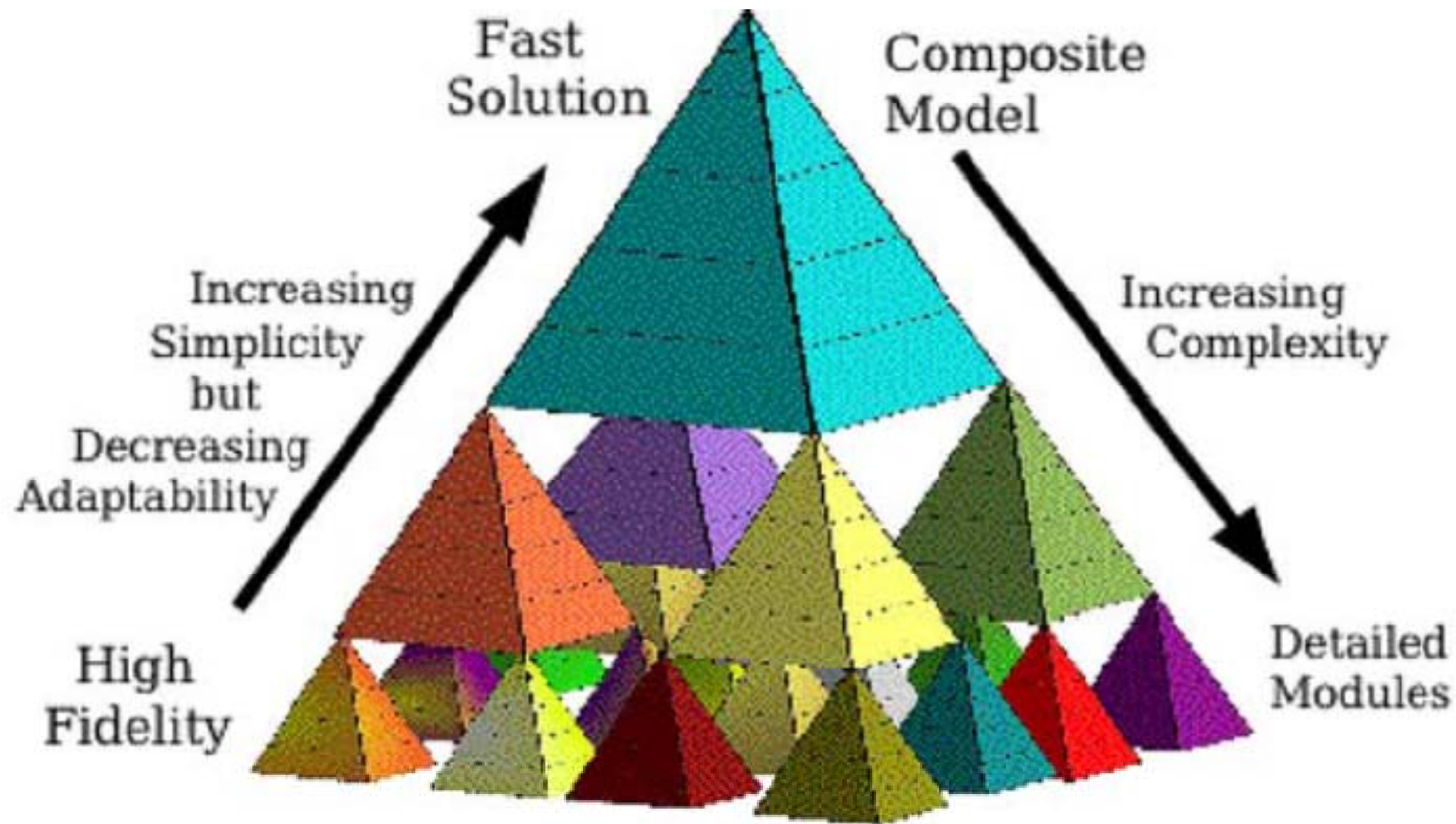
- Individual-based models have many parameters
 - Estimating all of the parameters can require much effort
 - Calibration generally underdetermined (large # of possible sets of parameter values that could calibrate well)
 - May need to make simplifying assumptions
- Pronounced individual-level stochastics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



Individual vs. Aggregate Models: Necessary Tradeoffs

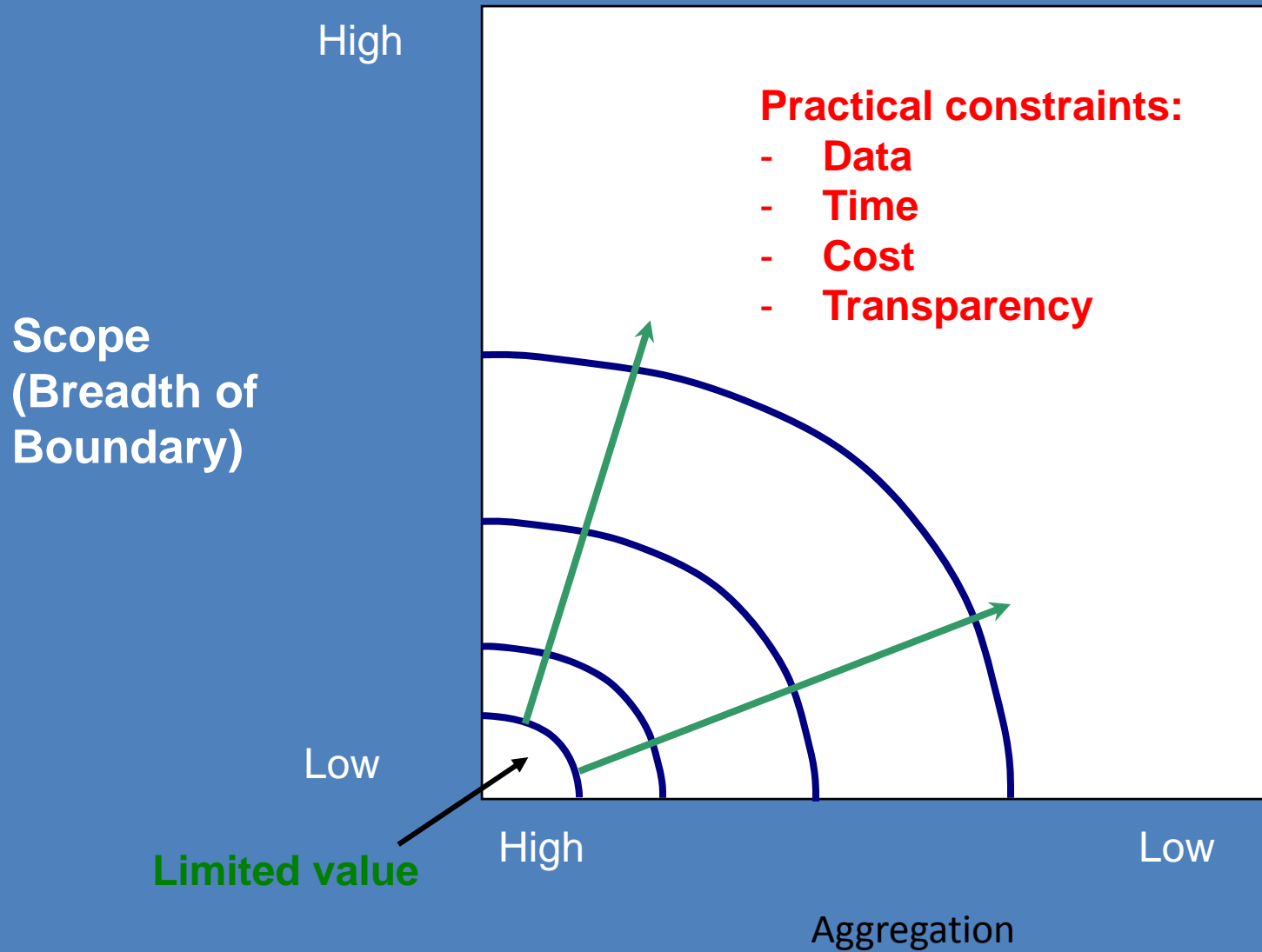
	Transition Generality	Network Represent ation	Calibration	Performance Issues				Capturing Learning/ Adapation
				Basal	Scaling with Population	Scaling with Heterogeneity	Need for Stochastics /Monte Carlo	
Individual Models	++	++				++		++
Aggregate Models		+	++	++	++		+	

- 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

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)

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