Network and Agent-based Model

Questions from Last Class?

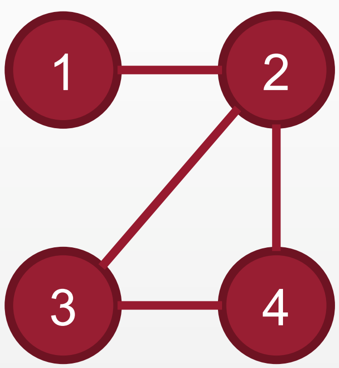
Compartmental Model Recap

- Population is divided up into compartments based on disease state/demographics
- Population mixes randomly
- Beta is a composite of contact and probability of infection given contact, and conventionally non-identifiable

Networks as a Concept

- What is a "network"
  - A conceptual way of representing relationships between things
  - Nodes: *Things*. People, places, etc.
  - Edges: Links between nodes
  - Occasionally these are called graphs, vertexes and arcs
    - Network science co-evolved in several different fields at about the same time
- Networks can be represented in a number of different ways
- A network's structure is sometimes called its "topology"

Representing Networks



Diagram

1	2
2	3
2	4
3	4

Edge List

$$\begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{pmatrix}$$

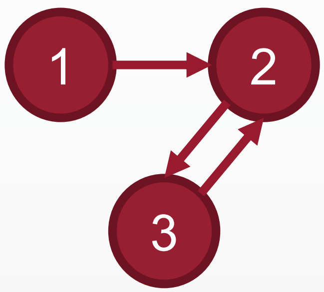
Adjacency Matrix

Representing Networks

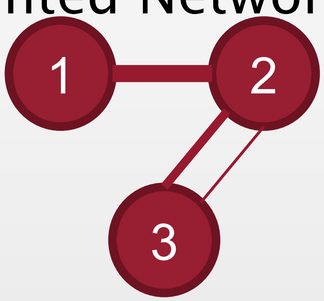
- Diagrams
  - Pros: Easily visualize the network structure, often look really cool
  - Cons: Can get difficult to interact with rapidly, “hairball” networks, not easily machine readable
- Edge Lists
  - Pros: Compact, expressive, easily machine readable
  - Cons: Less “human readable”
- Adjacency Matrix:
  - Pros: Matrix operations unlock all kinds of cool analysis techniques
  - Cons: Also less human reasable, less machine readable than edge lists
- Easy to go back and forth

Complications

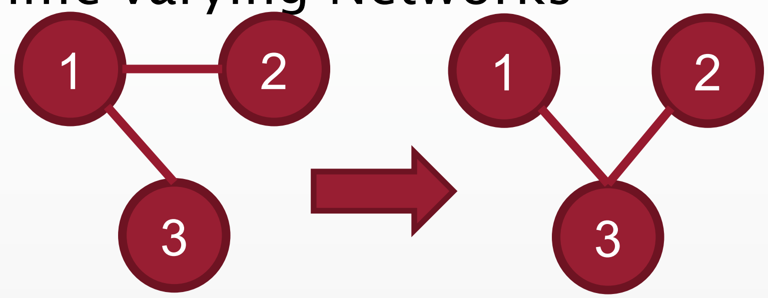
- Directed Networks



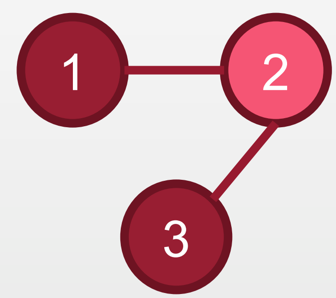
- Weighted Networks



- Time-varying Networks



- Different Types of Nodes/Edges



Very Complex Networks

- Multi-scale Networks
  - Networks of Networks
- A network can simultaneously be directed, weighted, time-varying, etc.

Many Networks, Many Questions

- Network Analysis is interested in the study of the shape and structure of networks
  - How are networks connected, do they form patterns, etc.
  - Simulating a process (like an epidemic) on a network is a small part of that
- Very active research field – sociology, physics, computer science and math all have their own takes

Basic Network Terminology

- Size: The number of nodes
- Density: The ratio of the number of edges in a network compared to the number of *possible* edges
- Path: A set of edges connecting node *I* and node *J*
- Shortest Path: A path with the least number of edges

More Terminology!

- Complete Graph: A network where all nodes are connected
- Connected Component: A portion of a network where there is a path between every pair of nodes
- Clique: Subsets of a network where every node in the clique forms a complete sub-graph
- Clustering Coefficient: The ratio number of neighbors of a node who are connected to each other to to the maximum possible number of such connections

Centrality

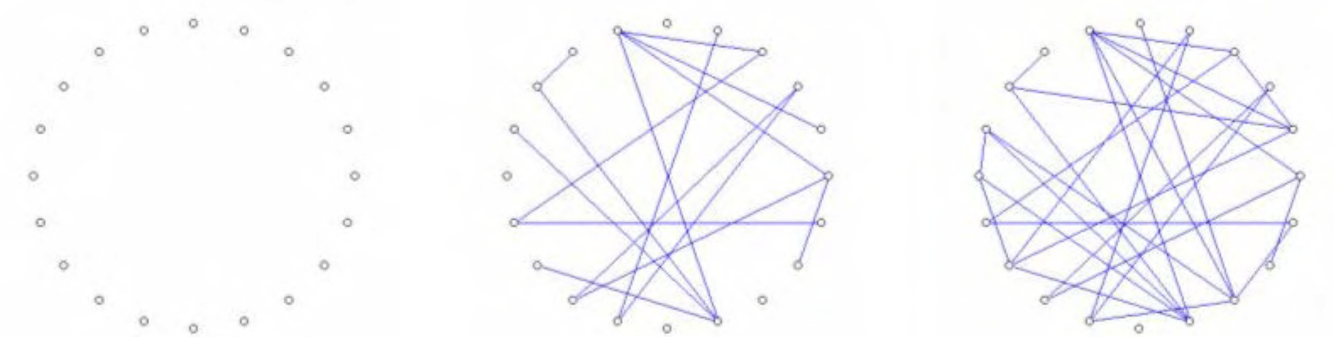
- Ways of describing a node *i*'s position in the network
  - Degree: How many other individuals connect to *i*?
    - In-degree: In a directed network, how many incoming connections
    - Out-degree: In a directed network, how many outgoing connections
  - Betweenness: How many shortest paths between individuals go through *i*?
  - Closeness: How many contacts away is *i* from all other individuals?
  - PageRank: How likely are you to end up at a particular node given you're wandering along paths with the occasional random jump. This is how Google got started.

Types of Networks

- Generated/Theoretical
- Empirical

Erdős–Rényi random graph:

- **Edges form between nodes with probability *p***



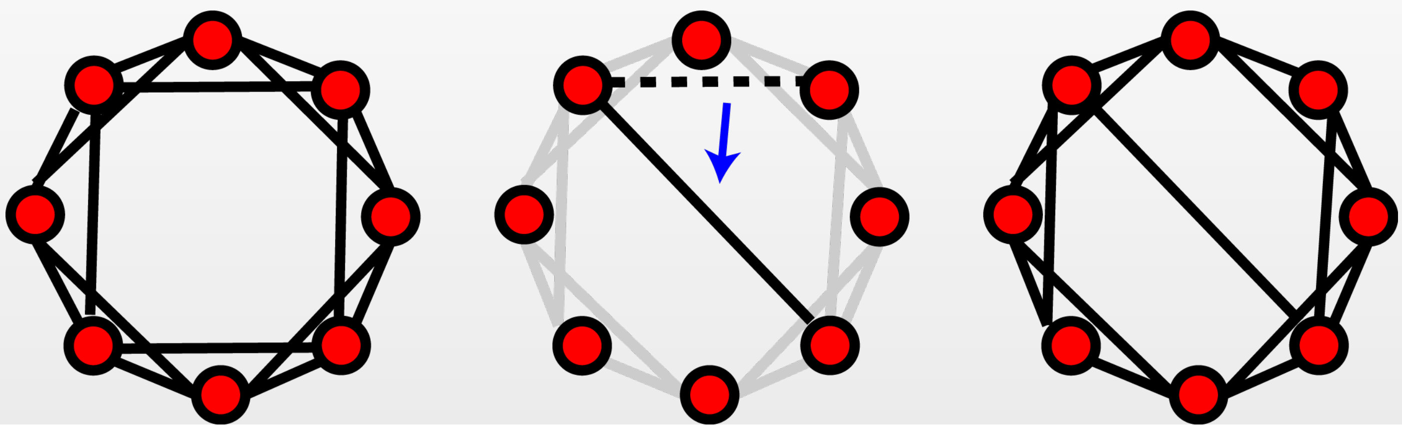
$p = 0$
(a)

$p = 0.1$
(b)

$p = 0.2$
(c)

Watts-Strogatz Small World Network

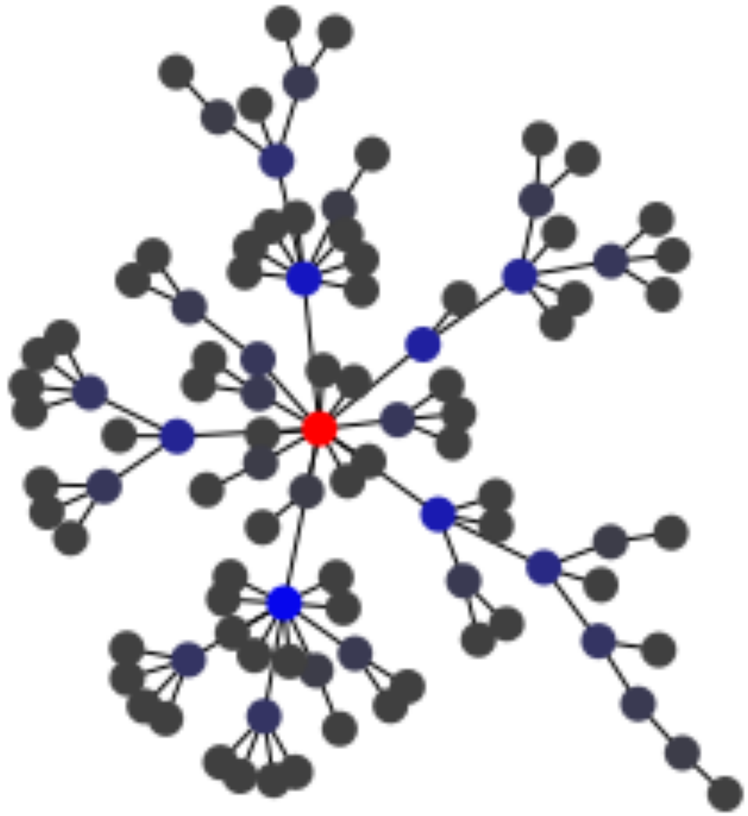
- Each node is connected to its *k* closest neighbors. There is another probability, *p*, that each edge will be "rewired", and instead become a random connection
- This results in a highly clustered network with occasional links between communities
- This is the "Six Degrees of Separation" concept



Barabási–Albert Preferential Attachment

- “Rich get richer”
- Slowly grow the network one node at a time
- Each node *preferentially attaches* to an existing node
- This results in a degree distribution that follows a power law
- Resembles some real systems, how *closely* it represents some real systems remains unclear
- Can result in interesting results – on average, if you pick someone at random, their friends have more friends than they do

$$p_i = \frac{k_i}{\sum_j k_j}$$



Empirical Networks

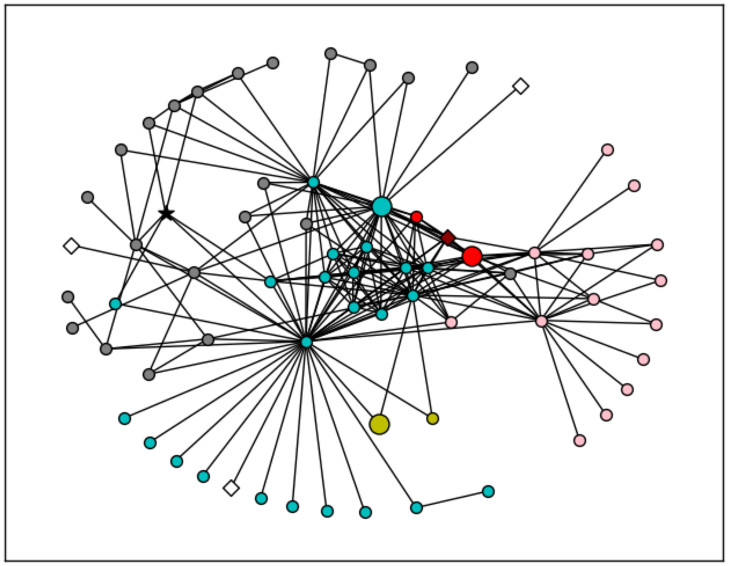
- These are networks that are collected from the real world
- These are often very useful, but are also often *very* expensive in terms of the time, effort and cost to collect them
- Extensive concerns about privacy
- *Network Epidemiology* by Martina Morris discusses the process behind collecting many networks

Surveys

- Ask people who they know
- Often simple and straightforward, but may vary on memory, how people define “friends”, etc.
- FluScale study in China is an example of this
- How do you reach vulnerable populations?
- Can build a network off each persons answers, or try to recruit those answers (”snowball” or respondent driven sampling)

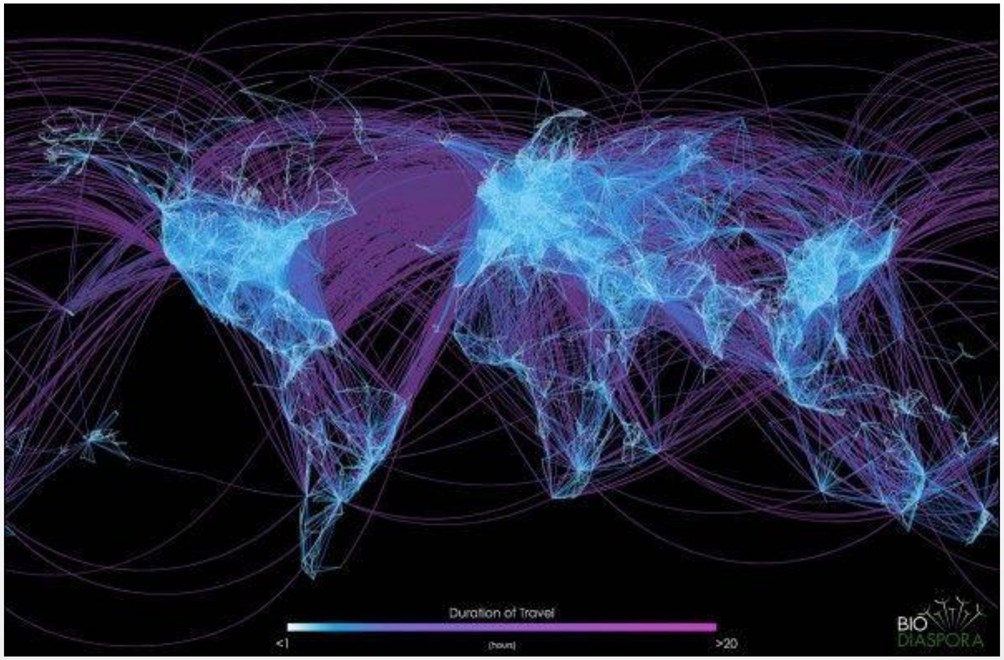
Colocation

- Assume people in the same place at the same time know each other
- Actors in the same movie, socialites at the same parties, people in the same work environment



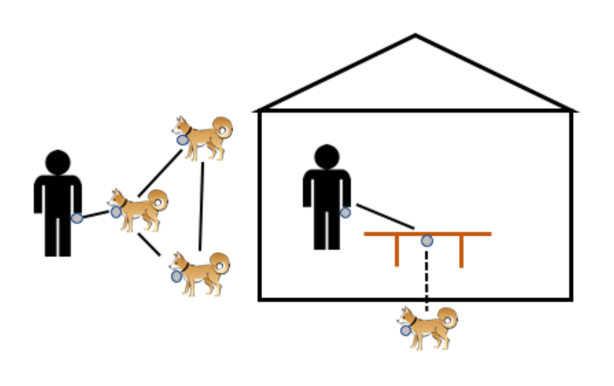
Scrape from Data

- Hospital transfer information
- Airline networks
- Livestock transactions
- Can often be expensive/proprietary



Sensors

- Can automatically collect who is near each other using sensors, mobile phones, etc.
- Assumes nearness = contact
- May be true for some diseases (influenza) and not for others (STIs)

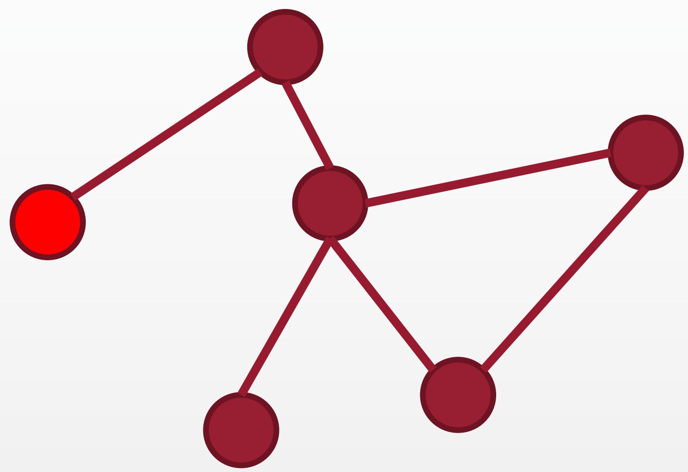


How Are Network Models Implemented?

- Proving things about networks is hard, and beyond the scope of this class
- Often done by simulation

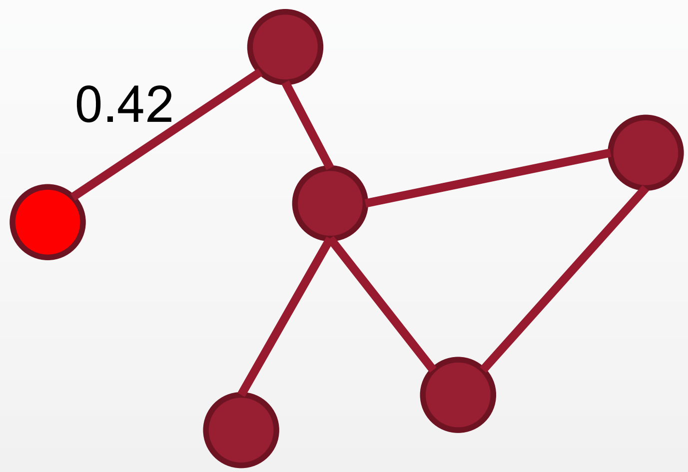
A Simple Network Model

- Start with one infected individual



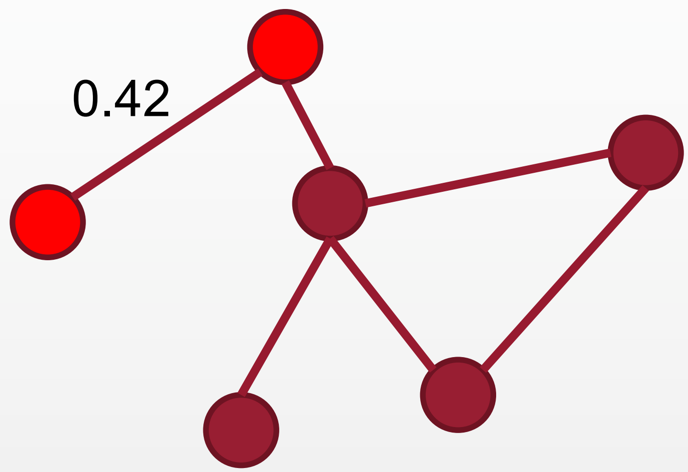
A Simple Network Model

- Spreads along edges to other nodes at probability p, where $p = 0.65$



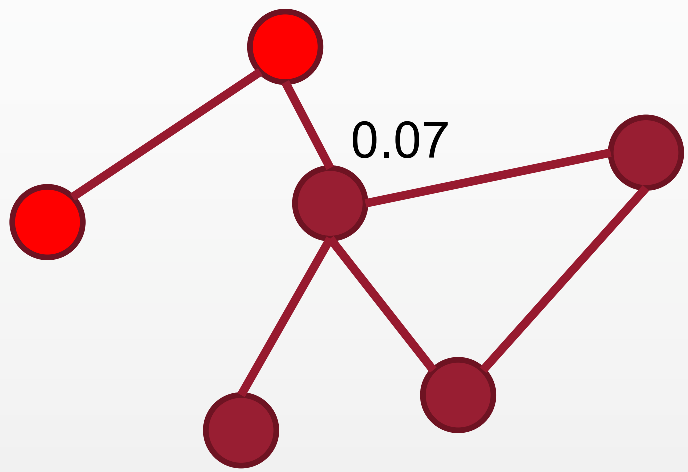
A Simple Network Model

- Spreads along edges to other nodes at probability p, where $p = 0.65$



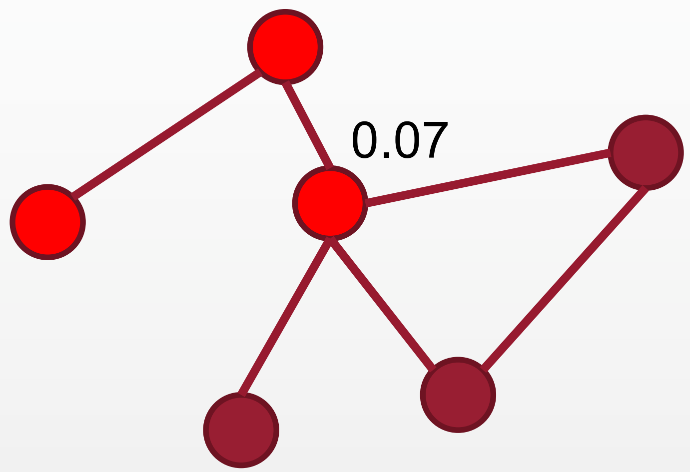
A Simple Network Model

- Spreads along edges to other nodes at probability p, where $p = 0.65$



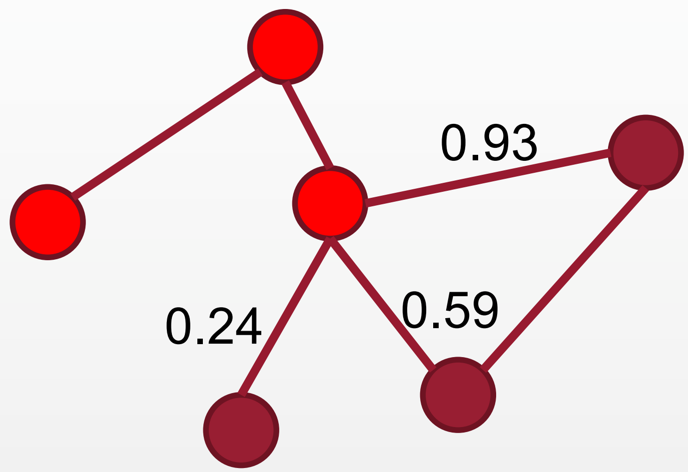
A Simple Network Model

- Spreads along edges to other nodes at probability p, where $p = 0.65$



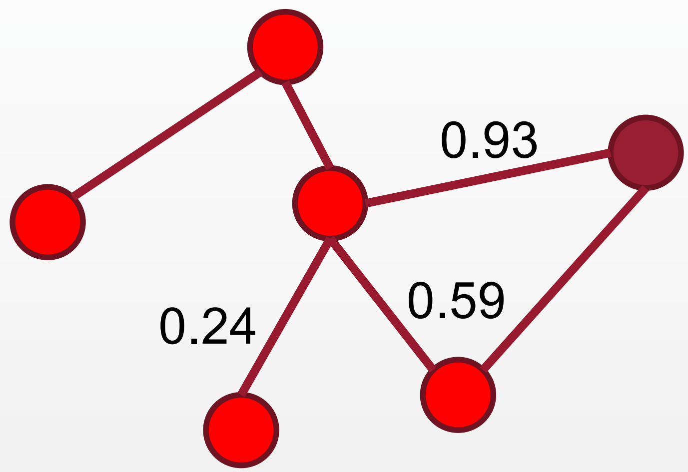
A Simple Network Model

- Spreads along edges to other nodes at probability p, where $p = 0.65$



A Simple Network Model

- Spreads along edges to other nodes at probability p, where $p = 0.65$



Differences Between Compartmental and Network Models

- Results make not agree
- This can be the result of:
  - Random chance (network models are by-and-large stochastic)
  - Network structure
- Contact pattern – note that each node has fewer interactions than they would in a compartmental model

Network Strengths

- More direct modeling of contact
- Can decouple contact and probability of infection transmission given contact
  - These two are nonidentifiable in a compartmental model
- They look really cool
- Built-in ability to have heterogeneity, stochasticity, etc.

Network Drawbacks

- Where does your network come from?
  - Some networks can be assumed
  - Likely better than mass action, but still possibly not right
- Collecting empirical networks is hard
- Much higher computational requirements
- Much harder to prove things
- Largely reproducible given the network, but many people don't make their networks public

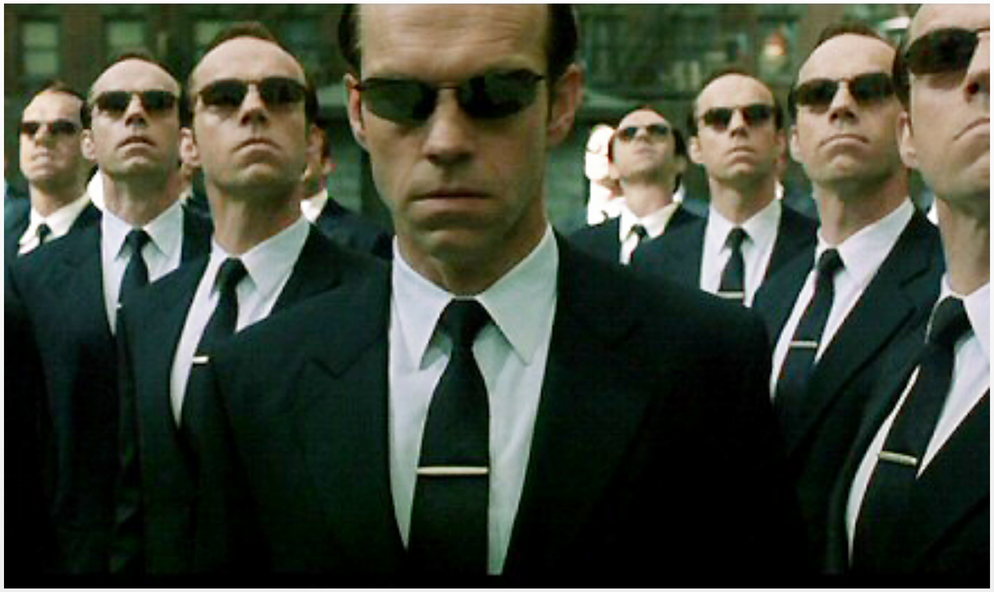
Agent Based Models

Recap

- Compartmental Models
  - Fast and “easy”
  - Random mixing
  - Representing heterogeneity is complicated
- Network Models
  - Harder to implement
  - Contact patterns are specified and fixed
  - Some heterogeneity can exist in the properties of nodes and edges

Enter the Agent-based Model

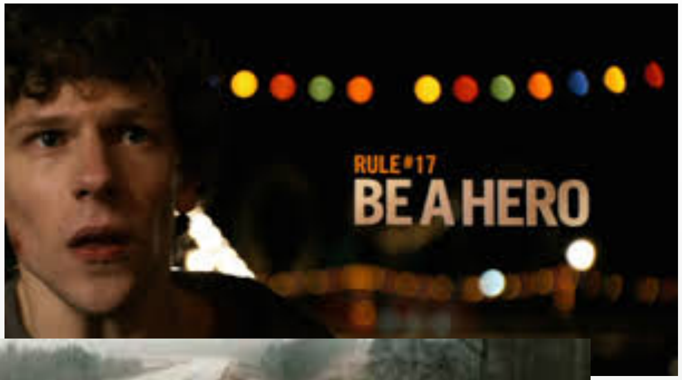
- Use a computer simulation to model lots of individuals in the same environment
- Population is modeled as a set of autonomous "agents" with relatively simple rules
- Contact is then an emergent property of behavior
  - How we interact instead of who we interact with




RULE #1:
CARDIO


RULE #31
CHECK THE
BACK SEAT


RULE #32
ENJOY THE
LITTLE
THINGS


RULE #17
BE A HERO

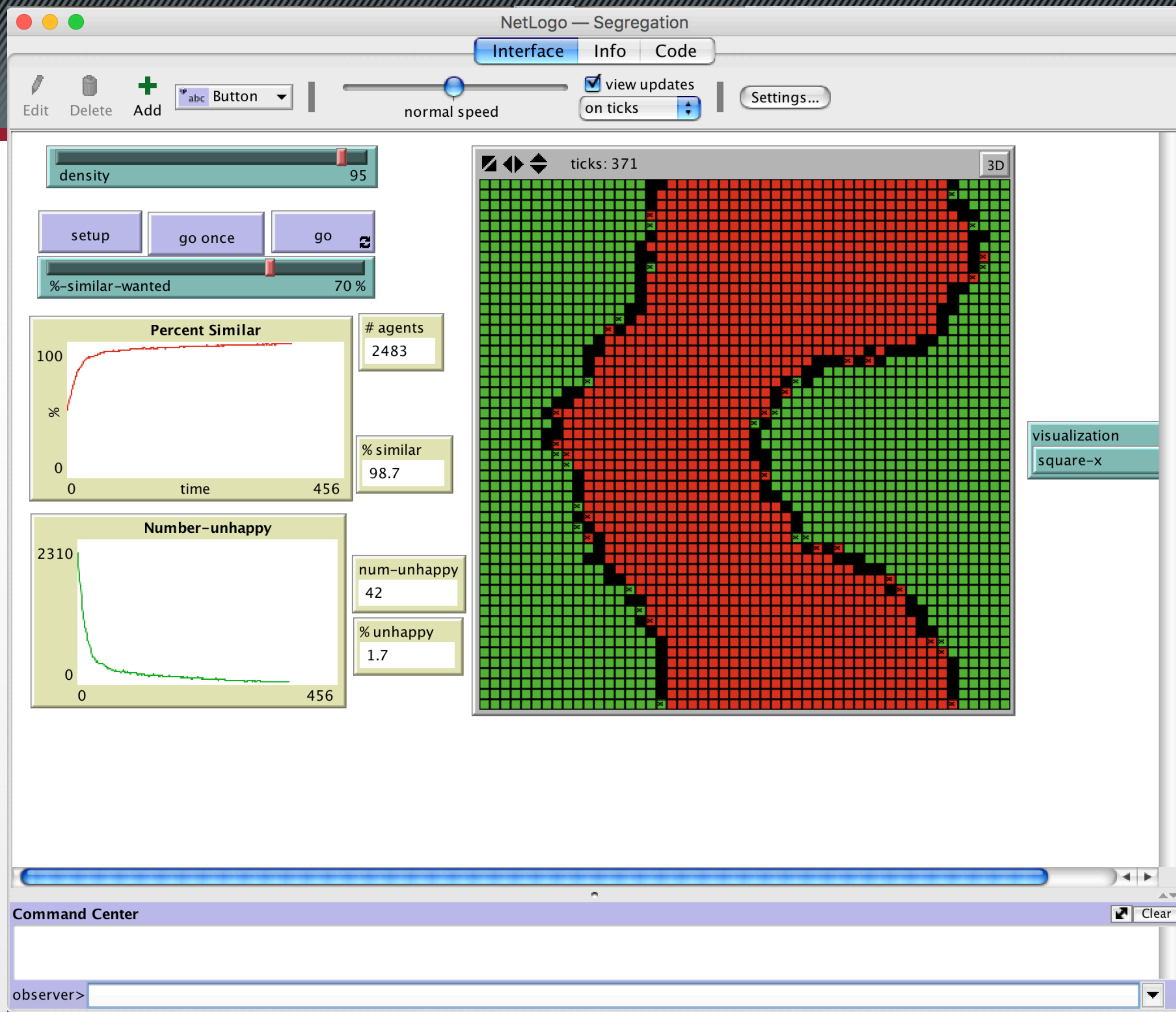

RULE #18
LIMBER
UP

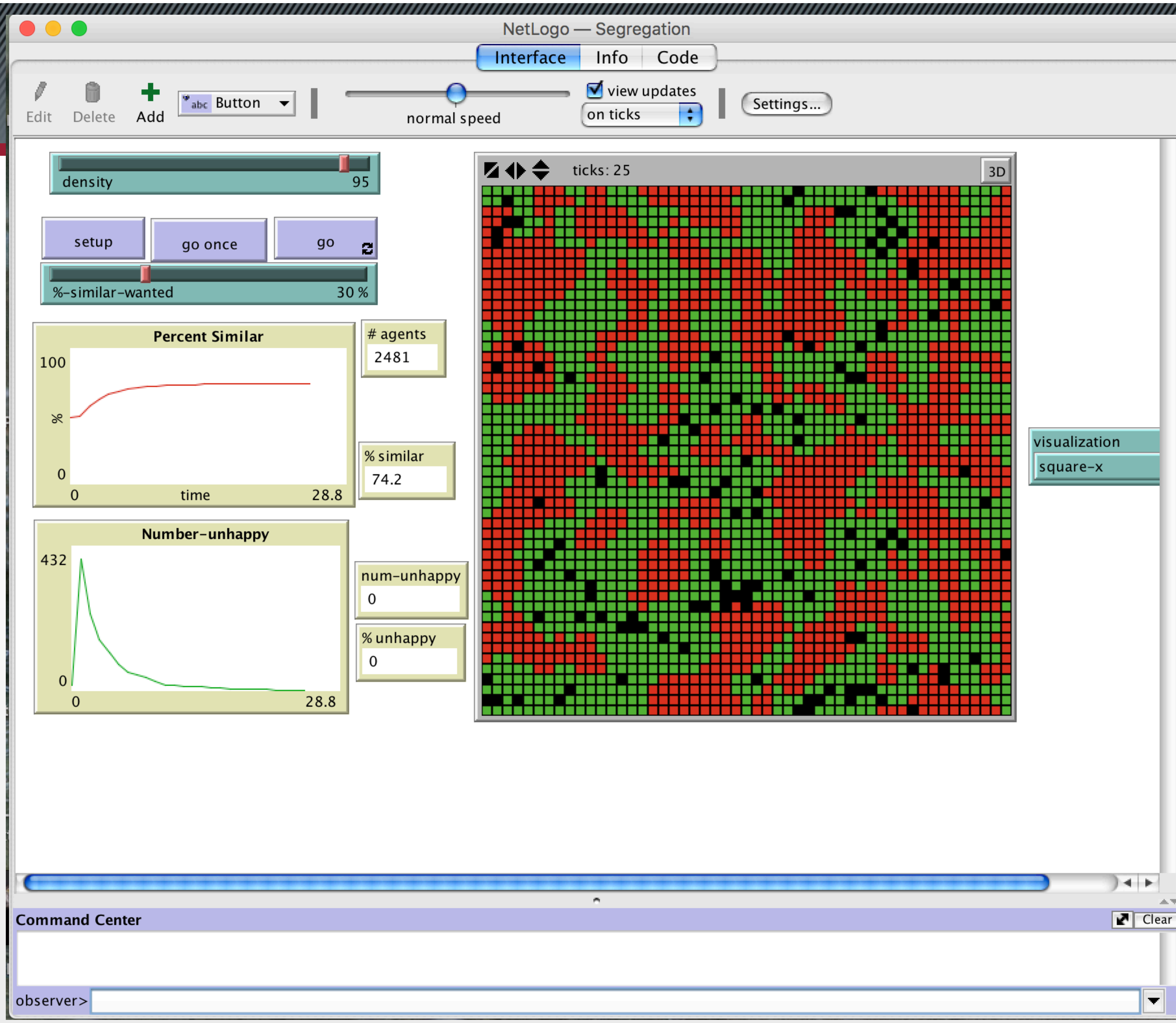
Why Is This Interesting

- This is a very flexible approach
- New kinds of randomness can be represented
  - Behaviors can be drawn individually from a distribution
- Complex results can arise from simple, low level interactions
- Can help discover patterns other models later describe

Schelling Segregation Model

- Remarkably simple model of preference for "like living with like"
- Two types of agent
  - Satisfied if at least *p* percent of neighbors are the same as them
  - Dissatisfied if lower, move to a new vacant location


NetLogo — Segregation
Interface
Info
Code
Edit
Delete
Add
Button
normal speed
view updates
on ticks
Settings...
density
95
setup
go once
go
%-similar-wanted
70 %
Percent Similar
100
%
0
0
time
456
agents
2483
% similar
98.7
Number-unhappy
2310
0
0
456
num-unhappy
42
% unhappy
1.7
ticks: 371
3D
visualization
square-x
Command Center
Clear
observer>

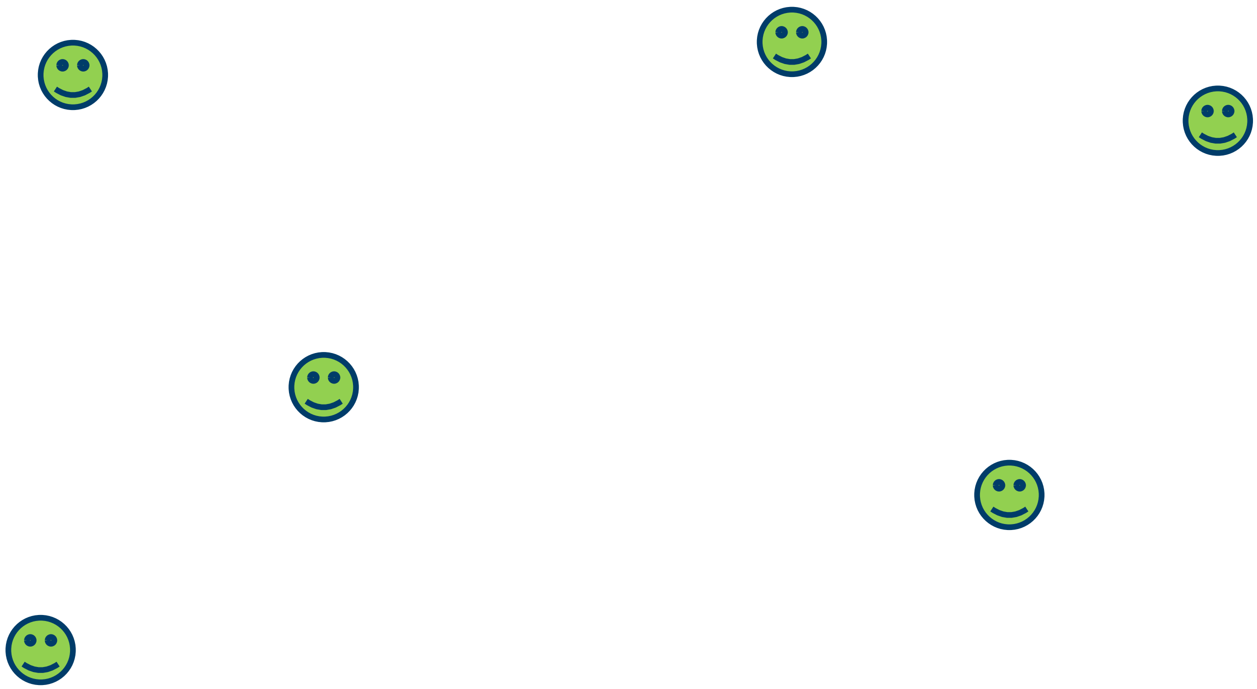

NetLogo — Segregation
Interface
Info
Code
Edit
Delete
Add
Button
normal speed
view updates
on ticks
Settings...
density
95
setup
go once
go
%-similar-wanted
30 %
Percent Similar
100
%
0
0
time
28.8
agents
2481
% similar
74.2
Number-unhappy
432
0
0
28.8
num-unhappy
0
% unhappy
0
ticks: 25
3D
visualization
square-x
Command Center
Clear
observer>

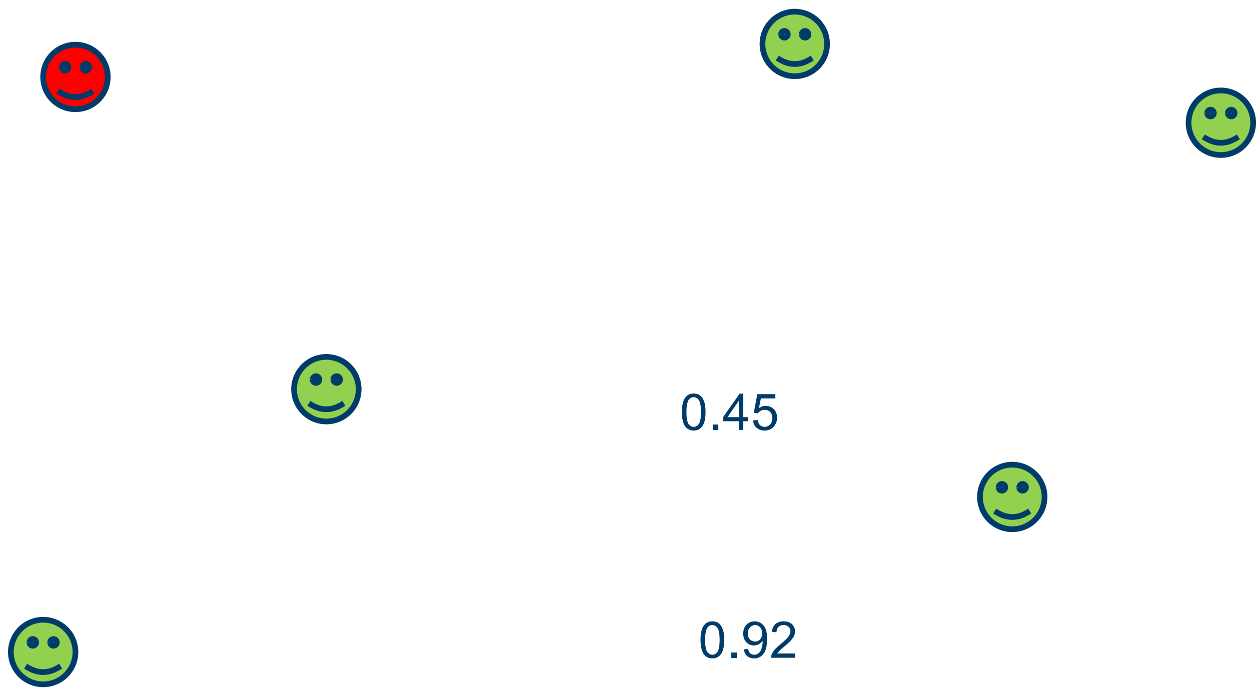
What Do We Get From This?

- Mild amounts of assortative preference can result in a highly segregated population
- "I'm willing to live in a neighborhood where 70% of my neighbors look not like me" seems like it would result in lots of diversity

The Fuzzy Grey Area

- Compartmental, Network and Agent-based models are often considered to be discrete entities.
- They really aren't
  - What if the behavior of an agent is "mix randomly"?
  - What if we make a compartment for every person?
  - What if nodes in a network add and remove links to one another based on rules?
  - What if we use an agent-based model to estimate the formation of a network?
  - What if we use a network model to represent the movement of agents?



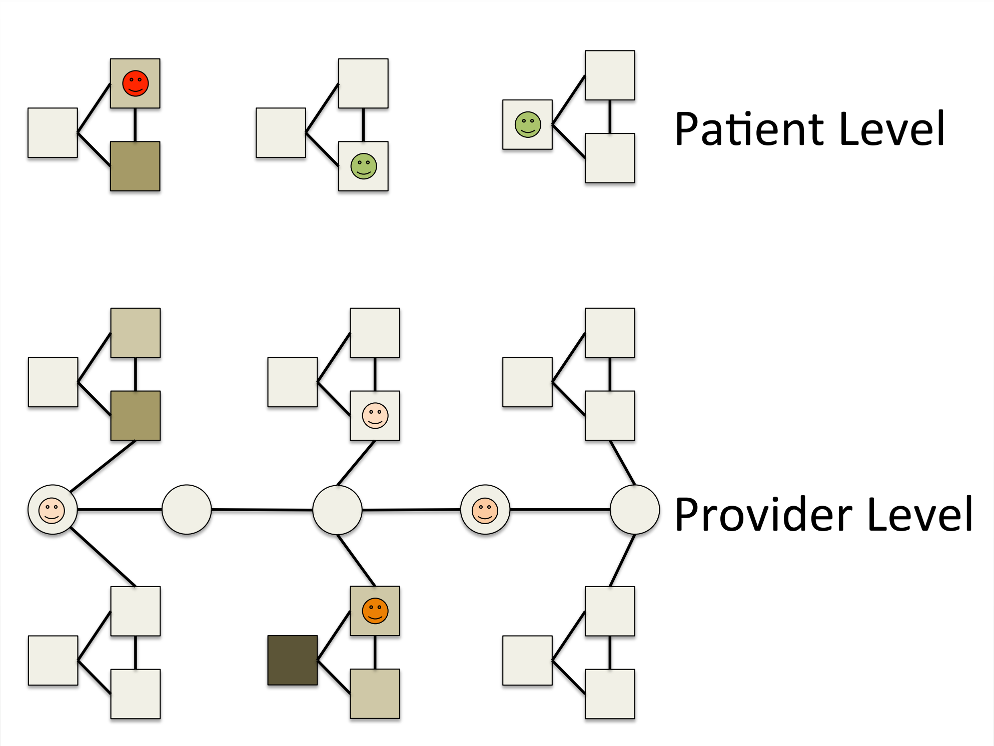


Transmission probability = 0.65

Where ABMs Are Particularly Strong

- Adding a type of stochasticity not present in other models
  - Random but rule-based mixing
  - Interactions with the environment
  - Positions, states and information about other agents
    - Responsive behavior
- Modeling different classes of individuals more easily
  - Draw parameters from a distribution, rather than a fixed value
  - Easily create a new type of agent by changing behavior rules

Environmental and State Sensing



Transmission probability (Civilian) = 0.65
Transmission probability (HCW) = 0.30
Treatment probability = 0.80

More Complexity

- What if p(Infection|HCW) was a distribution, representing experienced and inexperienced first responders?
- What if p(Infection|HCW) changed with time, representing fatigue?
- What if infected individuals move randomly *until* they see a HCW?
- What if they try to *avoid* HCWs?
  - This was the case for some Ebola patients
- How about adding terrain?

A Note of Caution

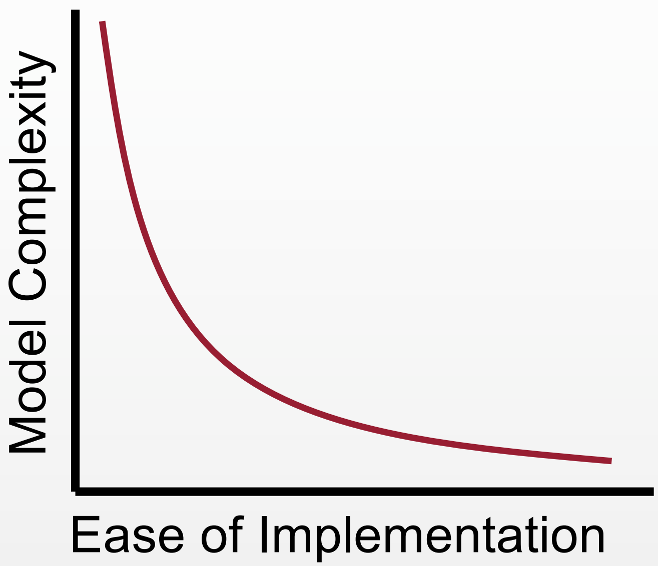
- Clearly, ABMs are a *very* powerful tool, and lend themselves well to sophisticated and complex models
  - Grouping and behavior processes, interaction with the environment, huge numbers of agents (a human body, an entire hospital, an entire healthcare system, an entire city…)
- It is easy to add complexity, it is hard to *implement* it
  - More complex models mean slower models
  - Parameter choices are difficult to find
- Easy to get carried away
  - Focus shifts to modeling the system, not the research question
- Randomness means you have to simulate the system *many* times

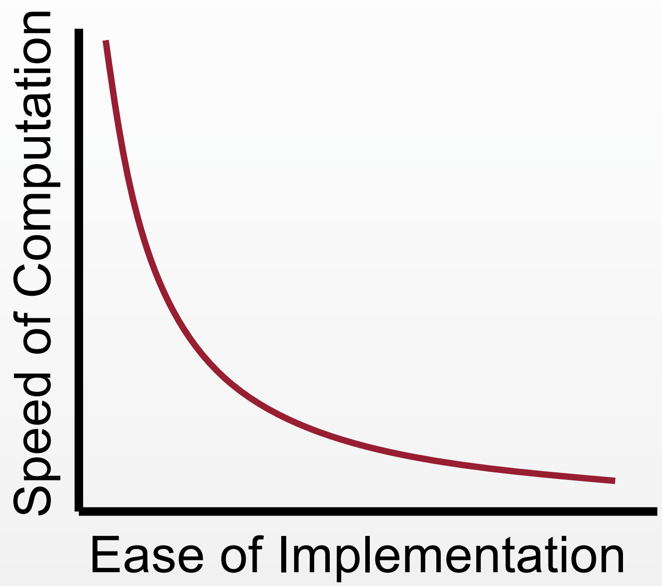
Other Tradeoffs

- Few analytical solutions
  - Simulation results instead of proofs
  - Those that do exist are *hard*
- Difficult to describe
  - Consider the figures in this presentation
  - Can use SIR-like flow charts, but harder to represent the whole population
    - No equations
    - Reproducibility is difficult
- Programming expertise
  - ABMs are software development projects as much as science

What Software Should I Use?

- Very common question
- Lots of possible options
- Open source, proprietary, graphical, etc.
- Could always break down and write your own
  - Lots of flexibility, lots of work
  - Isn't necessary just for learning
- Use what your colleagues/collaborators use

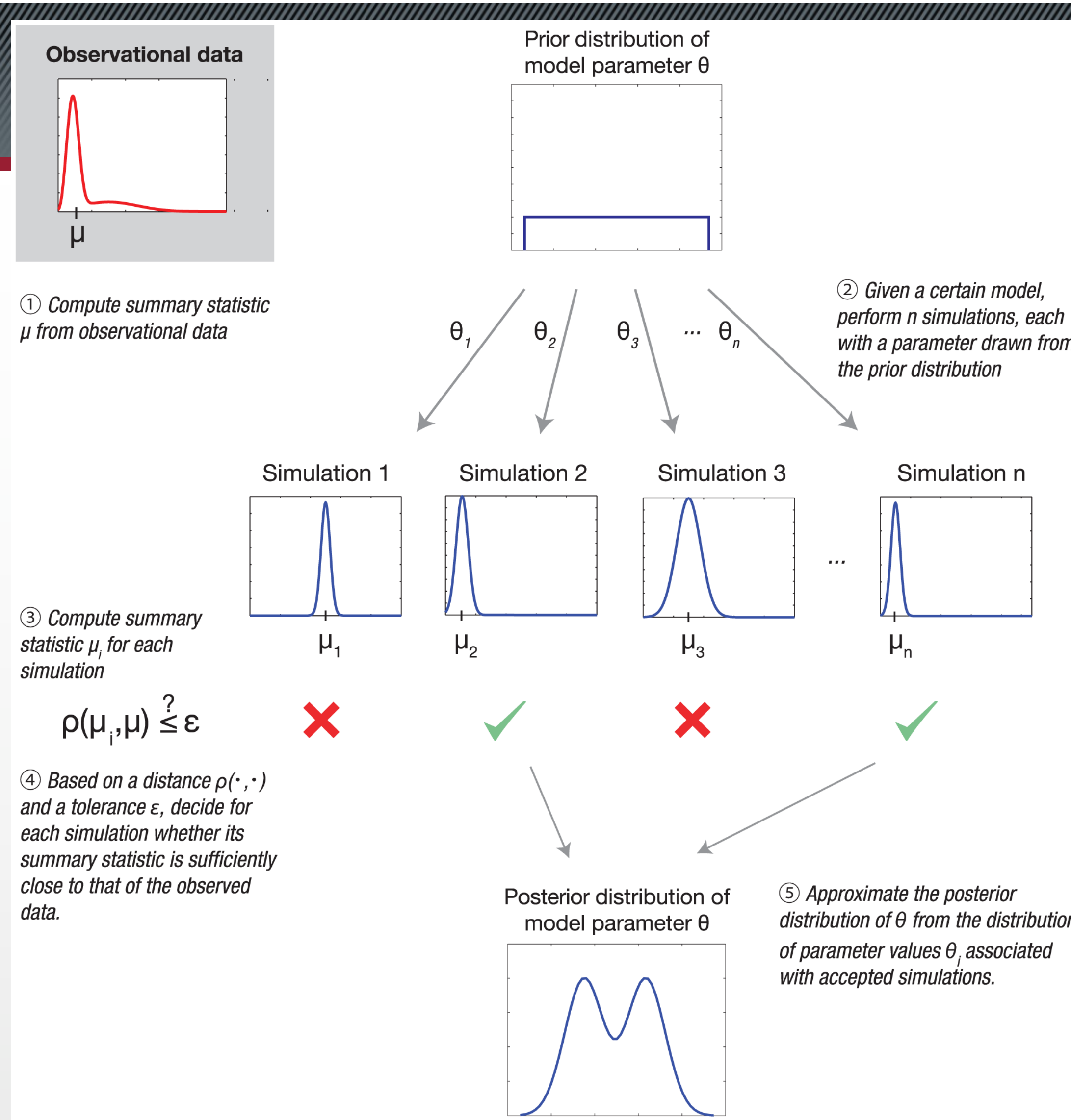

Model Complexity
Ease of Implementation


Speed of Computation
Ease of Implementation

Fitting Models to Data

- A whole multi-day workshop in its own right
- Much less straightforward with agent-based models
  - Multiple dimensions to try and fit
  - Stochasticity – does one run not fitting mean a bad fit, or randomness?
  - Approximate Bayesian Computation, particle filtering, pattern-oriented approaches (“calibrating to experience”), and many, many others

- Approximate Bayesian Computation
- In essence “Which values of a parameter we are interested in give us what we see in studies?”
- See Sunnåker et al. 2013. *PLoS Computational Biology*



Why This Is Useful

- Theoretical:
  - Likelihood-free inference
  - Acceptance may be made using summary statistics (with some loss of information), qualitative patterns (with lots of loss of information), etc. to constrain a model in ways that are difficult to capture using ML-methods.
    - Grimm calls this "Fitting to Patterns"
    - Simple observations about reality can rule out huge swathes of parameter space
- Practical:
  - Relatively straightforward to think about/implement

ABMs and Causal Inference

- Causal inference and Agent-based models sometimes feel at odds with one another
- Different heritage, different nomenclature, etc.
- Opinion:
  - They aren't
  - More strongly: Causal inference models *are* agent-based models with a series of constraints and assumptions imposed on them
  - ABMs provide *indisputably counterfactual scenarios*
  - But those scenarios may be about a fictional universe
  - How willing are you to step outside your data?