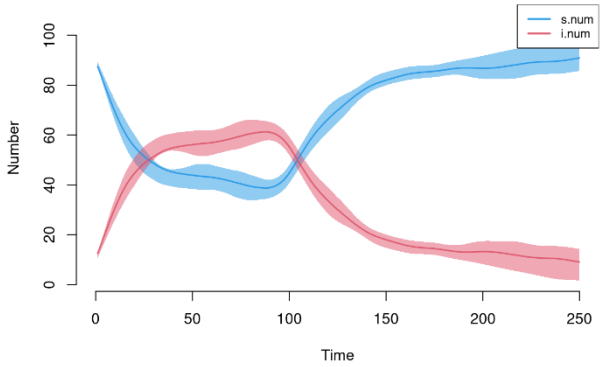


WELCOME!

NETWORK MODELING FOR EPIDEMICS

Martina Morris, Ph.D. University of Washington
Steven M. Goodreau, Ph.D. University of Washington
Samuel M. Jenness, Ph.D. Emory University

Development of methods, software, and learning materials presented throughout this course all supported by the US National Institutes of Health

Objectives for the course

Understand the principles and methods of network analysis relevant to infectious disease epidemiology

- Descriptive network analysis
- Statistical network analysis with ERGMs and TERGMs
- Empirical study designs for networks

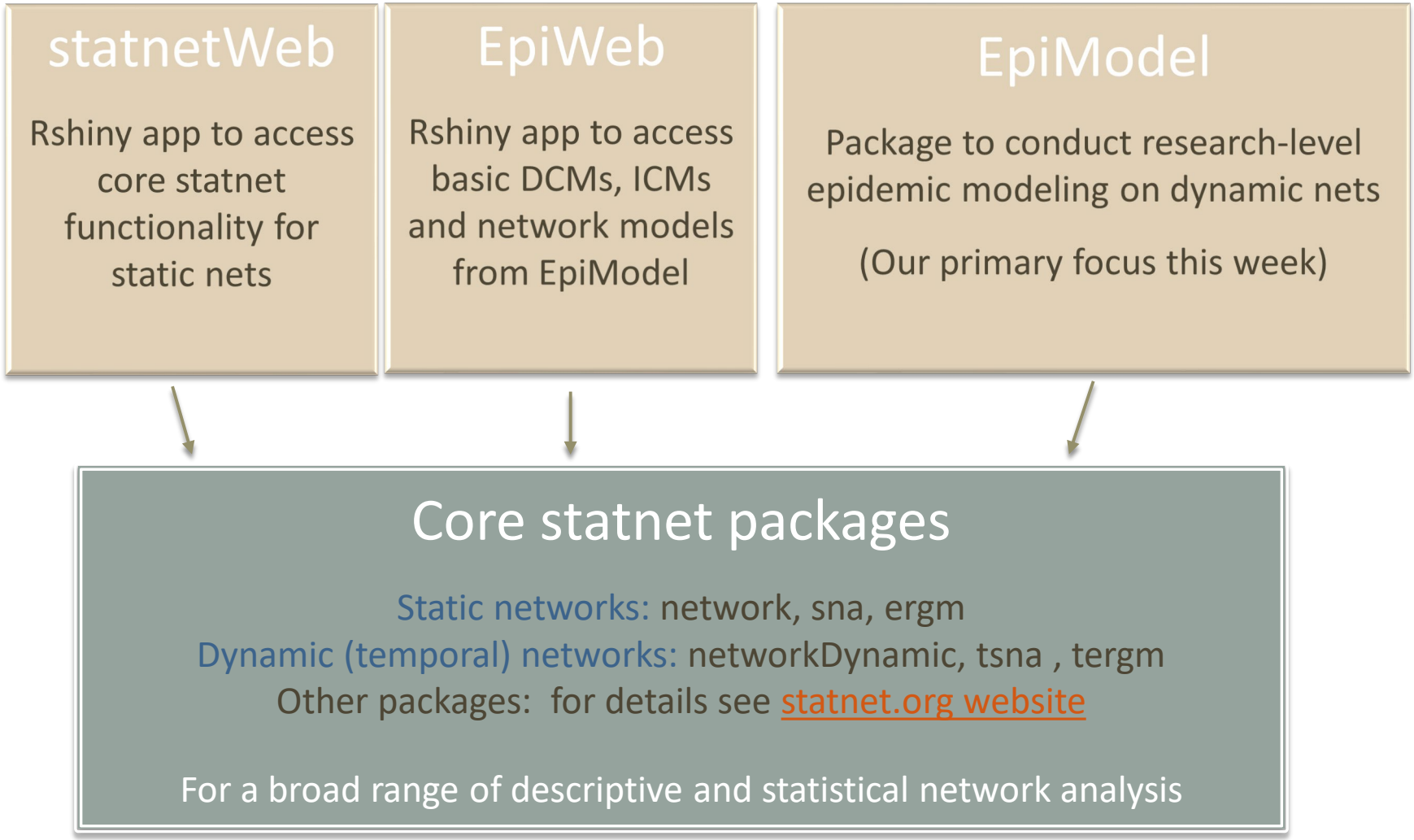
Develop the knowledge and software skills to run your own simple network transmission models, using R and the EpiModel package

Begin to learn how to extend EpiModel code for your own research applications

Lesson plan

Day	Module	Content
W	1	Intro, terms & concepts
	2	Statistical models for networks: theory
T	3	Statistical models for networks: practice
	4	Basic EpiModel in closed populations
	5	EpiModel: working with nodal attributes
	6	Data and network model parameterization
F	7	EpiModel in open populations (demography) pt. 1
	8	EpiModel in open populations (demography) pt. 2; visualization
	9	Extending EpiModel
	10	Discussion of projects; next steps; future resources

Software (all based in R)



Show of hands - who has experience:

- With epidemic modeling?
  - Using compartmental models? *
  - Using stochastic agent-based models? *
  - Using (full-fledged) network models? *
  - Using EpiModel?

What do we mean by these terms? We'll elaborate in a bit. For now just give your best answer.

Show of hands - who has experience:

- With R?

- With social network analysis?
  - Using descriptive methods?
  - Using statistical inference methods?
  - For static networks?
  - For dynamic networks?
  - Using the statnet suite of packages?

Whose research interests relate to:

- Human pathogens?
  - HIV?
  - Other sexually transmitted infection(s)?
  - Respiratory /airborne pathogen(s)?
  - Vector-borne pathogen(s)? (mosquitos, etc.)
  - Some other human pathogen?
- Animal pathogens?
- Diffusion of an intervention/behavior/information?
- Diffusion of something else entirely?

Background to epidemic modeling (1)

A lightning- fast overview

All* models, regardless of type, will contain the following ideas:

1. Time as a dimension over which the model unfolds
2. At least one type of element (aka agent; e.g. human beings)…
   - … of which there is a population…
   - … whose members are capable of being "infected"…
   - … and also capable of infecting others
3. At least one entity capable of doing the "infecting" (e.g. SARS-CoV2)
4. Some type of contact process by which the infection occurs
5. A record of whether and when the elements are infected

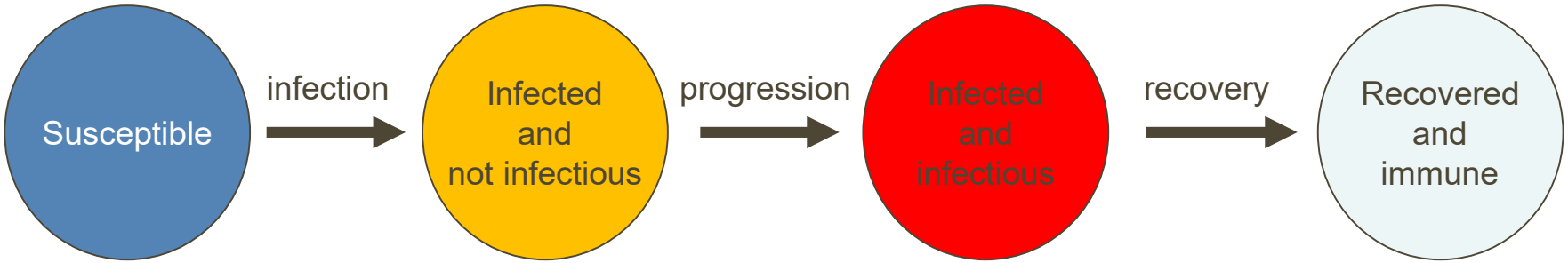
** pretty much; there are always weird exceptions to every rule*

Background to epidemic modeling (2)

A lightning- fast overview

Some models have additional infection statuses, e.g.

- recovered and immune
- infected but not yet infectious
- perhaps stages with different levels of infectiousness



Background to epidemic modeling (3)

A lightning- fast overview

Most but the very simplest of toy models will have:

1. Attributes of the elements (other than infection status), e.g.
   - demographic (sex, age….)
   - behavioral (level of sociality; occupation….)
   - clinical (tested or not; on treatment or not…)
   - geospatial (community; coordinates….)
2. Processes by which at least some of those attributes can change

Many consider attributes for the infectious agent as well, e.g.

- strain
- presence of specific mutations

Background to epidemic modeling (4)

A lightning- fast overview

Fundamental summary measure: R_0

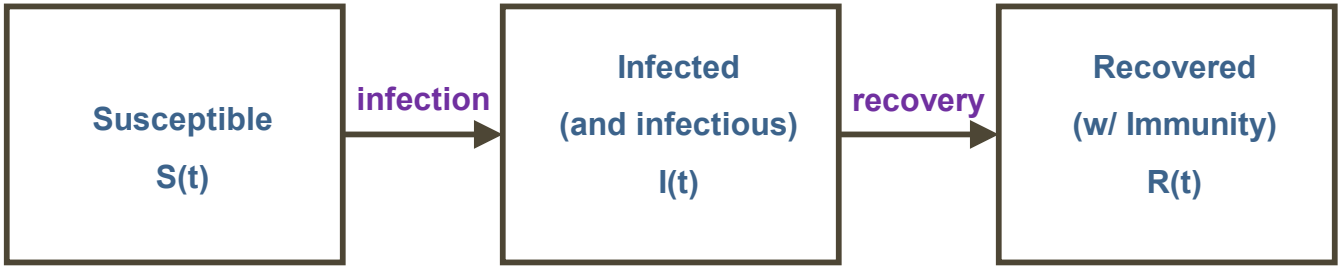
Captures the epidemic "persistence threshold" and velocity of transmission

Definition: The expected number of secondary infections generated by the first infected case in a population of susceptibles

Value of R_0	Implication
< 1	The first infected individual will on average infect < 1 person total. Transmission is too low for epidemic persistence
> 1	The first infected individual will on average infect >1 person total. Epidemic will typically grow and persist
= 1	Right on the threshold between persistence and extinction. Epidemic will typically just putter along

Deterministic compartmental modeling (DCMs)

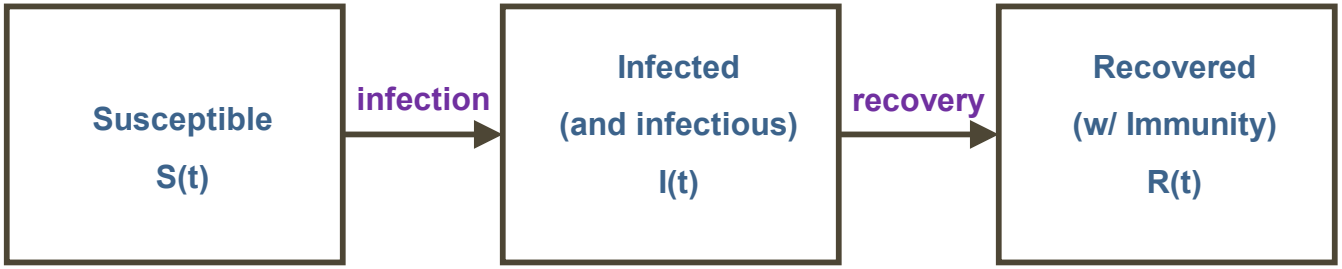
A lightning- fast overview



- Only the aggregate count in each state ("compartment") is represented, not individual persons
  - S(t) = # susceptible, etc.
  - Within each compartment, units are homogeneous
- Transitions ("flows") represent the aggregate count that moves from one compartment to another at any time point
  - Flows are represented by differential equations (or difference equations if in discrete time)

Deterministic compartmental modeling (DCMs)

A lightning- fast overview



Change in # of susceptible persons per time unit

$$\frac{dS}{dt} = -\beta\, SI/N$$

Infection flow

Change in # of infected persons per time unit

$$\frac{dI}{dt} = \beta\, SI/N - \rho I$$

Recovery flow

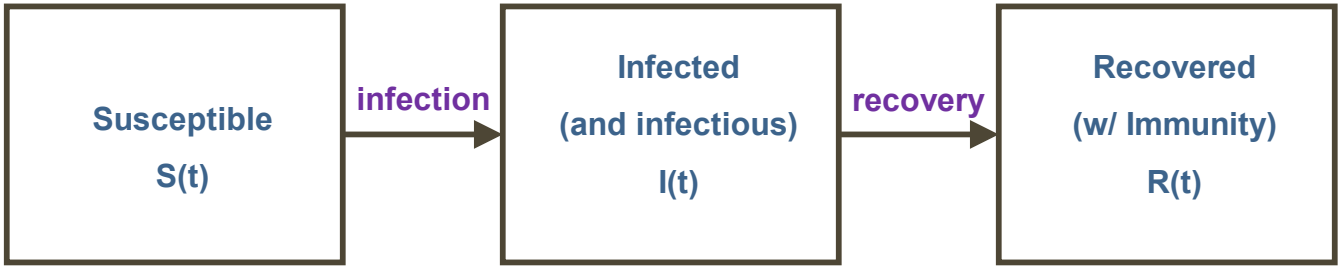
Change in # of recovered persons per time unit

$$\frac{dR}{dt} = \rho I$$

- β and ρ are model parameters
- Q: Where are the contact events?
- A: embedded in the infection flow

Deterministic compartmental modeling (DCMs)

A lightning- fast overview



Common notation for infections $\beta SI/N$ where β is called the "force of infection"

Can be disaggregated as: $\tau c SI/N$ where c = "contact rate"
τ = "transm. prob"

So: S susceptibles each have c contacts per unit time
I/N of the contacts are with infected
each susceptible-infected contact has probability τ of transmitting

Deterministic compartmental models

- Compartmental models are usually deterministic – each run gives the same result
- Measures = predicted counts (and represent the means of an equivalent stochastic process over an infinite number of runs)
- Compartments and flows can represent fractional persons

DCM strengths

- Familiar to many (and familiarity breeds comfort)
- Have a long body of literature identifying properties of different classes of models
- Simple models have simple closed form expressions for R_0
  - Intuitively, the number of secondary infections for the first case is:

    Contact rate/timestep * duration infected * transmission prob/contact

    c D τ
  - So for a simple SIR DCM: $R_0 = cD\tau$

DCM weaknesses

- Do not show the stochastic variation in a system
  - Stochastic CMs do exist, but only solve this one weakness
- Adding heterogeneity blows up quickly
  - Requires new compartments
    - e.g. adding 2 sexes means going from 3 compartments to 6: SF, IF, RF, SM, IM, RM
    - What if we wanted to add in 4 racial/ethnic groups? 5 ages? 5 categories of viral load? Testing? Treatment? Circumcision? Etc.
  - And if heterogeneity isn't in discrete categories?

DCM weaknesses

- Representing complex partnership network patterns is hard (or impossible, depending who you ask)
  - Non-random mixing on an attribute can be added into the incidence term easily enough
    - Raises additional questions in open populations where group sizes can change
  - But other partnership patterns are harder
    - Like a tendency to only have one partner at a time?
    - To be more (or less) less likely to have contact with your partner's partner?
    - Remember that compartments only considered people in the aggregate; individuals are not uniquely identified
- And there is no general method for jointly estimating the parameters of a system of partnerships like this

Individual-based models

- Represent each individual member of the population explicitly
- This might take the form of a data frame (in R speak)
  - Each row is an individual
  - Each column is an attribute
- Use code instead of equations to represent the relevant dynamic processes

IBM pseudocode

```
# Initial conditions
    # create a data.frame (nrow = # of agents, ncols = # of attributes)
    # assign infection status (S, I, R) as one attribute
    # assign all other attributes
# Simulate epidemic
for (at = 1:num.timesteps) {
     # infection
         # draw the number of contacts for that step
         # draw 1 pair of agents for each contact
         # filter to just the discordant SI pairs
         # flip coin for each pair to determine if transmission
         # do bookkeeping for new infections
     # recovery
         # identify infected elements
         # flip coin for each case to determine recovery
         # do bookkeeping for recoveries
      # update other attributes
         # exact code depends on the nature of the attribute
}
# process output
```

these can be made to depend on the attributes of the agents

IBM strengths

- Show the stochastic variability in these systems
- Can handle multiple forms of heterogeneity with relative ease
  - With individuals identified, they just get labeled
- Simple models have some simple closed form solutions for R_0
  - For examples, see this article

IBM weaknesses

- Representing heterogeneity *in the contact process* that creates the partnership network is still hard
  - Some example include queuing processes and "stub matching"
  - But these are not very realistic representations

- And here, too, there is no general method for jointly estimating the parameters of this complex process from data.

Final note on terminology

- Contacts vs. acts: a key distinction
  - E.g. think of sexual activity - when we say "# of contacts per year"
    - Does it mean number of sex acts?
    - Or numbers of different partners?

- From here on out, we will use the terms "acts" and "partners"
- This distinction matters for disease dynamics when there are repeated acts with the same person