DAY 1:

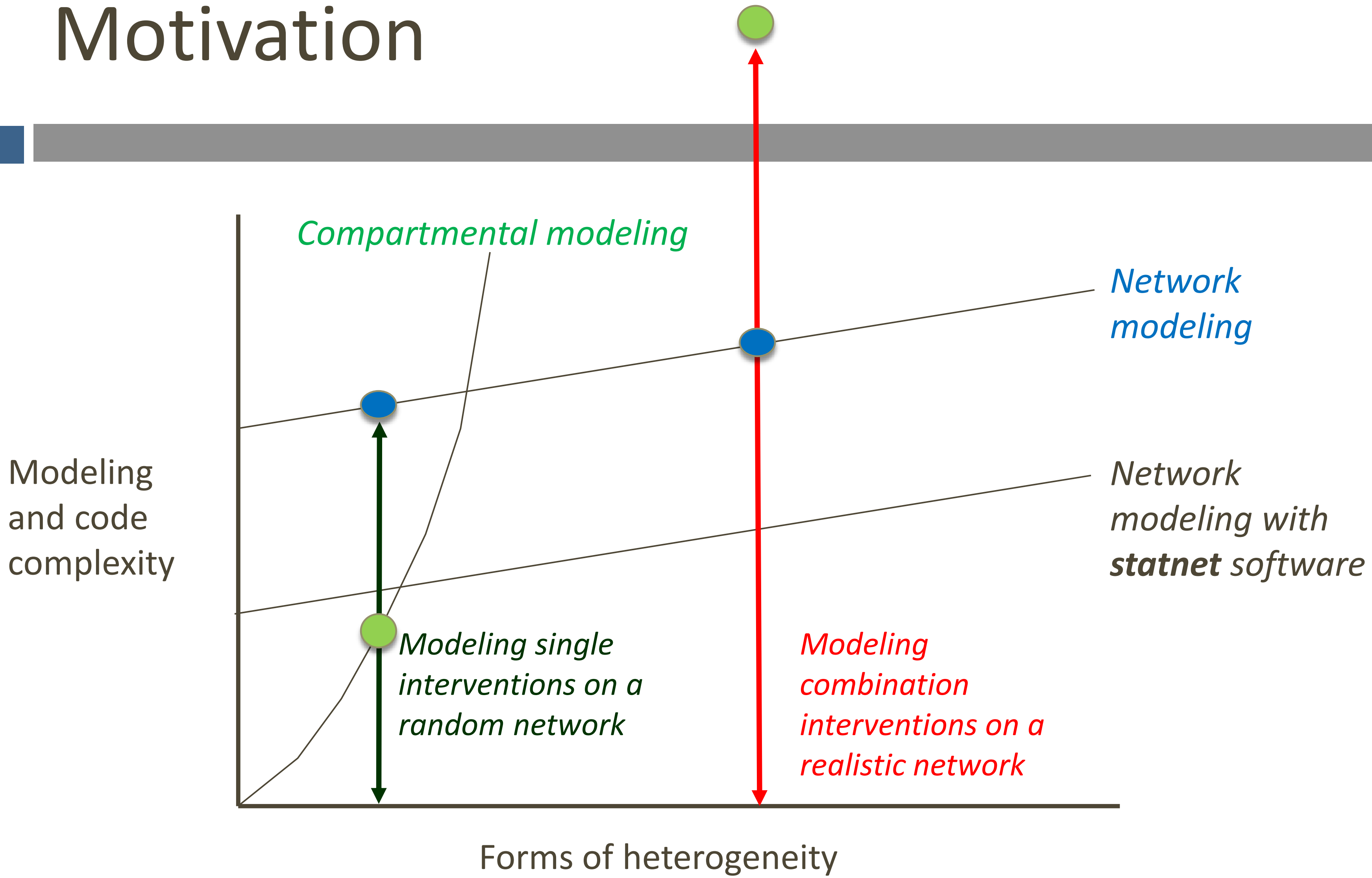
NETWORK THEORY AND TERMINOLOGY

NME WORKSHOP

Martina Morris, Ph.D.
Steven M. Goodreau, Ph.D.
Samuel M. Jenness, Ph.D.

Supported by the US National Institutes of Health

Motivation


Compartmental modeling
Network modeling
Modeling and code complexity
Network modeling with **statnet** software
Modeling single interventions on a random network
Modeling combination interventions on a realistic network
Forms of heterogeneity

Networks and Epidemics

- The key differences with the DCM are

Partnerships: Detailed control over formation, dissolution and network structure

Multiple partnership dynamics: The reachable path of infection over time

- Relevant epidemic contexts
  - Diseases that require proximity for transmission
    - E.g., childhood diseases, influenza, common cold
    - Households, workplaces, schools are places with repeated interactions
    - These cluster the network at an aggregate level, but there can be substantial contact across clusters in relatively short time frames (< 1 day)
  - Sexually transmitted infections
    - These structure the network at a micro level into components
    - There is often no contact across components, for long time frames (> 10 years)

Partnerships and R_0

- The traditional act-based R_0

$$\tau\alpha D = \frac{p(trans)}{act} \times \frac{acts}{time} \times time$$

Implies every act is independent

- Alternative partnership-based R_0

$$\varphi\kappa D = \frac{p(trans)}{pship} \times \frac{pships}{time} \times time$$

Implies every partnership is independent

Where: $\varphi = \{1 - (1 - \tau)^{\propto}\}$

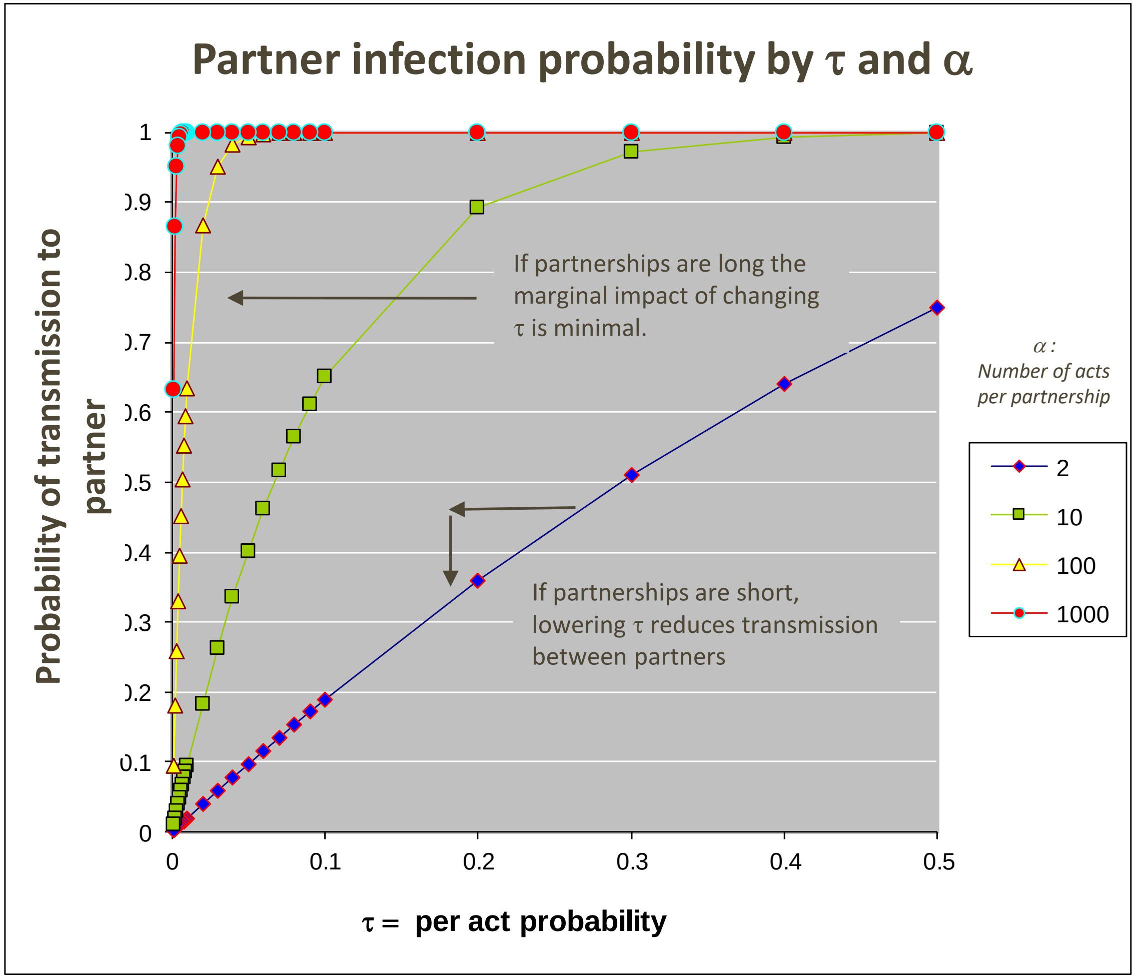
Partnerships vs. acts: transmission impact

Looking more closely:

$$\varphi = \{1 - (1 - \tau)^{\propto}\}$$

The more acts in a partnership, the less the effect of changes in τ.

Reductions in τ delay, rather than prevent transmission to a partner.



Partnerships are temporal

Sequence matters

- The *reachable path* in a transmission process is determined
  - Not by the partnerships at any single point in time
  - Nor by the cumulative total over time

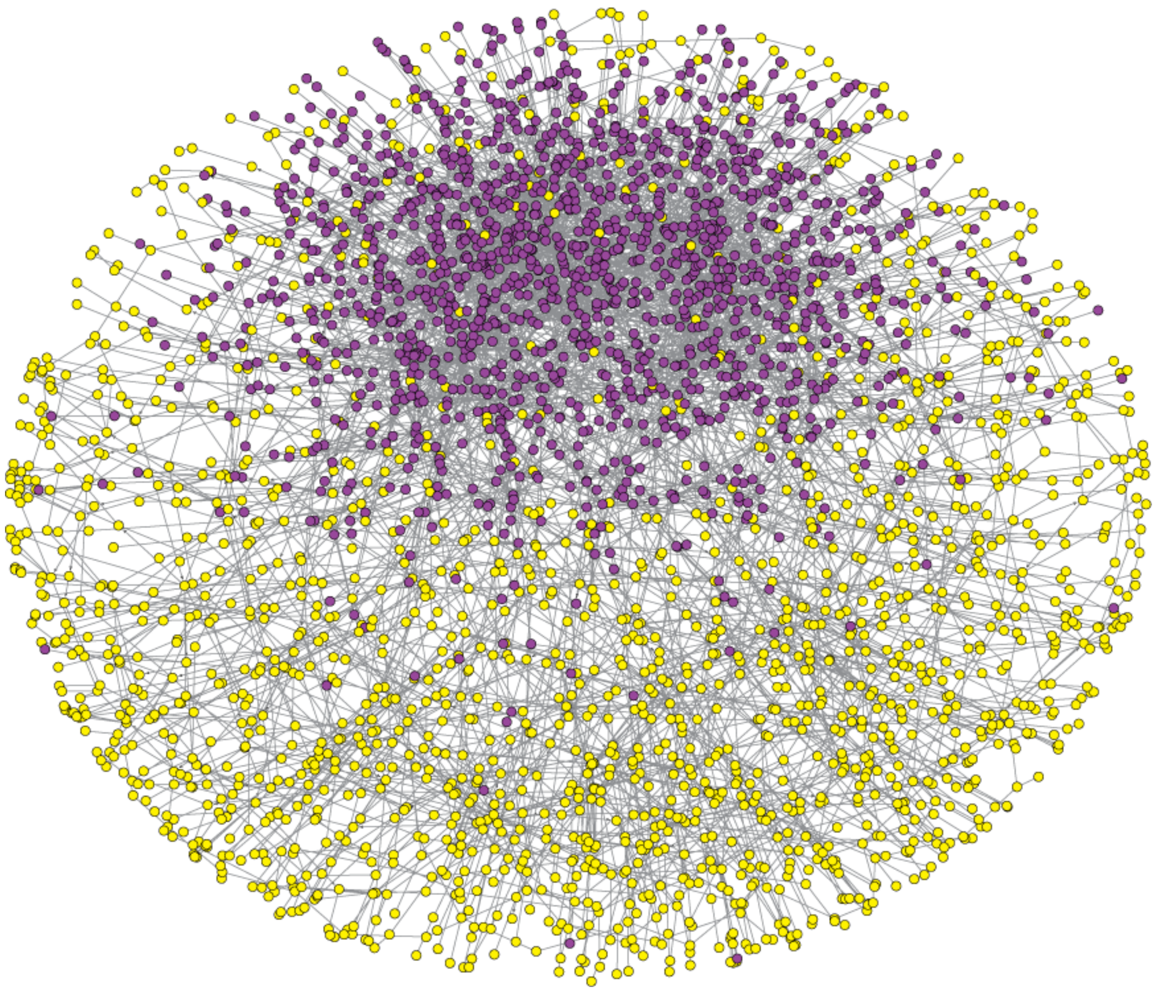
- By the *cumulative time ordered path of partnerships*

Cumulative connectivity: a network collapsed over time

10,000 nodes

2 demographic groups

10 years of partnership activity



99.7%

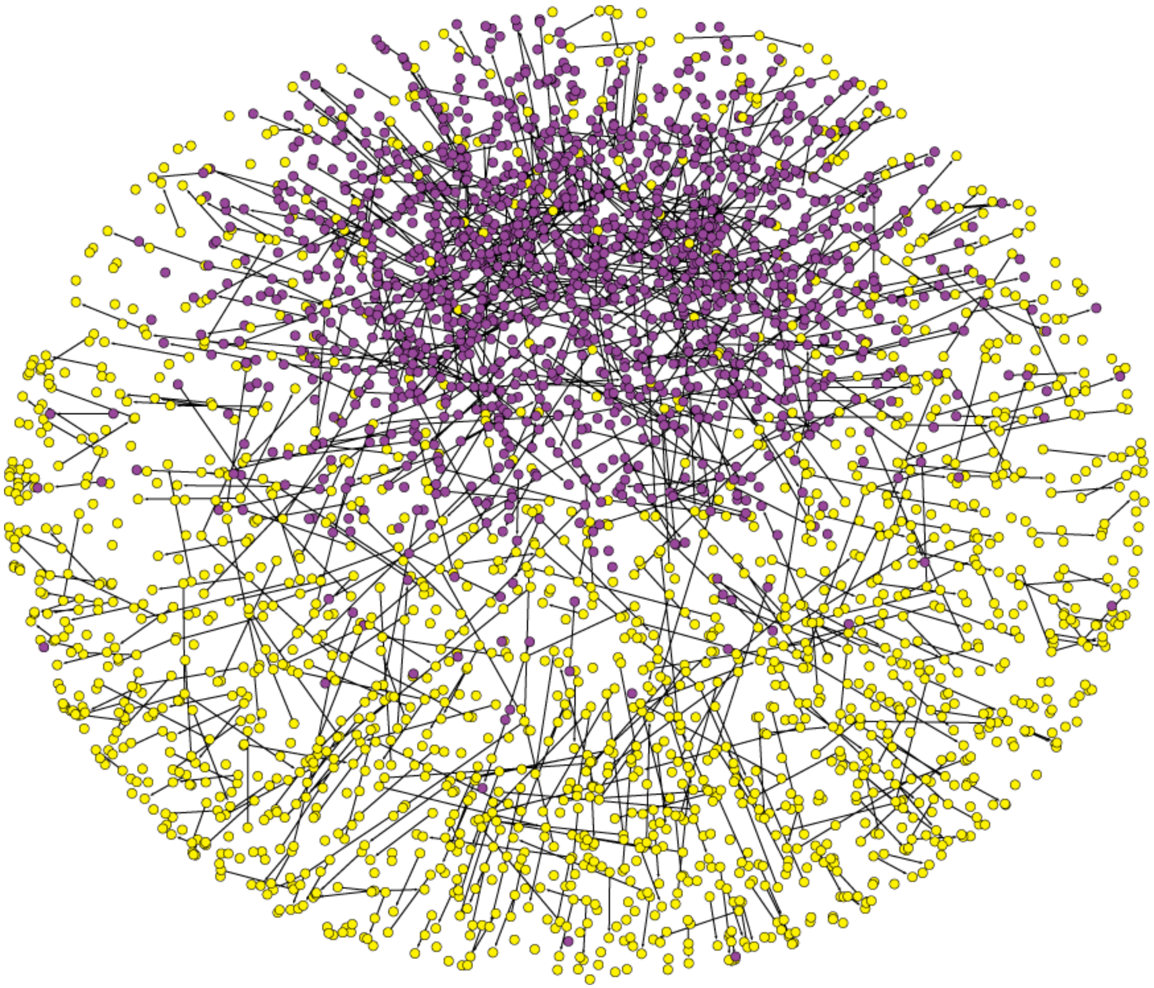
Connected

(nodes that can reach each other directly, or indirectly)

Daily connectivity: A momentary snapshot

Almost all components are size 2 or smaller

The largest components have 5-6 nodes



0.06%

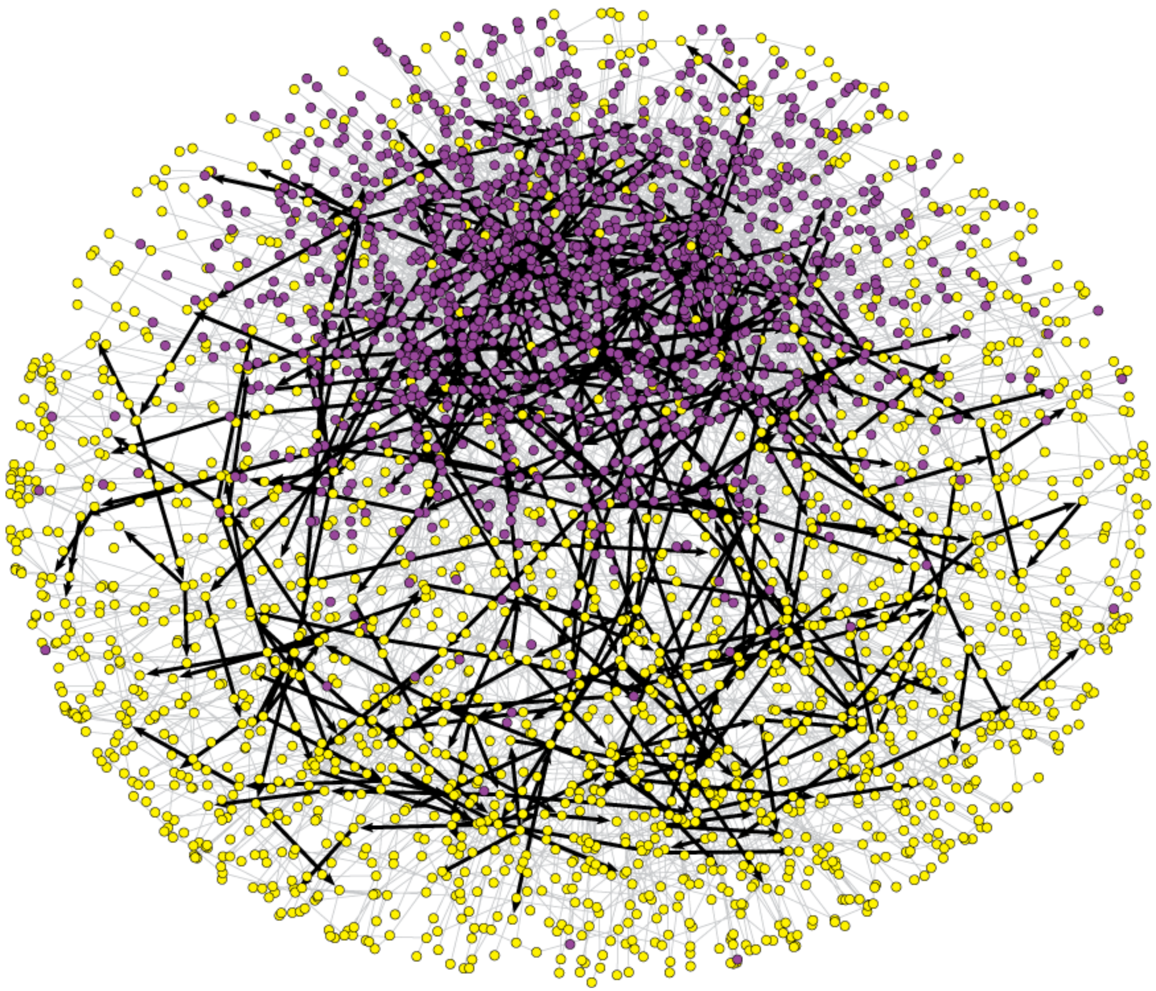
connected

The reachable path: different than both of these

The cumulative time-ordered path of partnerships

From 10 initial seeds

Over 10 years



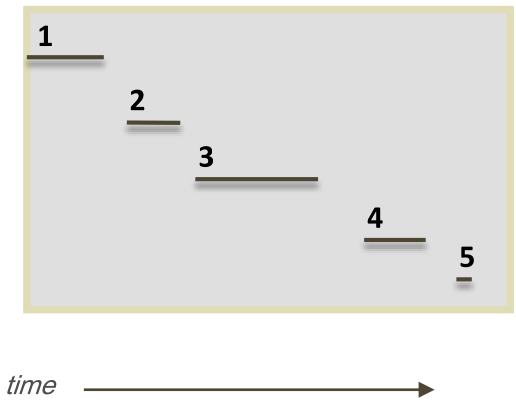
5.0%

Reachable

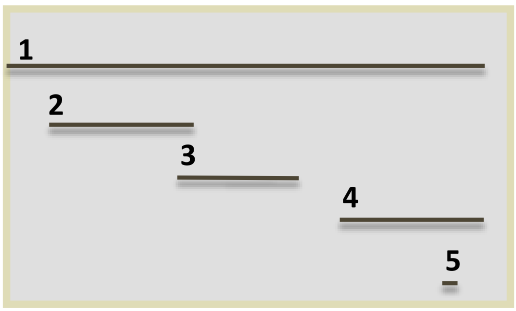
And timing matters

A sequence of partnerships can be either :

Serially monogamous



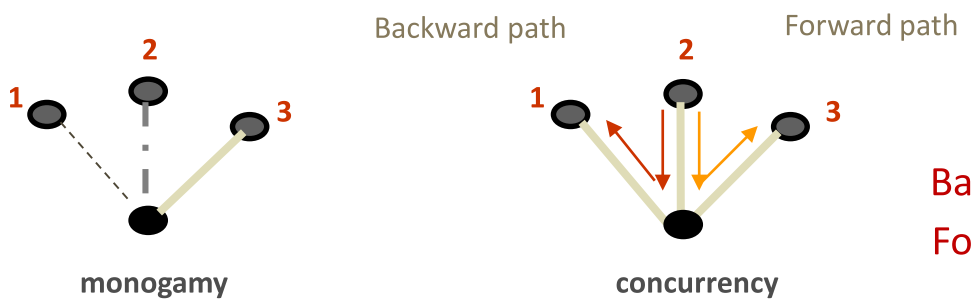
or Concurrent



Same contact rate (5/yr), but the sequence of start and end dates is different

Concurrency, reachable path and velocity

Concurrency removes the protection of sequence over time



Backward path: New chain of infection

Forward path: Shorter generation interval

Generates a unique *cross-sectional* network signature:

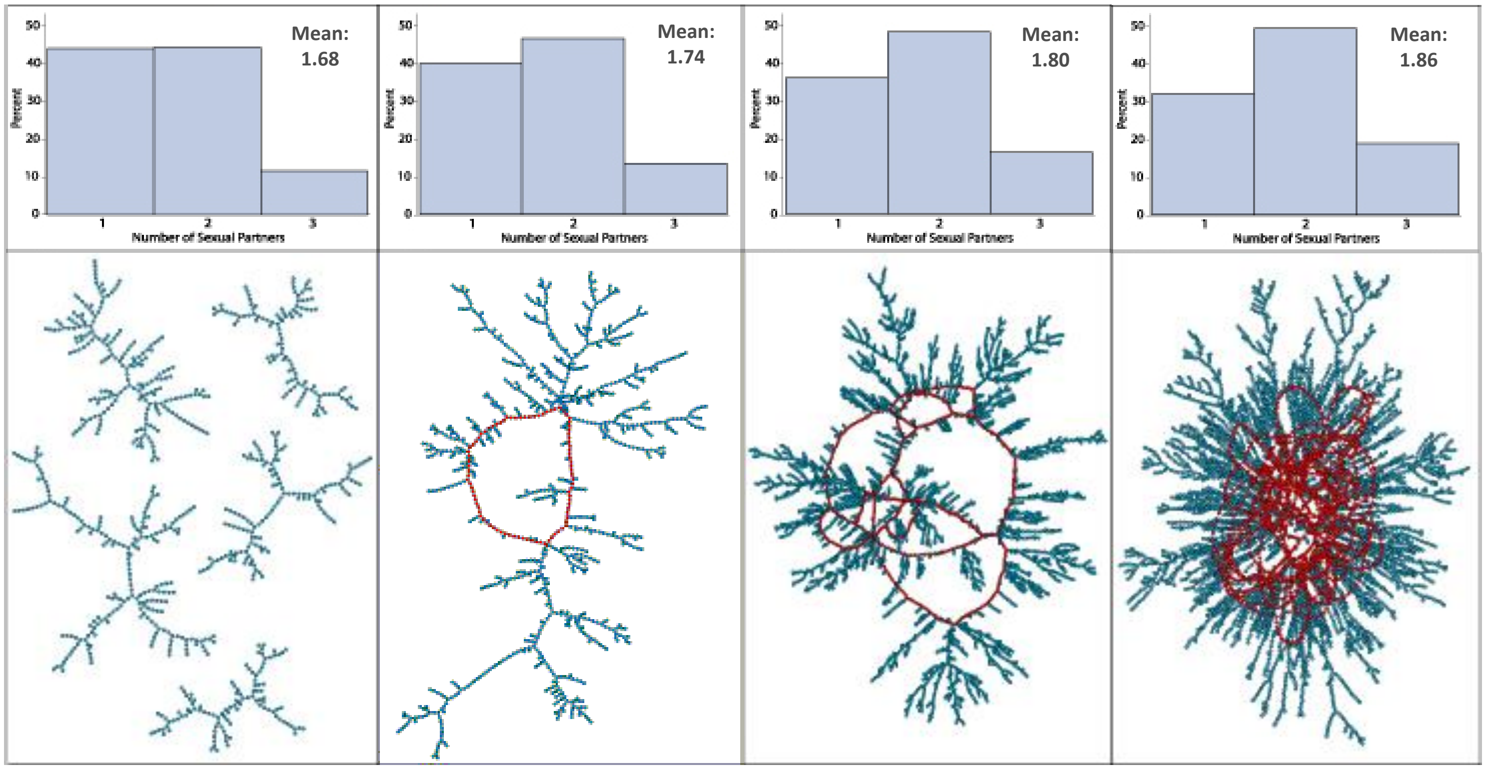
Larger components, the "concurrency superhighway" (Epstein, 2007)

Concurrency and the connectivity threshold

Number of ongoing partners on any particular day

Largest components

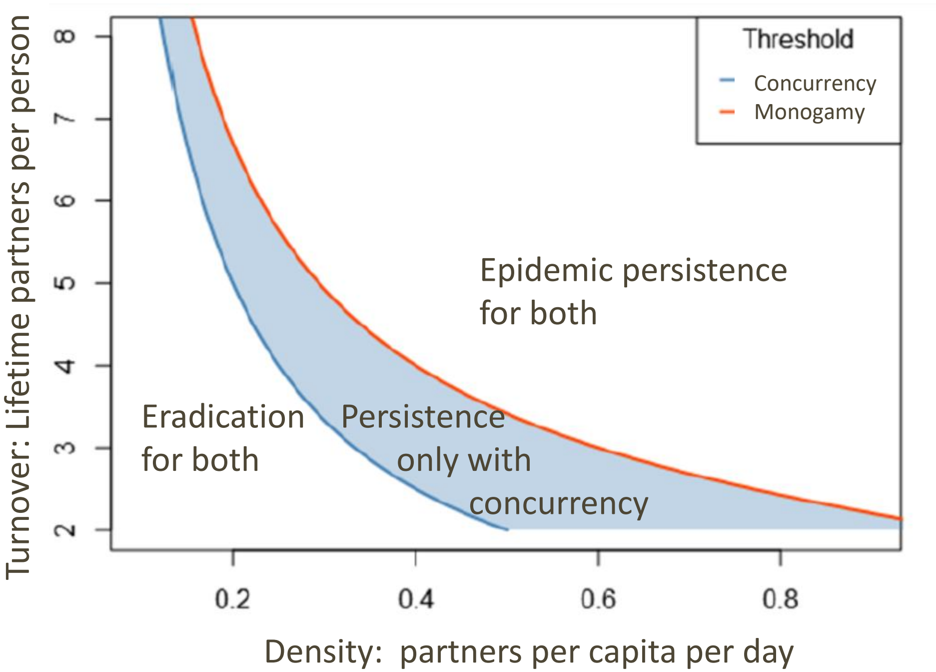
Bicomponents in red



In largest component:	2%	10%	41%	64%
In largest bicomponent:	0	1%	5%	15%

Ref: Morris, Goodreau and Moody 2007

Concurrency and R_0



- Epidemic persistence is a function of both static and temporal connectivity
  - **Density**: partners per capita on any day
  - **Turnover**: partners per capita over the lifetime
- Concurrency = more connectivity
  - For the same values of density and turnover
- So there is a region where epidemics persist only if there is concurrency

Armbruster et al. (2017)

Connectivity

- This is the key concept for understanding networks and epidemics
  - What is the network connectivity at any particular time?
  - What is the reachable path over time?

- It is also the key goal for the type of network modeling developed by our group
  - Generative models for network structure and dynamics
  - That can be estimated from data

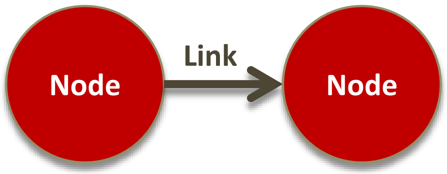
15 Network concepts and terminology

Through statnetWeb

Terminology

- **Node**: the entity of interest
  - for us, nodes represent people; also called actors or vertices

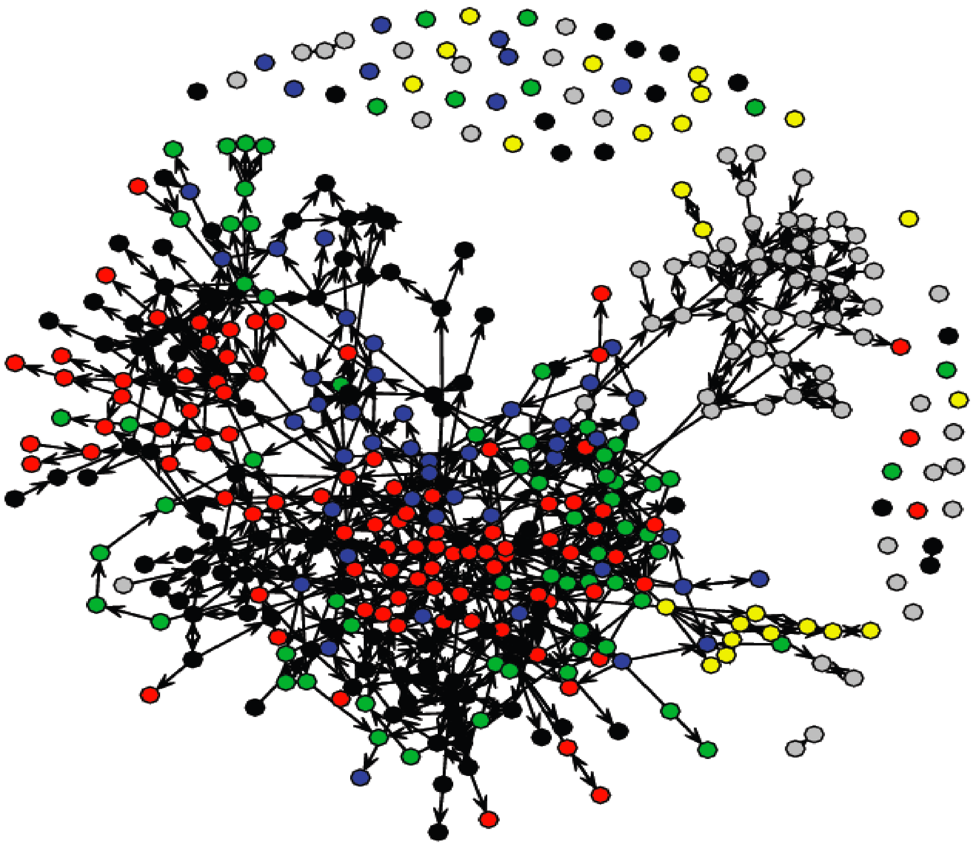
- **Link**: the relationship of interest
  - also called a tie, an edge, or a line



- **Network**: a collection of nodes and links
  - also called a graph

Nodes, links and networks

Beyond the pretty pictures, there are many different attributes of nodes, links and networks that have implications for the structures we can observe, and what we want to model



Types of nodes

- Individual units
  - Humans
  - Animals
  - Airports
  - Computers
  - Genes
- Collectivities
  - Countries, cities
  - Families
  - Species
  - Organs, Sensory systems

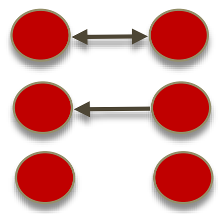
In social networks, a focal node is called “ego”, and the nodes linked to this focal node are “alters”

Types of links

- Social
  - Affective (like/dislike, trust/do not trust)
  - Kinship / social role (mother of, brother of, boss of)
  - Exchange (advice seeking, sexual intercourse, trade)
  - Cognitive (knows/does not know)
  - Affiliation (belongs to, is a member of)
- Physical
  - Road
  - Flight path
  - Wire / Wireless

Link properties

- Directed (e.g., likes)
  - Mutual
  - Asymmetric
  - Null
  - A directed graph is also called a di-graph
  - A directed edge is also called an arc



Nodes are now classified as senders and/or receivers

- Undirected (e.g., has sex with)

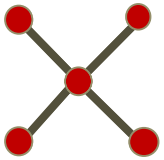
- Binary (0,1 on or off only)

- Signed and/or Valued (… -2, -1, 0, 1, 2 …)

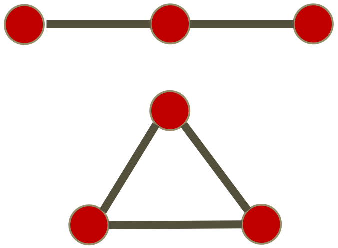
Configurations

Dyads

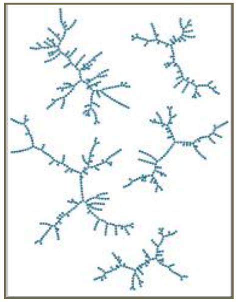
Stars



Triples & Triangles



Components



Any collection of nodes and links can be defined as a configuration

Levels of measurement

As we look at ways of describing network data, keep in mind the different levels of measurement

- Node level: *attributes of each individual node*
  - Examples: age, sex, infection state, degree
- Dyad level: *attributes of pairs or edges*
  - Examples: type of relationship, duration
- Component level: *subgraph attributes and distributions*
  - Examples: size, density, degree and geodesic distributions …
- Network level: *overall structural attributes and distributions*
  - Examples: density, degree, geodesic distribution …

E.g. Cycles

- Paths that lead back to the origin node
  - Cycle length k = number of lines in the cycle, “k-cycles”
  - Triangles are 3-cycles

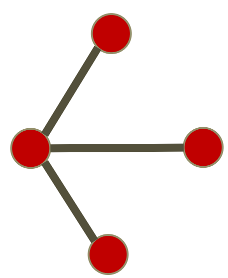
- Node level measure: Number of cycles a node is a member of

- Edge level measure: Number of cycles an edge is a member of

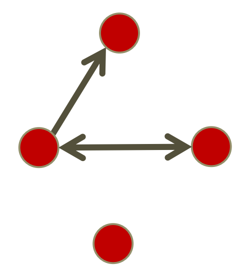
- Network level measure: The “cycle census” is a property of the network

Common network level measures

- Density: Fraction of all dyads that have an edge
- Isolate count: Number of nodes without any edges



Nodes: 4
Isolates: 0
Dyads: (4*3)/2 = 6
Density: 3/6 = .5



Nodes: 4
Isolates: 1
Dyads: 4*3 = 12
Density: 3/12 = .25

Types of networks

- Simplest form: 1-mode, undirected, binary ties, single relation

- 2-mode (aka *Bipartite*)
  - Two different types of nodes
  - Ties only allowed between groups

  Examples: Online network groups and persons (an Affiliation network)

  Heterosexual sex network

- Directed
  - Allows for a distinct set of in-ties and out-ties, and mutual

- Multiplex
  - More than one type of link possible

  Example: Sexual partnerships and needle sharing

Representing network data

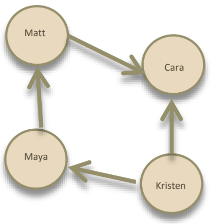
- Sociomatrix
  - (aka adjacency matrix)
  - simple but inefficient for large sparse nets (order n^2)

	Matt	Cara	Kristen	Maya
Matt	0	1	0	0
Cara	1	0	0	1
Kristen	0	1	0	1
Maya	1	0	0	0

- Edgelist

Matt	Cara
Cara	Matt
Cara	Maya
Kristen	Cara
Kristen	Maya
Maya	Matt

- Graph



Intro to statnetWeb

Core statnet packages

(network, sna, ergm, tergm, networkDynamic)

For various forms of descriptive and statistical network analysis

statnetWeb

User-friendly GUI to access main statnet functionality

Days 1-2

EpiModel

Package to conduct network-based epidemic modeling

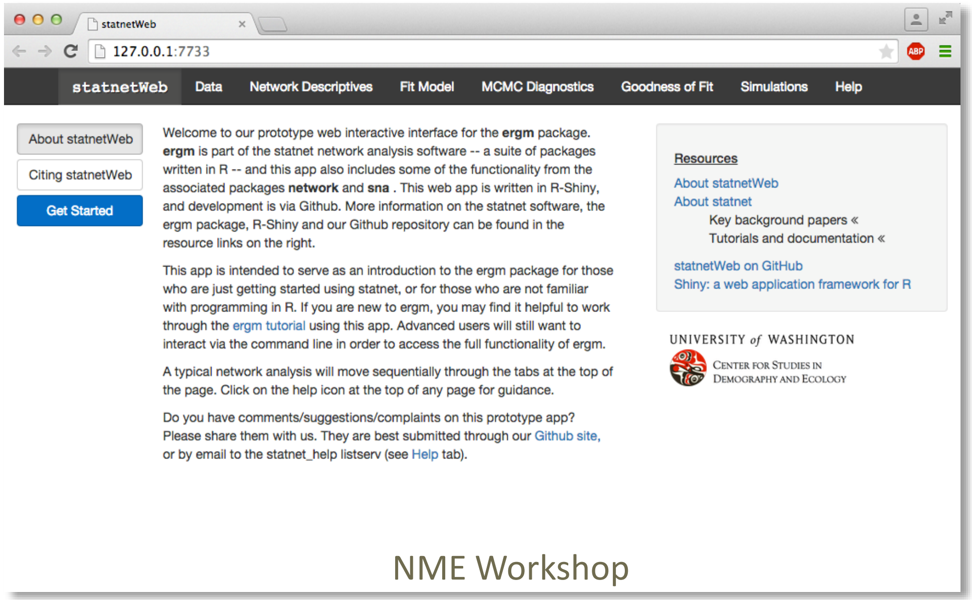
Both web GUI and command-line versions

Days 1-5

Intro to statnetWeb

statnetWeb is a graphical user interface for network analysis

- Runs in a web browser
- Wiki: https://statnet.csde.washington.edu/trac/wiki/statnetWeb



Accessing statnetWeb

THROUGH THE WEB

https://statnet.shinyapps.io/statnetWeb/

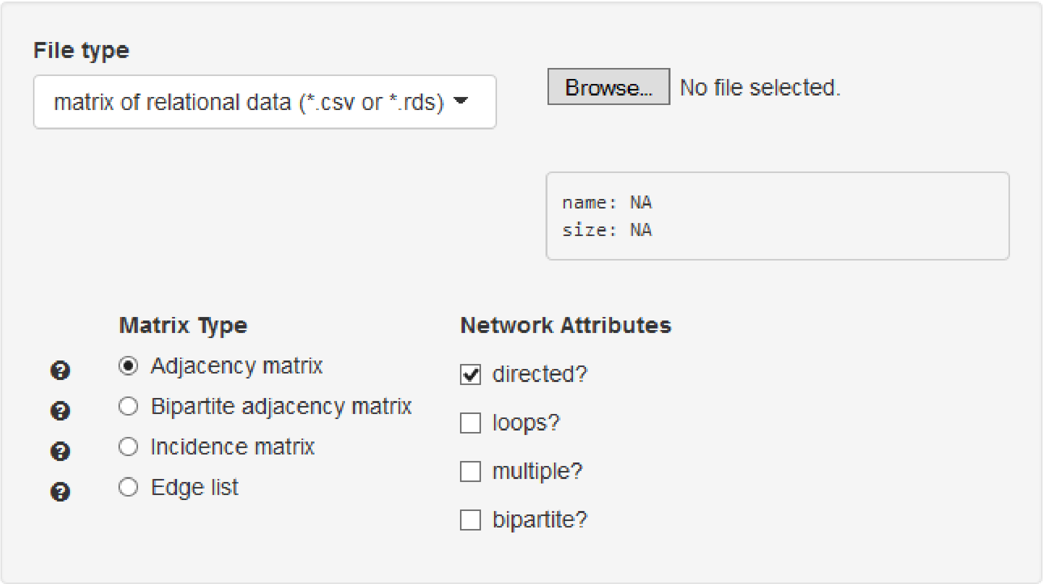
- No knowledge of R needed
- Everything runs on the host server
- Session will time-out after 15 minutes of inactivity
- Closing the browser window ends your session
  - Cannot restart where you left off
  - Download plots and summaries before exiting

THROUGH R

- Install the statnetWeb package
  - `install.packages("statnetWeb")`
- Load the package and launch the application
  - `library(statnetWeb)`
  - `run_sw()`
- No internet connection needed
- Everything runs on your local machine
  - Session will not time-out from inactivity
- Closing the browser window ends your session

Network data in statnetWeb

- On the “Data” page, you can upload networks from multiple sources
  - Internal: built-in networks
  - External: R, Excel, Pajek files

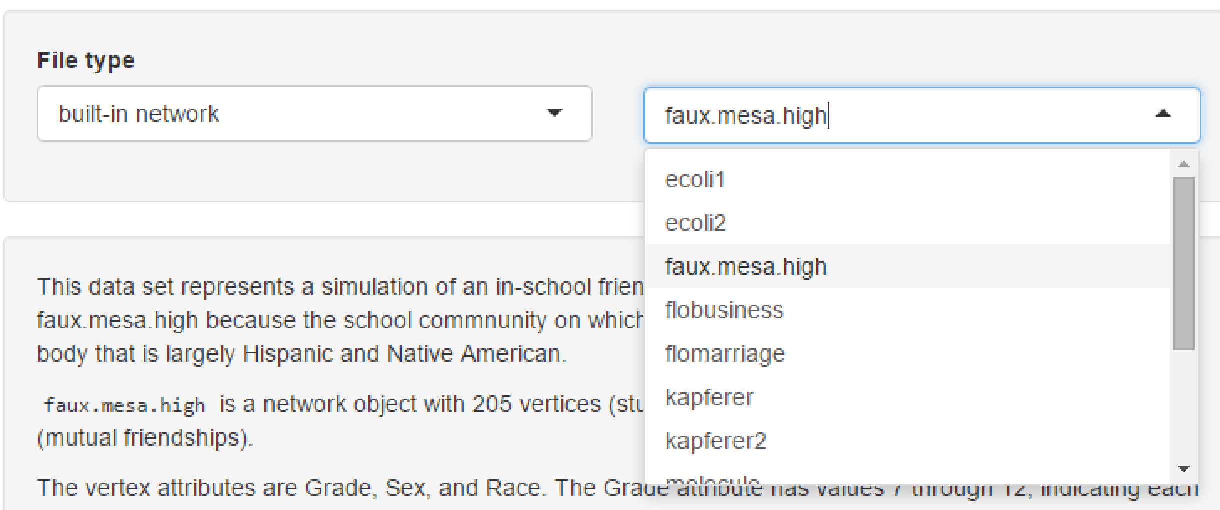


Examples in statnetWeb

- Load the "faux.mesa.high" network

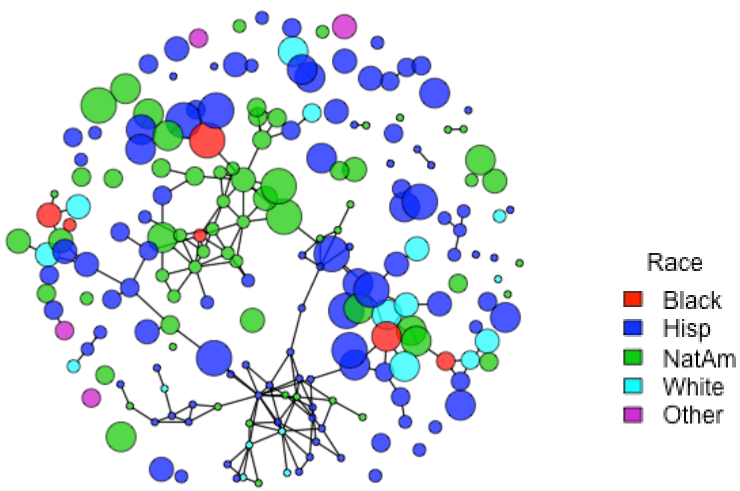
  High school network simulated from Add Health data

- We'll explore more network concepts using these data



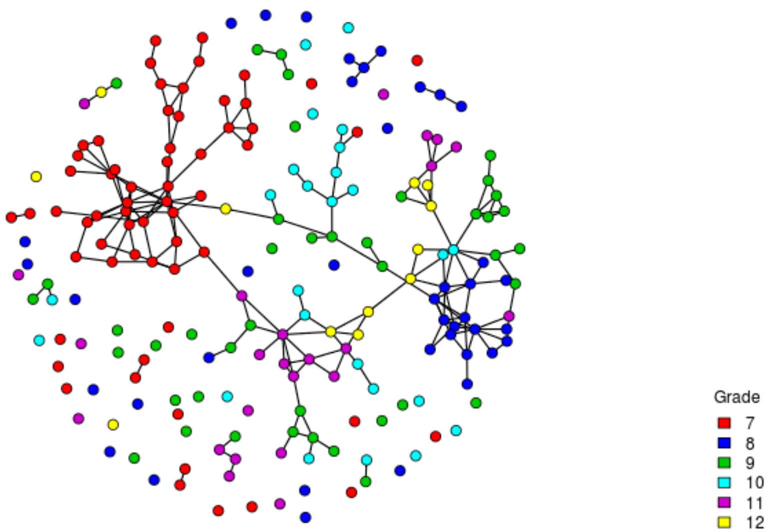
Attributes

- Individual nodes can have attributes like age, race, sex, etc.
- *Explore:*
  - *Color-code or size nodes on the network plot*
  - *Sort or search attributes in the interactive table*
  - *Look at histograms of attribute counts*
  - *What can you say about the structure of the network now?*



Node mixing by attribute

- Collapses the adjacency matrix into categories
- Cell counts = # links between nodes in row and col. categories



Mixing Matrix

Choose attribute

Grade

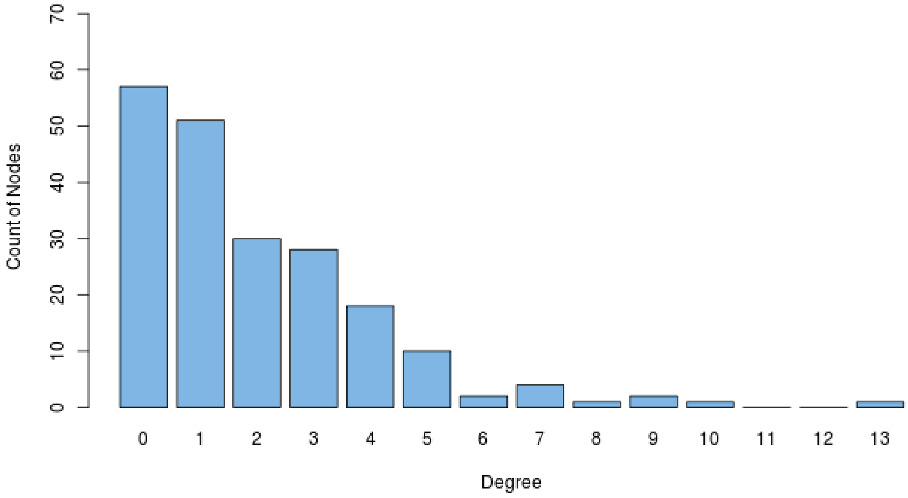
```
Note:  Marginal totals can be misleading
 for undirected mixing matrices.
     7   8   9  10  11  12
7   75   0   0   1   1   1
8    0  33   2   4   2   1
9    0   2  23   7   6   4
10   1   4   7   9   1   5
11   1   2   6   1  17   5
12   1   1   4   5   5   6
```

Measuring degree

- Node level: The number of edges "adjacent" to a node
  - Every node has a degree deg(i)
  - Di-graphs have in- and out- degrees, ideg(i) and odeg(i)
    - Indegree: the number of arcs that terminate at n_i
    - Outdegree: the number of arcs that orginate from n_i
- Network level: The degree distribution
  - Well-known parametric degree distributions: Uniform, Binomial, Poisson, Power-law
  - An empirical degree distribution may or may not resemble any of these

Degree distribution

- The degree distribution is a basic structural property
- To view it in statnetWeb:



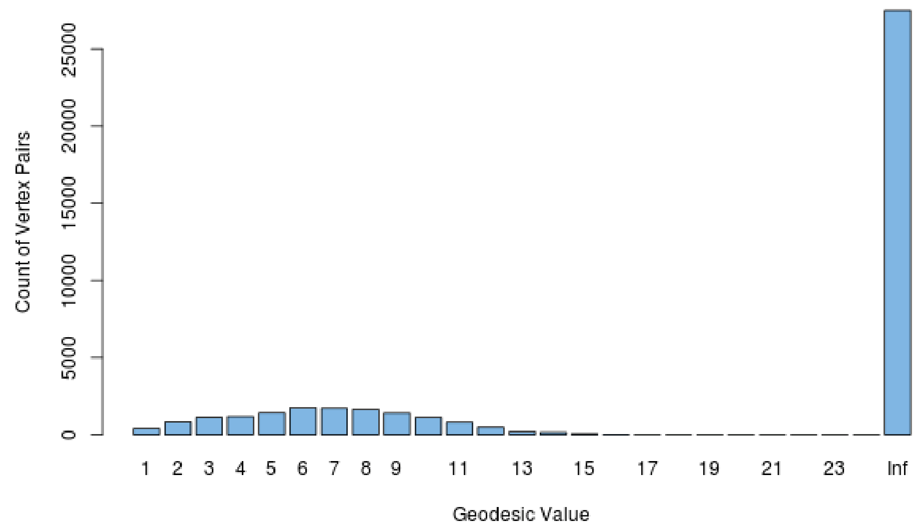
Connectivity measures: Geodesic

- Nodes are **reachable** if there is a path between them.

- A **geodesic** is the shortest path between two nodes
  - Two nodes have an infinite geodesic distance if they are unreachable

Geodesic distribution

- The geodesic distribution is another basic structural property of a network
- To view it in statnetWeb:

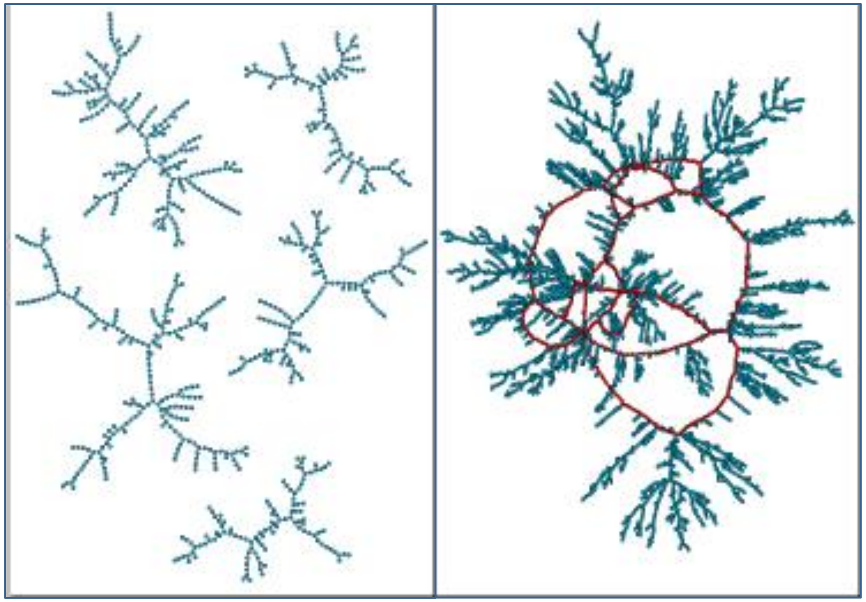


The last bar represents the node pairs with infinite geodesic distance

Connectivity measures: Components

- If some node pairs are unreachable, the graph will have multiple "**components**"
  - subgraphs of reachable nodes
- The component size distribution is another basic property of the network

Multiple components   1 giant component



NB: Think about how this connectivity comes to be created…

Description vs. Inference

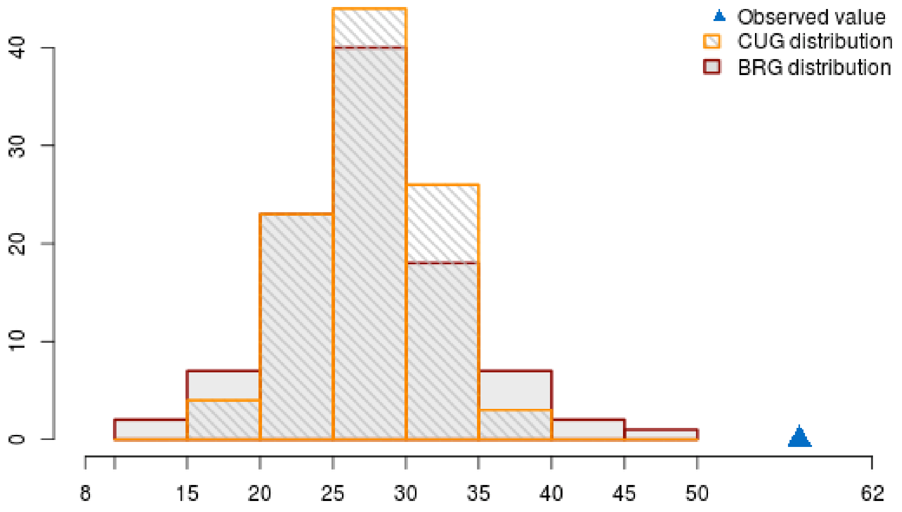
- So far we have been using descriptive statistics to explore our data
- Next, we might want to compare these statistics to what we would expect by chance
  - What do we mean “by chance”?
  - Can we use inferential statistics to draw more general conclusions?
- Need to define a reference distribution to act as the “null model”
- Two common null model distributions
  - Conditional Uniform Graph (CUG): Same density as the observed net
  - Bernoulli Random Graph (BRG): Same tie probability as the observed net

Using null models

- Select a summary measure for the observed data
- Compare it to the distribution simulated from a null model
- In statnetWeb:
  - We can conduct CUG tests for network summary measures
  - We can plot overlays on degree and geodesic distributions

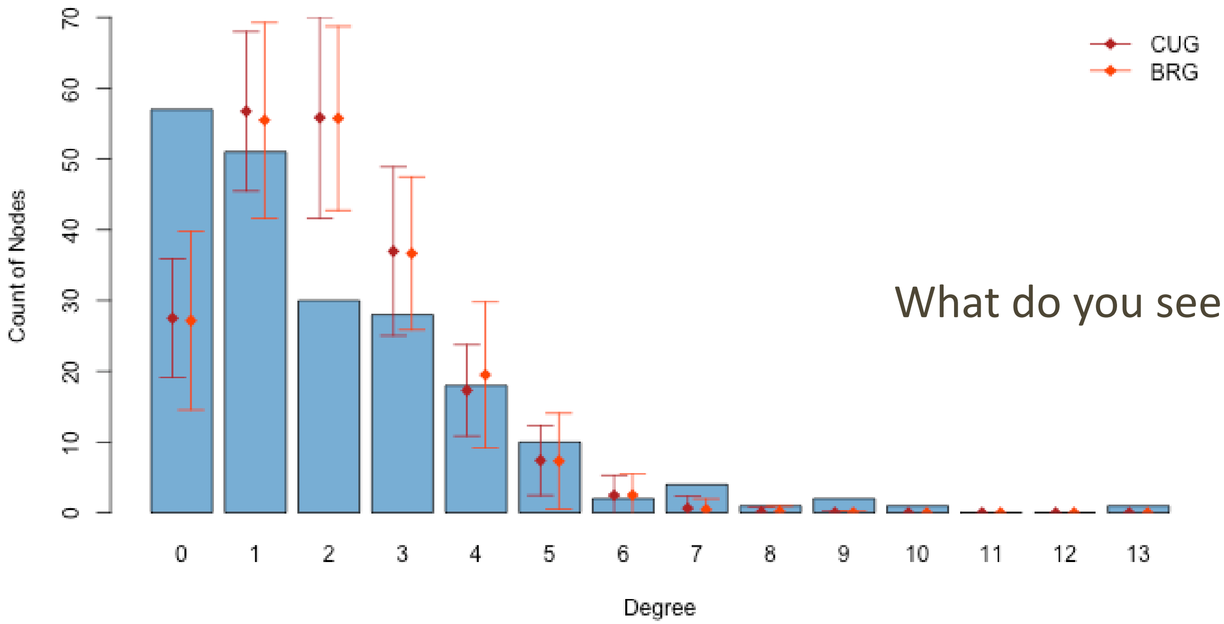
Conditional uniform graph tests

Compare the number of isolates in faux.mesa.high to what we would expect by chance



Degree distribution comparisons

In the degree distribution, add overlays for each null model



What do you see now?

Mean and 95% confidence intervals from 50 simulations are plotted

Where do we go from here?

Tomorrow:

Statistical Network Modeling
with Exponential-family Random Graph Models
(ERGMs)