



1

Network Theory & Terminology

We'll start with some motivation...

Assumptions redux

- One more assumption to tease out for ICMs
 - And it's similar to one we saw in the DCMs
- The intuition is there in the poker chips
 - Think about the draws from the bag
 - Yes, they assume random mixing
 - But what else?

ICMs: *Revisiting pairs*

- In the code, pairs are formed at Step 2
 - Draw the set of pairs that will have an act at that time step

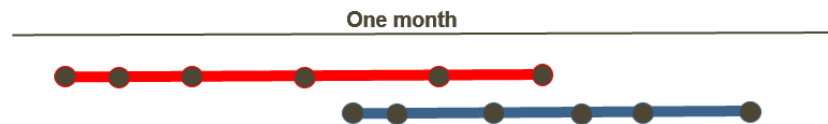
```
# Make edgelist of partnerships by ID number  
el <- t(replicate(acts, sample(1:num, 2)))
```

```
      [,1] [,2]  
[1,]    80    9  
[2,]     9   59  
[3,]     5   66  
[4,]     4   84
```

- At the next time step
 - What is the probability that any of these pairs of IDs will be drawn again?
 - So what is the implicit assumption?

Partnerships in the DCM framework

- That framework also has trouble representing persistent partnerships (as we saw in the earlier session)
 - If multiple acts occur within partnerships, DCMs take one of two forms.



1. Define a contact as an act. Model each act as a separate independent event, ignoring the persistent nature of the partnerships
2. Define a contact as a partnership. Compress all of the acts over the partnership into a single instance in time

Partnership assumptions 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: $\varphi = \{1 - (1 - \tau)^\alpha\}$

But ... do we care about this assumption?

- We make lots of assumptions in epidemic modeling
 - The goal is not to eliminate them
- But we should understand the impact they have
 - On the transmission system
 - And the epidemic outcomes of interest
- What are the impacts of partnership persistence?

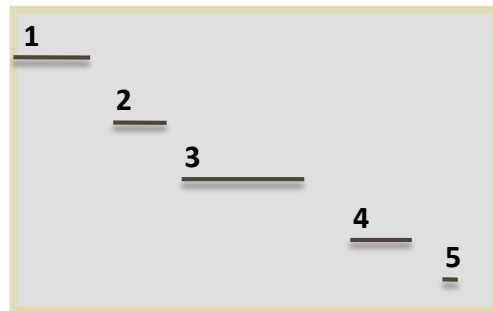
Persistent partnerships

- Change the connectivity of the transmission network
- In two ways
 - Cross-sectionally (at a moment in time)
 - Longitudinally

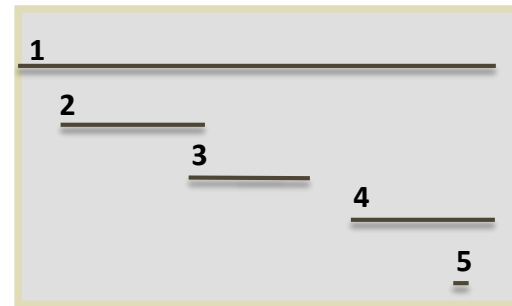
Partnership timing and sequence

A sequence of partnerships can be either :

Serially monogamous



or Concurrent

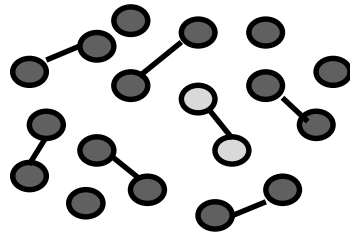


time →

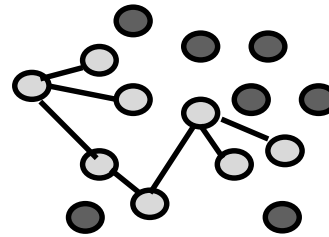
Same number of partnerships,
but the sequence of start and end dates is different

Concurrency and cross-sectional connectivity

A unique *cross-sectional* network signature:



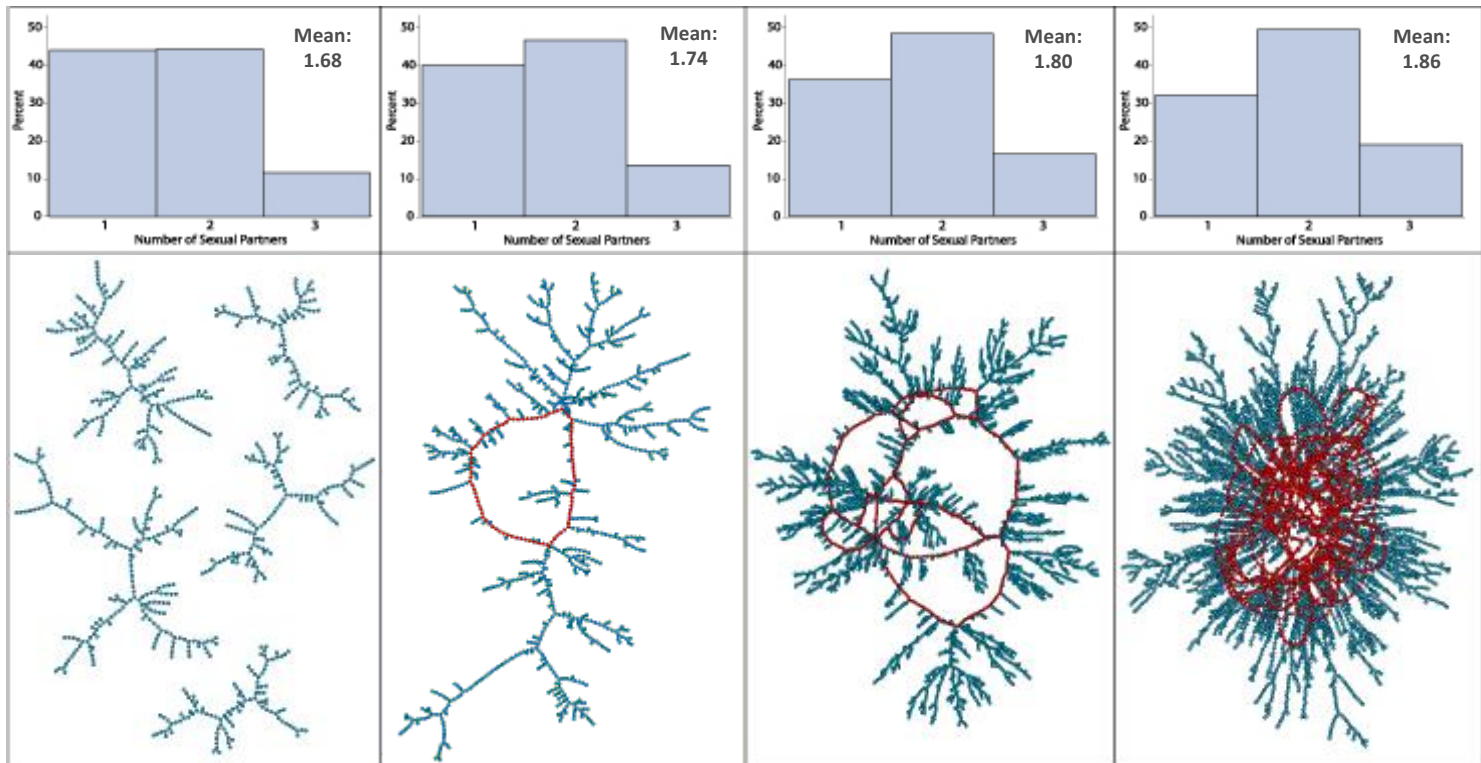
monogamy



concurrency

Concurrency and cross-sectional connectivity

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%

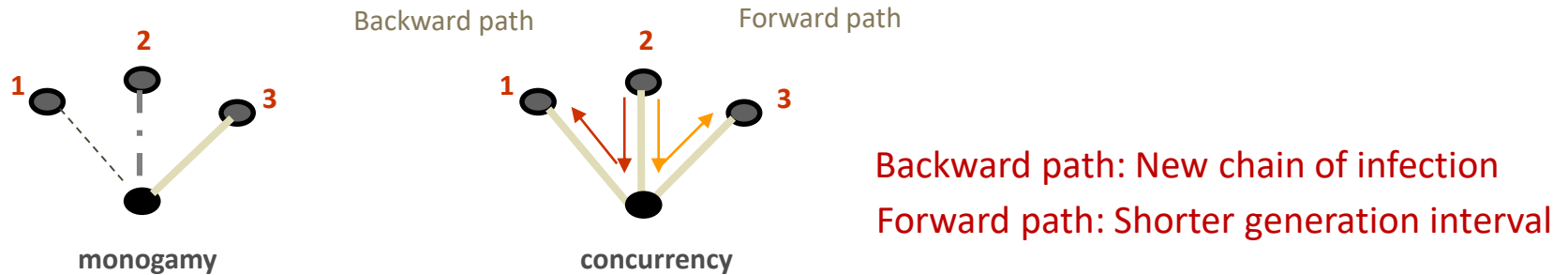
5%

15%

Ref: Morris, Goodreau and Moody 2007

Concurrency and longitudinal connectivity

Concurrency removes the protection of sequence over time



This changes two features of epidemic dynamics:

The reachable path of infection

Epidemic velocity

The reachable path

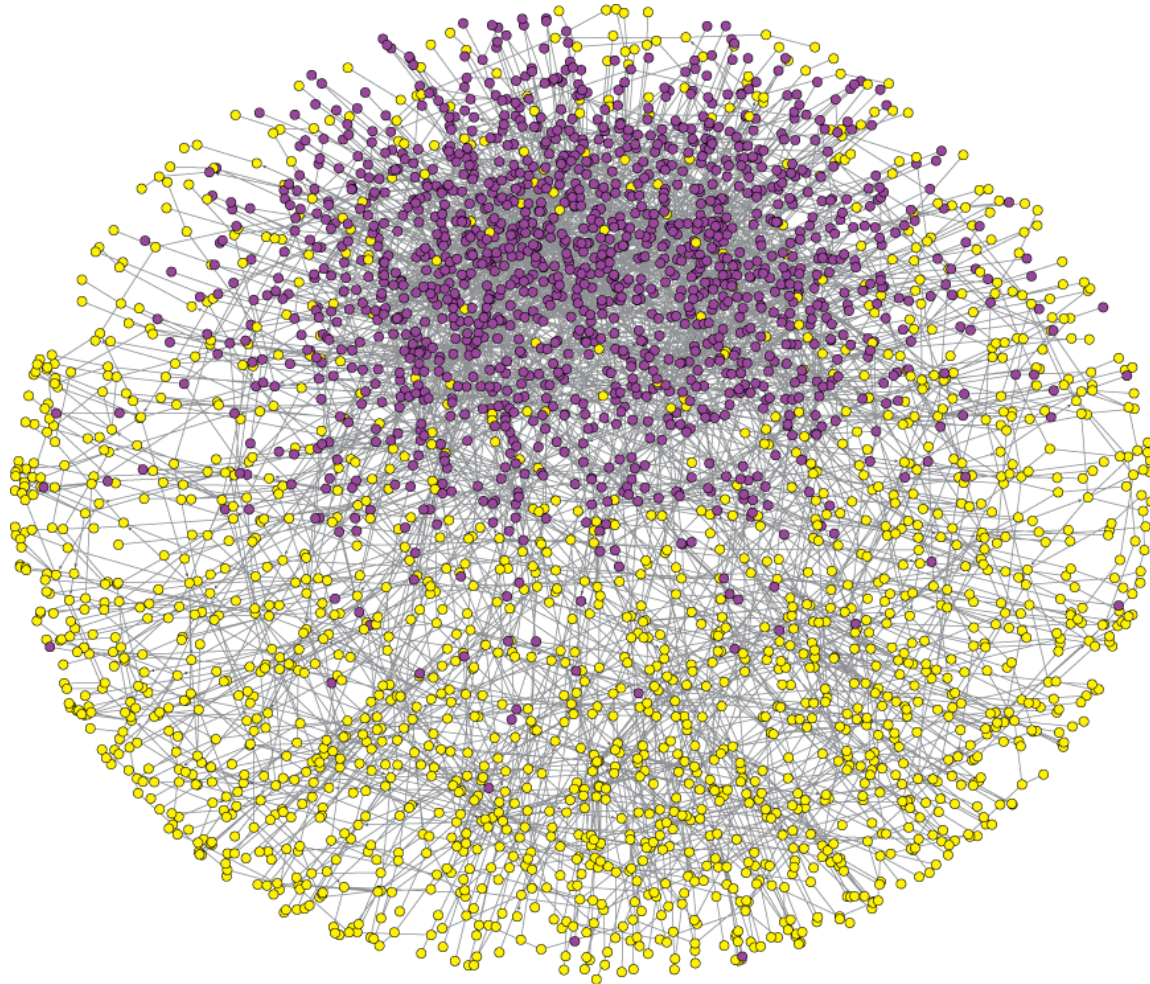
- The *reachable path* in a transmission system is determined
 - Not by the partnerships at any single point in time
 - Nor by the cumulative total over time
- By the *cumulative time ordered sequence of partnerships*

Cumulative connectivity: a network collapsed over time

10,000 nodes

2 demographic
groups

10 years
of partnership
activity



99.7%

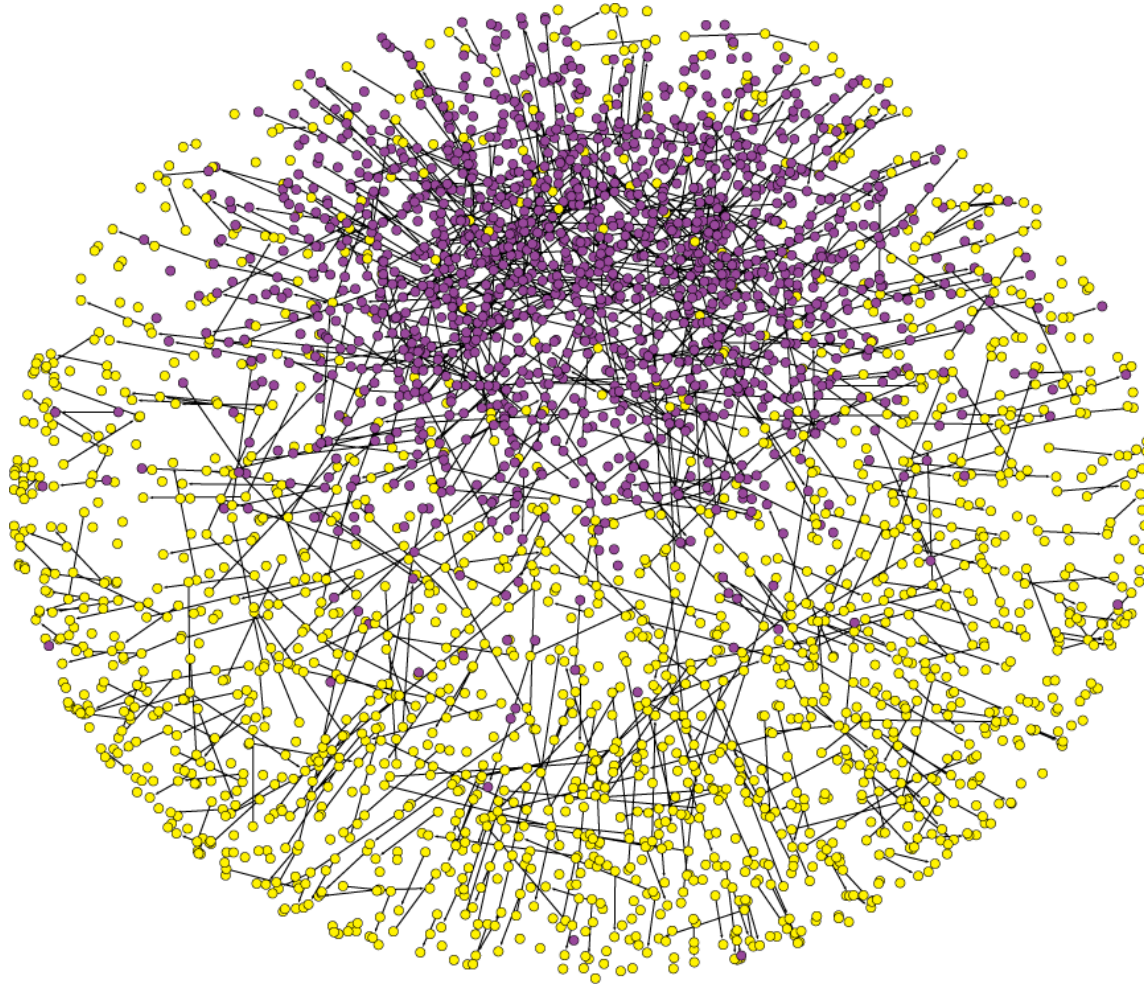
Connected

(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

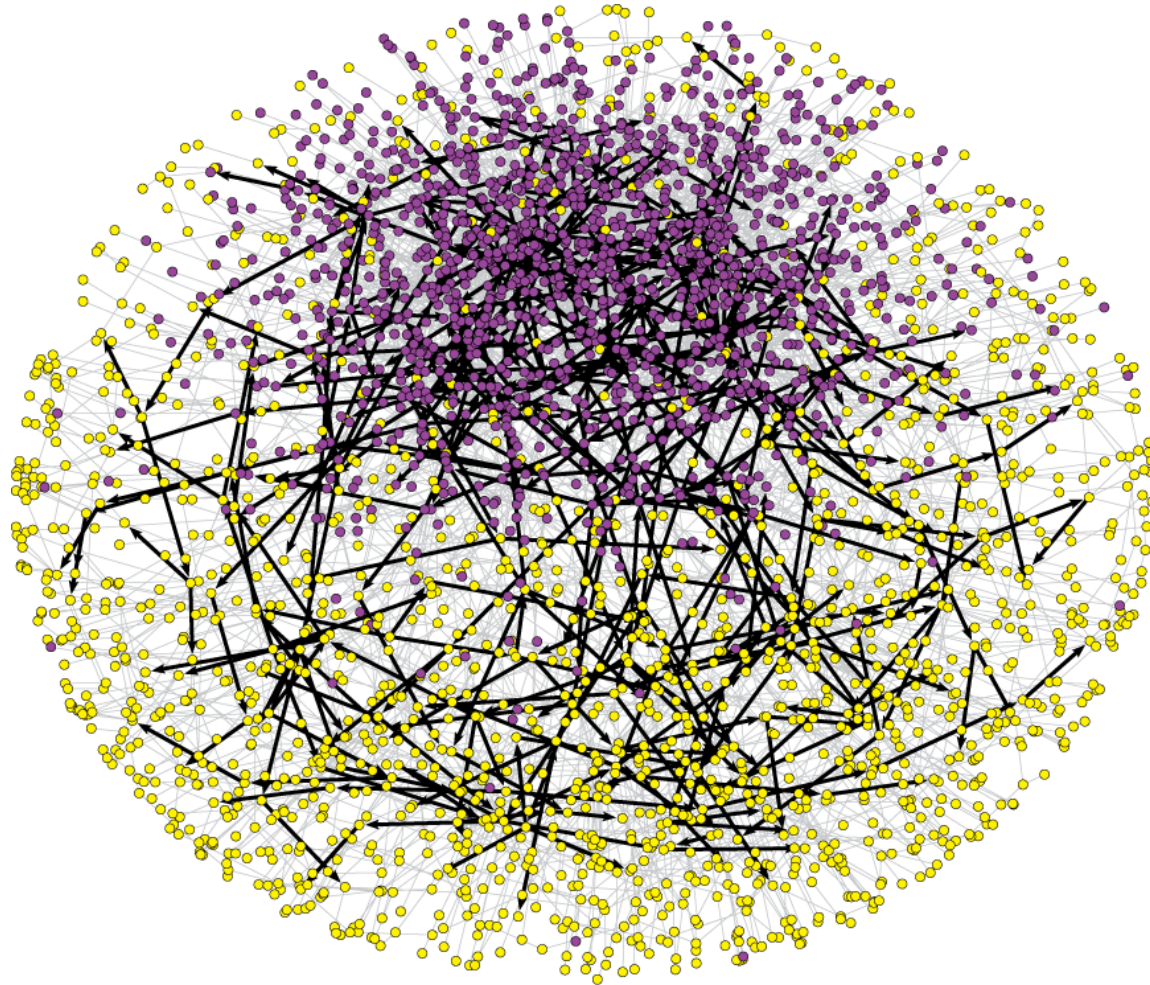
The reachable path: different than both of these

From 10 initial seeds

Trace out:

The cumulative time-ordered path of partnerships

Over 10 years

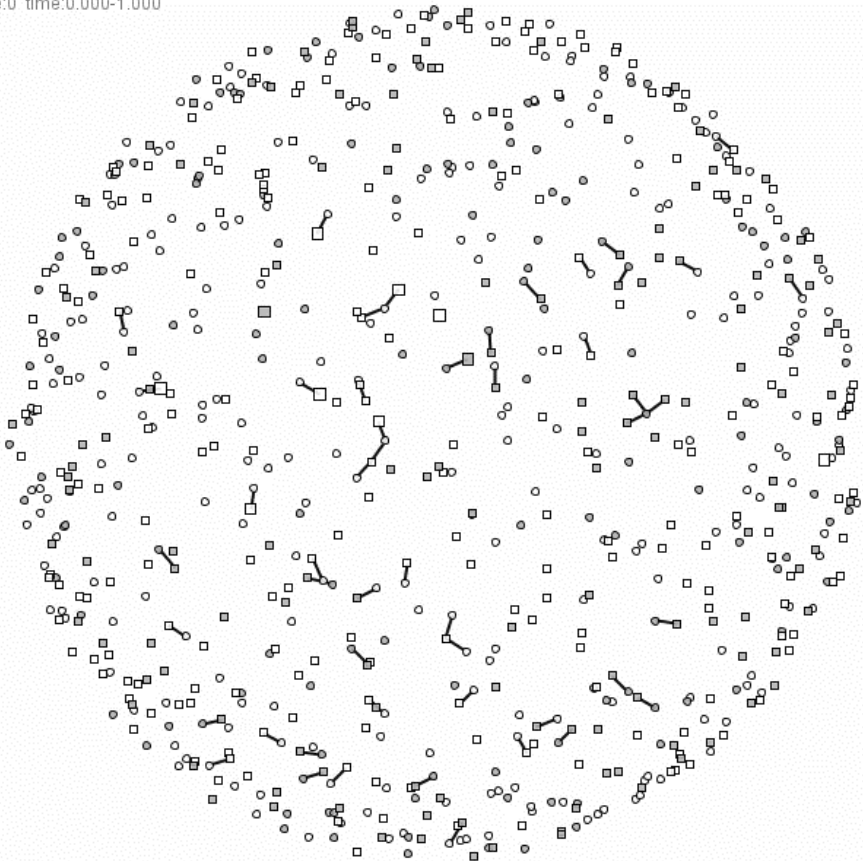


5.0%

Reachable

We track the growth of the forward reachable path (FRP)

slice:0 time:0.000-1.000



- Original network: 10,000 nodes
- Focus here on the ~600 that end up in the FRP
- 10 initial seeds (slightly larger squares)
- Tie color indicates:
 - BLUE: join FRP when dyad is mutually monogamous
 - RED: join FRP when at least one node has concurrent partner
 - BLACK: active tie, but dyad not in FRP yet

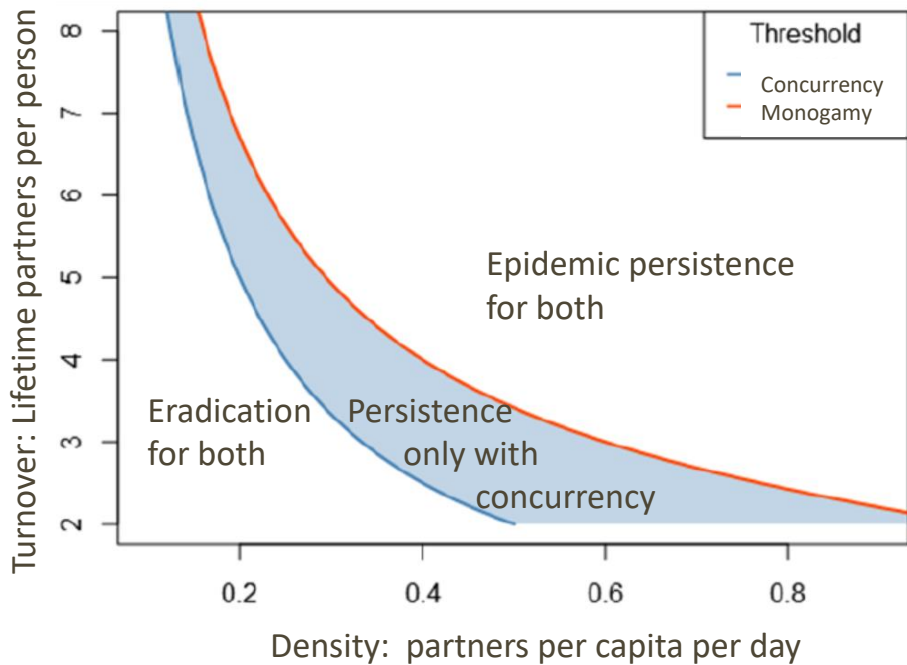
Only **5%** of ties are ever concurrent

But these account for **> 50%** of the FRP

Movie online at:

<https://www.youtube.com/watch?v=r3LYA5kirjA&t=214s>

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)

So yes, persistent partnerships matter

- For the connectivity of a transmission network
 - Cross-sectionally (with clear thresholds)
 - Longitudinally – via the reachable path and velocity
- And this is why we need network models
- Definition of a network model:
 - The network is on the LHS of an equation:
 - $\text{Network} = f(\text{some covariates})$
 - Statistical models for network structure and dynamics
 - Can be estimated from data
 - And used to simulate dynamic networks

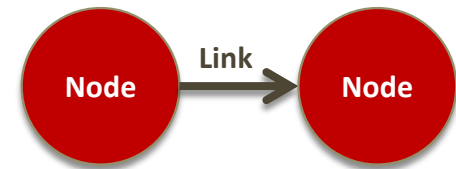
19

Network concepts and terminology

Through statnetWeb

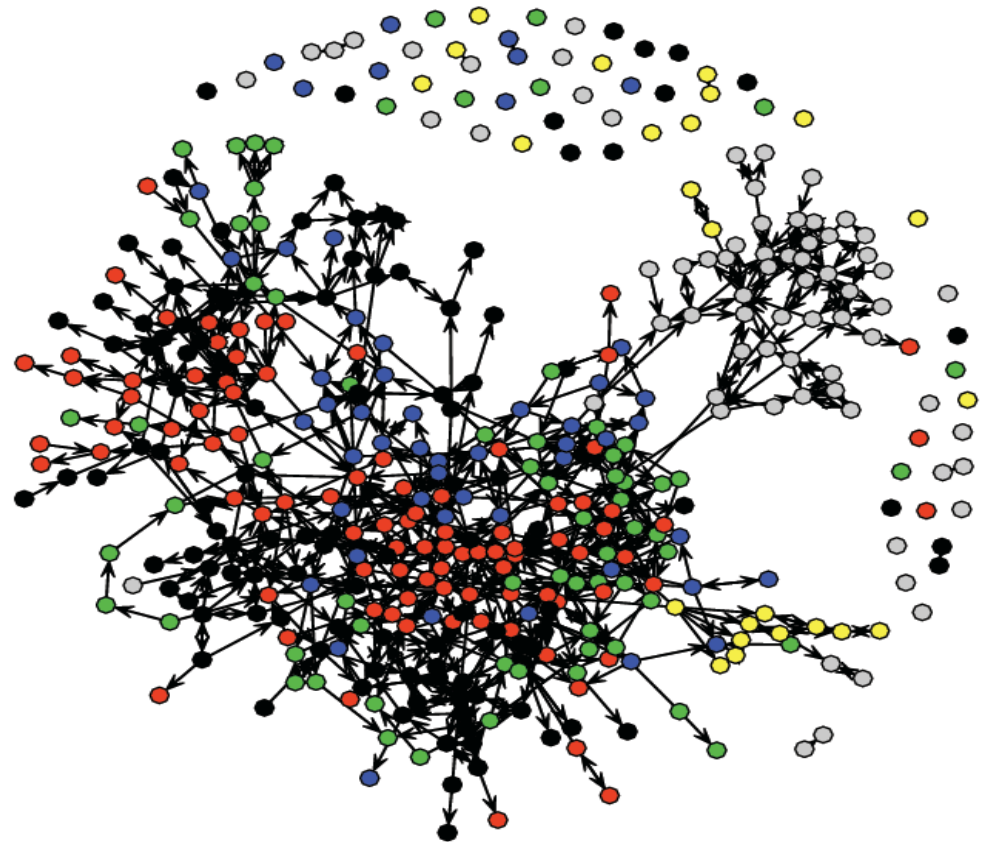
Terminology

- **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



Nodes are now classified as senders and/or receivers

- A directed graph is also called a di-graph

- A directed edge is also called an arc

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



- Binary (0,1 on or off only)

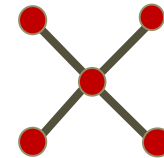
- Signed and/or Valued (... -2, -1, 0, 1, 2 ...)

Configurations

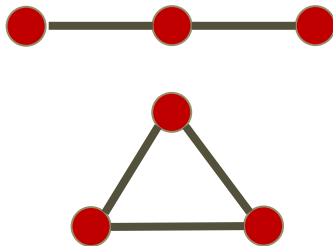
Dyads



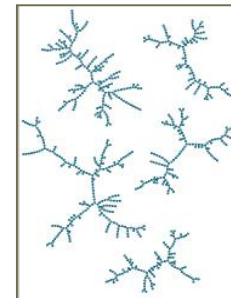
Stars



Triads



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

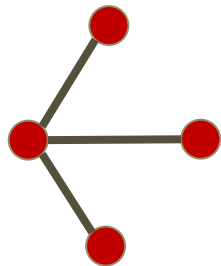
- Node level: *attributes of individual nodes*
 - Examples: age, sex, infection state, degree
- Dyad level: *attributes of pairs or edges*
 - Examples: type of relationship, duration
- Component level: *subgraph attributes and distributions*
 - Examples: size, density, degree and geodesic distributions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

E.g. Cycles

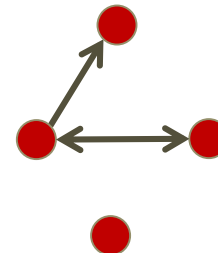
- Paths that lead back to the origin node
 - Cycle length k = number of lines in the cycle, “ k -cycles”
 - Triangles are 3-cycles
- Node level measure: Number of cycles a node is a member of
- Edge level measure: Number of cycles an edge is a member of
- Network level measure: The “cycle census”, the frequency of these configurations in 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: 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 (Affiliation network); heterosexual sex net
- Multiplex
 - More than one type of link possible

Example: Sexual partnerships and needle sharing
- Multilevel
 - Hierarchical/nested or Overlapping

Example: Grades within schools; persons working across multiple projects

Representing network data

■ Sociomatrix

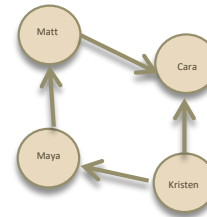
- (aka adjacency matrix)
- simple but inefficient for large sparse nets (order n^2)

	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



31

Explore with statnetWeb

Just a bit

Quick peek at statnetWeb

Core statnet packages

(network, sna, ergm, networkDynamic, tergm, tsna)

For descriptive and statistical network analysis, and simulation

statnetWeb

User-friendly GUI to access main
statnet functionality

Days 1-2

EpiModel

Package to conduct network-based
epidemic modeling

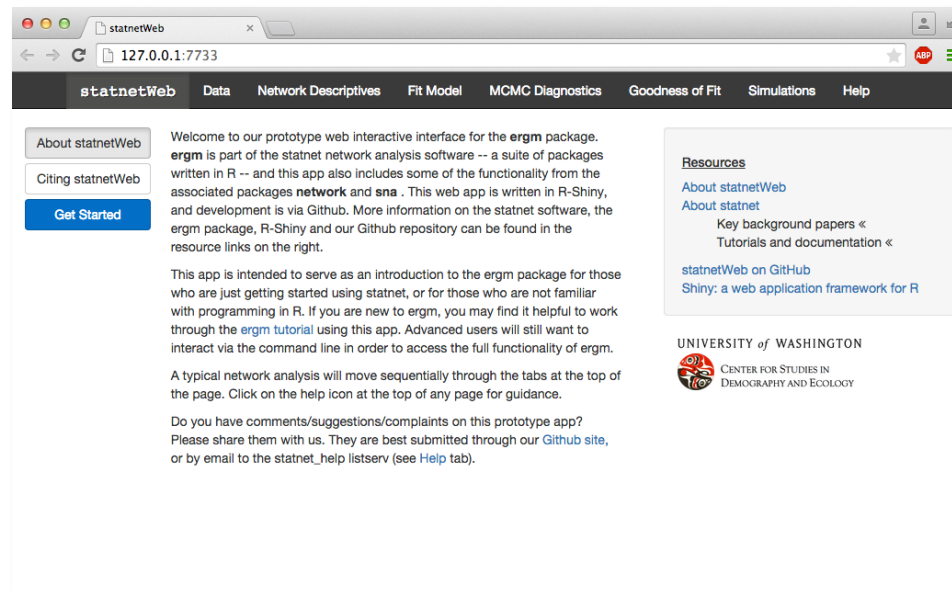
Both web GUI and command-line versions

Days 1-5

Intro to statnetWeb

statnetWeb is a graphical user interface for network analysis

- Runs in a web browser, like epiweb
- Wiki: <https://github.com/statnet/statnetWeb/wiki>



Using statnetWeb on your computer

- Install the statnetWeb package
 - `install.packages("statnetWeb")`
- Load the package and launch the application
 - `library(statnetWeb)`
 - `run_sw()`

Note: Closing the browser window ends your session, so if you want to save something, do it before you quit

Network data in statnetWeb

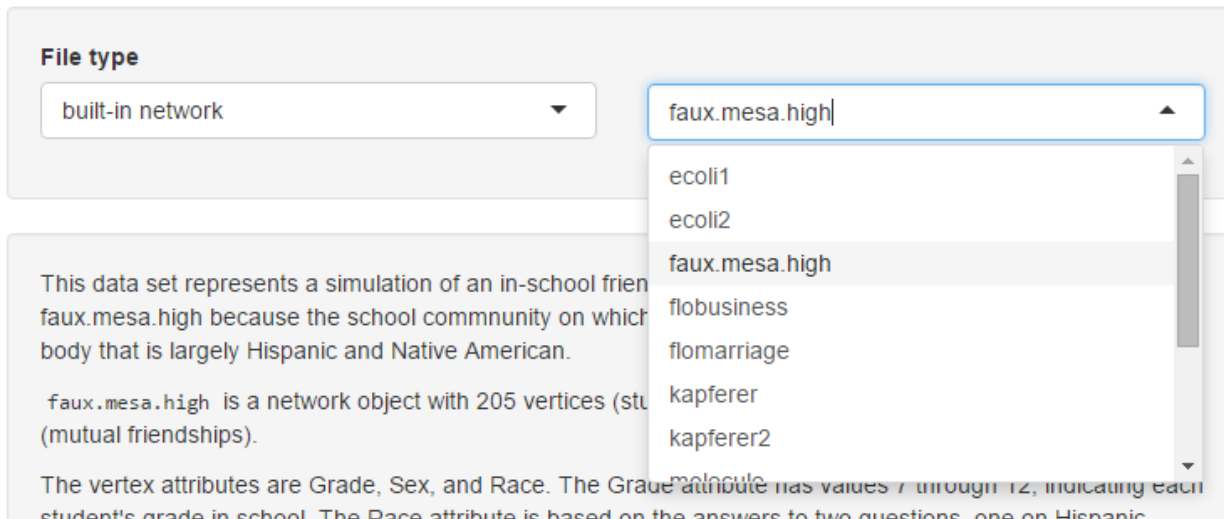
- On the “Data” page, you can upload networks from multiple sources

- Internal: built-in networks
- External: R, Excel, Pajek files

The screenshot shows the 'Data' page interface for uploading network data. It features a 'File type' dropdown menu set to 'matrix of relational data (*.csv or *.rds)', a 'Browse...' button, and a status message 'No file selected.'. Below this is a box for 'name: NA' and 'size: NA'. The 'Matrix Type' section includes radio buttons for 'Adjacency matrix' (selected), 'Bipartite adjacency matrix', 'Incidence matrix', and 'Edge list'. The 'Network Attributes' section includes checkboxes for 'directed?' (checked), 'loops?', 'multiple?', and 'bipartite?'.

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



The screenshot shows the statnetWeb interface. On the left, a 'File type' dropdown menu is set to 'built-in network'. On the right, a search input field contains 'faux.mesa.high', and a dropdown menu is open, listing several network datasets. The 'faux.mesa.high' dataset is highlighted in the list. Below the search field, there is a text box providing information about the 'faux.mesa.high' dataset, including its description and attributes.

File type
built-in network

faux.mesa.high

ecoli1
ecoli2
faux.mesa.high
flobusiness
flomarriage
kapferer
kapferer2
molecule

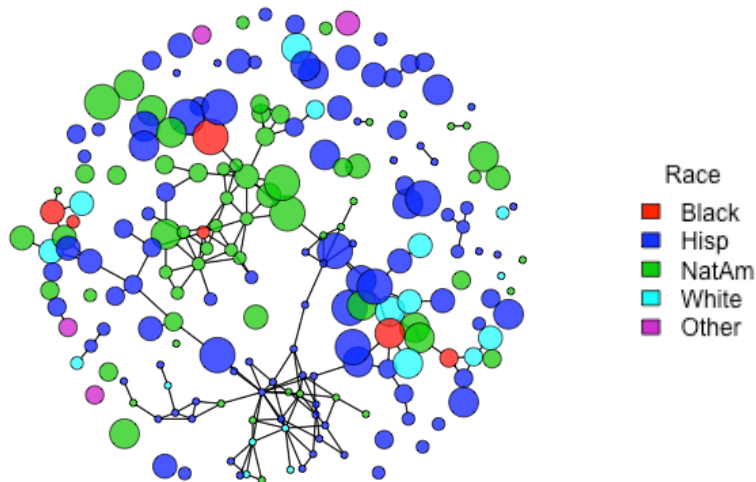
This data set represents a simulation of an in-school friendship network for faux.mesa.high because the school community on which it is based is largely Hispanic and Native American.

faux.mesa.high is a network object with 205 vertices (students) and 1000 edges (mutual friendships).

The vertex attributes are Grade, Sex, and Race. The Grade attribute has values 7 through 12, indicating each student's grade in school. The Race attribute is based on the answers to two questions, one on Hispanic

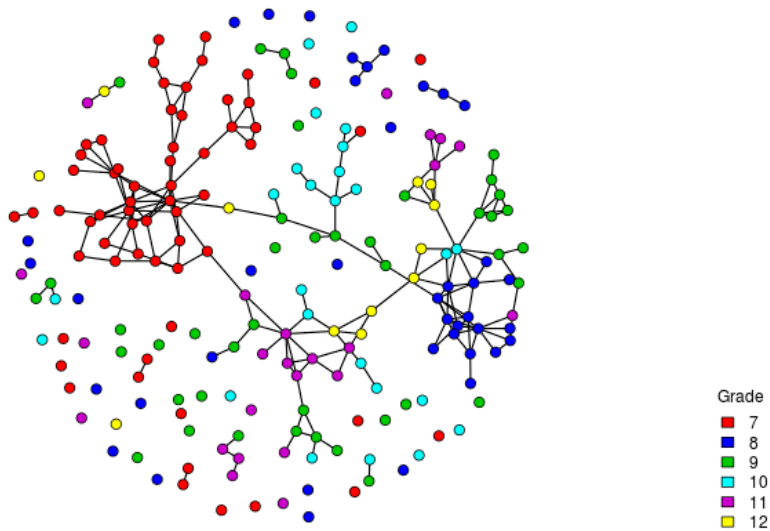
Attributes

- 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.

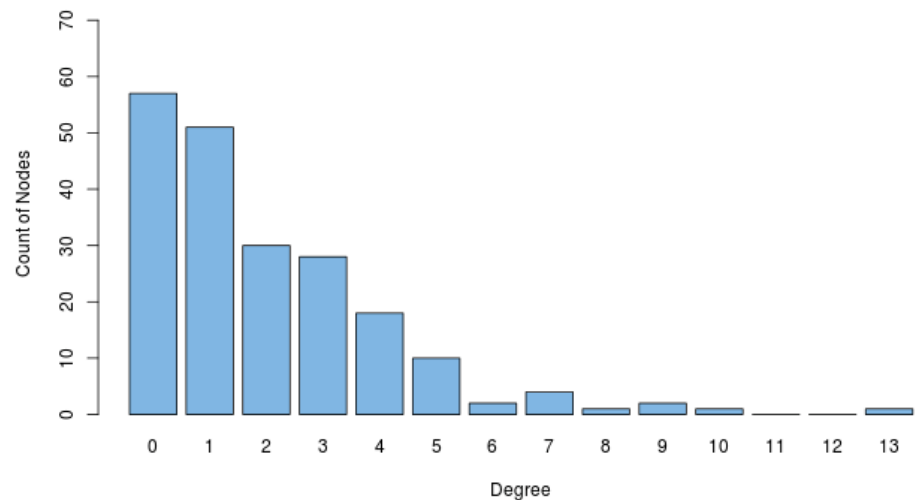
	7	8	9	10	11	12
7	75	0	0	1	1	1
8	0	33	2	4	2	1
9	0	2	23	7	6	4
10	1	4	7	9	1	5
11	1	2	6	1	17	5
12	1	1	4	5	5	6

Degree metrics

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

Degree distribution

- To view it in statnetWeb:

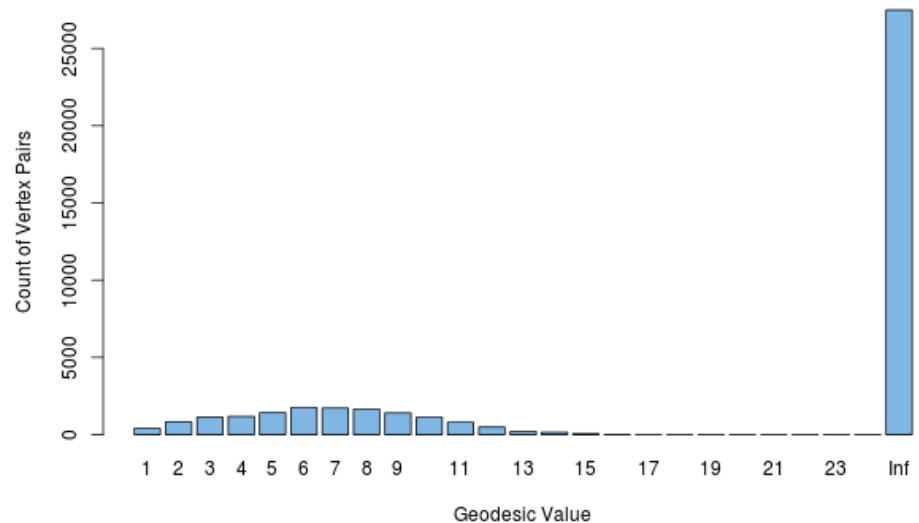


Connectivity measures: Geodesic

- 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

- To view it in statnetWeb:

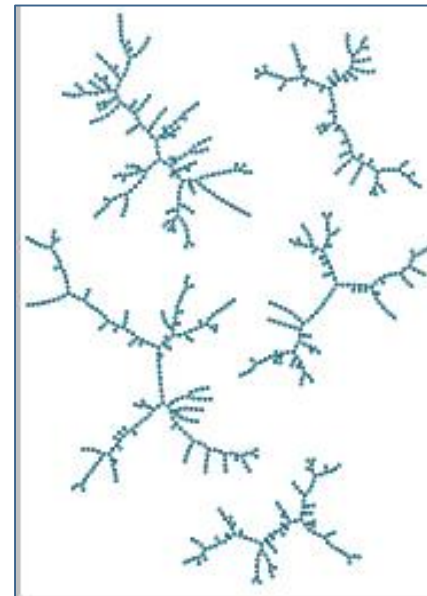


The last bar represents the node pairs with infinite geodesic distance

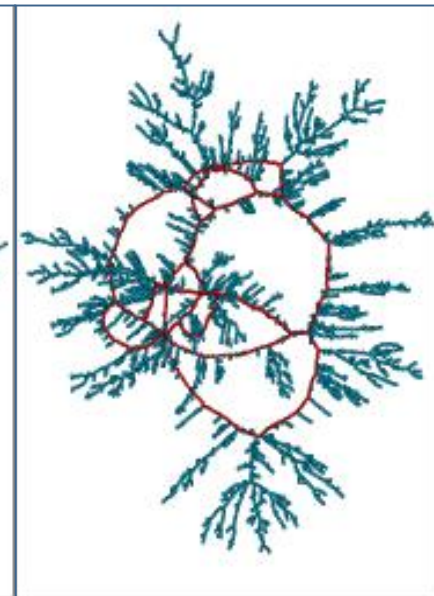
Connectivity measures: Components

- 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...

Where do we go from here?

- *Take a break now, for 15 min*
- *When we come back: Network data*
 - What is it
 - How do we collect it?
 - What would network data look like for your research topic?
- *Tomorrow*
 - Statistical Network Modeling

Other Resources

Concurrency simulation and intuition

- [Concurrency simulator](#) – an online Rshiny app
- [Concurrency tutorial](#) with exercises, and links to research papers