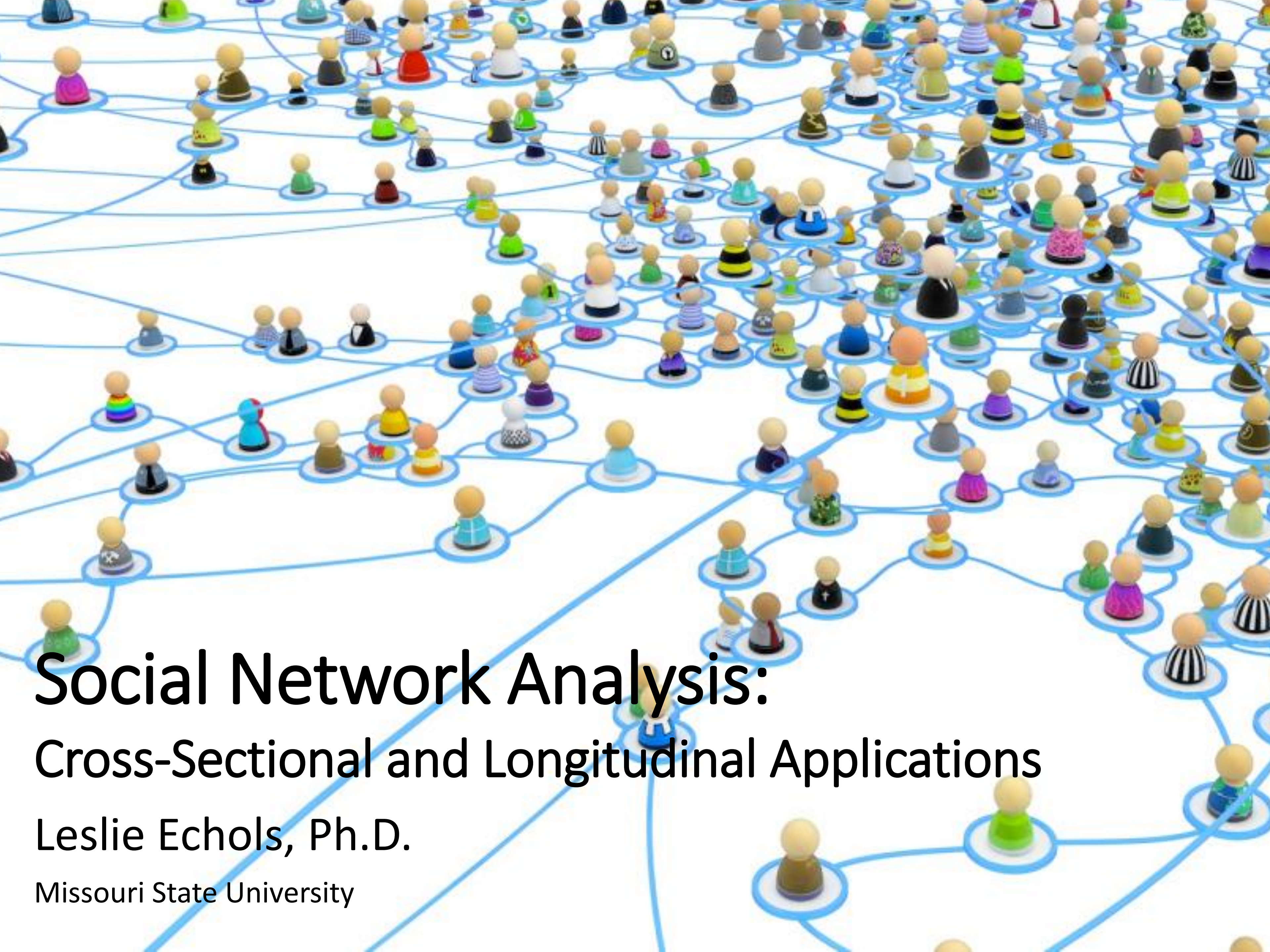


Social Network Analysis:

Cross-Sectional and Longitudinal Applications

Leslie Echols, Ph.D.

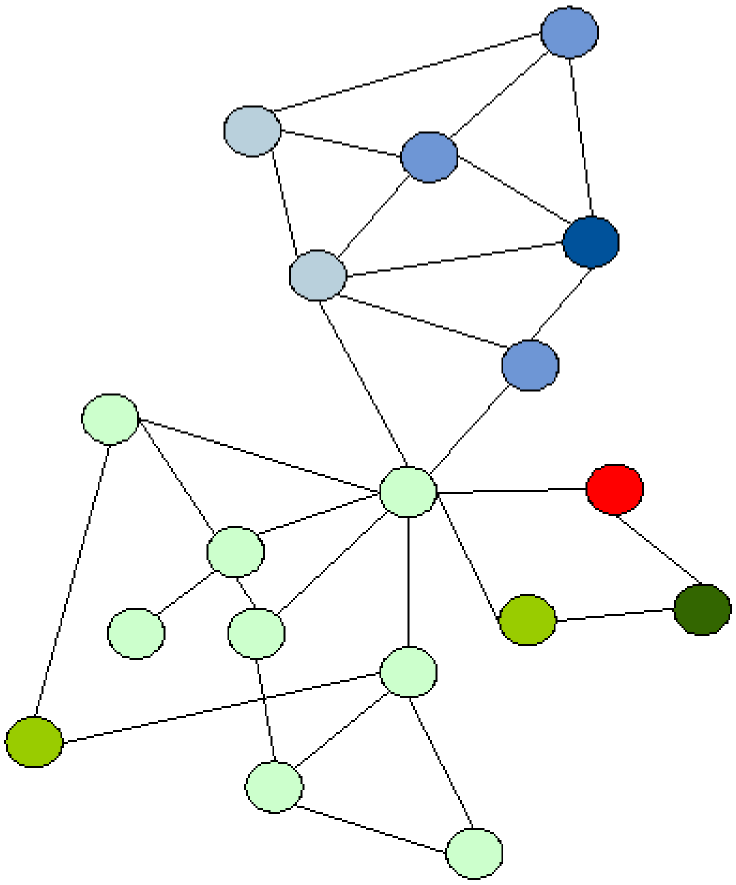
Missouri State University

What we'll cover…hopefully!

- Basics of social networks
  - Network/data structures
    - Compared to data structures for linear models
  - Questions that can be answered
- Examples of network studies
- Types of network analyses
  - Incorporating network statistics into traditional linear models
  - Modeling networks
- Network visualization

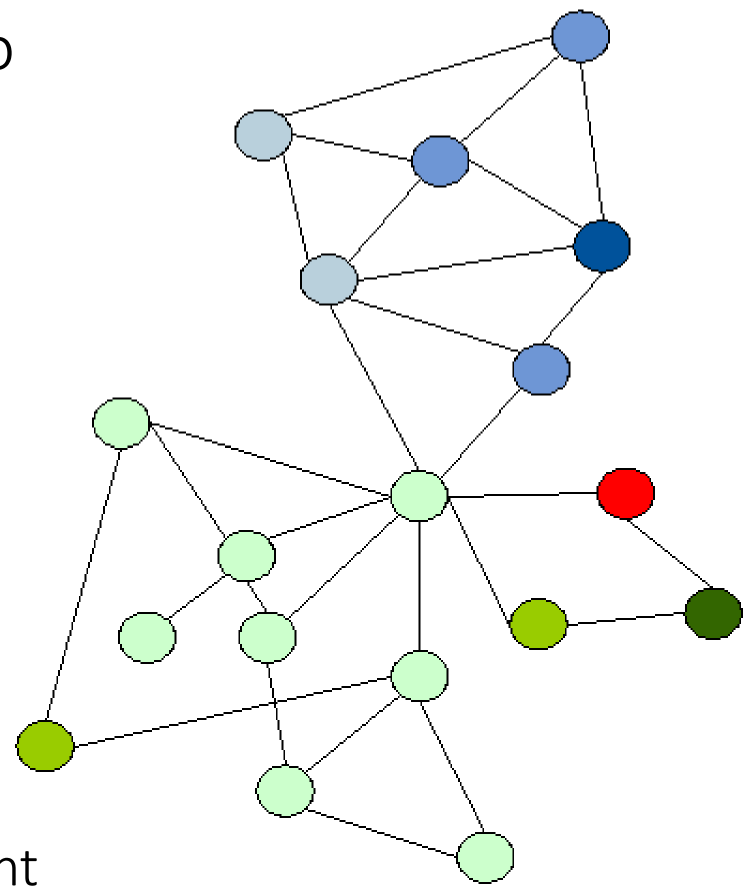
What are social networks?

- Interconnected ties or relationships among a set of individuals within some social boundary
  - Classroom or school
  - Nursing home or assisted living community
  - Prison
  - Community center
  - Daycare center
  - Business or organization
  - Reservation
  - Ward or auxiliary within a ward
  - Mission
  - Other bounded networks relevant to your research?



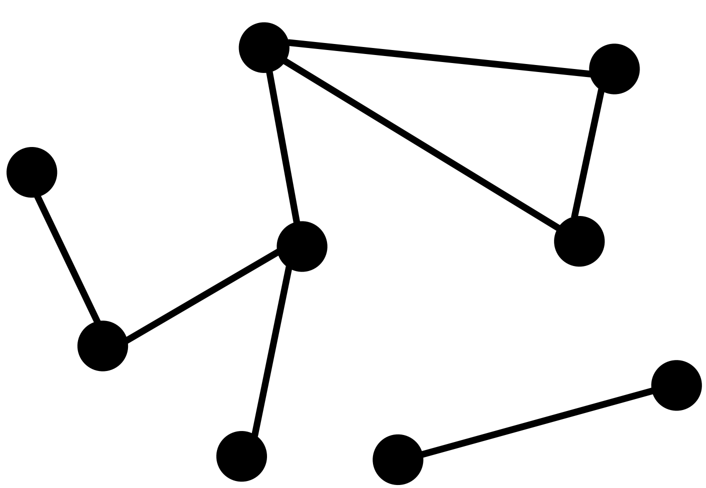
What are social networks?

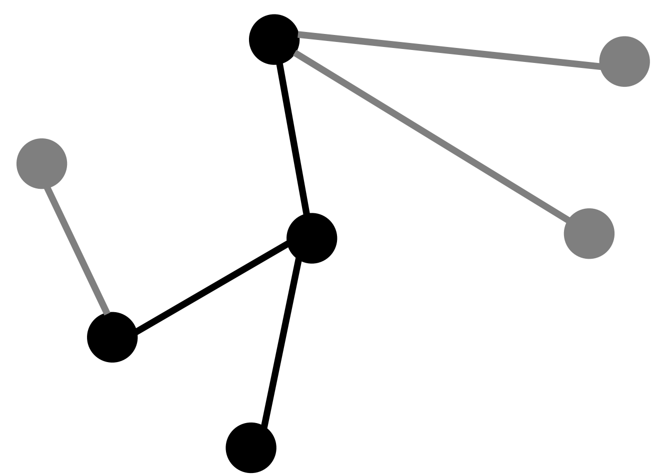
- A set of relations that apply to a set of actors
- Characterized by the type of relation
  - Friendship
  - Antipathy
  - Advice-seeking
  - Phone/e-mail communication
  - Gossip
  - Needle sharing
  - Choice of physician, financial advisor, beautician, etc.
  - Attendance at a meeting or event
  - Other types of relations relevant to your research?



Types of networks

- Complete networks
  - The set of ties among all pairs of nodes in a given (i.e., defined or bounded) network (e.g., friendships among all students in a classroom)
  - Every dyadic pair in the network has a value assigned to it
- Ego networks
  - Emphasizes single focal nodes and their ties to others (e.g., one student's friends)
  - Each ego potentially has a distinct set of alters





Why do social network research?

- When studying a social system
  - Not just the actors (individuals) of the system and their individual properties (composition of the system)
  - But how they relate to each other (structure of the system)
- When the variables of interest are the interdependencies between different actors (individuals) in the system, for example:

Traditional Approach	Network Approach
On average, how many friends do individuals have?	How likely is friendship in the network?
What are the predictors of friendship (e.g., what individual characteristics are associated with having friends)?	What network and dyadic characteristics influence friendship formation (e.g., do homophily, reciprocity, and transitivity constrain friendship choices)?
How are outcomes different for individuals with friends compared to individuals without friends?	How do friendships and individual characteristics co-evolve over time?

Traditional analytic approaches

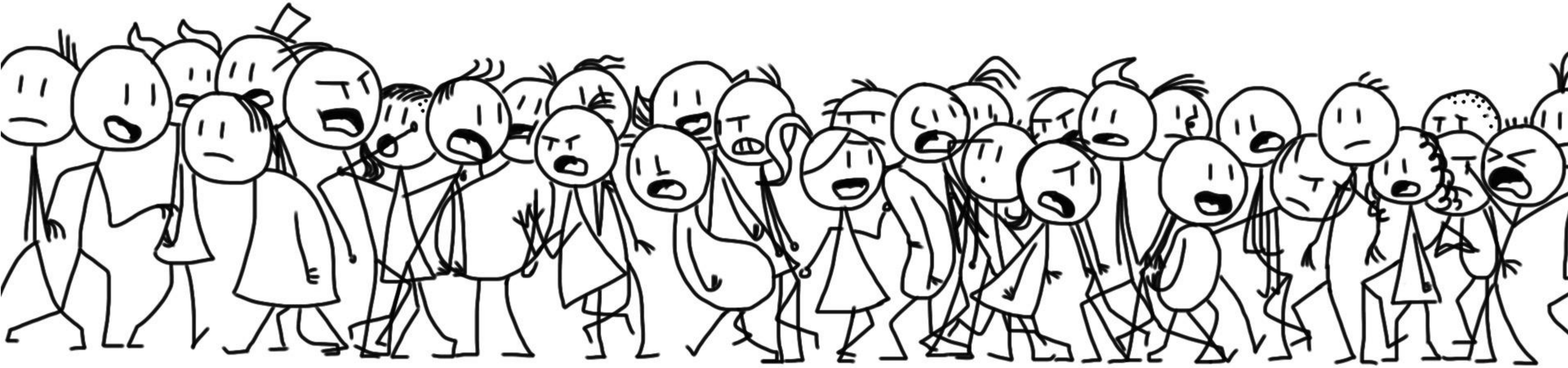
- Rely on a key (and required) underlying assumption that people in a study are unrelated or independent of one another
- Isolating individuals from their friends, families, jobs, and neighborhoods removes them from the source of the very behavior or characteristic we seek to understand

Traditional vs. network data analysis

- The major difference between conventional and network data is that conventional data focuses on actors and *attributes*, while network data focuses on actors and *relations*
  - How much of my neurotic behavior should be attributed to me?

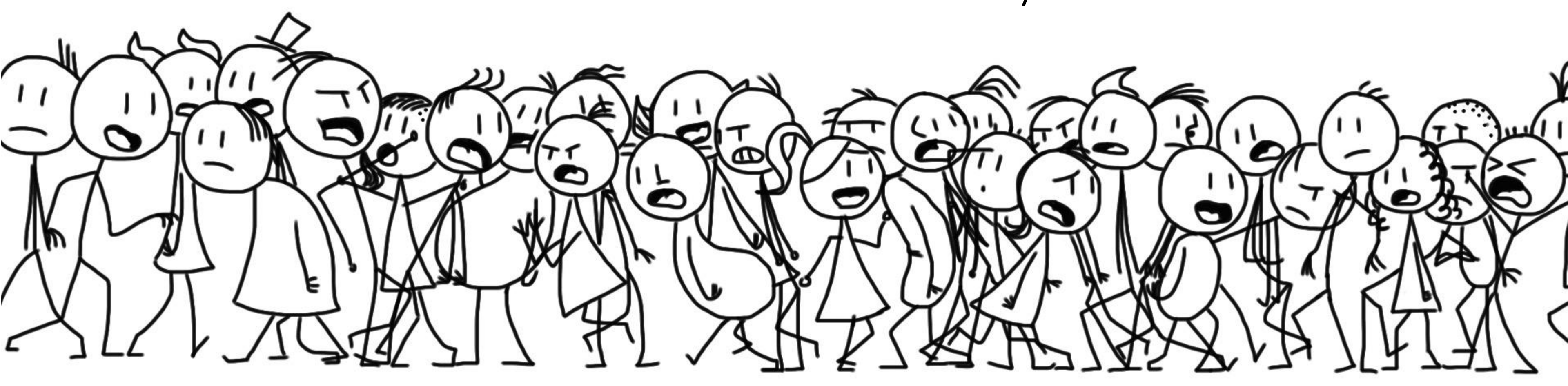
Traditional vs. network data analysis

- The major difference between conventional and network data is that conventional data focuses on actors and *attributes*, while network data focuses on actors and *relations*
  - How much of my neurotic behavior should be attributed to me?
  - And how much should be attributed to my social context?



Traditional vs. network data analysis

- Social network analysis can be used to understand if/how certain attributes influence certain relations
  - Are my friends neurotic because I want to hang out with people like me?
- Longitudinal social network analysis can be used to understand how certain relations influence certain attributes over time
  - Am I neurotic because I became like my friends?



Traditional vs. network data structures

- Traditional (rectangular) data

ID	Self-worth	Depression	Loneliness	Anxiety
1	2	3	2	3
2	1	3	4	3
3	3	1	3	2
4	2	3	3	2
5	2	3	3	4
6	4	2	3	2
7	1	4	4	4
8	1	3	4	3
9	2	3	4	3
10	4	2	2	2

Traditional vs. network data structures

- Traditional (rectangular) data

ID	Friend 1	Friend 2	Friend 3	Friend 4
1	2	3	9	7
2	1	3	6	5
3	5	7	4	1
4	5	6	8	9
5	2	6	9	10
6	10	3	5	8
7	3	2	10	9
8	9	3	4	7
9	5	8	2	1
10	9	3	5	6

Traditional vs. network data structures

- Traditional (rectangular) data

ID	Friend 1 Attribute	Friend 2 Attribute	Friend 3 Attribute	Friend 4 Attribute
1	2	3	9	7
2	1	3	6	5
3	5	7	4	1
4	5	6	8	9
5	2	6	9	10
6	10	3	5	8
7	3	2	10	9
8	9	3	4	7
9	5	8	2	1
10	9	3	5	6

Traditional vs. network data structures

- Network (matrix) data—binary

ID	1	2	3	4	5	6	7	8	9	10
1	--	1	1	0	0	0	1	0	1	0
2	1	--	1	0	1	1	0	0	0	0
3	1	0	--	1	1	0	1	0	0	0
4	0	0	0	--	1	1	0	1	1	0
5	0	1	0	0	--	1	0	0	1	1
6	0	0	1	0	1	--	0	1	0	1
7	0	1	1	0	0	0	--	0	1	1
8	0	0	1	1	0	0	1	--	1	0
9	1	1	0	0	1	0	0	1	--	0
10	0	0	1	0	1	1	0	0	1	--

Traditional vs. network data structures

- Network (matrix) data—weighted

ID	1	2	3	4	5	6	7	8	9	10
1	--	1	3	2	0	1	4	0	3	0
2	1	--	1	5	2	1	0	2	0	4
3	1	3	--	1	3	0	5	0	1	0
4	4	2	0	--	1	1	0	4	1	2
5	5	1	0	3	--	2	0	3	1	1
6	0	0	1	0	1	--	0	1	5	1
7	2	1	4	5	0	2	--	2	1	2
8	0	4	1	1	0	0	1	--	4	0
9	1	4	0	0	1	0	4	1	--	3
10	1	0	2	3	1	5	4	0	1	--

Types of matrices

- Binary
  - Contains only 1s and 0s in its cells
- Valued (weighted)
  - Convey the intensity of a tie by the values found within its cells
- One-mode
  - All actors are tied to one another according to one relation (e.g., friendship)
- Two-mode
  - Actors are tied to (or affiliated with) particular events (e.g., attendance at various social events)
  - Referred to as incidence, affiliation, or bipartite matrices

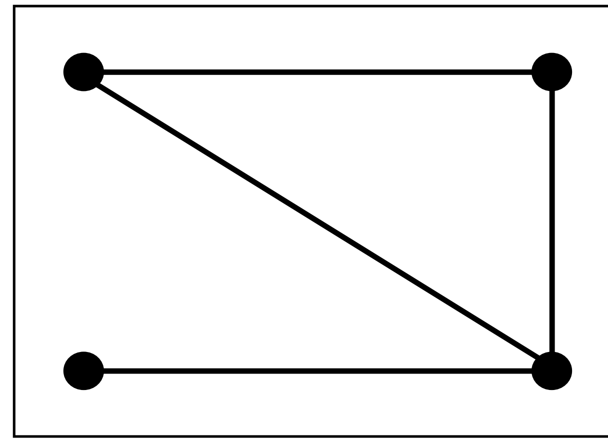
A note about notation

- In social network analysis, we use notation to designate which cell we are working with
  - The individual rows in a matrix are referred to as *i* (the sender)
  - The individual columns are referred to as *j* (the receiver)
  - $X_{i,j}$ refers to the value of the cell at row *i* column *j* (the tie, or an attribute of the tie, between *i* and *j*

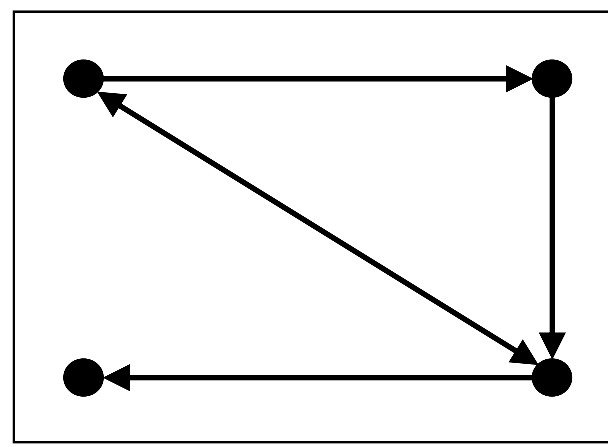
ID	1	2	3	4	5
1	--	1	1	0	0
2	1	--	1	0	1
3	1	0	--	1	1
4	0	0	0	--	1
5	0	1	0	0	--

Symmetric vs. asymmetric matrices

- A square matrix with as many rows and columns as there are actors in the data set
- Represents who is next to, or adjacent to whom, in the social space
- In a symmetric matrix both entries for the same set of partners will be the same ($X_{i,j} = X_{j,i}$)
- In an asymmetric matrix, the sender of a tie is the row and the target of the tie is a column, and the entries for the same set of partners may differ depending on their role as senders or receivers



Undirected (Symmetric) Graph



Directed (Asymmetric) Graph

The adjacency matrix

- Starting point for almost all network analyses
- Ignores the main diagonal (no self-nominations are permitted)
- Examining row and column vectors can give us information about social roles and behavioral tendencies in the network
  - Row vectors tell us how many nominations were made by each actor (outdegree)
  - Column vectors tell us how many nominations were given to (received by) each actor (indegree)

The adjacency matrix

- Who is most/least popular in this network?

ID	1	2	3	4	5	6	7	8	9	10
1	--	1	1	0	0	0	1	0	1	0
2	1	--	1	0	1	1	0	0	0	0
3	1	0	--	1	1	0	1	0	0	0
4	0	0	0	--	1	1	0	1	1	0
5	0	1	0	0	--	1	0	0	1	1
6	0	0	1	0	1	--	0	1	0	1
7	0	1	1	0	0	0	--	0	1	1
8	0	0	1	1	0	0	1	--	1	0
9	1	1	0	0	1	0	0	1	--	0
10	0	0	1	0	1	1	0	0	1	--

Questions that can be answered using SNA

- Cross-sectional
  - To whom are individuals connected?
  - What predicts a connection? Individual characteristics? Dyadic characteristics? Network characteristics?
  - Who has the most access to others in the network?
  - How well connected is a network overall?
- Longitudinal
  - How likely are connections to form, remain, or dissolve over time?
  - Do individual characteristics influence connections? Or do connections influence individual characteristics?

Examples of social network research

- Social integration and support among older adults
  - Abbott, Bettger, Hampton, & Kohler, 2012
  - Jang, Kim, Park, & Chiriboga, 2016
- Associations among prison inmates and implications for inmate safety and community reentry outcomes
  - Kreager et al., 2016
  - Schaefer, Bouchard, Young, & Kreager, 2017
- Role of social networks in religious life
  - Everton (2015)

Approaches to social network analysis

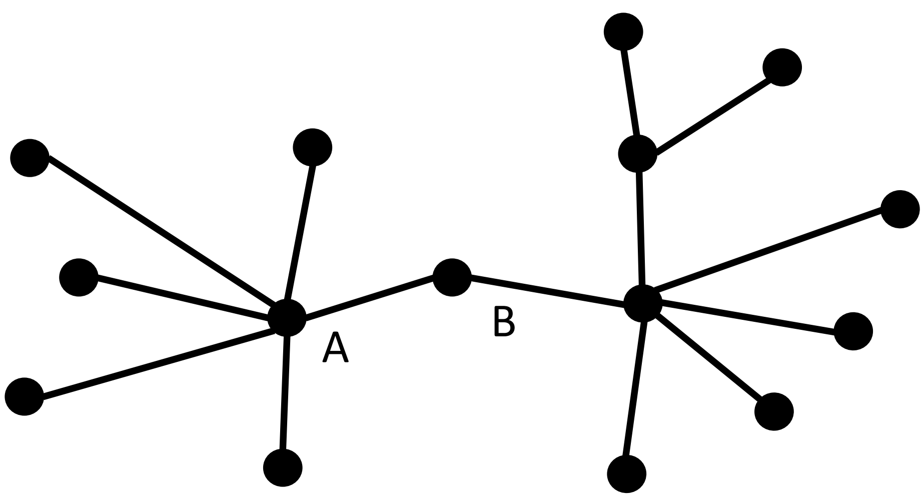
- Incorporating individual level network values as predictors in standard general linear models
  - Number of nominations given/received
  - Average distance to other individuals
  - Network position
- Network descriptives
  - Can be used as level 2 variables in multilevel models
- Using attributes of the network members along with characteristics of the whole network to predict binary ties or the strength of ties (ERGM)
- Using attributes of the network members and observed network ties across multiple time points to discern between processes of selection and influence (SIENA)

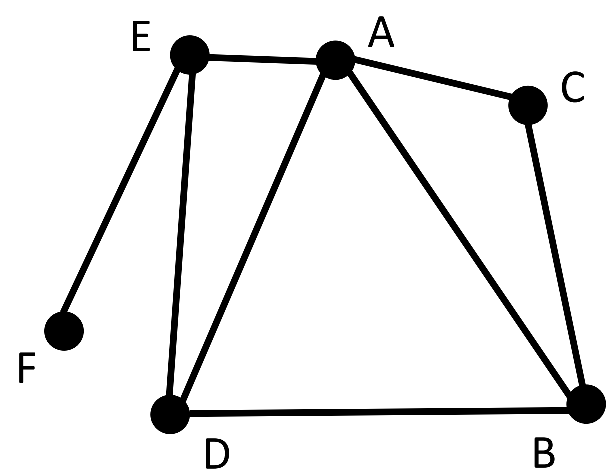
Level 1 variables

- Centrality
  - The number and length of pathways are important to understanding an individual's constraints and opportunities
  - Actors who have many pathways or short pathways to other actors may be more influential in the network
- Brokerage
  - Actors who mediate contact between two disconnected alters may play key roles in the network

Centrality measures

- People who occupy a central position in a network tend to be more visible, know many others, and be known by many others
- Centrality is characterized by degree, betweenness, and closeness
    - Different types of centrality signify different network roles and opportunities



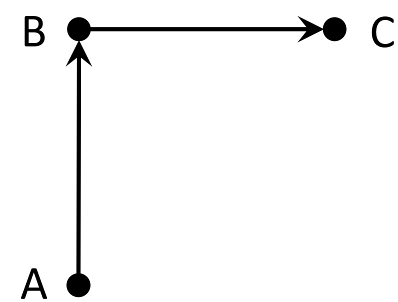


Brokerage analysis

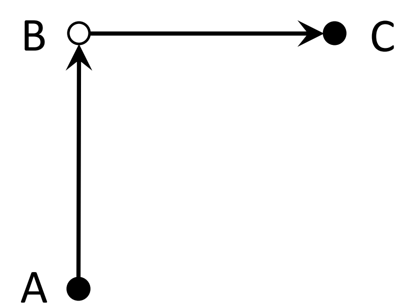
- Gould and Fernandez method
- Brokerage is the role played by a social actor who mediates contact between two alters
- Each vertex is assigned a brokerage score for each type of broker role
- Aggregate scores are also calculated for the entire network
- Requires two pieces of information
  - Adjacency matrix (graph object)
  - Attribute data (vector of class/group memberships)

Broker roles

- Within-group
  - *Coordinator* broker (the broker and the two alters belong to the same group)

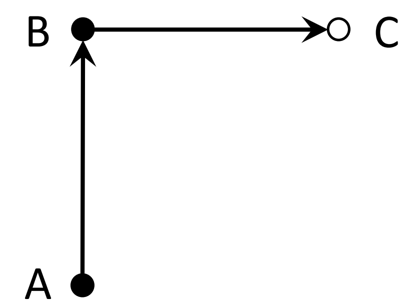


  - *Consultant* broker (the two alters belong to the same group but the broker belongs to another group)

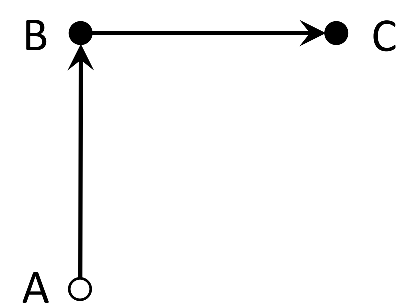


Broker roles

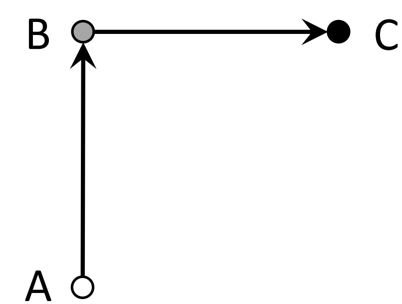
- Between-group
  - *Representative* broker (the broker and one alter belong to the same group and the other alter belongs to another group; the broker represents the interests of the first alter)



  - *Gatekeeper* broker (the broker and one alter belong to the same group and the other alter belongs to another group; the broker determines whether "outside" information is passed on)

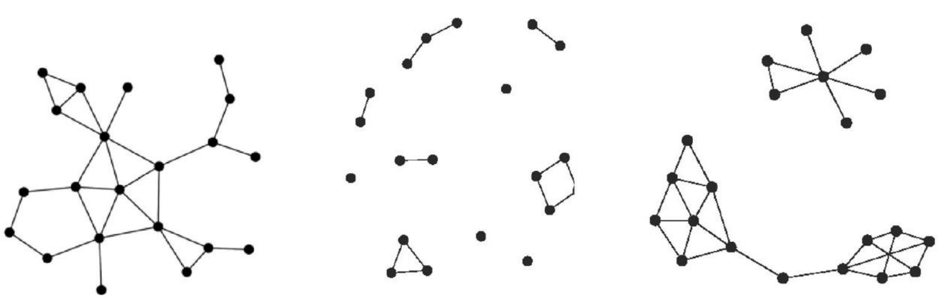


  - *Liaison* broker (the broker and the two alters all belong to different groups and the broker plays a neutral role)



Level 2 variables…boring?

- Aggregate measures of level 1 variables
  - Average number of incoming/outgoing ties
  - Average closeness between any two nodes
  - Total number of brokers by type
  - Note: to compare networks of different size you must normalize network statistics
- Density
  - Total observed ties out of total possible ties
  - More connections and shorter paths between actors mean information can travel quickly
  - Fewer connections mean the information may be novel
  - Network structure may influence group dynamics



Exponential random graph models (ERGM)

- Account for the presence (and absence) of network ties and so provide a model for network structure
- Considered to arise from local social processes, where by actors in the network form connections in response to other ties in their social environment
  - Homophily – the more similar you are to me, the more likely I am to choose you as a friend
  - Transitivity – a friend of my friend is my friend
  - Reciprocity – if you choose me as a friend I'm more likely to choose you as a friend

Kind of like logistic regression

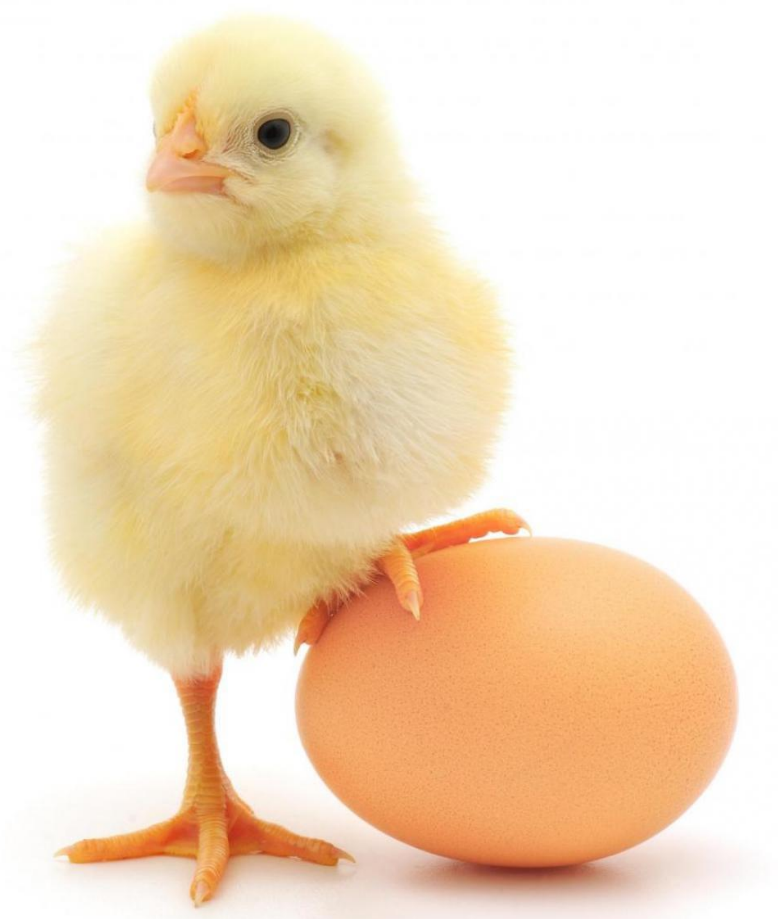
- ERGMs attempt to predict a binary variable (the presence or absence of a tie in a network) from a number of predictor variables
- Coefficients in the model indicate how important a variable is in determining the outcome
  - Interpreted as log odds (which can be transformed into odds or probabilities)

Log Odds of a Choosing a Friend Based on Racial Homophily and Mutual Friends

	School 1		School 2	
	Estimate	S.E.	Estimate	S.E.
Friendship (Edges)	-7.28	0.14***	-7.12	0.15***
Indegree	0.19	0.01***	0.20	0.01***
Reciprocity	3.45	0.12***	3.21	0.14***
Two-paths	-0.26	0.02***	-0.25	0.02***
Gender Homophily	1.99	0.11***	1.60	0.10***
Shared Classes	0.58	0.02***	0.41	0.02***
Racial Homophily				
Partial Homophily	0.35	0.13**	0.62	0.14***
Full Homophily	0.74	0.08***	0.78	0.09***
Mutual Friendship Nominations Received	0.92	0.06***	0.99	0.08***
Mutual Friendship Nominations Given				
No Racial Homophily (Cross-Race) Dyads				
Mutual Friends	1.27	0.07***	1.42	0.09***
Mutual Cross-Race Friends (race matches nominee)	1.49	0.17***	1.87	0.20***
Mutual Biracial Friends	1.00	0.18***	0.92	0.20***
Mutual Same-Race Friends (race matches nominator)	0.77	0.19*	0.50	0.23*
Partial Racial Homophily (Mono- to Biracial) Dyads				
Mutual Friends	-0.06	0.14	-0.52	0.17**
Mutual Biracial Friends	0.77	0.19*	0.51	0.42
Full Racial Homophily (Same-Race) Dyads				
Mutual Friends	-0.17	0.10	-0.45	0.14***

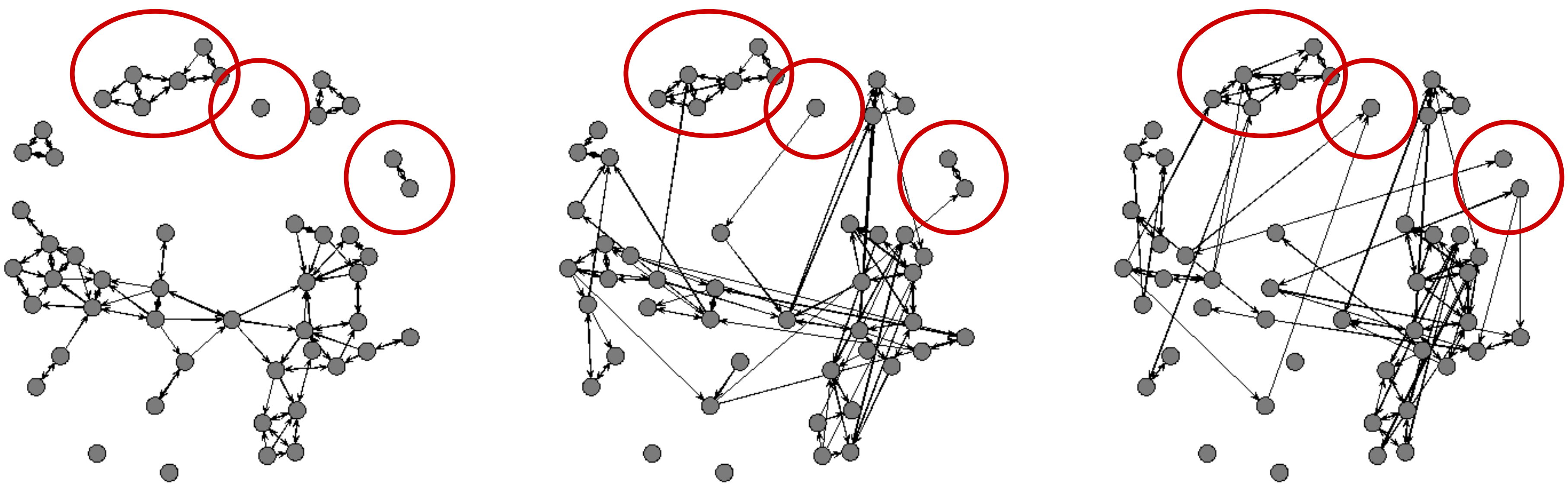
SIENA models

- Which came first, the tie (relationship) or the behavior?
  - Do I want to be your friend because I am like you *or* am I like you because you are my friend?
- Siena models attempt to answer this question (similar to cross-lagged models)



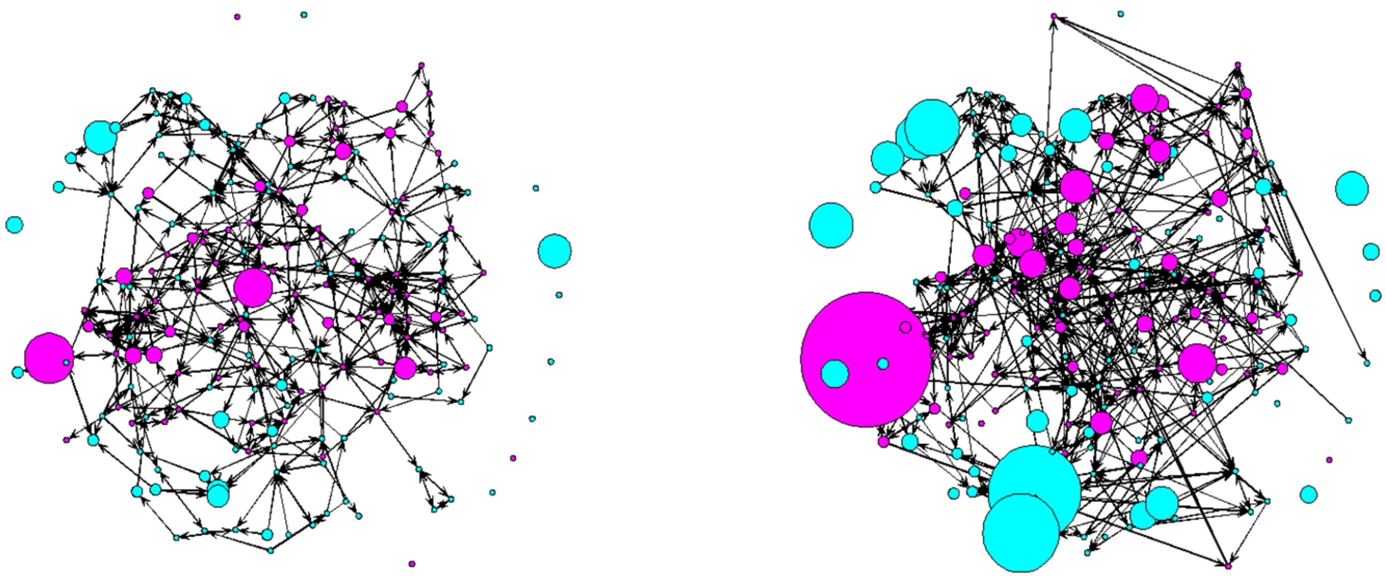
Dynamic nature of social networks

- Relational changes over time
  - Some ties may be established while other ties dissolve
  - Changes may be the result of structural position of actors (e.g., when friends of friends become friends), characteristics of actors or pairs of actors (i.e., actor/dyadic covariates), or residual random influences



Dynamic nature of social networks

- Behavioral changes over time
  - Associations with others may influence characteristics of actors in the network
  - Although these changes are referred to as behavioral dynamics, they can represent reputations, attitudes, feelings, or even physical and/or physiological states



False selection effects

- Dijkstra, Berger, & Lindenberg, 2011
- Early adolescents are likely to select friends similar to themselves in a number of ways (e.g., sex, race, social class, academic performance…and aggression)
  - Some of these relations are unidirectional (i.e., your race may influence who you choose as friends, but your friends can’t influence your race) while some may be bidirectional

False selection effects

CAUTION

- If similarity in aggression is the result of a selection process only, it may be the by-product of other factors
  - If aggression is influenced by gender, and gender influences friendship selection, aggression may look like a selection effect when it is not
  - If two aggressive adolescents who have a mutual friend are friends themselves, what appears to be a selection effect may actually be a transitivity effect
  - The moral of the story is that selection effects are easily inflated

Kind of like logistic regression…and then some

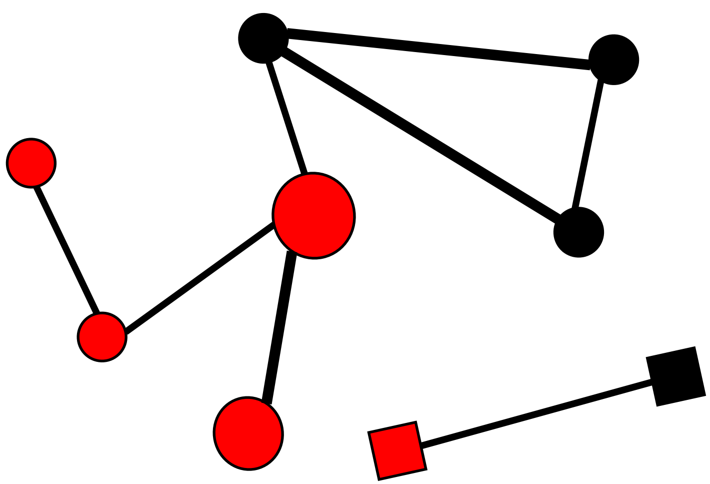
- Like ERGMs, Siena models attempt to predict the presence or absence of a tie in a network
  - Based on possible tie changes (e.g., formation, dissolution, maintenance)
- Include a rate function
  - For selection only models (one DV—tie) this refers to how often an actor gets the opportunity to change the state of a tie, whether or not he/she actually makes a change
  - For selection and influence models (two DVs—tie and behavior) this refers to how often an actor gets the opportunity to change the state of a tie *and* his/her behavior

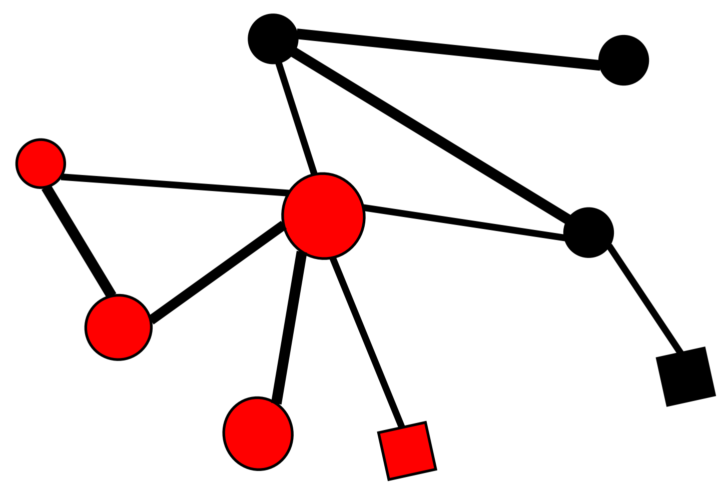
Data requirements (rules of thumb)

- At least 2 waves but usually less than 10
  - Between any pair of consecutive waves, the number of changes should not be too high
  - start by analyzing the transitions between each consecutive pair of observations
- At least 20 actors but no more than 300
  - the assumption that each actor is a potential network partner for any other actor might be implausible for networks with so many other actors that not all actors are aware of each other's existence
- Network data (in principle) should be complete
  - In practice 80% is close enough

Network visualization

- Change node color or shape for categorical variables
- Change node size for continuous variables
- increase thickness of line based on strength (or some value) of relationship
- Freeze nodes over time to see changes in ties and characteristics of nodes





Want to learn more?

- Upcoming workshops
  - SRA post-conference
    - 3 hours for $30
    - https://www.s-r-a.org/pre_postconferences
  - Missouri State
    - 3 days for $300
    - https://calendar.missouristate.edu/event/97776/185298
  - Stats Camp
    - 5 days for lots of $
    - https://www.statscamp.org/summer-camp/introduction-to-social-network-analysis-using-r-and-rsiena