Introduction to Social Network Analysis
SBSRC Interdisciplinary Workshop Series

Jeffrey A. Smith
University of Nebraska-Lincoln
Department of Sociology

Course Website:

https://jeffreysmith-sociology.com/workshops/introduction-to-social-networks/

The Macro Structure of the Course

- Introduction to network analysis (~ 3 hours)
  - The what, why and how of network analysis
- Lunch (1 hour)
  - Eating…
- Doing basic network analysis in R (~4 hours)
  - Introducing R
  - Dealing with network data in R
  - Plotting and basic measurement

Goals of This Course

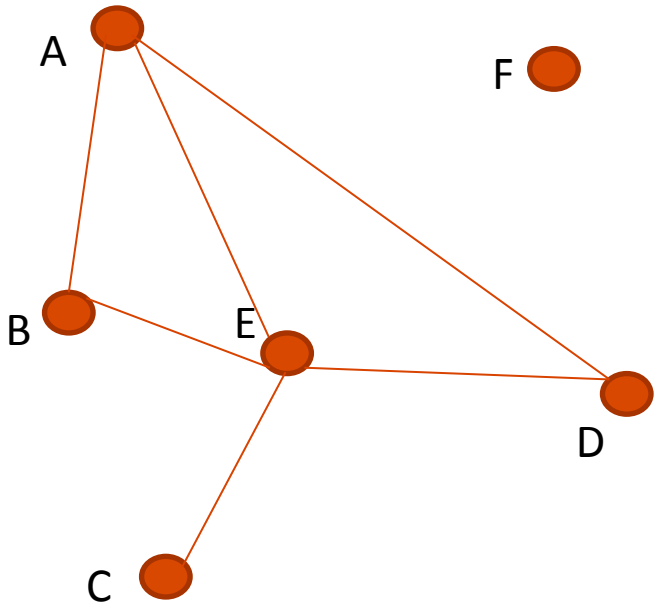
1. Understand the potential and limitations of a network approach
2. Understand the unique features of network data
3. Manipulate network data in R
4. Create an attractive network plot in R

- Name
- Department/Affiliation
- Why take a 1 day seminar on network analysis?

Part I. The ‘What’ of Network Analysis

Q) What is a network?

A) Network=set of links between objects



Preliminaries: Objects and Links

- Objects are often people but can be defined quite broadly
  - Organizations, animals, airports…
- Links are often interactions between people, but can also be defined quite broadly
  - Friendship, kinship, sex, knowing, advice, support, exchange/trades, fight, bully, etc…
  - Links are properties of the pair, not the individual

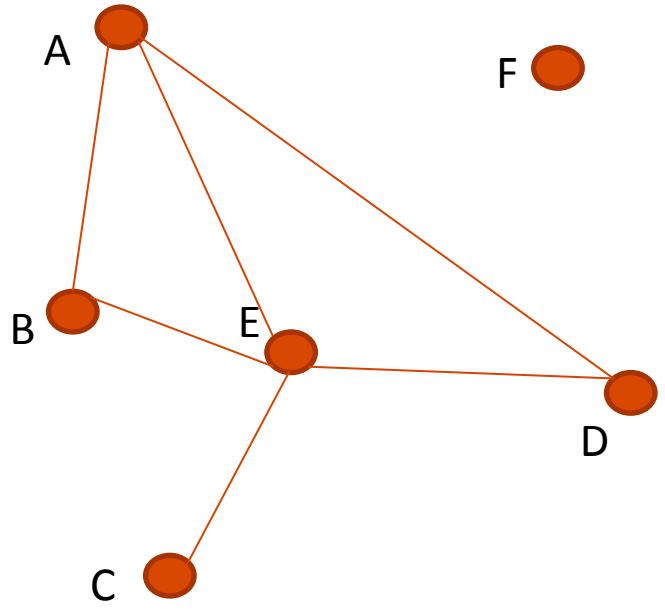
Examples of Links

Similarities			Social Relations				Interactions	Flows
Location	**Membership**	**Attribute**	**Kinship**	**Other role**	**Affective**	**Cognitive**		
e.g.,	e.g.,	e.g.,	e.g.,	e.g.,	e.g.,	e.g.,	e.g.,	e.g.,
Same spatial and temporal space	Same clubs	Same gender	Mother of	Friend of	Likes	Knows	Sex with	Information
	Same events	Same attitude	Sibling of	Boss of	Hates	Knows about	Talked to	Beliefs
	etc.	etc.		Student of	etc.	Sees as happy	Advice to	Personnel
				Competitor of		etc.	Helped	Resources
							Harmed	etc.
							etc.	

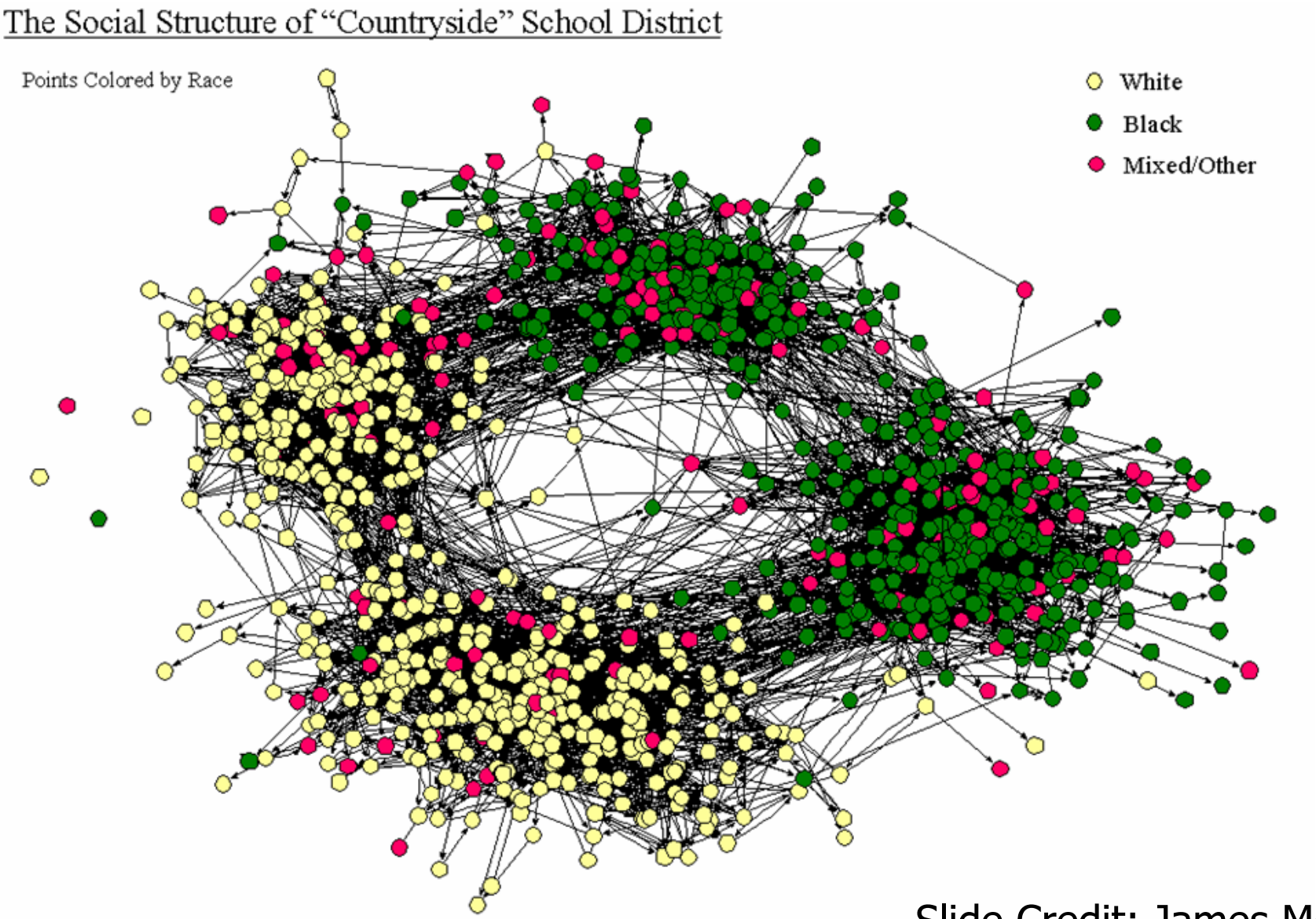
Credit: Borgatti et al. 2011

Clarifying Terminology

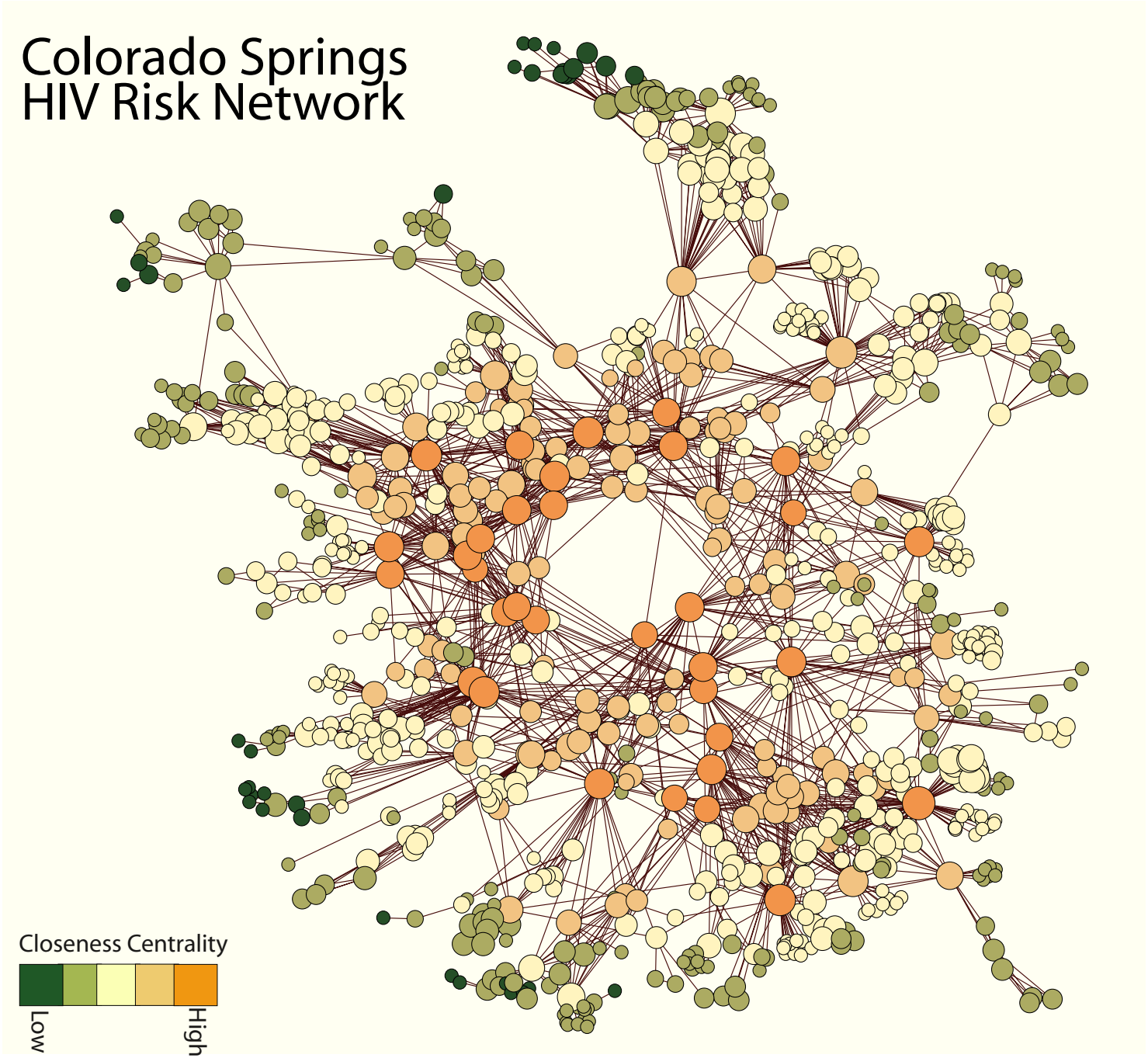
- Different words for the same thing
- Objects: nodes, actors, vertices
- Links: edges, ties, connections, relations

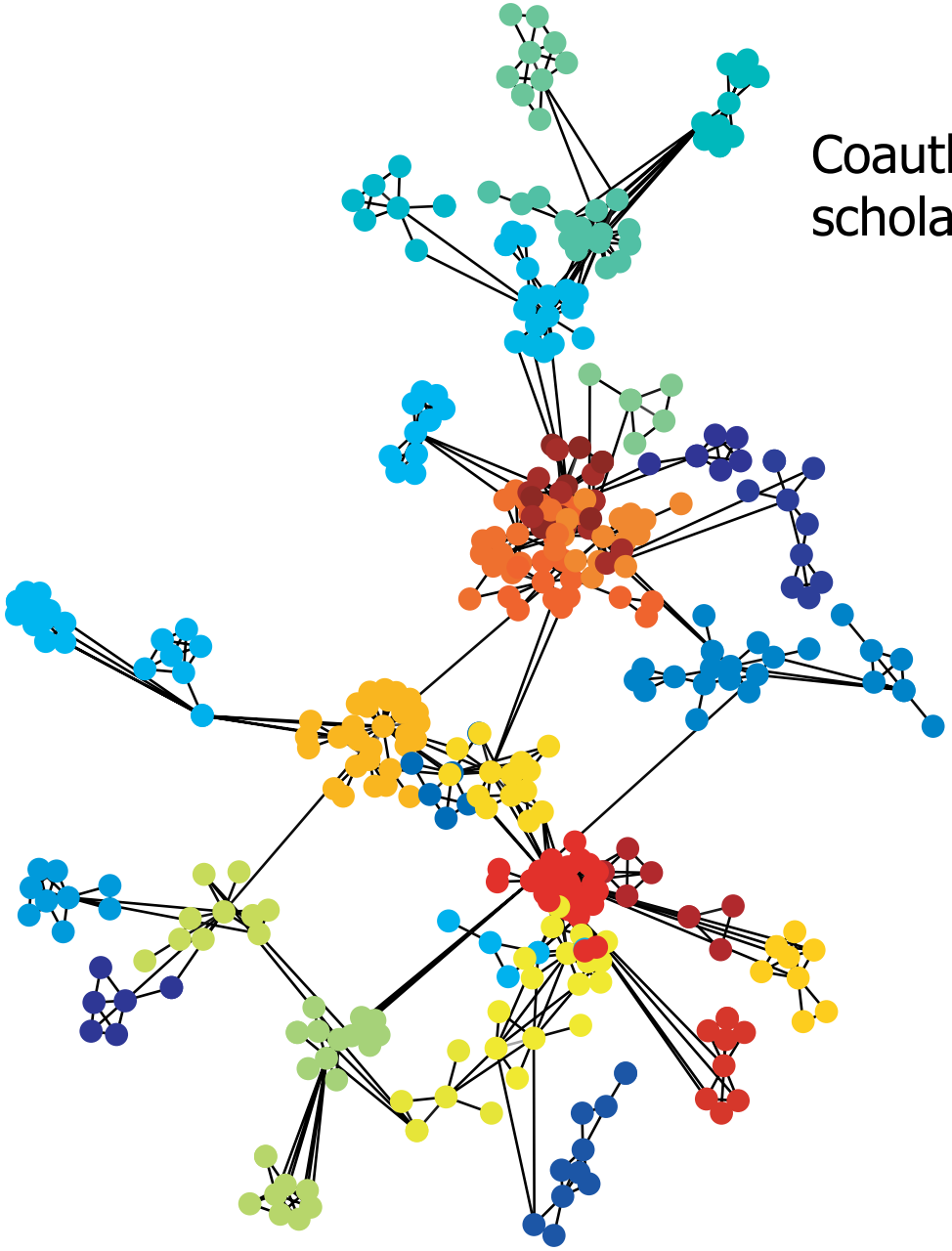


Example Networks



Slide Credit: James Moody


Colorado Springs
HIV Risk Network
Closeness Centrality
Low
High



Coauthorship between scholars in physics literature

Slide Credit: Porter et al. 2009

The ‘What’ of Network Analysis

- What is network analysis?
- Set of relational methods for studying the connections between actors
  - Based on computational methods and graphical imagery
  - Interested in the ties between actors rather than the actors themselves
  - Formally capture the complex pattern of connections amongst actors using empirical data

Part II. The 'Why' of Network Analysis

- "To speak of social life is to speak of the association between people – their associating in work and in play, in love and in war, to trade or to worship, to help or to hinder. It is in the social relations men establish that their interests find expression and their desires become realized." Peter M. Blau *Exchange and Power in Social Life*, 1964

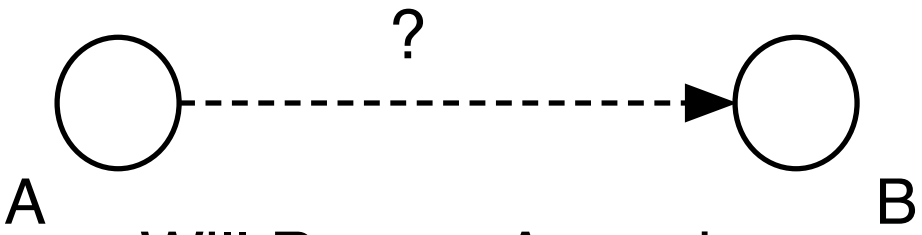
The ‘Why’ of Network Analysis

- For the last thirty years, empirical social research has been dominated by the sample survey. But as usually practiced, . . ., the survey is a sociological meat grinder, tearing the individual from his social context and guaranteeing that nobody in the study interacts with anyone else in it. It is a little like a biologist putting his experimental animals through a hamburger machine and looking at every hundredth cell through a microscope; anatomy and physiology get lost, structure and function disappear… If our aim is to understand behavior rather than simply record it, we want to know about primary groups, neighborhoods, organizations, social circles, and communities; about interaction, communication, role expectations, and social control.” Barton 1968, quoted from Freeman 2004

The ‘Why’ of Network Analysis

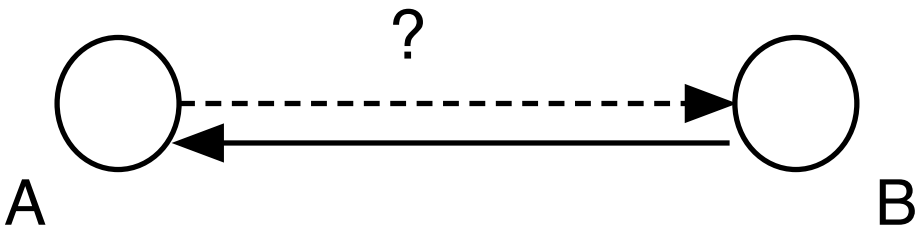
- Why study social networks?
  - Because individuals do not live in a vacuum
  - Because actors are interdependent
    - Need to study social world as a system
  - Network structure matters
    - Allows for explanations that focus on relationships and positions rather than individual traits

Why focus on Interdependences? The Simple Case of Two Actors



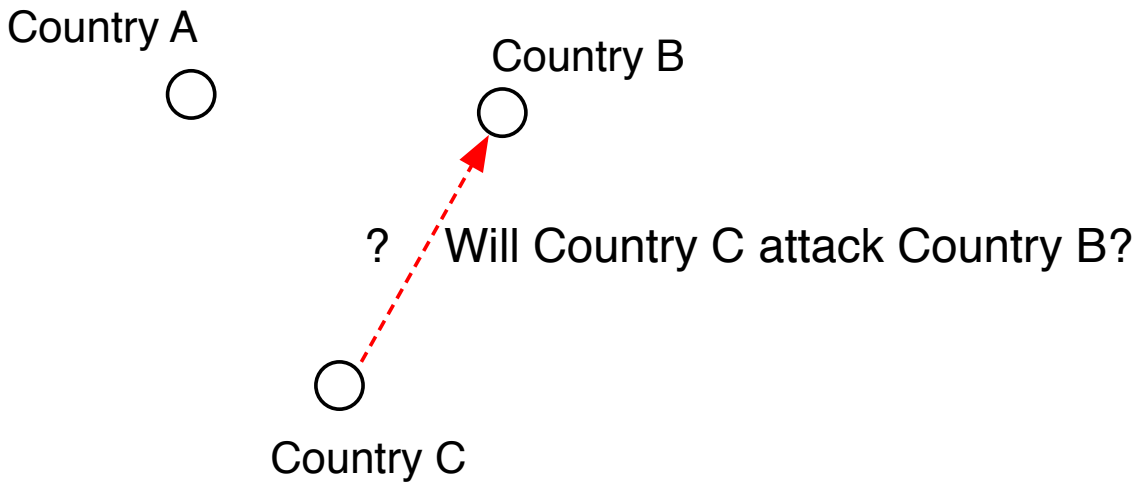
Will Person A send a gift to Person B?

Why focus on Interdependences? The Simple Case of Two Actors

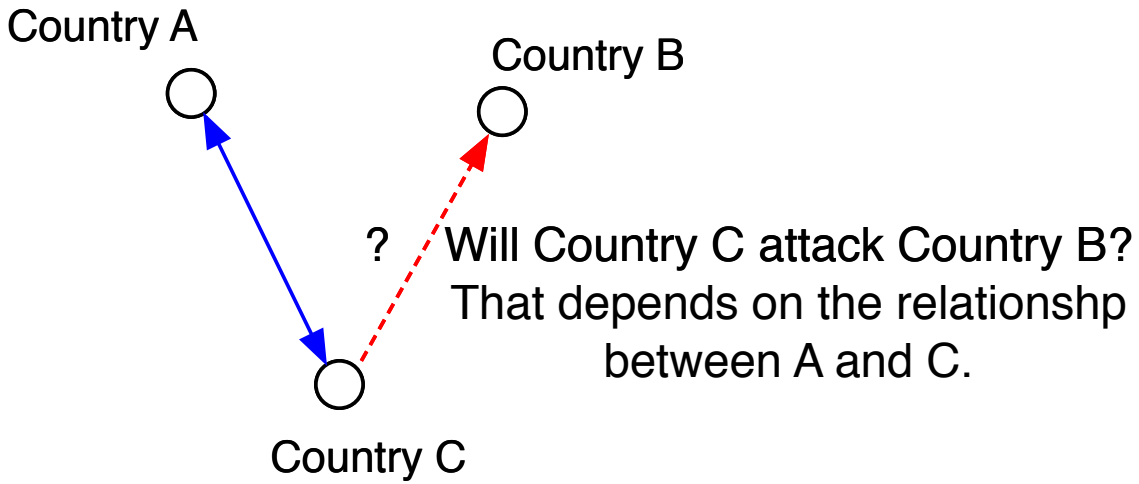


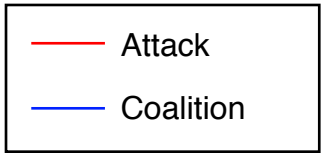
Will Person A send a gift to Person B? That depends if B sends gifts to A

Why focus on Interdependences? Adding a Third Actor

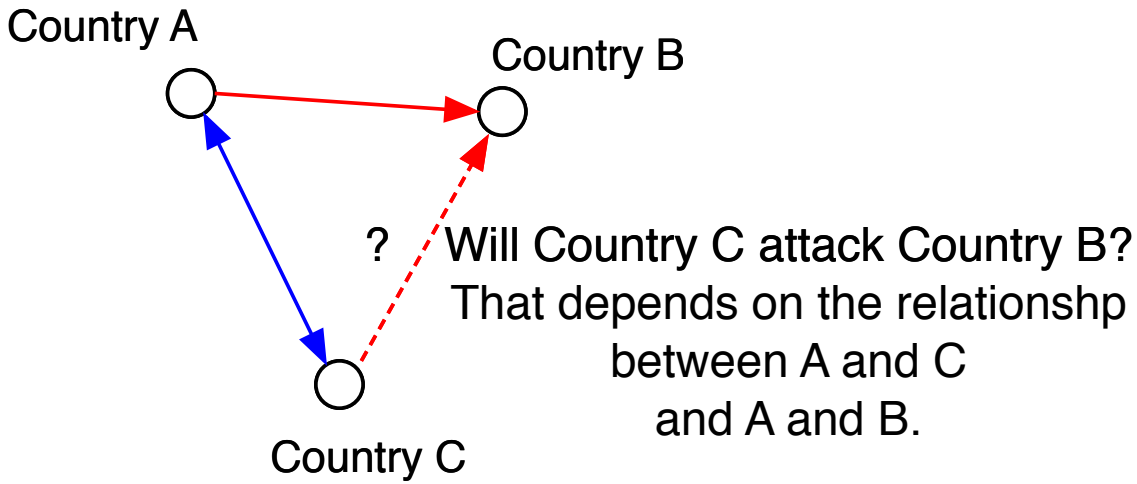


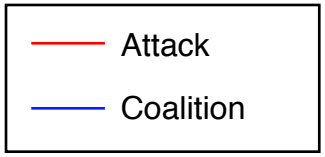
Why focus on Interdependences? Adding a Third Actor



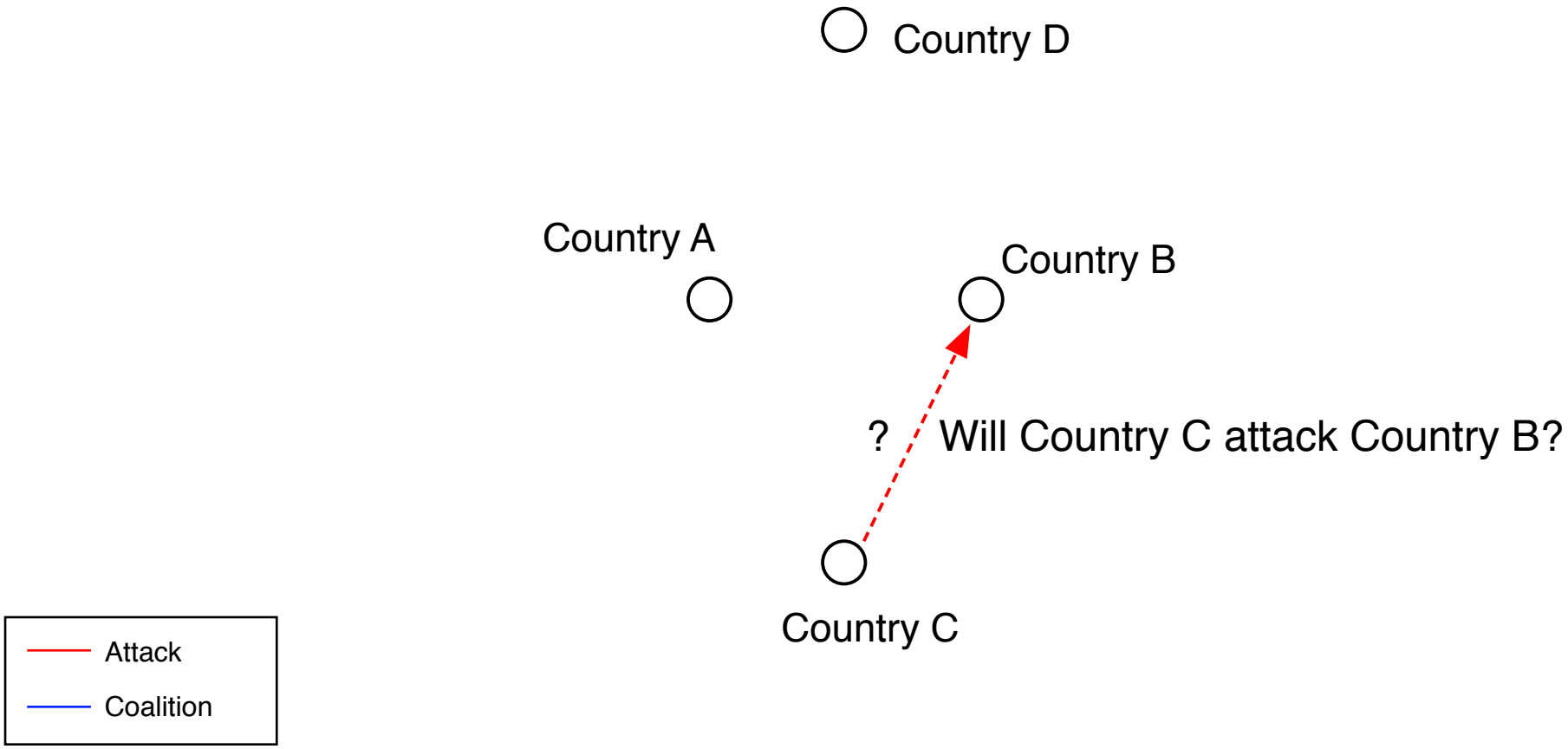


Why focus on Interdependences? Adding a Third Actor

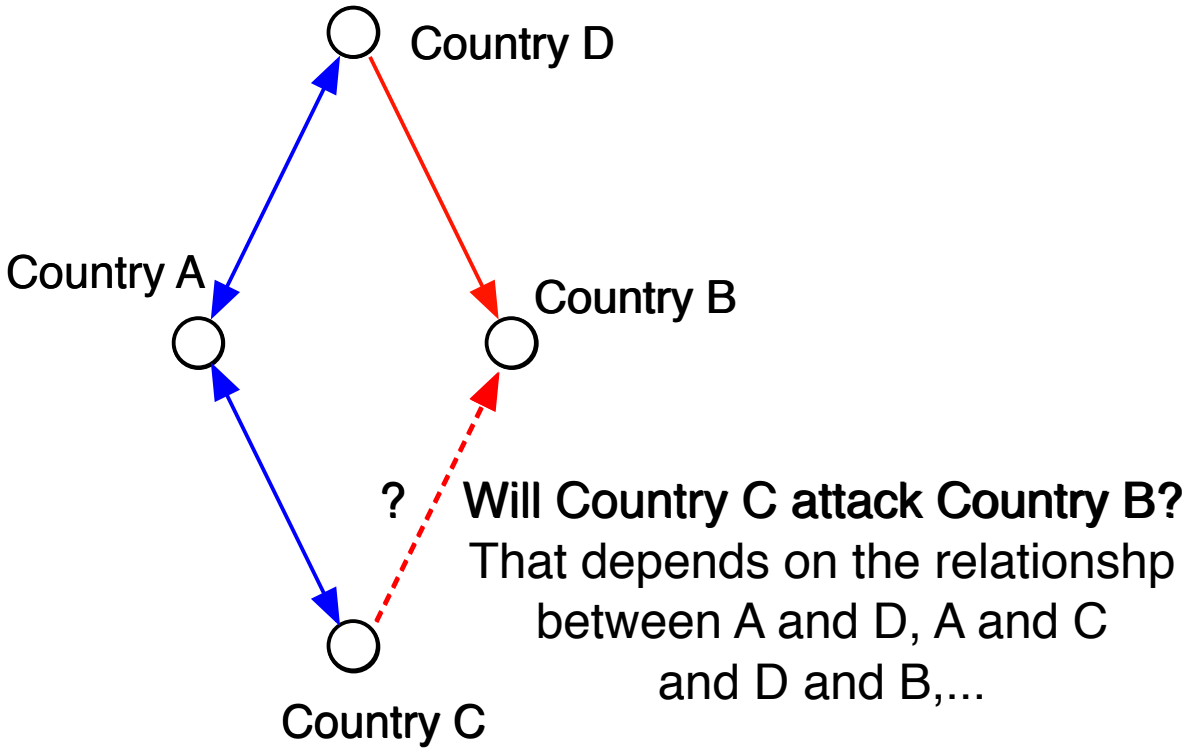




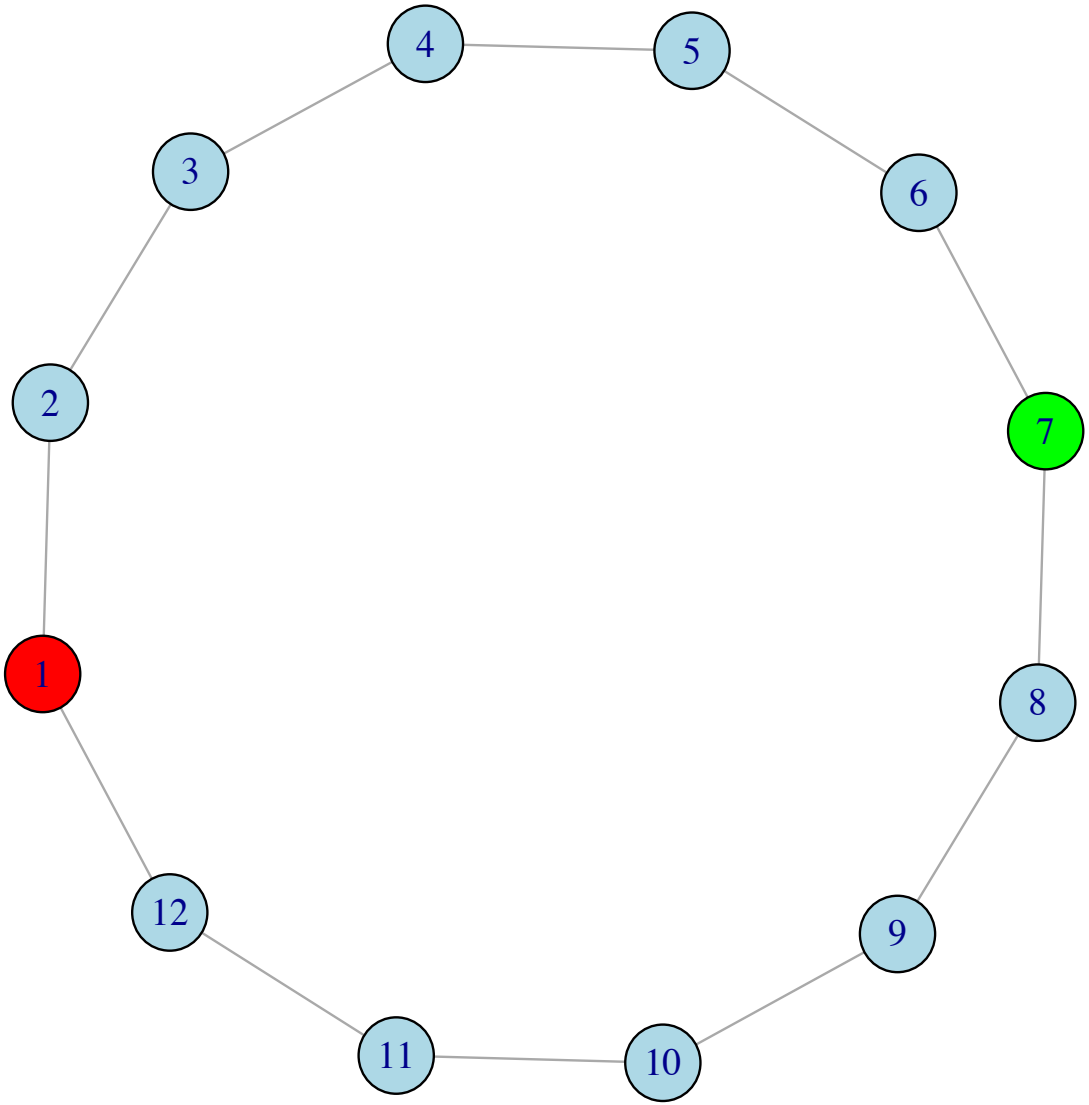
Why focus on Interdependences? Moving Beyond Triads



Why focus on Interdependences? Moving Beyond Triads

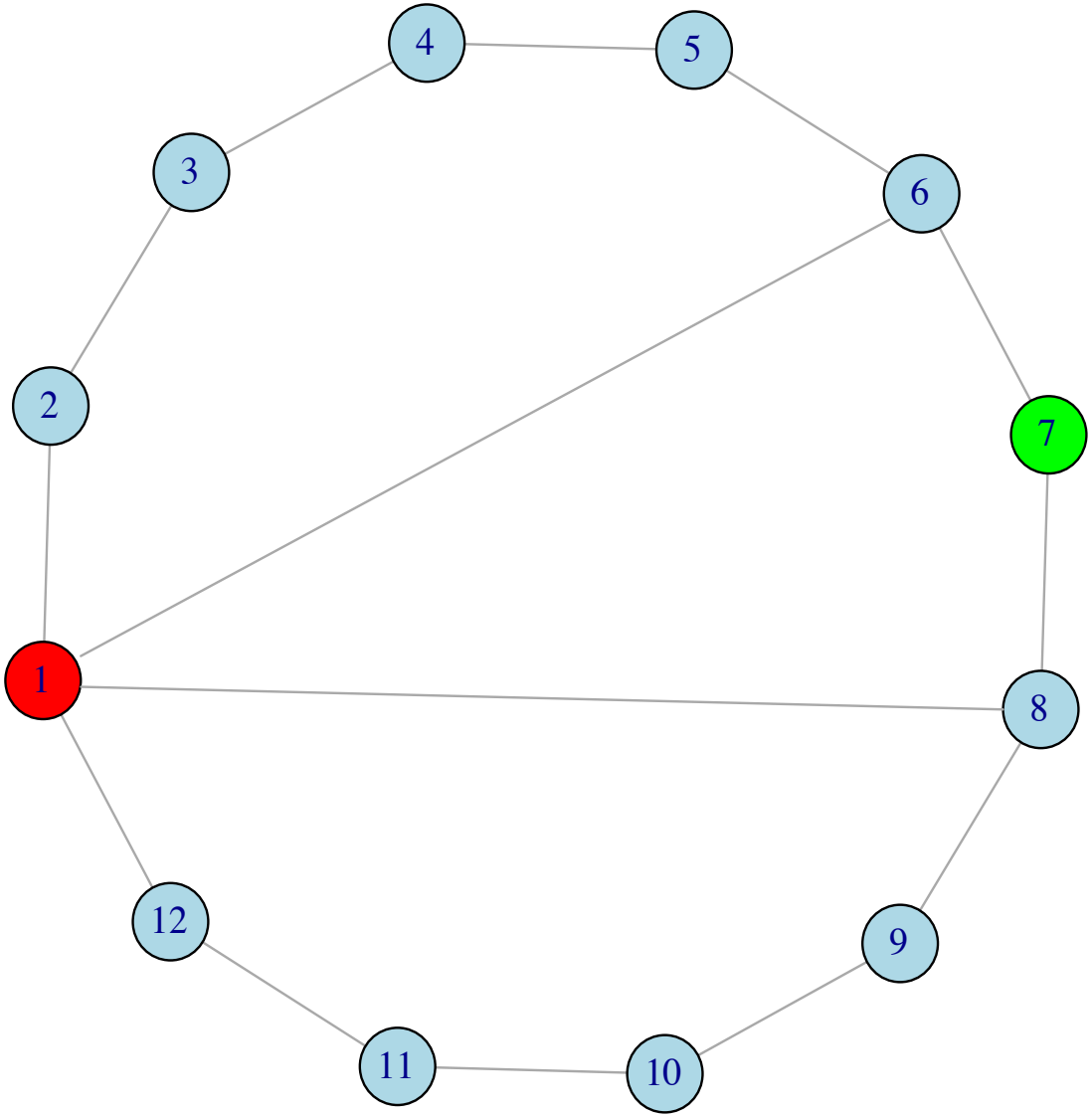


Why focus on Interdependences? Moving Way Beyond Triads



- What is the probability of a rumor going from 1 to 7?

Why focus on Interdependences? Moving Way Beyond Triads



- What is the probability of a rumor going from 1 to 7?

Example Questions

- Network as outcome
  - What underlying behavioral rules govern the formation of a tie?
  - Is the tendency toward homophily strong or weak?
  - What positions/roles exist?
  - How hierarchical is the network?
  - How/why do these things vary across context?

Example Questions

- Network as cause
  - How does being isolated (or embedded) affect health and mental health?
  - How does being in a particular network position lead to better/worse job leads?
  - How does the deviant behavior of peers affect our own behavior?
  - What is the potential for an epidemic?

III. The "How" of Network Analysis

- Methods of Collection
  - Observational
  - Survey
  - Archival
  - Digital records
    - Internet
    - Sensor data
    - Cell phone records

Assumption when collecting data

- Boundaries
- Captured meaningful relations on population of interest
- Missing data
  - Often assume we have no missing data
  - But this is almost never true!

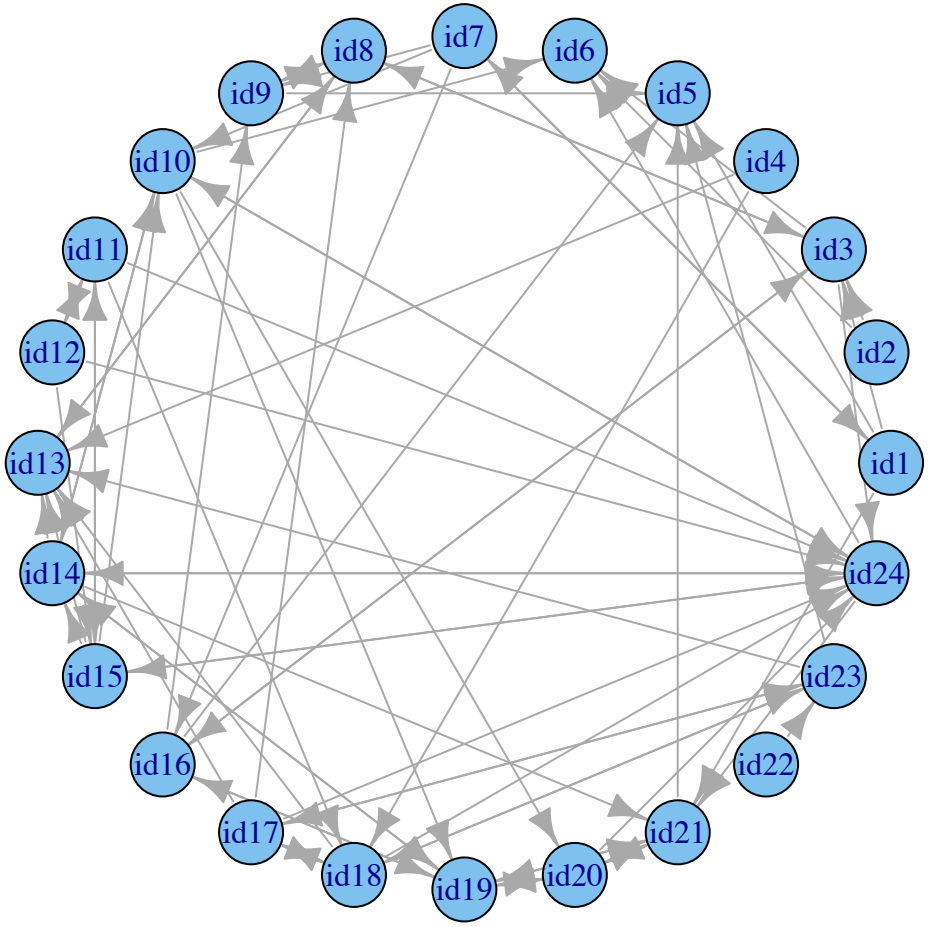
Representing Networks

- Need formal, systematic means of analyzing our network data
- Start with representing network data
  - Need both graphical and numerical representations

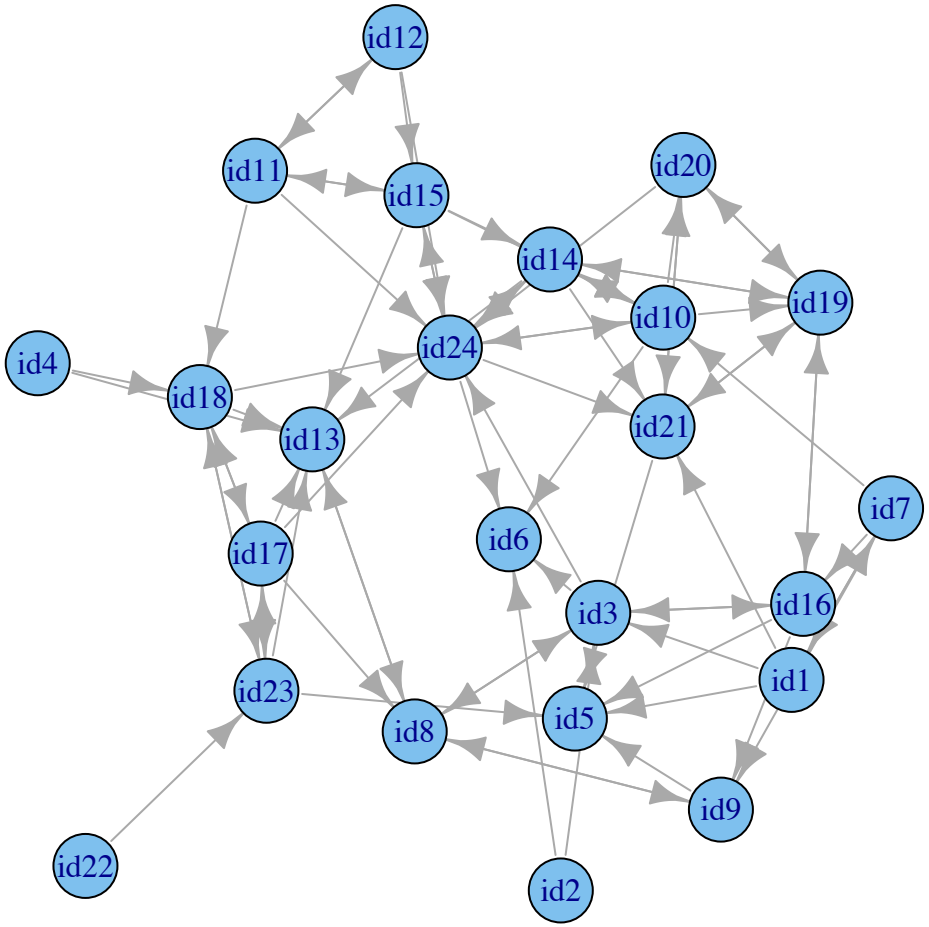
Graphs and Matrices

- Graphs and matrices offer two different ways of summarizing the same data
- Graphs
  - Advantage: offer an intuitive representation of network
  - Disadvantage: not analytically tractable
- Adjacency Matrix
  - Advantage: can calculate measure of interest from matrix
  - Disadvantage: not an intuitive (or efficient) way of representing the network

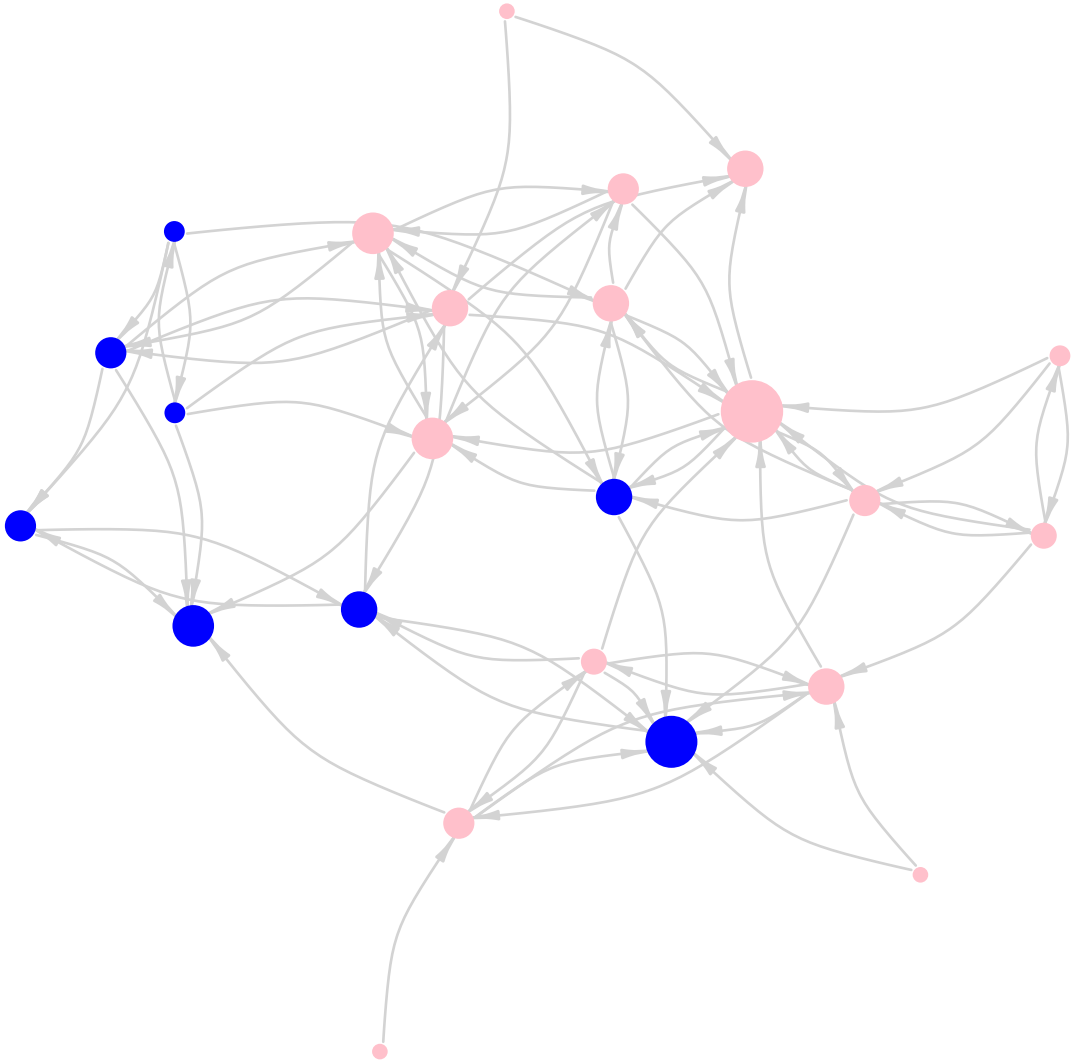
Not all figures are equally good…



Not all figures are equally good…



Not all figures are equally good…

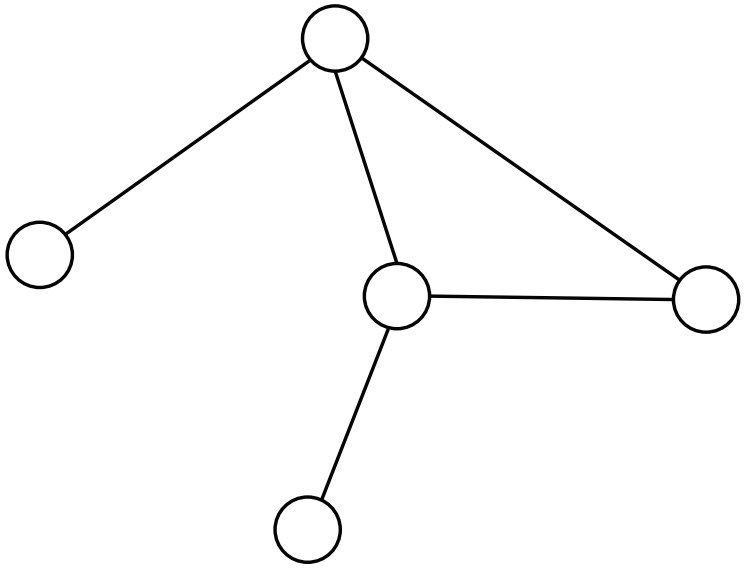


Graphical Representations

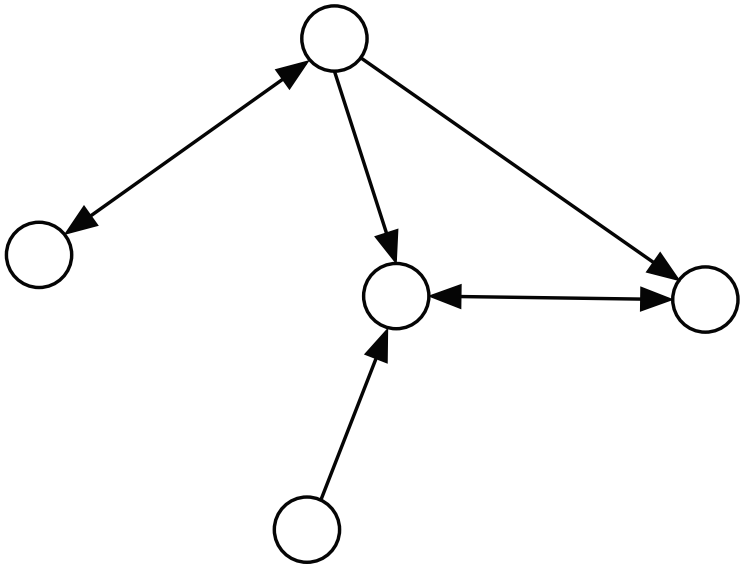
- Different types of networks yield different representations
  - Undirected versus directed
  - Binary versus valued
  - Single relation versus multiplex
  - One mode versus two-mode

Undirected or Directed (0/1 tie)

Undirected, Binary

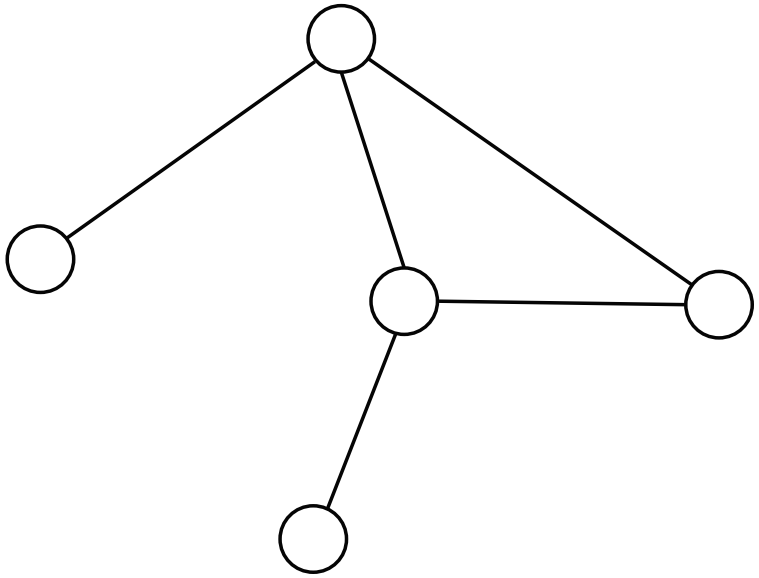


Directed, Binary

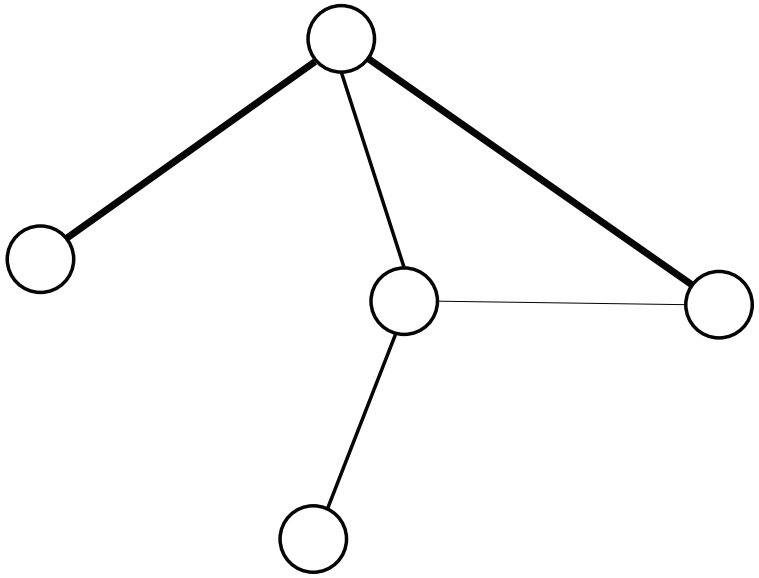


Binary or Valued (undirected)

Undirected, Binary

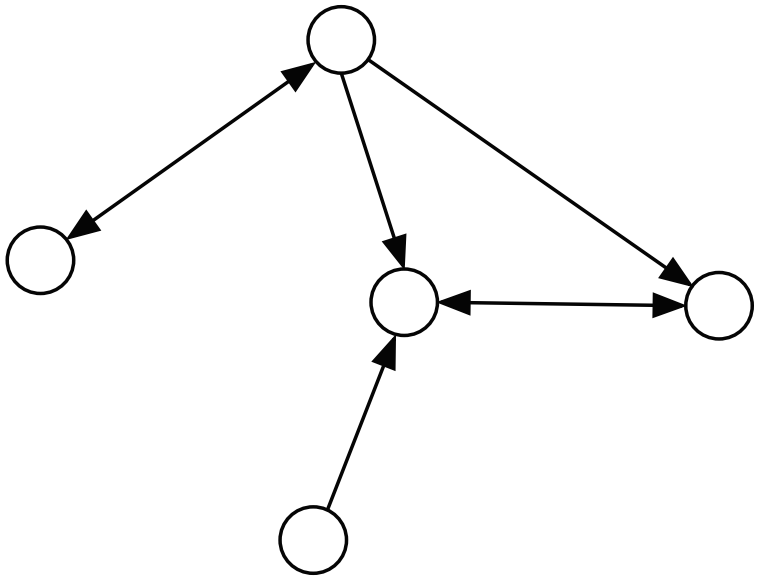


Undirected,Valued

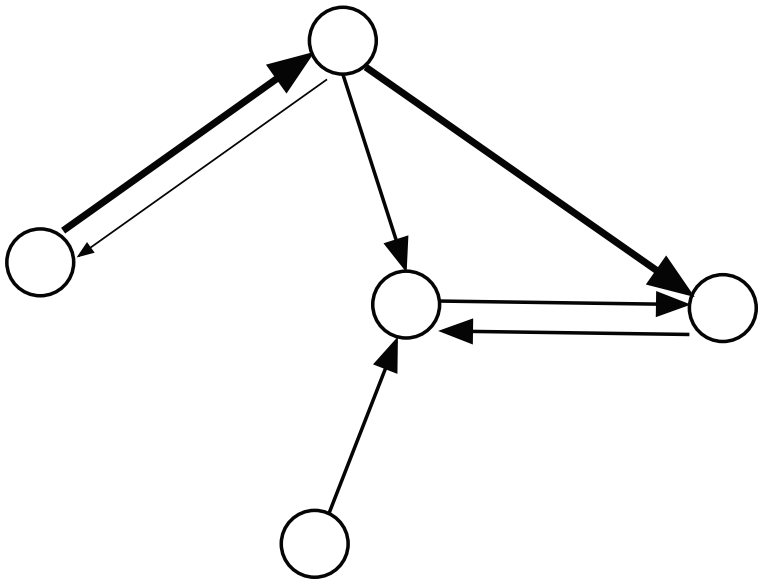


Binary or Valued (directed)

Directed, Binary

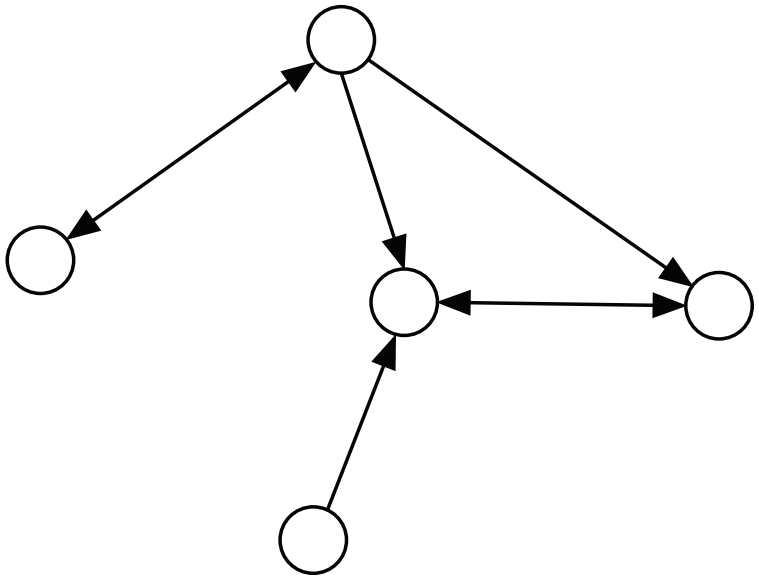


Directed, Valued

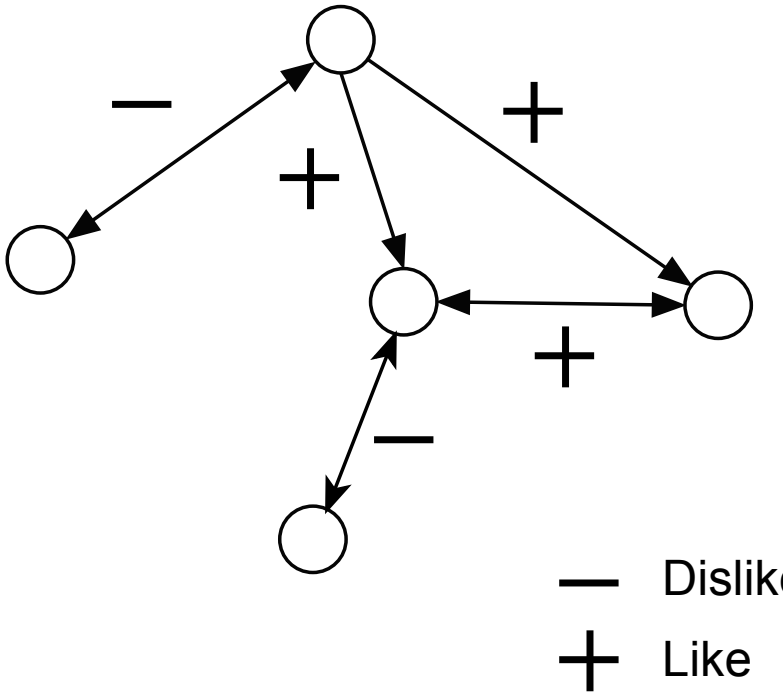


Binary or Signed (directed)

Directed, Binary

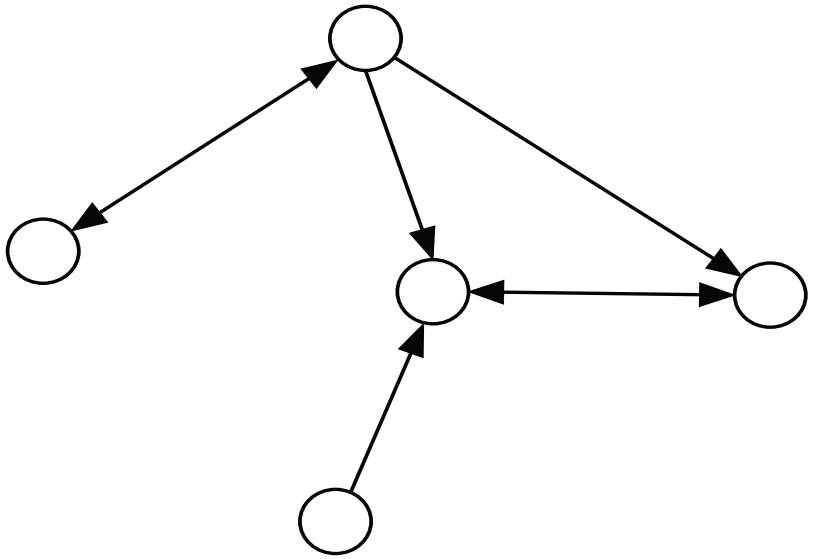


Directed, Signed

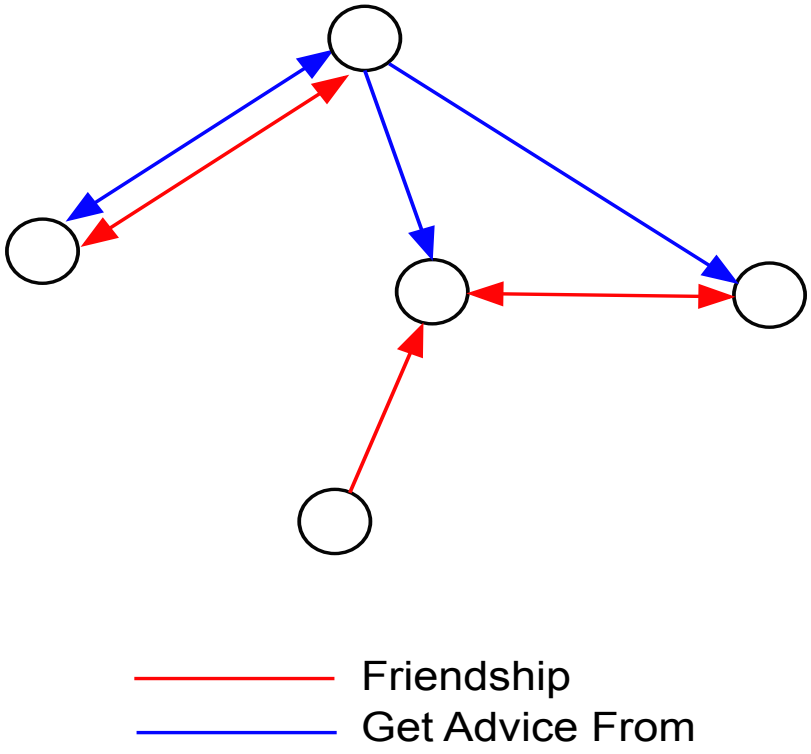


Single Relation or Multiplex (directed)

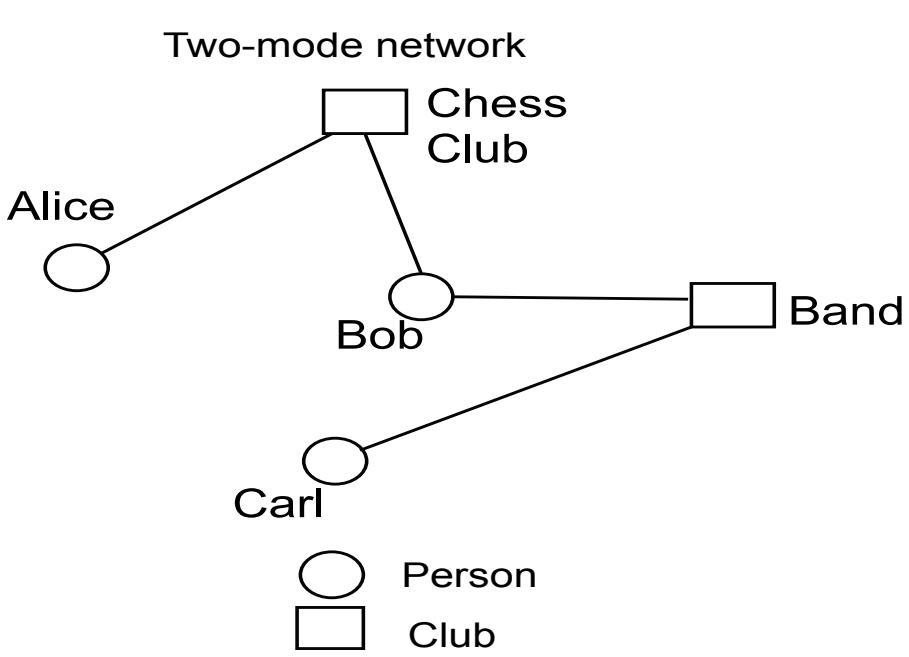
Directed, Binary, single relation

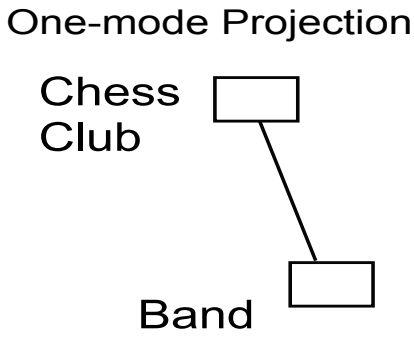


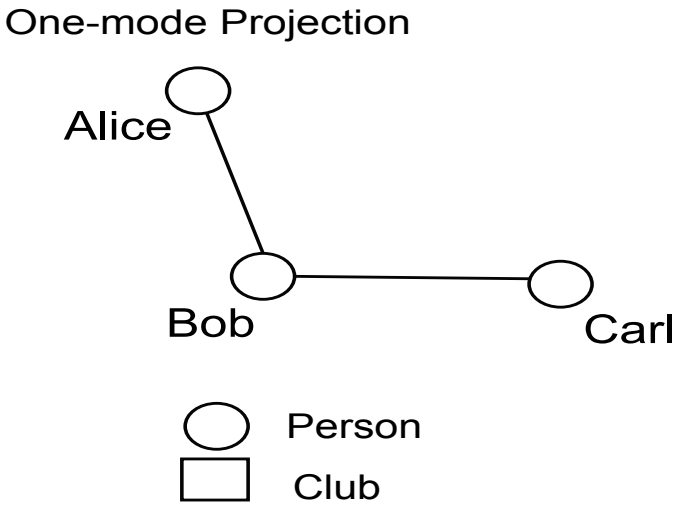
Directed, Binary, Muliplex



One Mode or Two Mode



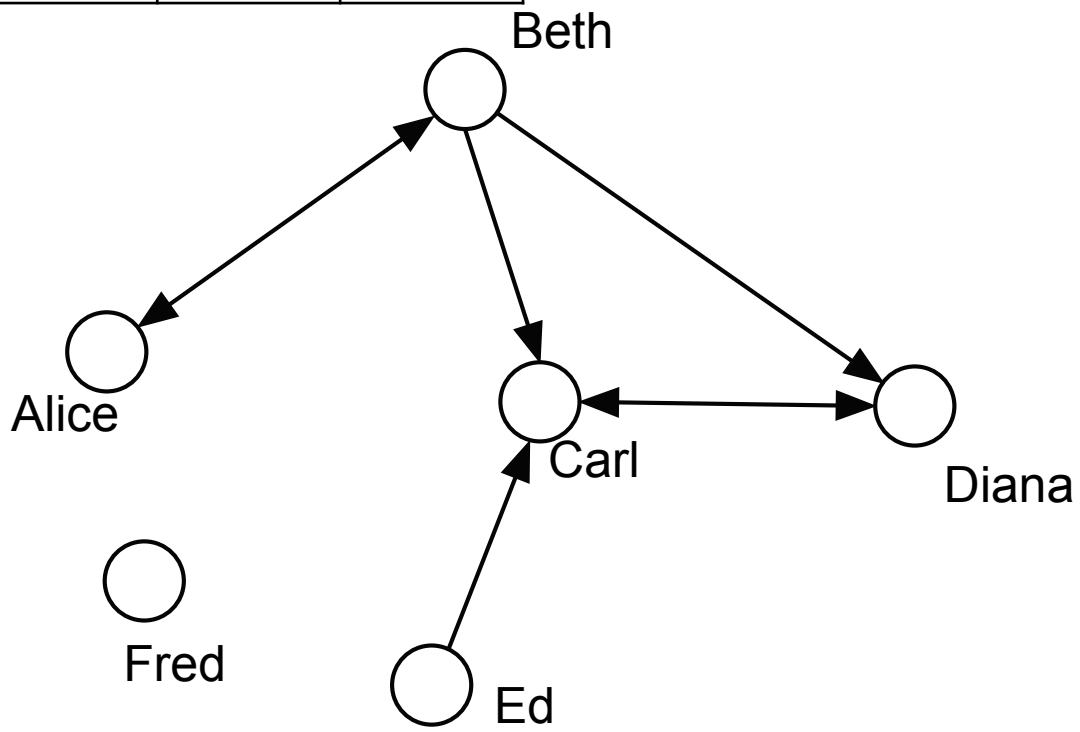




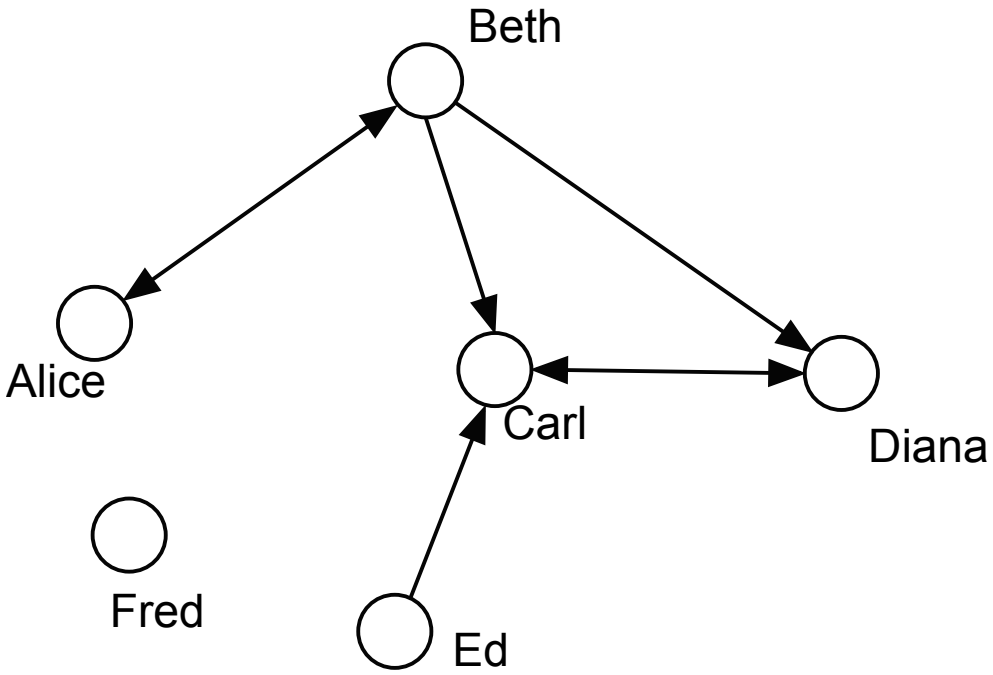
Matrix Representation

- Define a matrix X_{ij} such that the element i,j represents the existence/non-existence of a tie between i and j
- Row i corresponds to ties sent by actor i
- Column j corresponds to ties received by actor j

	Alice	Beth	Carl	Diana	Ed	Fred
Alice	0	1	0	0	0	0
Beth	1	0	1	1	0	0
Carl	0	0	0	1	0	0
Diana	0	0	1	0	0	0
Ed	0	0	1	0	0	0
Fred	0	0	0	0	0	0

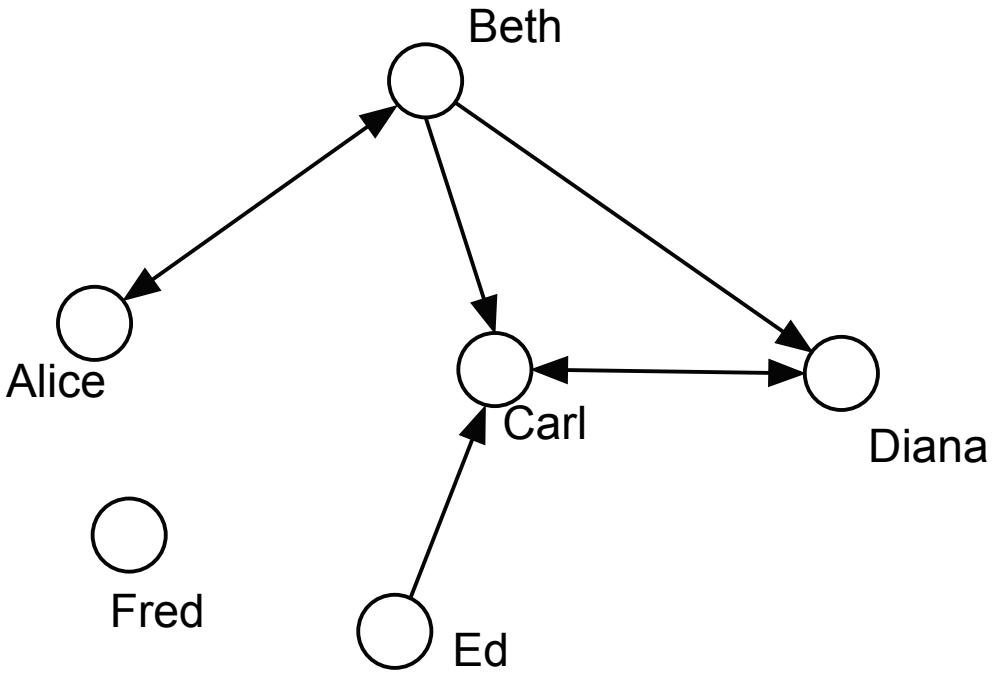


Alternative Representations: Edgelist



Sender	Receiver
Alice	Beth
Beth	Alice
Beth	Carl
Beth	Diana
Carl	Diana
Diana	Carl
Ed	Carl

Alternative Representation: Nomination List



ID	Nom1	Nom2	Nom3
Alice	Beth	.	.
Beth	Alice	Carl	Diana
Carl	Diana	.	.
Diana	Carl	.	.
Ed	Carl	.	.
Fred	.	.	.

Running through an example

- Let's say we collect the following information:
- Tim likes Sally, Harry, Mary
- Joe doesn't like anyone
- Sally likes Tim and Harry
- Harry likes Tim, Joe , and Sally
- Mary likes Joe
- What does the matrix look like for this network?

Tim likes Sally, Harry, Mary
Joe likes noone
Sally likes Tim and Harry
Harry likes Tim, Joe , and Sally
Mary likes Joe

	Tim	Joe	Sally	Harry	Marry
Tim					
Joe					
Sally					
Harry					
Marry					

What would the edgelist look like?

Tim likes Sally, Harry, Mary
Joe likes noone
Sally likes Tim and Harry
Harry likes Tim, Joe , and Sally
Mary likes Joe

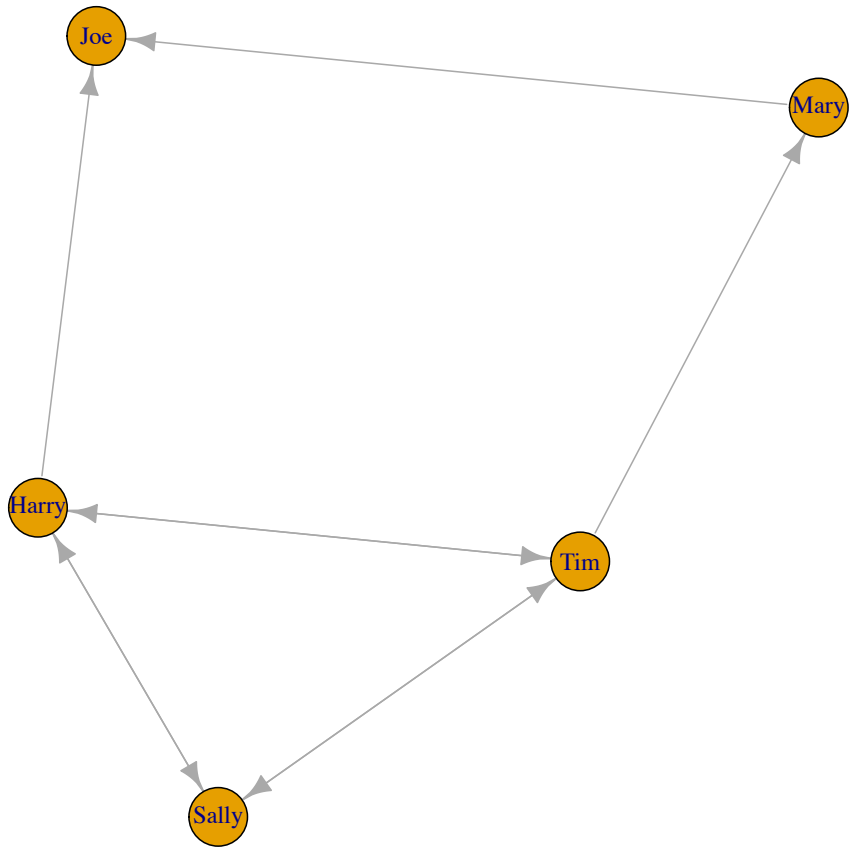
Sender	Receiver

Nomination List

ID	Nom1	Nom2	Nom3
Tim	Sally	Harry	Mary
Joe	.	.	.
Sally	Tim	Harry	.
Harry	Tim	Joe	Sally
Mary	Joe	.	.

Edgelist

Sender	Receiver
Tim	Sally
Tim	Mary
Tim	Harry
Sally	Tim
Sally	Harry
Harry	Tim
Harry	Joe
Harry	Sally
Mary	Joe



Matrix

	Tim	Joe	Sally	Harry	Mary
Tim	NA	0	1	1	1
Joe	0	NA	0	0	0
Sally	1	0	NA	1	0
Harry	1	1	1	NA	0
Mary	0	1	0	0	NA

Analyzing a Network

- Take the underlying matrix and calculate measures of interest
  - Examples: centrality, cohesion, group structure, roles/position, hierarchy, dynamics, diffusion
  - Measure becomes thing to predict or used as predictor of other variable of interest
- Could also try to model the network (predict ties between actors)
  - Local tendencies like transitivity and reciprocity

Measuring Network Structure and Diffusion Potential

- Network structure affects diffusion
- Networks measures related to diffusion
  - Density
  - Walks
  - Reachability
  - Components
  - Distance

Network Structure and Diffusion

- Higher density=more diffusion potential
- Higher reachability=more diffusion potential
  - Less “grouped”=more global diffusion potential
- Shorter paths=more likely diffusion
- More paths=more likely diffusion
  - Especially if independent paths

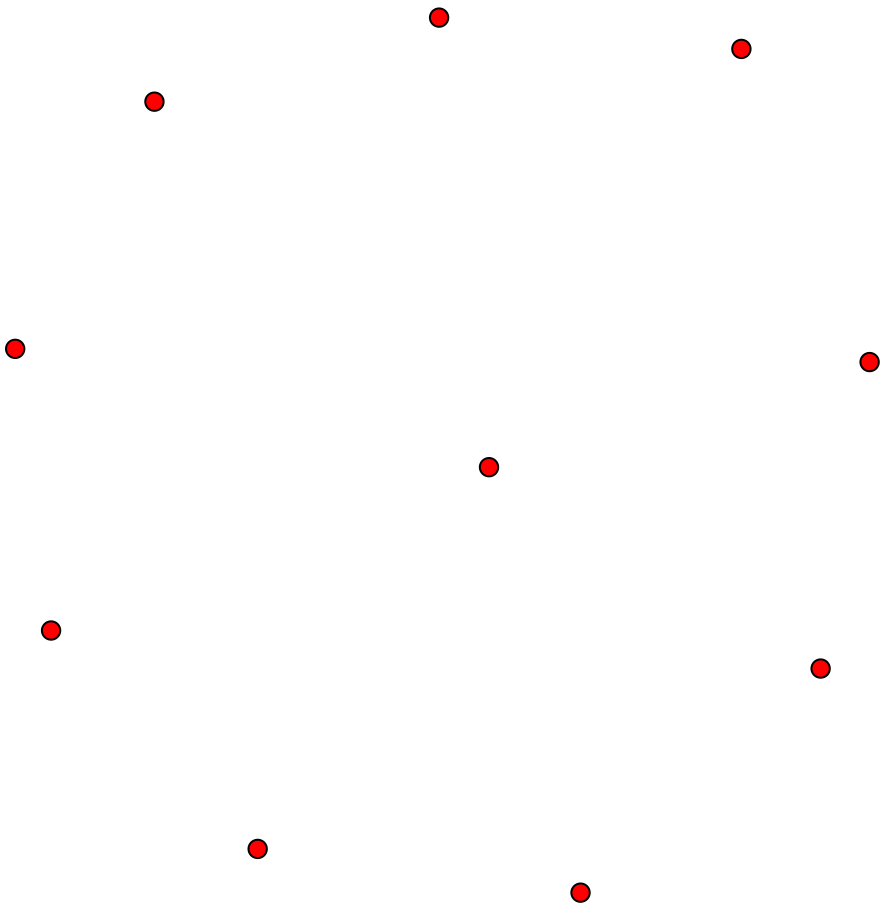
Some Simple Measures

- Density: total number of edges/total possible number of edges

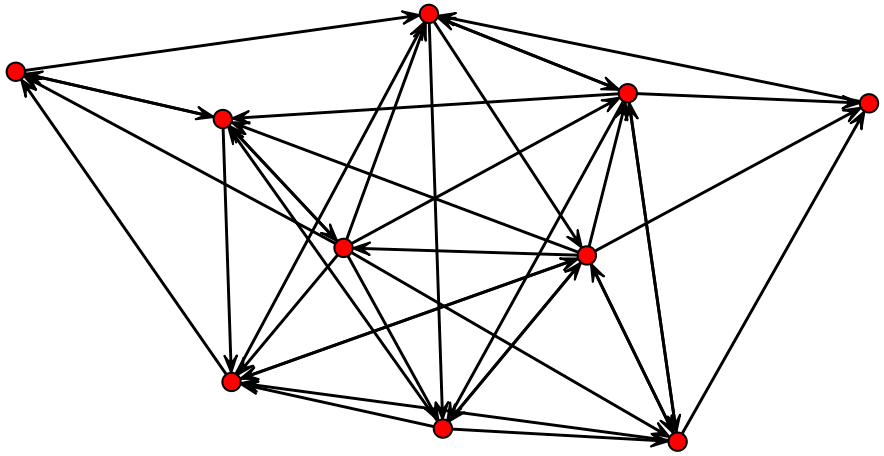
$$d = \sum_{i=1}^{n}\sum_{j=1}^{n} X_{ij} \Big/ (n * (n-1))$$

- Where X_{ij}=the network; n=number of people in network

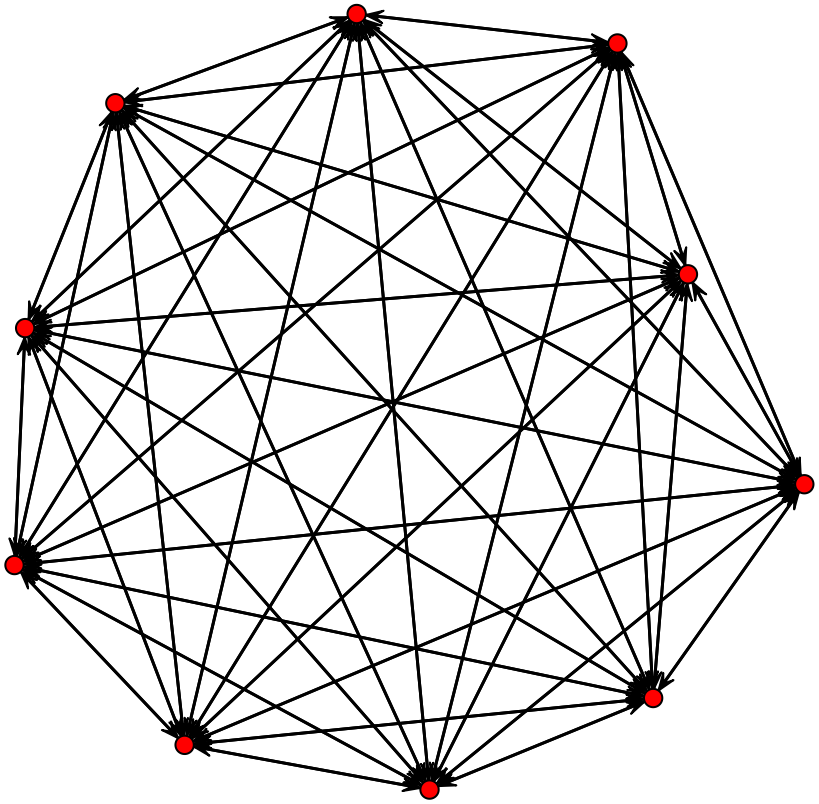
Density: 0/90=0



Density: 39/90=.433

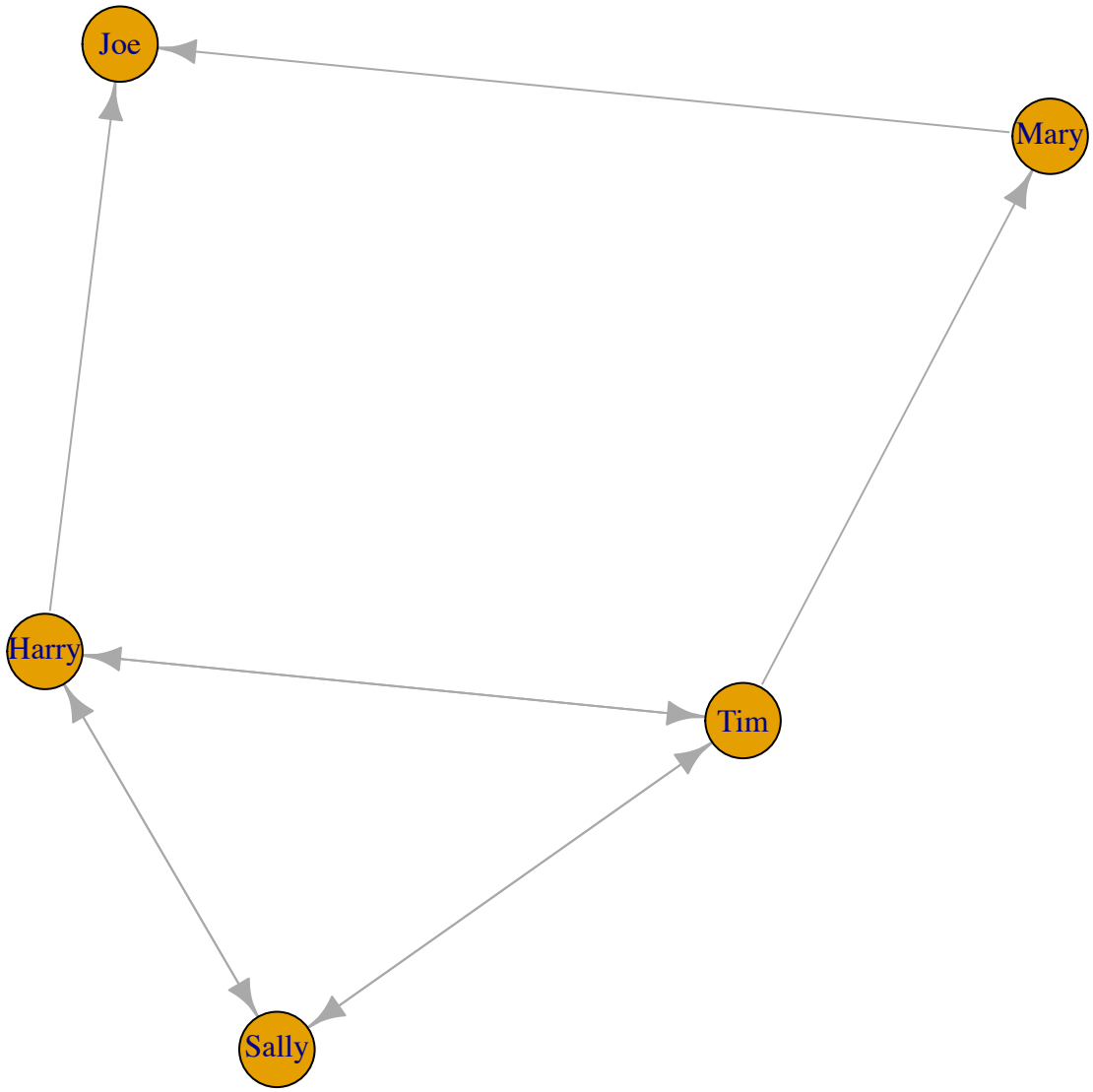


Density: 90/90=1

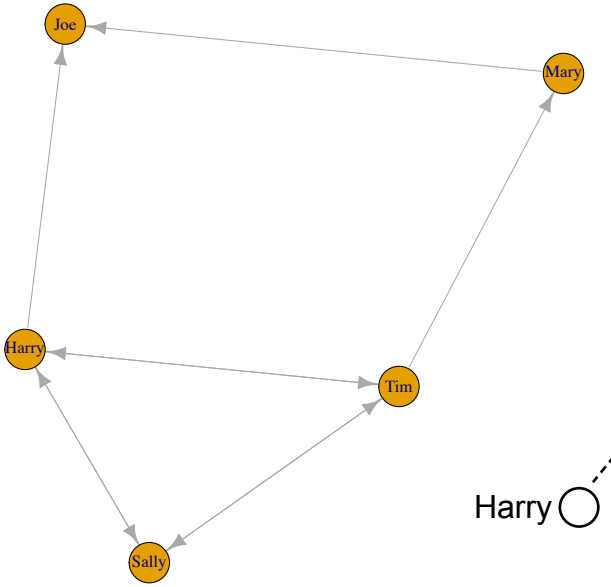


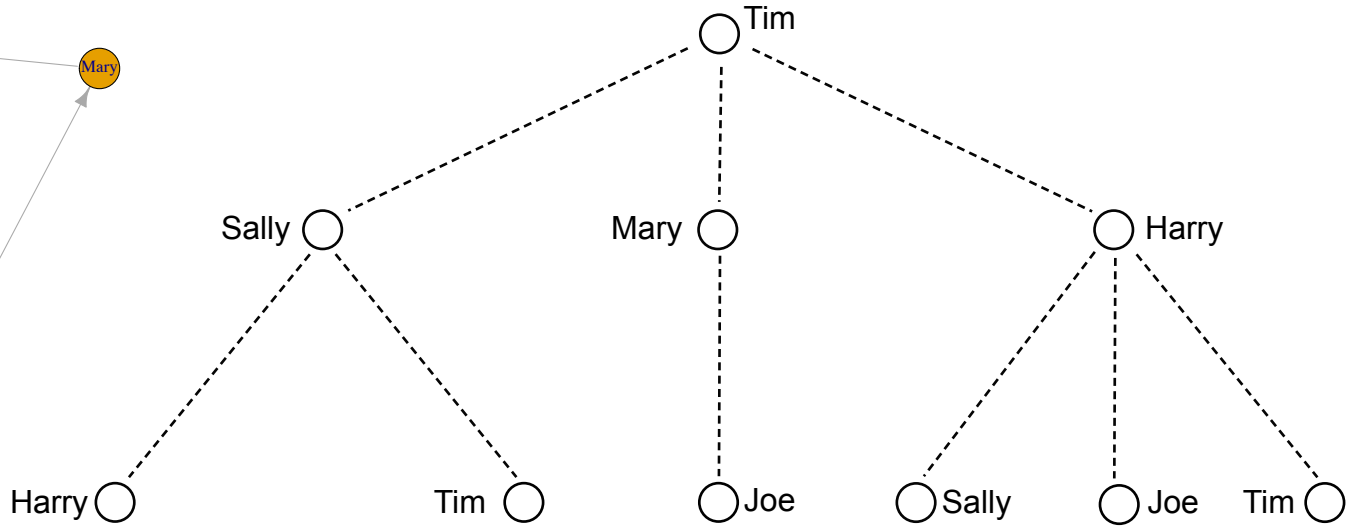
Walks

- Focus is on indirect connections between actors
  - What makes a network a social system
  - Important for diffusion
- Walks: sequence of nodes and edges that connect i to j
  - can go over same node more than once
  - can over same edge more than once
- Calculate if i and j are connected (indirectly) and how far

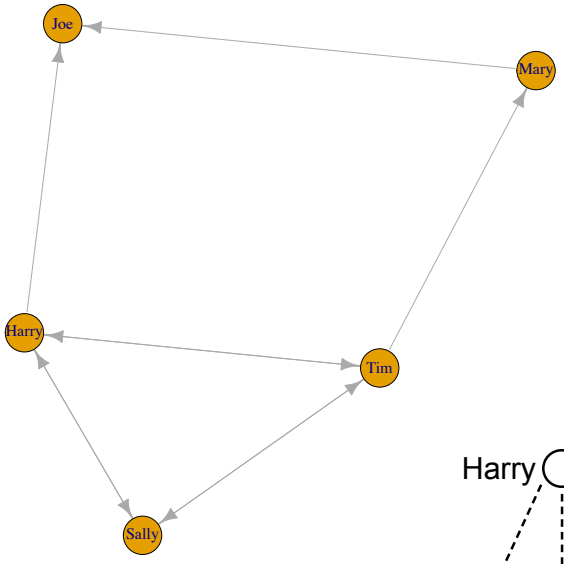

Joe
Mary
Harry
Tim
Sally

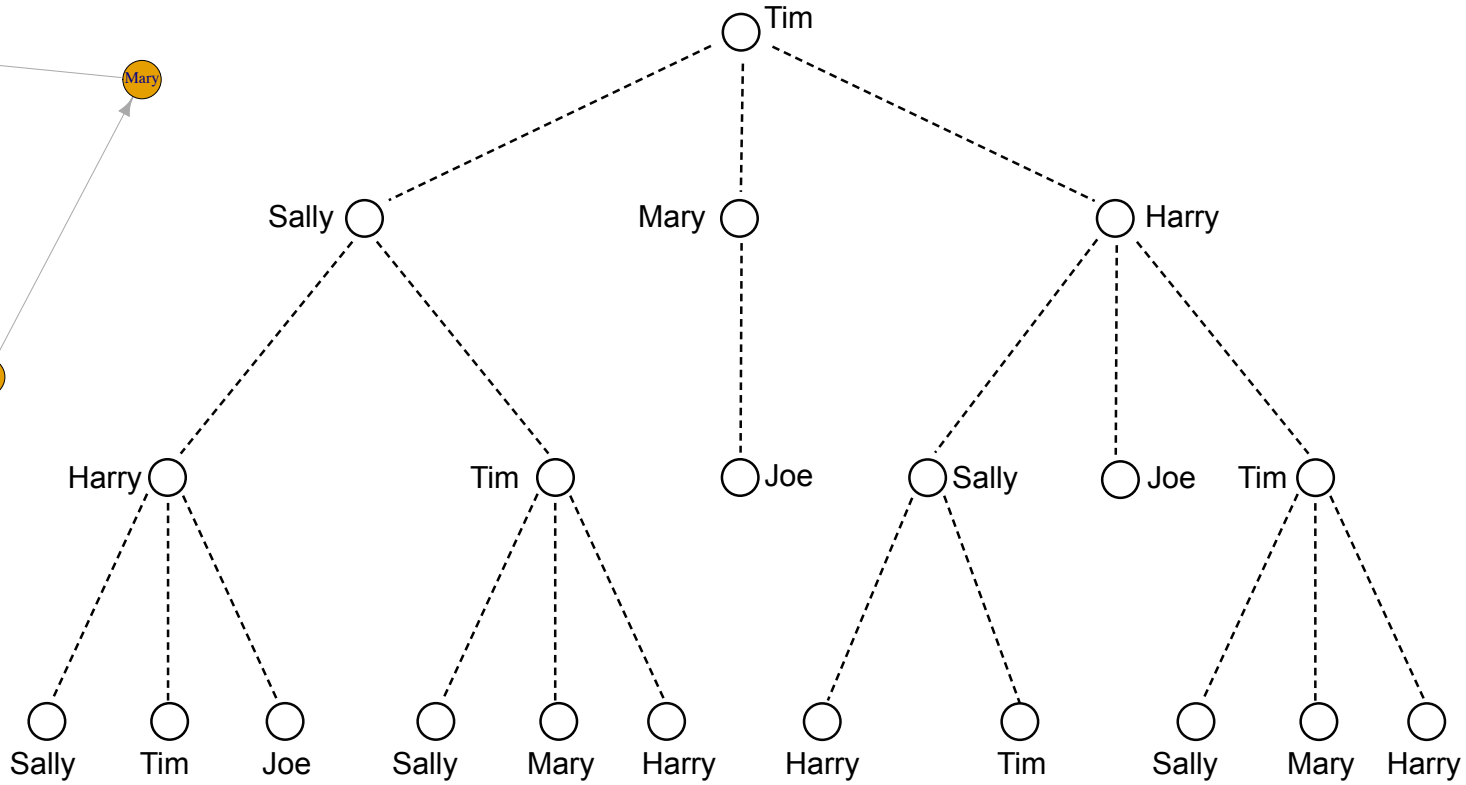
Walks of length 2, starting from Tim





Walks of length 3, starting from Tim





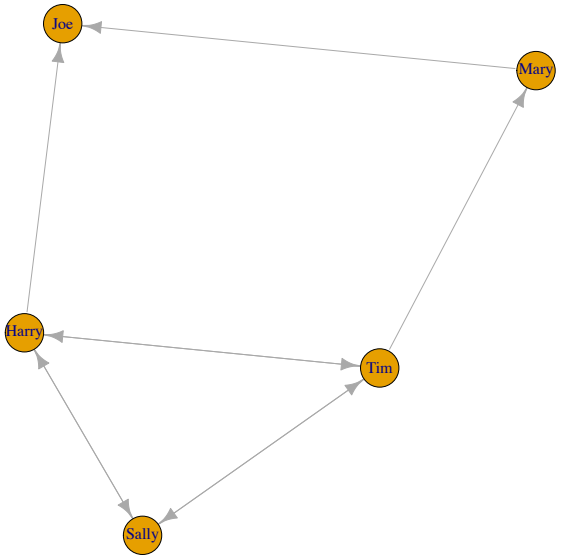
Walks continued

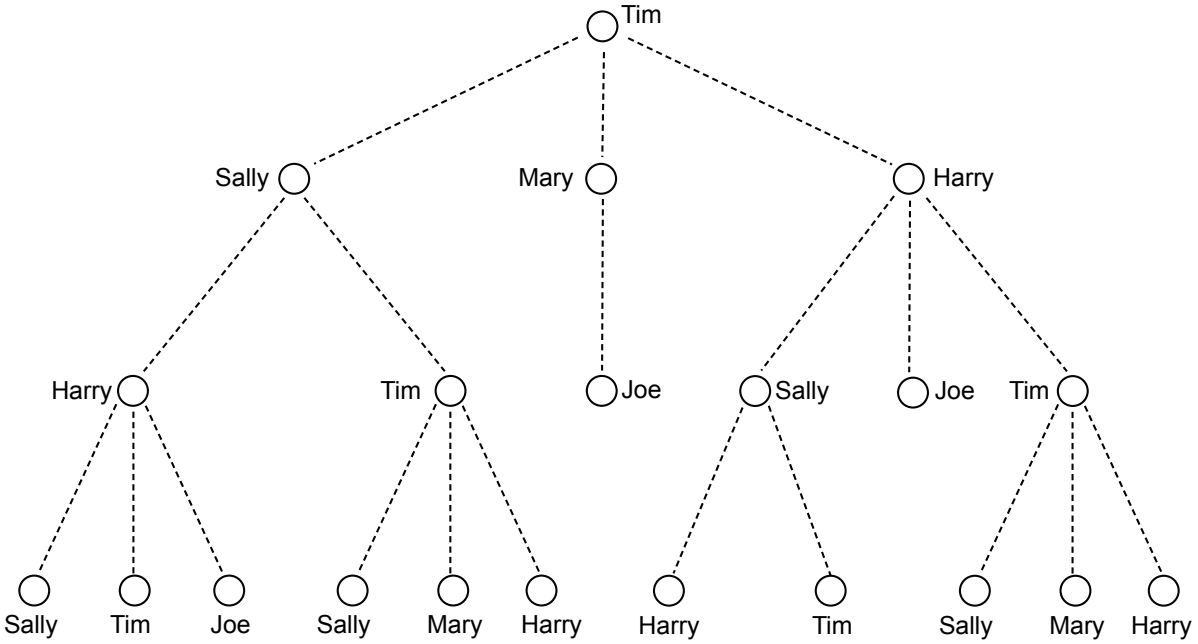
- Calculate number of walks between i and j of length p by multiplying adjacency matrix by itself
- X^2 tells us how many walks of length 2 there are between all i, j pairs
- X^p tells us how many walks of length p there are between all i, j pairs

Matrix Multiplication in One Slide

$$AB = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} e & f \\ g & h \end{bmatrix} = \begin{bmatrix} ae + bg & af + bh \\ ce + dg & cf + dh \end{bmatrix}$$

Number of Walks of Length 3



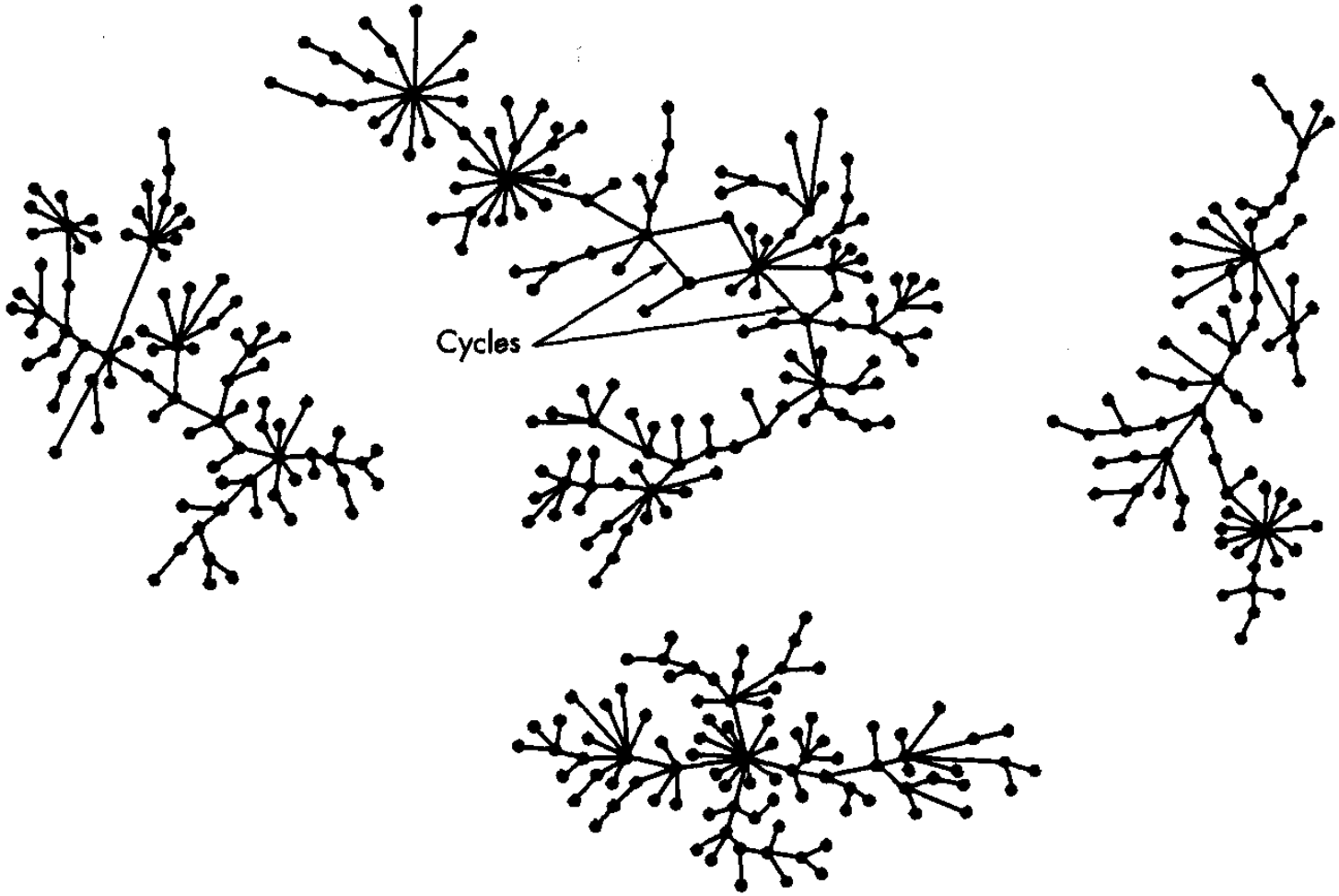


	Tim	Joe	Sally	Harry	Mary
Tim	2	1	3	3	2
Joe	0	0	0	0	0
Sally	3	2	2	3	1
Harry	3	3	3	2	1
Mary	0	0	0	0	0

Reachability and Components

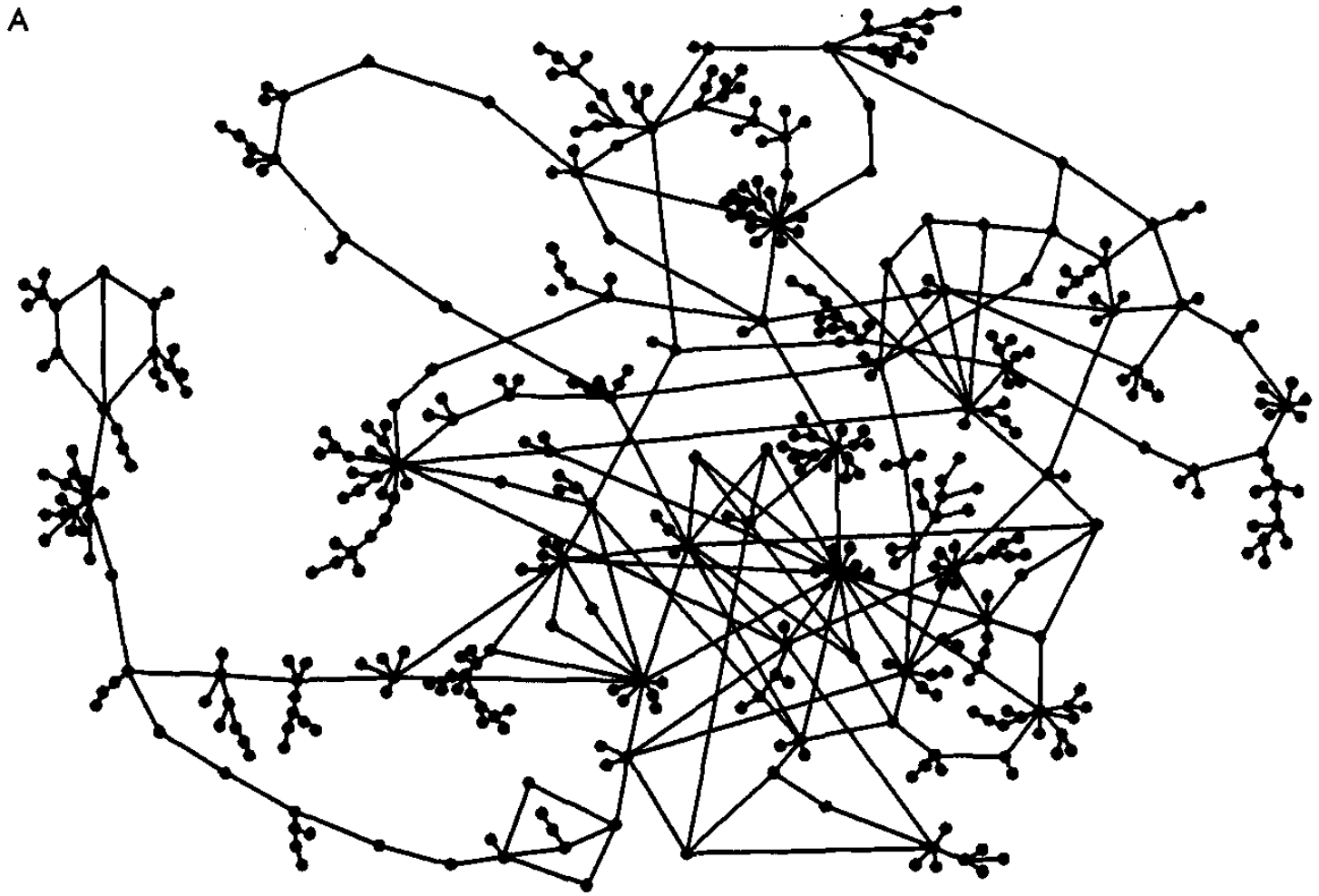
- Node i and j are said to be reachable if there is at least one sequence of edges that connects them
  - Minimal condition for global diffusion to occur
- Component is a set of nodes where everyone can reach everyone else

Lower Diffusion Potential



Source: Potterat, Muth, Rothenberg, et. al. 2002. *Sex. Trans. Infect* 78:152-158

Higher Diffusion Potential

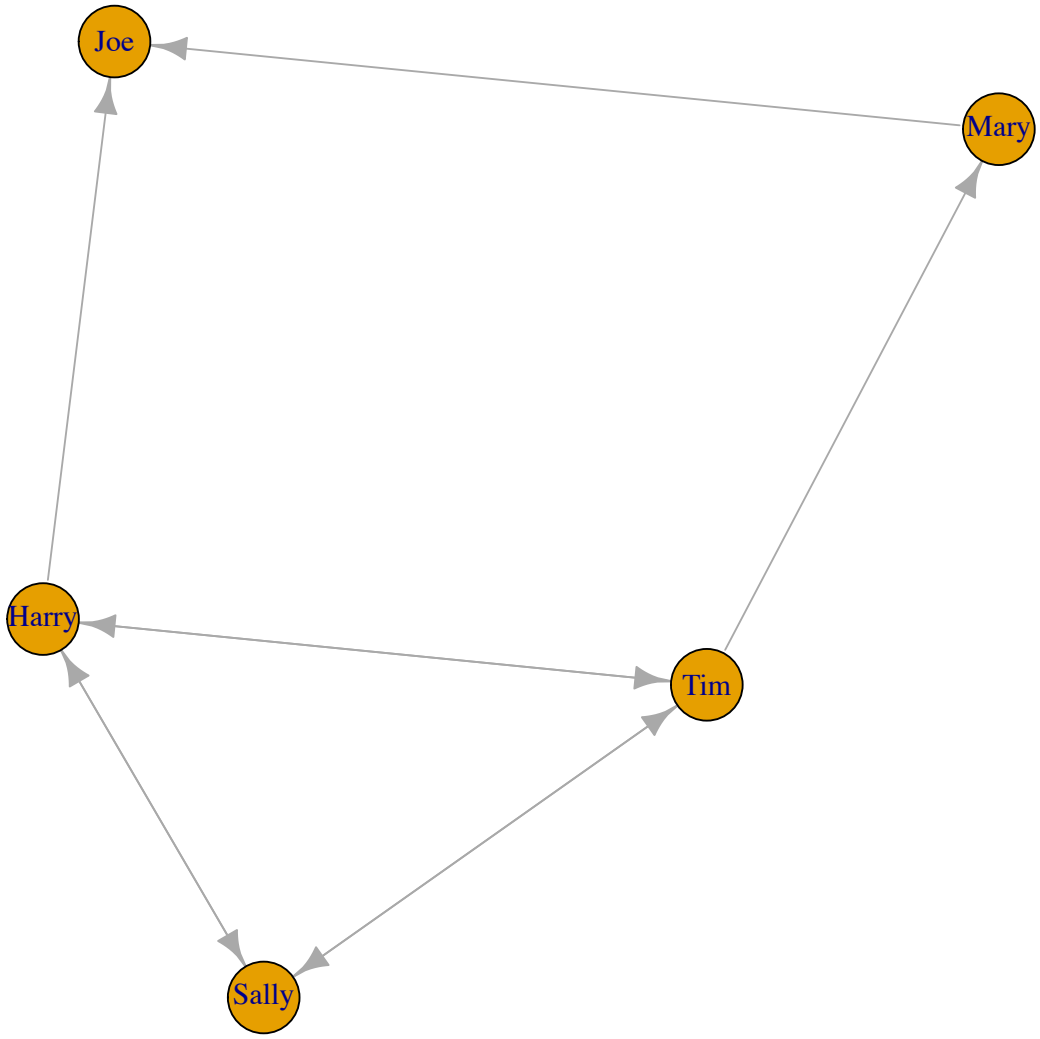


Source: Potterat, Muth, Rothenberg, et. al. 2002. *Sex. Trans. Infect* 78:152-158

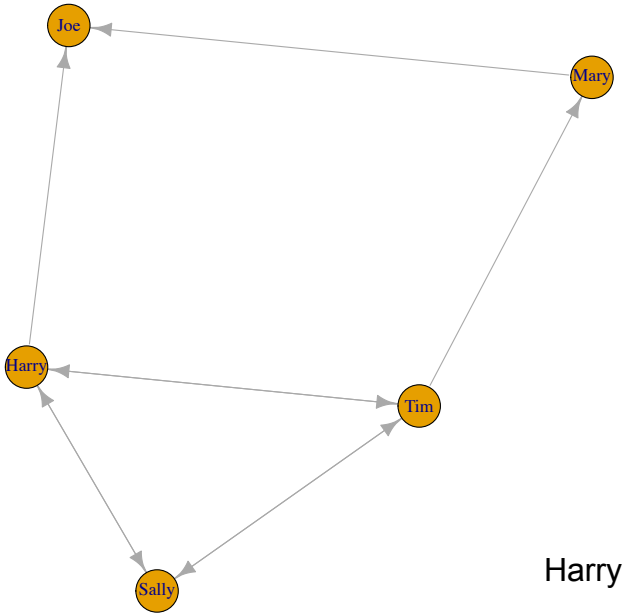
Distance

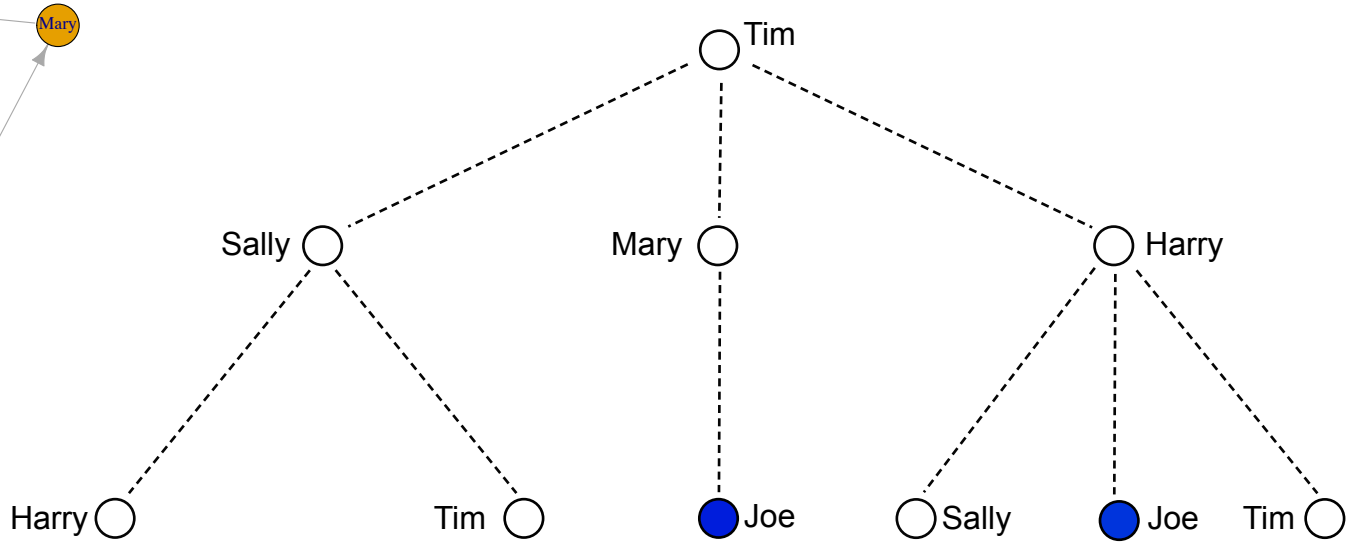
- Distance (D_{ij}) is the shortest path between i and j
  - With paths, not allowed to go over same node more than once
  - Not allowed to go over same edge more than once

Finding Distance between Tim and Joe?

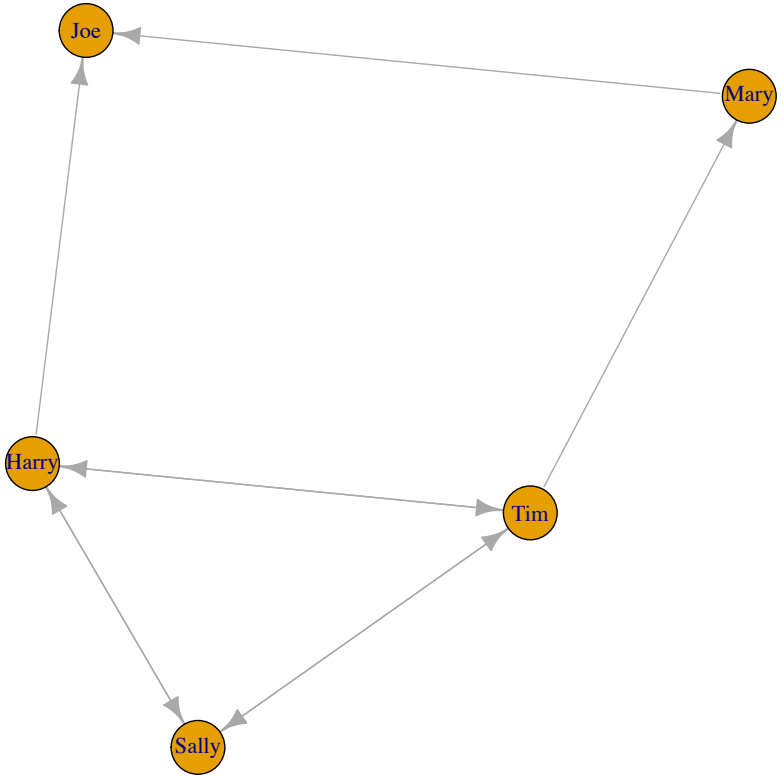


Finding Distance between Tim and Joe Using Breadth First Search (BFS)





Distance Matrix



	Tim	Joe	Sally	Harry	Mary
Tim	NA	2	1	1	1
Joe	Inf	NA	Inf	Inf	Inf
Sally	1	2	NA	1	2
Harry	1	1	1	NA	2
Mary	Inf	1	Inf	Inf	NA