Social Network Analysis

Paulo Serôdio

University of Essex
CSO, Sciences Po

July 2022

Content

- Introduction to Core Social Network Concepts
  - Overview of the field and the tools
  - Mathematical foundations
  - SNA Data & Survey Design
  - Centrality
  - Social Capital
  - Cohesion
  - Subgroups
  - Equivalence (Role & Position)
  - Hypotheses testing
- Introduction to network analysis in UCINET

History

Previous instructors: Steve Borgatti (Kentucky) & Rich DeJordy (Fresno State)

Structure

- **Monday** : Introduction, Fundamentals & Software
  - Algebra
  - Graph theory
  - Network data
  - Intro to UCINET
- **Tuesday** : Centrality, Centralization, Cohesion
- **Wednesday** : Local Neighbourhood & Ego-networks
- **Thursday** : Communities & subgroups
- **Friday** : Testing Hypotheses, Stochastic Models & Optional Topic

Objectives

- Build intuition
- Expose key concepts
- Highlight big questions
- provide abstract examples
- Some pointers to other studies
- *NOT* a substitute for technical work

Introduction

- Name
- Affiliation
- Discipline
- SNA Experience/Knowledge
- Phenomena of interest

What Defines SNA?

- Phenomenon studied
  - distinctive type of data
- Perspective taken
  - Perhaps one perspective, but multiple theories
- Methodological toolkit
  - new concepts, new tools

Reasoning about Networks

- What can achieve from studying networks?
  - Patterns and statistical properties of network data;
  - Design principles and models;
  - Understand the organisation of networks;
- How can we reason about networks?
  - **Empirical** : study data; measure and quantify;
  - **Mathematical** Models: graph theory & stats, distinguish surprising from expected phenomena
  - **Algorithms** : for hard computational challenges

how mathematicians reason about networks

- Mathematicians are concerned with the abstract structure of a graph
- Mathematicians define operations to analyze and manipulate graphs. Moreover, they develop theorems based upon structural axioms.

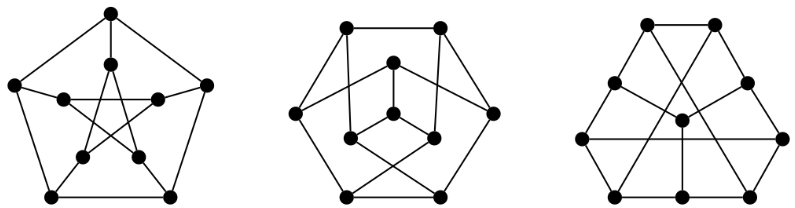
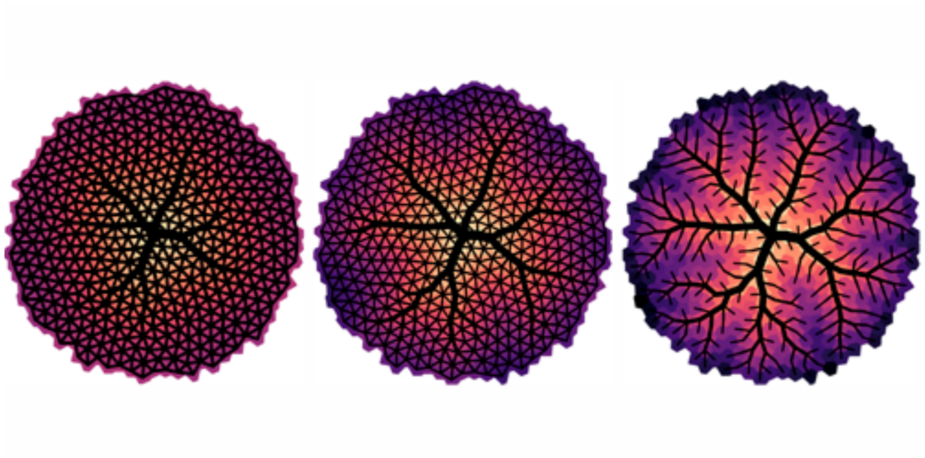


Figure 0.7: Three isomorphic drawings of the infamous Petersen graph!

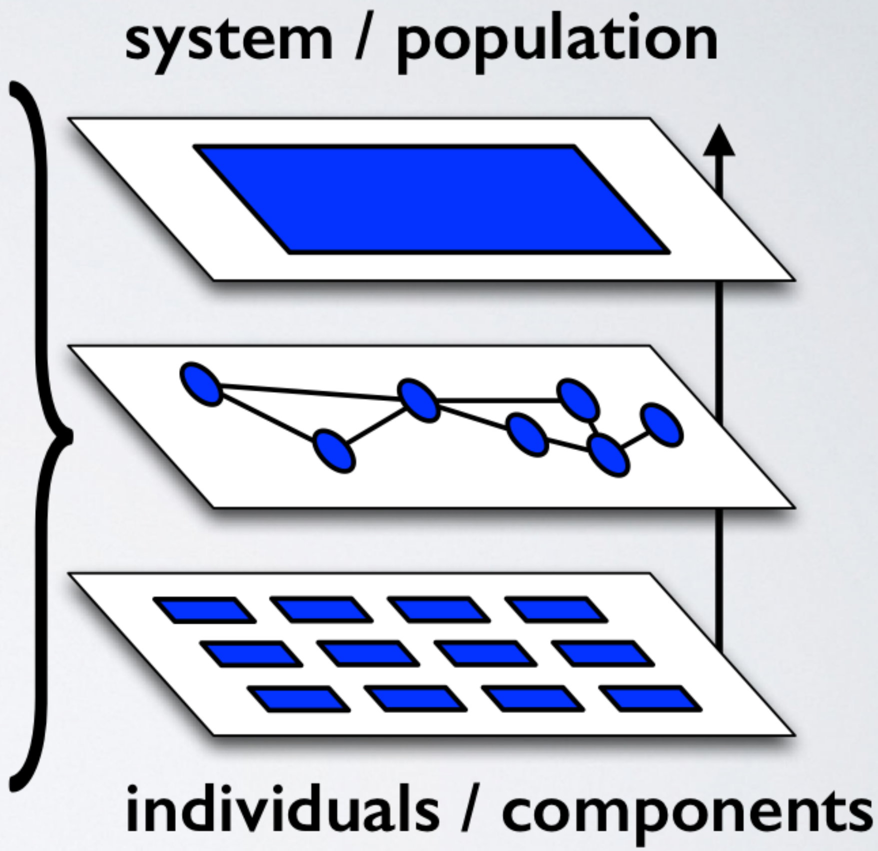
how physicists reason about networks

- Physicists are concerned with modeling real-world structures with networks.
- Physicists define algorithms that compress the information in a network to more simple values (e.g. statistical analysis).



what are networks?

- an approach
- a mathematical representation
- provide structure to complexity
- *structure above* individuals / components
- *structure below* system / population



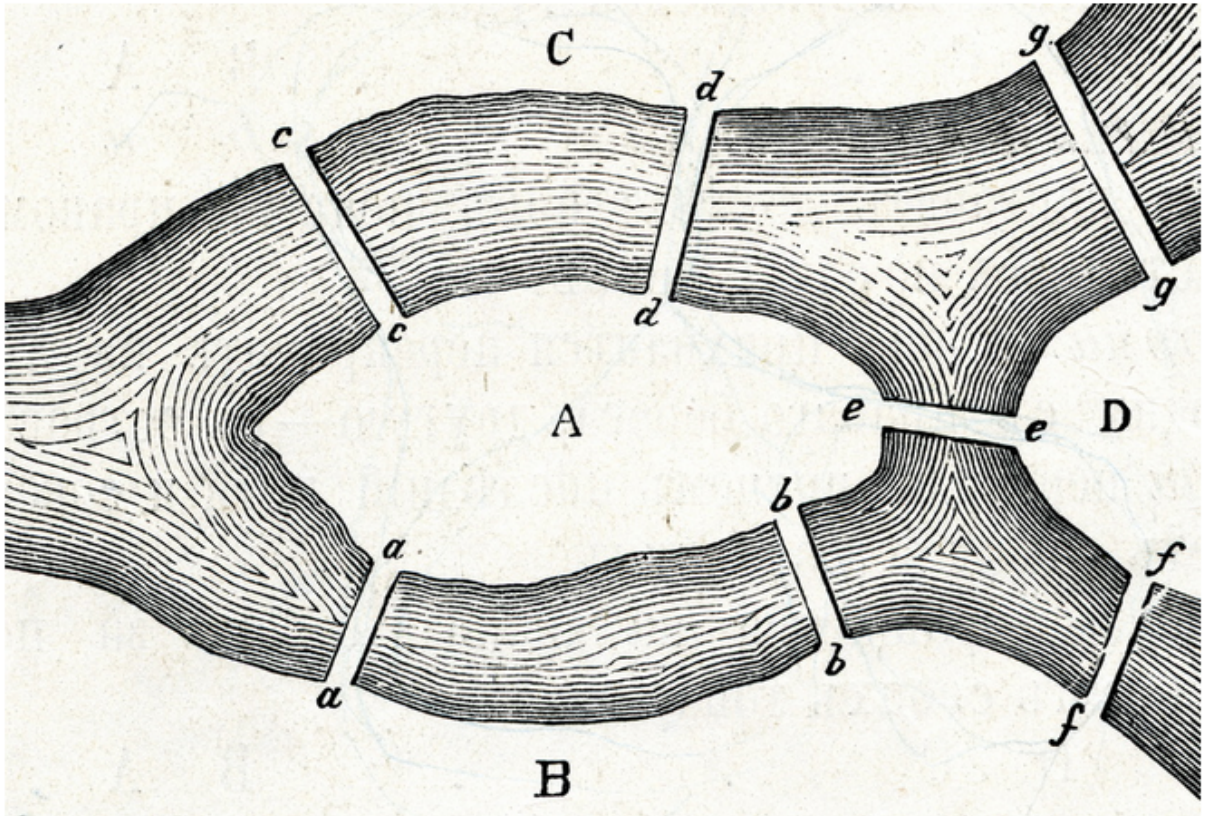
History of SNA

- 1736- Euler
- 1930s- Sociometry
- 1940s Psychologists
- 1950s & 60s Anthropologists
- 1970s Rise of Sociologists
  - Small Worlds, Strength of weak ties
- 1980s IBM computation
  - Computer programs developed
- 1990s Ideas spread
  - UCINET released, spread of network analyis to multiple fields, social capital, embedded ties
- 2000s Physicists jump on the bandwagon

Graph Theory Beginnings: Leonard Euler

- Swiss mathematician and logician (1707 - 1783)
- Network analysis begins with solution to the “Bridges of Königsberg” question in 1735

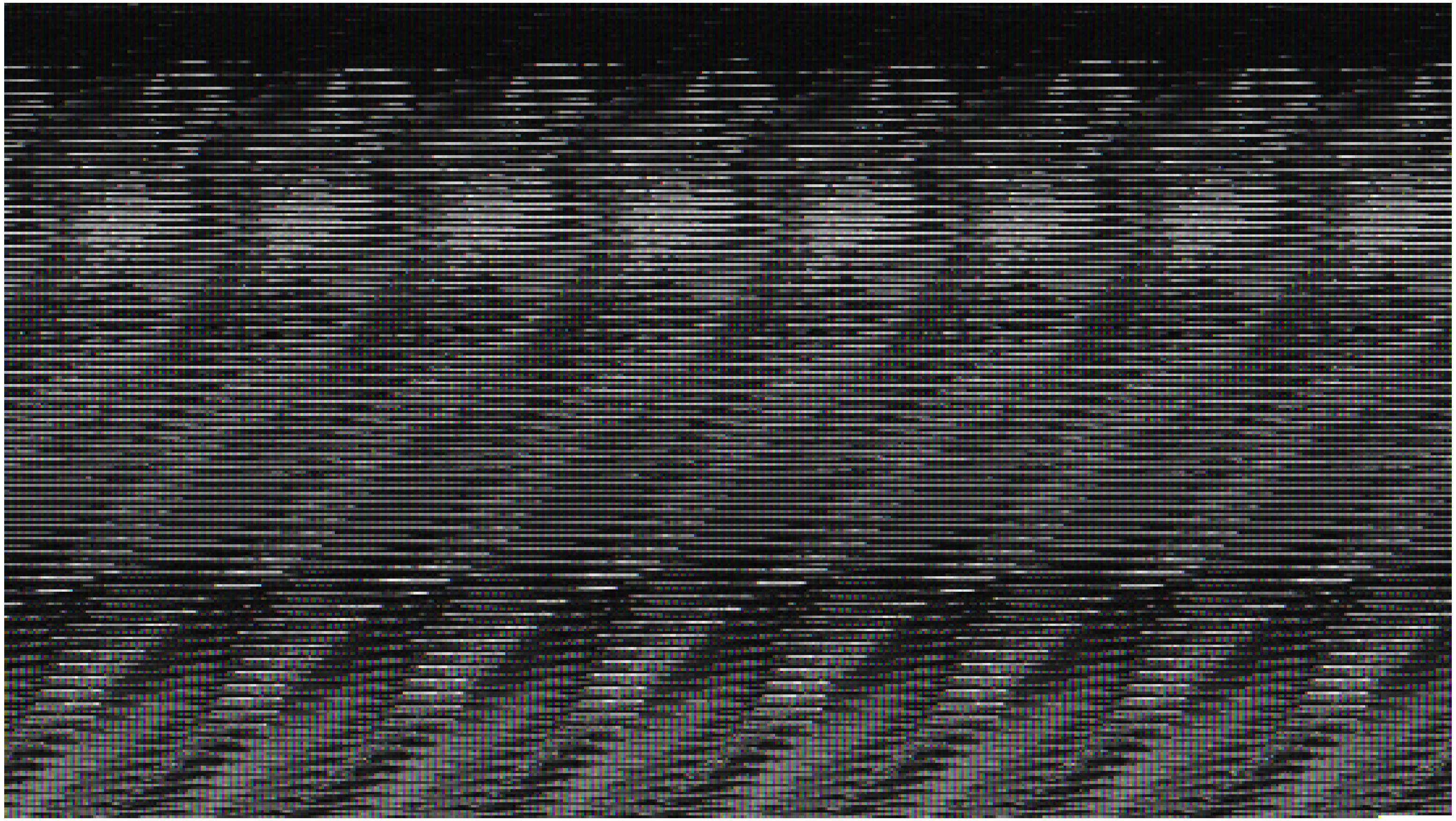
The Seven Bridges of Königsberg



Big Question: Can one walk across all seven bridges and never cross the same one twice?

Definition: an Euler path walks through a graph without revisiting edges; an Euler circuit is an Euler path that starts and stops at the same vertex.

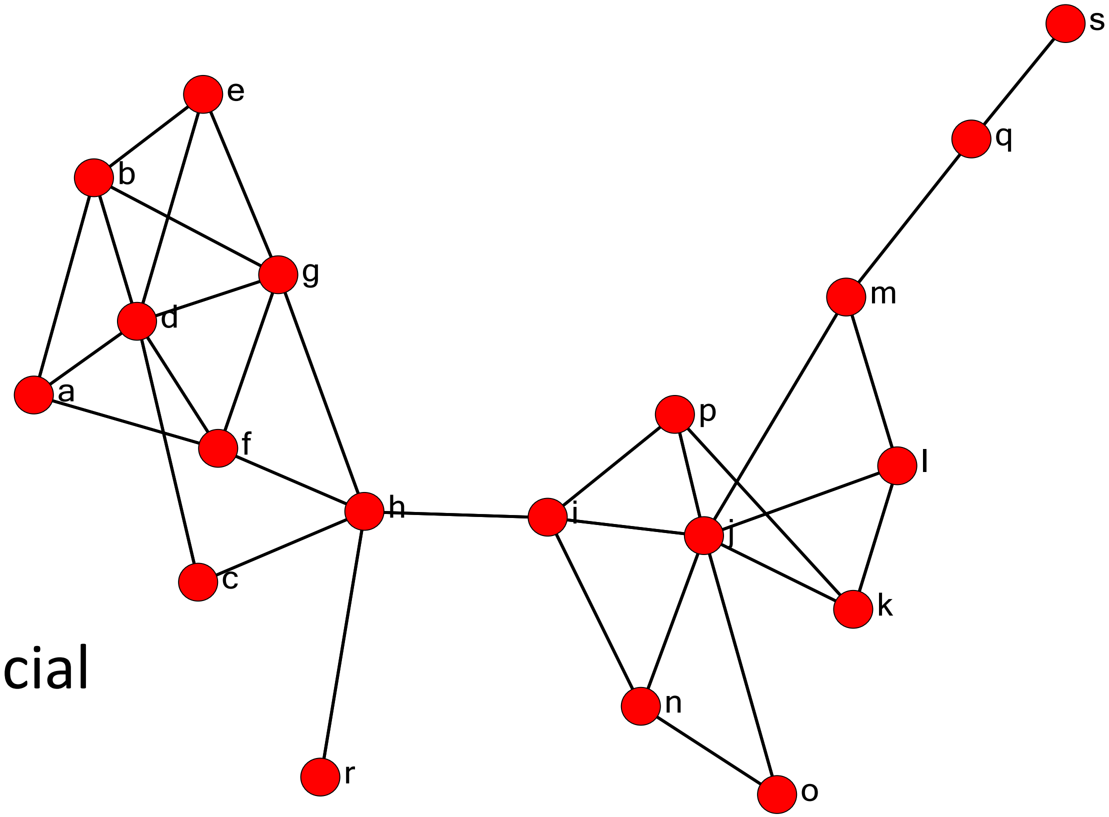
Euler theorem: if a graph has an Euler circuit, then every **vertex** has even degree.



What is a network?

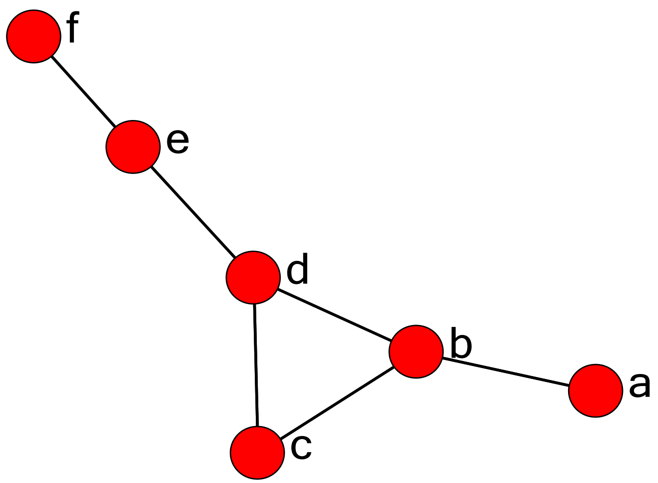
Network

- Set of nodes
- Set of ties among them
- Ties interlink through common nodes
  - Resulting in paths
- In social network analysis, ties typically represent a social relation
  - E.g., kinship, family



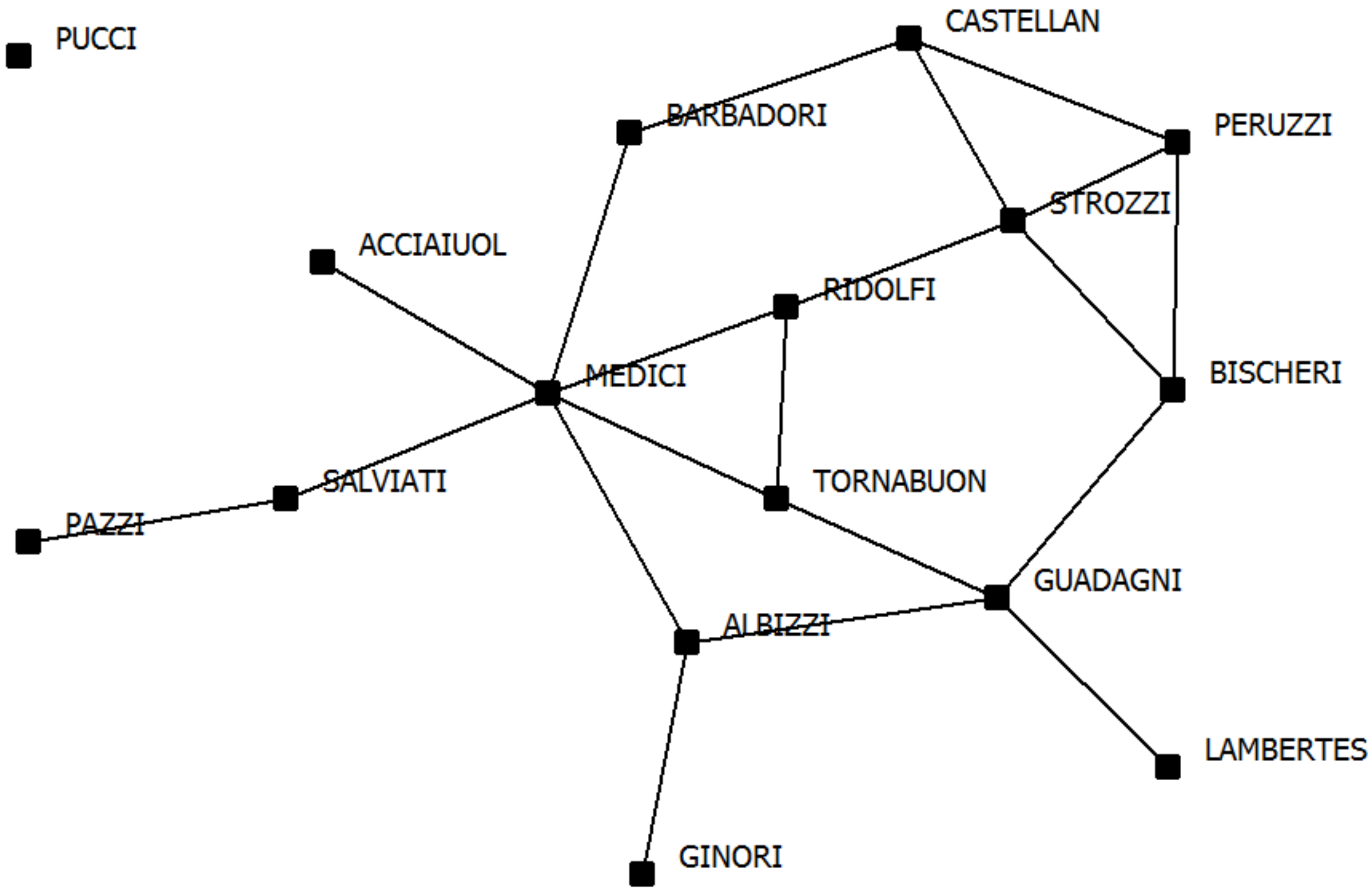
Adjacency matrix

- Can represent a network as a node-by-node matrix
    - Typically 1s and 0s, could be strengths of tie



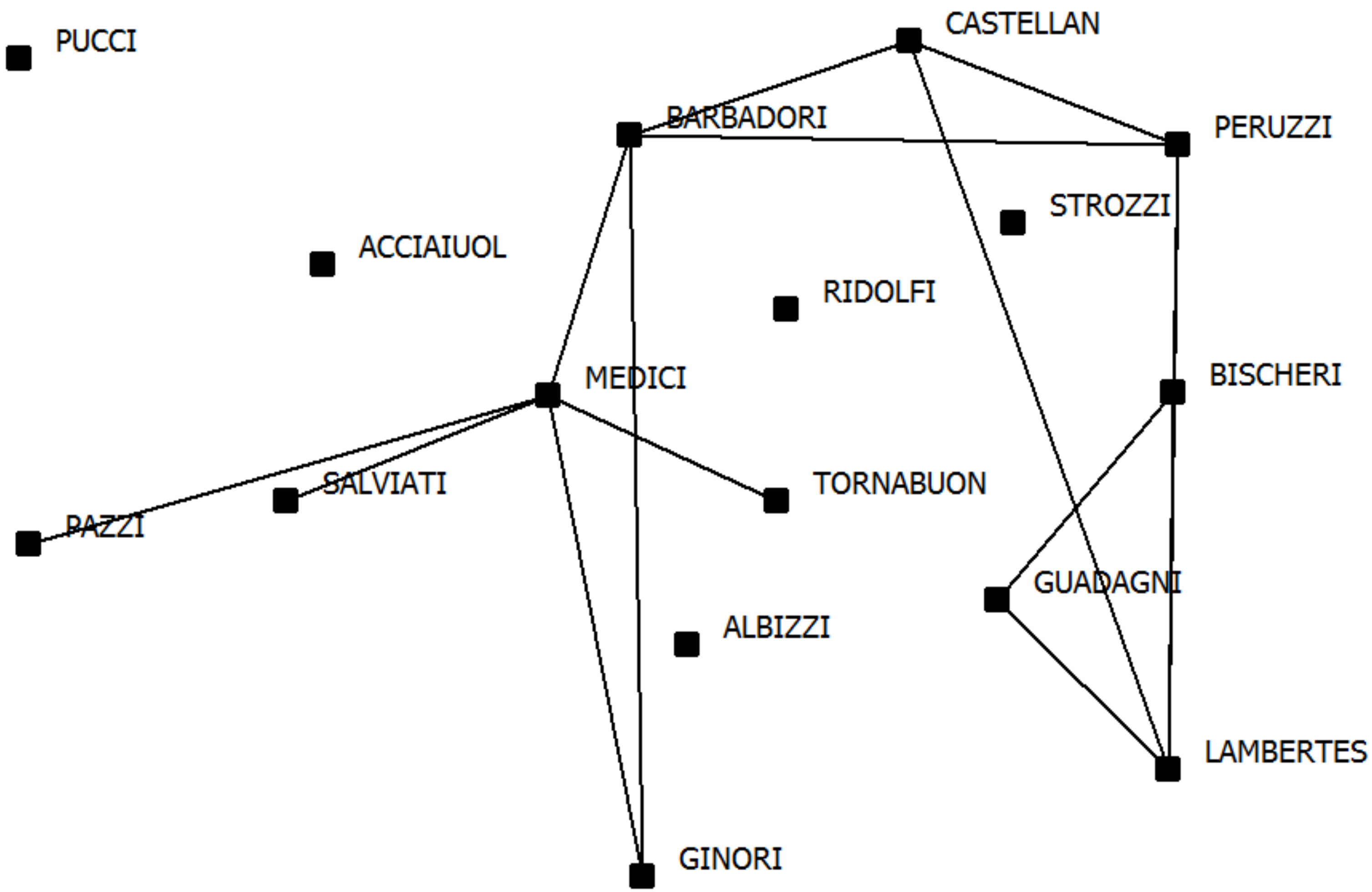
	a	b	c	d	e	f
a		1	0	0	0	0
b	1		1	1	0	0
c	0	1		1	0	0
d	0	1	1		1	0
e	0	0	0	1		1
f	0	0	0	0	1	

Marriage ties between families



Padgett & Ansell (1991). Marriage ties among Florentine families during the Renaissance

Business ties between families

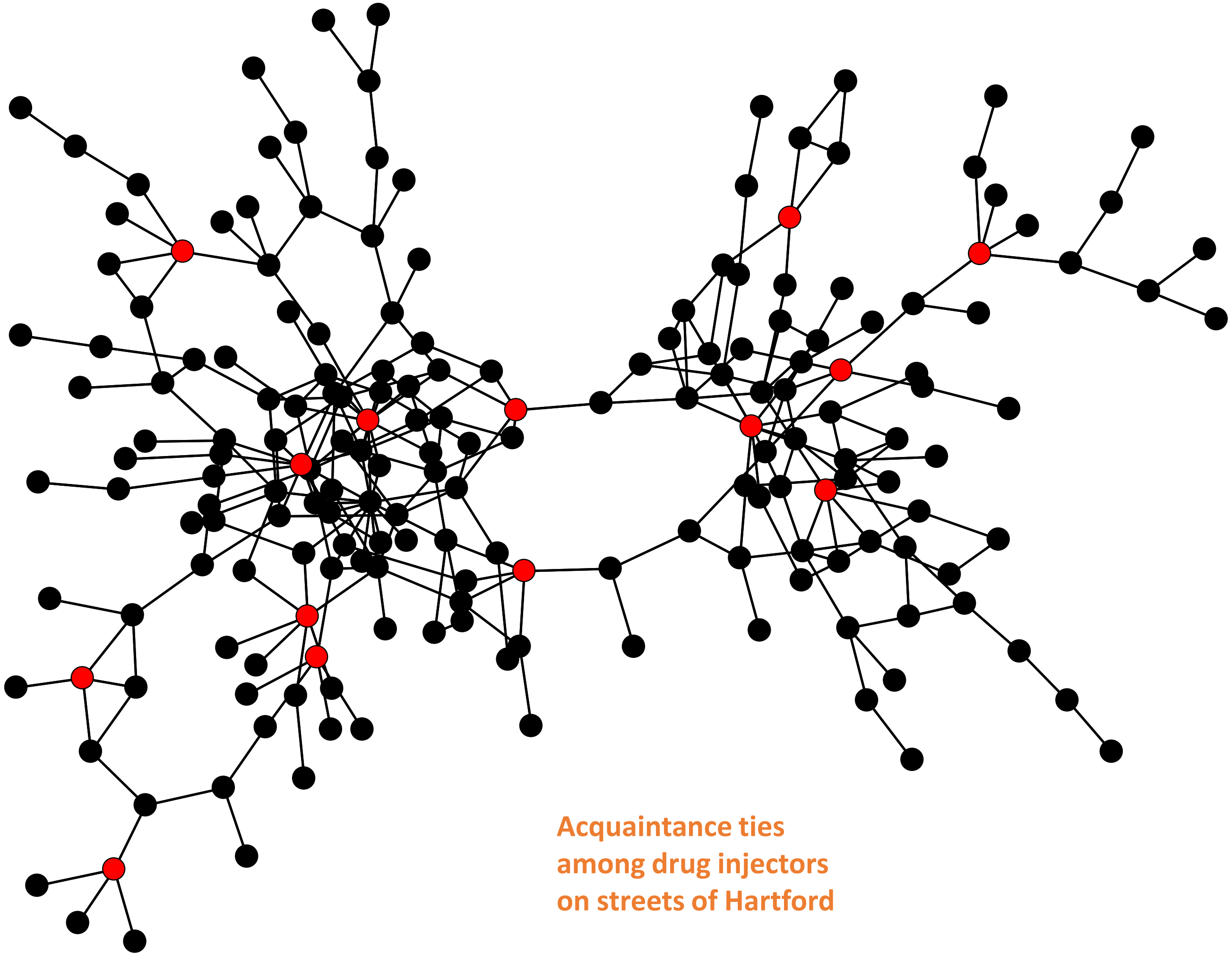


Dyadic variables

- A given type of relation, such as marriage, can be seen as a dyadic variable that describes the relationship between every pair of nodes
- A dyadic variable assigns a value to each pair of nodes

Dyad	Married	Business
ACCIAIUOLI-GUADAGNI	0	0
GUADAGNI-STROZZI	0	0
PUCCI-STROZZI	0	0
BISCHERI-SALVIATI	0	0
ACCIAIUOLI-GINORI	0	0
GUADAGNI-RIDOLFI	0	0
MEDICI-TORNABUONI	1	1
CASTELLANI-SALVIATI	0	0
BARBADORI-GUADAGNI	0	0
CASTELLANI-LAMBERTESCHI	0	1
ACCIAIUOLI-ALBIZZI	0	0
GUADAGNI-PUCCI	0	0
LAMBERTESCHI-STROZZI	0	0
MEDICI-PUCCI	0	0

Acquaintance network



Acquaintance ties
among drug injectors
on streets of Hartford

Is this a network?

Book Co-purchasing

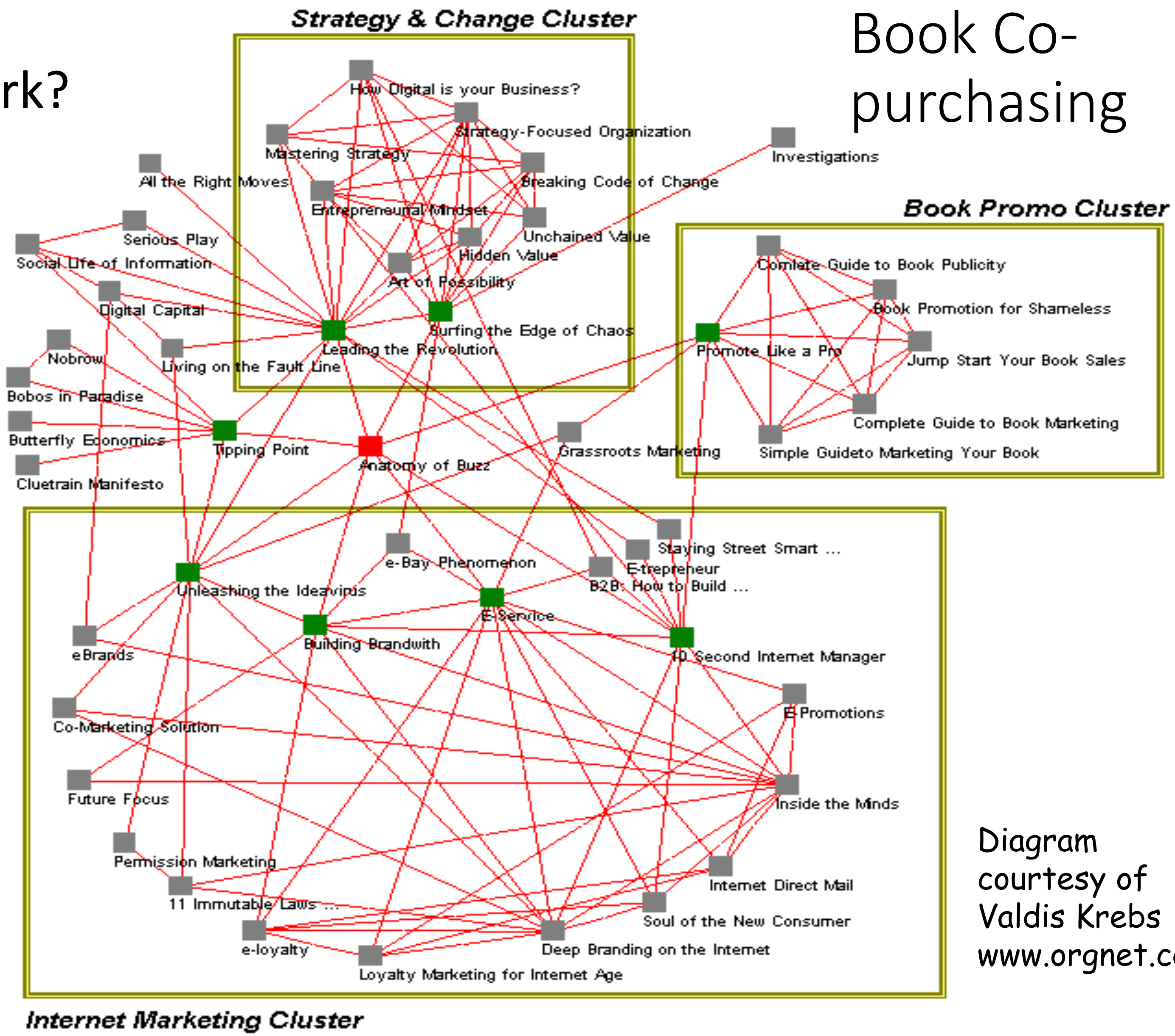
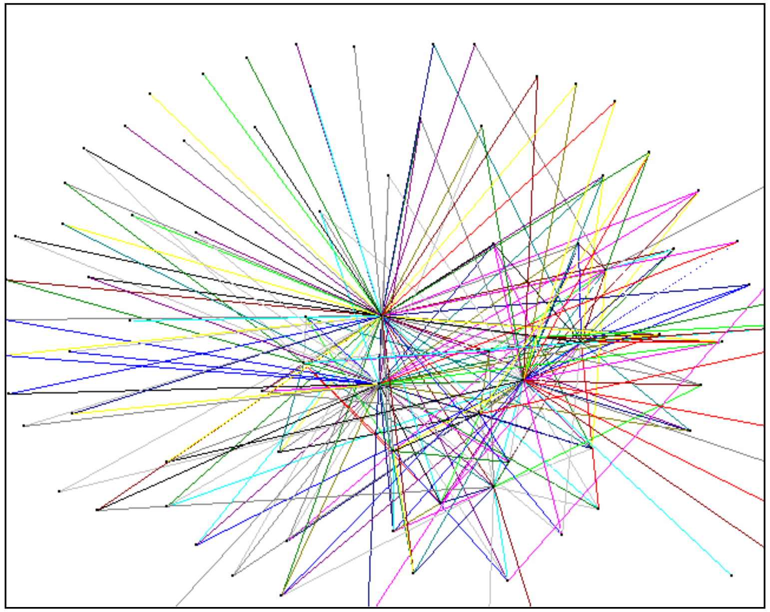


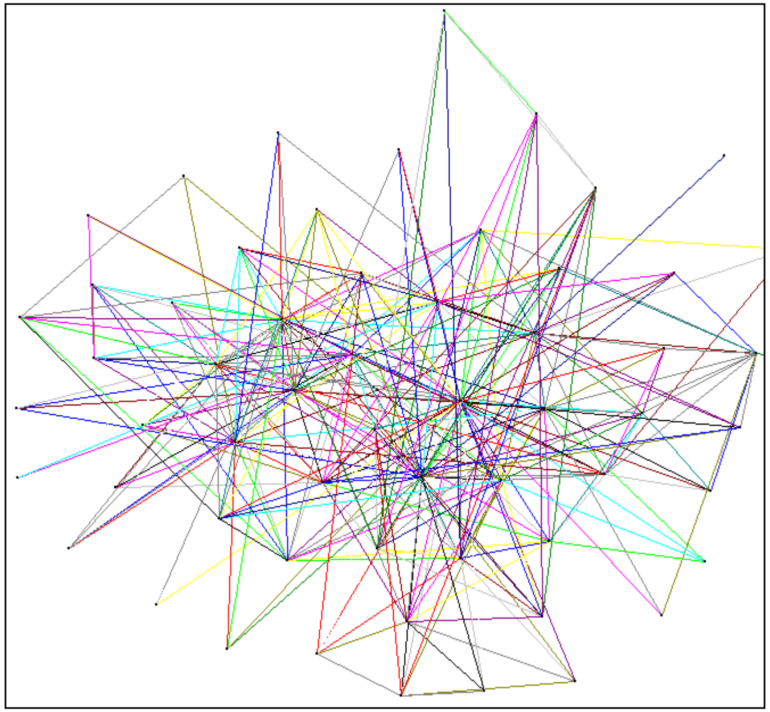
Diagram courtesy of Valdis Krebs www.orgnet.com

Comparing airlines' route structures

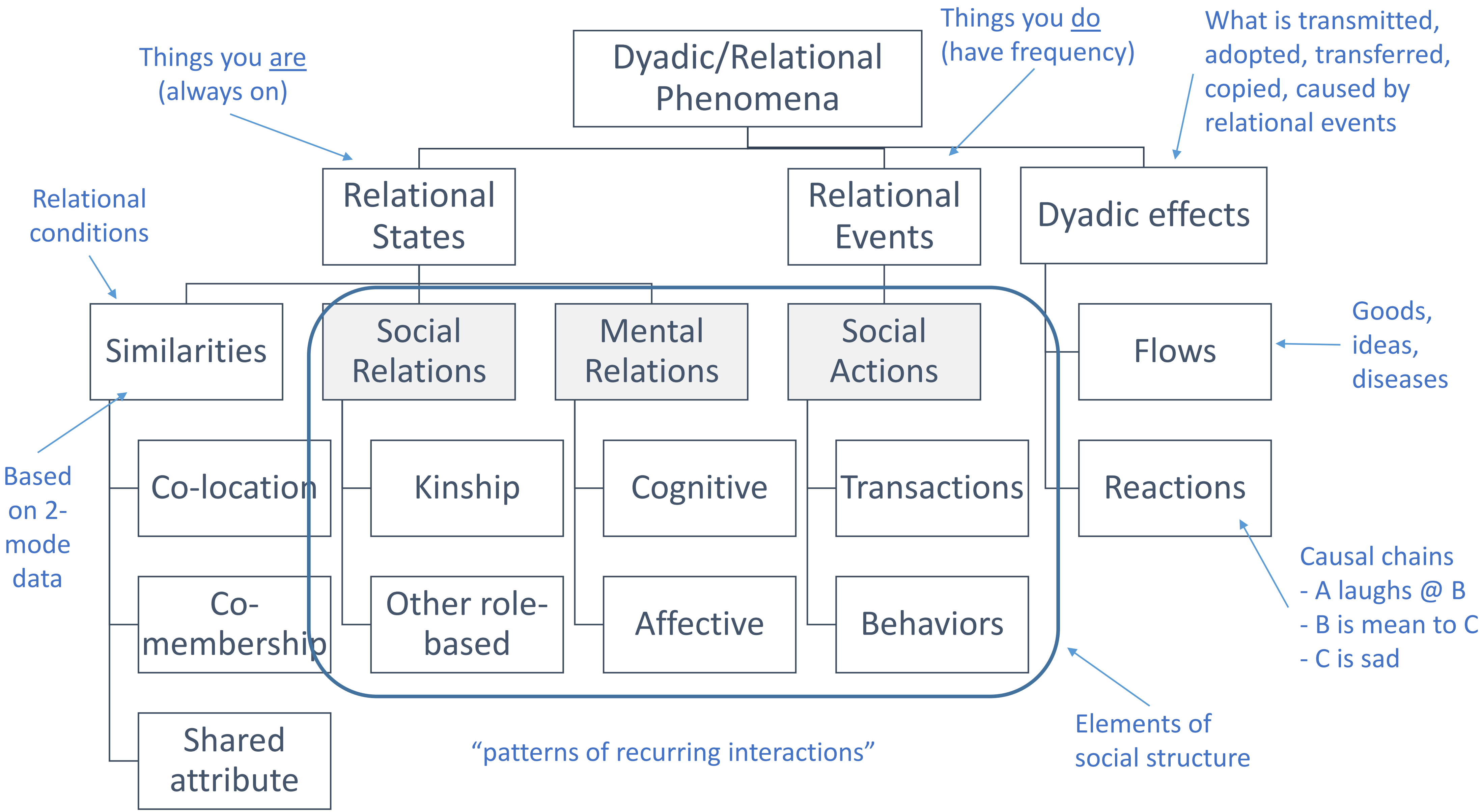
Major Carrier



"Discount" Airline



Note: Route maps defined around one specific hub only
Source: Industry data, BCG analysis


Dyadic/Relational Phenomena
Things you are (always on)
Things you do (have frequency)
What is transmitted, adopted, transferred, copied, caused by relational events
Relational States
Relational Events
Dyadic effects
Relational conditions
Similarities
Social Relations
Mental Relations
Social Actions
Flows
Goods, ideas, diseases
Based on 2-mode data
Co-location
Kinship
Cognitive
Transactions
Reactions
Causal chains
- A laughs @ B
- B is mean to C
- C is sad
Co-membership
Other role-based
Affective
Behaviors
Shared attribute
"patterns of recurring interactions"
Elements of social structure

Entailed interactions

- Friendship carries with it certain norms about how the friends will behave toward each other
  - Rights and obligations
  - Expectations
- Kinship ties have these too
- Professor / student
- So this means that a given “base relation” entails a variety of interactions
  - And base relations also have a variety of different functions, e.g., material aid, emotional support, advice, etc.

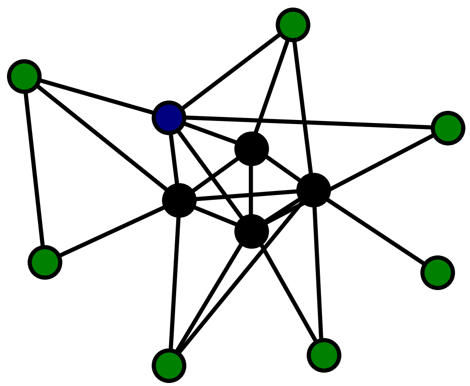
Multiplexity

- A given dyad (pair of persons) can be connected by more than one kind of base relation at the same time
  - E.g., both kin and co-worker
- I wouldn't classify being friends and talking often as multiplex
  - Because the base relation entails the talking

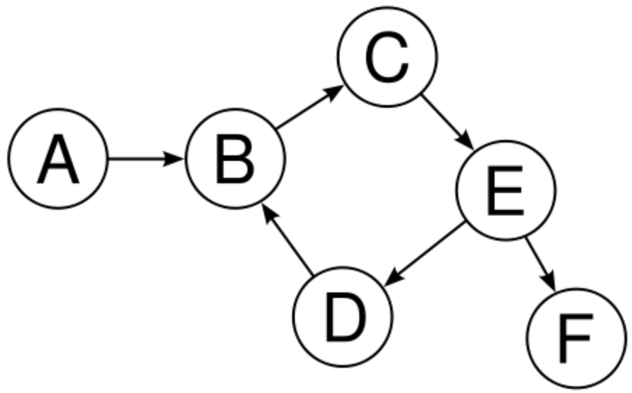
Multiplex relationship → MEDICI-TORNABUONI

Dyad	Married	Business
ACCIAIUOLI-GUADAGNI	0	0
GUADAGNI-STROZZI	0	0
PUCCI-STROZZI	0	0
BISCHERI-SALVIATI	0	0
ACCIAIUOLI-GINORI	0	0
GUADAGNI-RIDOLFI	0	0
MEDICI-TORNABUONI	1	1
CASTELLANI-SALVIATI	0	0
BARBADORI-GUADAGNI	0	0
CASTELLANI-LAMBERTESCHI	0	1
ACCIAIUOLI-ALBIZZI	0	0
GUADAGNI-PUCCI	0	0
LAMBERTESCHI-STROZZI	0	0
MEDICI-PUCCI	0	0

Directed and undirected

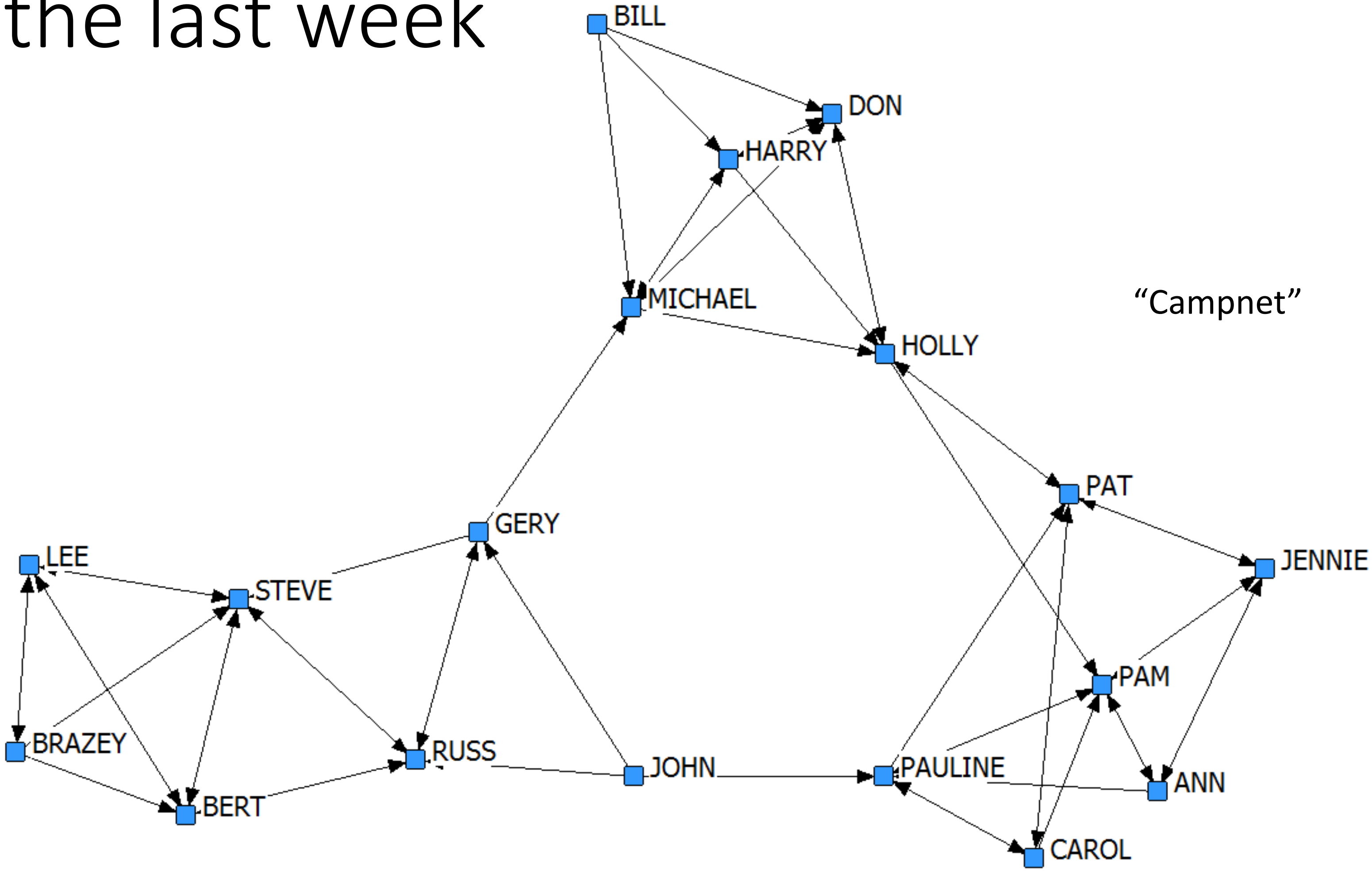


undirected



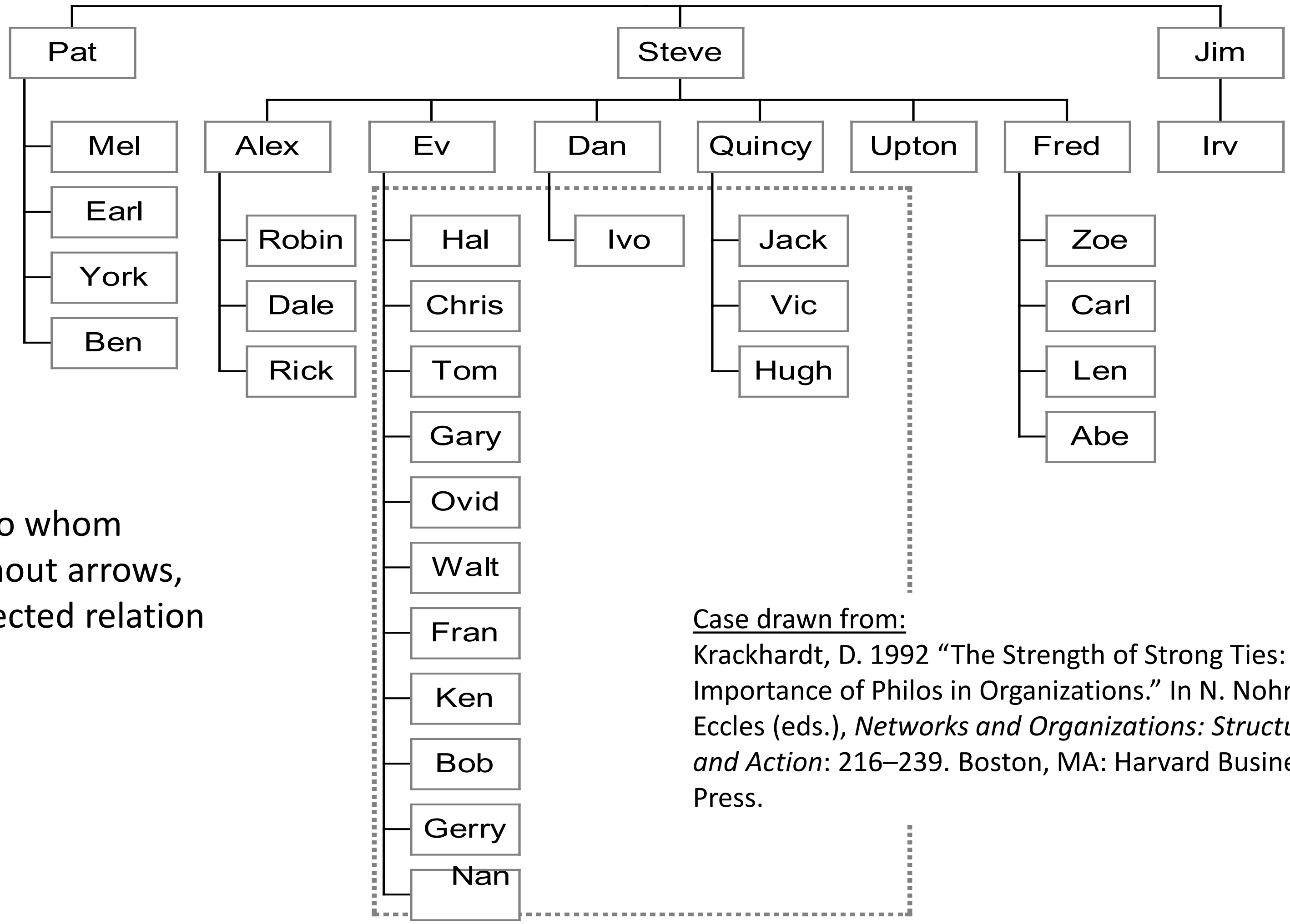
directed

The 3 people you interacted with the most over the last week



"Campnet"

Organization chart



Who reports to whom

- Drawn without arrows, but is a directed relation

<u>Case drawn from:</u>
Krackhardt, D. 1992 "The Strength of Strong Ties: The Importance of Philos in Organizations." In N. Nohria & R. Eccles (eds.), *Networks and Organizations: Structure, Form, and Action*: 216–239. Boston, MA: Harvard Business School Press.

2-mode data: who attended what event

Names of Participants of Group I	Code Numbers and Dates of Social Events Reported in *Old City Herald*													
	(1) 6/27	(2) 3/2	(3) 4/12	(4) 9/26	(5) 2/25	(6) 5/19	(7) 3/15	(8) 9/16	(9) 4/8	(10) 6/10	(11) 2/23	(12) 4/7	(13) 11/21	(14) 8/3
1. Mrs. Evelyn Jefferson	×	×	×	×	×	×		×	×					
2. Miss Laura Mandeville	×	×	×		×	×	×	×						
3. Miss Theresa Anderson		×	×	×	×	×	×	×	×					
4. Miss Brenda Rogers	×		×	×	×	×	×	×						
5. Miss Charlotte McDowd			×	×	×		×							
6. Miss Frances Anderson			×		×	×		×						
7. Miss Eleanor Nye					×	×	×	×						
8. Miss Pearl Oglethorpe						×		×	×					
9. Miss Ruth DeSand					×		×	×	×					
10. Miss Verne Sanderson							×	×	×			×		
11. Miss Myra Liddell								×	×	×		×		
12. Miss Katherine Rogers								×	×	×		×	×	×
13. Mrs. Sylvia Avondale							×	×	×	×		×	×	×
14. Mrs. Nora Fayette						×	×		×	×	×	×	×	×
15. Mrs. Helen Lloyd							×	×		×	×	×		
16. Mrs. Dorothy Murchison								×	×					
17. Mrs. Olivia Carleton									×		×			
18. Mrs. Flora Price									×		×			

Figure 1. Davis, Gardner and Gardner (1941) *Deep South* women-by-events matrix.

Co-participation data

	E1	E2	E3	E4	E5	E6	E7	E8	E9	E10	E11	E12	E13	E14
EVELYN	1	1	1	1	1	1	0	1	1	0	0	0	0	0
LAURA	1	1	1	0	1	1	1	1	0	0	0	0	0	0
THERESA	0	1	1	1	1	1	1	1	1	0	0	0	0	0
BRENDA	1	0	1	1	1	1	1	1	0	0	0	0	0	0
CHARLOTTE	0	0	1	1	1	0	1	0	0	0	0	0	0	0
FRANCES	0	0	1	0	1	1	0	1	0	0	0	0	0	0
ELEANOR	0	0	0	0	1	1	1	1	0	0	0	0	0	0
PEARL	0	0	0	0	0	1	0	1	1	0	0	0	0	0
RUTH	0	0	0	0	1	0	1	1	1	0	0	0	0	0
VERNE	0	0	0	0	0	0	1	1	1	0	0	1	0	0
MYRNA	0	0	0	0	0	0	0	1	1	1	0	1	0	0
KATHERINE	0	0	0	0	0	0	0	1	1	1	0	1	1	1
SYLVIA	0	0	0	0	0	0	1	1	1	1	0	1	1	1
NORA	0	0	0	0	0	1	1	0	1	1	1	1	1	1
HELEN	0	0	0	0	0	0	1	1	0	1	1	1	0	0
DOROTHY	0	0	0	0	0	0	0	1	1	0	0	0	0	0
OLIVIA	0	0	0	0	0	0	0	0	1	0	1	0	0	0
FLORA	0	0	0	0	0	0	0	0	1	0	1	0	0	0

2-mode “affiliations” data
person by event

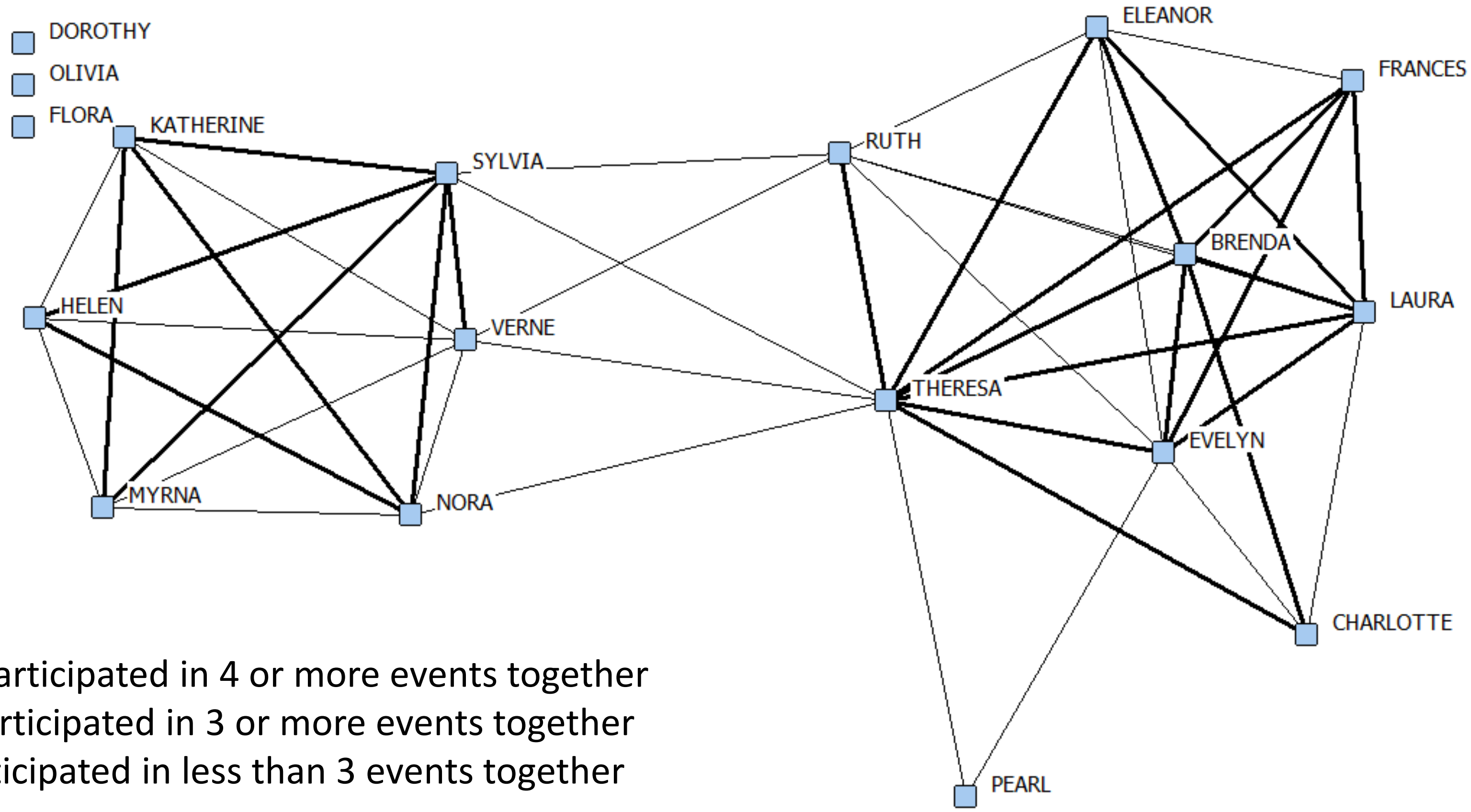
Valued adjacency matrix

	EVE	LAU	THE	BRE	CHA	FRA	ELE	PEA	RUT	VER	MYR	KAT	SYL	NOR	HEL	DOR	OLI	FLO
EVELYN	8	6	7	6	3	4	3	3	3	2	2	2	2	2	1	2	1	1
LAURA	6	7	6	6	3	4	4	2	3	2	1	1	2	2	2	1	0	0
THERESA	7	6	8	6	4	4	4	3	4	3	2	2	3	3	2	2	1	1
BRENDA	6	6	6	7	4	4	4	2	3	2	1	1	2	2	2	1	0	0
CHARLOTTE	3	3	4	4	4	2	2	0	2	1	0	0	1	1	1	0	0	0
FRANCES	4	4	4	4	2	4	3	2	2	1	1	1	1	1	1	1	0	0
ELEANOR	3	4	4	4	2	3	4	2	3	2	1	1	2	2	2	1	0	0
PEARL	3	2	3	2	0	2	2	3	2	2	2	2	2	2	1	2	1	1
RUTH	3	3	4	3	2	2	3	2	4	3	2	2	3	2	2	2	1	1
VERNE	2	2	3	2	1	1	2	2	3	4	3	3	4	3	3	2	1	1
MYRNA	2	1	2	1	0	1	1	2	2	3	4	4	4	3	3	2	1	1
KATHERINE	2	1	2	1	0	1	1	2	2	3	4	6	6	5	3	2	1	1
SYLVIA	2	2	3	2	1	1	2	2	3	4	4	6	7	6	4	2	1	1
NORA	2	2	3	2	1	1	2	2	2	3	3	5	6	8	4	1	2	2
HELEN	1	2	2	2	1	1	2	1	2	3	3	3	4	4	5	1	1	1
DOROTHY	2	1	2	1	0	1	1	2	2	2	2	2	2	1	1	2	1	1
OLIVIA	1	0	1	0	0	0	0	1	1	1	1	1	1	2	1	1	2	2
FLORA	1	0	1	0	0	0	0	1	1	1	1	1	1	2	1	1	2	2

1-mode co-attendance matrix

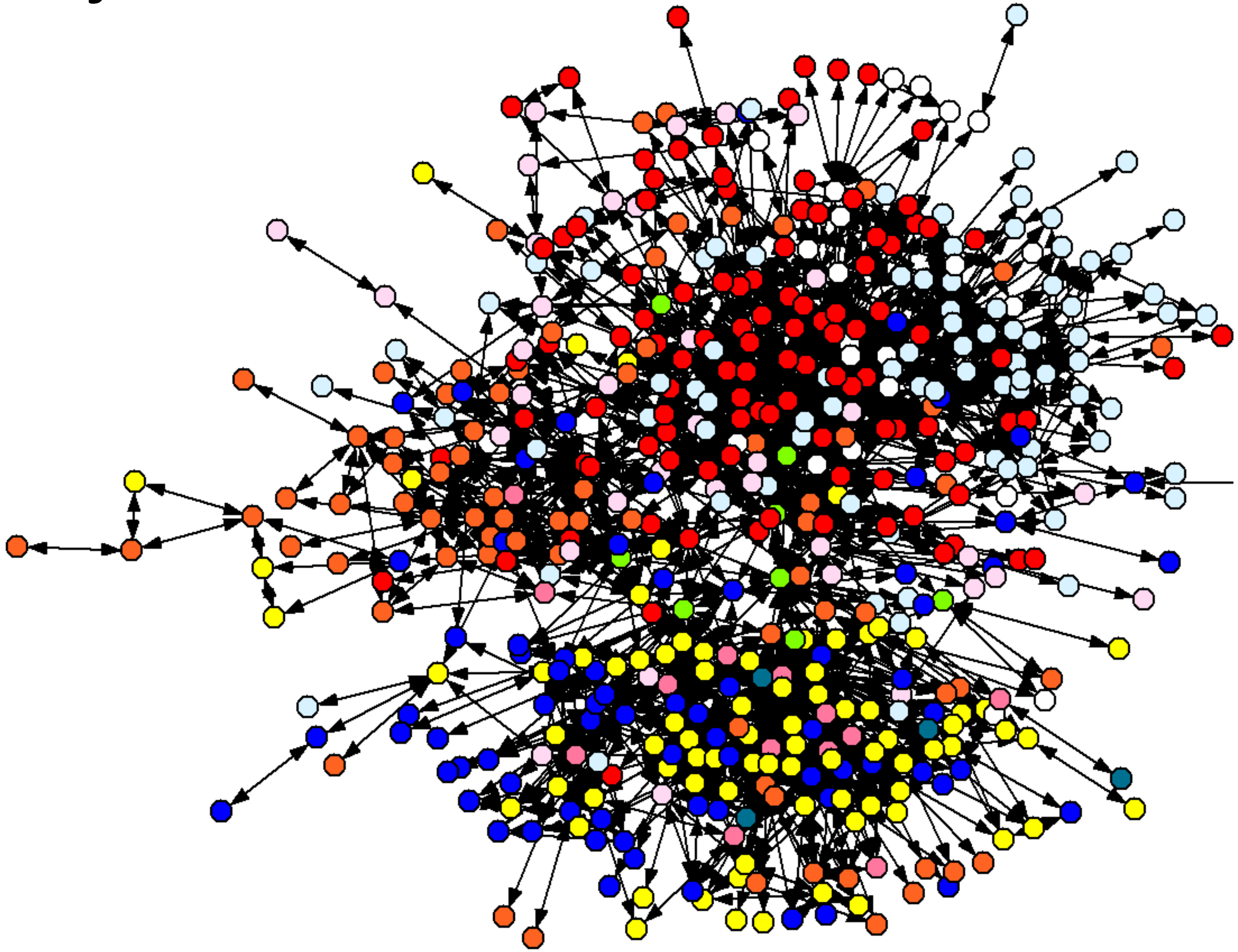
Co-participation in events

Valued network



Thick line: participated in 4 or more events together
Thin line: participated in 3 or more events together
No line: participated in less than 3 events together

Project collaboration network



Project collaboration among 960 scientists

ID	Initials	Description	Color	Symbol
1	BHS	Behavioral Sciences	Yellow	
2	CCG	Communication Analysts	Lime Green	
3	DCL	Washington at large	Red	
4	ES	Economics	Blue	
5	HEW	Health, Educ. & Welfare	Pink	
6	IS	Int'l Studies	White	
7	MS	Mgmt. Sci	Orange	
18	SRG	Survey Research	Midnight Blue	
22	STAT	Statisticians	Salmon	
23	TAS	Tech & Applied Sci	L sky blue	

Bank Wiring Room

- Hawthorne Studies
- Western Electric Plant
- 1920s & 1930s

Roethlisberger, FJ, and WJ Dickson. 1939. **Management and the Worker** (Cambridge: Harvard University Press.

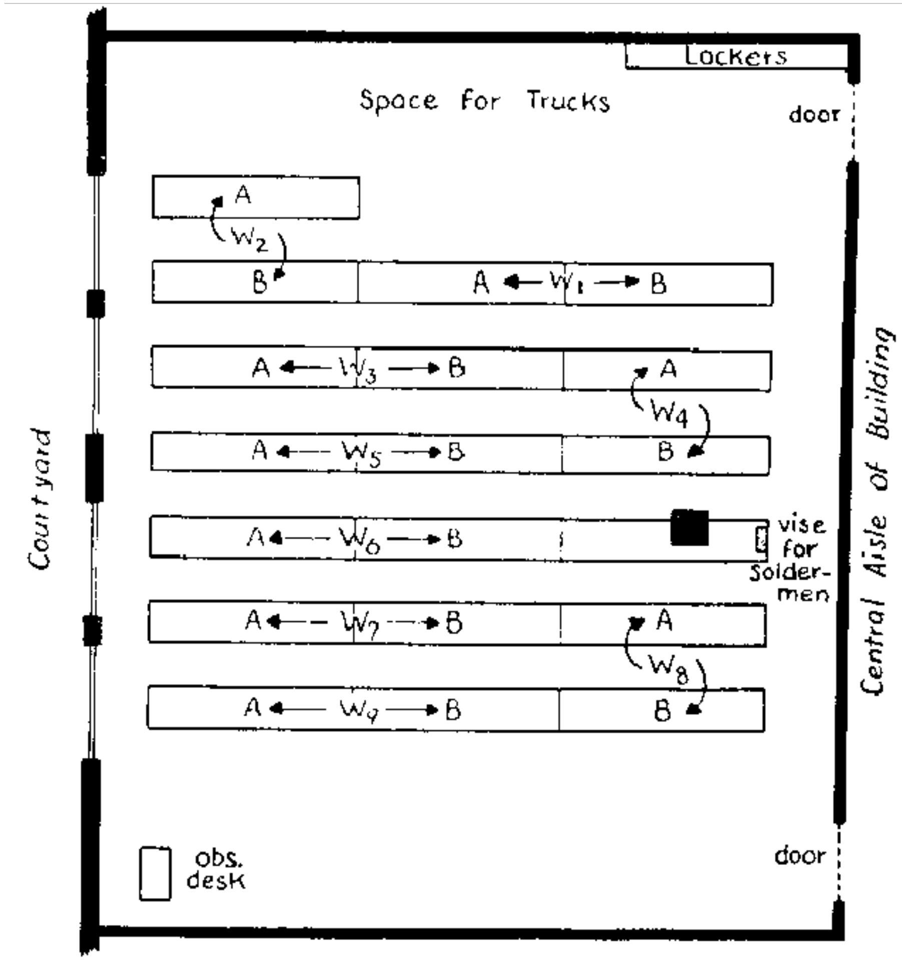


Figure 34
Diagram of Observation Room Showing Wiremen's Positions (A & B)

Game Playing Relations

Roethlisberger, FJ, and WJ Dickson. 1939.
Management and the Worker. Harvard University Press.

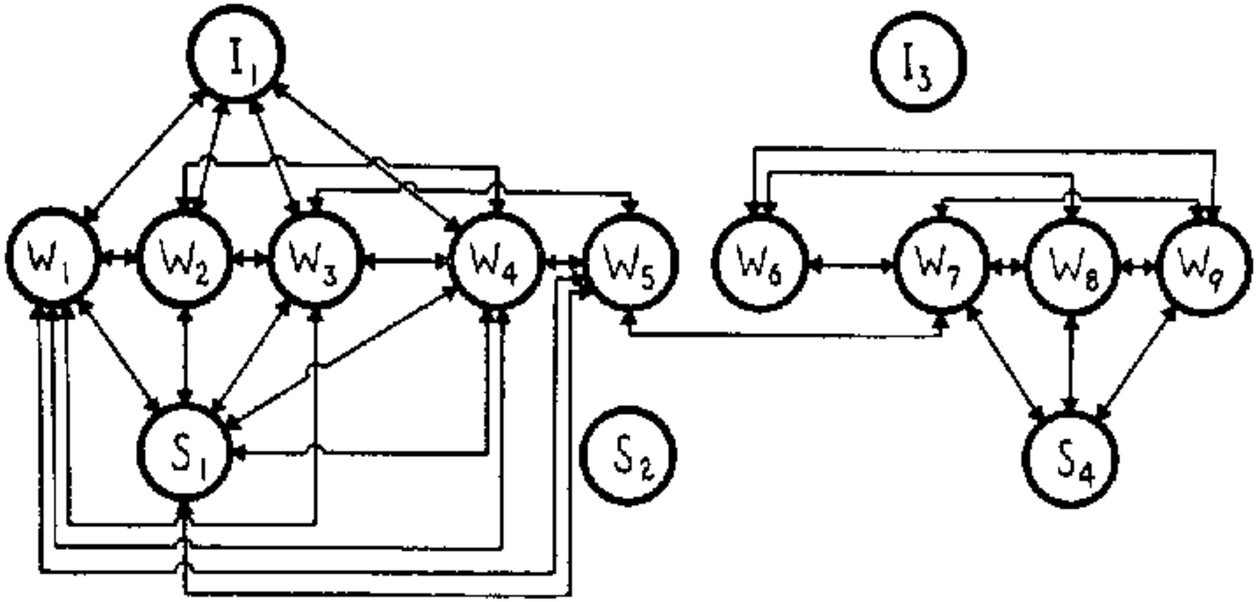
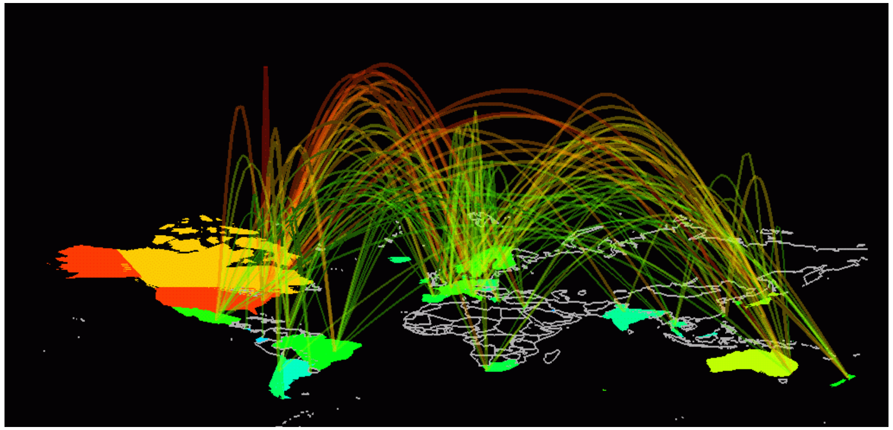


Figure 39
Participation in Games

Networks: why do we care?

Networks are everywhere



So maybe we should try to understand them?

- A molecule is a network of atoms
- A brain is a network of neurons
- A body contains many networks, including the circulatory system
- Genes form regulatory networks that turn other genes on and off
- Firms are networks of individuals, passing along information, orders and coordinating efforts
- Buildings contain many networks, including heating/cooling, plumbing, electrical
- Economies are networks of firms and other agents buying and selling
- Societies are networks
- Countries contain many networks, e.g., transportation systems, phone systems
- The internet is a network
- Ecosystems are networks of species eating each other, creating environments for each other, etc.

But …

- Networks are also a lens
- We see networks everywhere because we like to think that way
- A network is created any time a researcher says
  - I’m interested in this set of people,
  - And, I define a tie as …. [having the same color hair] [having met before] [etc]
- Don’t want to over-reify networks
- And yet …

Network mechanisms

Consider the case of AIDS

- 1981 CDC aware of increasing number of cases of opportunistic illnesses like Kaposi's sarcoma
- Virtually all cases were gay men
  - Syndrome initially named Gay-Related Immune Deficiency (GRID)
- Logistic regression of opportunistic illness on being gay
- Proposed mechanism
  - Stigmatized identity causes stress, leading to weakened immune system

Subject ID	Age	Gay	Rare Cancer
1	33	0	0
2	27	0	0
3	89	1	1
4	34	0	0
5	56	1	0
6	23	0	0
7	54	0	0
8	12	1	1
9	45	0	0
10	67	0	0
11	43	1	1
12	21	1	0

Network logic

Contagion | diffusion | influence mechanisms

Subject ID	Age	Gay	Rare Cancer
1	33	0	0
2	27	0	0
3	89	0	0
4	34	0	0
5	56	1	0
6	23	0	0
7	54	0	0
8	12	1	1
9	45	0	0
10	67	0	0
11	43	1	1
12	21	1	0

Network structure provides backcloth that enables and constrains flows

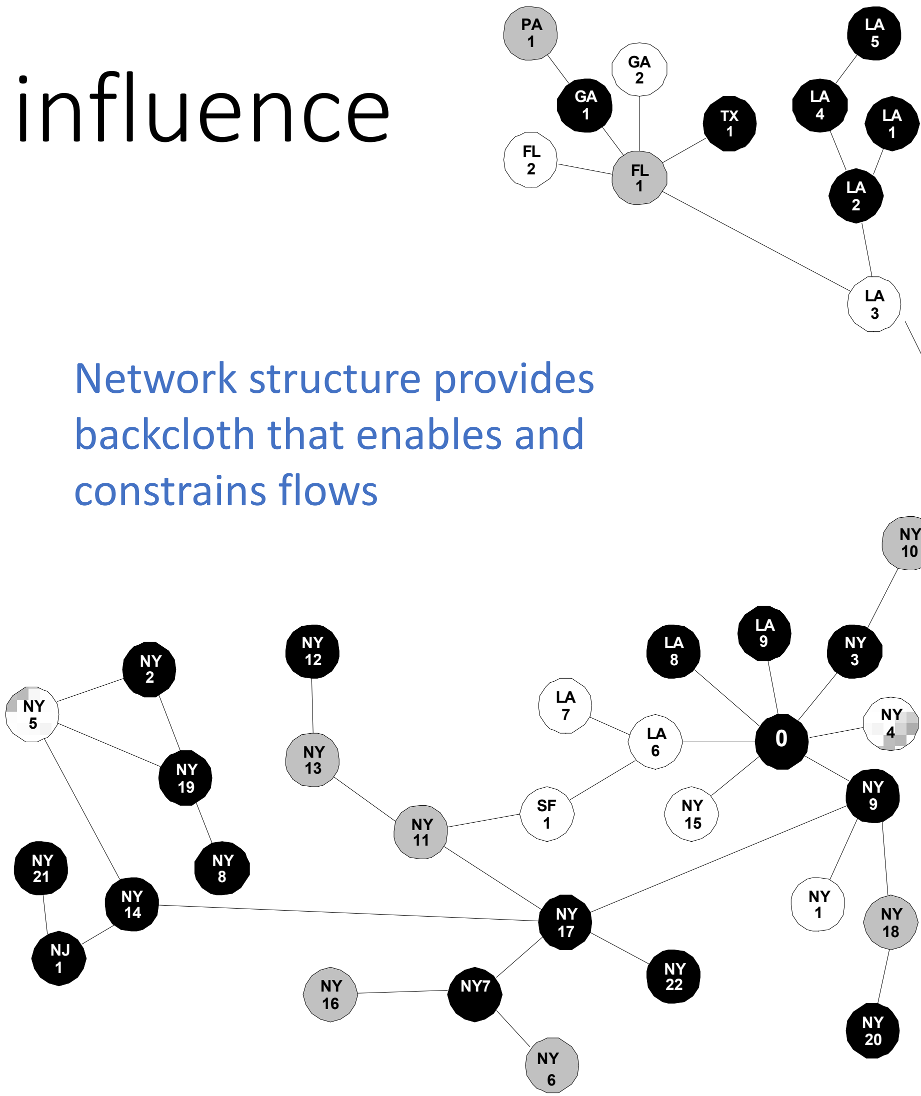
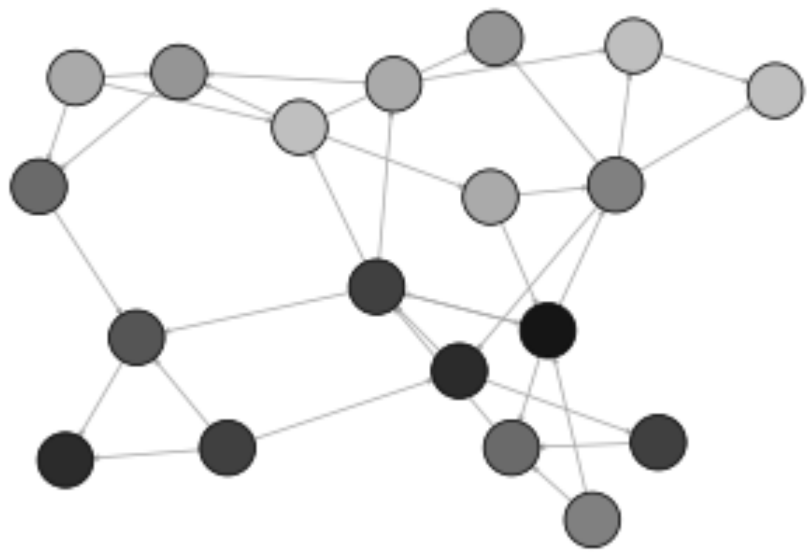
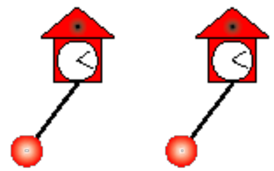


Diagram by Bill Darrow, CDC

Network models of style

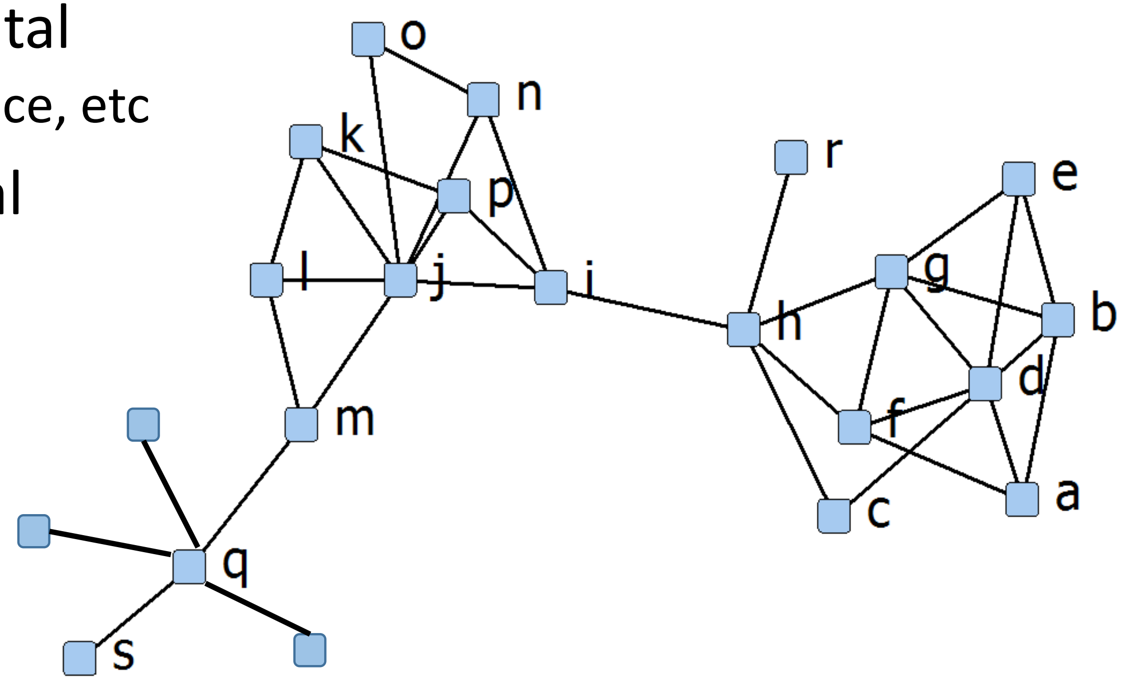


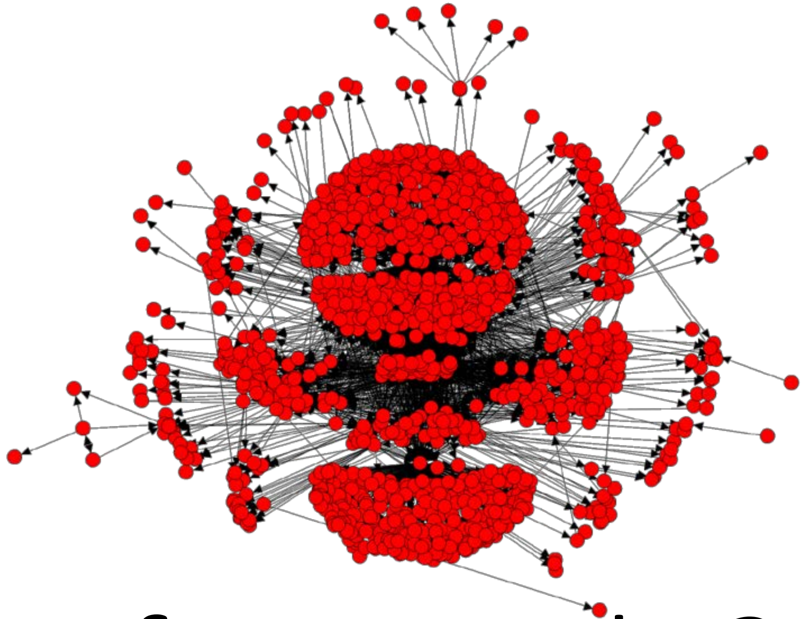
- Why do people …?
  - Wear the clothes they do
  - Speak the way they do
  - Believe the things they do
  - Do things the way they do
  - Etc.
- Partly individual reasons (maximize utility function), but partly contagion/influence from people they know
  - Contagion, diffusion, adoption of innovation, common fate



Modeling achievement

- Why some individuals/organizations are more successful than others
- Standard answer is human capital
  - Motivation, education, intelligence, etc
- Network answer is social capital
  - Position in the network
- Bridging/Brokering positions
  - Access to non-redundant info
  - Freedom of action
  - Combine knowledge from one group to that of another





What are the consequences of networks?

Diffusion & influence

- Networks provide a system of pipes through which things can flow
  - Information
  - Goods
  - Money
  - Infections
- Interpersonal influence processes
  - I adopt vaping, you adopt vaping, your other friend adopts…
  - Eating patterns – e.g., so-called obesity contagion

Coordination & access to resources

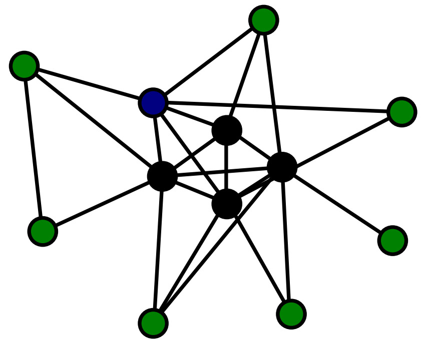
- Like common culture, social networks bind people together so they can accomplish more than individuals working alone
  - Can literally link arms
  - More figuratively can agree/ally with each other, vote together
  - Dependencies, kinship ties lead to help
  - Ties bind people together to create superordinate entities, like bureaucracies
- Entrepreneur can use friends'
  - Money
  - Computer expertise
  - Time
  - Access to city council

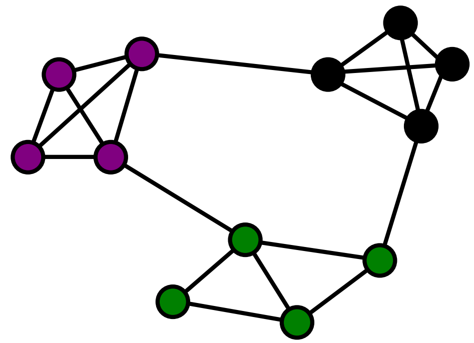
Network theory provides explanations for …

- Style
  - Why people have the particular beliefs, behaviors, and belongings they do
  - Generic research question: explain hetero/homogeneity
  - Generic network explanation: contagion, diffusion, interpersonal influence processes
    - Contagion of obesity, happiness, etc
    - Diffusion of innovations
    - Spread of disease
    - Fads and fashion
    - Social conformity
- Success
  - Achievement and reward
  - Why some people are more successful than others
  - Generic research question: explain differential success
  - Generic network explanation: social capital
    - Ties provide access to resources
    - Certain positions in social structures are advantageous
    - Coordination & collaboration
    - Innovation knowledge creation

Levels of analysis -- Organized by most to least number of units

- Dyad level – $O(n^2)$
  - Units are pairs of persons
  - Variables are things like presence of absence of a certain kind of tie between each pair of persons in network
- Node level – O(n)
  - Units are persons
  - Variables are things like the number of friends each person has
- Group/network level – O(1)
  - Units are whole networks (e.g., teams, firms or countries)
  - Variables are things like the density of trust ties, or the average number of degrees of separation between members of the group





Dyad level

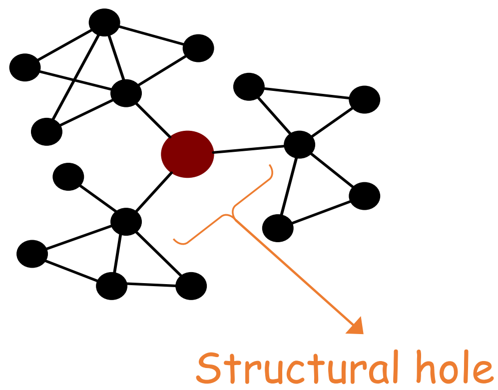
- Raw network data are dyadic
- for each pair of persons we measure
  - whether they have a tie or not (are they friends?)
  - How strong the relationship is (how close are they?)
  - Other aspects of the tie
    - How long have they been friends?
    - How often do they talk?
  - Measurement can undirected or directed
    - Undirected: are they co-workers? If A is coworker of B, then B is coworker of A
    - Directed: advice. Does A give advice to B? If so, maybe B does not give advice to A

Dyad level : antecedents and consequences

- Consequences
  - If A has tie to B, and A knows something, they may tell B, and now both know it
  - So, a consequence of the tie is similarity/homogeneity
    - I have same info as you
    - I adopt same shoes as you
- Antecedents
  - What determines which pair are friends are which are not?
  - Often look to attributes of the individuals
  - So, an antecedent of the tie is similarity

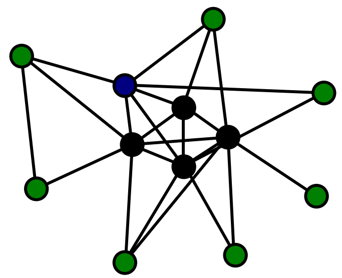
Node level: antecedents and consequents

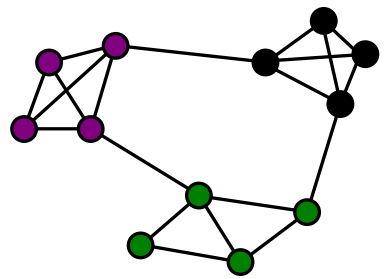
- Consequences
  - Employees with more friends in the higher levels of the organization get promoted earlier and have better raises
  - In management the canonical hypothesis is that managers with more structural holes perform better and get rewarded better
- Antecedents
  - Individuals with more outgoing personalities tend to be more central in the organizational network
  - People with ability to interact productively with diverse kinds of people are more likely to ties to people who are not tied to each other



Group level

- Consequences
  - Teams with more centralized communication networks solve problems more quickly
- Antecedents
  - Teams with greater demographic homogeneity more likely to have core/periphery network structures rather than clumpy structures





Antecedents and consequences

Antecedents

- Socio/cultural/psychological processes that give rise to social ties, interactions, exchanges
  - What determines who is connected to whom?
  - Why do some people have more ties than others?
  - Why does the network have the structure it does?
- **Theory of networks**

Consequences

- Mechanisms that translate ties, positions, structure into outcomes
  - How does the tie between two actors affect what happens between them?
  - How does centrality translate into power?
  - How does network structure determine diffusion speed?
- **Network theory**

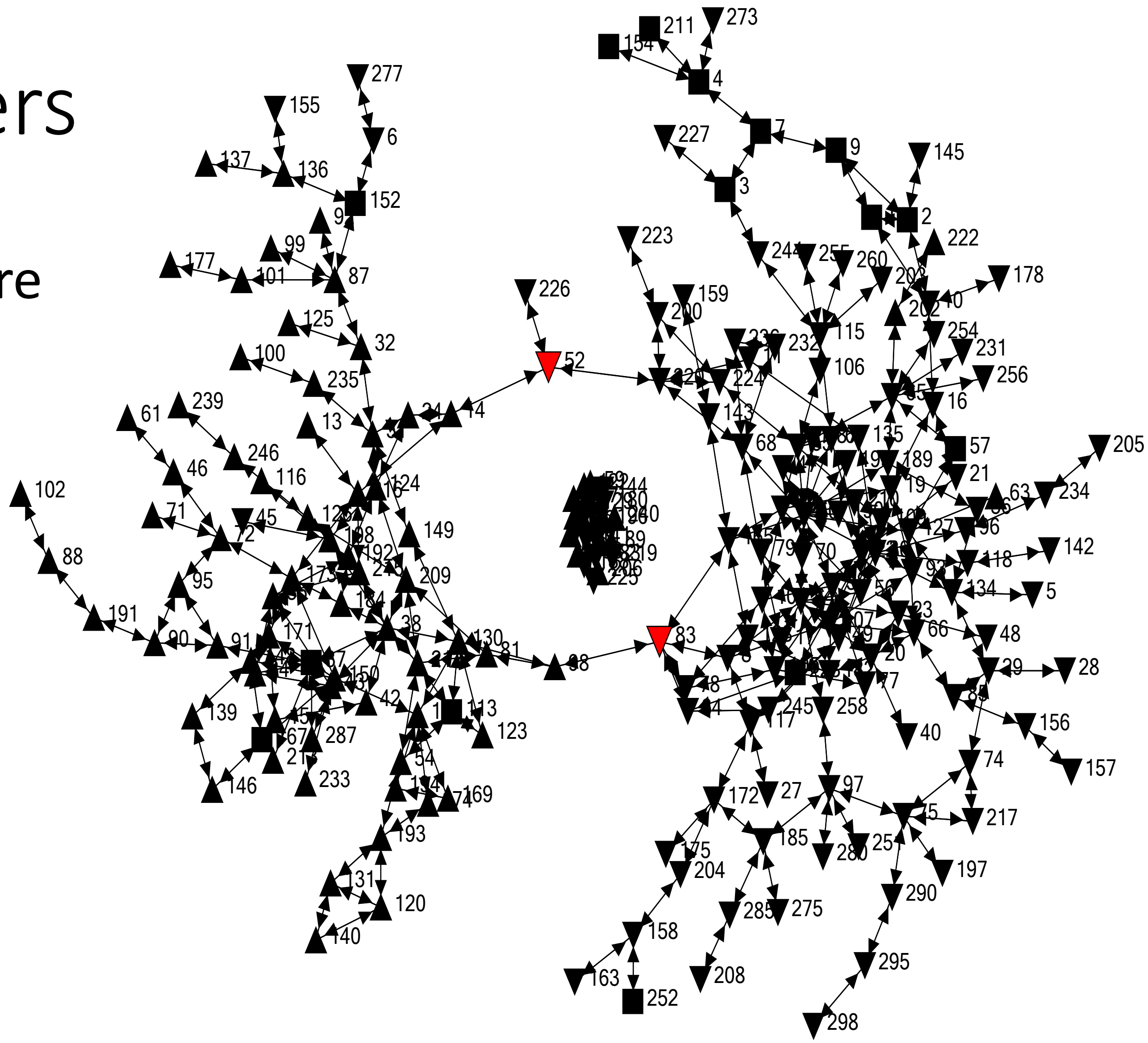
Types of studies

	Dyad Level	Node Level	Group Level
Theory of Networks (Antecedents)	Understanding who becomes friends with whom	Explaining why some people are more liked than others	Explaining why some groups have more centralized network structures
Network Theory (Consequences)	Predicting similarity of opinion as a function of friendship	Explaining why some employees rise through the ranks faster than others as a function of social ties	Predicting team performance as a function of structure of trust network within team

Characteristics of network thinking

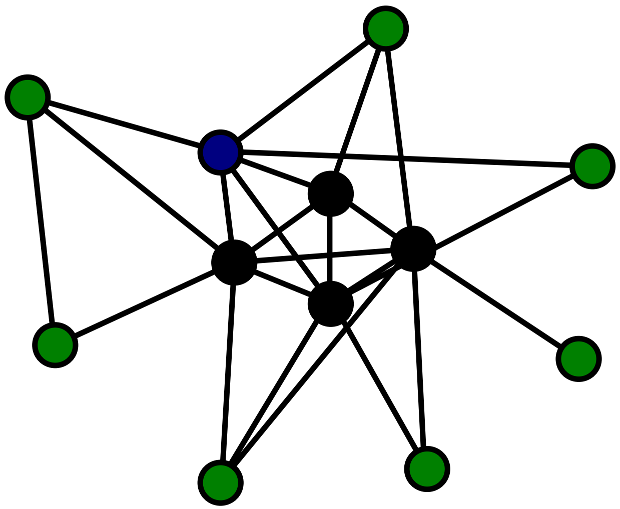
Structure matters

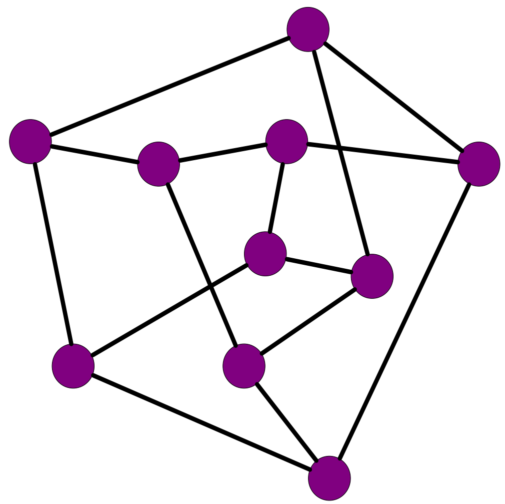
- This is a fragile structure easily broken up

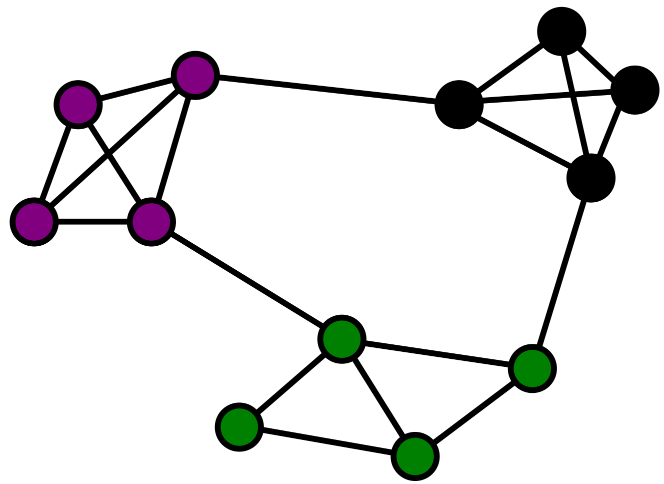


Which networks are good for what?

- Consequences of these structures for the organization and for nodes





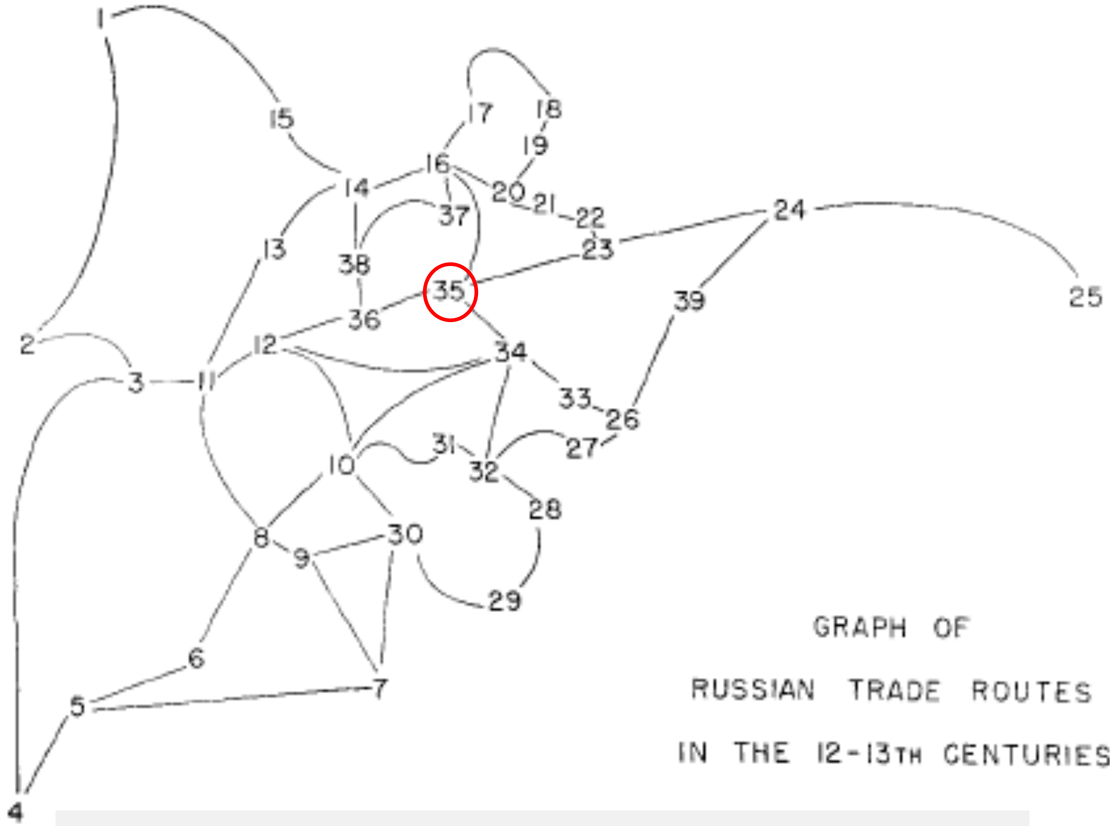


Position matters: the emergence of Moscow

- Pitts (1979) study of 12th century Russia and the later emergence of Moscow
- Why did Moscow come to dominate?
  - Great man theory
  - Resource richness

Position matters

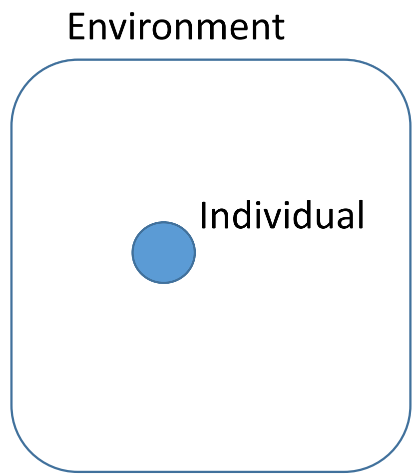
- Rivers enable trade between city-states
  - System of rivers creates network of who can trade directly and indirectly with whom
  - What happens in the network is a function of global paths and position
  - Moscow very high in betweenness centrality

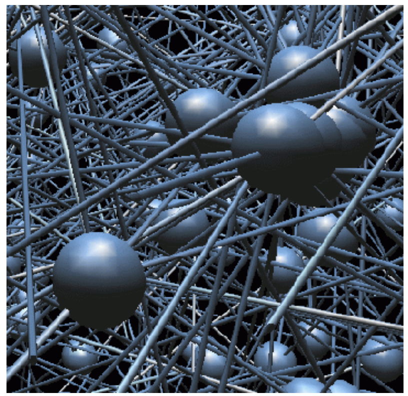


Nodes have high betweenness to the extent they are along the shortest paths between pairs of nodes

SNA as open systems perspective

- Importance of an individual's environment
  - To explain individual outcomes, must take into account the node's social environment in addition to internal characteristics
  - In SNA, the environment is conceptualized as network
  - An emphasis on structure relative to agency
  - Consistent with an open systems perspective
- The contrast is with an essentialist/dispositional perspective
  - Predict individual's outcomes using other characteristics of the individual
  - Employee's success a function of ability and motivation

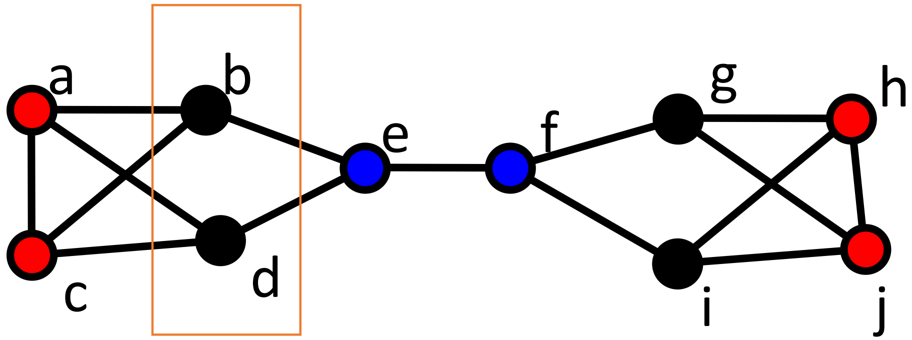




We are all embedded in a thick web of relations

Environment as location in network

- Many fields have concept of environment affecting the individual
  - Turbulent/differentiated environments in organizational theory
- In networks, the environment is conceptualized as other agents
- And these agents are connected to each other and to ego in a particular pattern/structure



Traits versus environment

- Traditionally, social science has focus on attributes of individuals to predict individual outcomes
  - Income as a function of education
  - Essentialist, dispositional, closed system perspective
- SNA looks not only at your own attributes, but also the attribs of the people in your life

Variables (attributes)

Cases (entities)

	Age	Sex	Education	Income
1001				
1002				
1003				
1004				
1005				
…				

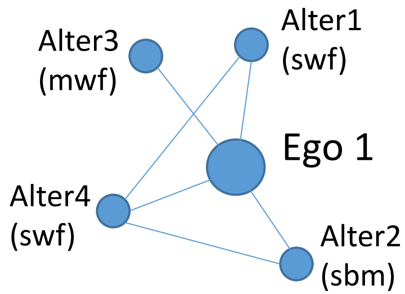
Research designs

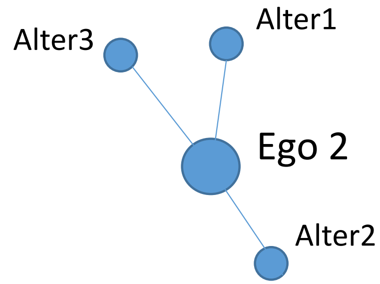
Whole network / sociocentric design

- Start with a set of people (typically a “natural” group such as a gang or a department)
- Collect data on the presence/absence (or strength) of ties of various kinds among all pairs of members of the set
  - Who doesn’t like whom; How frequently each pair of persons have a conversation
  - Typically collected via survey: respondent presented with roster of people to select/rate
- Issues
  - The set of persons needs to be some kind of census – can’t randomly pick sample of 100 persons from the population of all Americans
  - The set can’t be too big
  - Problems with inferential validity – how to generalize results?

Personal network / egocentric design

- Select random sample of respondents/subjects
  - Call them egos
- For each subject, identify the set of persons in that subject's life
  - Call them alters
- For each alter, determine their individual characteristics
  - E.g., ask ego how old the alter is, whether they use drugs, etc.
- For each alter, determine the nature of the relationship with ego
  - E.g., ask ego how often they talk to alter, whether alter is a neighbor, etc.
- For pairs alters, determine their relationships to each other
  - E.g., ask ego whether alter 1 is friends with alter 2, etc.





Issues with personal network design

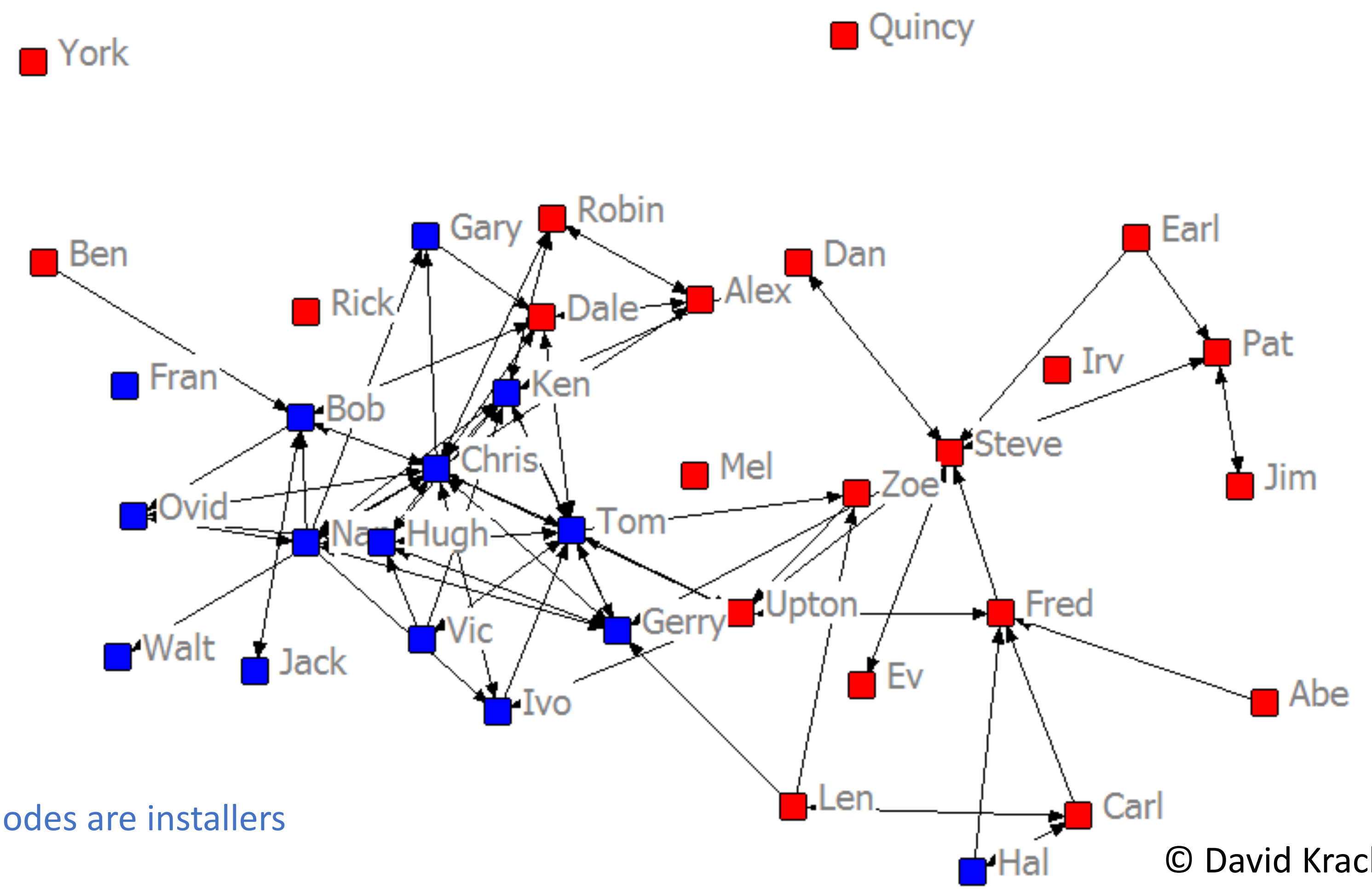
- Can use random samples, enabling generalizability of findings
- Can study very large populations
- Can’t say anything about network structure, or position of nodes within the structure
- Typically collected via survey, so all of the information about alters is obtained from ego’s perceptions
    - May be inaccurate
    - But maybe it is ego’s perception that matters …

Cognitive social structures (CSS) design

- A blend of whole network and personal network designs
- Start with natural group of persons as in whole network design
- Ask each person to indicate not only their own relationship with each other person, but also their perception of the relationships among all pairs of persons
- Result is a perceived network from each member of the network
- Issues
  - Tedious for the respondent – can only be used with small groups
  - Extremely rich data. Can calculate accuracy of each person's perceptions. Study effects of social perceptions

"Who would this person consider to be a personal friend? Please place a check next to all the names of those people who that person would consider to be a friend of theirs"

Friendship network -- ilas

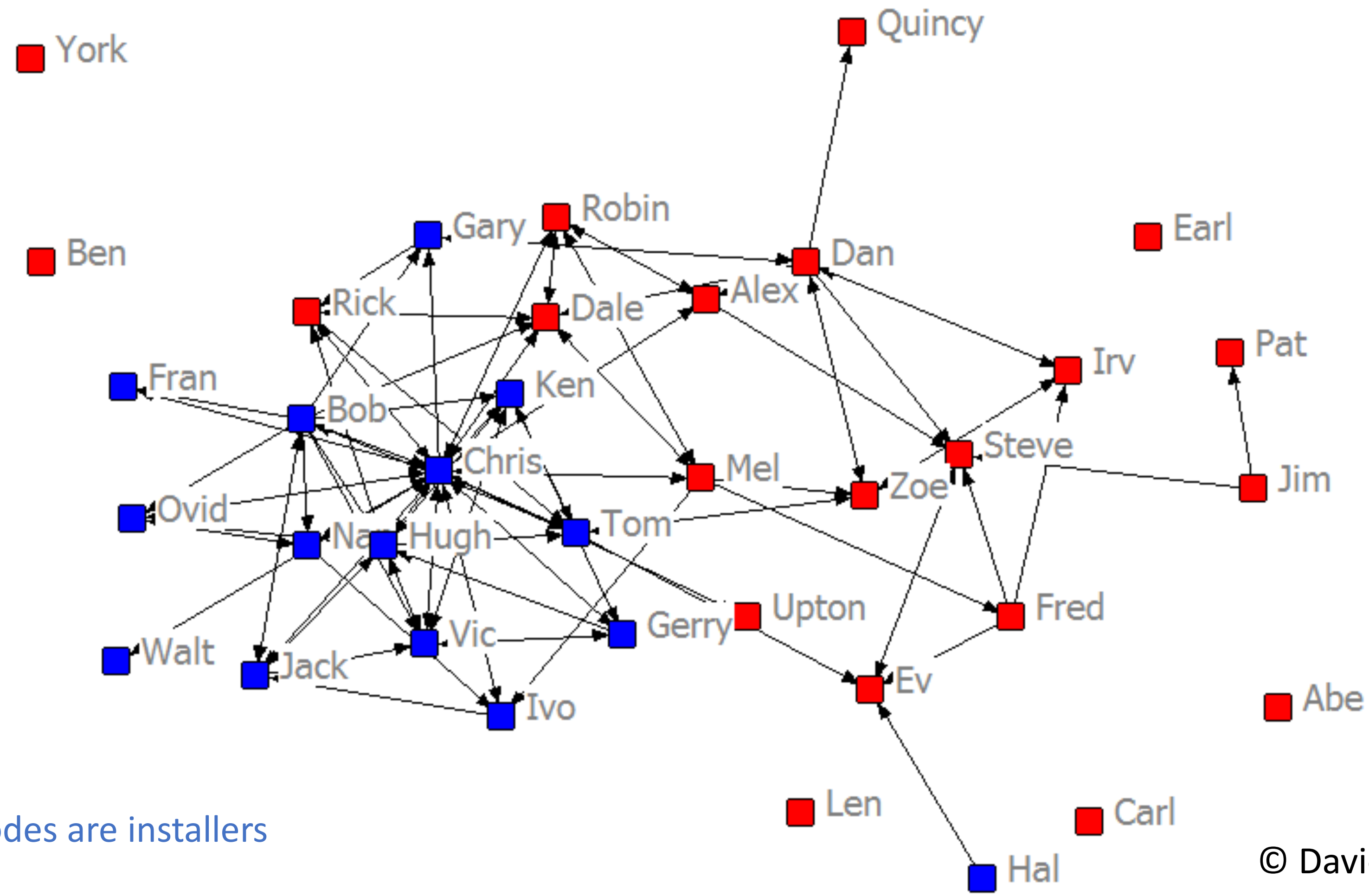


Blue nodes are installers

Node	Indeg
Chris	10
Ken	8
Tom	8
Gerry	7
Steve	6
Hugh	6
Bob	5
Dale	5
Nan	5
Fred	4
Ovid	4
Upton	4
Ivo	3
Pat	3
Robin	3
Alex	3
Carl	2
Gary	2
Zoe	2
Dan	2

"Who would this person consider to be a personal friend? Please place a check next to all the names of those people who that person would consider to be a friend of theirs"

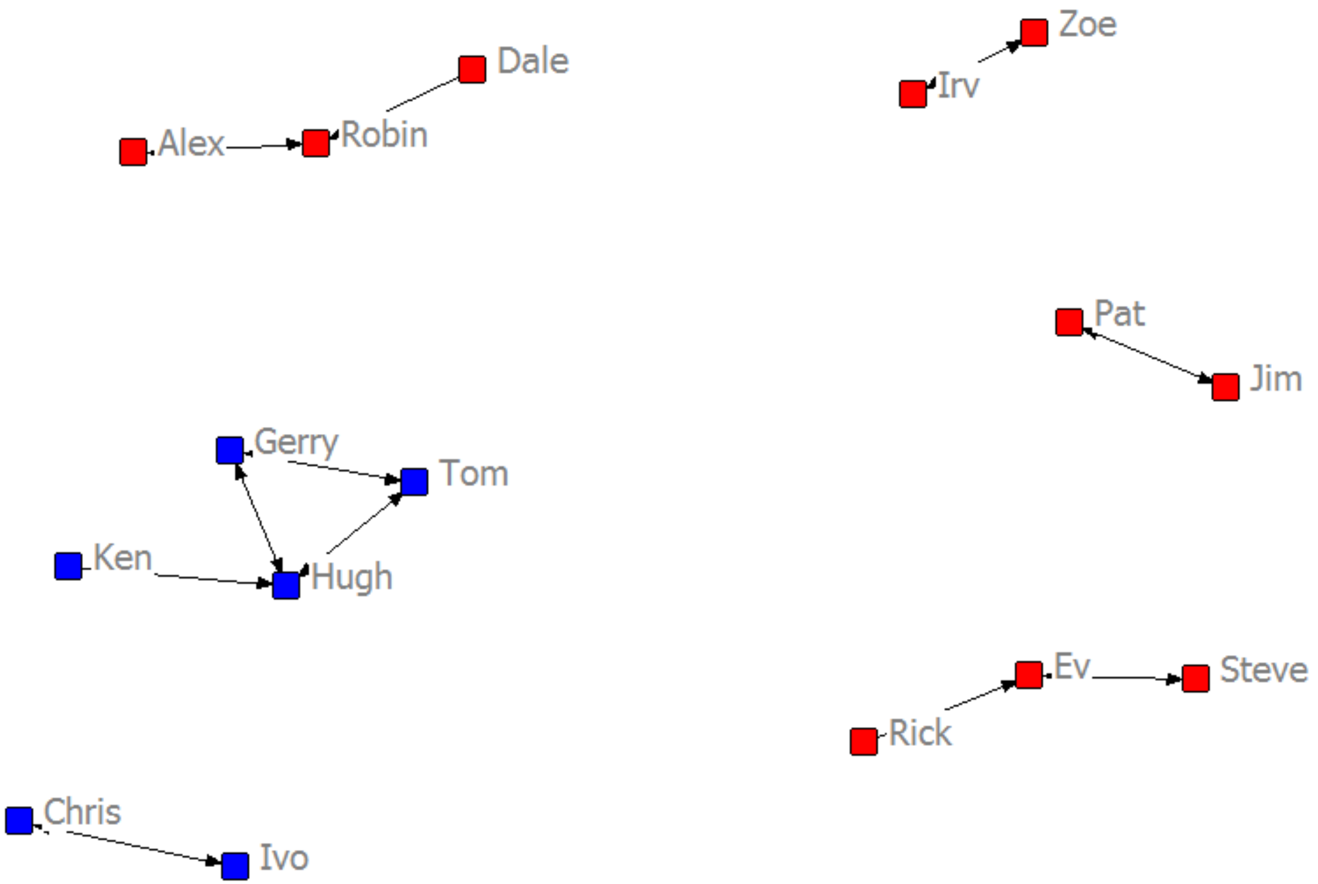
Chris's perception of the friendship network



Blue nodes are installers

"Who would this person consider to be a personal friend? Please place a check next to all the names of those people who that person would consider to be a friend of theirs"

Ev's perception of the friendship network



Fundamentals of Network Analysis

- **Data** structure
- **Matrix** Algebra
- Set and **graph** theory

Defining & Describing a network

- In social network analysis, we draw on two major areas of mathematics regularly:
  - **Matrix Algebra**
    - Tables of numbers
    - Operations on matrices enable us to draw conclusions we couldn’t just intuit
  - **Graph Theory**
    - Branch of discrete math that deals with collections of ties among nodes and gives us concepts like paths

Network vs. Case Perspective

- One of the biggest differences between the SNA perspective and more traditional social science perspectives is the nature of the data
  - Instead of individual cases, where we collect the same information for a bunch of people
  - Here, we collect information about the interaction of pairs of people

Mainstream Logical Data Structure

- 2-mode rectangular matrix in which rows (cases) are entities or objects and columns (variables) are attributes of the cases
- Analysis consists of correlating columns
  - Emphasis on explaining one variable

ID	Age	Education	Salary
1			
2			
3			
4			

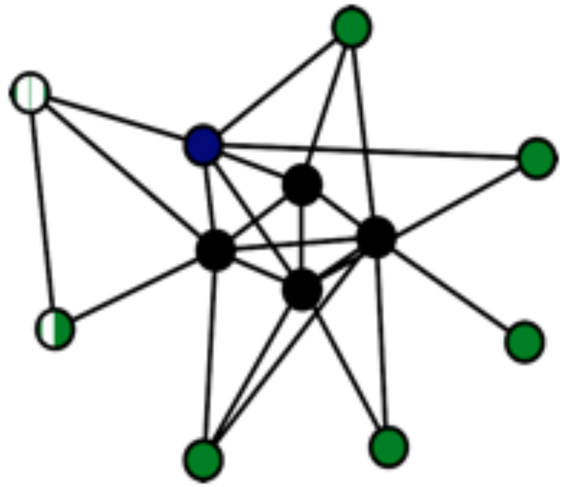
Network Logical Data Structures

Friendship

	Jim	Jill	Jen	Joe
Jim	-	1	0	1
Jill	1	-	1	0
Jen	0	1	-	1
Joe	1	0	1	-

Proximity

	Jim	Jill	Jen	Joe
Jim	-	3	9	2
Jill	3	-	1	15
Jen	9	1	-	3
Joe	2	15	3	-



- Multiple relations recorded for the same set of actors
- Each relation is a variable
  - variables can also be defined at more aggregate levels
- Values are assigned to pairs of actors
- Hypotheses can be phrased in terms of correlations between relations
  - Dyadic-level hypotheses

Network description

describing networks

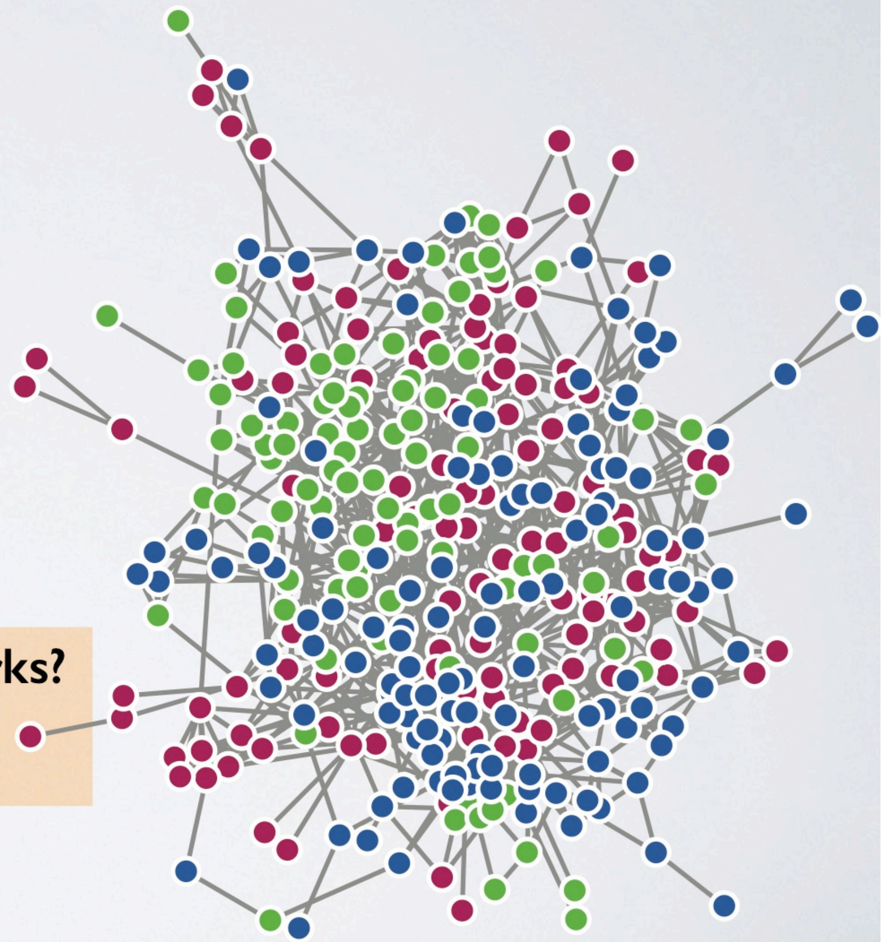
what networks look like

questions:

- **how are the edges organized?**
- **how do vertices differ?**
- **does network location matter?**
- **are there underlying patterns?**

what we want to know

- **what processes shape these networks?**
- **how can we tell?**

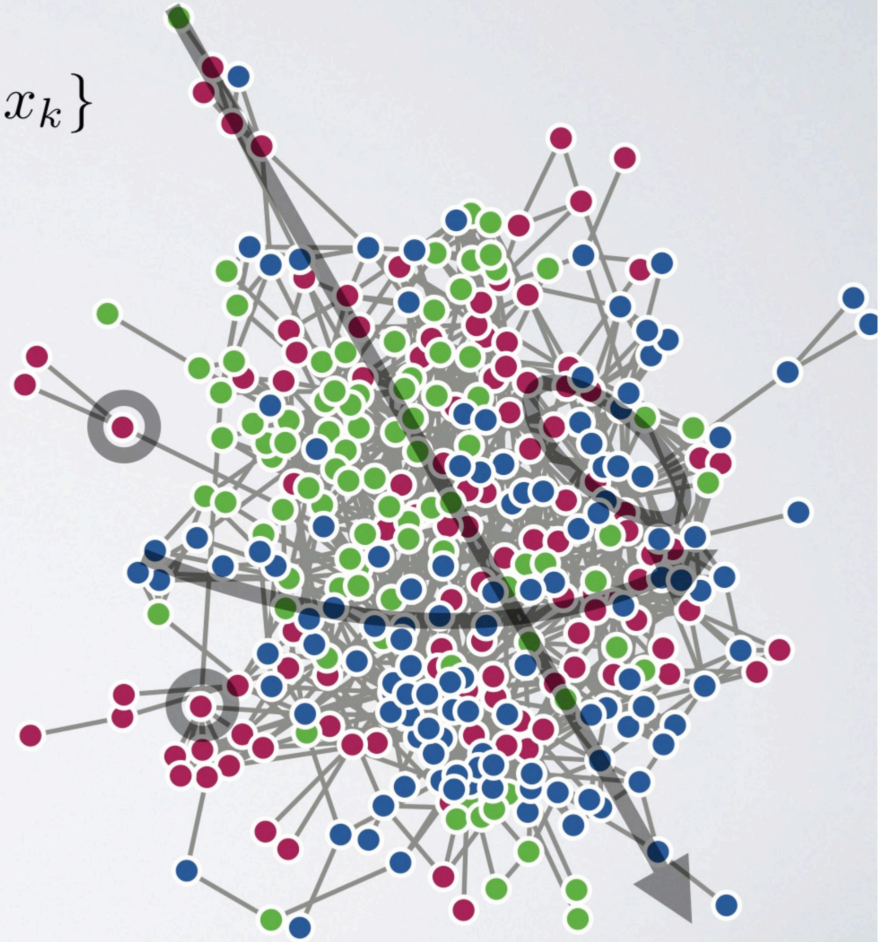


describing networks

a first step : **describe its features**

$$f : G \to \{x_1, \ldots, x_k\}$$

- **degree distributions**
- **short-loop density (triangles, etc.)**
- **shortest paths (diameter, etc.)**
- **vertex positions**
- **correlations between these**



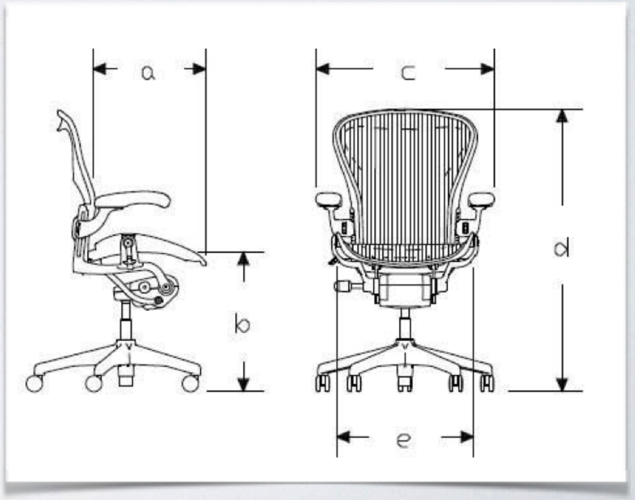
describing networks

a first step : **describe its features**

$$f : \text{object} \rightarrow \{x_1, \ldots, x_k\}$$

- **physical dimensions**
- **material density, composition**
- **radius of gyration**
- **correlations between these**

helpful for exploration, but not what we want…



describing networks

what we want : **understand its structure**

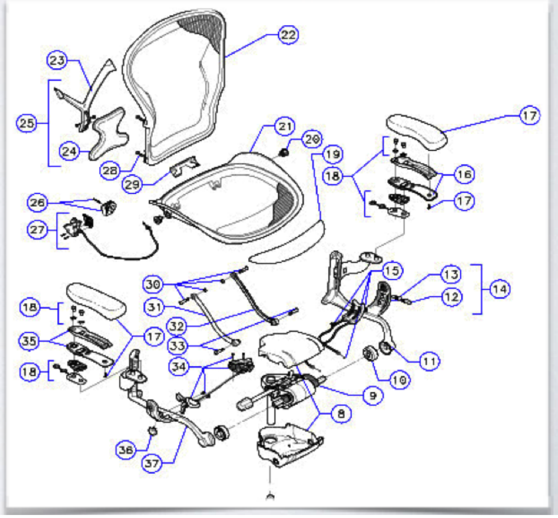
$$f : \text{object} \rightarrow \{\theta_1, \ldots, \theta_k\}$$

- **what are the fundamental parts?**
- **how are these parts organized?**
- **where are the degrees of freedom $\vec{\theta}$?**
- **how can we define an abstract class?**
- **structure — dynamics — function?**

what does ***local-level structure*** look like?

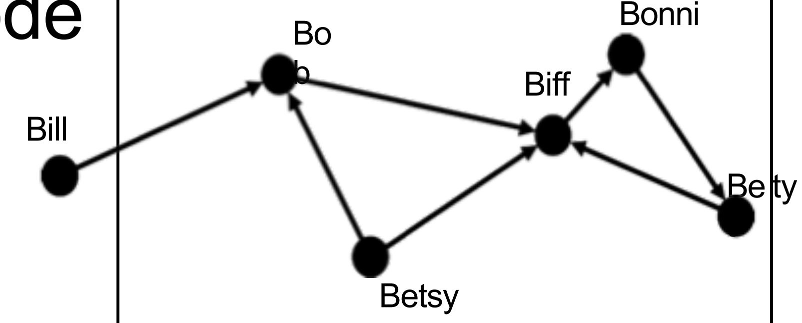
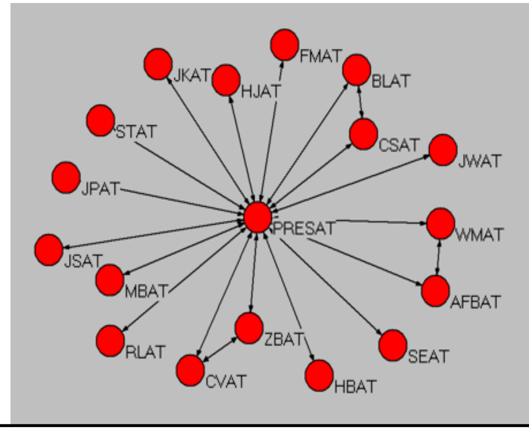
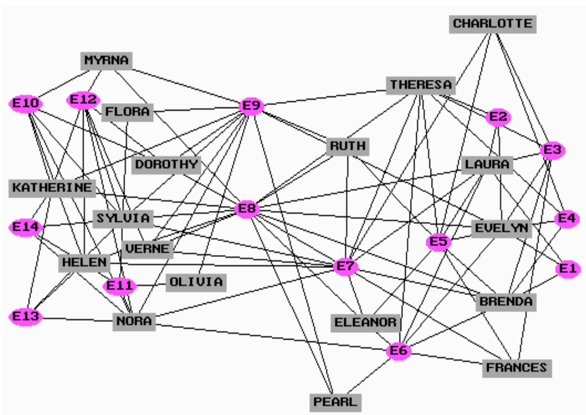
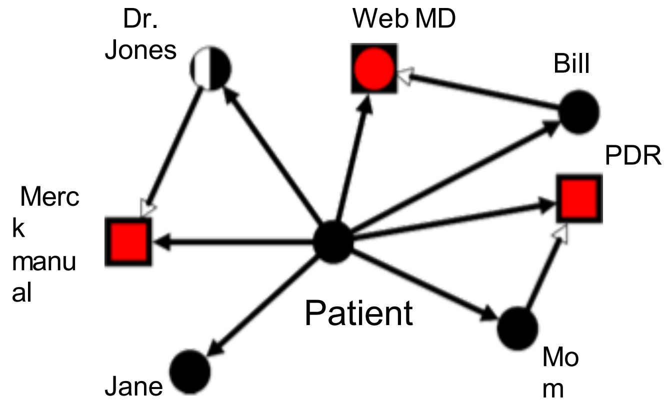
what does ***large-scale structure*** look like?

how does ***structure constrain*** function?

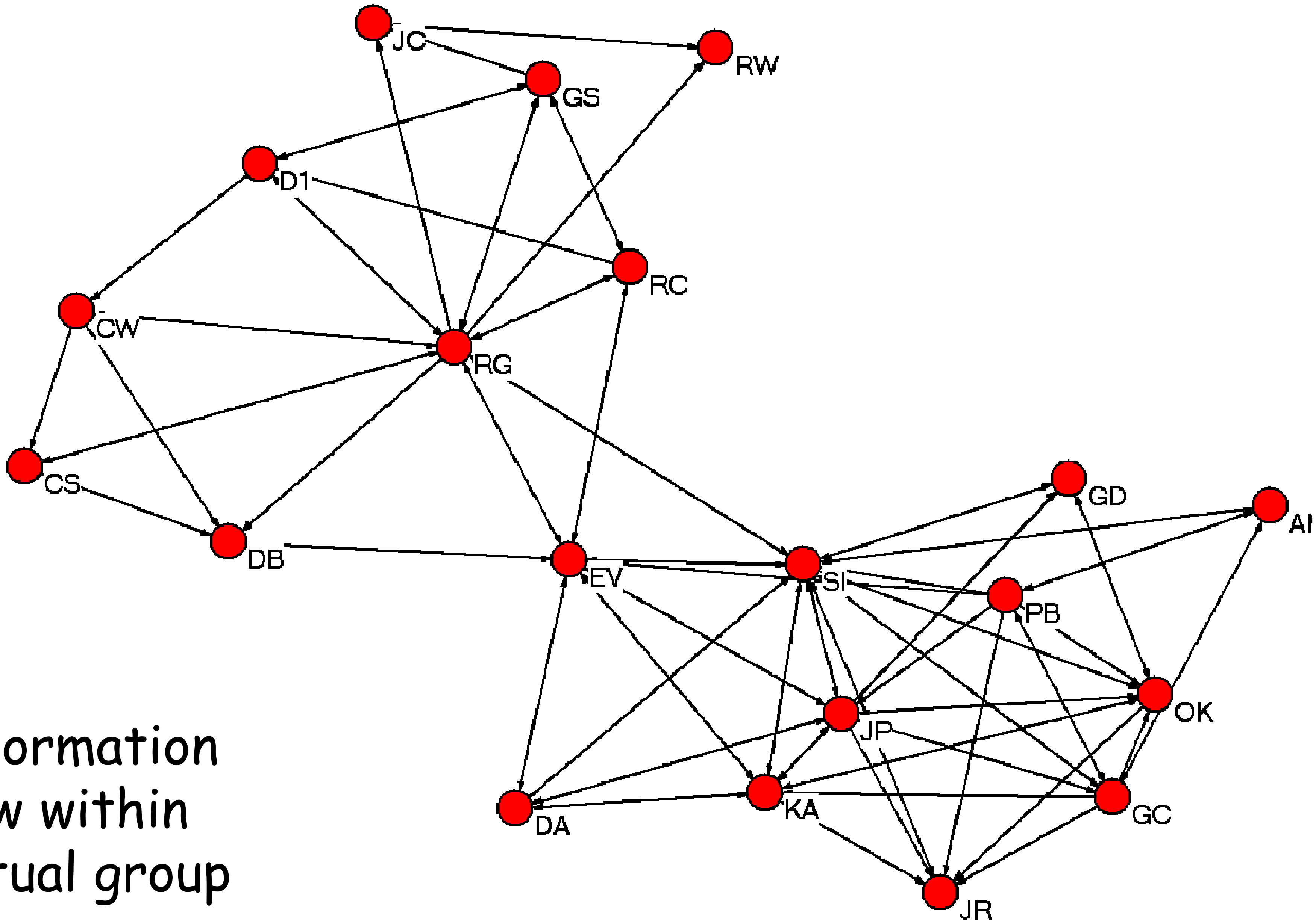


Network Representation

Kinds of Network Data

	Complete	Ego
1-mode		
2-mode		

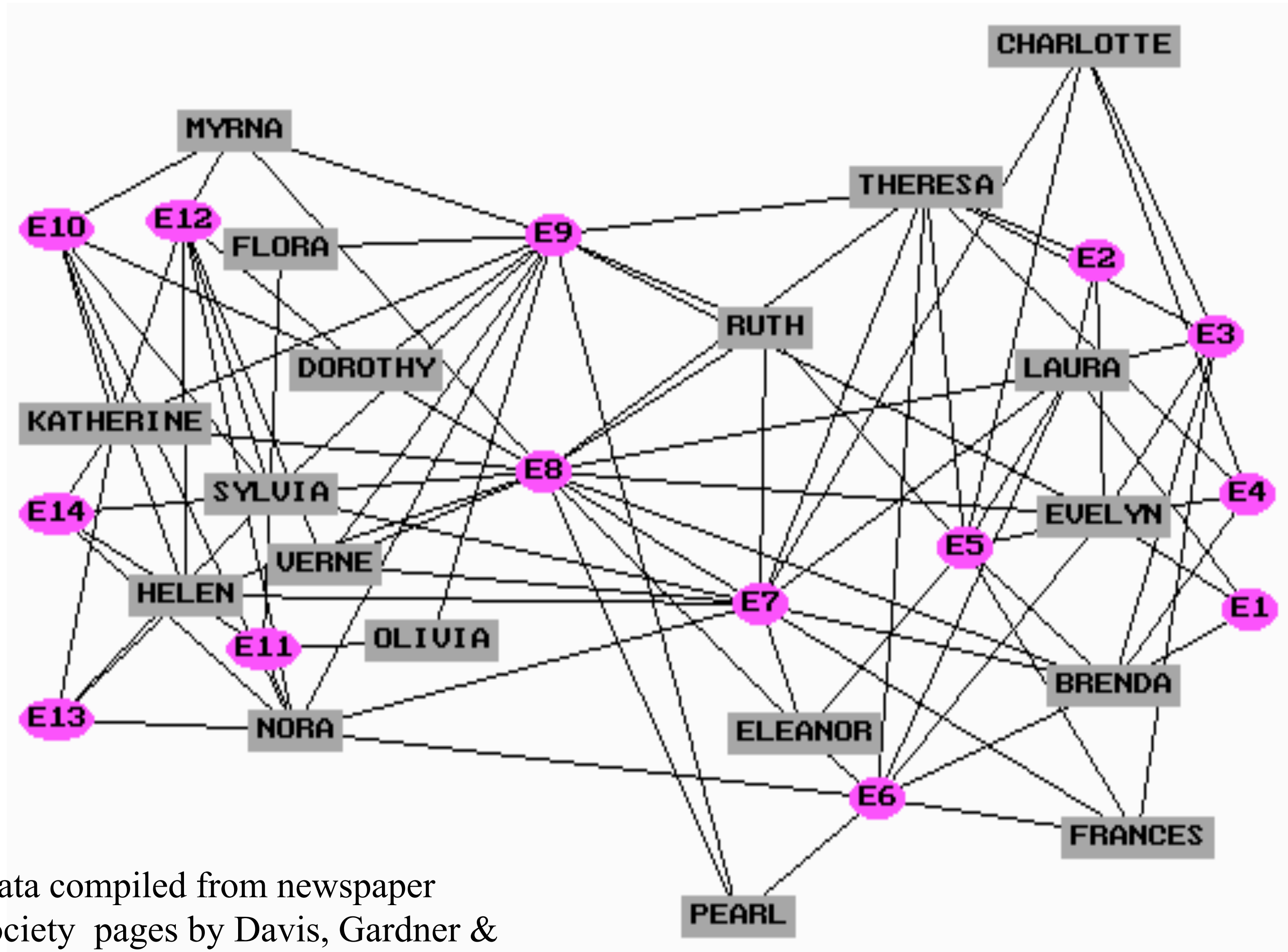
1-mode Complete Network



Information flow within virtual group

Data collected by Cross

2-mode Complete Network

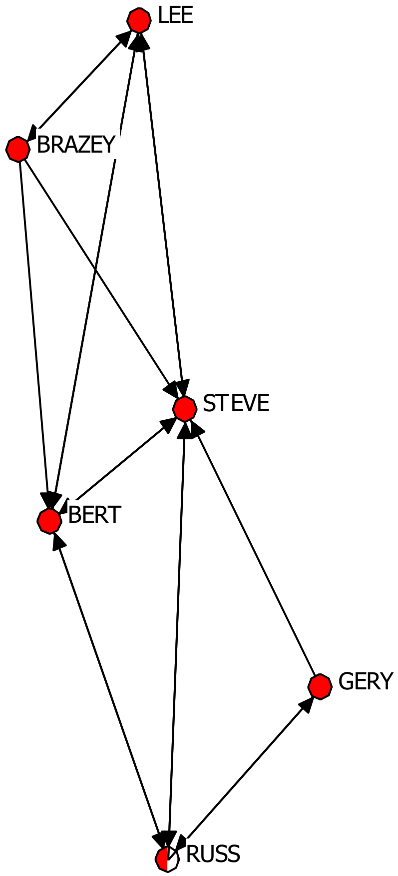


Data compiled from newspaper society pages by Davis, Gardner & Gardner

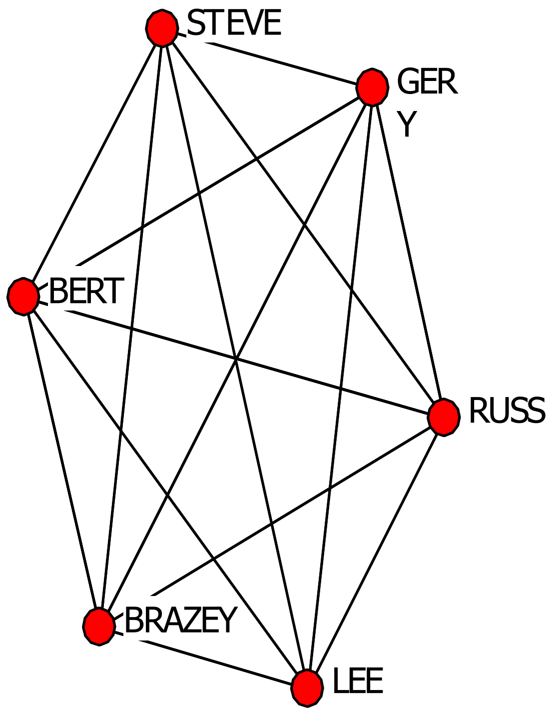
Complete Network Data vs. Complete Graph

- The term "Complete Network Data" refers to collecting data for/from all actors (vertices) on the graph
  - The opposite if Ego-Network or Ego-Centric Network data, in which data is collected only from the perspective an individual (the ego)
- The term "Complete Graph" refers to a graph where every edge that could exist in the graph, does:
  - For all i, j ($j>i$), $v(i,j) = 1$

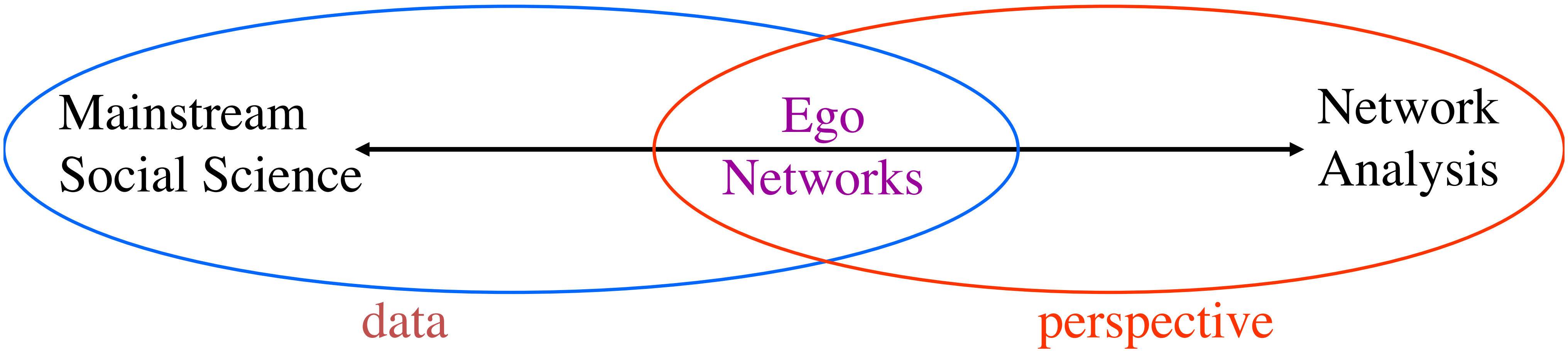
Complete Network Data



Complete Graph



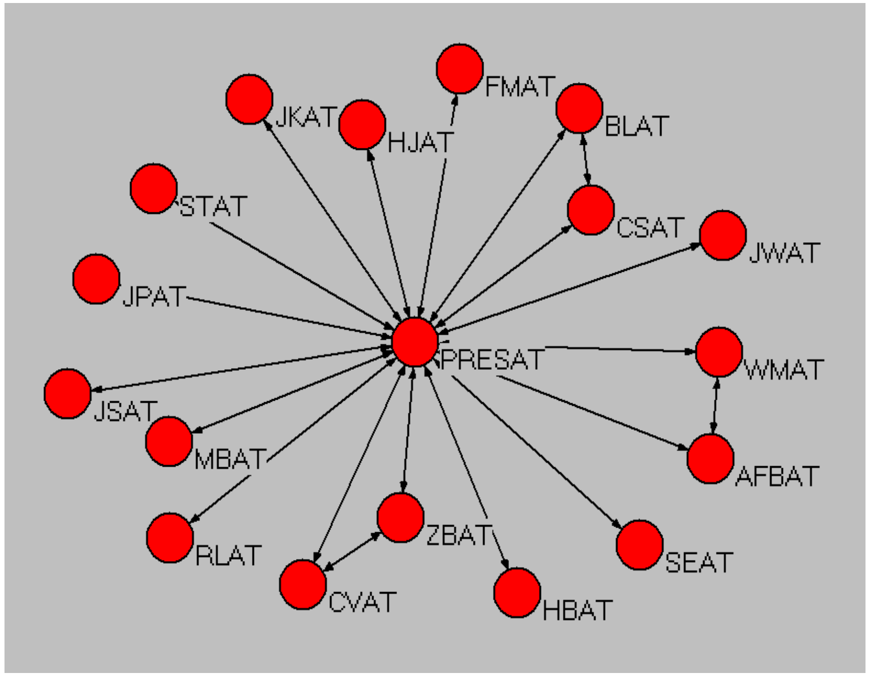
Ego Network Analysis



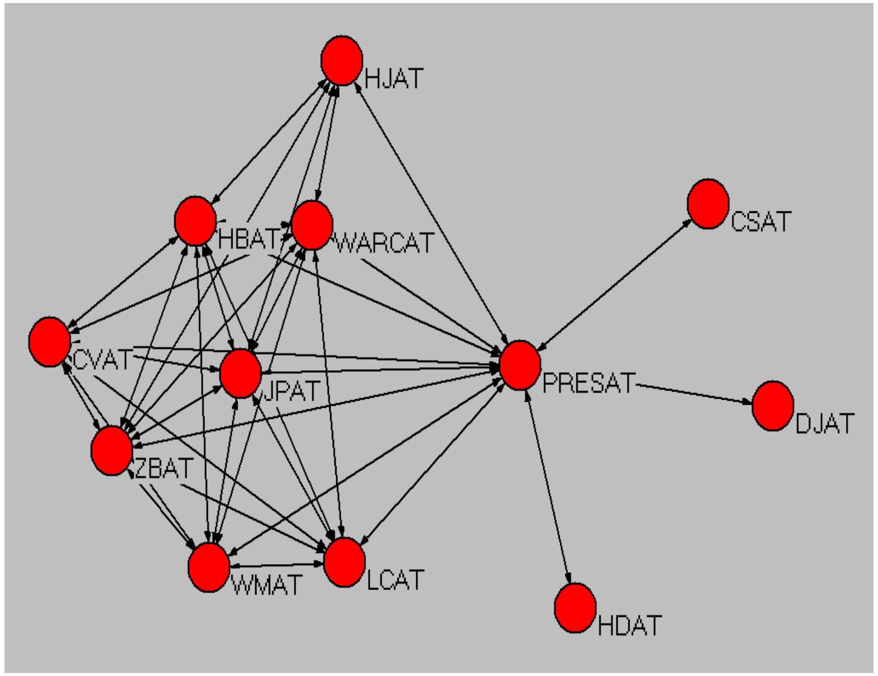
- Combine the perspective of network analysis with the data of mainstream social science

1-mode Ego Network

Carter Administration meetings



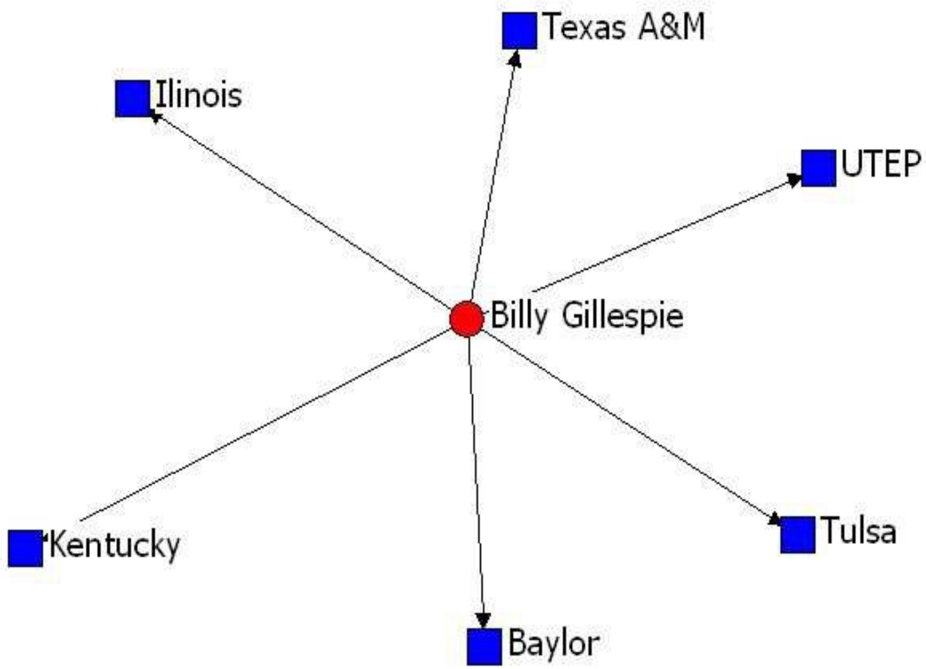
Year 1



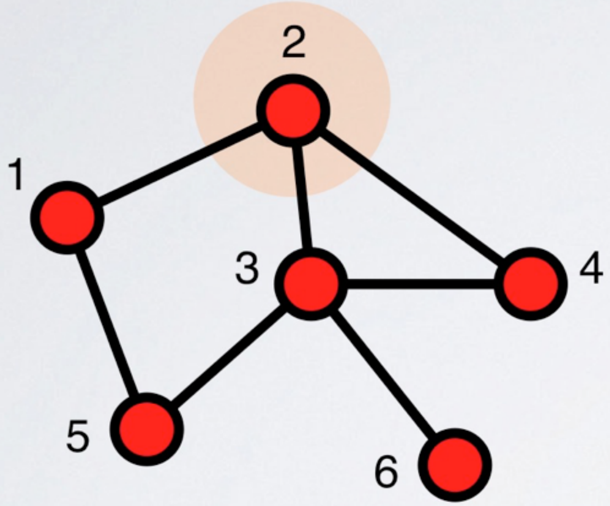
Year 4

Data courtesy of Michael Link

2-mode Ego Network



representing networks – simple undirected



undirected

unweighted

no self-loops

adjacency matrix

A	1	2	3	4	5	6
1	0	1	0	0	1	0
2	1	0	1	1	0	0
3	0	1	0	1	1	1
4	0	1	1	0	0	0
5	1	0	1	0	0	0
6	0	0	1	0	0	0

adjacency list

A

$1 \rightarrow \{2, 5\}$

$2 \rightarrow \{1, 3, 4\}$

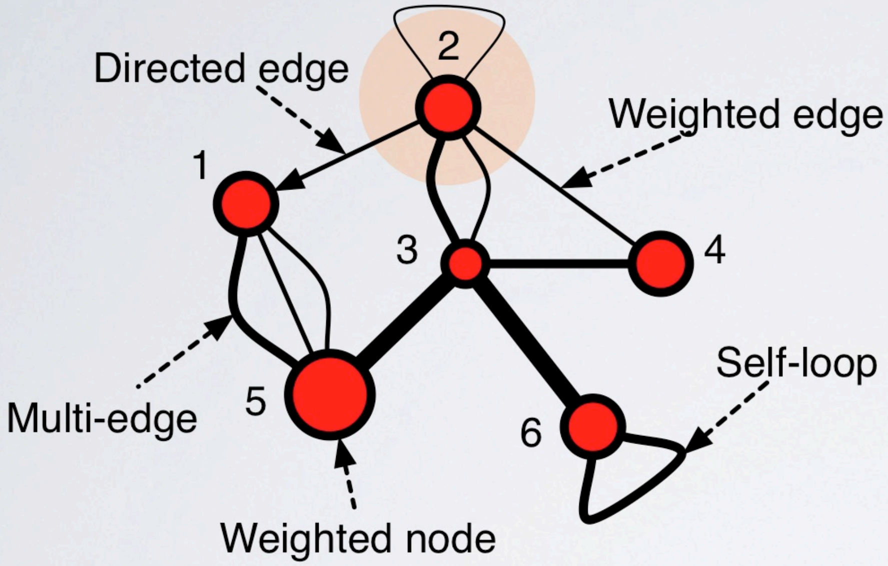
$3 \rightarrow \{2, 4, 5, 6\}$

$4 \rightarrow \{2, 3\}$

$5 \rightarrow \{1, 3\}$

$6 \rightarrow \{3\}$

representing networks – complex



adjacency matrix

A	1	2	3	4	5	6
1	0	0	0	0	$\{1,1,2\}$	0
2	1	$\frac{1}{2}$	$\{2,1\}$	1	0	0
3	0	$\{2,1\}$	0	2	4	4
4	0	1	2	0	0	0
5	$\{1,1,2\}$	0	4	0	0	0
6	0	0	4	0	0	2

adjacency list

A	
1	$\rightarrow \{(5,1),(5,1),(5,2)\}$
2	$\rightarrow \{(1,1),(2,\frac{1}{2}),(3,2),(3,1),(4,1)\}$
3	$\rightarrow \{(2,2),(2,1),(4,2),(5,4),(6,4)\}$
4	$\rightarrow \{(2,1),(3,2)\}$
5	$\rightarrow \{(1,1),(1,1),(1,2),(3,4)\}$
6	$\rightarrow \{(3,4),(6,2)\}$

representing networks – directed networks

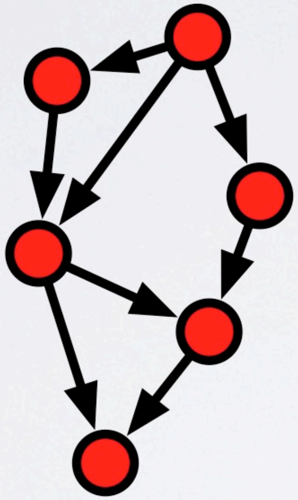
$$A_{ij} \neq A_{ji}$$

citation networks

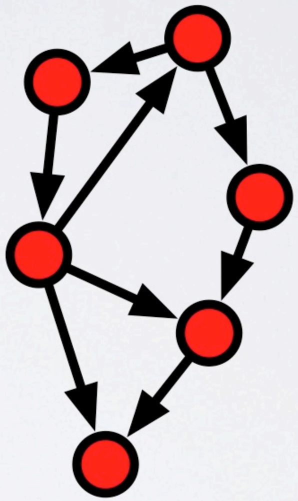
foodwebs*

epidemiological

others?



directed acyclic graph



directed graph

WWW

friendship?

flows of goods, information

economic exchange

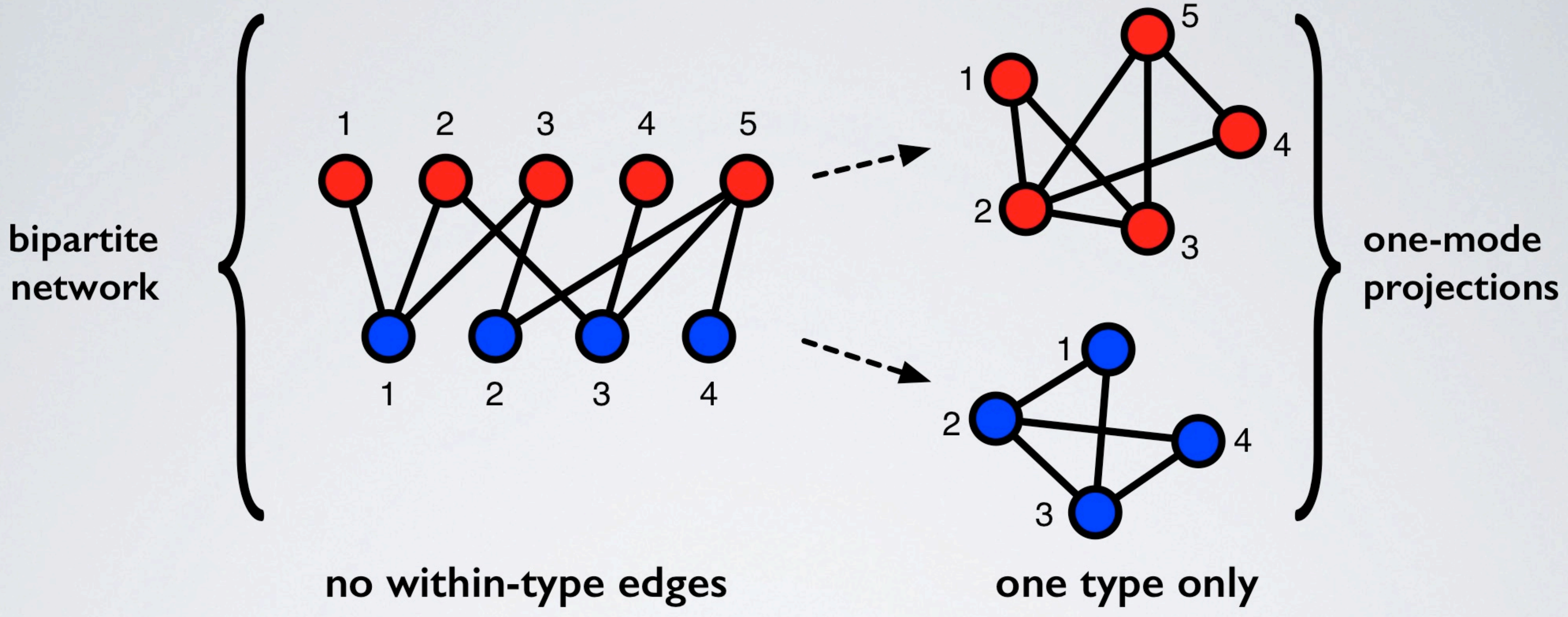
dominance

neuronal

transcription

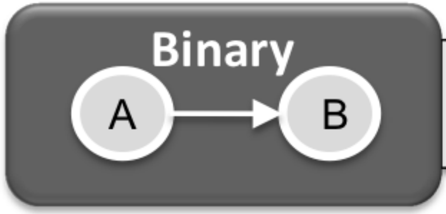
time travelers

representing networks – bipartite networks

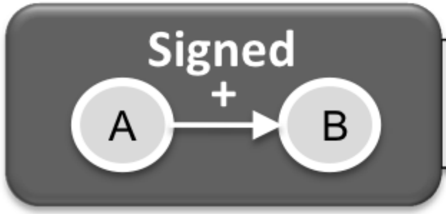


authors & papers	people & locations (checkins)
actors & movies/scenes	metabolites & reactions
musicians & albums	genes & substrings
people & online groups	words & documents
people & corporate boards	plants & pollinators

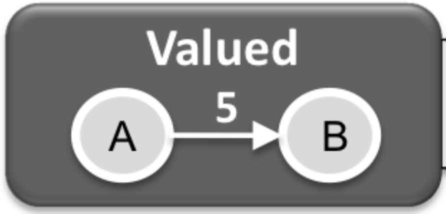
representing networks – link types



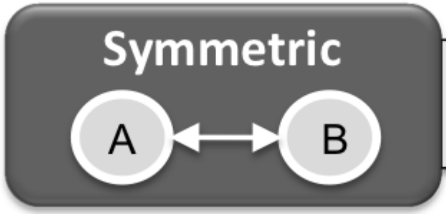
- **Link or no link** (1 or 0)



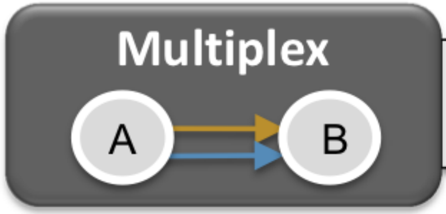
- **Positive or negative** (+, - or 0)



- **Weighted links** (each link is assigned a value)

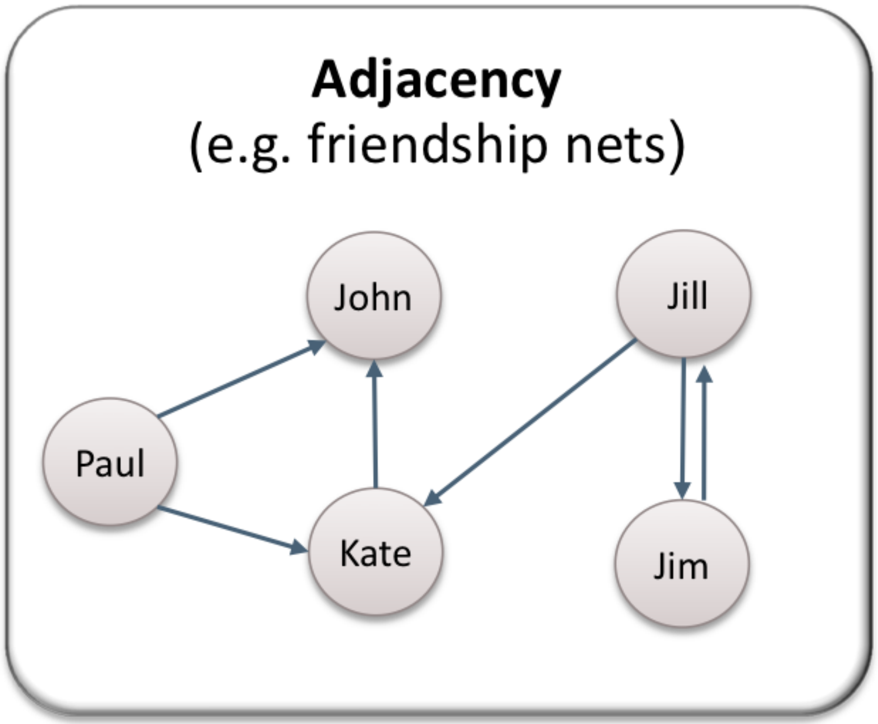


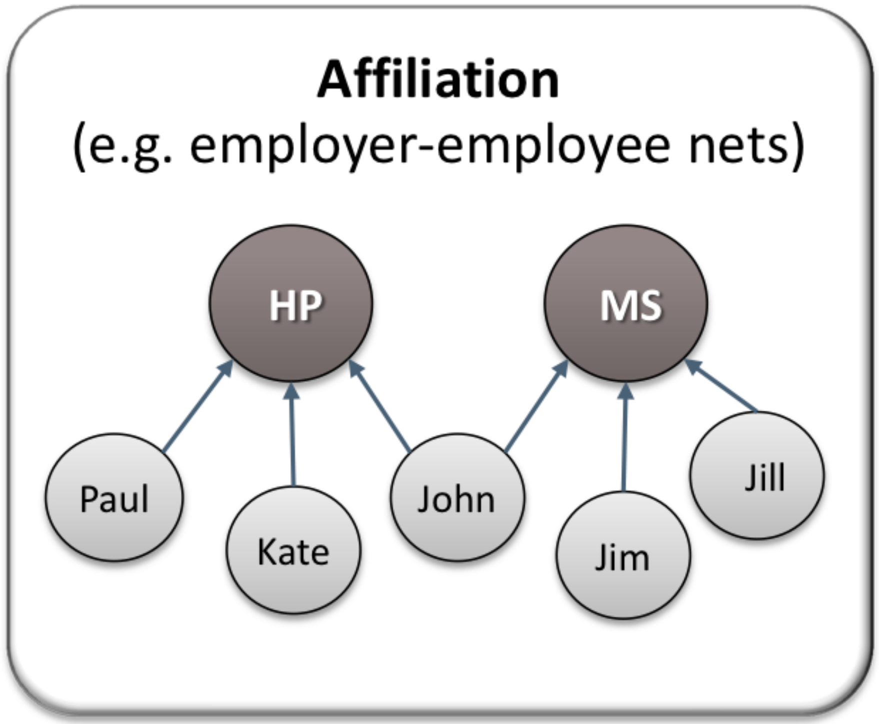
- **Directed vs. undirected** (undirected = symmetric)



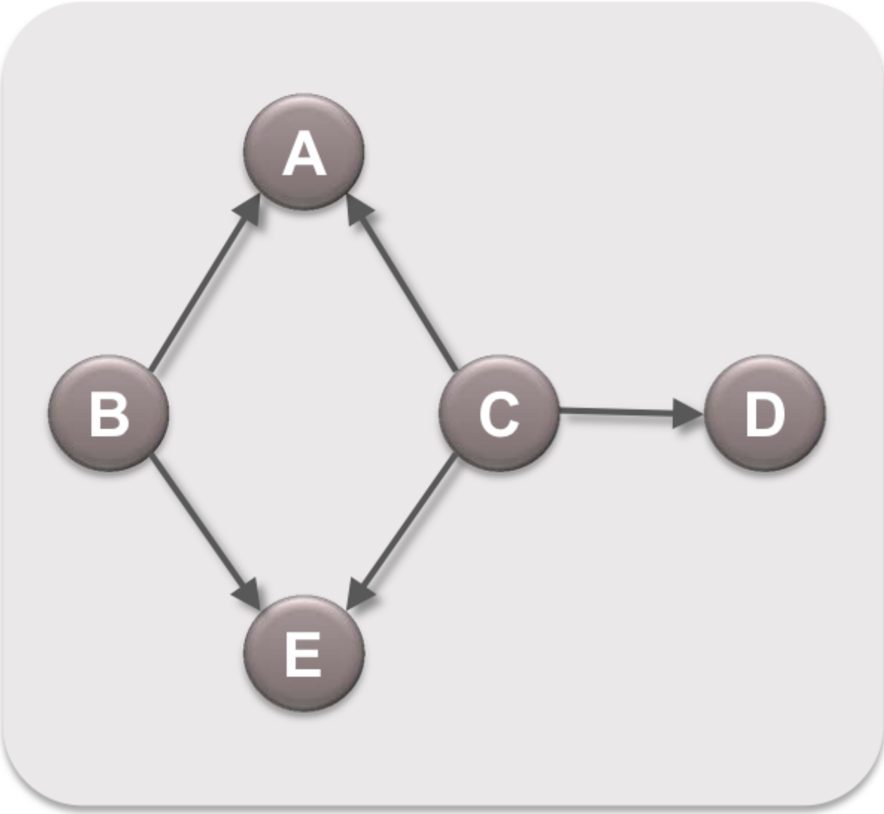
- **Multiplex** (more than one link type)

representing networks – network modes



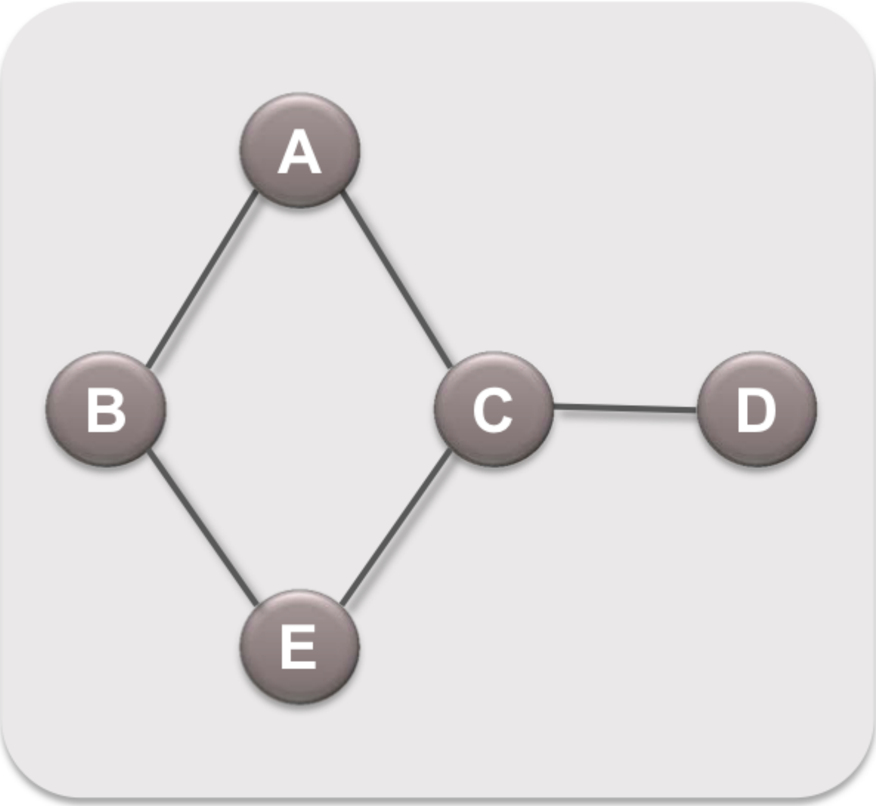


representing networks – directed networks



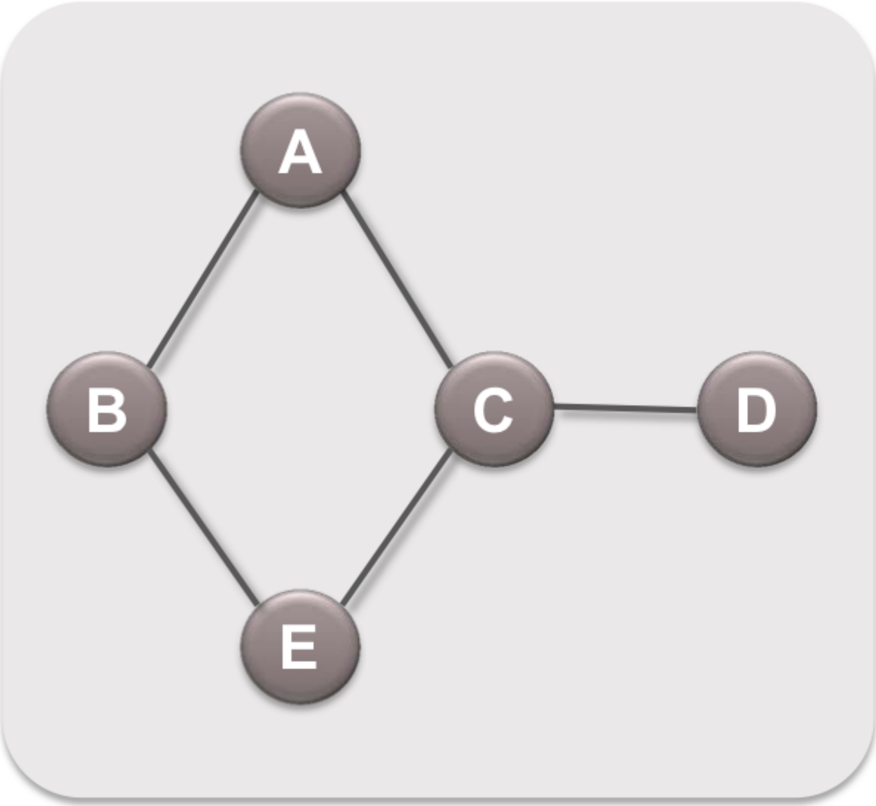
	A	B	C	D	E
A	0	0	0	0	0
B	1	0	0	0	1
C	1	0	0	1	1
D	0	0	0	0	0
E	0	0	0	0	0

representing networks – symmetric networks



	A	B	C	D	E
A	0	1	1	0	0
B	1	0	0	0	1
C	1	0	0	1	1
D	0	0	1	0	0
E	0	1	1	0	0

representing networks – affiliation networks



	A	B	C	D	E
A	0	1	1	0	0
B	1	0	0	0	1
C	1	0	0	1	1
D	0	0	1	0	0
E	0	1	1	0	0

Network Data

storing network data

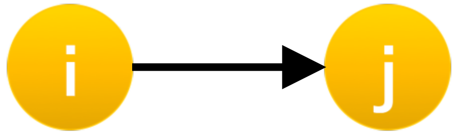
1. Adjacency matrix

2. Edgelist

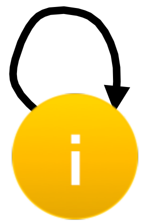
3. Adjacency/node list

1. Adjacency Matrix

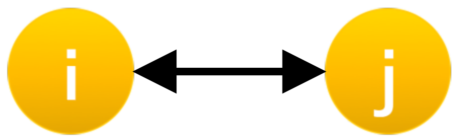
- Representing edges (who is adjacent to whom) as a matrix
  - A_{ij} = 1 if node *i* has an edge to node *j*
    = 0 if node *i* does not have an edge to j



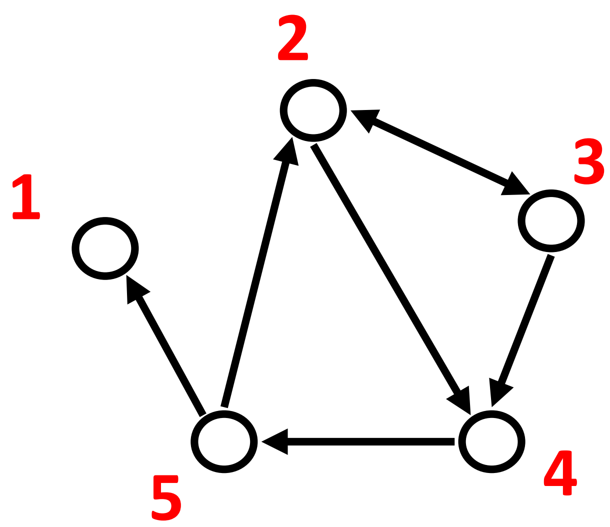
  - A_{ii} = 0 unless the network has self-loops



  - $A_{ij} = A_{ji}$ if the network is undirected, or if *i* and *j* share a reciprocated edge



1. Adjacency Matrix



$$A = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 \end{pmatrix}$$

Issues:

1. Your dataset will likely contain network data in a non-matrix format;
2. Large, sparse networks take way too much space if kept in a matrix format

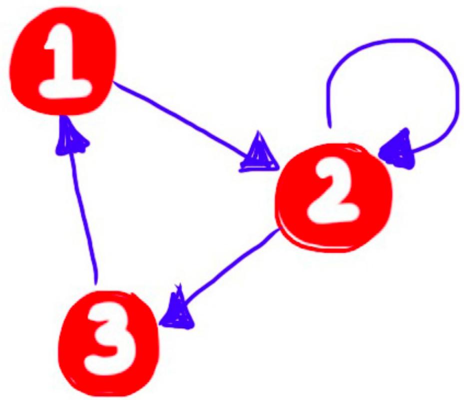
1. Adjacency Matrix

Which adjacency matrix represents this network?

A) $$\begin{bmatrix} 0 & 1 & 1 \\ 0 & 1 & 1 \\ 1 & 0 & 0 \end{bmatrix}$$

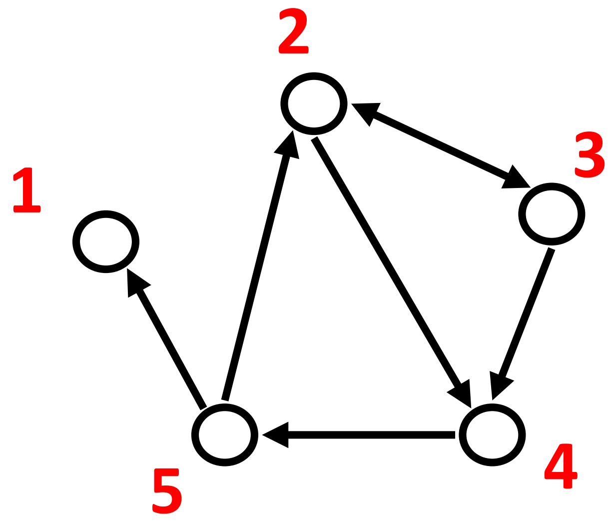
B) $$\begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$$

C) $$\begin{bmatrix} 0 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 0 \end{bmatrix}$$

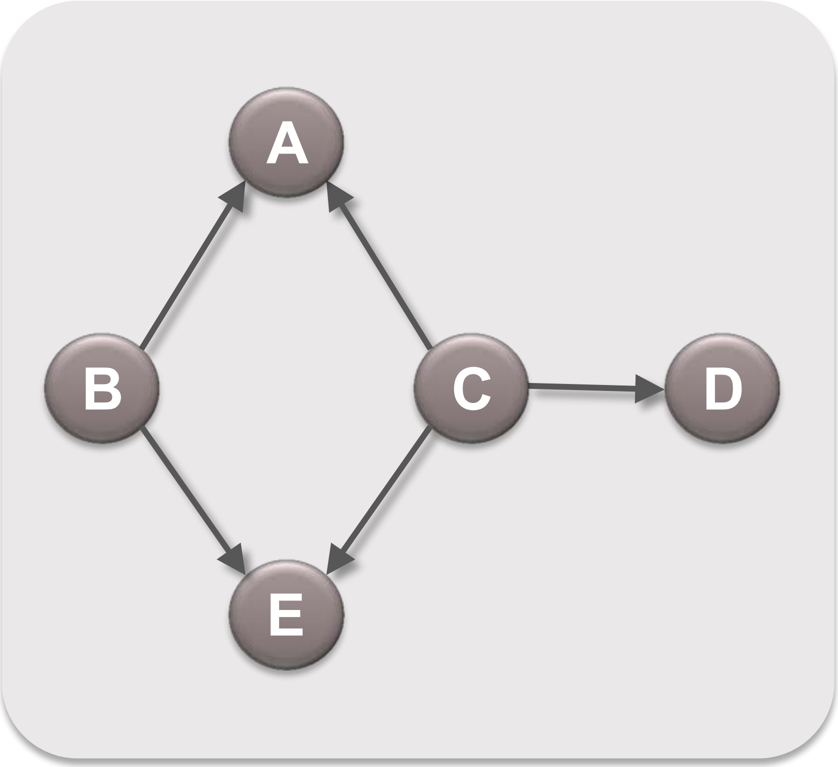


2. Edge list

- Edge list
  - 2, 3
  - 2, 4
  - 3, 2
  - 3, 4
  - 4, 5
  - 5, 2
  - 5, 1



2. Edge List (with weights)

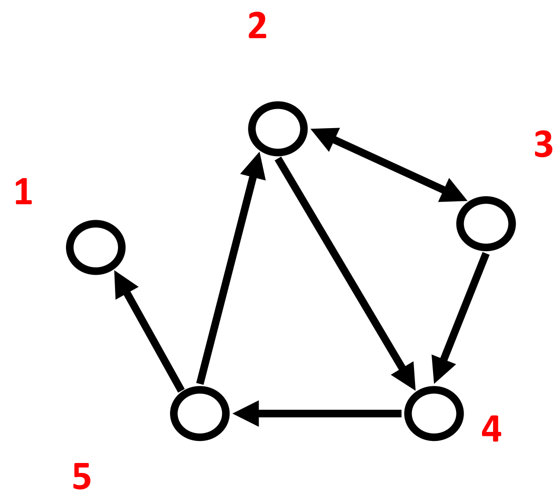


Source	Destination	Weight
B	A	1
B	E	1
C	A	1
C	E	1
C	D	1

Note: Weights are optional.

3. Adjacency list | Node list

- Adjacency list
  - is easier to work with if network is
    - large
    - sparse
  - quickly retrieve all neighbors for a node
    - 1:
    - 2: 3 4
    - 3: 2 4
    - 4: 5
    - 5: 1 2



Matrix Algebra

Matrix Algebra

- Matrix Concepts, Notation & Terminologies
- Adjacency Matrices
- Transposes
- Matrix Operations

Matrices

- Symbolized by a capital letter, like A
- Each cell in the matrix identified by row and column subscripts: a ij
    - First subscript is row, second is column

ID	Age	Gender	Income
Mary	a_11		
Bill			
John		a_32	
Larry			

Vectors

- Each row and each column in a matrix is a vector
- – Vertical vectors are column vectors, horizontal are row vectors
- Denoted by lowercase bold letter: **y**
- Each cell in the vector identified by subscript $\boldsymbol{x}_i$

Ways and Modes

- Ways are the dimensions of a matrix.
- Modes are the sets of entities indexed by the ways of a matrix

	Event 1	Event 2	Event 3	Event 4
EVELYN	1	1	1	1
LAURA	1	1	1	0
THERESA	0	1	1	1
BRENDA	1	0	1	1
CHARLO	0	0	1	1
FRANCES	0	0	1	0
ELEANOR	0	0	0	0
PEARL	0	0	0	0
RUTH	0	0	0	0
VERNE	0	0	0	0
MYRNA	0	0	0	0

2-way, 2-mode

	Mary	Bill	John	Larry
Mary	0	1	0	1
Bill	1	0	0	1
John	0	1	0	0
Larry	1	0	1	0

2-way, 1-mode

Proximity Matrices

- Proximity Matrices record “degree of proximity”.
- Proximities are usually among a single set of actor (hence, they are 1-mode), but they are not limited to 1s and 0s in the data.
- What constitutes the *proximity* is user-defined.
    - Driving distances are one form of proximities, other forms might be number of friends in common, time spent together, number of emails exchanged, or a measure of similarity in cognitive structures.

Proximity Matrices

- Proximity matrices can contain either *similarity* or *distance* (or *dissimilarity*) data.
    - Similarity data, such as number of friends in common or correlations, means a larger number represents more similarity or greater proximity
    - Distance (or dissimilarity data) such as physical distance means a larger number represents more dissimilarity or less proximity

Transposes

- The transpose M' of a matrix M is the matrix flipped on its side.
  - The rows become columns and the columns become rows
  - So the transpose of an m by n matrix is an n by m matrix.

Transpose Example

M	Tennis	Football	Rugby	Golf
Mike	0	0	1	0
Ron	0	1	1	0
Pat	0	0	0	1
Bill	1	1	1	1
Joe	0	0	0	0
Rich	0	1	1	1
Peg	1	1	0	1

M^T	Mike	Ron	Pat	Bill	Joe	Rich	Peg
Tennis	0	0	0	1	0	0	1
Football	0	1	0	1	0	1	1
Rugby	1	1	0	1	0	1	0
Golf	0	0	1	1	0	1	1

Dichotomizing

- X is a valued matrix, say 1 to 10 rating of strength of tie
- Construct a matrix Y of ones and zeros s.t. $y_{ij} = 1$ if $x_{ij} > 5$, and $y_{ij} = 0$ otherwise

	EVE	**LAU**	**THE**	**BRE**	**CHA**
EVELYN	8	6	7	6	3
LAURA	6	7	6	6	3
THERESA	7	6	8	6	4
BRENDA	6	6	6	7	4
CHARLOTTE	3	3	4	4	4

	EVE	LAU	THE	BRE	CHA
EVELYN	1	1	1	1	0
LAURA	1	1	1	1	0
THERESA	1	1	1	1	0
BRENDA	1	1	1	1	0
CHARLOTTE	0	0	0	0	0

Symmetrizing

- When matrix is not symmetric, i.e., $x_{ij} \neq x_{ji}$
- Symmetrize various ways. Set y_{ij} and y_{ji} to:
  - Maximum(x_ij, x_ji): union rule;
  - Minimum(x_ij, x_ji): intersection rule;
  - Average (x_ij+x_ji)/2
  - Lowerhalf: choose x_{ij} when $i > j$ and x_{ji} otherwise

Symmetrizing Example

What rule are we using here?

	ROM	BON	AMB	BER	PET	LOU
ROMUL_10	0	1	1	0	3	0
BONAVEN_5	0	0	1	0	3	2
AMBROSE_9	0	1	0	0	0	0
BERTH_6	0	1	2	0	3	0
PETER_4	0	3	0	1	0	2
LOUIS_11	0	2	0	0	0	0

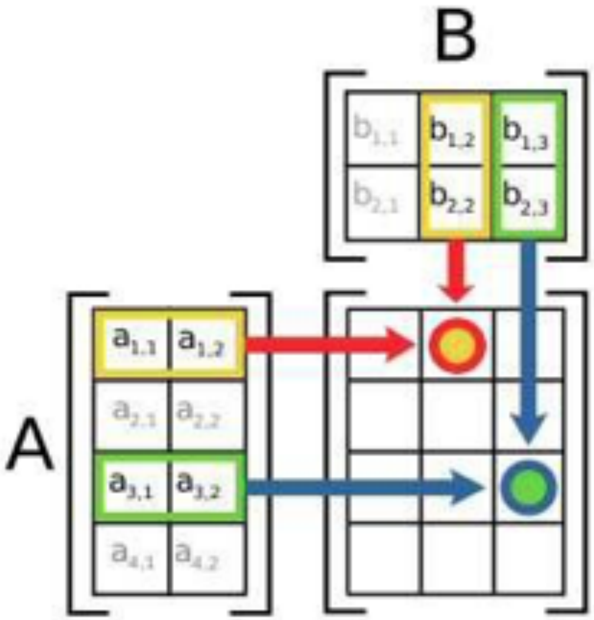
	ROM	BON	AMB	BER	PET	LOU
ROMUL_10	0	1	1	0	3	0
BONAVEN_5	1	0	1	1	3	2
AMBROSE_9	1	1	0	2	0	0
BERTH_6	0	1	2	0	3	0
PETER_4	3	3	0	3	0	2
LOUIS_11	0	2	0	0	2	0

Matrix Multiplication

- Matrix products are not generally commutative (i.e., AB does not usually equal BA)
- Notation: $C = AB$
- only possible when the number of columns in A equals number of rows in B; these are said to be comformable. It is calculated as:

$$c_{ij} = \sum a_{ik} * b_{kj} \quad \forall k$$

Matrix multiplication example i



$$\begin{bmatrix} 1 & 0 & 2 \\ -1 & 3 & 1 \end{bmatrix} \cdot \begin{bmatrix} 3 & 1 \\ 2 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 \times 3 + 0 \times 2 + 2 \times 1 & 1 \times 1 + 0 \times 1 + 2 \times 0 \\ -1 \times 3 + 3 \times 2 + 1 \times 1 & -1 \times 1 + 3 \times 1 + 1 \times 0 \end{bmatrix} = \begin{bmatrix} 5 & 1 \\ 4 & 2 \end{bmatrix}$$

Matrix multiplication example ii

Skills	Math	Verbal	Analytic
Kev	1.00	.75	.80
Jeff	.80	.80	.90
Lisa	.75	.60	.75
Kim	.80	1.00	.85

Items	Q1	Q2	Q3	Q4
Math	.50	.75	0	.1
Verbal	.10	0	.9	.1
Analytic	.40	.25	.1	.8

- Given a Skills and Items matrix calculate the "affinity" that each person has for each question
- Kev for Question 1 is:
  = 1.00 * .5 + .75* .1 + .80 * .40
  = .5 + .075 + .32 = **0.895**
- Lisa for Question 3 is:
  = .75 * .0 + .60* .90 + .75 * .1
  = .0 + .54 + .075 = **0.615**

Affin	Q1	Q2	Q3	Q4
Kev	0.895	0.95	0.755	0.815
Jeff	0.840	0.825	0.810	0.880
Lisa	0.735	0.75	0.615	0.735
Kim	0.840	0.813	0.985	0.860

Assessing node's environment

X

	a	**b**	**c**	**d**	**e**	**f**
a	0	1	0	1	1	1
b	0	0	1	0	0	0
c	1	1	0	1	0	0
d	0	1	1	0	1	1
e	1	0	0	0	0	0
f	1	1	0	0	1	0

A

	hrs	**$**	**lib**
a	3	50	1
b	9	10	4
c	3	5	3
d	7	30	5
e	1	20	2
f	5	5	4

XA

	hrs	**$**	**lib**
a	22	65	15
b	3	5	3
c	19	90	10
d	18	40	13
e	3	50	1
f	13	80	7

- Hrs and $ columns of XA give social access to resources
- Lib column gives how liberal the person's social environment is

Boolean matrix multiplication

- Values can be 0 or 1 for all matrices
- Products are dichotomized

Would have been a 2 in regular matrix multiplication

	Mary	Bill	John	Larry
Mary	0	1	0	1
Bill	1	0	1	0
John	0	0	0	1
Larry	0	0	0	0

A

	Mary	Bill	John	Larry
Mary	0	0	1	1
Bill	1	0	1	0
John	0	0	0	1
Larry	0	1	0	0

B

	Mary	Bill	John	Larry
Mary	1	1	1	0
Bill	0	0	0	1
John	0	1	0	0
Larry	0	0	0	0

AB

Composition of relations

- We represent each social relation (e.g., F= friend of, B = boss of) as a matrix
- To create the compound relation friend of the boss of (FB), we just multiply the two matrices

F

	a	b	c	d
a	0	1	0	1
b	1	0	1	0
c	1	1	0	1
d	1	0	1	0

x

B

	a	b	c	d
a	0	0	1	1
b	1	0	0	0
c	0	1	0	0
d	0	0	0	0

=

FB

	a	b	c	d
a	1	0	0	0
b	0	1	1	1
c	1	0	1	1
d	0	1	1	1

Composition of relations

Hard for *a* to borrow any subordinates

F

	a	b	c	d
a	0	1	0	1
b	1	0	1	0
c	1	1	0	1
d	1	0	1	0

x

B

	a	b	c	d
a	0	0	1	1
b	1	0	0	0
c	0	1	0	0
d	0	0	0	0

=

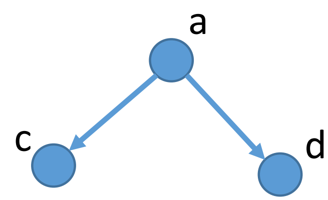
FB

	a	b	c	d
a	1	0	0	0
b	0	1	1	1
c	1	0	1	1
d	0	1	1	1

Everyone is friends with their boss

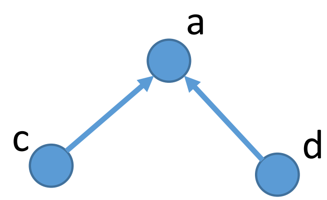
- FB(c,a) = 1 (or cFBa) means that person c is friend of someone (namely b) who is the boss of a. i.e., c is friends with a's boss
- FB(a,a) = 1 (or aFBa) means person a is friends with someone (b again) who is a's boss. i.e., *a* is friends with her boss
- FB(b,d) = 1, so person b is friends with someone (a) who is the boss of d

Converse of a relation



“boss of”

- In relational terms, the converse of a relation is the reciprocal role
  - Converse of “boss of” is “subordinate of”
- In graph terms, we are just reversing the direction of arrows
- In matrix terms, we are transposing matrix
  - Construct B’ (reports to) from B (is the boss of)



“reports to”

B

	a	b	c	d
a	0	0	1	1
b	1	0	0	0
c	0	1	0	0
d	0	0	0	0

Boss of

B'

	a	b	c	d
a	0	1	0	0
b	0	0	1	0
c	1	0	0	0
d	1	0	0	0

Reports to

To transpose a matrix, write each row as a column

Composition of relations – with converse

- To create the compound relation friend of the subordinate of (FB’), we just post-multiply F by the transpose of B
- FB’(c,a) = 1 (or cFB’a) means that person c is friend of someone (namely *d*) who is a subordinate of *a*. i.e., c is friends with *a*’s subordinate
- FB’(a,a) = 1 (or aFB’a) means person *a* is friends with someone (*d*) who is her subordinate. i.e., *a* is friends with one of her direct reports.

F

	a	b	c	d
a	0	1	0	1
b	1	0	1	0
c	1	1	0	1
d	1	0	1	0

x

B'

	a	b	c	d
a	0	1	0	0
b	0	0	1	0
c	1	0	0	0
d	1	0	0	0

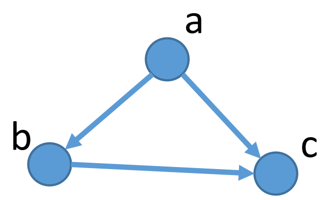
=

FB’

	a	b	c	d
a	1	0	1	0
b	1	1	0	0
c	1	1	1	0
d	1	1	0	0

Everybody likes a’s subordinates

Transitivity



- L = "likes someone", uLLv means u likes someone who likes v

L

	a	b	c	d
a	0	1	0	1
b	0	0	1	0
c	0	1	0	0
d	1	0	1	0

x

L

	a	b	c	d
a	0	1	0	1
b	0	0	1	0
c	0	1	0	0
d	1	0	1	0

=

LL

	a	b	c	d
a	1	0	2	0
b	0	1	0	0
c	0	0	1	0
d	0	2	0	1

Note diagonal of LL is all 1s, so everyone is lucky enough to like someone who likes them

LL(d,b) = 2 indicates d likes 2 people who like b

- If *a* likes b and b likes c, does that mean *a* likes c?
- If matrix L = matrix LL, then L is a transitive relation, in keeping with balance theory

Products of matrices & their transposes

- XX' = product of matrix X by its transpose

$$(XX')_{ij} = \sum_k x_{ik} x_{jk}$$

- Computes sums of products of each pair of rows (cross-products)
- Similarities among rows

	1	2	3	4
Mary	0	1	1	1
Bill	1	0	1	0
John	0	0	0	1
Larry	0	0	0	0
Tina	1	1	1	0

X

	Mary	Bill	John	Larry	Tina
1	0	1	0	0	1
2	1	0	0	0	1
3	1	1	0	0	0
4	1	0	1	0	0

X'

	Mary	Bill	John	Larry	Tina
Mary	3	1	1	0	1
Bill	1	2	0	0	1
John	1	0	1	0	0
Larry	0	0	0	0	0
Tina	2	2	0	0	2

XX'

Multiplying a matrix by its transpose

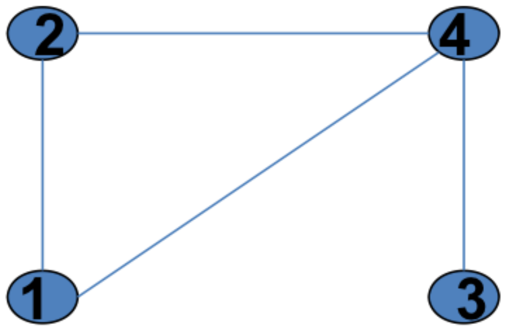
	E1	E2	E3	E4	E5	E6	E7	E8	E9	E10	E11	E12	E13	E14
EVELYN	1	1	1	1	1	1	0	1	1	0	0	0	0	0
LAURA	1	1	1	0	1	1	1	1	0	0	0	0	0	0
THERESA	0	1	1	1	1	1	1	1	1	0	0	0	0	0
BRENDA	1	0	1	1	1	1	1	1	0	0	0	0	0	0
CHARLOTTE	0	0	1	1	1	0	1	0	0	0	0	0	0	0
FRANCES	0	0	1	0	1	1	0	1	0	0	0	0	0	0
ELEANOR	0	0	0	0	1	1	1	1	0	0	0	0	0	0
PEARL	0	0	0	0	0	1	0	1	1	0	0	0	0	0
RUTH	0	0	0	0	1	0	1	1	1	0	0	0	0	0
VERNE	0	0	0	0	0	0	1	1	1	0	0	1	0	0
MYRNA	0	0	0	0	0	0	0	1	1	1	0	1	0	0
KATHERINE	0	0	0	0	0	0	0	1	1	1	0	1	1	1
SYLVIA	0	0	0	0	0	0	1	1	1	1	0	1	1	1
NORA	0	0	0	0	0	1	1	0	1	1	1	1	1	1
HELEN	0	0	0	0	0	0	1	1	0	1	1	1	0	0
DOROTHY	0	0	0	0	0	0	0	1	1	0	0	0	0	0
OLIVIA	0	0	0	0	0	0	0	0	1	0	1	0	0	0
FLORA	0	0	0	0	0	0	0	0	1	0	1	0	0	0

	EV	LA	TH	BR	CH	FR	EL	PE	RU	VE	MY	KA	SY	NO	HE	DO	OL	FL
E1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
E2	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
E3	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0
E4	1	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
E5	1	1	1	1	1	1	1	0	1	0	0	0	0	0	0	0	0	0
E6	1	1	1	1	0	1	1	1	0	0	0	0	0	1	0	0	0	0
E7	0	1	1	1	1	0	1	0	1	1	0	0	1	1	1	0	0	0
E8	1	1	1	1	0	1	1	1	1	1	1	1	1	0	1	1	0	0
E9	1	0	1	0	0	0	0	1	1	1	1	1	1	1	0	1	1	1
E10	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	0	0	0
E11	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	1	1
E12	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	0	0	0
E13	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0
E14	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0

	EVE	LAU	THE	BRE	CHA	FRA	ELE	PEA	RUT	VER	MYR	KAT	SYL	NOR	HEL	DOR	OLI	FLO
EVELYN	8	6	7	6	3	4	3	3	3	2	2	2	2	2	1	2	1	1
LAURA	6	7	6	6	3	4	4	2	3	2	1	1	2	2	2	1	0	0
THERESA	7	6	8	6	4	4	4	3	4	3	2	2	3	3	2	2	1	1
BRENDA	6	6	6	7	4	4	4	2	3	2	1	1	2	2	2	1	0	0
CHARLOTTE	3	3	4	4	4	2	2	0	2	1	0	0	1	1	1	0	0	0
FRANCES	4	4	4	4	2	4	3	2	2	1	1	1	1	1	1	1	0	0
ELEANOR	3	4	4	4	2	3	4	2	3	2	1	1	2	2	2	1	0	0
PEARL	3	2	3	2	0	2	2	3	2	2	2	2	2	2	1	2	1	1
RUTH	3	3	4	3	2	2	3	2	4	3	2	2	3	2	2	2	1	1
VERNE	2	2	3	2	1	1	2	2	3	4	3	3	4	3	3	2	1	1
MYRNA	2	1	2	1	0	1	1	2	2	3	4	4	4	3	3	2	1	1
KATHERINE	2	1	2	1	0	1	1	2	2	3	4	6	6	5	3	2	1	1
SYLVIA	2	2	3	2	1	1	2	2	3	4	4	6	7	6	4	2	1	1
NORA	2	2	3	2	1	1	2	2	2	3	3	5	6	8	4	1	2	2
HELEN	1	2	2	2	1	1	2	1	2	3	3	3	4	4	5	1	1	1
DOROTHY	2	1	2	1	0	1	1	2	2	2	2	2	2	1	1	2	1	1
OLIVIA	1	0	1	0	0	0	0	1	1	1	1	1	1	2	1	1	2	2
FLORA	1	0	1	0	0	0	0	1	1	1	1	1	1	2	1	1	2	2

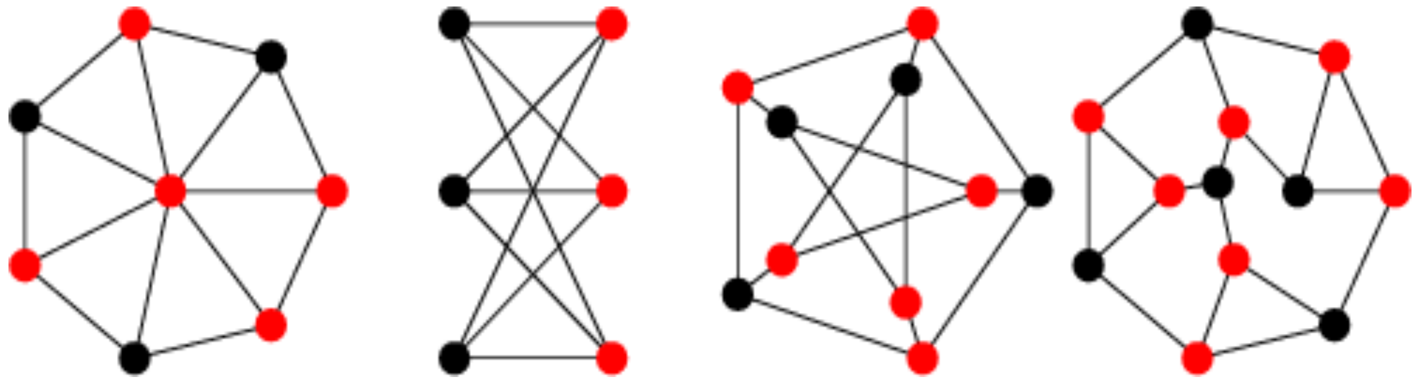
squaring an adjacency matrix

$$g = \begin{pmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix}$$



$$g^2 = \begin{pmatrix} 2 & 1 & 1 & 1 \\ 1 & 2 & 1 & 1 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 3 \end{pmatrix}$$

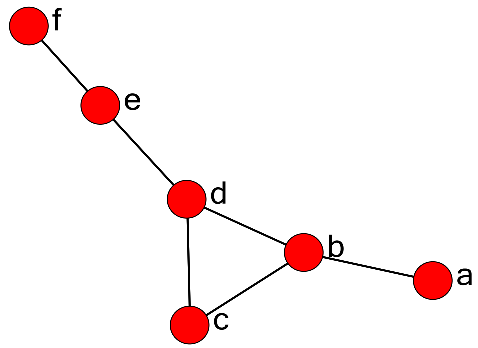
number of walks of length 2 from i to j



Graph Theoretic Concepts

Intro to graph terminology

- Nodes
  - Aka vertices or points in more mathematical work
  - Actors, agents, egos, alters, contacts in more sociological work
  - Nodes can individuals or collective actors, such as countries
  - In social network analysis, nodes typically have agency
- Ties
  - Aka edges, arcs or lines in more technical work
  - Links, bonds, direct connections etc in more sociological work
  - Ties are typically binary: they link exactly two nodes

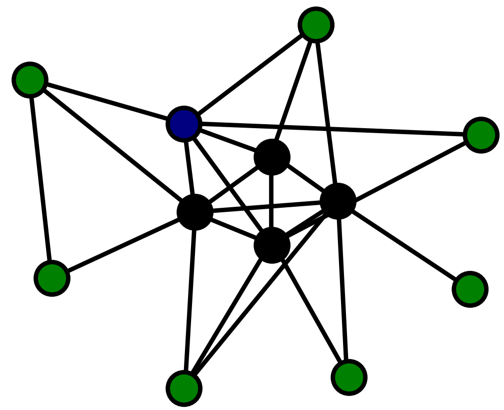


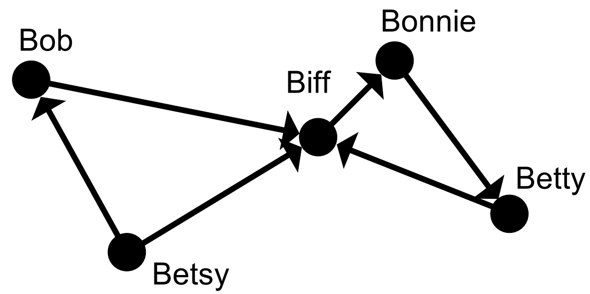
A graph

- G(V,E) is …
- A set of vertices V, together with …
- A set of edges E
- The edges are binary, meaning they have exactly two endpoints
  - They are 2-tuples
- If the edges are k-tuples (where $k > 2$), they comprise a hyper-graph

Directed and undirected graphs

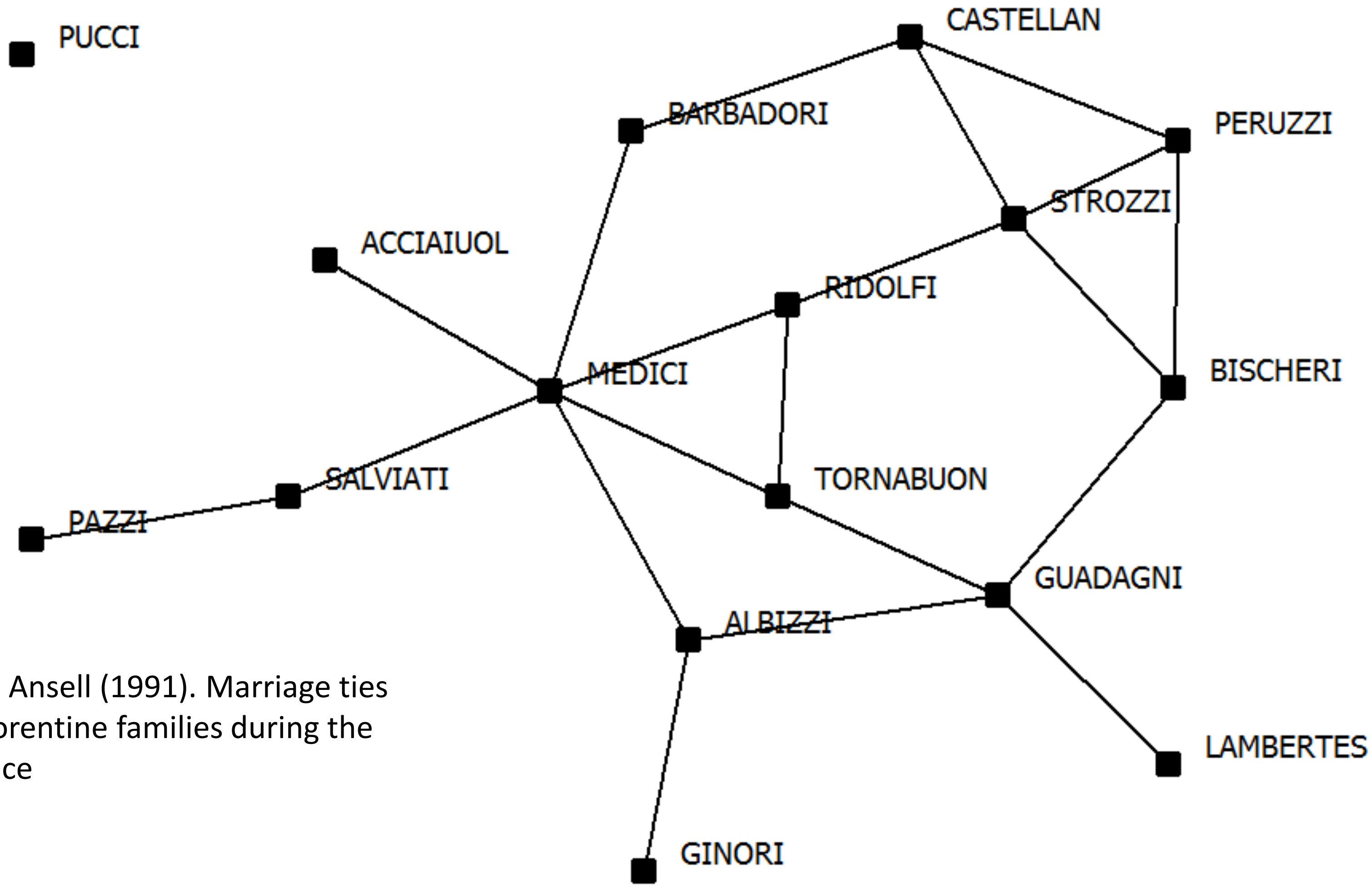
- Graphs can be directed or undirected*
- Undirected
  - In an undirected graph, the ties don't have direction – two nodes u and v are connected by a tie, but it doesn't matter whether you say u has tie to v or v has tie to u.
    - E.g., married, taking same class, siblings
  - The ties are called edges
- Directed
  - Ties (which are called arcs) have direction. If u has a tie to v, it may or may not be true that v has a tie to u
    - Gives advice to; sends an email to; thinks well of
  - Directed graphs often called digraphs
- An undirected graph is <u>like</u> a directed graph in which all arcs are reciprocated, but technically there is a difference
  - In an undirected graph, non-reciprocity is impossible/insensible





*But in some usages graph refers to both, m to the species as a whole, while other times it contrasts with 'woman'

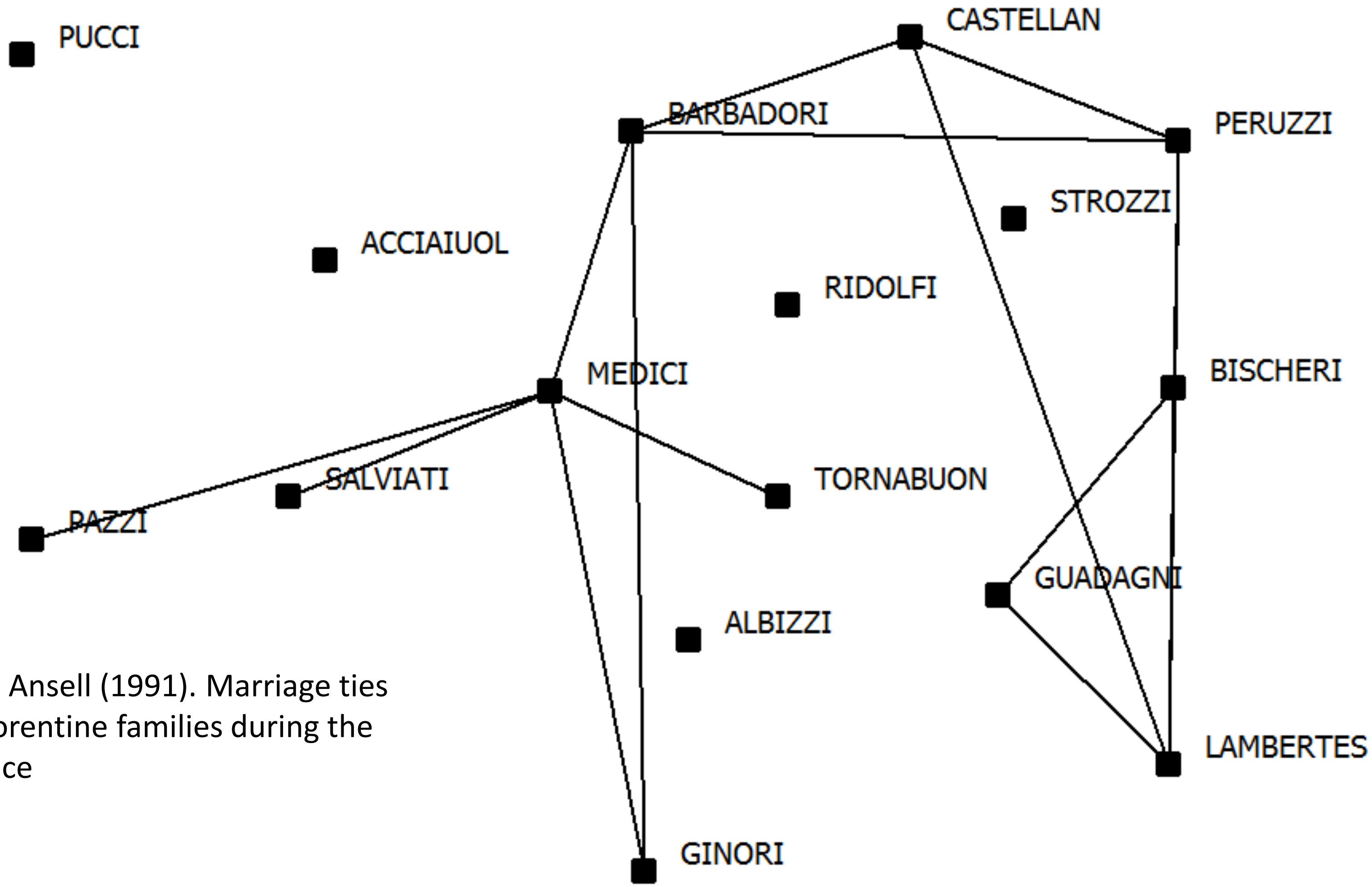
Marriage ties between families



Padgett & Ansell (1991). Marriage ties among Florentine families during the Renaissance

Business ties between families

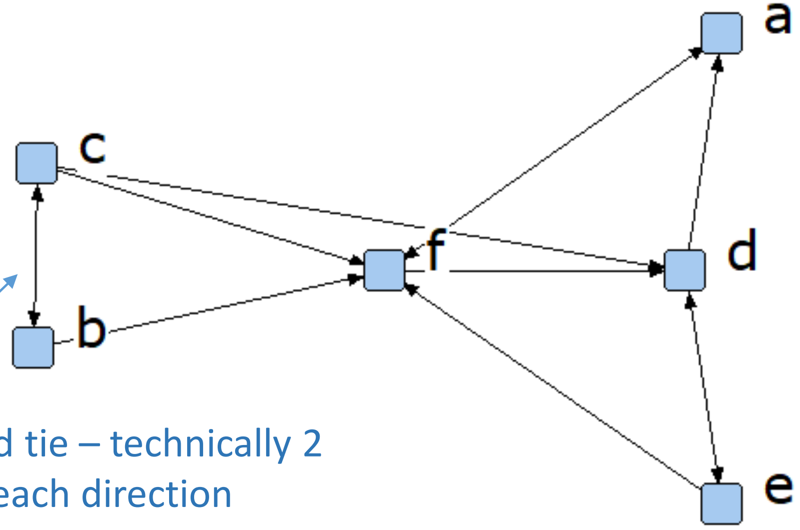
->draw padgett



Padgett & Ansell (1991). Marriage ties among Florentine families during the Renaissance

Directed networks

- In directed graphs, ties have direction, and need not be reciprocated
- Adjacency matrix is not symmetric



Reciprocated tie – technically 2 ties, one in each direction

	a	b	c	d	e	f
a	0	0	0	0	0	1
b	0	0	1	0	0	1
c	0	1	0	1	0	1
d	1	0	0	0	1	0
e	0	0	0	1	0	1
f	1	0	0	1	0	0

Consider “likes” and “seeks advice from”

Transpose Adjacency matrix

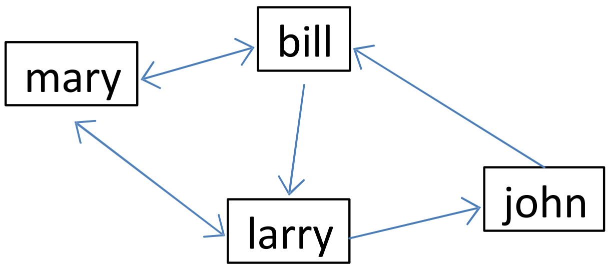
- In directed graphs, interchanging rows/columns of adjacency matrix effectively reverses the direction & meaning of ties

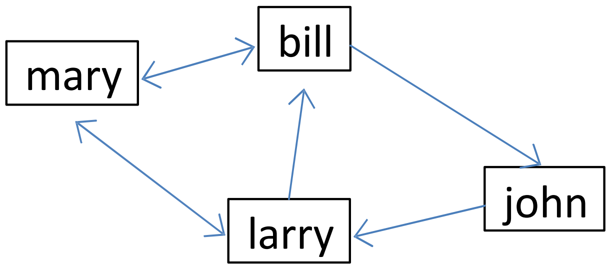
	Mary	Bill	John	Larry
Mary	0	1	0	1
Bill	1	0	0	1
John	0	1	0	0
Larry	1	0	1	0

Gives money to

	Mary	Bill	John	Larry
Mary	0	1	0	1
Bill	1	0	1	0
John	0	0	0	1
Larry	1	1	0	0

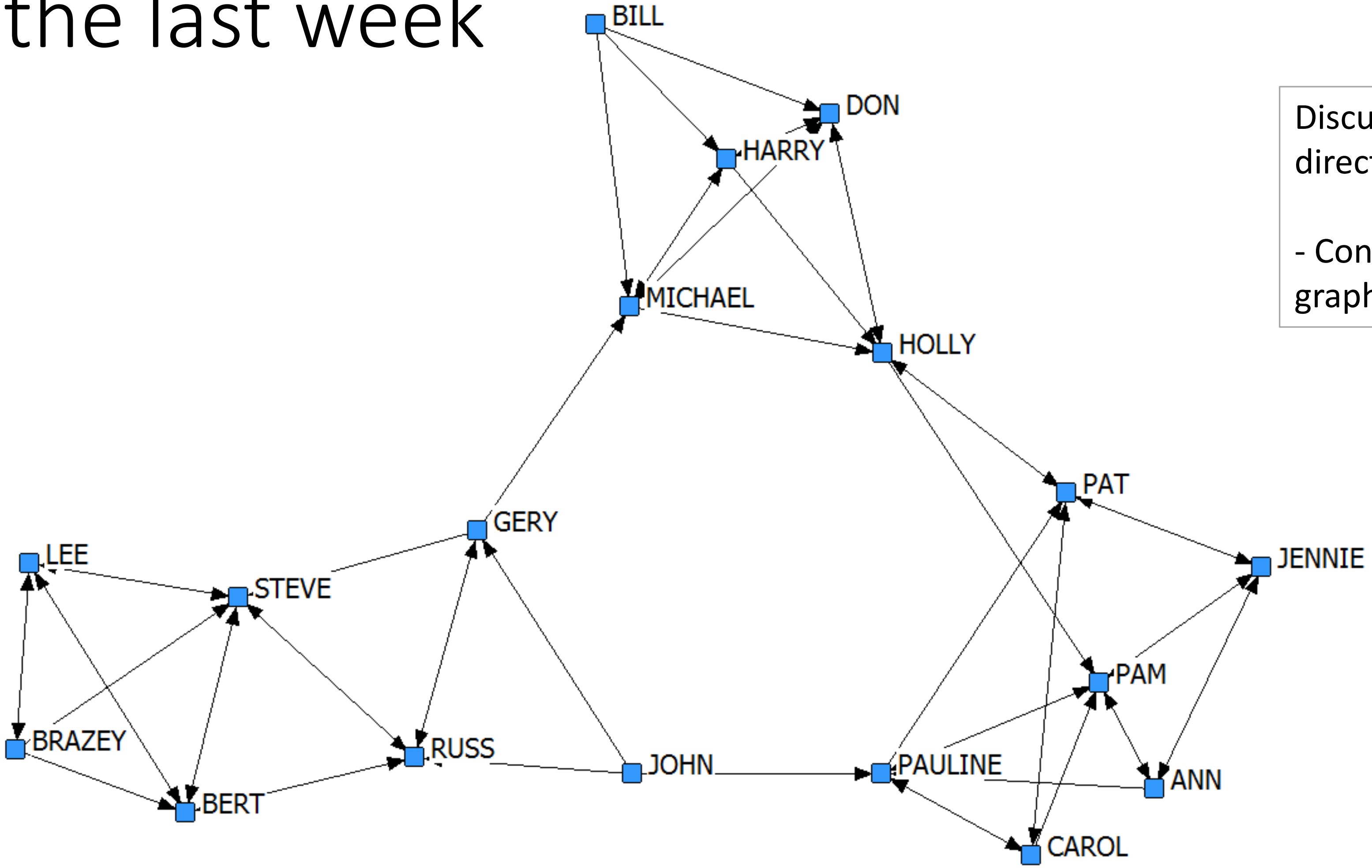
Gets money from





How meaningful are arrows?

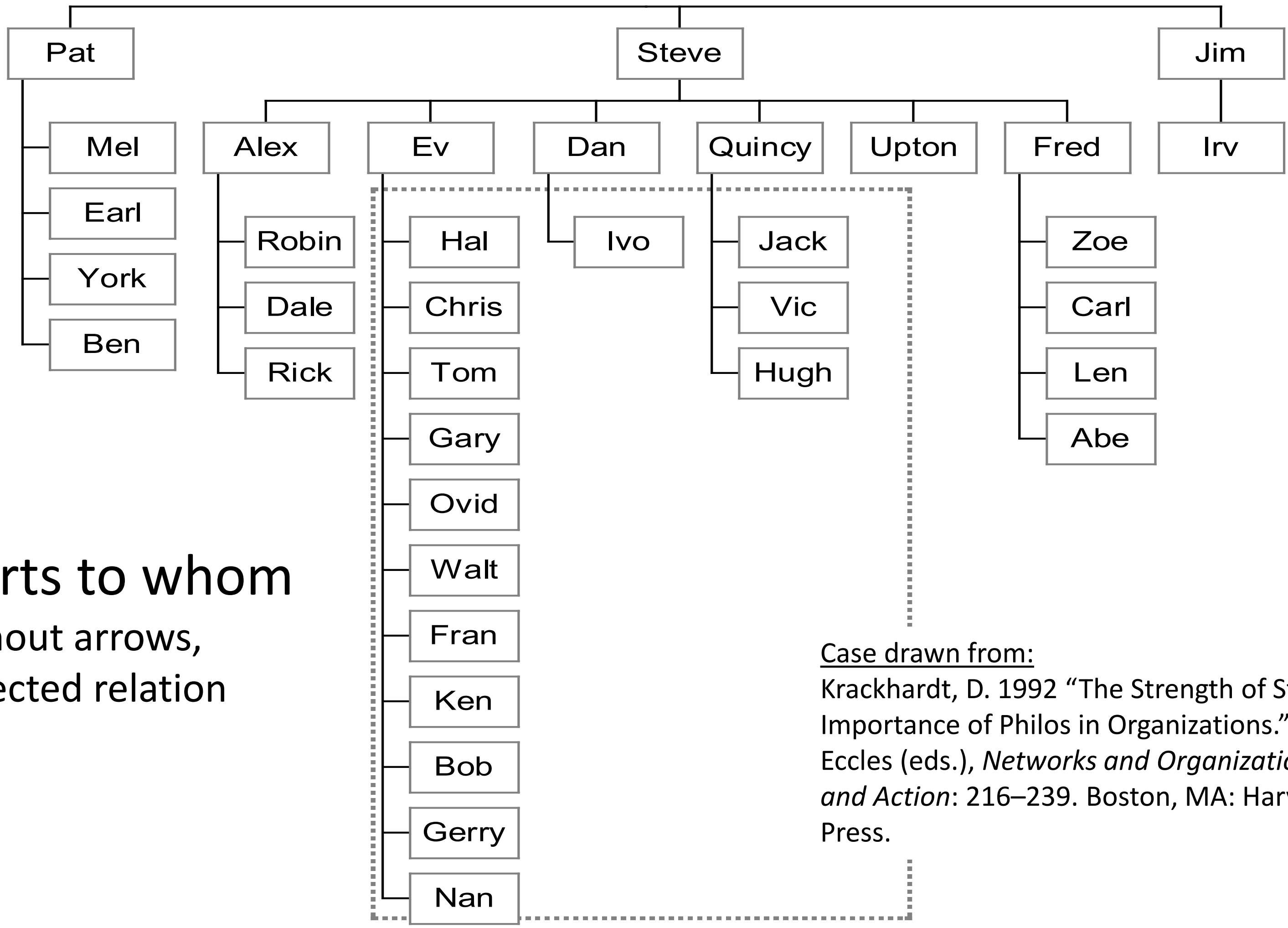
The 3 people you interacted with the most over the last week



Discuss reversing direction

- Converse of the graph

->Draw krack-high-tec



Who reports to whom

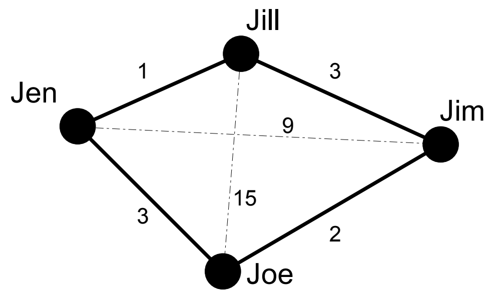
- Drawn without arrows, but is a directed relation

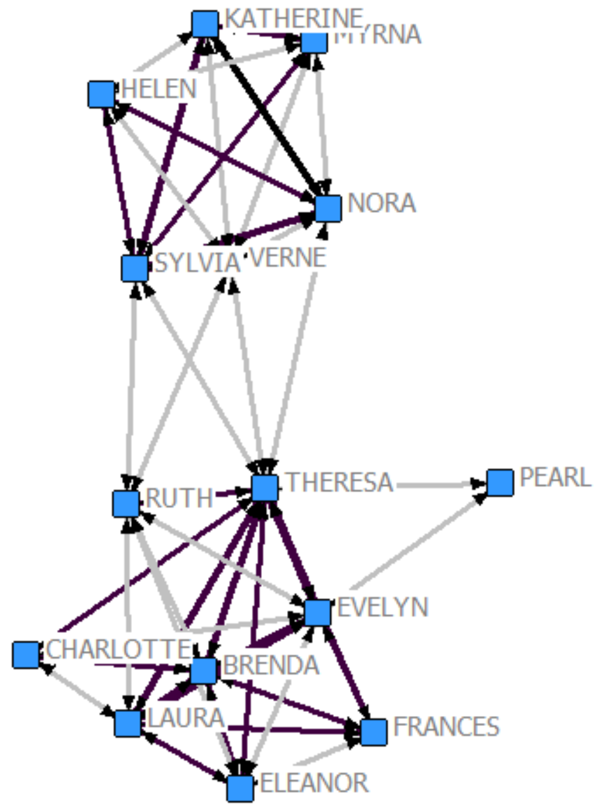
Case drawn from:

Krackhardt, D. 1992 “The Strength of Strong Ties: The Importance of Philos in Organizations.” In N. Nohria & R. Eccles (eds.), *Networks and Organizations: Structure, Form, and Action*: 216–239. Boston, MA: Harvard Business School Press.

Valued networks

- We can attach values to ties, representing quantitative properties of the relationship
- G(V,E,F), where F is a function delivering real values
  - Strength of relationship
  - Information capacity of tie
  - Rates of flow or traffic across tie
  - Distances between nodes
  - Probabilities of passing on information
  - Frequency of interaction





Valued Adjacency Matrix

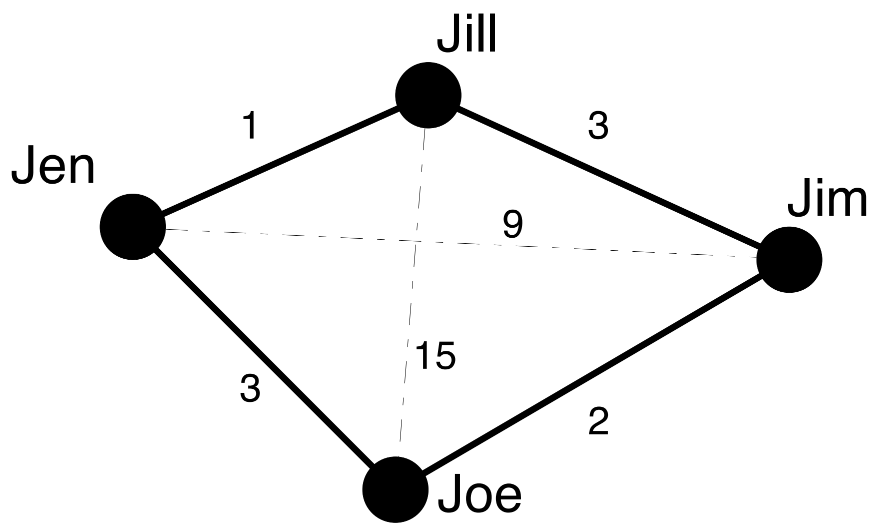
Dichotomized

	Jim	Jill	Jen	Joe
Jim	-	1	0	1
Jill	1	-	1	0
Jen	0	1	-	1
Joe	1	0	1	-

Distances btw offices

	Jim	Jill	Jen	Joe
Jim	-	3	9	2
Jill	3	-	1	15
Jen	9	1	-	3
Joe	2	15	3	-

- The diagram below uses solid lines to represent the adjacency matrix, while the numbers along the solid line (and dotted lines where necessary) represent the proximity matrix.
- In this particular case, one can derive the adjacency matrix by dichotomizing the proximity matrix on a condition of $p_{ij} <= 3$.

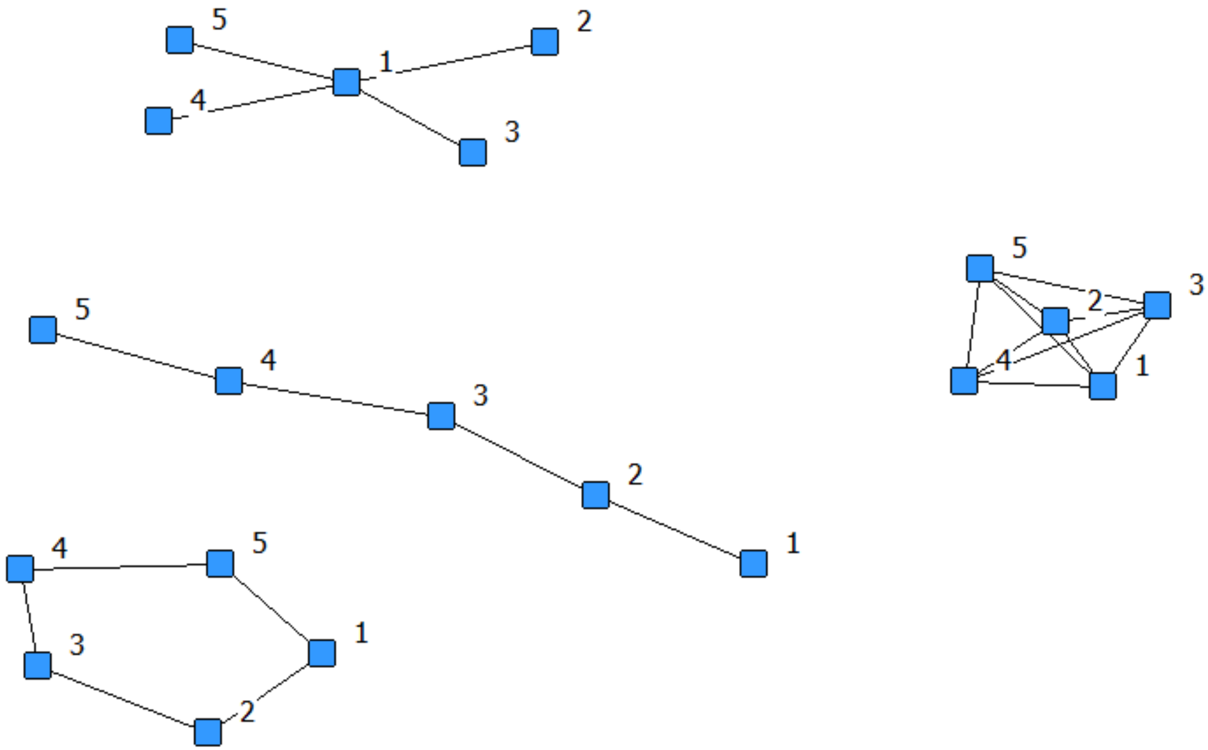


Reflexive graphs

- A reflexive tie is a tie from a node to itself
  - Self-loops
- Reflexive graphs are ones in which ties from a node to itself is allowed
- Normally only used when nodes represent collective agents such as cities
  - Number of phone calls between US cities

Some well-known graphs

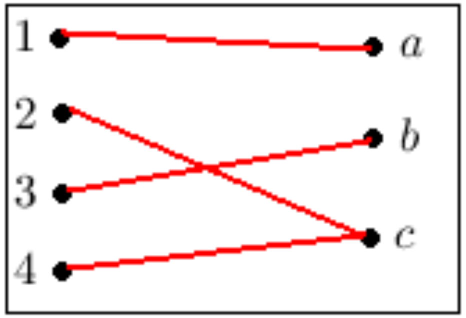
- Line/path
- Circle/cycle
- Clique
- Star



Expressing the presence of a tie

- Suppose you have an undirected graph G(V,E)
  - To express that u and v have a tie in this graph we can write (u,v) ϵ E or, if there multiple graphs under discussion, (u,v) ϵ E(G)
  - It is irrelevant whether we write (u,v) ϵ E or (v,u) ϵ E
- If G(V,E) is directed, then
  - (u,v) ϵ E means u has a tie to v.
  - If it also true that (v,u) ϵ E , we say the u--v tie is reciprocated

Mathematical relations



- A graph can also be viewed as a mathematical relation
- Wikipedia:
  - In mathematics, a **binary relation** over two sets *A* and *B* is a set of ordered pairs (*a*, *b*), consisting of elements *a* of *A* and elements *b* of *B*. That is, it is a subset of the Cartesian product *A* × *B*. It encodes the information of relation: an element *a* is related to an element *b*, if and only if the pair (*a*, *b*) belongs to the set. Binary relation is the most studied form of relations among all n-ary relations.
  - A graph is a special case where A and B are the same set
- Just a set of pairs of things. To say that u and v are tied by a given relation we can write, as before
  - (u,v) ϵ E(G)
  - But is also convenient to write uEv, which says u has the relation with v

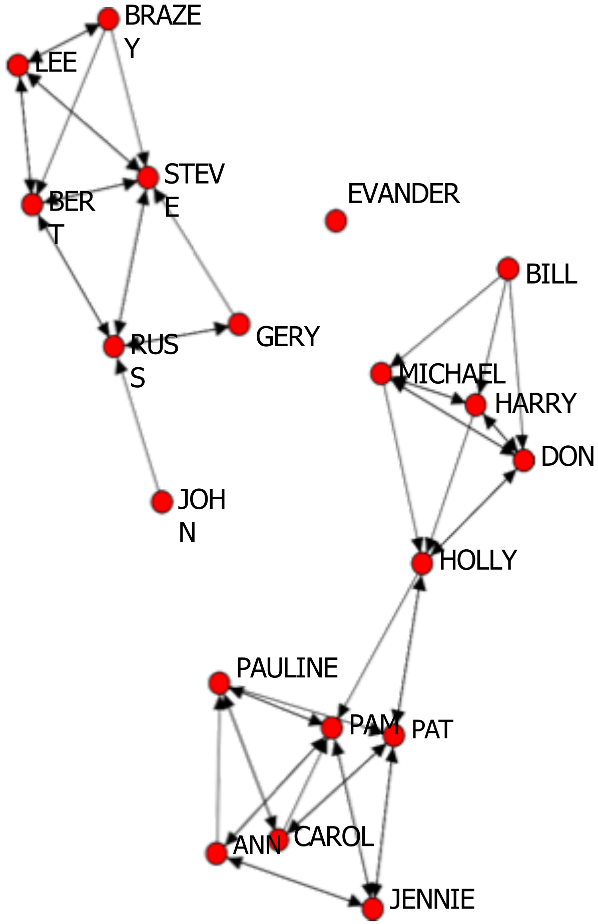
Relational terminology

- Suppose B is the relation "is the brother of" and F is the relation "is the father of"
  - uBv means u is the brother of v
  - yFx means y is the father of x
- We can define a compound relation BF as "is the brother of someone who is the father of"
  - uBFx means u is the brother of the father of x
- So BF is the uncle relation
  - U = BF
  - zUx means z is the uncle of x

Relational terminology – cont.

- The relation FF is the father of the father of
  - uFFv means that u is the grandfather of v
- We use F‘ to indicate the converse of a relation F
- If F means is the father of, then F’ means is the child of
  - uFv if and only if vF’U
- The compound relation F’F means ‘the child of the father of’
  - uF’Fv means that u is the child of someone who is the father of v.
  - Who are u and v to each other? They are siblings
- The relation FF’ is the father of the child of
  - uFF’v means that u is the father of someone who is the son of v
  - In other words u and v are co-parents to each other – they have the same children

Node-related concepts

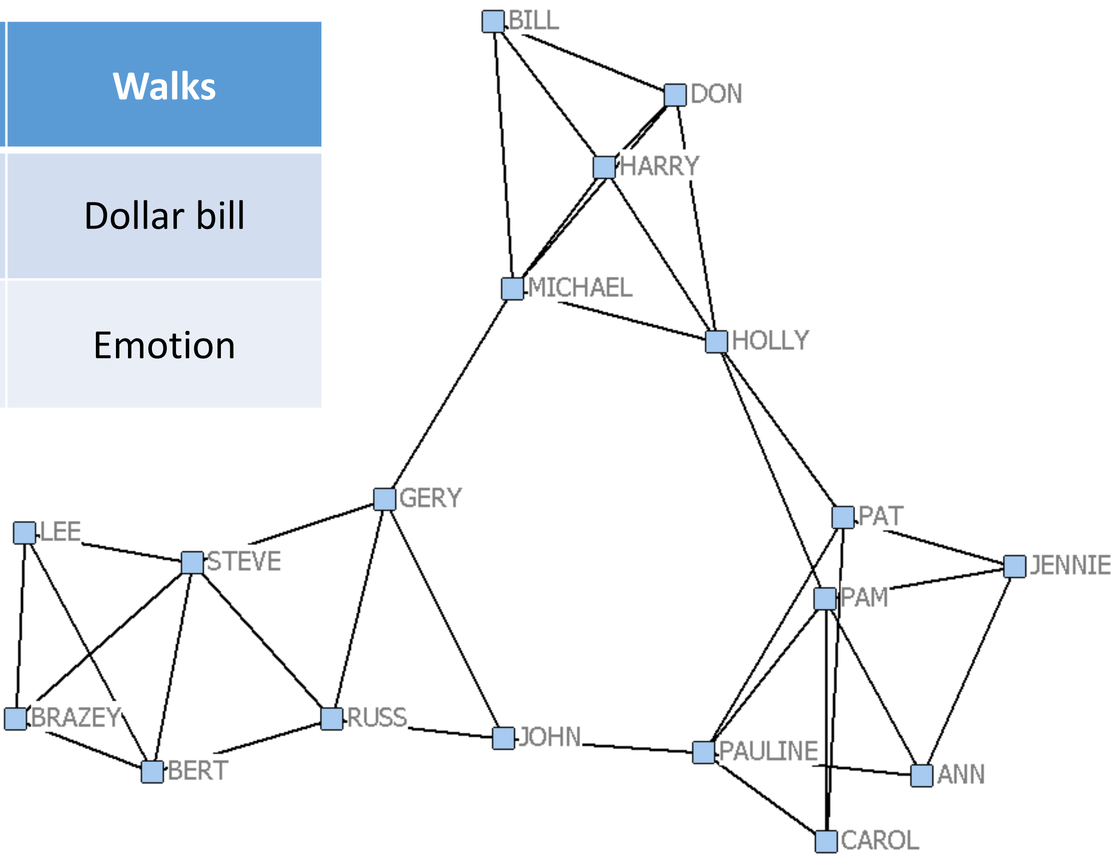


- **Degree**
  - The number of ties incident upon a node
  - In a digraph, we have indegree (number of arcs to a node) and outdegree (number of arcs from a node)
- **Pendant**
  - A node connected to a component through only one edge or arc
    - A node with degree 1
    - Example: John
- **Isolate**
  - A node which is a component on its own
    - E.g., Evander

How do things move?

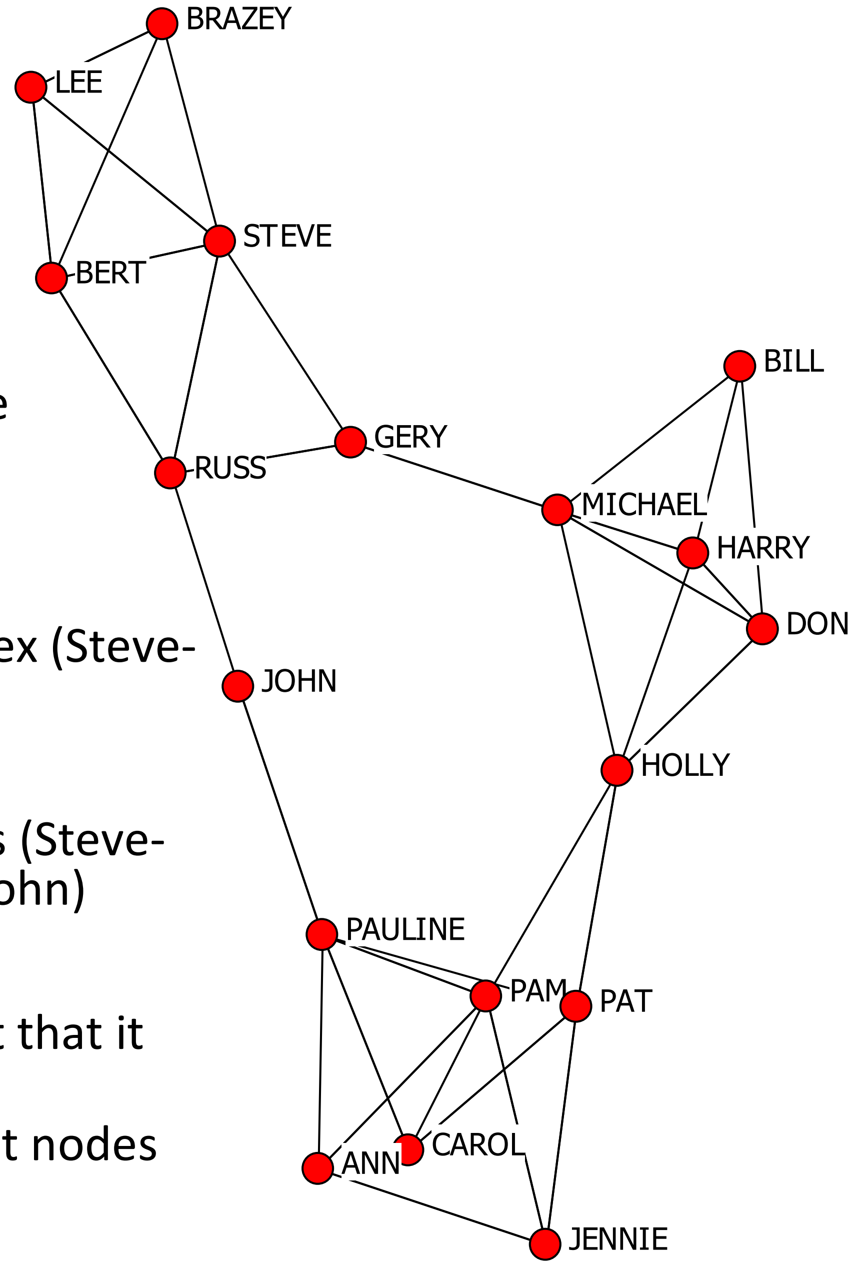
	Paths	Trails	Walks
Move	Snail mail	Used paperback	Dollar bill
Copy	Virus	Gossip	Emotion

- Path – can't revisit a node or a line
- Trail – can't revisit a line
- Walk – unrestricted
- Every path is a trail, every trail is a walk



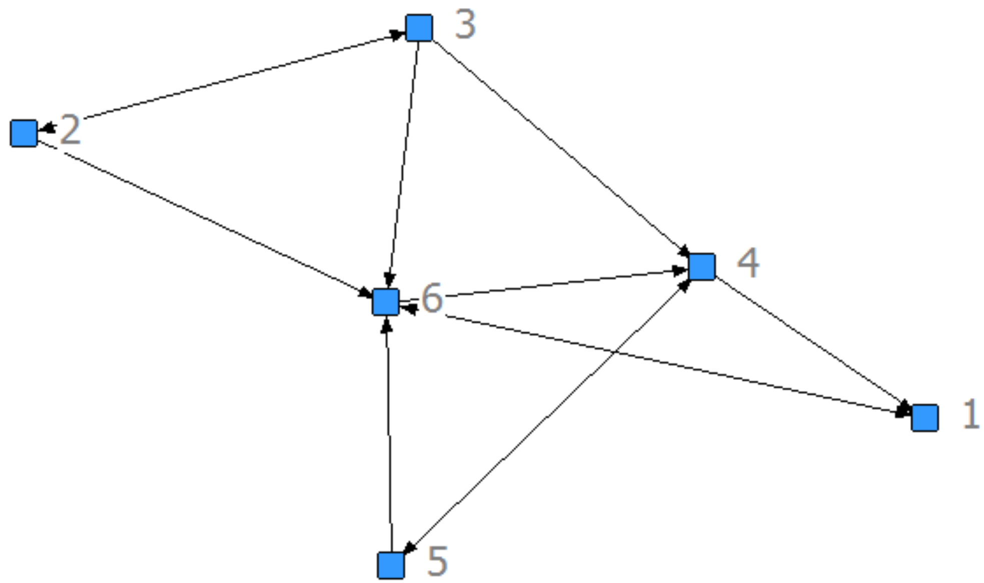
Graph traversals

- Walk
  - Any unrestricted traversing of vertices across edges (Russ-Steve-Bert-Lee-Steve)
- Trail
  - A walk restricted by not repeating an edge or arc, although vertices can be revisited (Steve-Bert-Lee-Steve-Russ)
- Path
  - A trail restricted by not revisiting any vertex (Steve-Lee-Bert-Russ)
- Geodesic Path
  - The shortest path(s) between two vertices (Steve-Russ-John is shortest path from Steve to John)
- Cycle
  - A cycle is in all ways just like a path except that it ends where it begins
  - Aside from endpoints, cycles do not repeat nodes
  - E.g. Brazey-Lee-Bert-Steve-Brazey



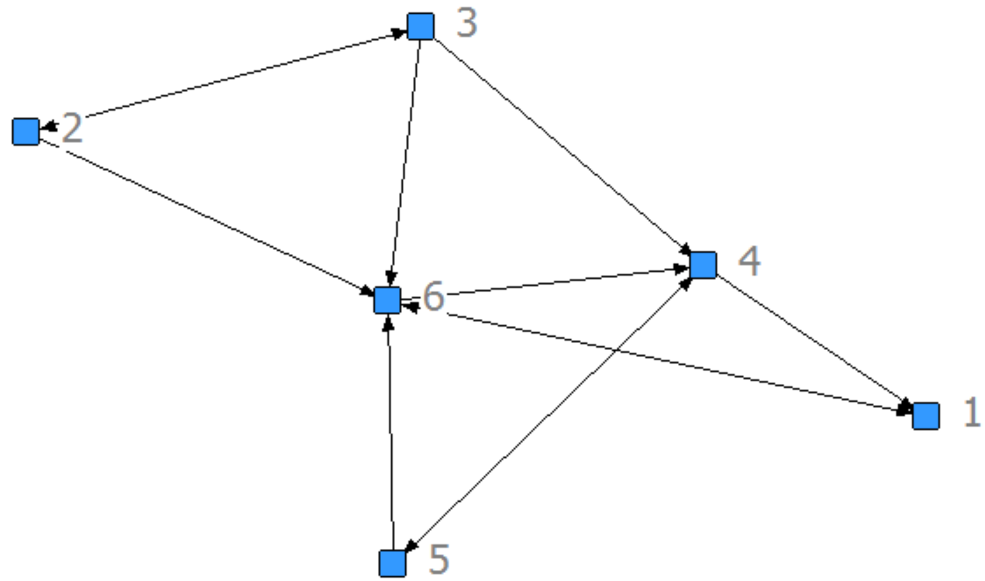
Path

- A path is of sequence of incident lines (together with the nodes they connect) in which no node occurs more than once
  - Can’t revisit a node
- 3-4-1-6 is a path
- 3-4-1-6-4 is not
- Length of a path is defined as the number of lines in it
  - Path 2-3-4-1 is length 3
- The shortest path from u to v is called a *geodesic*



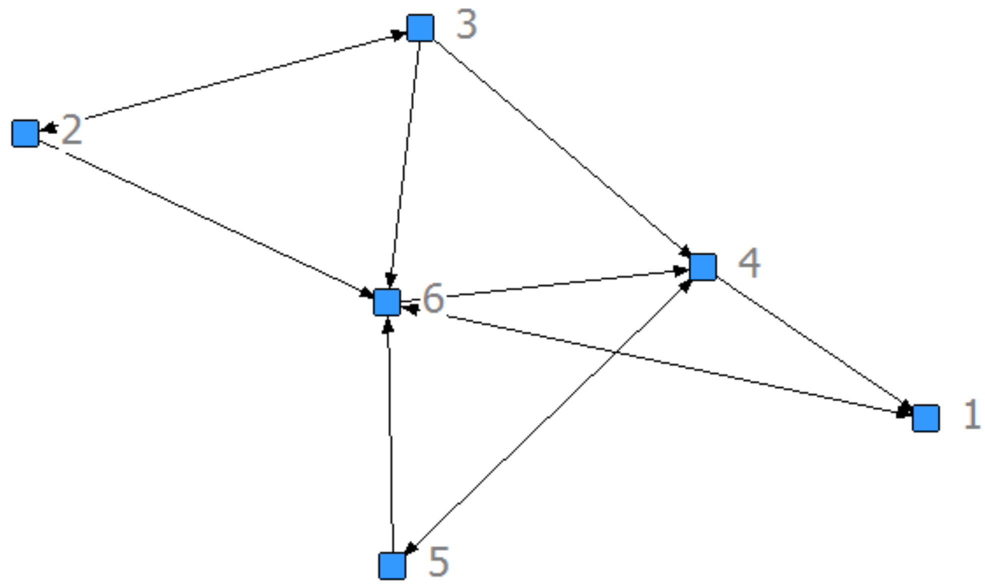
Trail

- Trail is a sequence of incident lines such that no line occurs more than once
  - Nodes can be revisited, but lines can't
- 3-4-1-6-4-5-6 is a trail
- 3-4-1-6-4-5-6-4 is not



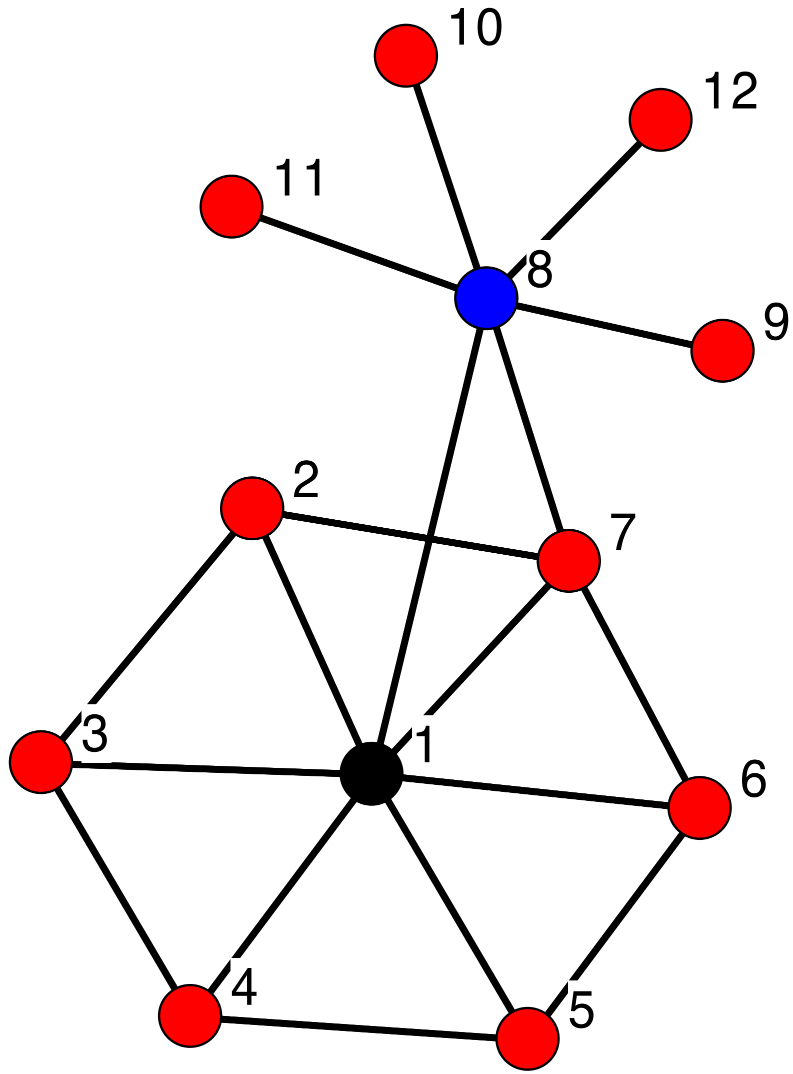
Walks

- Walks are unrestricted sequences of incident edges.
  - Can revisit any node or line
  - 2-3-2-6 is a walk
  - 2-3-4-6 is not (must obey direction)
- Every path is a trail, every trail is a walk, every path is a walk
- All of these are ways that things can traverse a graph, can flow through the graph



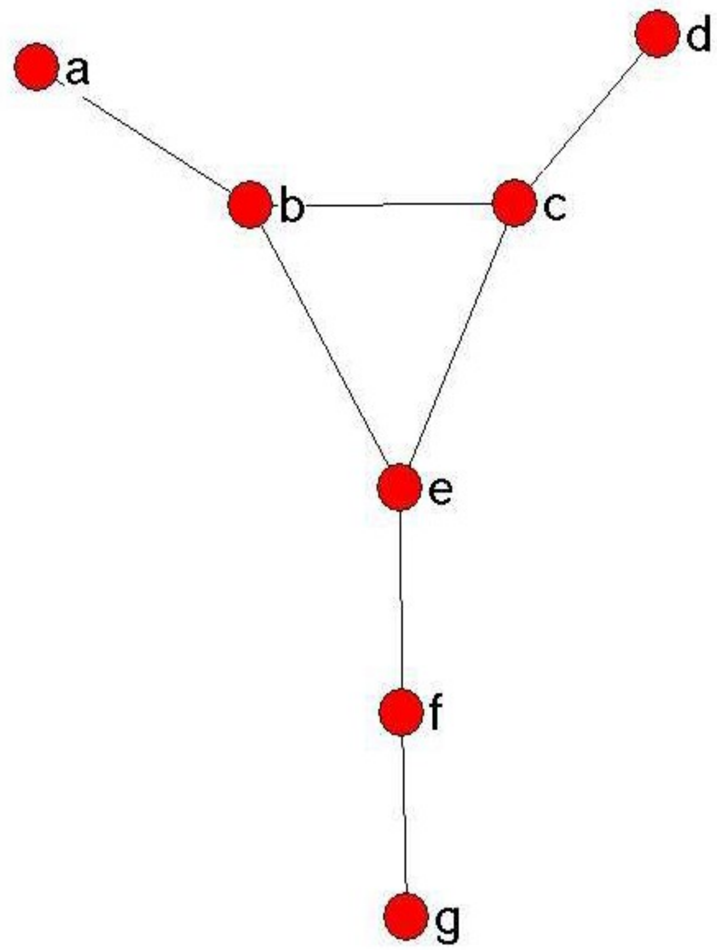
Length & Distance

- Length of a path (or any walk) is the number of links it has
- The **Geodesic Distance** (aka graph-theoretic distance) between two nodes is the length of the shortest path
  - Distance from 5 to 8 is 2, because the shortest path (5-1-8) has two links



Geodesic Distance Matrix

	a	b	c	d	e	f	g
a	0	1	2	3	2	3	4
b	1	0	1	2	1	2	3
c	2	1	0	1	1	2	3
d	3	2	1	0	2	3	4
e	2	1	1	2	0	1	2
f	3	2	2	3	1	0	1
g	4	3	3	4	2	1	0



Path lengths

- We can think of LL as L^2. If $L^2(a,c) > 0$, it means there exists a path (technically, a walk) from a to c that is exactly 2 links long
  - If we compute LLL or L^3, then $L^3(a,c) > 0$ means there exists at least one walk from a to c that is exactly 3 links long
- More generally if $L^k(i,j) > 0$, it means there is at least one walk from i to j that is exactly k links long
  - $L^k(i,j) = 7$ means there are 7 different walks from i to j that are of length k

L

	a	b	c	d
a	0	1	0	1
b	0	0	1	0
c	0	1	0	0
d	1	0	1	0

x

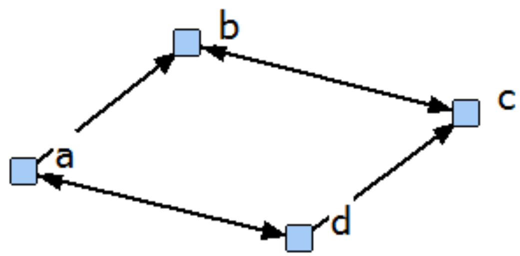
L

	a	b	c	d
a	0	1	0	1
b	0	0	1	0
c	0	1	0	0
d	1	0	1	0

=

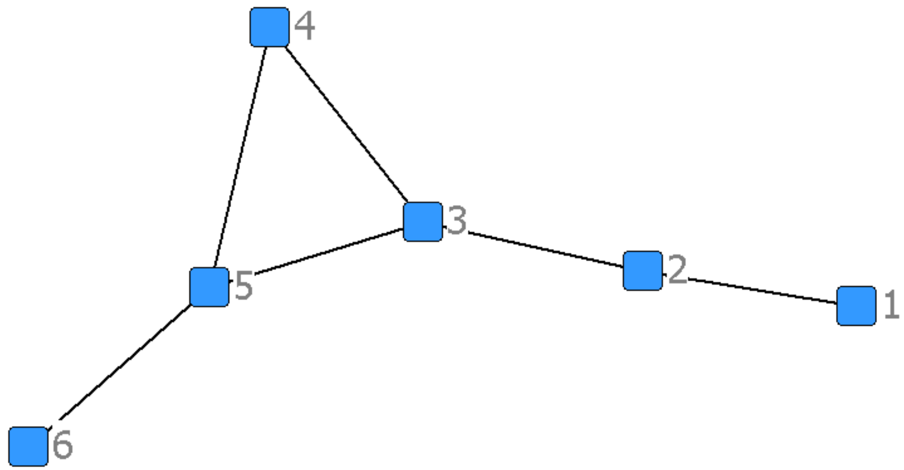
LL

	a	b	c	d
a	1	0	2	0
b	0	1	0	0
c	0	0	1	0
d	0	2	0	1



Matrix powers example

Note that shortest path from 1 to 5 is three links, so $x_{1,5} = 0$ until we get to X^3



	1	2	3	4	5	6
1	0	1	0	0	0	0
2	1	0	1	0	0	0
3	0	1	0	1	1	0
4	0	0	1	0	1	0
5	0	0	1	1	0	1
6	0	0	0	0	1	0

X

	1	2	3	4	5	6
1	1	0	1	0	0	0
2	0	2	0	1	1	0
3	1	0	3	1	1	1
4	0	1	1	2	1	1
5	0	1	1	1	3	0
6	0	0	1	1	0	1

X^2

	1	2	3	4	5	6
1	0	2	0	1	1	0
2	2	0	4	1	1	1
3	0	4	2	4	5	1
4	1	1	4	2	4	1
5	1	1	5	4	2	3
6	0	1	1	1	3	0

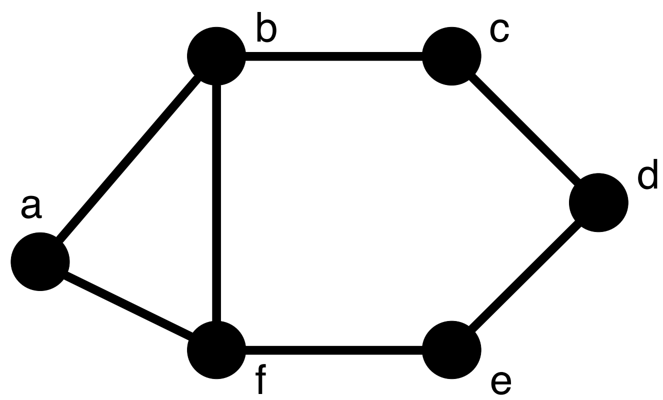
X^3

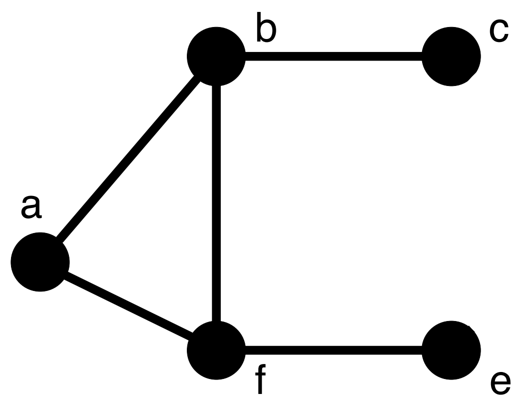
	1	2	3	4	5	6
1	2	0	4	1	1	1
2	0	6	2	5	6	1
3	4	2	13	7	7	5
4	1	5	7	8	7	4
5	1	6	7	7	12	2
6	1	1	5	4	2	3

X^4

Subgraphs

- Set of nodes
  - Is just a set of nodes
- A subgraph
  - Is set of nodes together with ties among them
- An induced subgraph
  - Subgraph defined by a set of nodes
  - Like pulling the nodes and ties out of the original graph

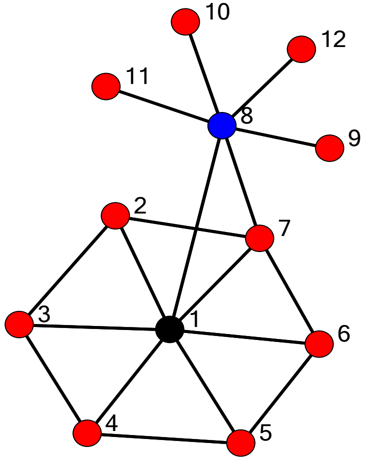




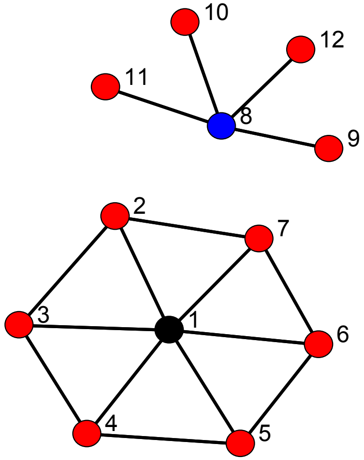
Subgraph induced by considering the set {a,b,c,f,e}

Connected vs disconnected graphs

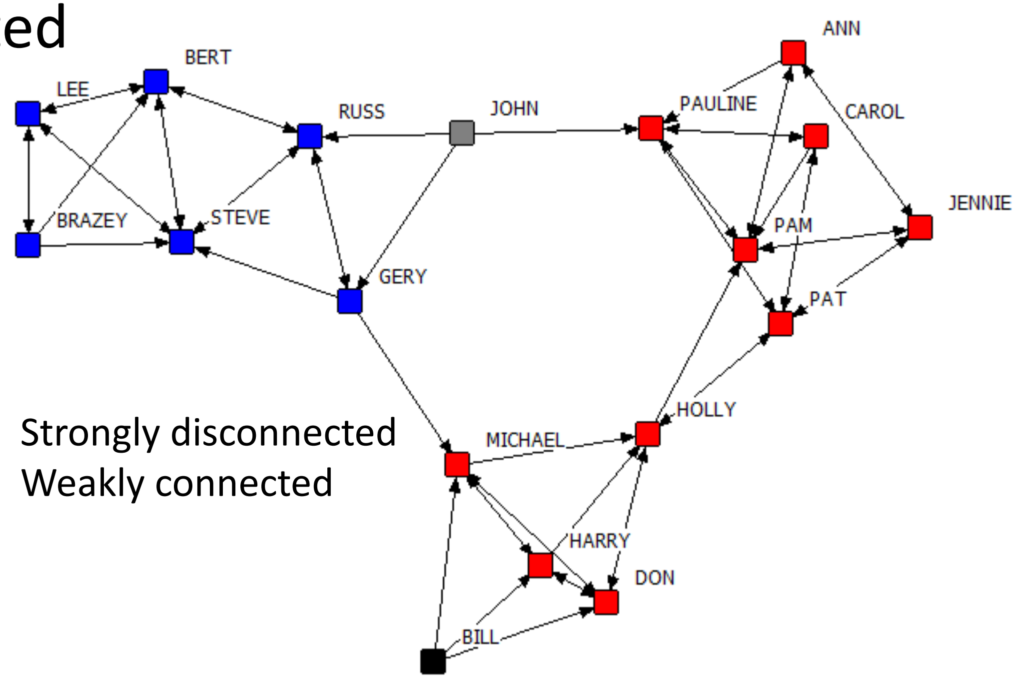
- A graph is connected if you can reach any node from any other – i.e., there exists a path from one to the other
- Directed graphs are often disconnected



Connected



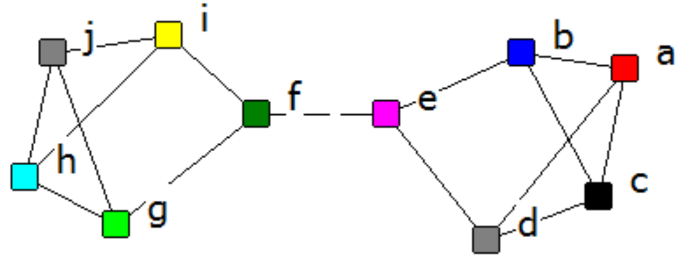
Disconnected



Strongly disconnected
Weakly connected

Component

- Maximal sets of nodes in which every node can reach every other by some path (no matter how long)
  - Coherent fragments of a graph
- A graph with a single component is called a connected graph
- Weak vs strong components
  - A weak component is where we ignore the direction of the arcs



Removing F-E tie would create a network with 2 components

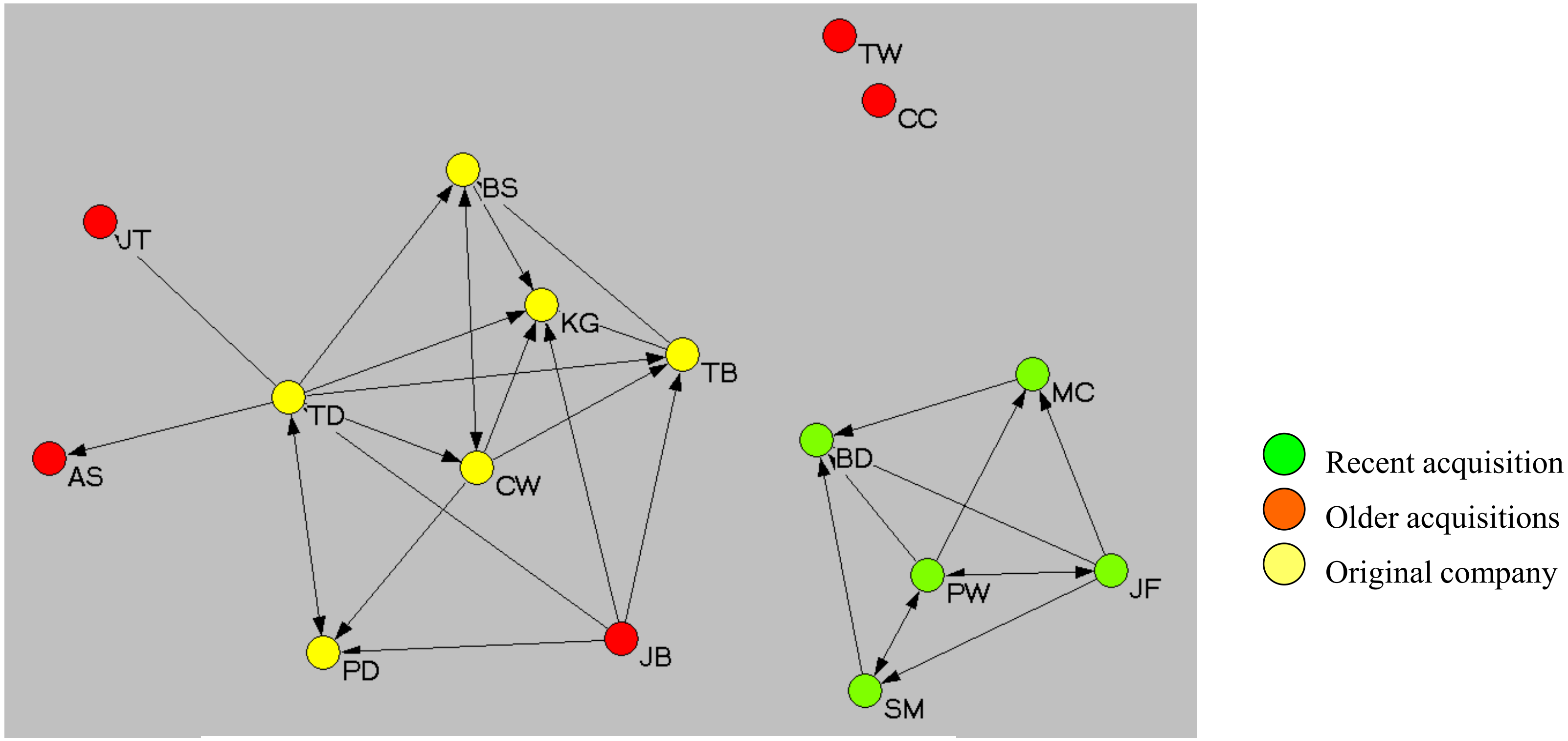
It is relations (types of tie) that define different networks, not components. A network that has two components remains one (disconnected) network.

Components in Directed Graphs

- Strong component
  - There is a directed path from each member of the component to every other
- Weak component
  - There is an undirected path (a weak path) from every member of the component to every other
  - Is like ignoring the direction of ties – driving the wrong way if you have to

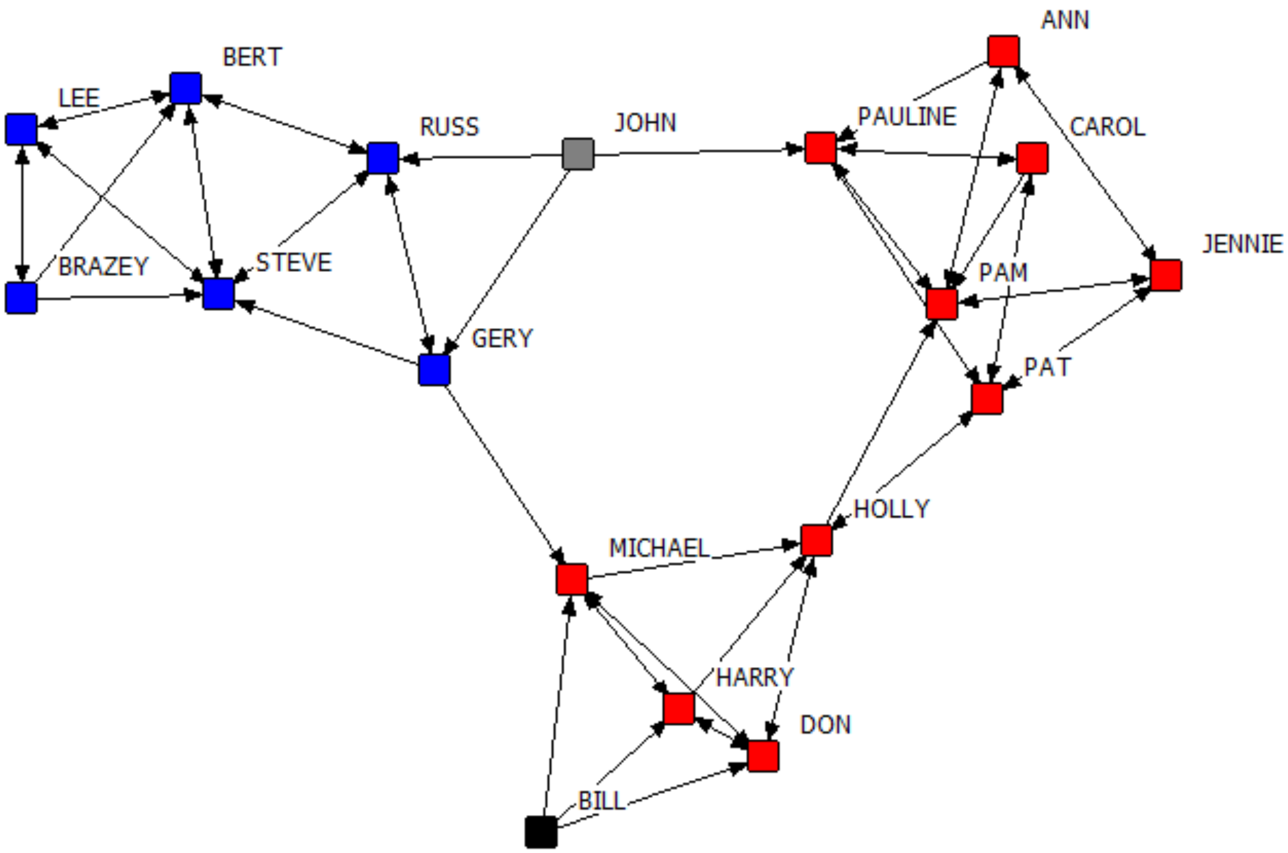
A network with 4 weak components

Who you go to so that you can say 'I ran it by _____, and she says ...'



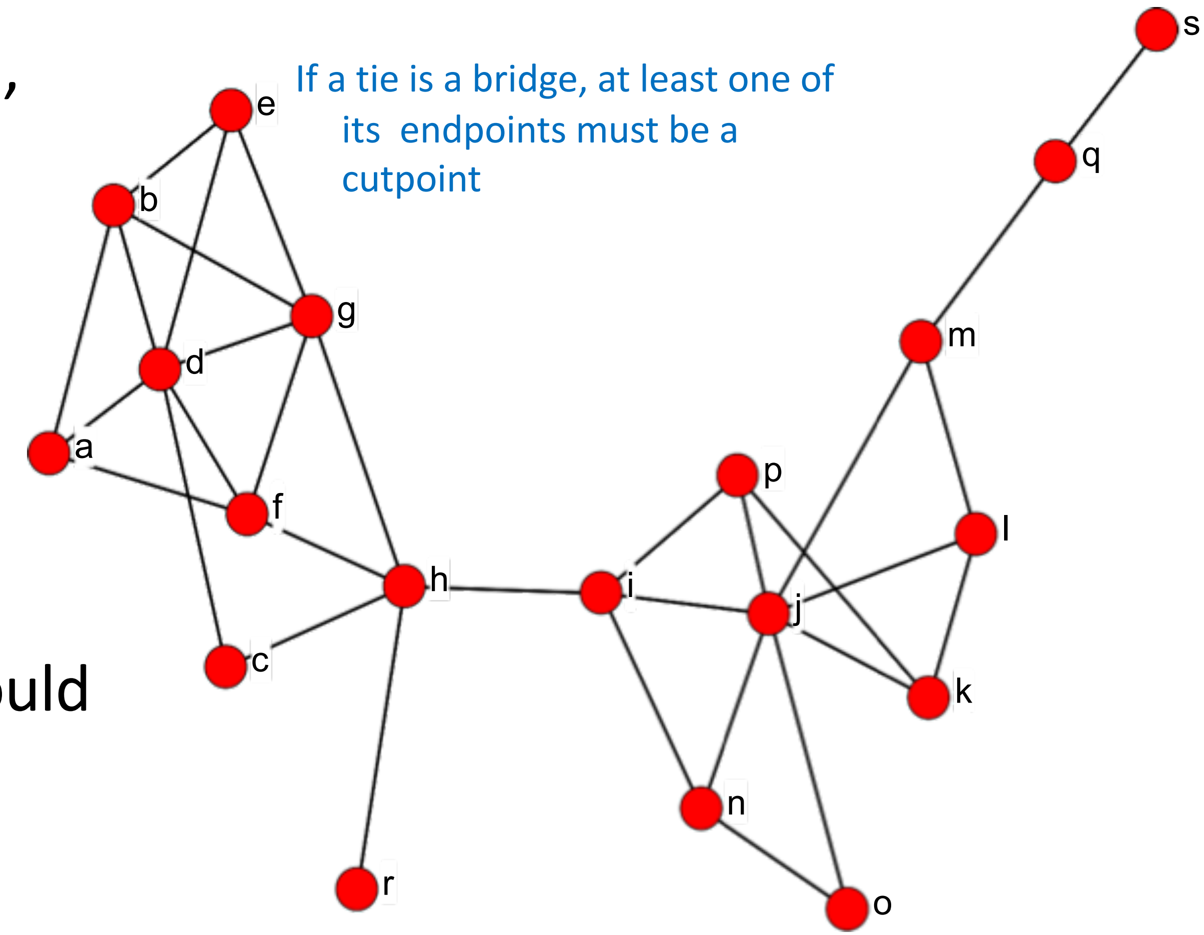
Data drawn from Cross, Borgatti & Parker 2001.

1 weak component, 4 strong components



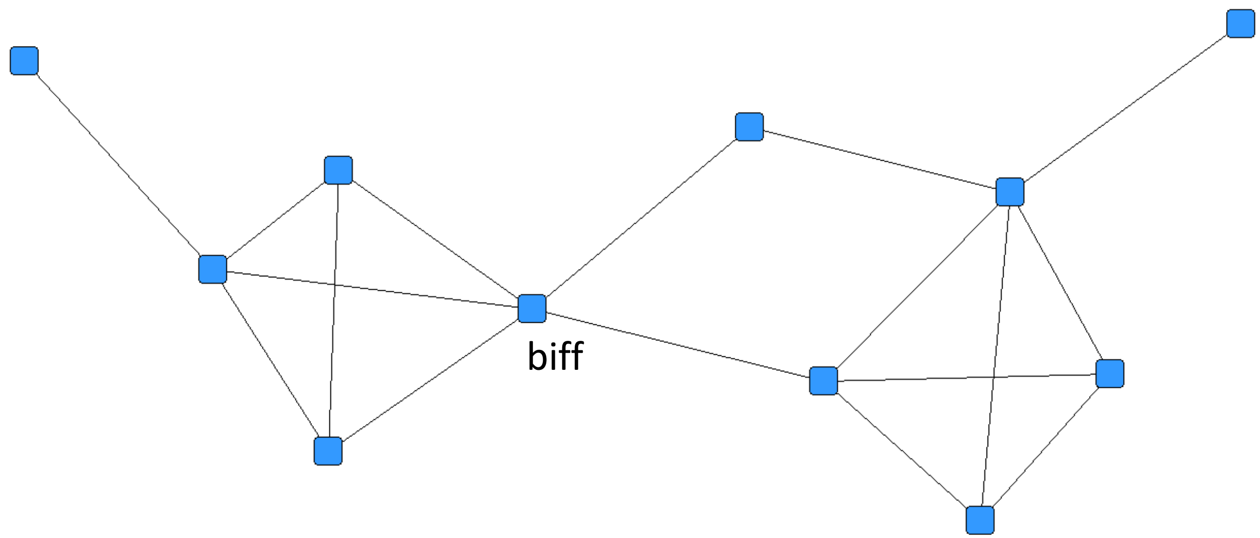
Cutpoints and Bridges

- Cutpoint
  - A node which, if deleted, would increase the number of components
- Bridge
  - A tie that, if removed, would increase the number of components



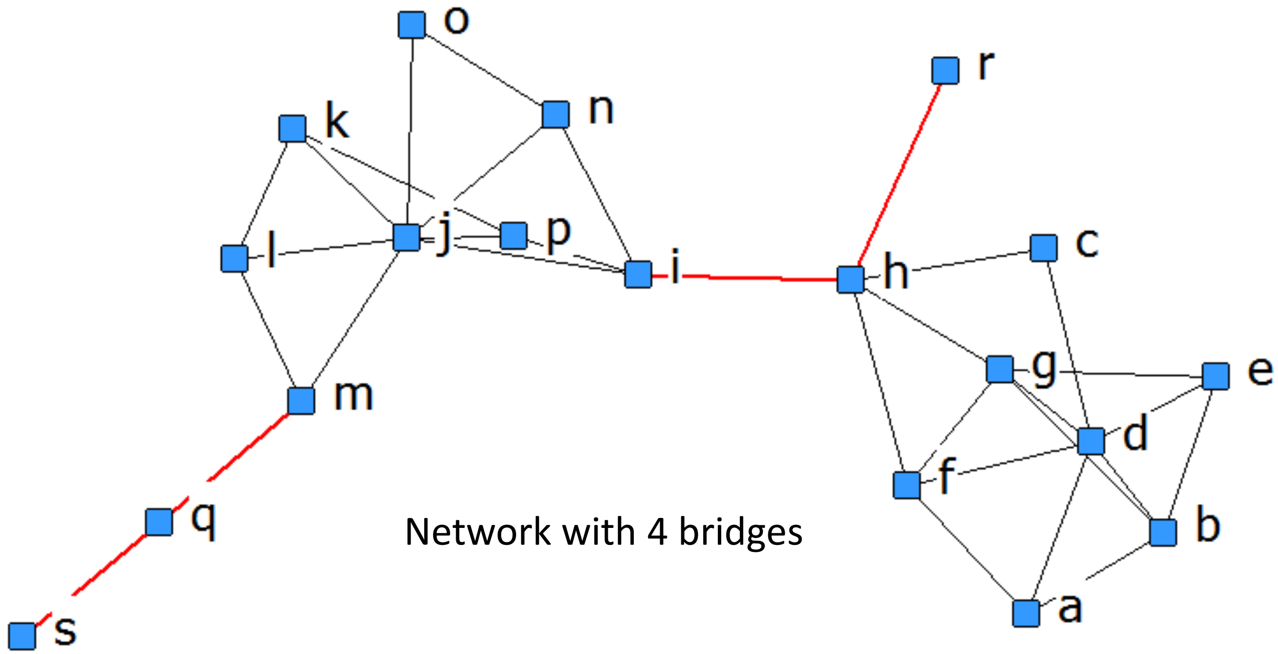
Cutpoints

- Nodes which, if deleted, would increase the number of components in the network
  - Removing Biff would disconnect the network (create 2 components)



Bridge

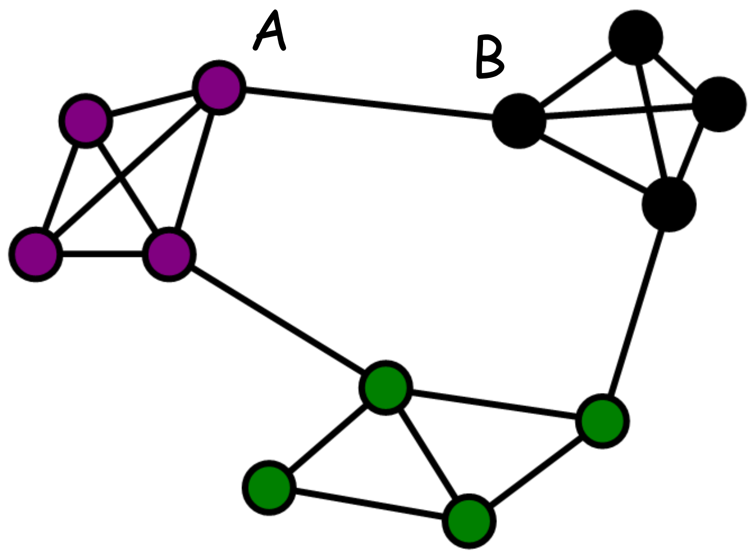
- An edge which, if removed, would increase the number of components in the network



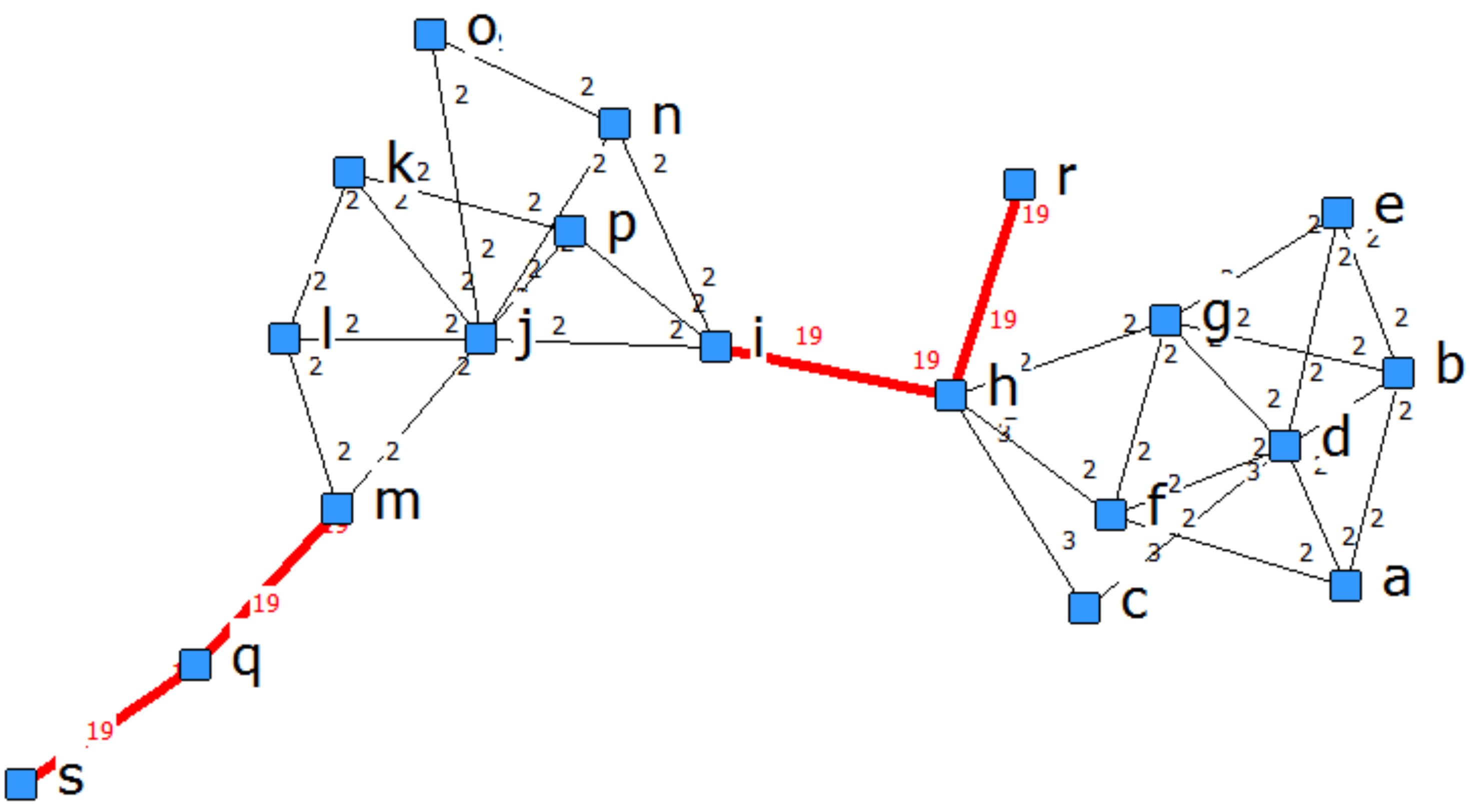
Network with 4 bridges

Local Bridge of Degree K

- An edge that connects nodes that would otherwise be a minimum of *k* steps apart
  - The A—B tie is local bridge of degree 5
- Loss of relationship between A and B would effectively, though not actually, disconnect A from B



Local bridges of degree k



Getting the Data in UCINET

- Four options:
  - DL Files
    - Text files of various formats that can be created easily by geeks and nerds
  - Excel Files/Grid format
    - UCINET has a spreadsheet tool that easily interacts with Excel or can allow manual entry if network is not too large
  - VNA Files
    - Text files that allow for a single-file that contains both dyadic and nodal attribute data
  - Import Text Via Spreadsheet tool
    - A new tool in UCINET that lets you do DL file formats in a spreadsheet tool

DL Files

- These are the most versatile
- There are multiple formats:
  - Full Matrix
  - Nodelist
  - Edgelist
- Each has its advantages

DL Data Formats

Dl n=5 Format = edgelist Labels embedded Data: billy john 6 john billy 1 john jill 2 jill mary mary billy 5 mary jill mary jill Best for data coming from a relational databases or if you have valued data. Values are added if repeated and default to 1	Dl n=5 Format = nodelist Labels embedded Data: billy john john billy jill jill mary mary billy jill This method is best for BINARY data **NOTE: This is a dichotomized version of the others**	Dl n=5 Format = edgelist2 Labels embedded Data: billy Essex 4 john Cambridge 2 jill Oxford 3 mary Leeds 6 This is the same as the edgelist format, except the nominating node (the first column) is of a different MODE than the nominated node (the second column). **There is also nodelist2**

VNA Files

- These **CAN** combine in one file both:
  - Nodal (attribute) data **and**
    - e.g., Age, gender, Education Level
  - Network/Relational/Dyadic data
    - E.g., Communicates with, Trusts
- Can have textual data
  - NetDraw will preserve the labels
  - UCINET will transform them to numbers

Sample VNA File

*Node data
"ID", "Gender", "Role"
"HOLLY" “FEMALE" “STUDENT"
“STEVE" “MALE" “TEACHER"
"CAROL" “FEMALE" “STUDENT"
…
*Tie data
FROM TO "campnet"
"HOLLY" "PAM" 1
"HOLLY" "PAT" 1
"BRAZEY" "STEVE" 1
"BRAZEY" "BERT" 1
"CAROL" "PAM" 1
"PAM" "ANN" 1
"PAT" "HOLLY" 1

Excel/Data Grid

- Excel is the "Universal Translator"
- UCINET has a Data Grid tool that
  - Looks like excel
  - Reads excel files
  - Works really well with Excel Cut&Paste
    - As long as you click in the right place for pasting your data

Some tricks

- If the network is small (not too many people)
  - I use excel
  - Create a comma-separated full-matrix-style file and cut and paste into the data grid
  - Manually create attribute file in UCINET (#s only)
- If the network is larger
  - I create an edgelist DL file for the network only
  - And a VNA file just with node data (attributes)
  - Then I:
    - Import the DL file into UCINET (creating ##h & ##d files)
    - Open the vna file as an attribute file
    - If I want to do attribute-based analyses in UCINET, I export the Attributes as a UCINET dataset (will translate text to numbers automatically for me- but I can't control them)

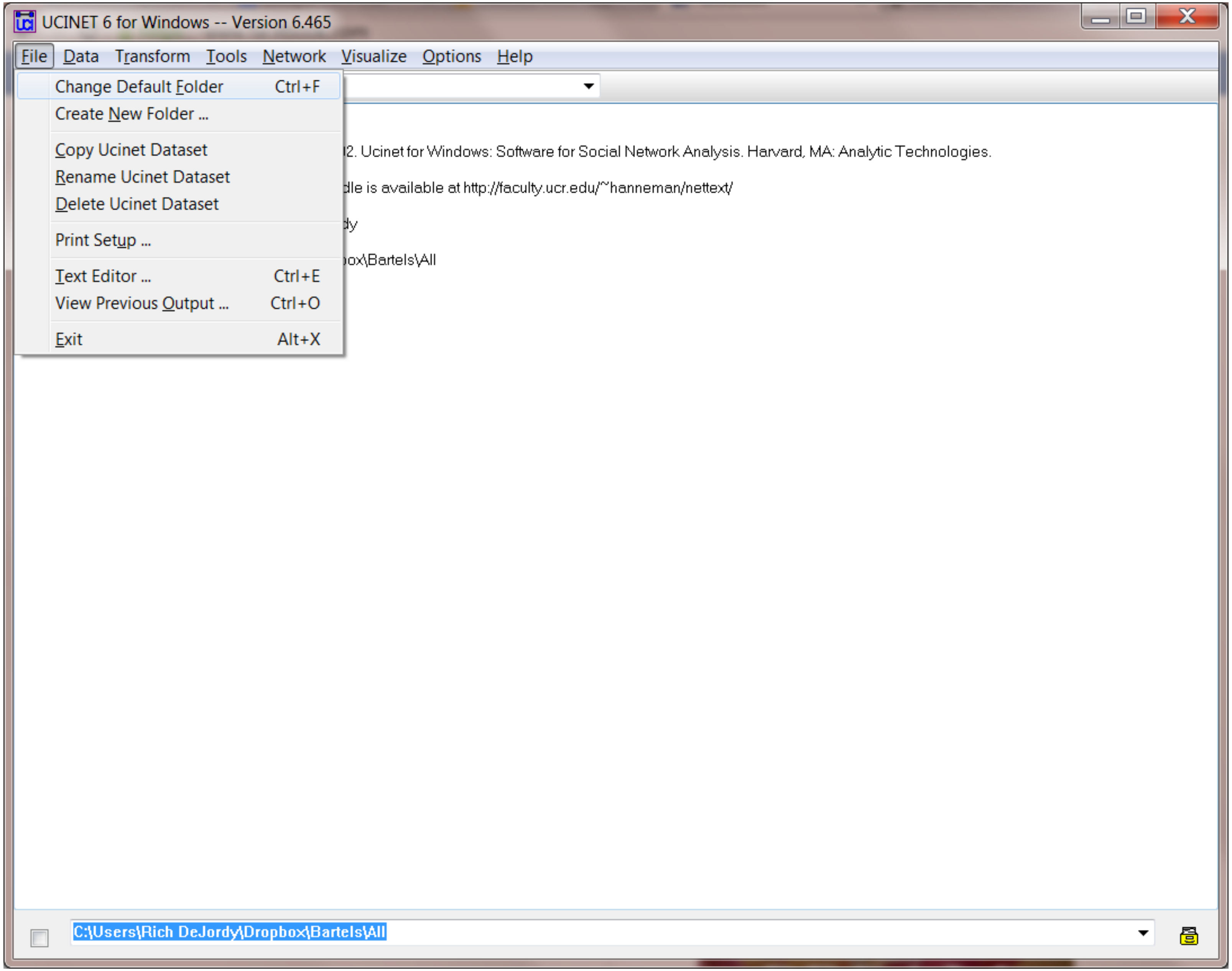
Where to find the importing

- In UCINET
  - Data | import | DL
  - Data | Import | VNA
  - Data | Spreadsheets | Matrix (Ctrl-S)
  - Data | Import via Spreasheet | DL
- In NetDraw
  - File | Open | Ucinet DL Text file
  - File | Open | VNA text file
  - NetDraw can work with the text files (no UCINET dataset). UCINET does not.

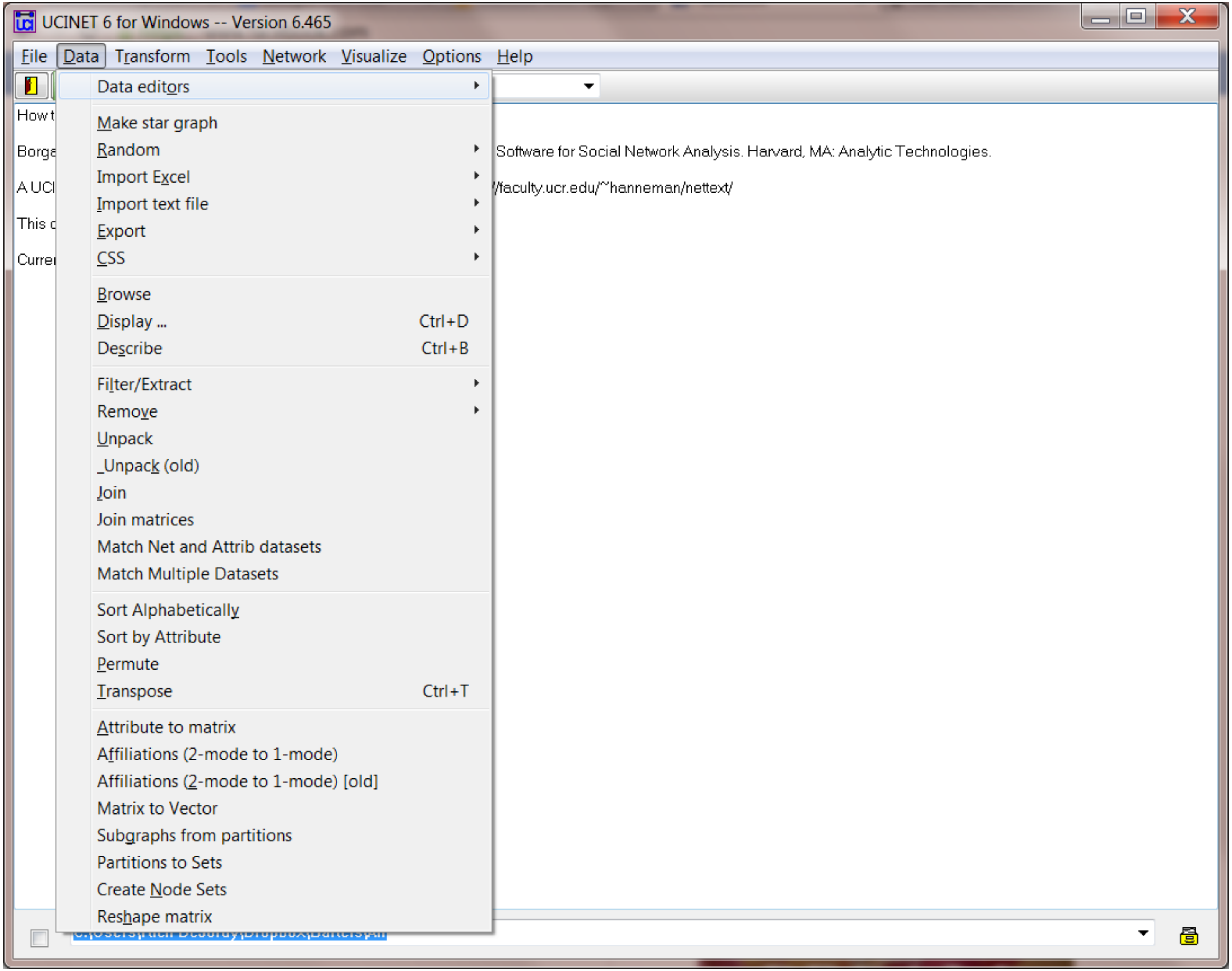
If you forget the format

- Just Export one of the Sample files
  - For DL files
    - From UCINET go to
      Data | Export | DL
  - For VNA files
    - From NetDraw, load the data and go to
      File | Save Data as | VNA | Complete

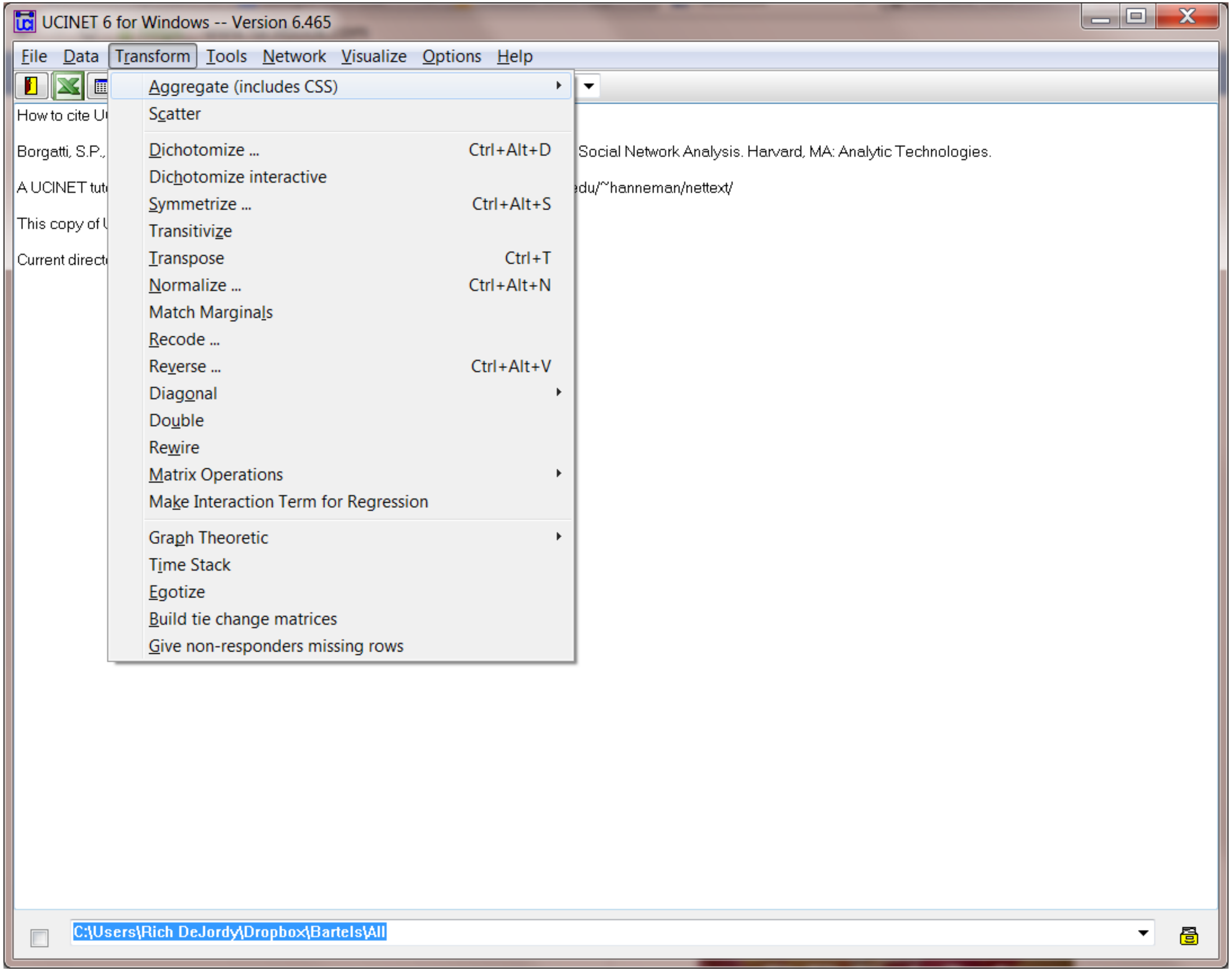
UCINET File Menu



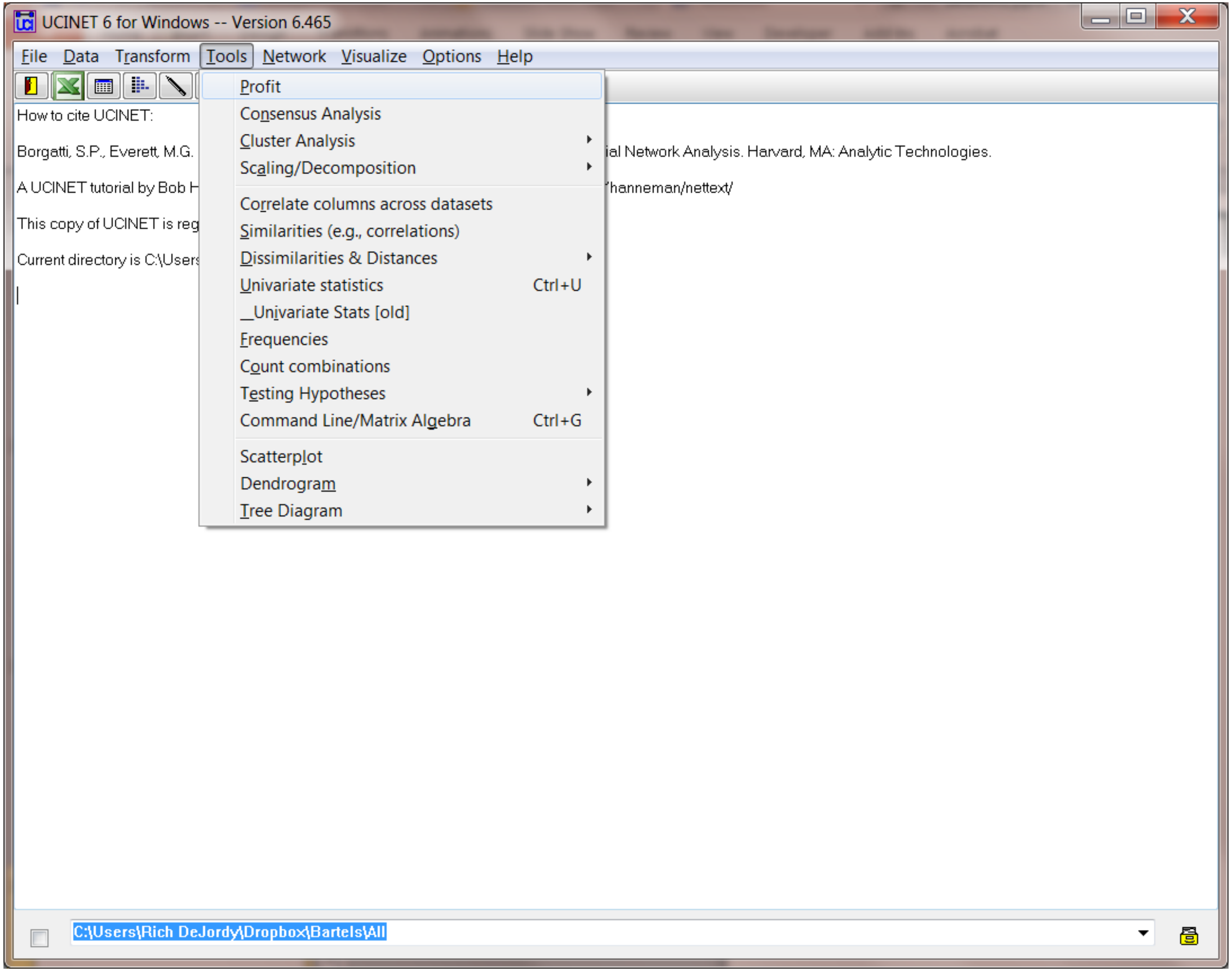
UCINET Data Menu



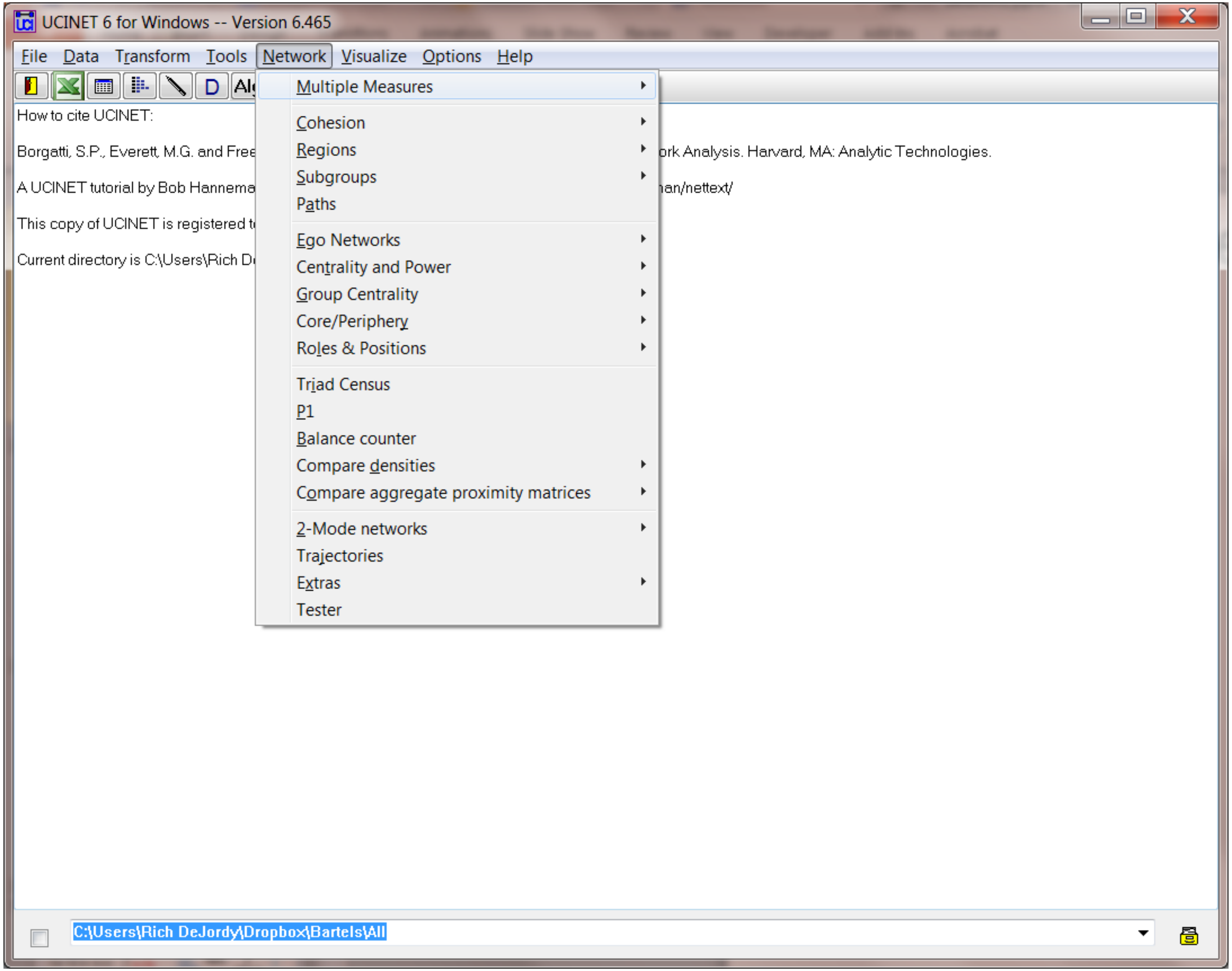
UCINET Transform Menu



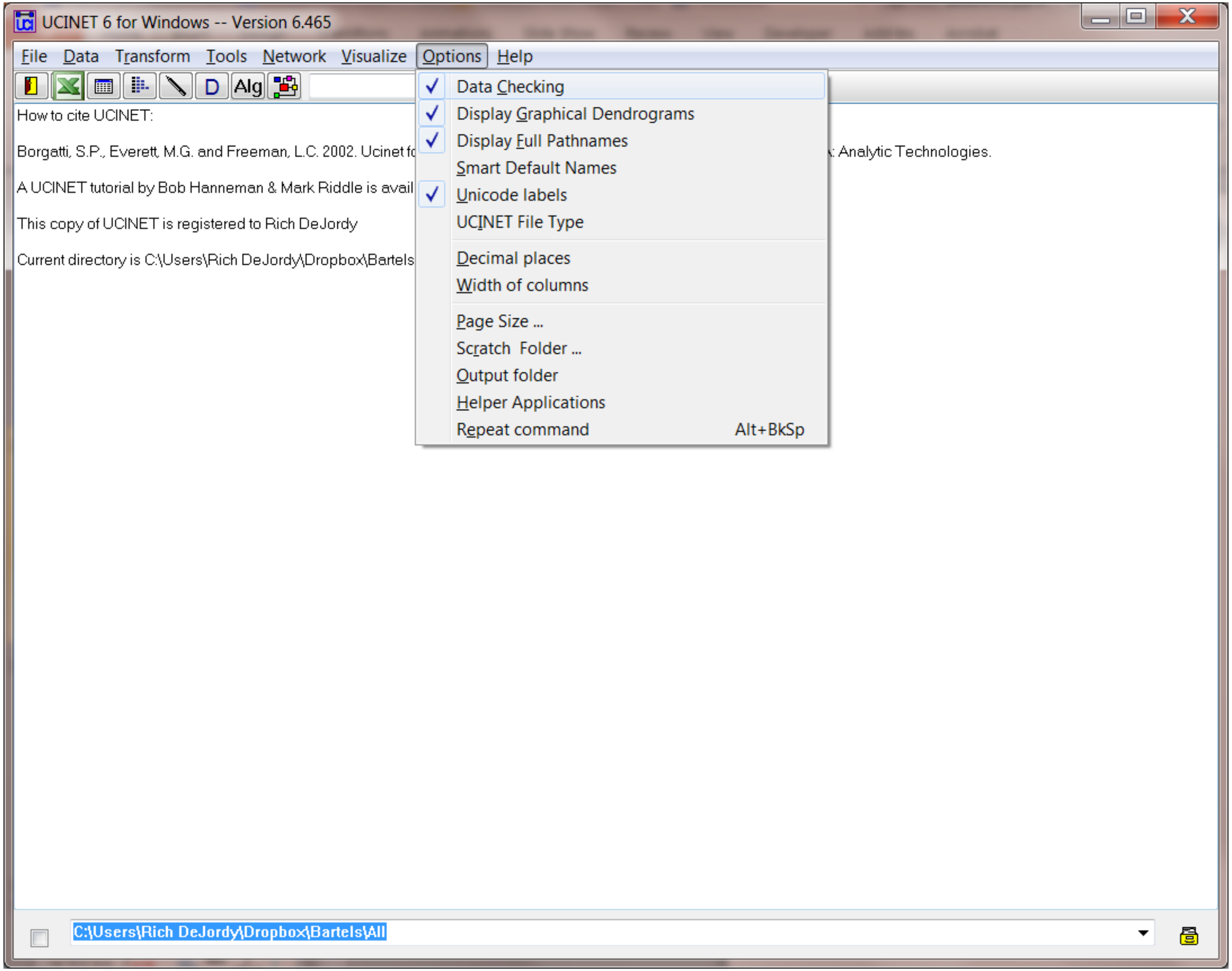
UCINET Tools Menu



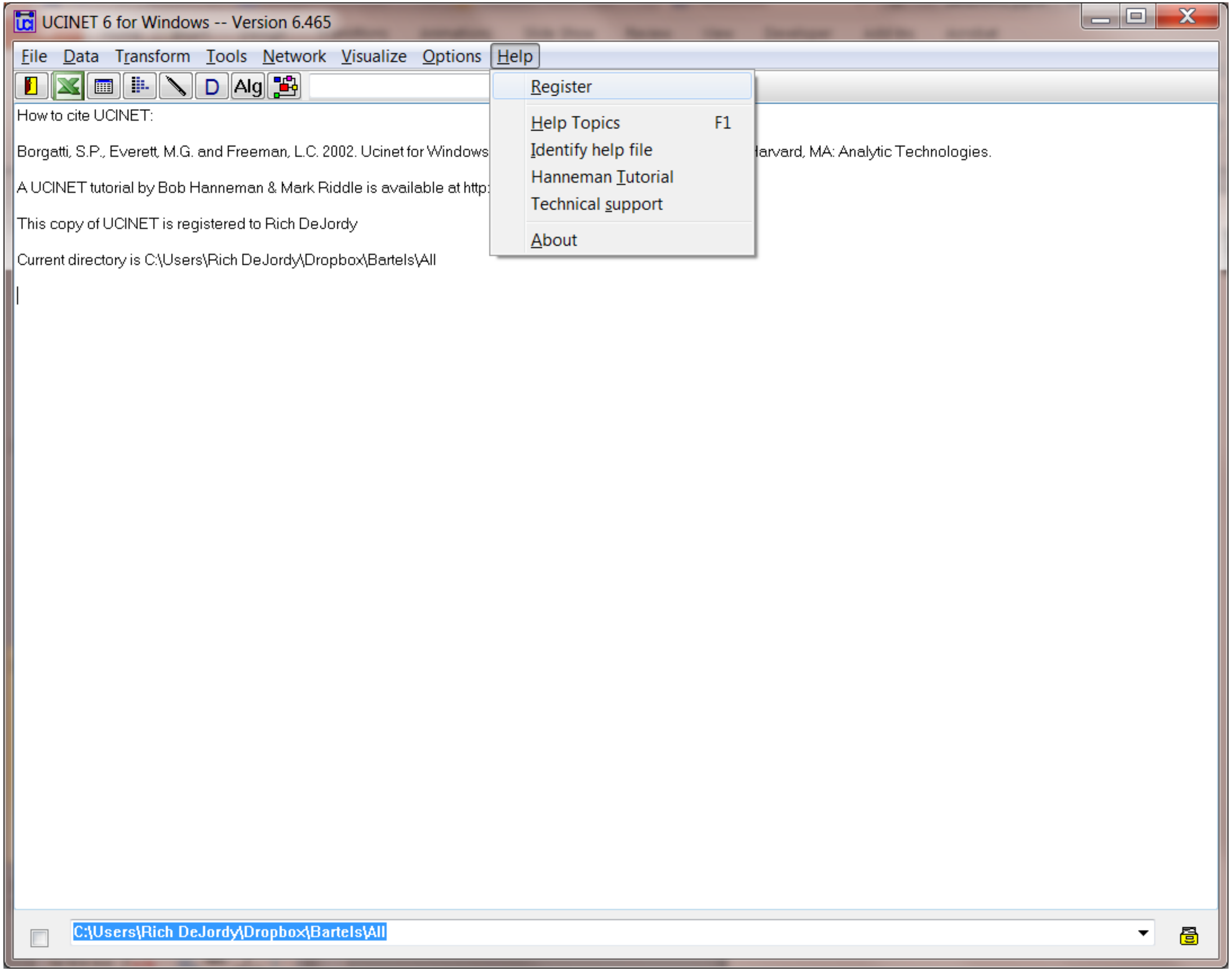
UCINET Network Menu



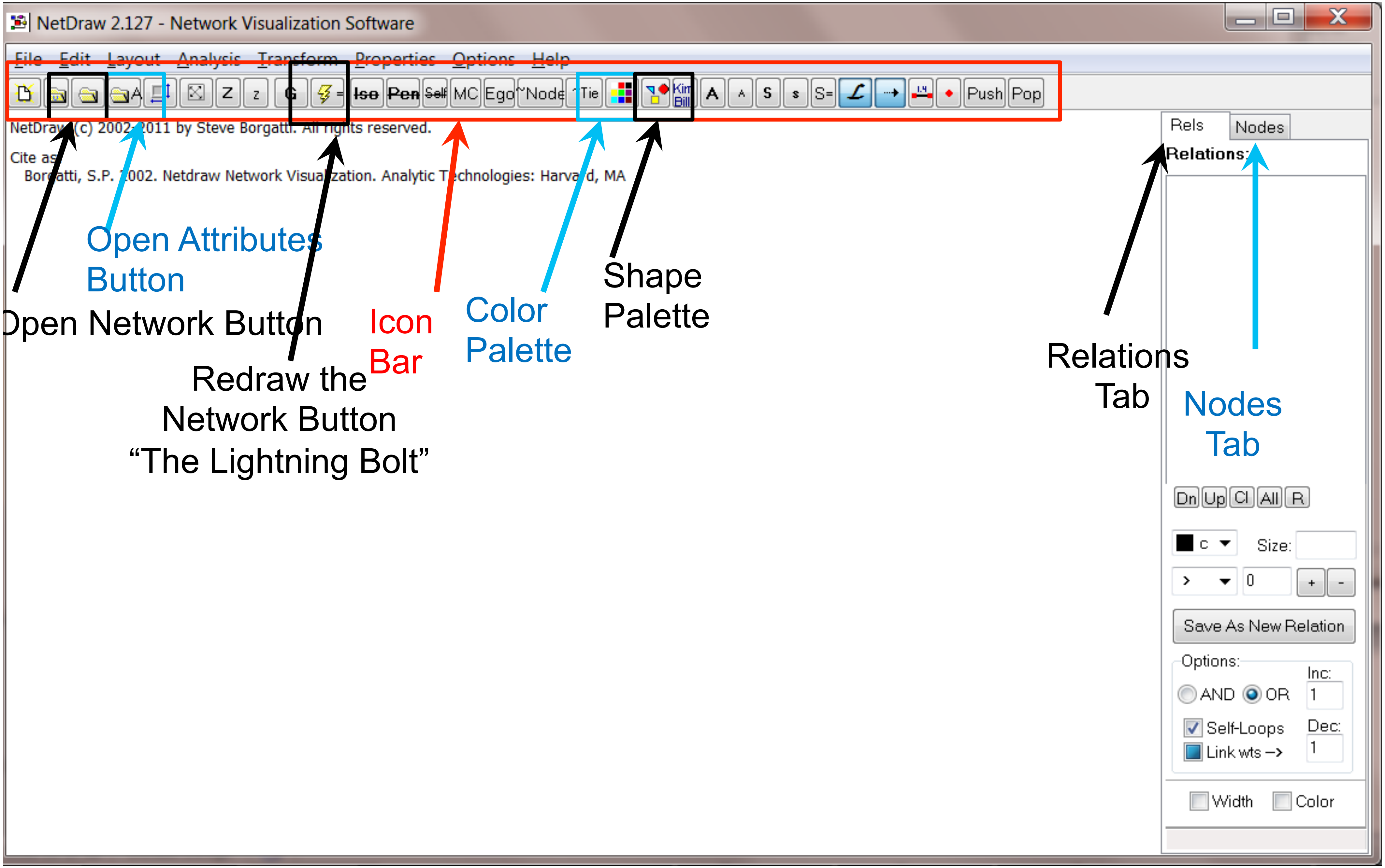
UCINET Options Menu



UCINET Help Menu



NetDraw


NetDraw 2.127 - Network Visualization Software
File Edit Layout Analysis Transform Properties Options Help
NetDraw (c) 2002-2011 by Steve Borgatti. All rights reserved.
Cite as:
Borgatti, S.P. 2002. Netdraw Network Visualization. Analytic Technologies: Harvard, MA
Open Attributes Button
Open Network Button
Redraw the Network Button "The Lightning Bolt"
Icon Bar
Color Palette
Shape Palette
Relations Tab
Nodes Tab
Rels
Nodes
Relations:
Size:
Save As New Relation
Options:
AND
OR
Inc:
Self-Loops
Dec:
Link wts -->
Width
Color