STATISTICAL HORIZONS

Introduction to Social Network Analysis

Stephen Borgatti, Ph.D.

Upcoming Seminar:
April 28-29, 2017, Philadelphia, Pennsylvania

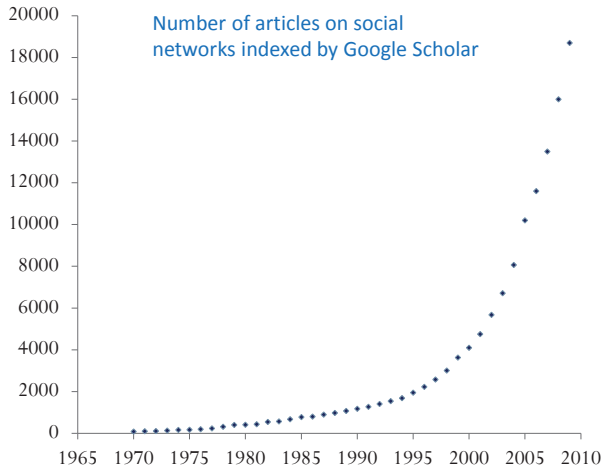
Introduction to SNA

Statistical Horizons ● Social Network Analysis

Steve Borgatti ● LINKS Center ● University of Kentucky

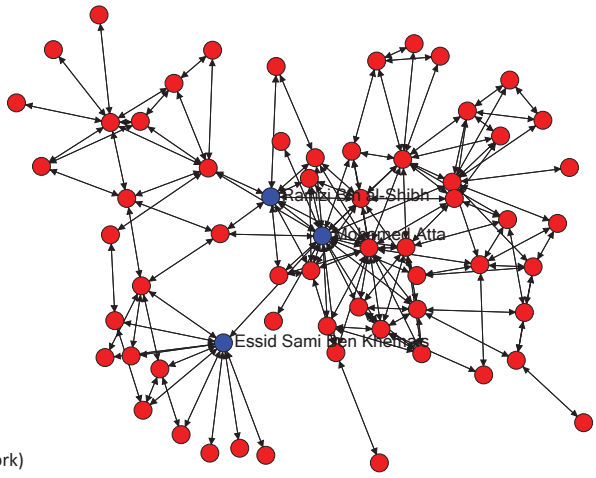
http://tinyurl.com/statisticalhorizons2016

Rise in popularity of network research



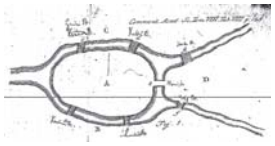
Lots of applied interest

- Health sciences
  - Epidemiology and patient support
- Management consulting companies
  - Boston Consulting Group (BCG)
  - Booz Allen Hamilton
  - McKinsey (through ex-student Rob Cross)
  - Arthur Andersen
  - CFAR (specialists in hospitals)
- Other companies
  - Merck, Pfizer, Novartis
  - BankBoston
  - Towers Perrin
  - Price Waterhouse (forensics & change)
- US govt
  - JWAC, US Army HTS, DTRA, NSA (both blue and red team work)
  - Civilian management



It's not exactly new

1700s- Euler
1930s- Moreno's Sociometry
Hawthorne studies
1940s Psychologists
Clique formally defined
1950s & 60s Anthropologists
Kinship analysis; society as network
1970s Rise of Sociologists
Small Worlds, Strength of weak ties; *Social Networks*; INSNA;
Sunbelt conference
1980s IBM computation
Computer programs developed
1990s Multi-disciplinary diffusion
Spread of network analysis to multiple fields; Social capital & embeddedness in vogue
2000s Physicists' "new science"
Scale free, small worlds, etc.



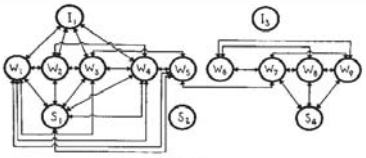
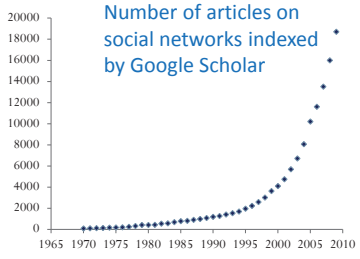
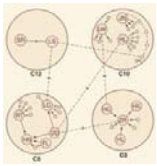
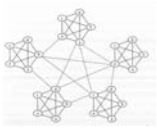


FIGURE 39
PARTICIPATION IN GAMES

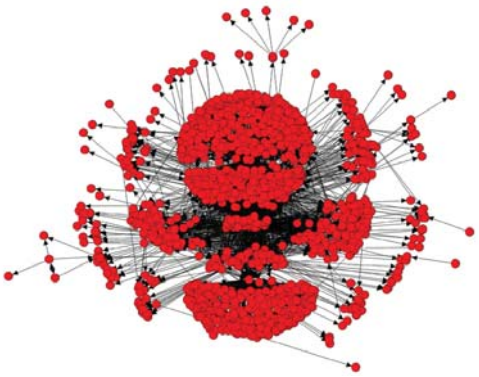


Number of articles on social networks indexed by Google Scholar





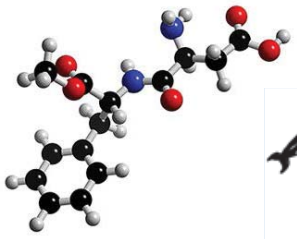
Why do we care?



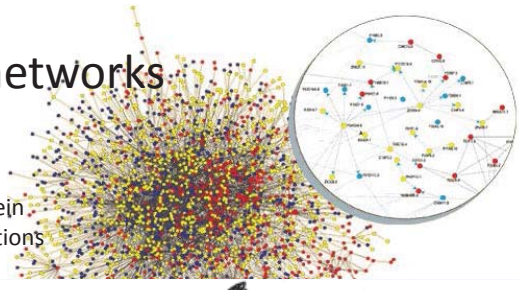
Nature organizes itself as networks

"network science"

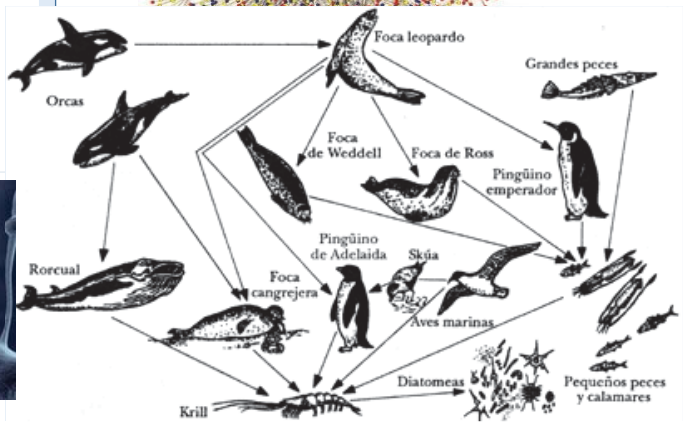
Aspartame sweetener



Neurons in the brain



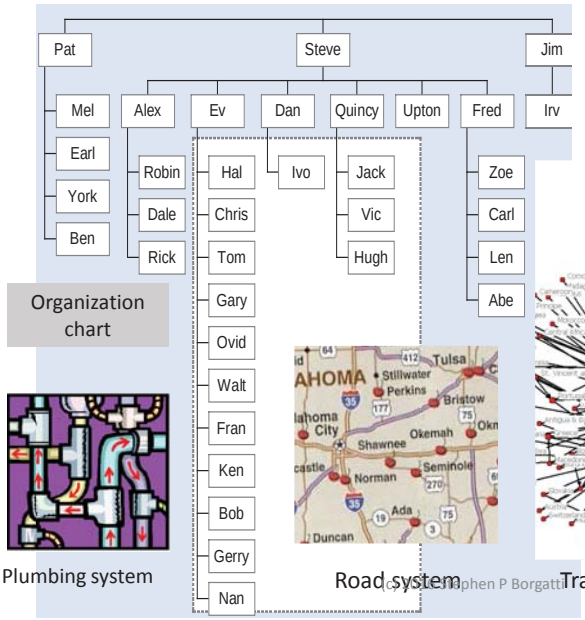
Protein reactions



Food web

… and so do humans

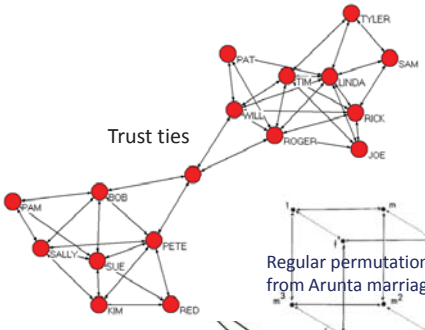
An effective way of organizing



Organization chart

Plumbing system

Road system



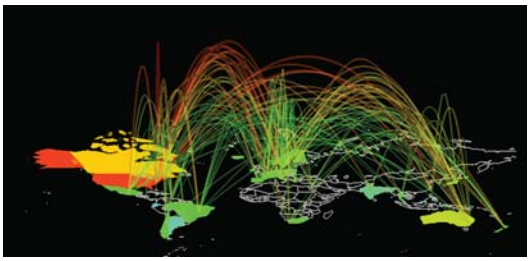
Trust ties

Regular permutation group arising from Arunta marriage system.

Trade relations among nations

Networks are everywhere

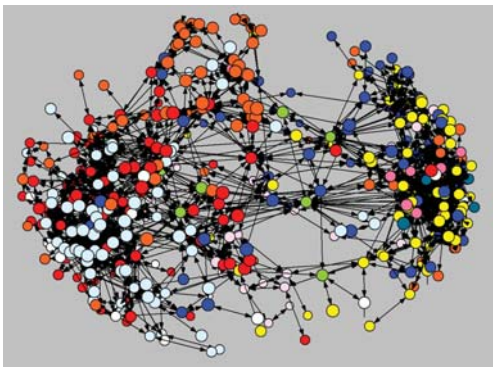
*or are they just a way of seeing the world?



- 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
- 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.

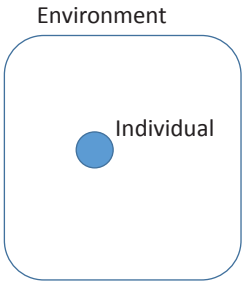
Characterizing SNA

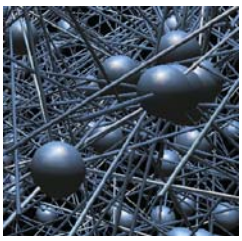
Characterizing network theorizing



Contextual

- 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

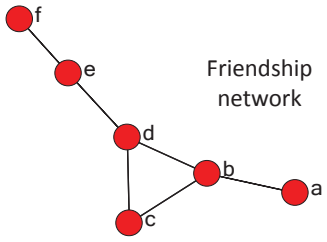




We are all embedded in a thick web of relations

Relational

- Traditionally, social science has focus on attributes of individuals to predict individual outcomes
  - Income as a function of education
- SNA puts the focus on relationships between individuals



Friendship network

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

Cases (entities)

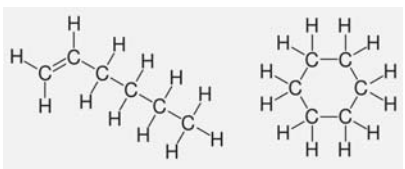
Variables (attributes)

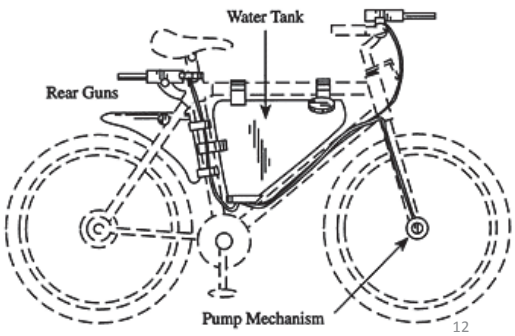
	Age	Sex	Education	Income
1001				
1002				
1003				
1004				
1005				
…				

Canonical data matrices
- Person by attribute
- Person by person

Structural

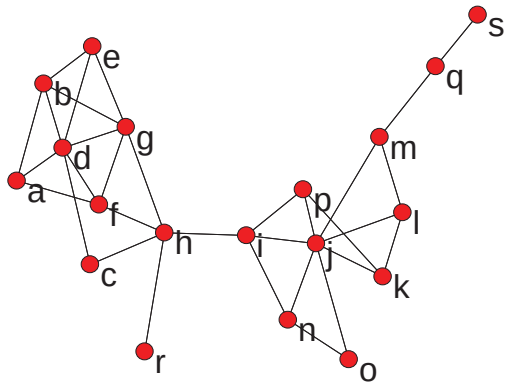
- It's not just relational (ties) but structural (pattern of ties)
  - To understand function, need to know more than list of elements. It's how they are connected
  - Non-reductionist, emergent flavor
- Indirect effects propagate
  - Power grids





Positional

- A node's position in a network determines in part the opportunities and constraints that it will face
  - Risk of news, risk of infection
  - Sense of identity
  - Individual social capital
- **Backcloth / traffic distinction**
  - Social ties provide conduits along which traffic can flow
  - A node's position in the network has significant implications for …
    - How early it encounters something flowing
    - How frequently it receives what is flowing
    - With what certainty it is reached



Summarizing the network perspective

Key Dimensions

- Contextual
  - It's the environment, stupid!
- Relational
  - By environment, we mean ties to others
- Structural
  - It's a network
  - Concepts and metrics for characterizing the network
- Positional
  - Location, location, location

The Flow Model

- Network theory largely …
  - Regards flows as the key mechanism underlying outcomes
  - Assumes the data we collect are about the roads that enable flows
  - Most of the conceptual machinery (e.g., centrality measures) is about calculating expected flows given the network structure and given some assumptions about how things flow

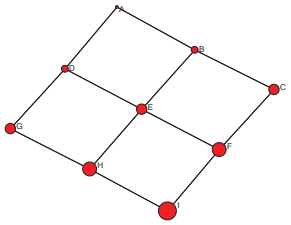
Network research has sought to explain …

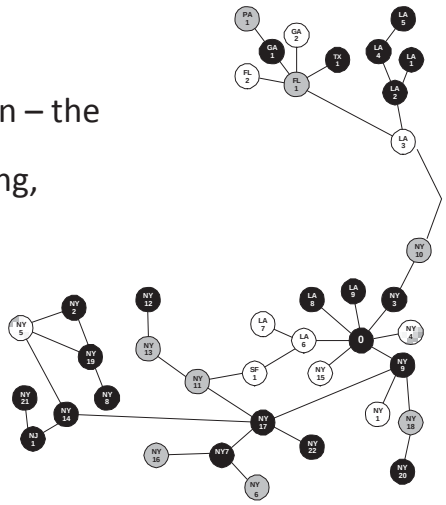
- Homogeneity
  - Why people have similar beliefs, behaviors, and belongings
  - Generic network explanation: contagion, diffusion, interpersonal influence processes
    - Contagion of obesity, happiness, etc
    - Diffusion of innovations
    - Spread of disease
    - Fads and fashion
    - Social conformity
- Achievement and reward
  - Why some people are more successful than others
  - Generic network explanation: social capital
    - Ties provide access to resources
    - Certain positions in social structures can be exploited for gain

Big Idea #1 -- Contagion

Discovery of HIV: Sexual contacts among gay men w/ unusual cancers, traced by Bill Darrow of the CDC

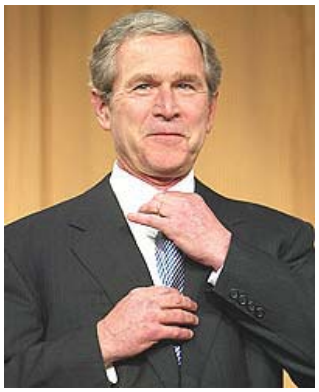
- Individuals influence each other
  - Infect each other with diseases, ideas, behaviors
- As a result, we observe network autocorrelation – the tendency for adjacent nodes to have similar characteristics such as opinions, ways of dressing, food preferences
- Flows of information, money
- The case of AIDS





Big Idea #2 – Social capital

- Why are some individuals more successful than others?
  - Attributes such as intelligence, motivation
    - Human capital
  - Who they know, who they owe
    - Social capital
  - Social ties provide access to resources the individual doesn't own/control directly



Types of network research

Antecedents and consequences

Mainstream Network Research

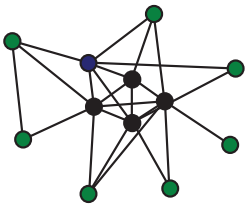
- Antecedents
  - Social processes that give rise to social ties, interactions, exchanges
    - And higher level constructs like popularity or network structure
  - Theory of networks
- Consequences
  - Mechanisms that translate ties into outcomes
    - Not just ties but network position and network structure
  - Network theory

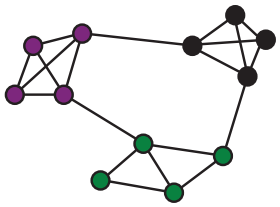
Cognitive SNA

- Antecedents
  - How ties & network structures are perceived by 3rd parties
- Consequences
  - Consequences of these perceptions
  - E.g., Being perceived to be friends with a high status other affects judgments of your influence (even more than actual friendship with high status other)

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





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

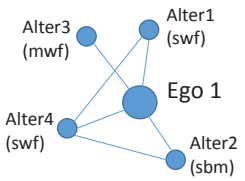
Research designs

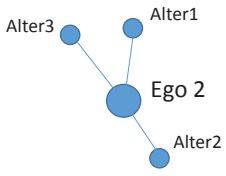
Whole network or sociocentric design

- 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 or 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.





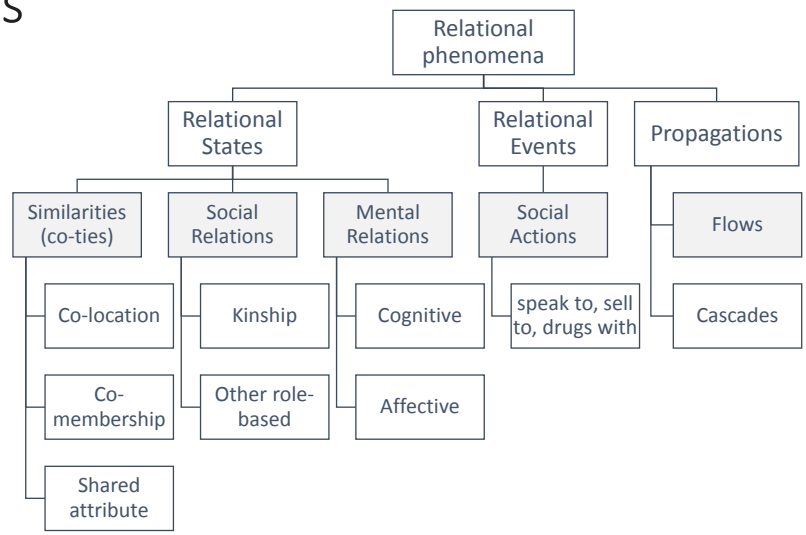
Cognitive social structures (CSS) design

- 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

Design comparisons

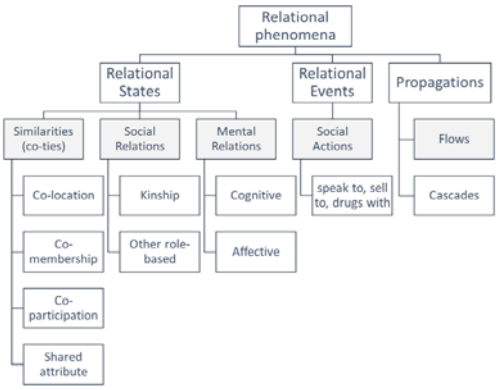
Full	Personal	Cognitive
Can compute all of the stats you can compute with personal design, plus more Compute global network measures like centrality Introduces significant challenges for statistical significance due to autocorrelation	Can use random samples and standard statistics to study large populations Can characterize node's network neighbor, e.g. demographic composition of friends Respondents (and alters) can be anonymous Tie data can be richer than in Full because of few names and anonymity Alter data is from ego's pov	Can do everything you can do with full network Can study perception of networks and how this impacts ego outcomes If survey-based, very tedious data collection -- requires small networks.

Kinds of ties



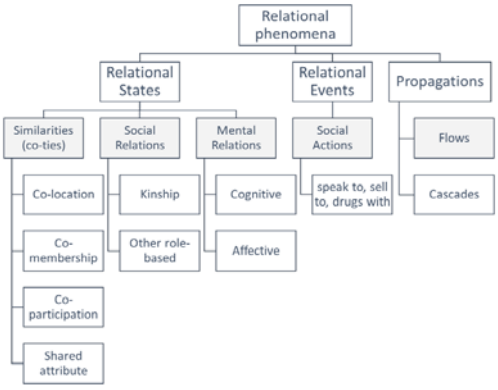
Relational states

- Relational states include …
  - Co-ties such as co-members of P&T
  - Kinship ties like son of
  - Other role-based ties like boss of, student of, friend of
- Relational states have an always-on character
  - While they hold, they hold continuously
- Relational states are things you are
  - I am my father's son



Relational events

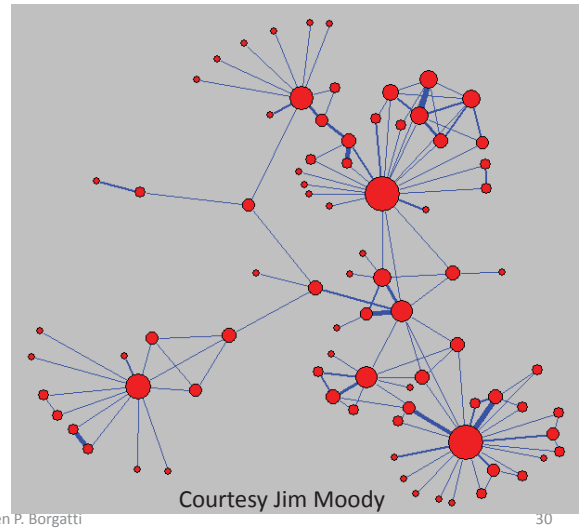
- Relational events include …
  - Have meeting with, send email to, ask question of
  - sex with, inject with, shake hands with
  - Transactions, e.g., a sale
- Relational events are discrete and transitory
  - They happen, then they are gone
- Relational events are things you count up, not things you are
  - # of lunches together vs son of



- Old definitions of networks sound like events, as in "recurring patterns of relations"
  - but include relations like friendship as examples

Building networks on events versus states

- Co-authorship network from Moody (2002)
  - Line shown between two nodes if they publish a paper together
- Note connectedness of network, ability for information to flow everyone to everyone
- But some node located more advantageously than others
  - Structural holes; centrality


Courtesy Jim Moody