

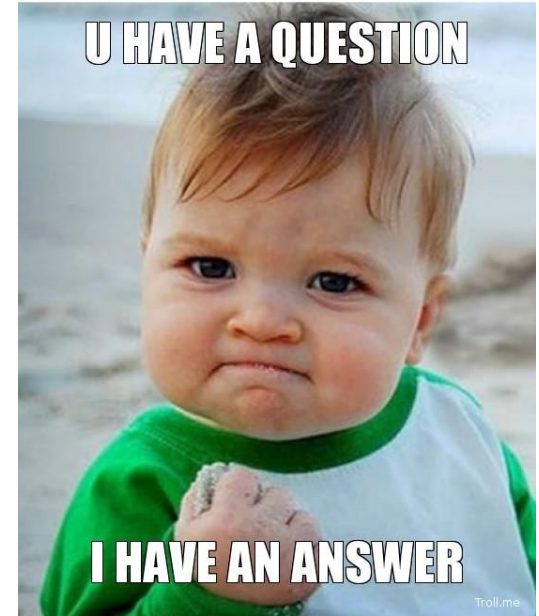
Introduction to Network Analysis

Ann McCranie
Workshop in Methods
December 2, 2016

These slides: <http://bit.ly/2gQrOoM>

Key questions today.

1. What is network analysis?
2. Why would I want to use it in my research?
3. What are the major approaches - and what are some of the common forms of analysis?
4. What kind of data do I need?
5. What kind of training do I need?
6. What sort of analysis tools would I use?
7. What are the resources available to me to get started?
8. What's happening at IU in network analysis?



What is network analysis?

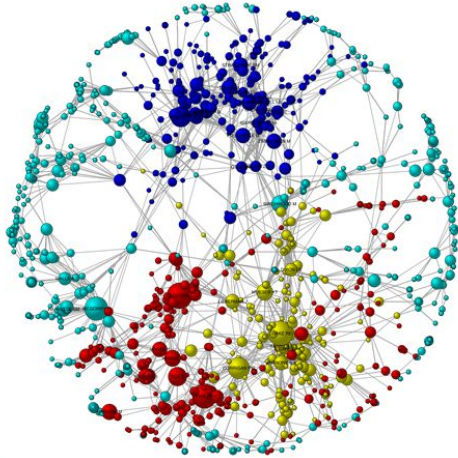
- It is of interdisciplinary interest – no discipline owns it.
 - Sometimes called “network science,” particularly outside social sciences.
- It’s not a unitary type of analysis: think of it more as a toolbox with many different widgets.
- It combines many statistical approaches you are already familiar with and some you probably aren’t.
- It’s not theory per se, but it should be informed by it
- I think of it as a “relational perspective”





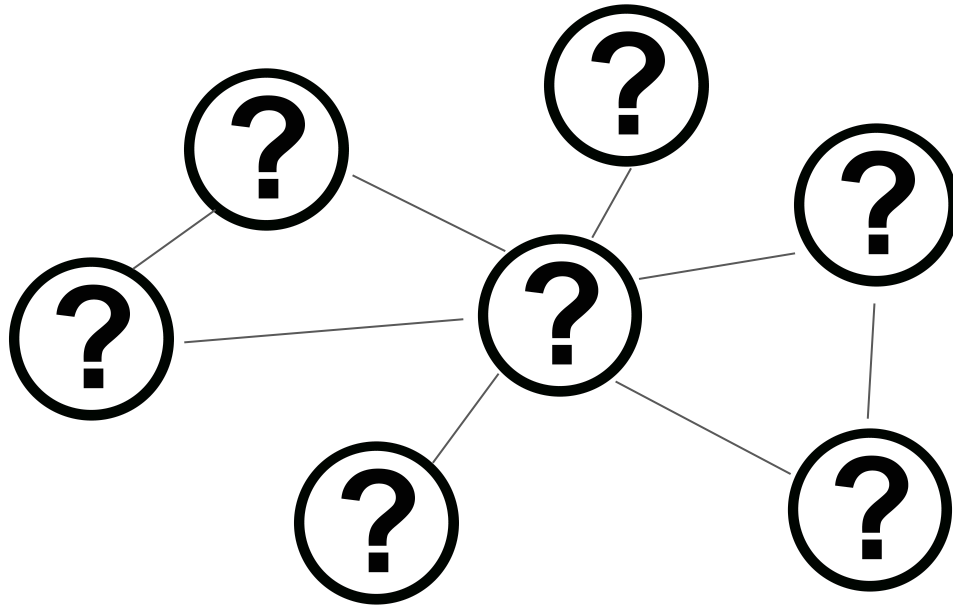
INDIANA UNIVERSITY

Relational perspective



- Actors and their actions are interdependent
- Ties between actors are channels for flow of something... resources, affect, information, etc.)
- Network ties both enable and constrain actions
- *Structure* is viewed as “*enduring patterns of relations*”

Why would I want to use network analysis?



What are the major types of network analysis?

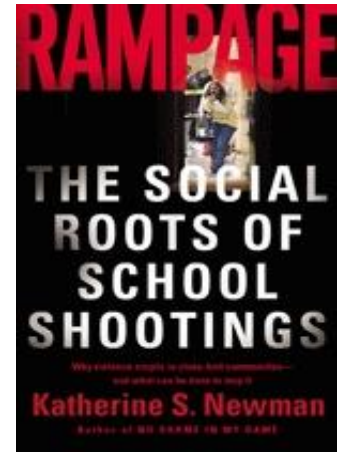
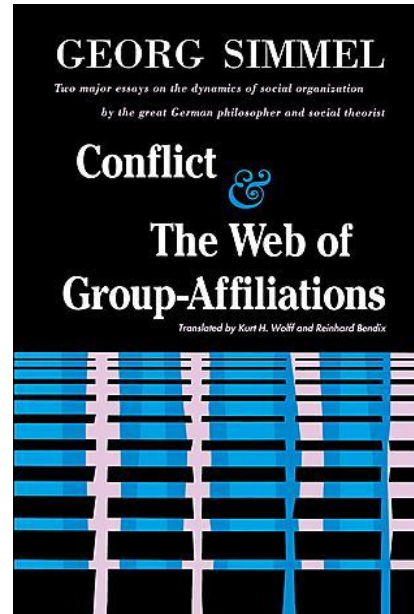
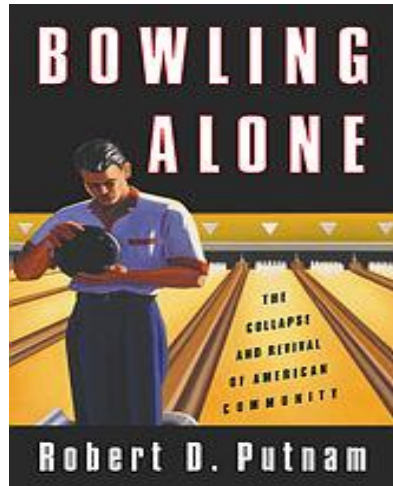
I like to think of them in three somewhat fuzzy categories:

1. Network-inspired
2. Egocentric or “personal networks”
3. “Whole” or “complete” (sometimes called “sociometric” or “sociocentric”)

“Network-inspired”

- This is work that takes seriously:
 - The embedding of people (or social “entities”) into larger groups of people and/or entities
 - The important interdependencies of BOTH individuals AND the relationships between them
 - The complexity of considering all of this at once

Examples of network-inspired work



Egocentric Networks

- By “ego,” we just mean that this is centered on an individual, and not a group
 - Egocentric work is typically thought of as work that collects information about a person’s social network ties without collecting information **from** those network ties
 - One recent excellent example: The Important Matters question on the GSS

The General Social Survey Question

From time to time, most people discuss important matters with other people. Looking back over the last six months—who are the people with whom you discussed matters important to you? Just tell me their first names or initials. IF LESS THAN 5 NAMES MENTIONED, PROBE: Anyone else?

J.M. McPherson, L. Smith-Lovin and M. Brashears, Social isolation in America, *American Sociological Review* 71 (3) (2006), pp. 353–375. (also, see errata, 2008)

ONLINE SUPPLEMENT
to article in
AMERICAN SOCIOLOGICAL REVIEW, VOL. 71 (JUNE:353–375)

Table 1. Size of Discussion Networks: 1985 and 2004 (1985 N =1531, 2004 N = 1426)

	Total Discussion Network		Kin Network ^a		Non-Kin Network ^a	
	1985	2004	1985	2004	1985	2004
0	8.10%	22.60%	24.40%	37.80%	36.10%	52.60%
1	14.80%	19.80%	29.70%	30.70%	22.30%	22.30%
2	14.70%	19.80%	22.60%	16.40%	19.80%	14.40%
3	21.60%	17.10%	13.10%	9.50%	12.70%	6.20%
4	15.40%	9.20%	6.70%	4.40%	6.10%	3.10%
5	20.00%	6.60%	3.50%	1.30%	3.00%	1.40%
6+	5.40%	4.80%				
Mean	3.06	2.12	1.58	1.16	1.39	.89
Mode	3	0	1	0	0	0
SD	1.88	1.99	1.45	1.38	1.49	1.35

^a Information on kinship was collected on the first five alters cited. Therefore, the sum of kin and non-kin alters is not equal to the overall network size distribution.

SEE ALSO: Paik and Sanchagrin (2013, ASR) for discussions of the problems with these findings.

J.M. McPherson, L. Smith-Lovin and M. Brashears, Social isolation in America, *American Sociological Review* 71 (3) (2006), pp. 353–375. (also, see errata, 2008)

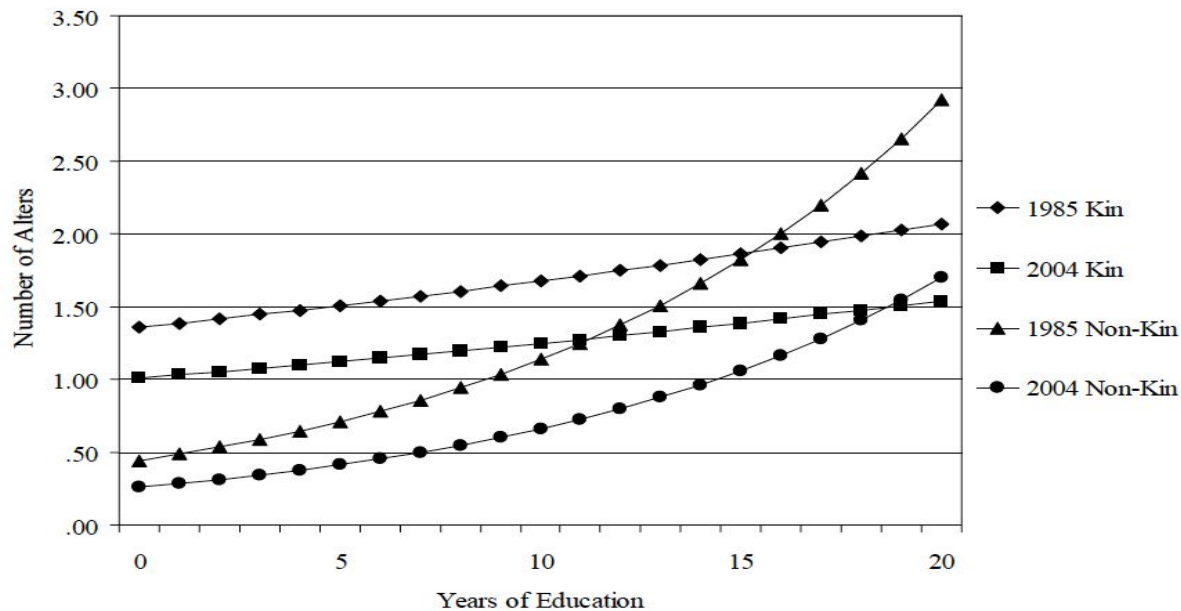
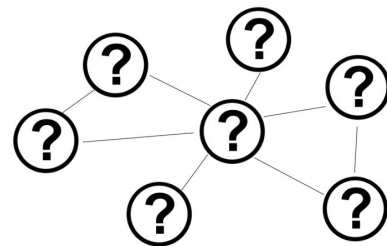


Figure 1. Ego Network Size for Kin and Non-Kin Ties, 1985 and 2004

Some of the things you can do with egocentric data

- Number of ties/alters mentioned
- Homogeneity of alters (on characteristics such as gender, race, etc.)
- Composition of alters (For instance, kin, non-kin)
- Ego's perceived network of the relationships between the alters (actually a complete network)
- Centrality measures of alters can be constructed
- Density measures
- Strength of ties between ego and alter AND between alter and alter
- Change over time (attrition, turnover)



Chicago Health and Social Life Survey

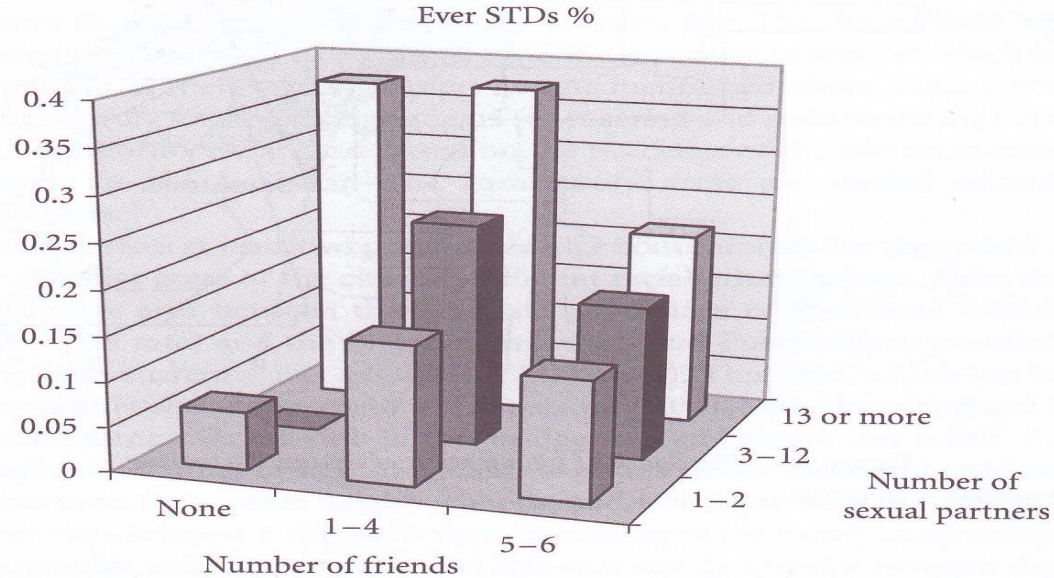


Figure 1.4. Effects of Social Networks on STDs

Source: Laumann et al. (2004).

It's not such different data...

- Social support and social capital literature often focuses on egocentric data
- You likely already have many statistics tools at your disposal to do egocentric analysis (although there are, of course, many sophisticated and specific ways to deal with the data).
- The key point here is that you can assume that individuals in your study or sample are independent of one another, at least in respect to the social network information that you are collecting about them
- This assumption is important, and is the key difference between egocentric and “complete” data

Whole/Complete Networks

- Sometimes people call this “sociometric” or “sociocentric” to contrast to egocentric
- Centered on a population, a defined group of people or social entities.
- Requires some sort of **boundary** to your population
- In this work you are either collecting information on direct relationships/interactions/transfers/etc. with people or on shared affiliations between individuals.
- The key here is that you are collecting information on the the relationships - the interdependencies - of individuals in a group.

Key Terms in Complete Network Analysis

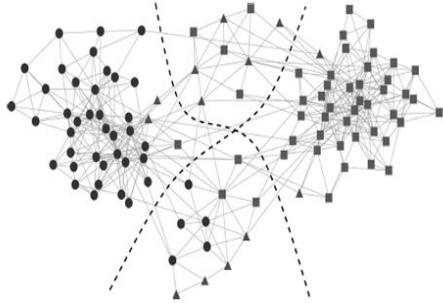


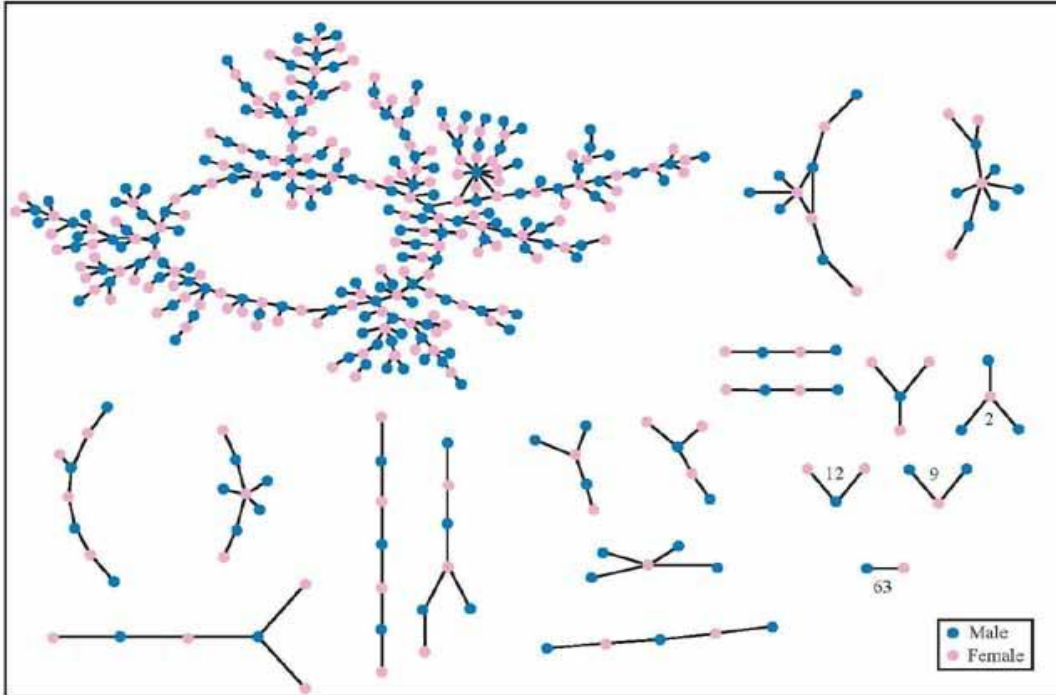
Fig. 3. Krebs' network of books on American politics. Vertices represent books and edges join books frequently purchased by the same readers. Dashed lines divide the four communities found by the algorithm, and shapes represent the political alignment of the books (circles are liberal, squares are conservative, and triangles are centrist or unaligned).

- Node/Actor/Vertex: The “individual”
- Tie/Line/Edge/Arc: The relationship between them
- Matrix: An array of elements, rows and columns, that represent the relationships between nodes in the network
- Attribute: A favored term for a characteristic of an individual (race, class, gender...)
- Generator: The question used to garner relationship information

Newman (2006)
“Modularity and
community structure in
networks”

Whole/complete network examples

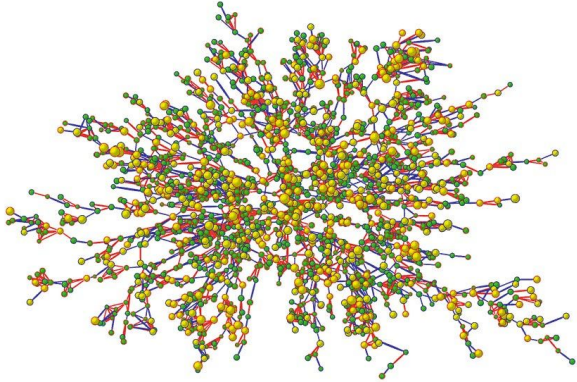
The Structure of Romantic and Sexual Relations at "Jefferson High School"



Each circle represents a student and lines connecting students represent romantic relations occurring within the 6 months preceding the interview. Numbers under the figure count the number of times that pattern was observed (i.e. we found 63 pairs unconnected to anyone else).

Bearman PS, Moody J, Stovel K.
"Chains of affection: The structure of adolescent romantic and sexual networks"
American Journal of Sociology, Vol. 100, No. 1.

Framingham Heart Study & Networks



- Obesity Spreads through networks:
- [July 2007: The New England Journal of Medicine. “The Spread of Obesity in a Large Social Network over 32 Years” by Nicholas A. Christakis and James H. Fowler.](#)

Framingham Heart Study Background

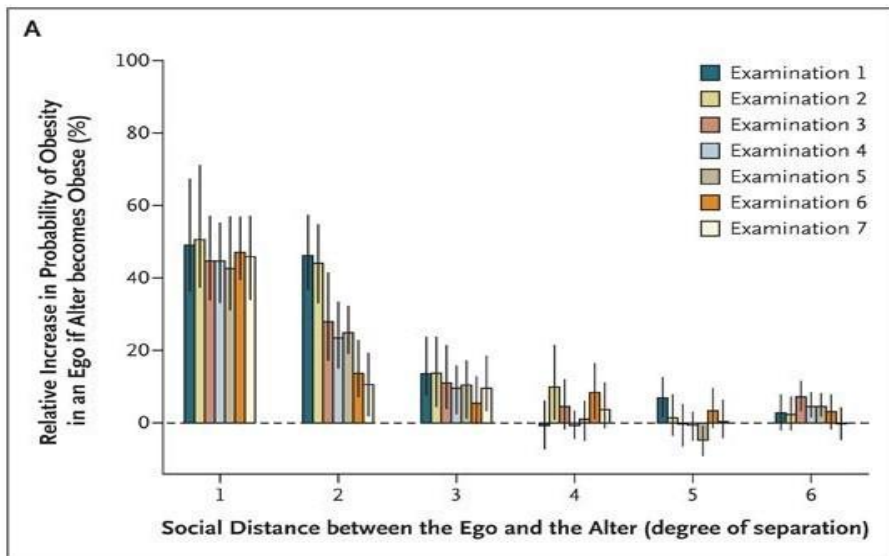
- Took place in Framingham, MA from 1948, still continuing
- Is one of the most famous medical health studies, gave us such findings as the connection of cholesterol and blood pressure with cardiovascular problems
- These network studies involves 5124 people in the "Offspring Cohort", followed from 1971 to 2003
- For contact information, asked who their friends were (also collected: family, spouses, neighbors)
- A network was constructed - each person had about 7.5 friends, family, or spousal ties)

The Spread of Obesity Among Friends and Family

(Christakis and Fowler 2007)

- Major findings:
 - It's not just that people selectively choose people as friends or spouses who weigh the same as they do...
 - A person's risk of becoming obese increases by the weight status of friends (57% increase), siblings (40%) and spouse (37%). The weight of their neighbors didn't seem to matter.
- Also influential: the weight of the friends of their friends...

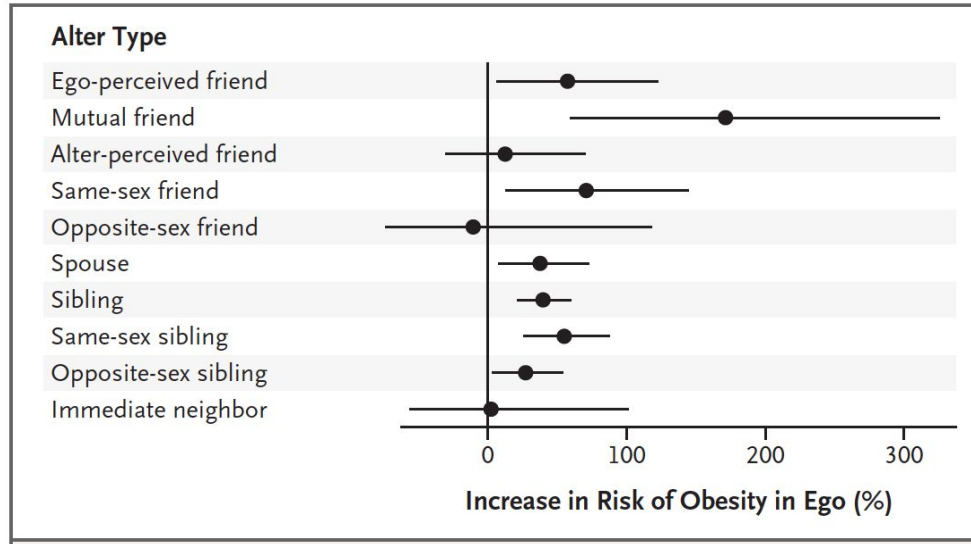
The Spread of Obesity, Implications



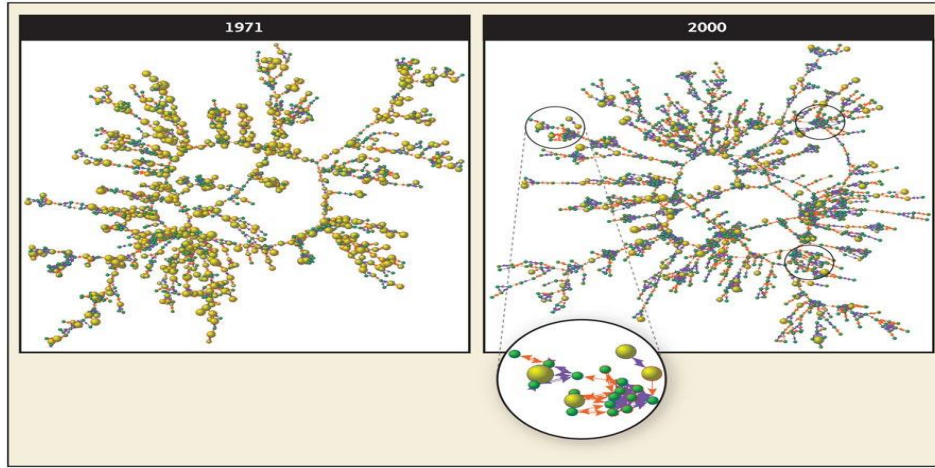
- Obesity has an important relational and social component. "Social distance" matters in this case more than physical distance.
- Approaches to deal with this problem (which the authors argue is a public health problem, not just a clinical health problem of individuals) could benefit from efforts to involve peers.

But it's not just negative risk we share...

The Spread of Obesity, Implications



The Cessation of Smoking

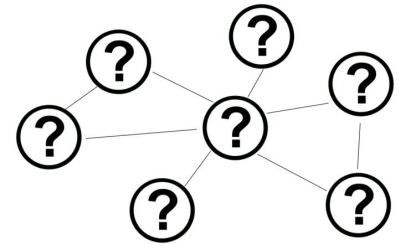


<http://content.nejm.org/cgi/content/full/358/21/2249/DC1>

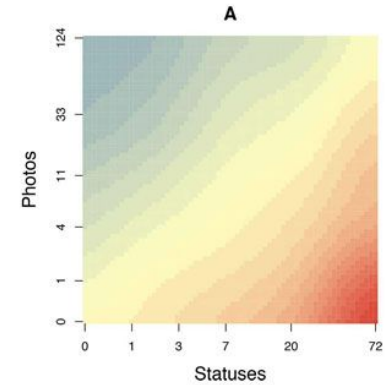
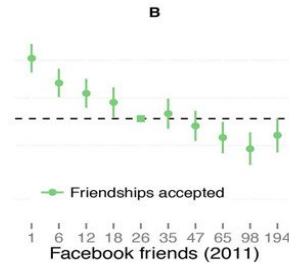
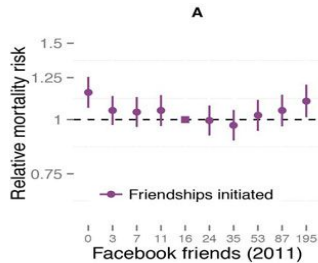
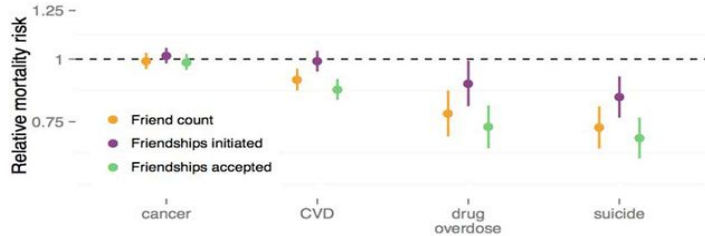
- Whole groups quit at once.
- Over time, smokers move to the periphery.
- Spouse: 67%; Sibling: 25%; Friend: 36%; Coworker at small workplace: 34%.
- Having more education meant influence was higher.

The Cessation of Smoking, cont.

- This study suggests that the decision to quit smoking is not made solely by isolated individuals.
- Collective pressures work.
- Social policies (like smoking laws, taxes) might be an adjunct, but one of the big ways to influence people to smoke is to get other people not to smoke.
- More education means that people wield more influence - could start to explain differences in SES and risk behavior.



Newer Work: Social Media Use and Health



Mortality increases (red) as the ratio of statuses to photos declines

From: Hobbes, Burkes, Christakis and Fowler (2016) PNAS

<http://www.pnas.org/content/113/46/12980>

Mortality decreases as the individual receives and accepts friendship requests for overall # of friends. Also declines, but less, for friend requests sent.

Robust Action and the Rise of the Medici, 1400-1434

John F. Padgett; Christopher K. Ansell. *The American Journal of Sociology*, Vol. 98, No. 6. (May, 1993), pp. 1259-1319.

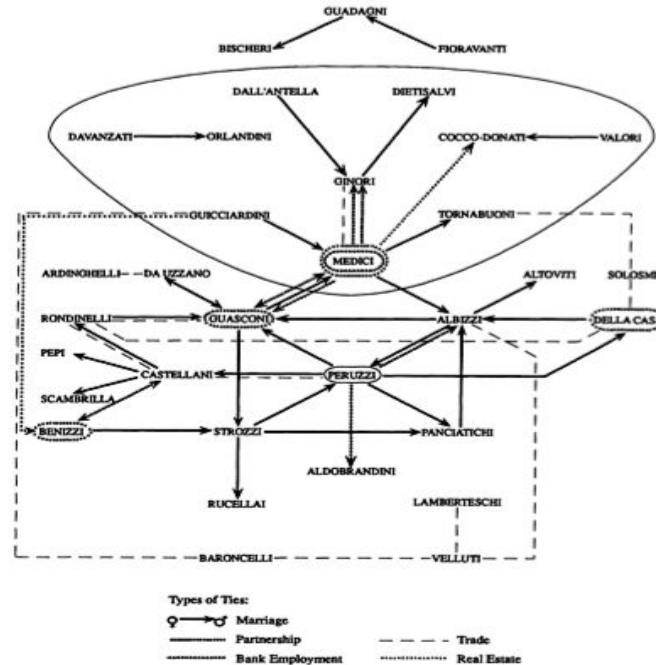


FIG. 2a.—Marriage and economic blockmodel structure (92 elite families)

Basic Network "Dimensions"

- *Size*: Typically refers to the number of nodes a network has
- *Density*: What percentage of possible ties the network actually has
- *Directionality*: Whether or not a relationship can flow in both directions (directed, asymmetric) or whether it by definition must be the same for both parties (undirected, symmetric)
- *Value*: Whether the tie is simply present or not (unvalued, unweighted, binary) or whether there is a value attached to the ties (valued, weighted, continuous, etc.)
- *Mode*: Whether the tie is about a direct relationship between two actors (one mode, adjacency) or whether it is about an actor and an event (two mode, affiliation)
- *Multiplexity*: Whether the relationship is of a single type, or whether there are multiple relations between the same actors
- *Time*: Whether the data is cross sectional or time-dependent, "dynamic"

When someone asks you what “type” of network you have...

- Relations can be directed or undirected. In a directed relation the tie from Actor A to Actor B is made “independently” from the tie from Actor B to Actor A.
 - Directed: “From whom do you seek advice?”
 - Unidirected: “To whom are you married?”
- They can be valued or binary.
 - Valued: Jane and Joe publish ten articles together. (10)
 - Binary: Jane seeks advice from Joe (1), but Joe doesn’t seek advice from Jane (0).
- They can also be multiplex - there are multiple types of relationships at once.
 - Friends AND/OR business partners AND/OR enemies AND/OR mentors





Many kinds of networks

- *Affective/evaluative* (e.g., friendship, liking, nomination, respect)
- *Support* (sending emotional or social support)
- *Transfers* of resources (business transactions, lending, borrowing)
- *Sharing* (information, infections)
- *Interaction* (working together, shaking hands, sending messages)

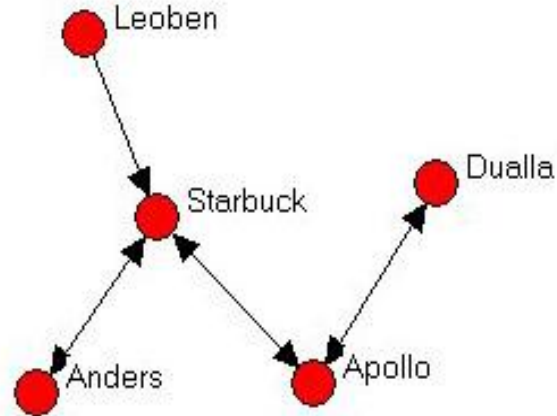


More kinds of networks...

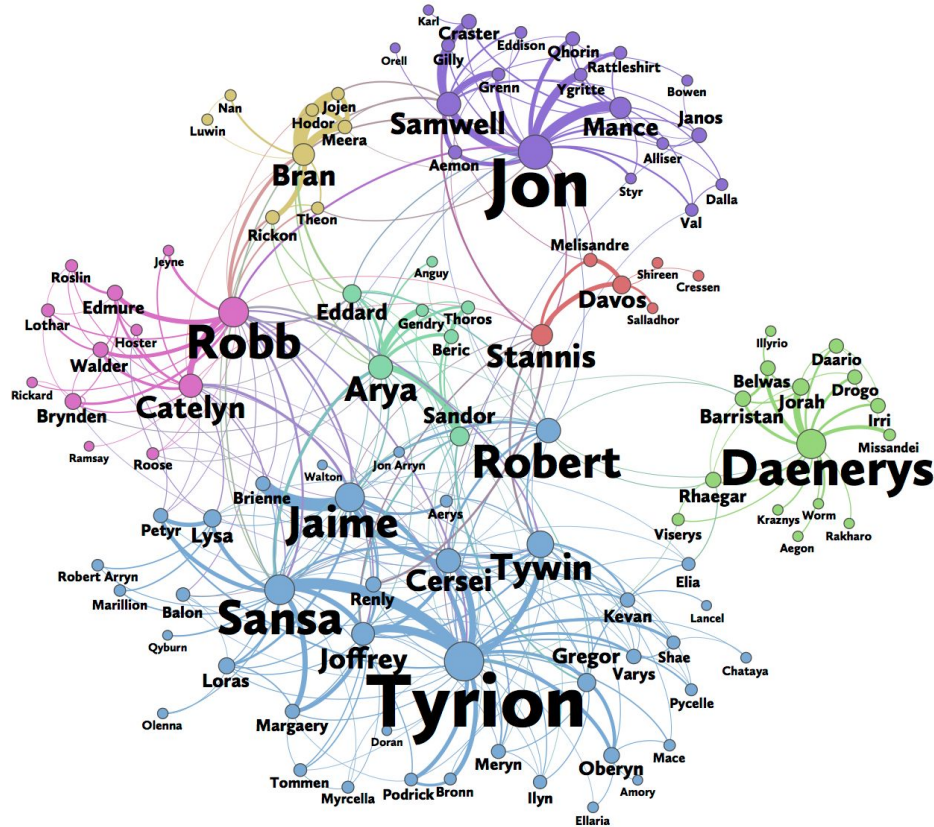
- *Movement* between places or statuses (migration, social or physical mobility)
- *Physical connection* (a road, river, or bridge connecting two points)
- *Formal relations* (organizational chart)
- *Genealogical* (kinship or descent)
- And a special kind:
 - *Association or affiliation* (jointly attending an event, belonging to the same group)

What sociometric data looks like

		1	2	3	4	5
	A	D	S	A	L	
		-	-	-	-	-
1	Apollo	0	1	1	0	0
2	Dualla	1	0	0	0	0
3	Starbuck	1	0	0	1	0
4	Anders	0	0	1	0	0
5	Leoben	0	0	1	0	0



A little more current pop culture reference



From “A Network of Thrones” based on third book, a Storm of Swords.

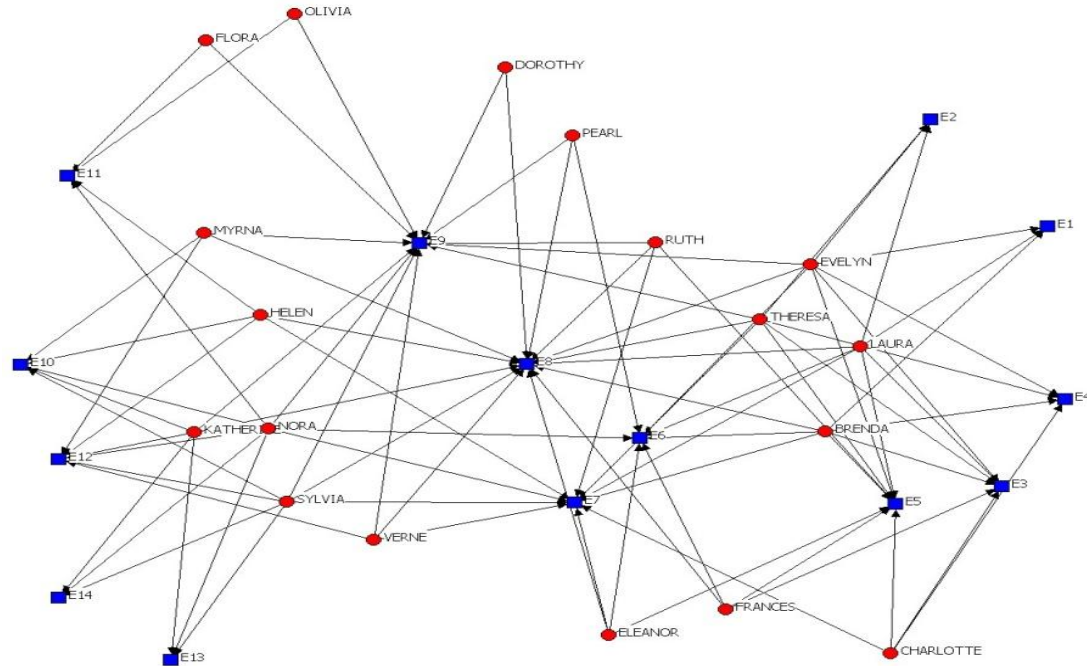
<https://www.macalester.edu/~abeverid/thrones.html>

Data can also be “one-mode” or “two mode”

- Affiliations represent non-direct relationships between actors.
- Called “two-mode” because it has “actors” and “events”
- Two-mode data is often transformed into one-mode data.
 - Caution: this assumes that some sort of aggregation of “events” equals some strength of a “relationship” between actors. Think about this carefully before you proceed down this path with your own work.

Classic Affiliation dataset: Davis Southern Ladies

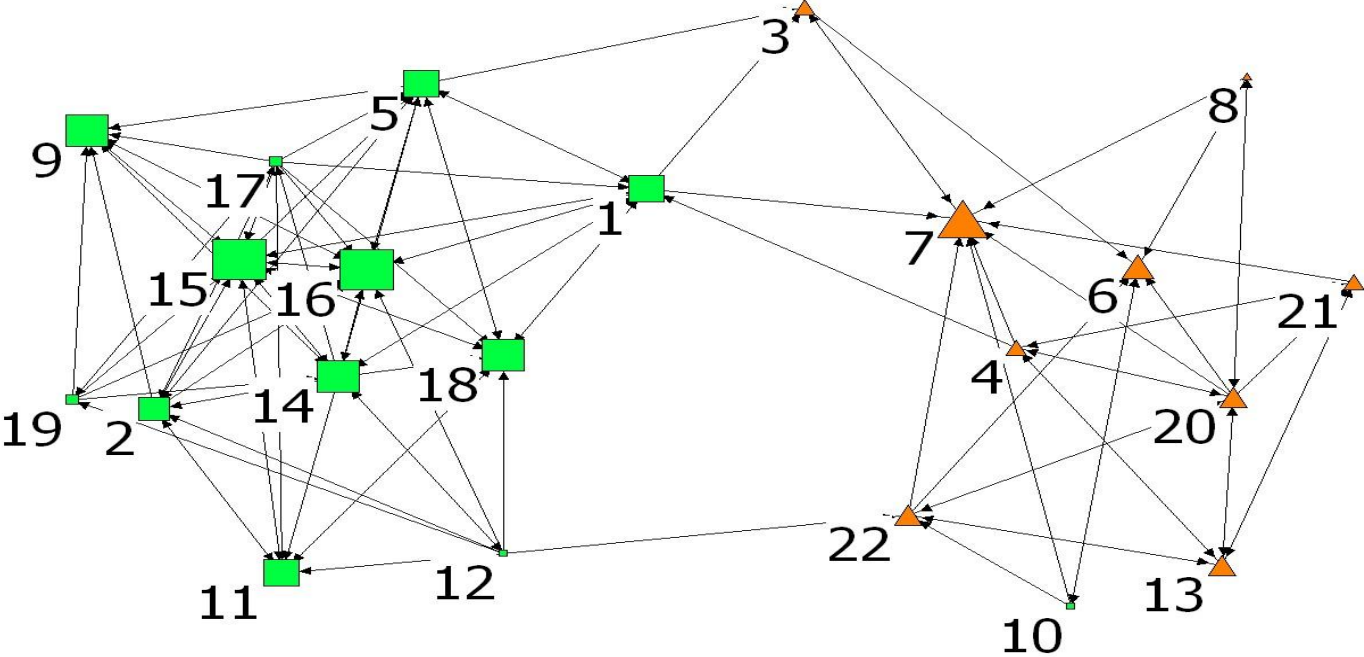
		1	2	3	4	5	6	7	8	9	0	1	2	3	4
	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E
1	EVELYN	1	1	1	1	1	0	1	1	0	0	0	0	0	0
2	LAURA	1	1	1	0	1	1	1	1	0	0	0	0	0	0
3	THERESA	0	1	1	1	1	1	1	1	1	0	0	0	0	0
4	BRENDA	1	0	1	1	1	1	1	1	0	0	0	0	0	0
5	CHARLOTTE	0	0	1	1	1	0	1	0	0	0	0	0	0	0
6	FRANCES	0	0	1	0	1	1	0	1	0	0	0	0	0	0
7	ELEANOR	0	0	0	0	1	1	1	1	0	0	0	0	0	0
8	PEARL	0	0	0	0	0	1	0	1	1	0	0	0	0	0
9	RUTH	0	0	0	0	1	0	1	1	1	0	0	0	0	0
10	VERNE	0	0	0	0	0	0	1	1	1	0	0	1	0	0
11	MYRNA	0	0	0	0	0	0	0	1	1	1	0	1	0	0
12	KATHERINE	0	0	0	0	0	0	0	1	1	1	0	1	1	1
13	SYLVIA	0	0	0	0	0	1	1	1	1	1	0	1	1	1
14	NORA	0	0	0	0	0	1	1	0	1	1	1	1	1	1
15	HELEN	0	0	0	0	0	0	1	1	0	1	1	1	0	0
16	DOROTHY	0	0	0	0	0	0	0	1	1	0	0	0	0	0
17	OLIVIA	0	0	0	0	0	0	0	0	1	0	1	0	0	0
18	FLORA	0	0	0	0	0	0	0	0	1	0	1	0	0	0



Three basic levels of analysis

1. Individual nodes (or even ties!)
2. Subgroups
3. Complete network

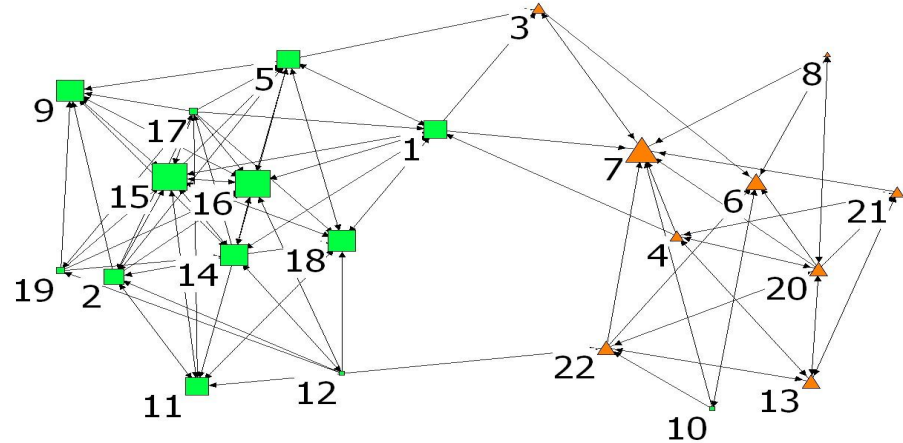
Fifth Grade Network



Levels of Analysis of Sociometric Data

Individual

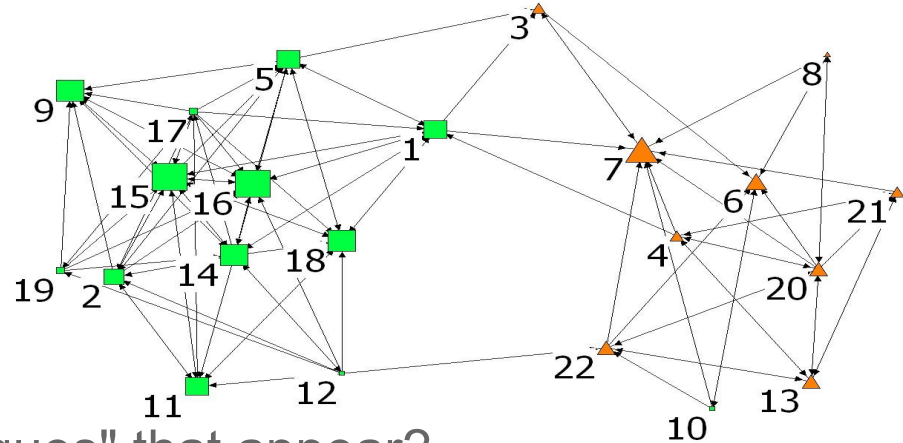
- Example: Who is the most popular student in this class? How popular is each student?



Levels of Analysis of Sociometric Data

Subgroup

- Example: What are the differences between the girls and boys? How are the groups connected? How are they different?

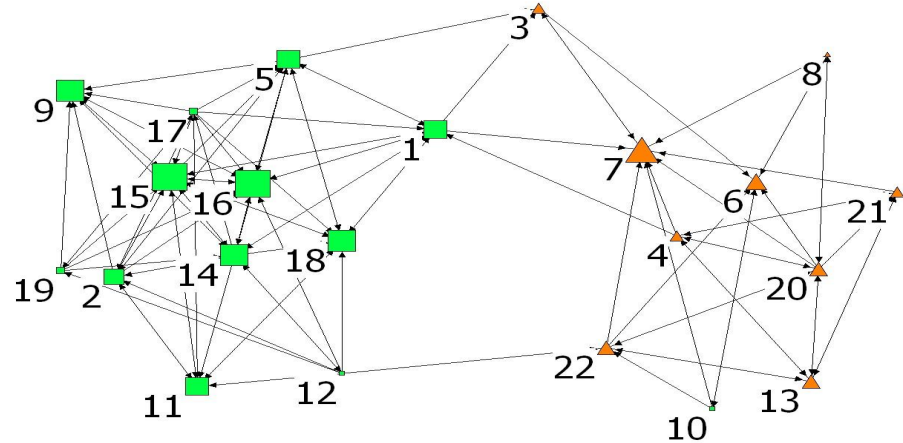


- Are there any obvious "cliques" that appear?

Levels of Analysis of Sociometric Data

Whole Network

- Example: How cohesive is this network? How dominated is it by a popular kid?

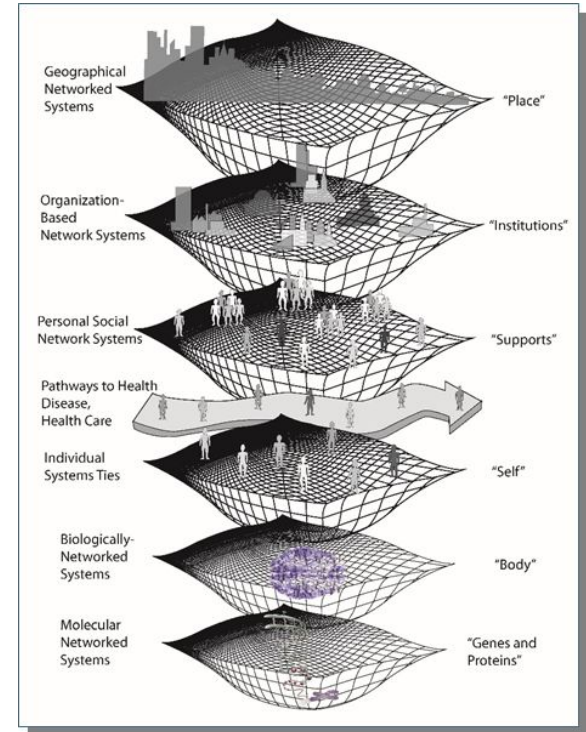


Another way of thinking of levels: Social Symbiome

The Social Symbiome, based in Network and Complex Systems science, for human health, disease, and health care.

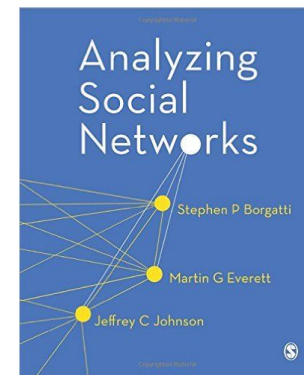
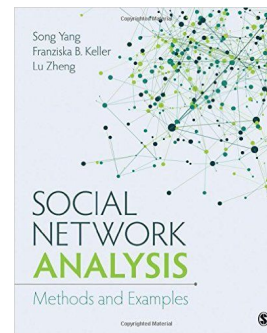
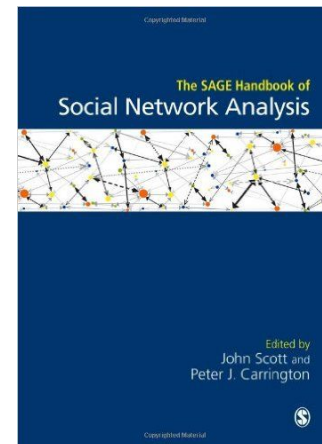
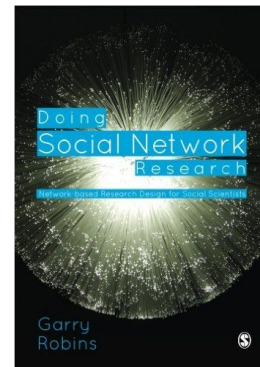
Pescosolido, B., S. Olafsdottir, B. Perry, O. Sporns, E. Meslin, T. Grubestic, J. Martin, L. Koehly, W. Pridemore, A. Vespignani, T. Foroud, & A. Shekhar. 2015. "The Social Symbiome Framework: Linking *Genes-to-Global* Cultures in Public Health Using Network Science." In *Handbook of Applied Systems Science*, Z. Neal, ed. New York: Routledge, Inc.

Pescosolido, B.A. 2015. "Linking the Social Brain to the Social World through Network Connections." Pp. 247-279 in *Social Neuroscience: Brain, Mind, and Society*. R. Schutt, L.J. Seidman, and M.S. Keshavan, eds. Cambridge, MA: Harvard University Press.



Common types of analysis

- Centrality
- Dyads and triads
- Subgroup analysis
- Clustering, Grouping, community detection
- Positional Analysis - block modeling
- Network characteristics
- Statistical modeling of networks
- **VISUALIZATION**



Visualization can take different forms

<http://www.global-migration.info/>

Visualizing migration between world regions with an interactive circle hierarchy.



What kind of data do I need?

For sociometric data, you must have high quality data.

- Ideally, 100% response rates
- Quality is seriously compromised for some types of measures quickly the lower you drop - some measures are more sensitive
- A rule of thumb is 85-90% at very minimum response, but really, even that is problematic
- Don't plan a study you know you can't get anywhere near that
- Don't try to switch to "egocentric" for a low response sociometric study

Can be collected through surveys, archive, observation.

Key questions today.

1. What is network analysis?
2. Why would I want to use it in my research?
3. What are the major approaches - and what are some of the common forms of analysis?
4. What kind of data do I need?
5. What kind of training do I need?
6. What sort of analysis tools would I use?
7. What are the resources available to me to get started?
8. What's happening at IU in networks?

Refer to the handout,
also available:

<http://bit.ly/2gDXCwh>

Spring 2017 IUB courses

Informatics

I590 Network Science (Online)

INFO-I 590 TOPICS IN INFORMATICS (3 CR)
VT: NETWORK SCIENCE
33905 Ahn Y

Coming up, Fall 2017:
Intro to Networks,
Wasserman, P657

Economics

ECON-E 724 SEMINAR IN ECONOMIC THEORY (3 CR)
VT: NETWORK FORMATION GAMES
10567 11:15A-12:30P TR WY 329
Page F

Getting started!

- Textbooks
- Datasets
- Conferences
- Software

My guide:

<http://bit.ly/2gDXCwh>

Introduction to Network Analysis

Workshop in Methods

December 2, 2016

Instructor: Ann McCranie, Indiana University Network Science Institute (IUNI)



This document can be viewed at <http://bit.ly/2gDXCwh>. You are also welcome to download slides and lab materials from my previous longer workshops here, organized by topic here: <http://annmccranie.net/site/CPSR.html>. *Opinions expressed are purely my own, and not those of IUNI.*

Recommended Introductory Textbooks for Social Network Analysts

There are many excellent textbooks specific to particular fields (economics, physics, health, security studies) or for specific software packages (such as Pajek or R), but the first three books below (from left) provide excellent high-level treatments of network analysis, particularly whole or complete analysis. They are suitable for a graduate level course, or an independent researcher trying to learn about the field. The fourth book *Another text*, forthcoming in 2017 from Cambridge, by Perry, Pescosolido, and Borgatti, will address egocentric networks specifically.



More listed on this regularly updated Amazon List: <http://a.co/aU8k0De>

Journals that have network focus (primarily social)

Network Science: <https://www.cambridge.org/core/journals/network-science>

Social Networks <http://www.journals.elsevier.com/social-networks/>

Journal of Complex Networks <http://comnet.oxfordjournals.org/>

Connections http://insna.org/connections_archives.html

Journal of Social Structure: <https://www.cmu.edu/joss/> (Online only, undergoing restructuring)

Social Network Analysis and Mining: <http://www.springer.com/13278>

What's happening at IU in this area?

- Talk series: Network Science Talks
- Open Science Forums (here in Spring)
- 170 faculty who work in networks
 - Open to people interested, not just those already involved!
- PhD minor
- Phd in Networks and Complex Systems in Informatics
- NetSci 2017 in Indianapolis, hosted by IUNI

JOIN IUNI-DISCUSSION-L, follow us on Twitter and Facebook

