

Introduction to Social Network Analysis

SICSS Norrköping 2025

Carl Nordlund

www.carlnordlund.net

Purpose/content

Introduction to social(-scientific)
network analysis

Seeing and thinking in terms of
relational data

Basic toolkit of (descriptive)
methods

Hands-on workshop/lab (R)

First principles



Introductions

Carl Nordlund

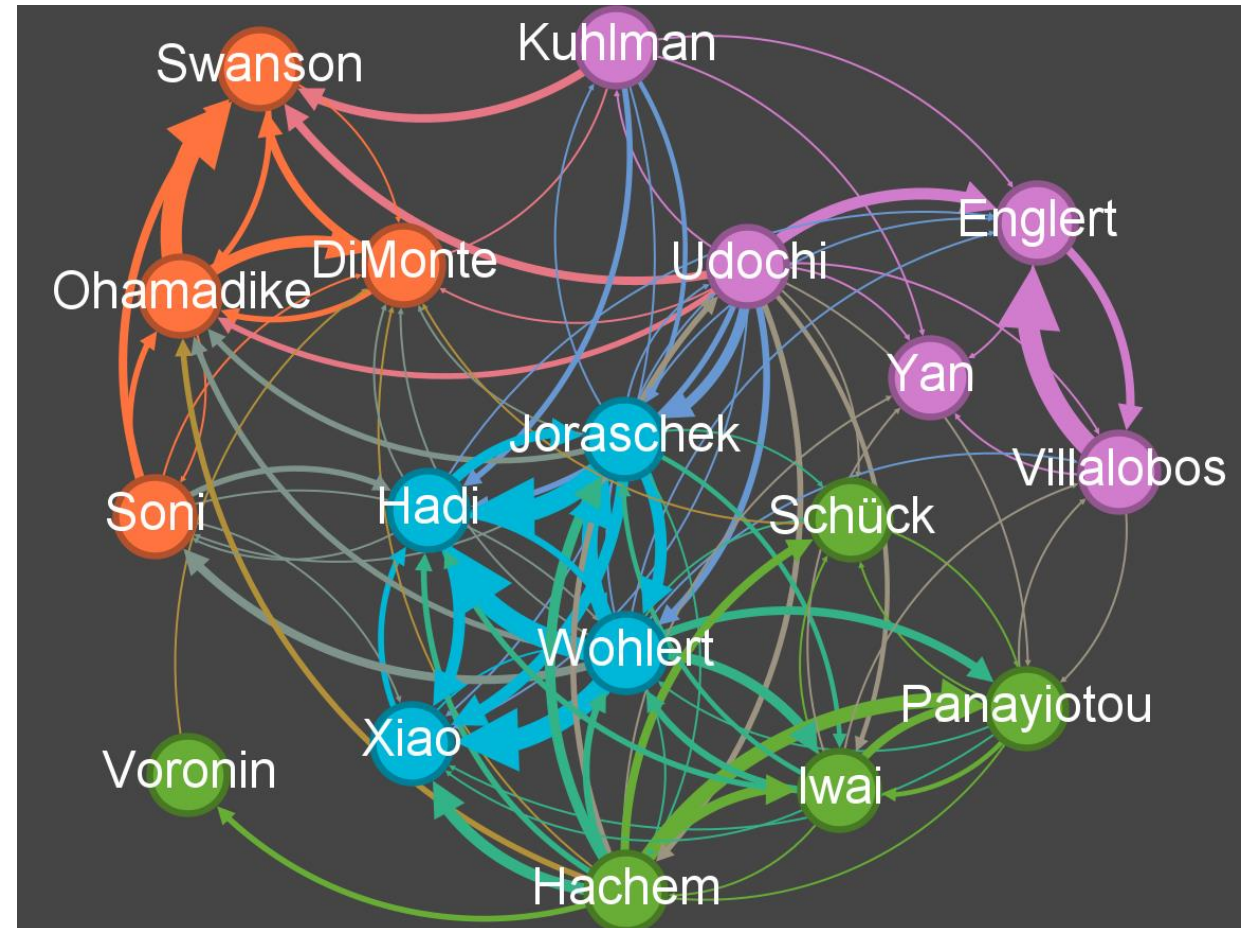
- Associate professor @ IAS
- PhD human ecology, Lund university
- MSc comp sci studies
- Governo Insikt AB
- carlnordlund.net

Students of SICSS 2025

- Roster with names (from website)
- Based on this, I conducted some network analysis to see how you are all connected

The SICSS class of 2025 (version A)

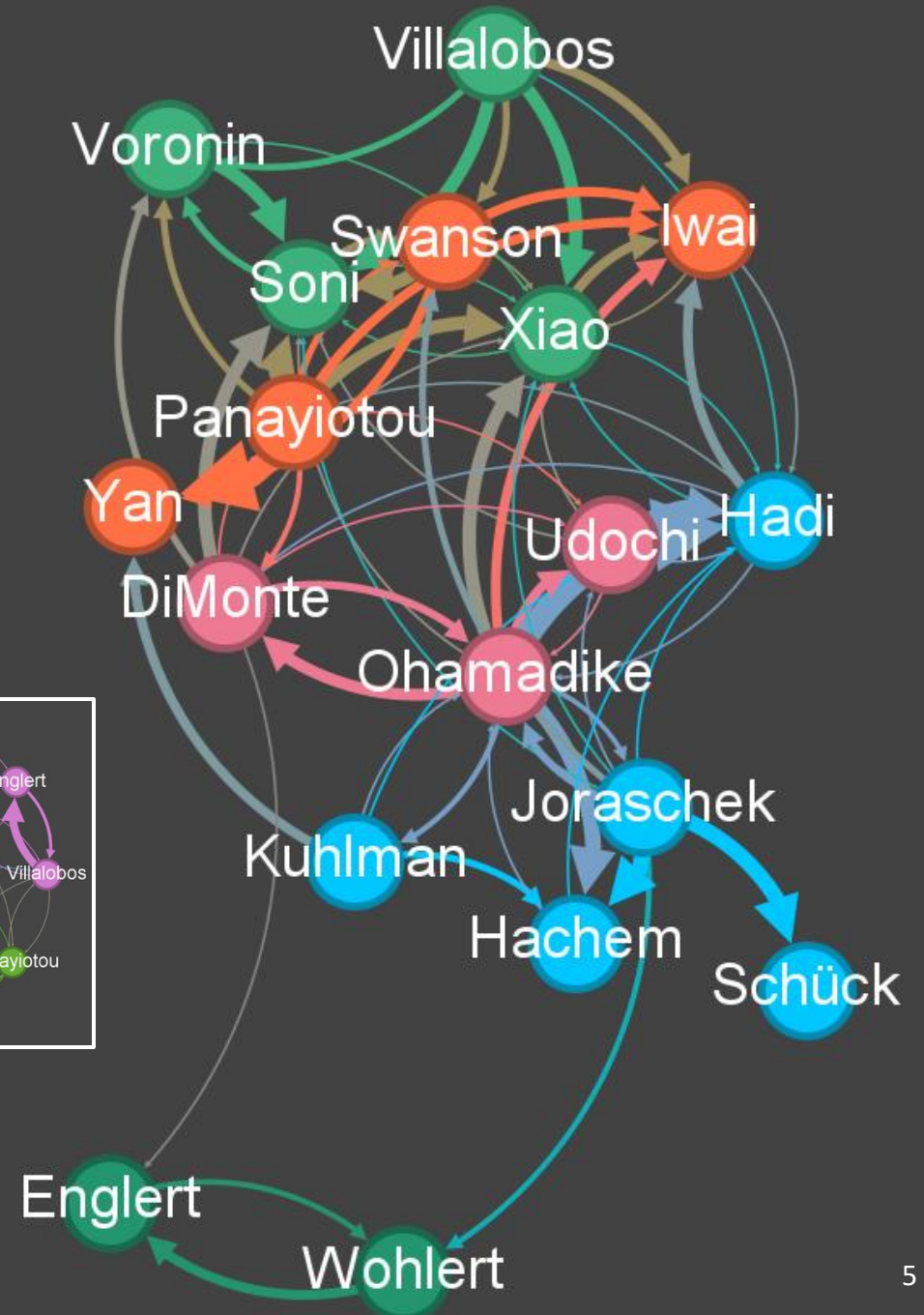
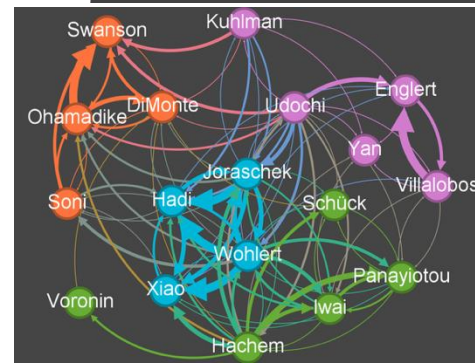
- Directional, valued (normalized) network
- Mostly non-reciprocal (full self-ties)
- Four communities (Girvan-Newman)
- Valued blockmodeling: ambiguous findings



The SICSS class of 2025 (version B)

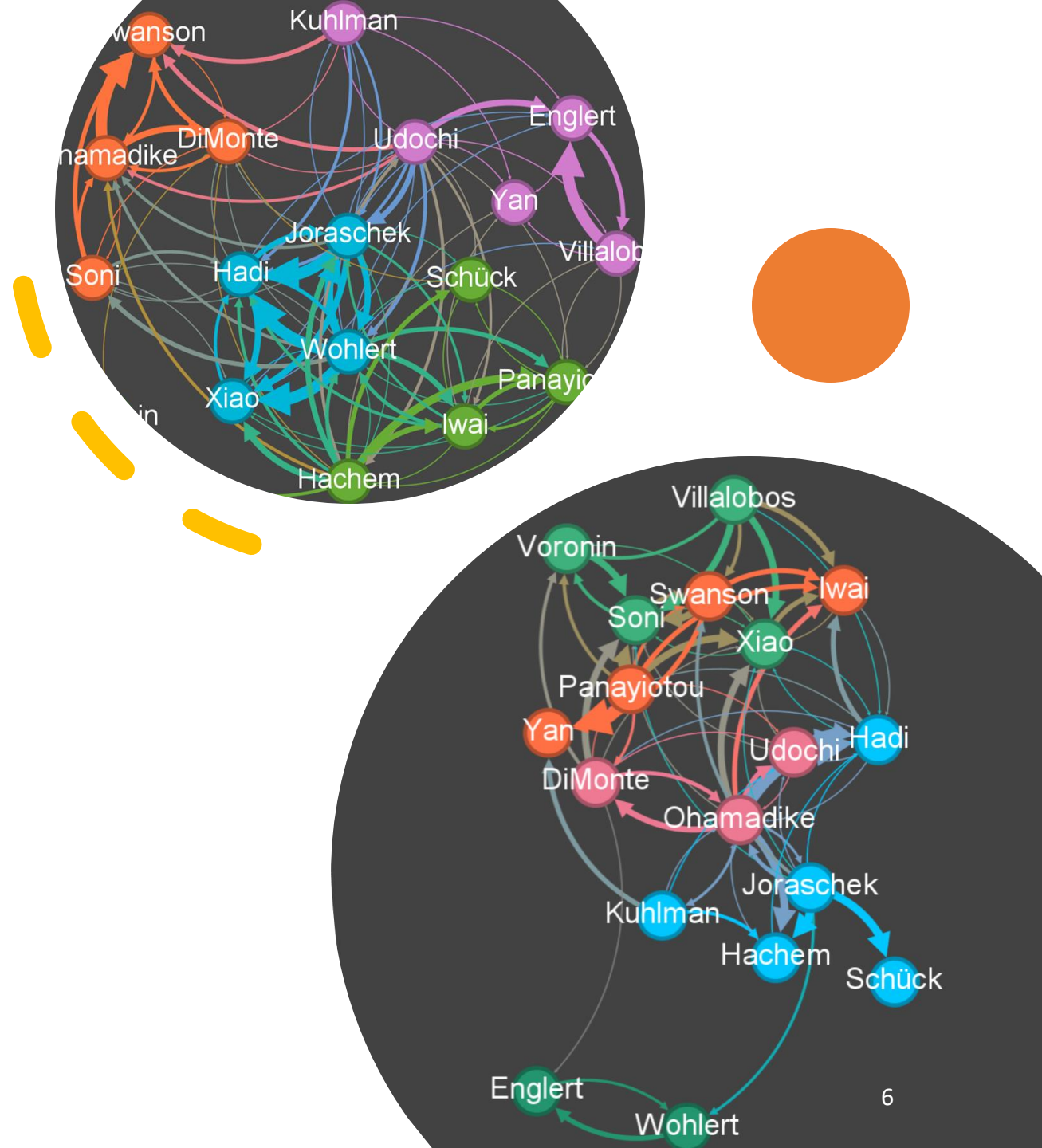
- Different data than version A
- Less dense
- Five communities
- Previously connected: now disconnected
- Some notable full ties (100%)

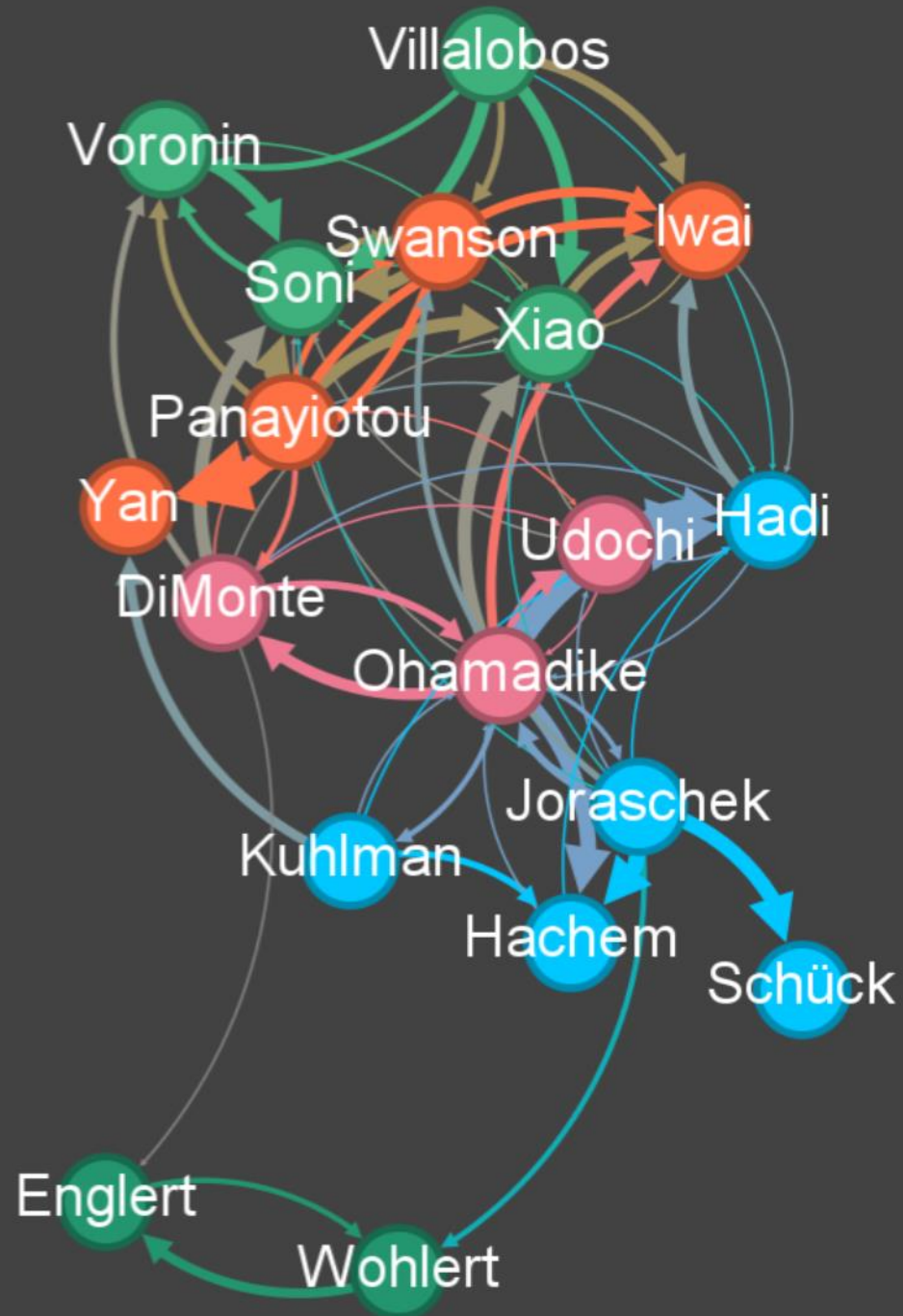
What is this weird stuff?!



These are so-called Peter networks

- Captures the share of shared letters in your first (A) and last (B) names
- Edge weight(a->b): how much of b's name that can be reconstructed given the letters in a's name
- Prune away all edges with weights < 0.5
- 'Relations' solely derived from letters as found in class roster!
- Subsequent community detection and interpretations





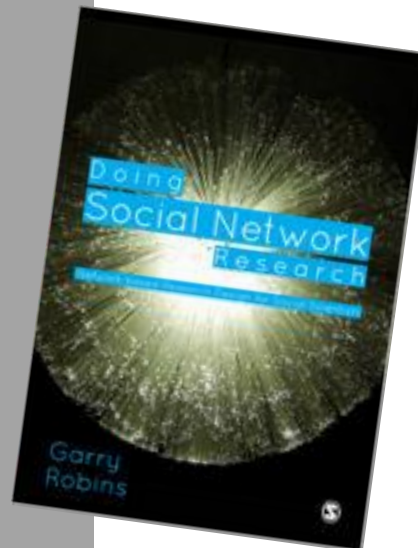
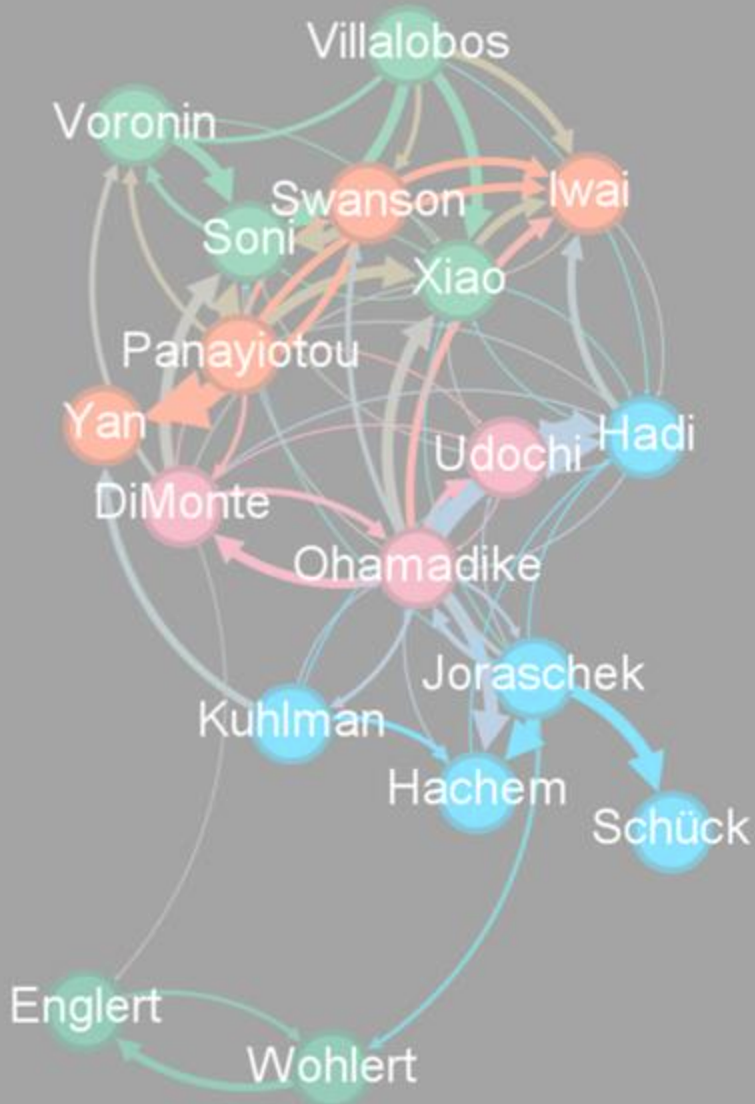
Usefulness of Peter networks?

- None. Zero. Zilch.
- A fantastic example of generated pseudo-relational data with zero social-scientific use
- ...that still can be generated, visualized, and analyzed in various ways

...which paves the way to...

Reasons for network analysis


- To understand how individuals (or other social entities) in certain social position have different individual outcome
- To understand how individuals affect social structure
- To facilitate this: (as always) intricate interplay between:
 - Research question
 - Data
 - Methods
 - Theory
- Peter networks: perhaps for visualization methods



Robins (2015), Doing Social Network Research:
Network-based Research Design for Social Scientists

Traditional panel data

A




Age: 35
Income: 40k
Education: BSc
Gender: Male

B



Age: 20
Income: 30k
Education: MA
Gender: Female

C




Age: 41
Income: 42k
Education: n/a
Gender: Female

D




Age: 52
Income: 53k
Education: n/a
Gender: Male

E



Age: 21
Income: 34k
Education: MA
Gender: Male

F



Age: 56
Income: 28k
Education: PhD
Gender: Female

Properties (attributes)
of individual entities

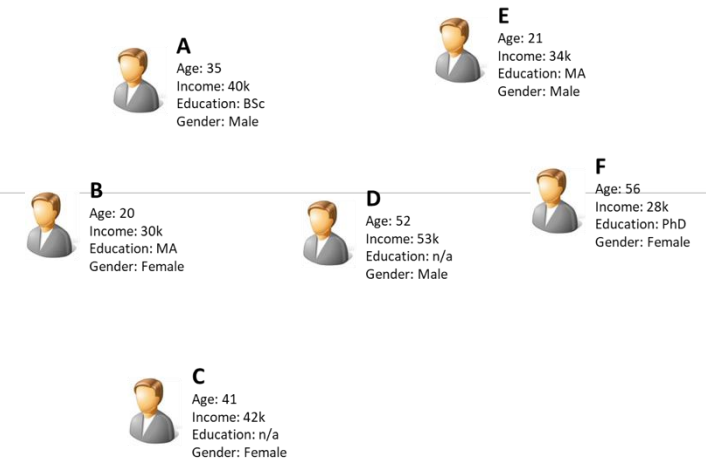
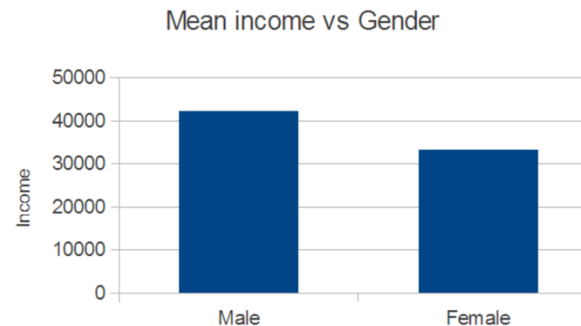
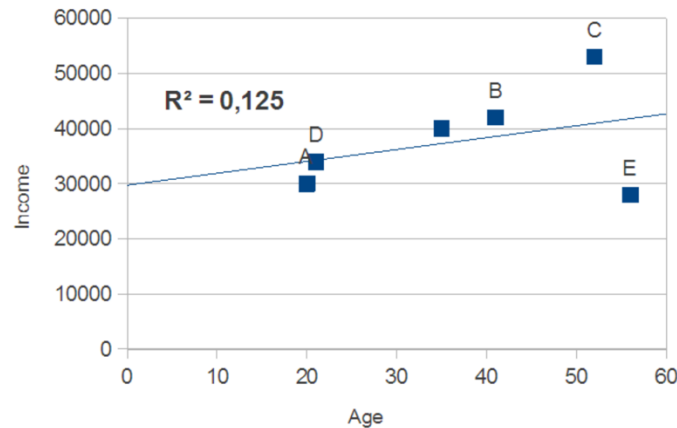
Usually far more
than six

Label	Age	Income	Education	Gender
A	35	40k	BSc	Male
B	20	30k	MA	Female
C	41	42k	n/a	Female
D	52	53k	n/a	Male
E	21	34k	MA	Male
F	56	28k	PhD	Female

Traditional panel data

- Comparing two or more 'attributes' among entities
- Identifying would-be associations between variables
- Methods typically assume independence between entities and their attributes
- Sampling possible

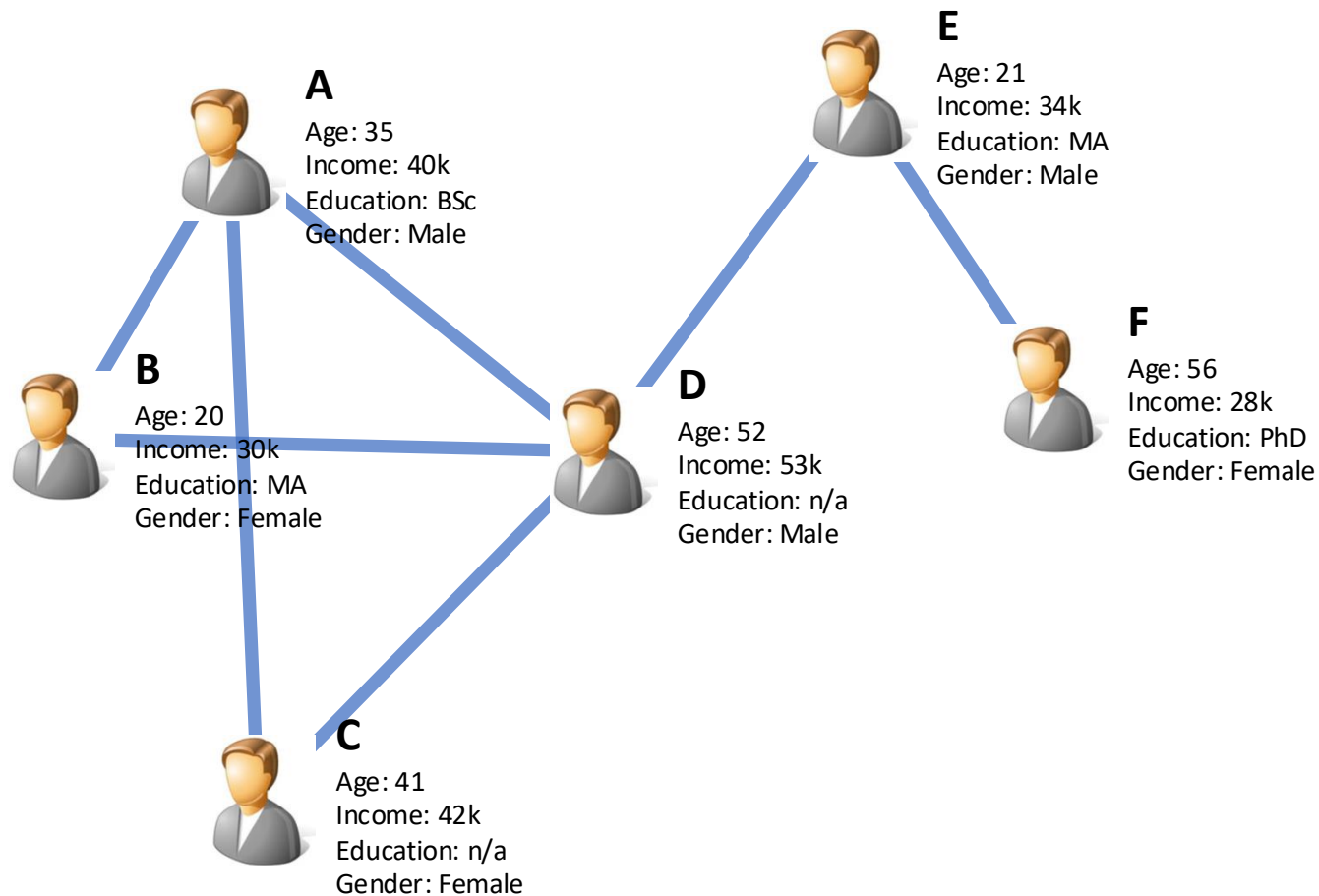
Examining statistical associations



Label	Age	Income	Education	Gender
A	35	40k	BSc	Male
B	20	30k	MA	Female
C	41	42k	n/a	Female
D	52	53k	n/a	Male
E	21	34k	MA	Male
F	56	28k	PhD	Female

Relational data

- Entities: actors / nodes / vertices (here: social ones, e.g. individuals)
- Data: sets of relations *between* entities (e.g. friendship)



Friendship (reciprocated) [matrix-format]

	A	B	C	D	E	F
A	-	1	1	1	0	0
B	1	-	0	1	0	0
C	1	0	-	1	0	0
D	1	1	1	-	1	0
E	0	0	0	1	-	1
F	0	0	0	0	1	-

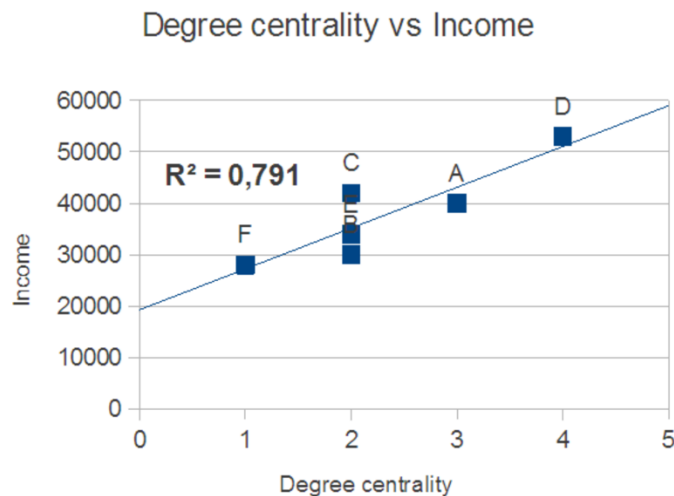
Relational data

- Entities: actors / nodes / vertices
- Data: sets of relations between entities (e.g. friendship)

Measuring relations

Degree centrality

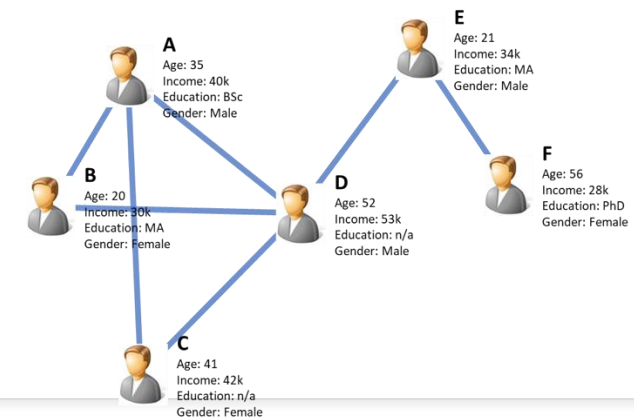
Rudimentary individual (micro-level) measure of centrality:
Number of ties that a node has



- Linking to attributional data
- Likely needs proper null model (rewiring)
- Mere associations

Friendship (reciprocated) [matrix-format]

	A	B	C	D	E	F
A	-	1	1	1	0	0
B	1	-	0	1	0	0
C	1	0	-	1	0	0
D	1	1	1	-	1	0
E	0	0	0	1	-	1
F	0	0	0	0	1	-



Relational data

- Capturing the in-between
- Inter-dependence of observations
 - Removing a 'data point' can have grand implications
- Sampling typically problematic
 - System boundaries for full populations important
- Supplementing 'conventional' (attributional) data

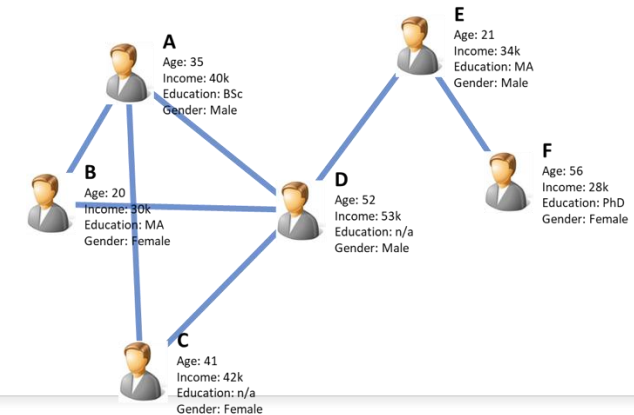
Ethical considerations!

The Robins reminders

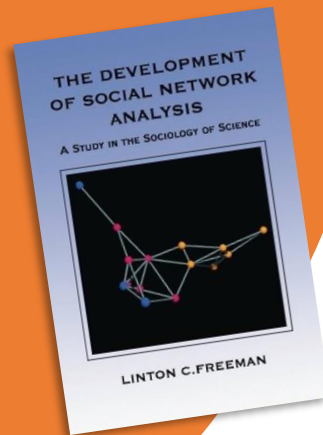
- Incorporating network data and metrics far more than a mere methodological extension
- Explicit theoretical commitments about structure and dependence: *that the in-between matters!*
- Going beyond *Hobbesession* with individual, assumedly independent entities to understand social world

Friendship (reciprocated) [matrix-format]

	A	B	C	D	E	F
A	-	1	1	1	0	0
B	1	-	0	1	0	0
C	1	0	-	1	0	0
D	1	1	1	-	1	0
E	0	0	0	1	-	1
F	0	0	0	0	1	-

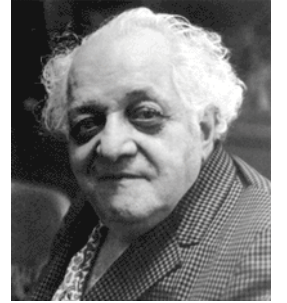


Historical roots and precedents



Behavioural sciences and sociology

- Moreno/Jenning 1930's sociometry
- Classical reference on history of SNA

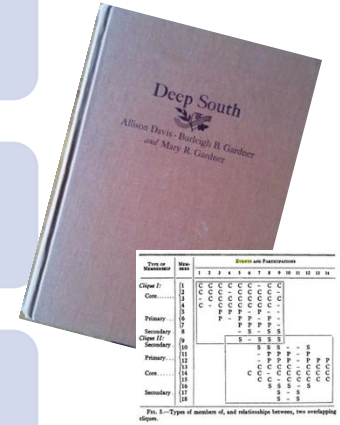


Earlier precedents (1920s)

- John Almack's (1922), Beth Wellman (1926): homophilic choices in schools

Anthropology

- The bipartite Deep South networks of Davis, Gardner, Gardner (1941)



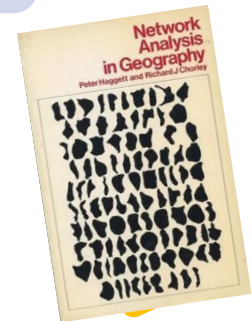
Human geography

- Kansky (1963)
- Haggett/Chorley (1969)

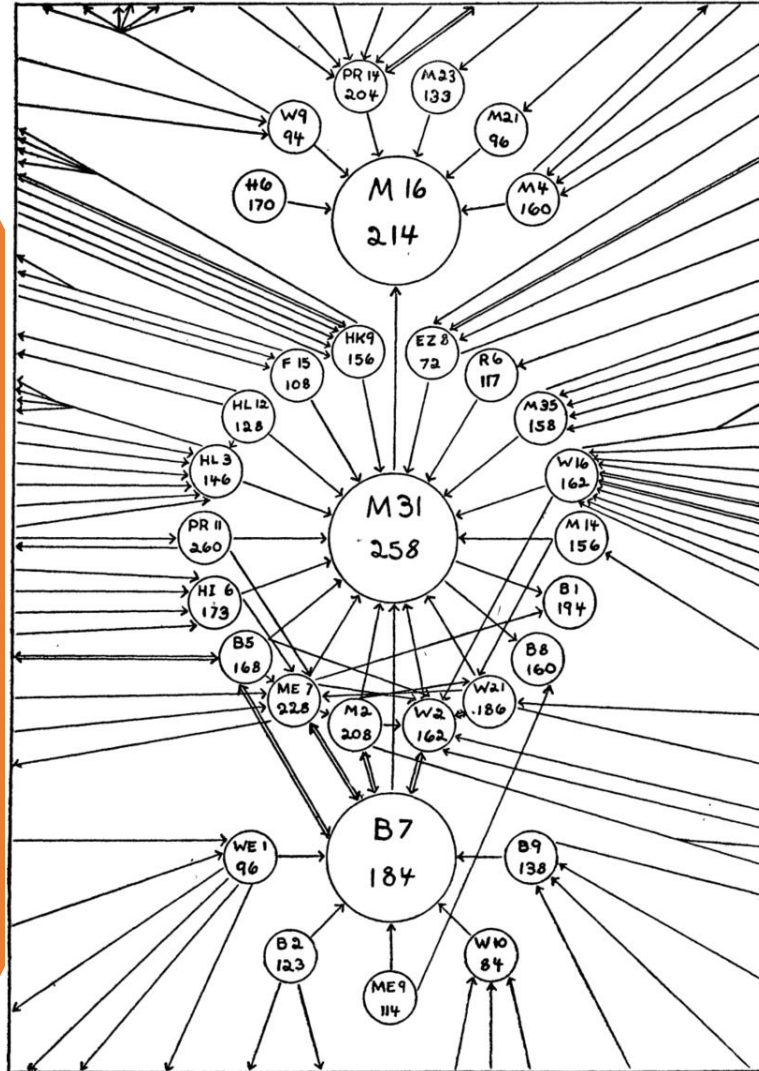
Relational sociology of 1970's

- (Harrison) White, Burt, Breiger, Wellman, Arabie etc.

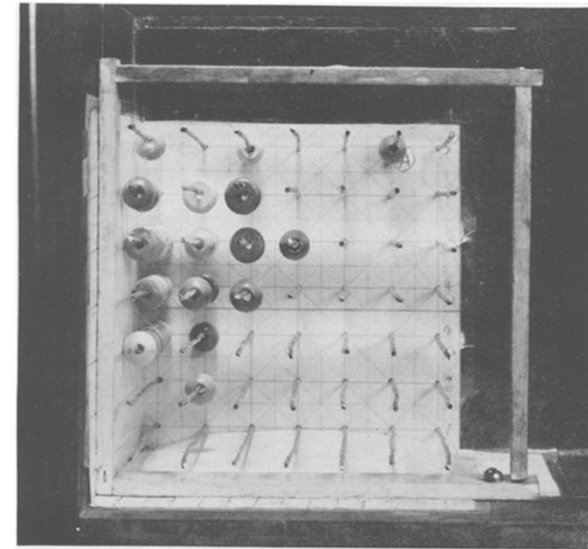
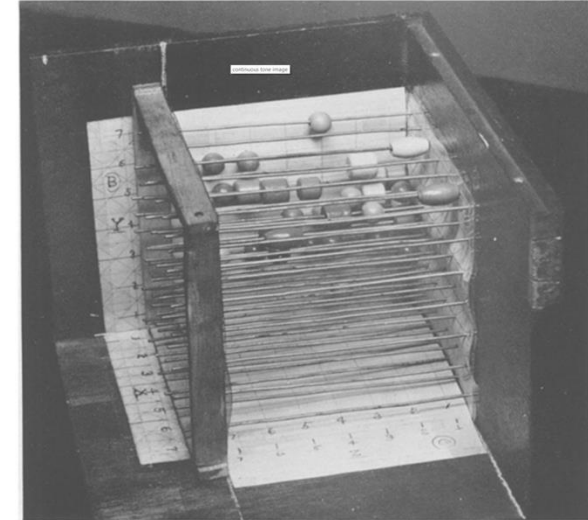
Freeman, L. (2000). Visualizing Social Networks, Journal of Social Structure, 1(1)
Freeman, L. (2004). The Development of Social Network Analysis. Empirical Press.



Historical roots and precedents



Lundberg & Steele (1938), Social Attraction-Patterns in a Village, *Sociometry*, 1 (3/4), p. 387



Chapin (1950), Sociometric Stars as Isolates, *American Journal of Sociology*, 56(3), p. 265ff

Network terminology

Actor / Node / Vertex

Entities of interest

Often individuals, but also organizations, groups, cities/countries, words etc.

Often have properties/attributes (e.g. age, income, gender etc)

Relations / Edges / Links / Dyads

Connecting pairs of actors/nodes

Each set of relations represent different contextual relational types (e.g. friendship, kinship, working relation, road connection, amount of bilateral trade)

Network

Collection of (at least) one set of nodes and (at least) one set of edges connecting pairs of nodes



Types of relations



Directional vs. Symmetric

- **Directional**
Likes, dislikes, trade flow, calls, supervises, talks to, friendship nomination
- **Symmetric**
Dialogue, kinship, friendship (reciprocal), road connection, co-presence, co-affiliation

Binary vs. Valued

- **Binary**
Likes, dialogue, alliance, cited in, kinship
- **Valued / weighted**
Number of phone calls, degree of friendship (Likert-scaled), number of citations, value of trade flows

Also signed

Types and Basic transformations

	Binary	←	Valued
Symmetrical	John and Sue are friends USA and Canada has a treaty NGO A and B share resources		John and Sue met 4 times Dist Malmö-Lund is 10 km Politicians A and B has 5 collabs
↑			
Directional	Sven dislikes John Denmark invaded Sweden John likes Sven		John gave Sue 8 appples Denmark invaded Sweden twice Five politicians left party A to B

...and various kinds of normalizations, null models (CM, rewiring, backboning etc), dichotomization, pruning etc

Data formats

- Keep tabs on directionality, edge values etc.
- Supplement with actor attributes

Edgelist

from	to	value
1	9	1
2	6	1
2	7	1
2	9	1
3	5	1
3	8	1
4	7	1
4	11	1
4	15	1
5	3	1

For large (sparse) networks
Unclear directionality
Optional value columns

(Socio)matrix

	c1	c2	c3	c4	p11	p12	p21	p22	p31	p32	p33	p41	p42
c1		1	0	1	1	1	0	0	0	0	0	0	0
c2	1		1	0	0	0	1	1	0	0	0	0	0
c3	0	1		1	0	0	0	0	1	1	1	0	0
c4	1	0	1		0	0	0	0	0	0	0	1	1
p11	1	0	0	0		0	0	0	0	0	0	0	0
p12	1	0	0	0	0		0	0	0	0	0	0	0
p21	0	1	0	0	0	0		0	0	0	0	0	0
p22	0	1	0	0	0	0	0		0	0	0	0	0
p31	0	0	1	0	0	0	0	0		0	0	0	0
p32	0	0	1	0	0	0	0	0	0		0	0	0
p33	0	0	1	0	0	0	0	0	0	0		0	0
p41	0	0	0	1	0	0	0	0	0	0	0		0
p42	0	0	0	1	0	0	0	0	0	0	0	0	

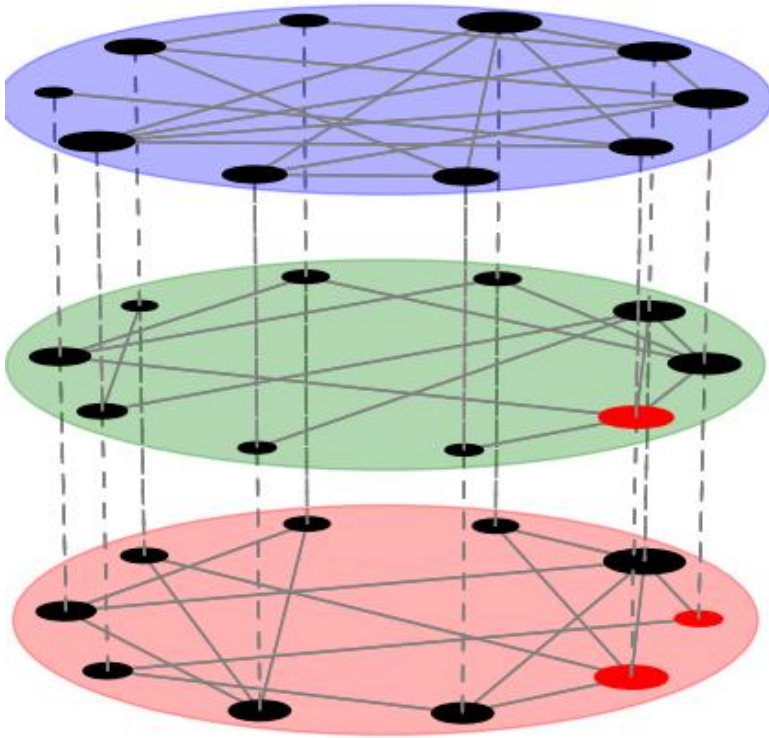
For (relatively) small, dense networks
All relational types (but takes space)
Useful for blockmodeling methods

Nodelist (1)

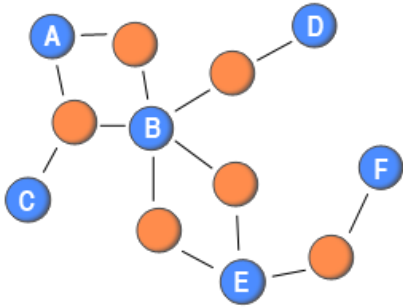
c1	c2	c4	p11	p12	
c2	c1	c3	p21	p22	
c3	c2	c4	p31	p32	p33
c4	c1	c3	p41	p42	
p11	c1				
p12	c1				
p21	c2				
p22	c2				
p31	c3				
p32	c3				
p33	c3				
p41	c4				
p42	c4				

For large binary networks
Unclear directionality

Multilayer networks



- Same set of nodes
 - Individuals, countries, organizations etc
- Multiple/different sets of relations
 - Different types of relations among individuals (e.g. Breiger, Boorman, Arabie 1974)
 - Bilateral trade between countries, different commodities (e.g. Snyder & Kick 1979)
- Multiplex, Multi-relational, "tofts" (types-of-ties)
 - Attempt at generalization/formalization (Kivelä et al 2014)



(from Opsahl 2011)

Bipartite networks

A.k.a. 2-mode networks, affiliation networks

- Also N-mode and N-partite networks

Two sets of nodes / actors

- Society ladies and social events (Davis et al 1941)
- Interlocking directorates (directors on corporate boards)
- Authors and scientific papers
- Company branches and cities (Taylor et al; GaWC papers)
- Hashtags and tweets

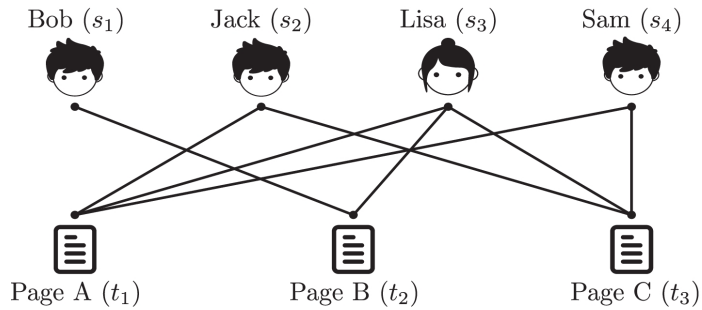
Relations only between these two sets

- No relations within each set
- Possibility for many-to-many relations

Ties typically binary and symmetric

Relations or a set of attributes (indicator variables)?

- A paper has a set of authors
- A social event is visited by a set of women
- A woman attends a set of social events

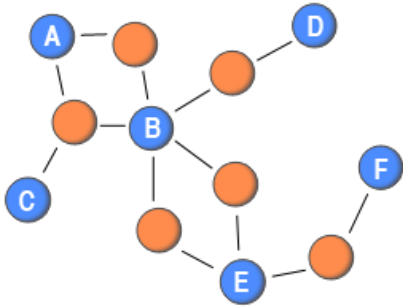


(from Zhang et al 2023)

TYPE OF MEMBERSHIP	MEMBERS	EVENTS AND PARTICIPATIONS													
		1	2	3	4	5	6	7	8	9	10	11	12	13	14
Clique I:	Core.....	1	C	C	C	C	C	C	-	C	C				
		2	C	C	C	-	C	C	C	C	C				
		3	-	C	C	C	C	C	C	C	C				
		4	C	-	C	C	C	C	C	C	C				
		5	-	C	C	C	C	C	C	C	C				
Primary...		6	P	-	P	P	-	P	-						
		7	P	-	P	P	P	P	-						
		8	-	S	-	S	S	S							
Clique II:	Secondary...	9			S	-	S	S	S						
		10			S	S	S	S	-	-	S				
		11			-	P	P	P	-	P					
	Primary...	12			-	P	P	P	-	P	P	P			
		13			C	C	C	C	-	C	C	C	C		
	Core.....	14			C	C	-	C	C	C	C	C	C	C	
		15			C	C	-	C	C	C	C	C	C	C	
		16					S	S	S	-	S				
	Secondary...	17					S	-	S						
		18					S	-	S						

FIG. 5.—Types of members of, and relationships between, two overlapping cliques.

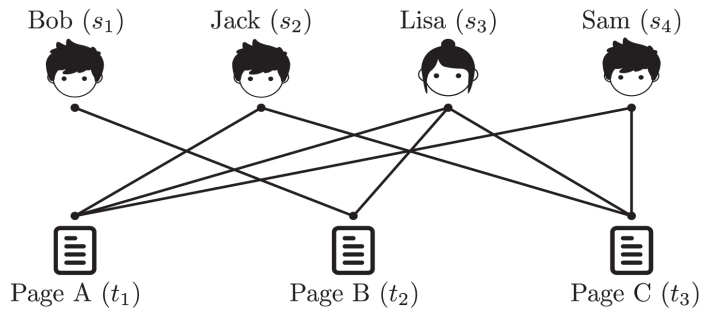
(from Davis et al 1941)



(from Opsahl 2011)

Bipartite networks: methods and projections

- Fewer methods for analyzing 2-mode networks
 - (though Borgatti & Everett 1997; Doreian et al 2004)
- Often projected: converted into 1-mode (unipartite) network
- Different projection method: classical approach by matrix multiplication of transpose

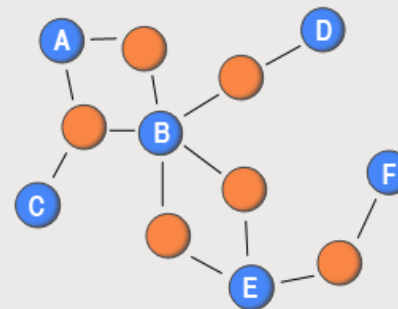


(from Zhang et al 2023)

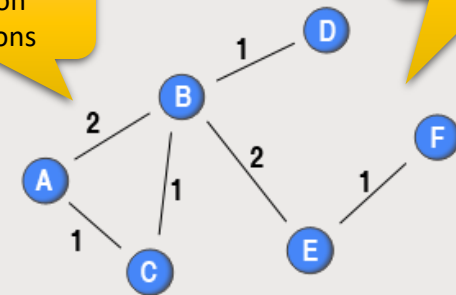
TYPE OF MEMBERSHIP	MEMBERS	EVENTS AND PARTICIPATIONS													
		1	2	3	4	5	6	7	8	9	10	11	12	13	14
Clique I:	1	C	C	C	C	C	C	-	C	C					
Core.....	2	C	C	C	-	C	C	C	C	C					
	3	-	C	C	C	C	C	C	C	C					
	4	C	-	C	C	C	C	C	C	C					
Primary...	5		P	P	P	P	P	-							
	6		P	-	P	P	P	P	-						
	7			P	P	P	P	P	-						
Secondary...	8		-	S	-	S	S	S							
Clique II:	9		S	-	S	S	S	S							
Secondary...	10		S	S	S	S	-	-	S						
	11		-	P	P	P	P	-	P						
Primary...	12		-	P	P	P	P	-	P	P	P				
	13		C	C	C	C	C	-	C	C	C	C			
Core.....	14		C	C	-	C	C	C	C	C	C	C	C		
	15		C	C	-	C	C	C	C	C	C	C	C	C	
	16			S	S	S	-	S							
Secondary...	17		S	-	S										
	18		S	-	S										

FIG. 5.—Types of members of, and relationships between, two overlapping cliques.

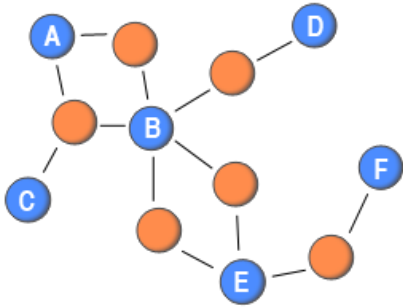
(from Davis et al 1941)



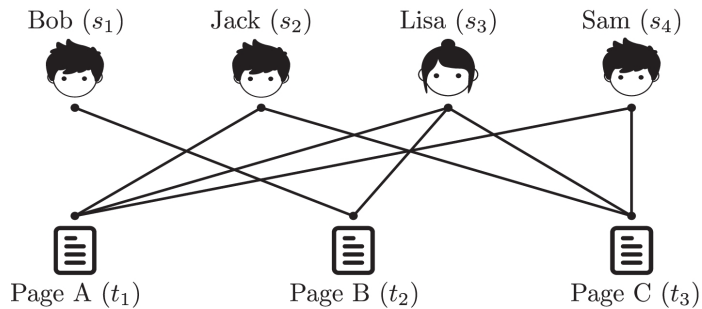
A and B have 2 common affiliations



D and F have no shared affiliations

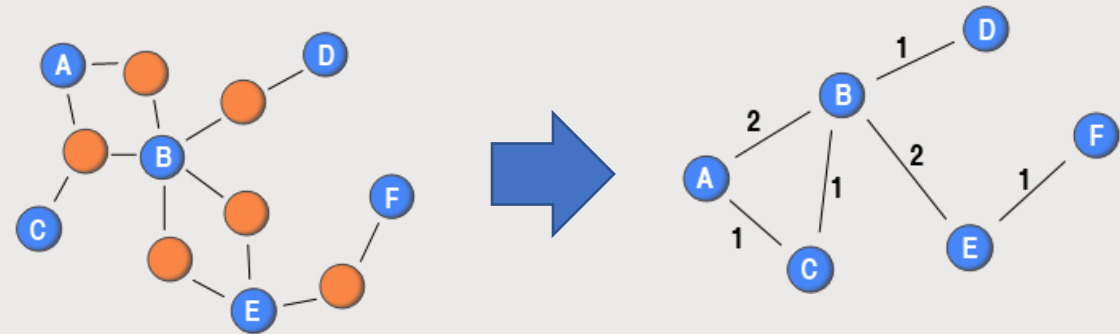


(from Opsahl 2011)



(from Zhang et al 2023)

Bipartite networks: methods and projections



- Projections must make theoretical sense!
 - Peter networks (slightly different projection) does not
- Ladies at social events (Davis et al 1941)
 - Assumption of social exposure at events: feasible
 - Projected network: valued symmetrical – number of shared affiliations
 - We are not observing that individuals interact, but we can identify which individuals that are (and are not) exposed to each other
 - Could be dichotomized as deemed fit (minimum number of shared events/exposure to infer a social tie/exposure)

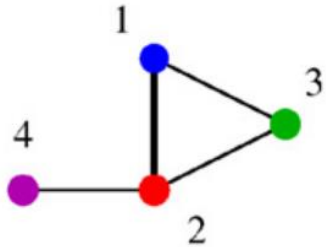
TYPE OF MEMBERSHIP	MEMBERS	EVENTS AND PARTICIPATIONS													
		1	2	3	4	5	6	7	8	9	10	11	12	13	14
Clique I:	1	C	C	C	C	C	C	-	C	C					
	2	C	C	C	-	C	C	C	C	C					
	3	-	C	C	C	C	C	C	C	C					
	4	C	-	C	C	C	C	C	C	C					
	5	-	C	C	P	P	-	P	-	-					
Primary...	6	P	-	P	-	P	-	P	-	-					
	7	P	-	P	P	P	P	-	-	-					
	8	-	S	-	S	S	S	-	-	-					
Secondary...	9		S	-	S	S	S	-	-	-					
Clique II:	10									S					
	11									P	P	-	P		
	12									-	P	P	-	P	P
	13									C	C	C	-	C	C
	14									C	C	-	C	C	C
Core.....	15									C	C	-	C	C	C
	16									S	S	-	S		
	17									S	-	S	-	S	
	18									S	-	S	-	S	
	19									S	-	S	-	S	

FIG. 5.—Types of members of, and relationships between, two overlapping cliques.

(from Davis et al 1941)

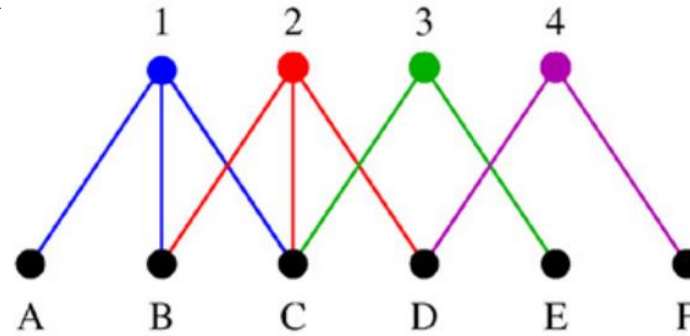
• Edge inflation phenomena
• Null models recommended (simulation)

	1	2	3	4
1	(3)	2	1	0
2	2	(3)	1	1
3	1	1	(2)	0
4	0	1	0	(2)



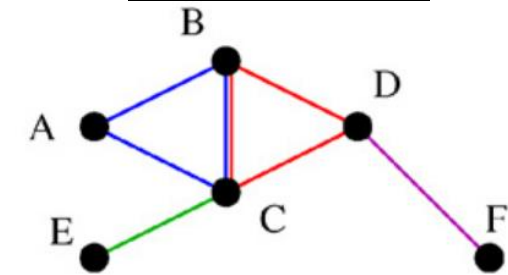
$$P_2 = A^T \cdot A$$

	1	2	3	4
A	1	0	0	0
B	1	1	0	0
C	1	1	1	0
D	0	1	0	1
E	0	0	1	0
F	0	0	0	1



$$P_1 = A \cdot A^T$$

	A	B	C	D	E	F
A	(1)	1	1	0	0	0
B	1	(2)	2	1	0	0
C	1	2	(3)	1	1	0
D	0	1	1	(2)	0	1
E	0	0	1	0	(1)	0
F	0	0	0	1	0	(1)

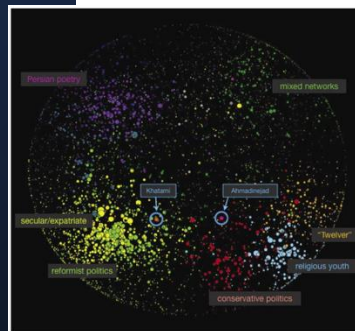


(from Latapy et al 2008)

Bipartite networks: methods and projections

- Bipartite projections have directionality!
 - How actors A-F affiliates with 1-4: network of relations between A-F
 - How affiliations 1-4 share actors: a network of relations between 1-4
- Southern ladies-event data:
 - Network between ladies, capturing number of shared events
 - Network between events, capturing number of shared ladies

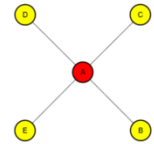
Getting network data



(Kelly & Etling 2008)

• Sociometric surveys

- Collect ego networks of each individual (name generators)
- Merge all ego networks into larger network



• Snowball sampling

- Initial study population (zone 1)
- New actors mentioned (next zone)
- Query new actors
- Prune outer layer (no info about their ties)

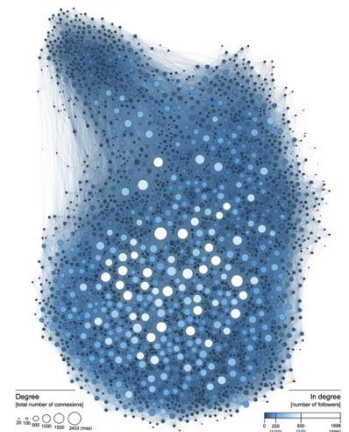
	A	C	D	F	G	B	E	H	
A		1	1	0	1	0	0	0	✓
C	1		0	0	1	0	1	1	✓
D	1	0		1	1	1	0	0	✓
F	0	1	1		0	0	0	1	✓
G	1	1	0	1		0	1	0	✓
B							?	?	
E									
H							?	?	

• Importance of well-defined system boundaries!

- Interactions within individual school classes feasible
- Assuming inter-class relations can be ignored
- Full-population typically needed

• Databases and digital trace data

- Email log files
- Social media ties (1- and 2-mode)
- Employment roster
- Transportation networks



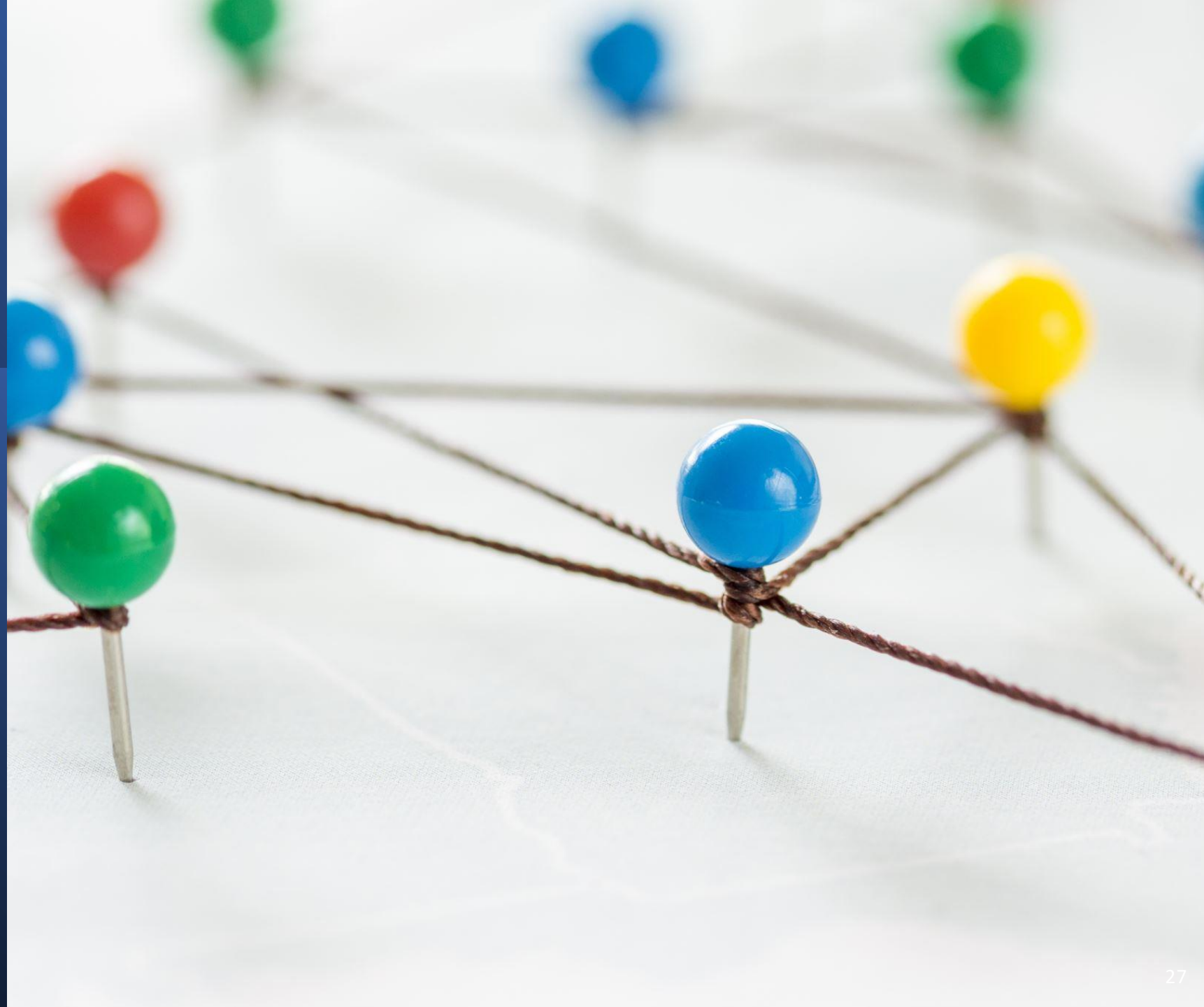
(Grandjean 2015)

A photograph of a breakfast table. On the left, a white coffee cup sits on a saucer. Next to it is a folded newspaper. On the right, a white plate holds two golden-brown pastries. The background is blurred, showing a vase of flowers and a window. The word "Break!" is written in white text across the center of the image.

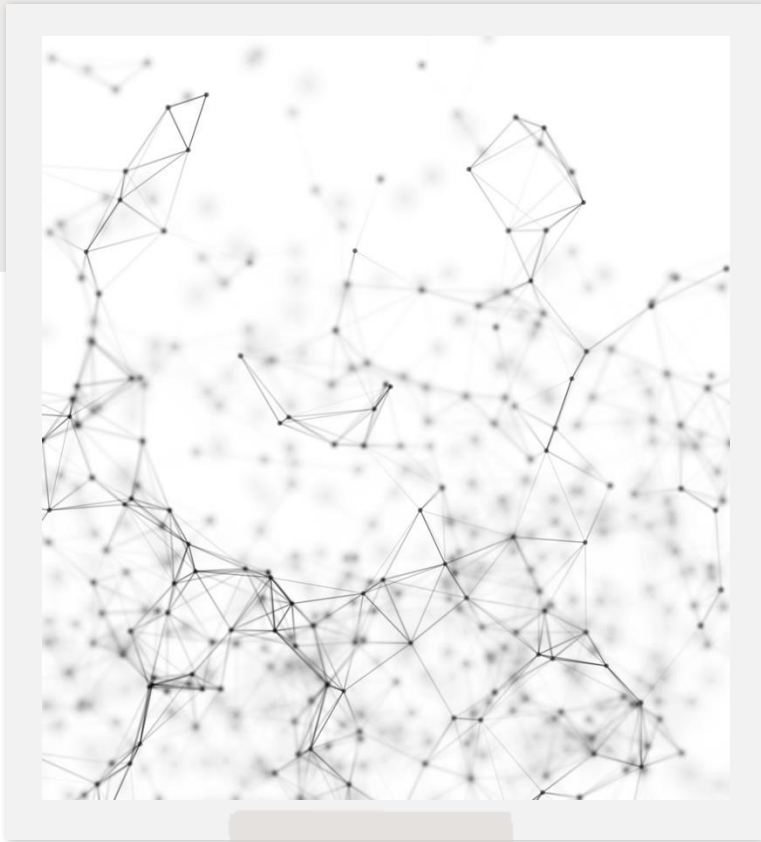
Break!

Network metrics and methods

(Descriptive only)

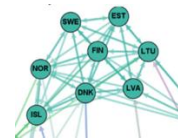


Network metrics and methods



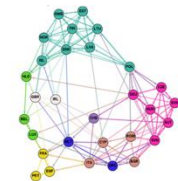
Micro-level

- Structural properties of individual actors



Meso-level

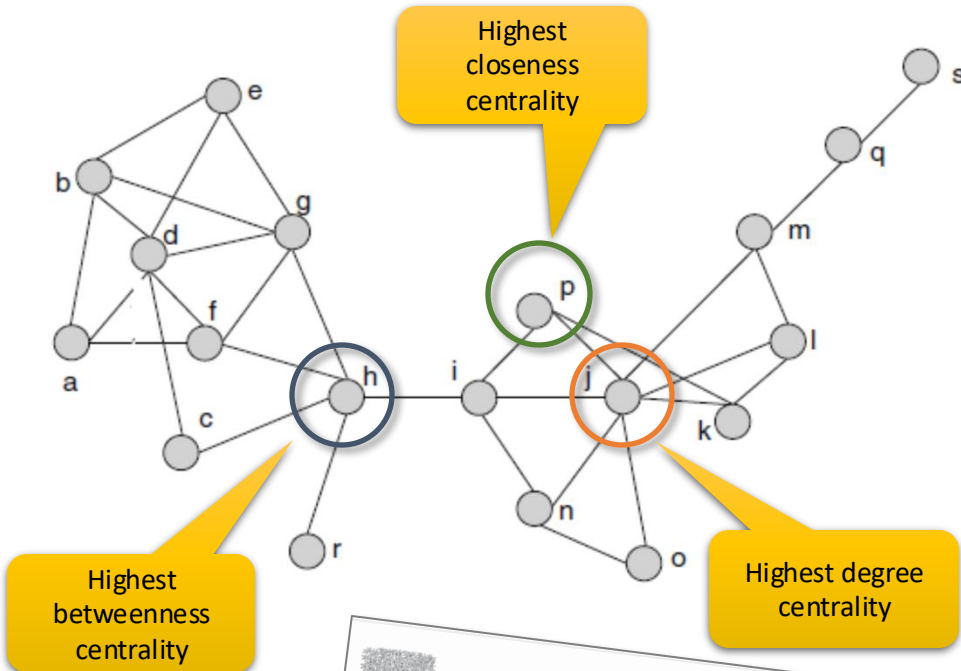
- Properties and similarities of actor subsets



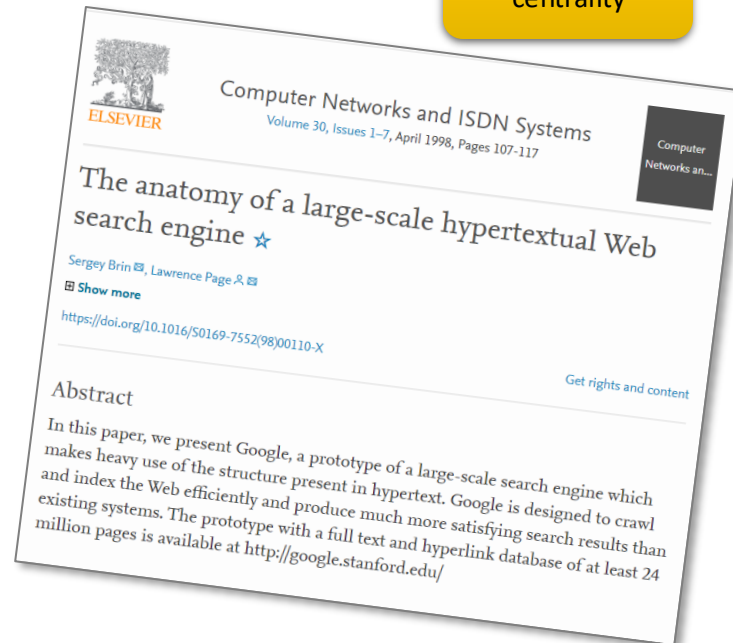
Macro-level

- Properties of the network as a whole

Centrality



- **Katz centrality**
Extension of degree centrality (taking non-nearest neighbors into account)
- **PageRank**
Extension of Katz centrality (adjusting an inflationary aspect of Katz)



Degree centrality

- Number of ties that an actor is connected to

Betweenness centrality

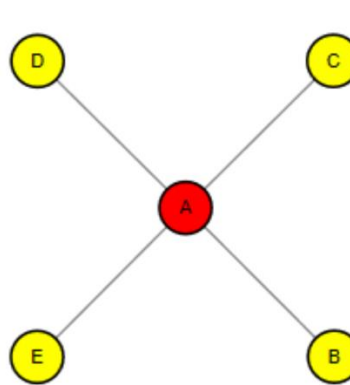
- How "between" an actor is
- Number of shortest paths between all pairs of actors that passes through an actor

Closeness centrality

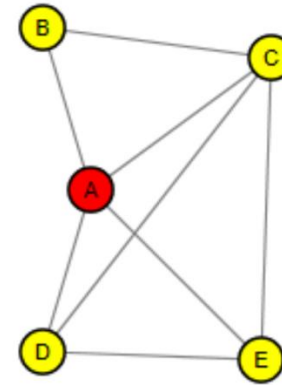
- How "close" an actor is to all others
- Sum of shortest path from an actor to all other actors

Clustering coefficient

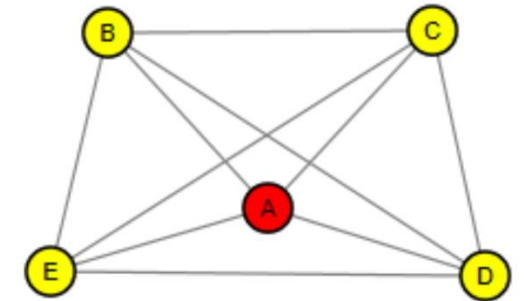
- How connected are my alters?
- Density of alter-alter ties in an ego-network



0.00



0.67



1.00

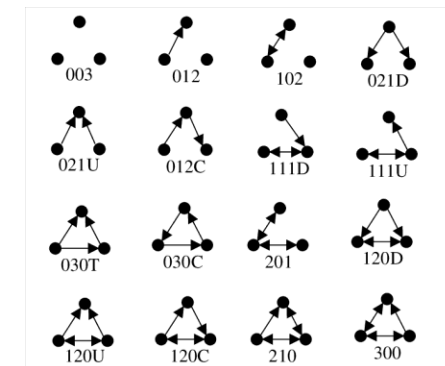
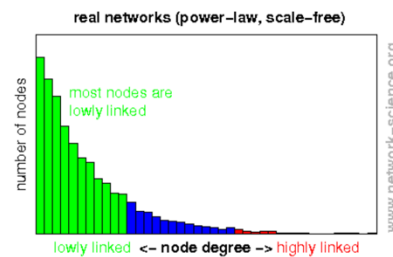
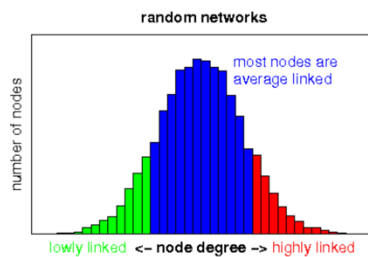
Cohesive Subgroups (communities)

- Group actors in a network based on these having relatively more ties with each other than with other non-group actors
- Multiple different heuristics
 - Girvan-Newman
 - Clique overlap
 - Cores / k-Cores / k-Plex / Factions etc (see, e.g., Wasserman & Faust 1994)
- Girvan-Newman algorithm (2002)
 - Identify highest 'edge-betweenness'
 - Remove that edge
 - Calculate modularity (goodness-of-fit measure of community structures)
 - Repeat (recalculate edge betweenness)



Macro-level metrics

- Properties of the network as a whole
 - Nbr of (disconnected) components
 - Density (existing/possible ties)
 - Diameter (longest shortest-path)
 - Reciprocity
- Centralization
 - Average and distribution of nodal centralities
- Triad census



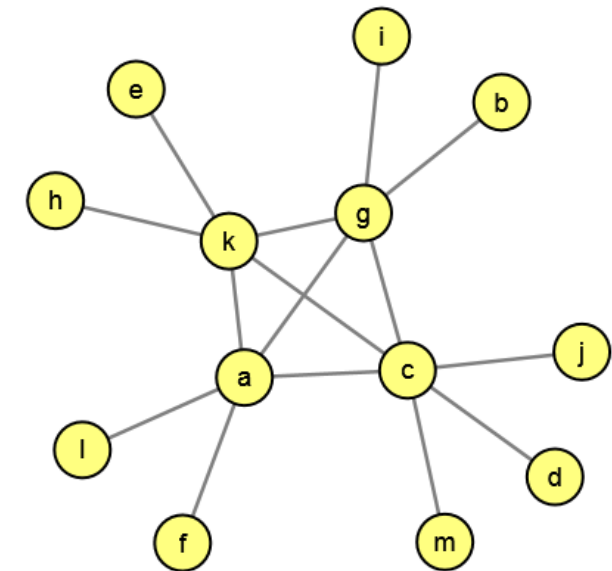
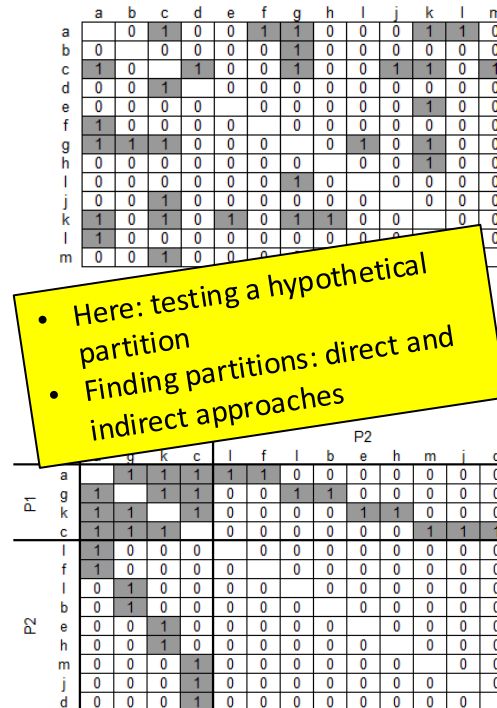
Role-analysis (blockmodeling)

- Given a rich, complex network of (social) relations:
- Identifying roles in networks
 - Partition actors based on relational similarities
 - Formal way to capture notion of "social role" (Harrison White 1970s)
- Hospital example: patients, doctors, nurses, admin
 - Doctors interact with other doctors, and patients
 - Nurses interact with doctors, patients and admin
 - Admin interact with patients
 - Patients interact with all, except other patients
- More complex relational patterns
 - Beyond community detection
 - Core-periphery, transitive, hierarchy
 - "Custom" roles and structures



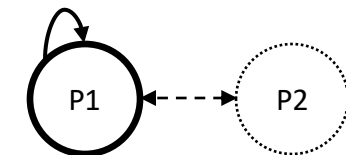
Role-analysis (blockmodeling)

- Classical core-periphery structure
 - Seemingly two types of actors
- Core actors
 - Connected with several other core actors
 - Connected with a subset of peripheral actors
- Peripheral actors
 - NOT connected with other peripheral actors
 - Connected to singular core actors
- Re-shuffling original sociomatrix
 - According to our hypothetical separation
- Identifying blocks
 - Representing relational patterns within and between equivalent actors



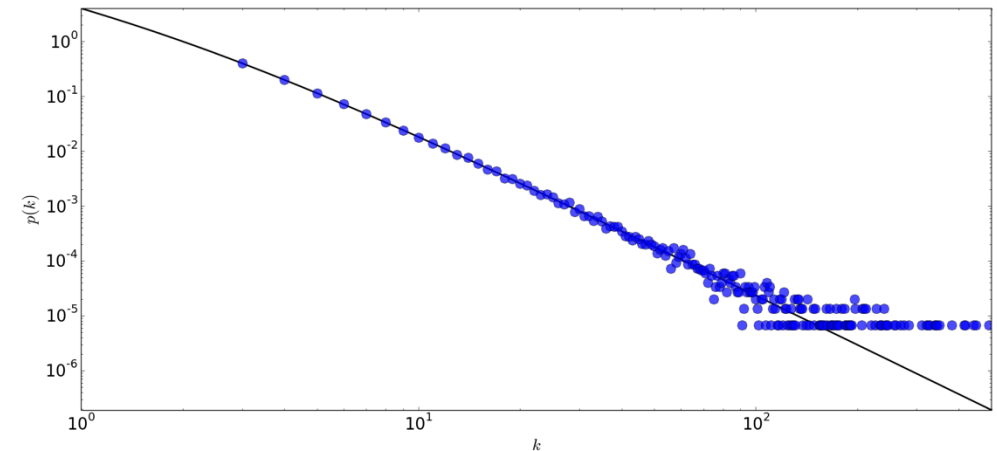
(Galtung 1971)

	P1	P2
P1	com	reg
P2	reg	nul



Ideas about network dynamics

- Assortativity/homophily
 - Actors with similar attributes more likely to form connections (e.g. McPherson 2001)
 - Typically yields community structures
- Preferential attachment
 - High-degree actors more likely to connect with new actors
 - Associated with degree distribution following a power law
- Grand Unifying Network Theory
 - Certain dynamics and mechanisms that exist in all (or most) types of networks (cf. Brandes et al 2013)



Social Network Analysis and CSS

- Plethora of ways to describe/define what CSS is and how it relates to
 - Quantitative social-science at large
 - Data science
- This conceptual pluralism likely good for now
 - Emerging field
- Possible distinction between CSS and Data Science concerning the triad between
 - Research question (including theory)
 - Data
 - Methods



Social Network Analysis and CSS

- In CSS, we tend to start off with the RQ cog
- Building on previous research, we formulate RQs of social-scientific and/or societal relevance
 - Is GDP growth linked to positionality in trade networks?
 - Are private schools better than public?
 - Are inter-ethnic families happier than endogamic ones?
- To address these: pair up methods with data
- Circling back to RQ at the end



Social Network Analysis and CSS

- Postulating that data science is more exploratory
- Given (typically large and complex) datasets
 - More exploratory, open-minded approach
 - Apply methods to prune out insights data
- Although general topic area given by the data
 - Perhaps more open to *finding* signals
 - Seed to an interesting research question
- Formulating question (or topic)
- Connecting with previous research



Social Network Analysis and CSS

- Ethnographic cycle (Gladwin 1989)
 1. Ask ethnographic questions
 2. Collect data
 3. Analyze data
 4. Discover better question (repeat from 1)
- In context of CSS (and SNA)
 1. Research question
 2. Database, scraping, wrangling
 3. Methods and analysis
 4. Discover better questions (repeat from 1)

