GENEVA GRADUATE INSTITUTE

INSTITUT DE HAUTES ÉTUDES INTERNATIONALES ET DU DÉVELOPPEMENT

GRADUATE INSTITUTE OF INTERNATIONAL AND DEVELOPMENT STUDIES

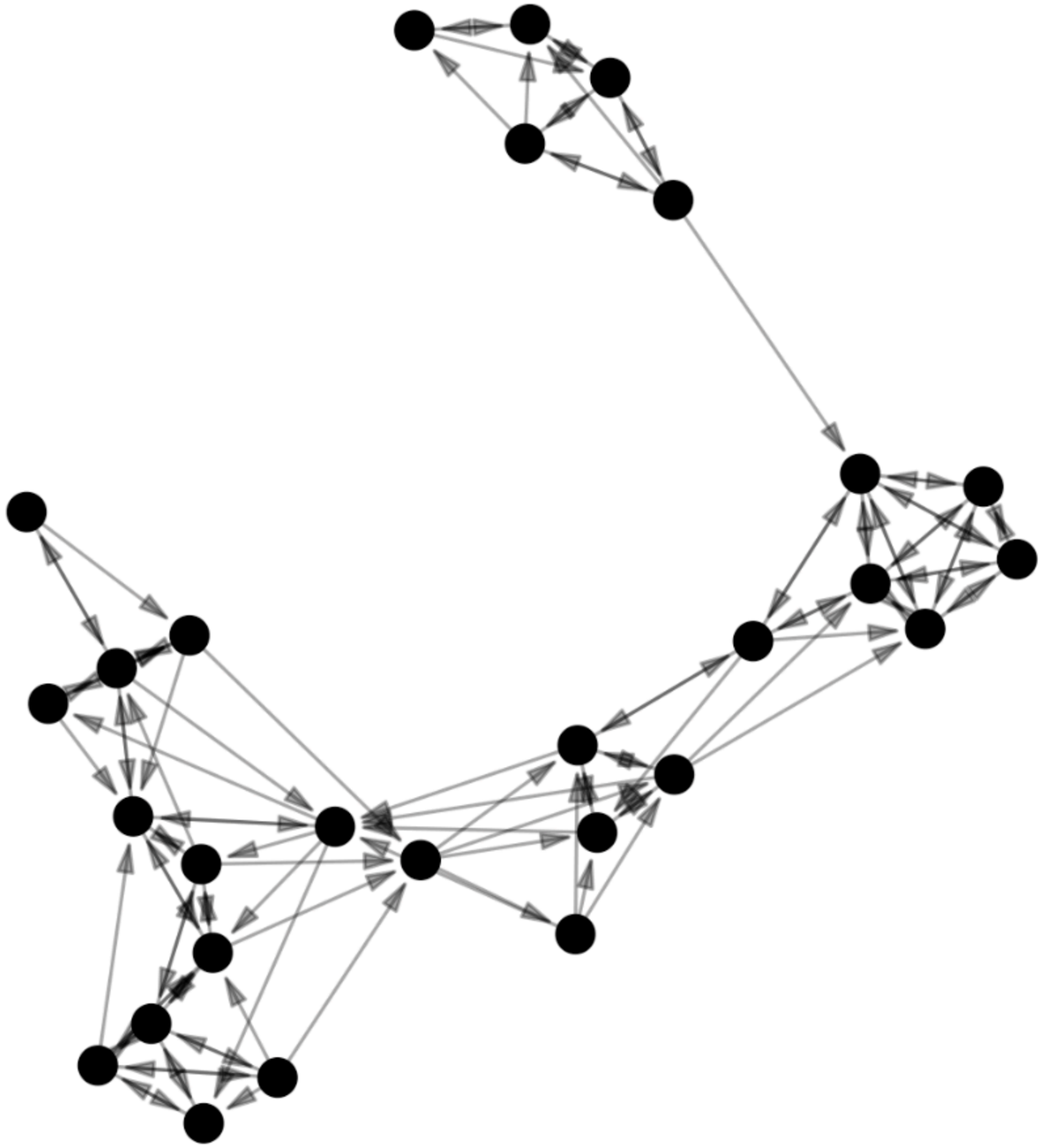
Relations

Social Networks Theories and Methods

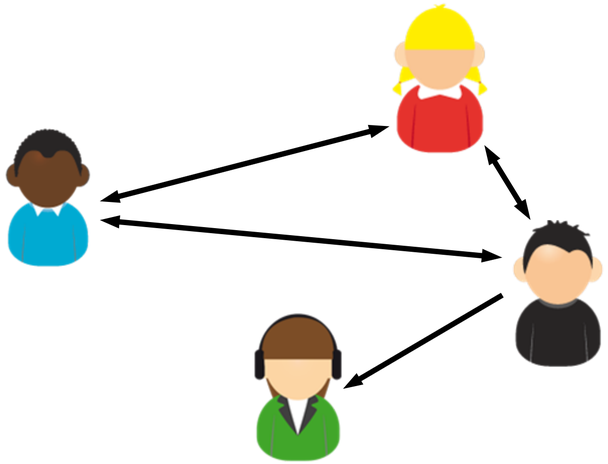
James Hollway

Memory check

- How would you describe this network?



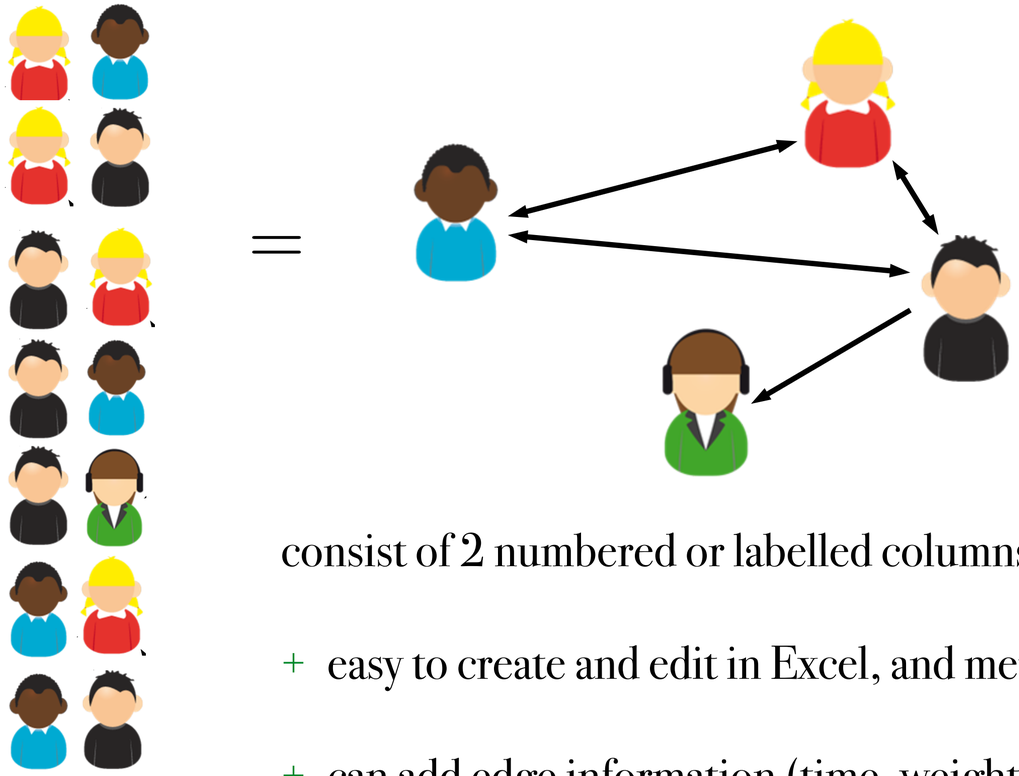
Graph



+ graphs are pretty and often quite natural to interpret

- graphs can quickly become complex and details difficult to discern

Edgelist <-> Graph



consist of 2 numbered or labelled columns (if directed then ordered)

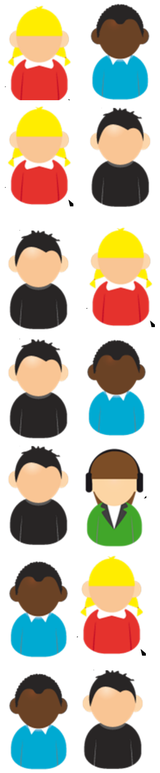
\+ easy to create and edit in Excel, and memory efficient

\+ can add edge information (time, weight, etc) as additional columns

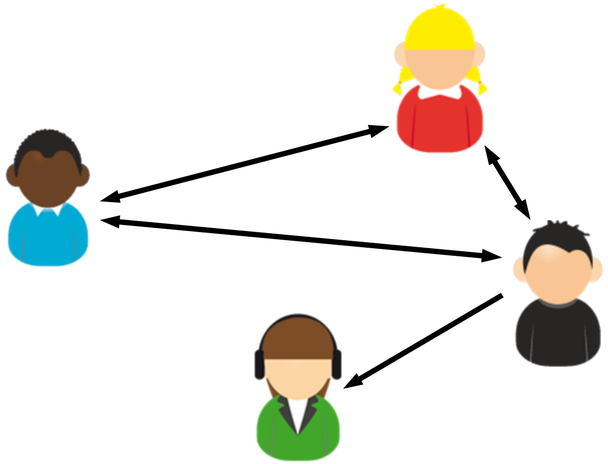
\- cannot deal with isolates/nodal attributes without additional data objects

\- more complicated statistics than degree difficult to calculate

Edgelist <-> Graph <-> Matrix



=



=

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

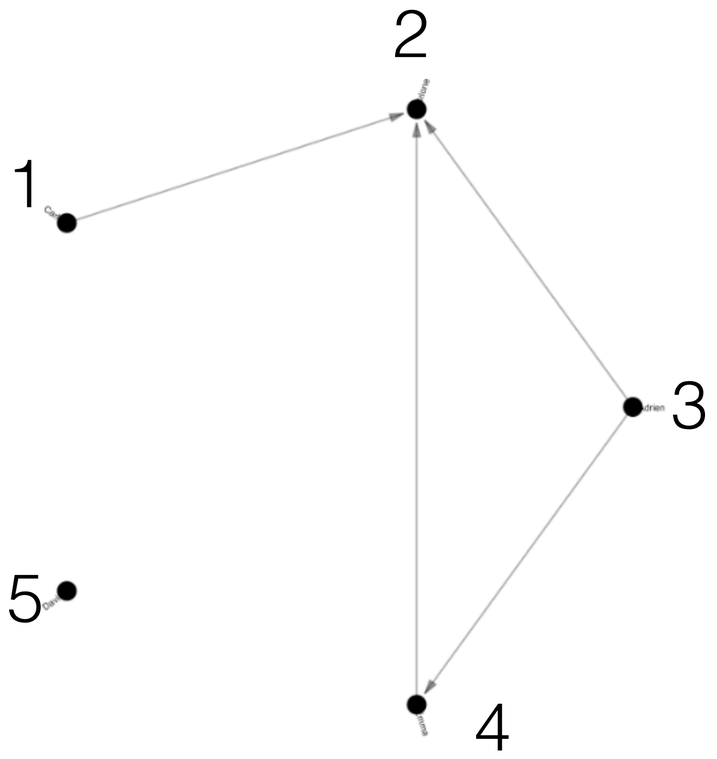
rows and columns correspond to (ordered) senders and recipients

1 indicates a tie from sender to recipient

0 indicates absence of a tie (diagonal 0s in simple network)

- memory inefficient for sparse networks
- somewhat incomprehensible
+ encodes all relational information (ties, weights, isolates, etc)
+ flexible, quick analysis

Let's try a few




```
##      [,1] [,2] [,3] [,4]
## [1,]    0    0    1    0
## [2,]    1    0    0    0
## [3,]    1    0    1    1
## [4,]    0    1    1    1
```

```
##   from to
## 1    1  2
## 2    1  4
## 3    2  3
## 4    2  1
## 5    3  4
## 6    4  1
```

Turn into matrix, edgelist, and graph respectively

Sources

Self-reports

- Surveys (e.g. Hogan et al 2016)
  - using e.g. Network Canvas
  - name generators vs rosters
- Interviews (e.g. Bellotti 2014)

Constructions

- Web-scraping
  - Chrome or Firefox extensions
  - Rcrawler, rvest and vosonSML packages
  - Python (e.g. to access Twitter API)
- Manual/automatic text coding
  - Considerable literature-based datasets
  - Can extract relationships (e.g. similarities) in or across documents

Observation

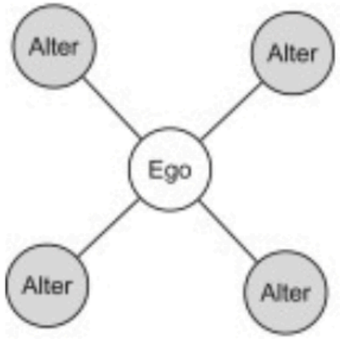
- Participant observation (e.g. Wyatt et al 2011)
- RFID badges (e.g. Elmer et al 2018)

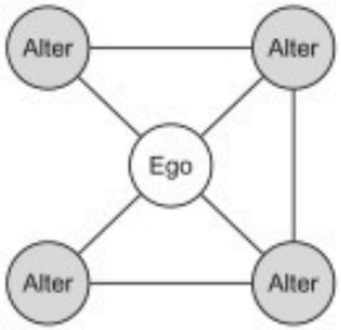
Archives

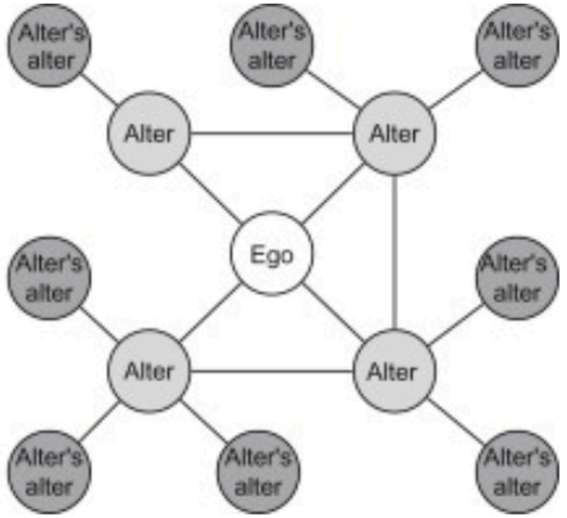
- Individual
  - Diary research (e.g. Fu 2008)
  - Historical records (e.g. Padgett & Ansell 1993)
  - Social media (e.g. Golder et al 2007)
  - Communication logs (e.g. Goldberg et al 2016)
- Organizational
  - Publication or patent records (e.g. Lazega et al 2008, Goetze 2010)
  - Agreement data between countries (e.g. Hollway and Koskinen 2016)

Ego networks

Usually name generator, name interpreter, and sometimes name interrelater.







- can analyse larger networks (sampling)
- compatible with traditional methods (and designs)
- can study intersecting social circles (focal nodes)

ONE-PERSON	NUCLEAR	NUCLEAR	NUCLEAR	NUCLEAR	COMPOSITE	NUCLEAR	NUCLEAR	NUCLEAR	COMPOSITE
One-person household	Married couple	Married couple, kid	Married couple, 2 kids	Parent, 1 kid	Householder, friend or partner	Married couple, 3 kids	Parent, 2 kids	Married couple, 4 kids	Parent, kid, friend or partner
									

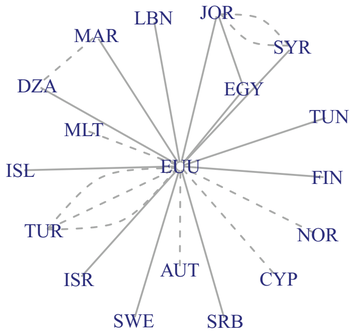
NUCLEAR	COMPOSITE	EXTENDED	COMPOSITE	EXTENDED	EXTENDED	EXTENDED	EXTENDED	EXTENDED	EXTENDED
Parent, 3 kids	Householder, non-relative	Householder, sibling	Parent, 2 kids, friend or partner	Householder, parent	Parent, kid, grandkid	Householder, grandkid	Married couple, grandkid	Married couple, kid, grandkid	Householder, relative
									

COMPOSITE	COMPOSITE	NUCLEAR	NUCLEAR	EXTENDED	COMPOSITE	COMPOSITE	EXTENDED	EXTENDED	EXTENDED
Group of 3 friends	Parent, 3 kids, friend or partner	Married couple, 5 kids	Parent, 4 kids	Married couple, grandparent	Householder, friend, non-rel.	Parent, kid, non-relative	Parent, kid, 2 grandkids	Married couple, relative	Married couple 2 kids, grandkid
									

EXTENDED	EXTENDED	EXTENDED	COMPOSITE	COMPOSITE	EXTENDED	COMPOSITE	EXTENDED	EXTENDED	COMPOSITE
Married couple, 1 kid, 2 grandkids	Married couple, 2 kids, relative	Householder, kid, parent	Non-related group of 3	Married couple, non-relative	Married couple, kid, parent	Group of 4 friends	Parent, 2 kids, grandkid	Parent, kid, grandparent	Married couple, kid, non-relative
									

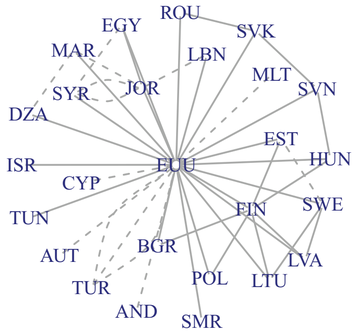
EXTENDED	EXTENDED	EXTENDED	EXTENDED	NUCLEAR	EXTENDED	EXTENDED	NUCLEAR	EXTENDED	COMPOSITE
Married couple, relative	Married couple, sibling	Married couple, 2 kids, parent-in-law	Parent, kid, sibling	Parent, kid, stepkid	Householder, 2 parents	Married couple kid, parent-in-law	Married couple, 6 kids	Married couple, 2 grandkids	Parent, 4 kids, friend or partner

EU 1989



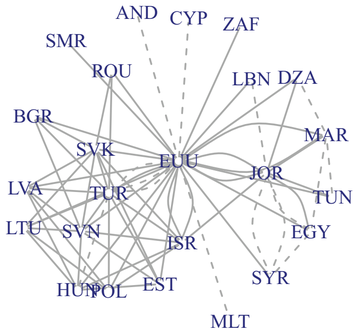
EU's Bilateral PTAs = 19
Transitivity = 0.06

EU 1994



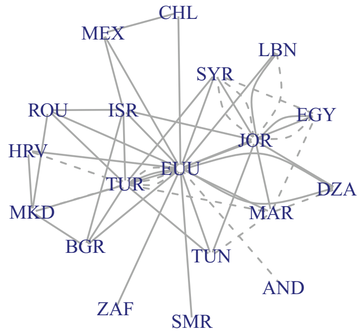
EU's Bilateral PTAs = 27
Transitivity = 0.15

EU 1999



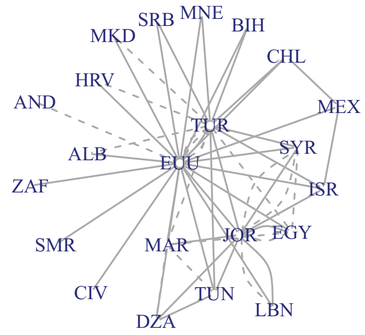
EU's Bilateral PTAs = 29
Transitivity = 0.41

EU 2004



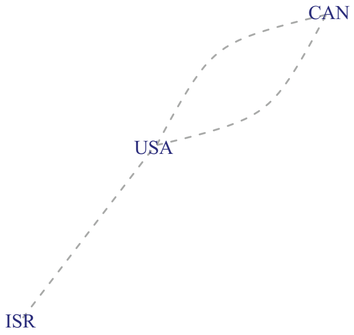
EU's Bilateral PTAs = 22
Transitivity = 0.36

EU 2009



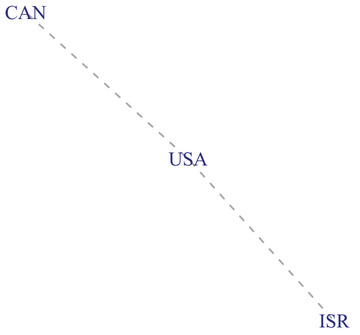
EU's Bilateral PTAs = 24
Transitivity = 0.31

USA 1989



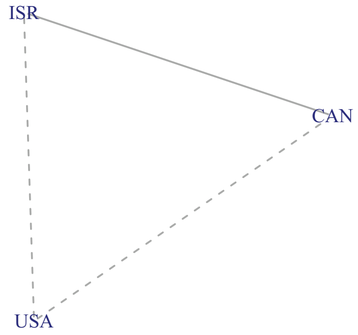
USA's Bilateral PTAs = 3
Transitivity = 0

USA 1994



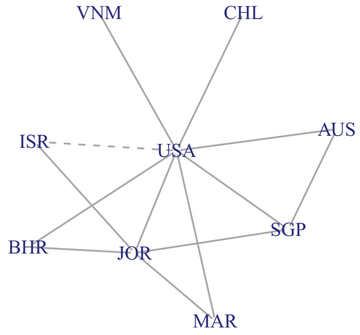
USA's Bilateral PTAs = 2
Transitivity = 0

USA 1999



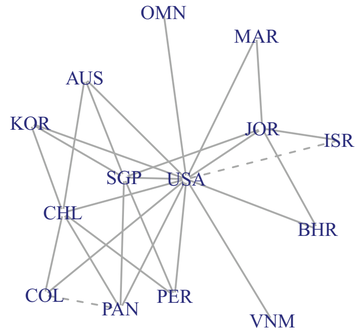
USA's Bilateral PTAs = 2
Transitivity = 1

USA 2004



USA's Bilateral PTAs = 8
Transitivity = 0.33

USA 2009

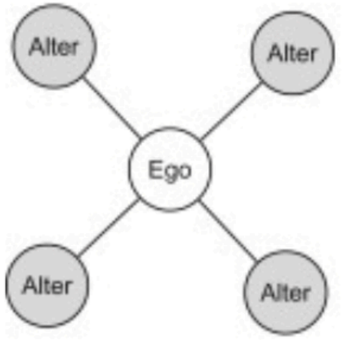


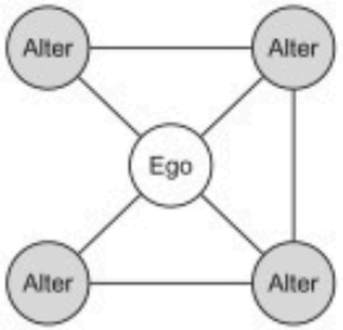
USA's Bilateral PTAs = 13
Transitivity = 0.32

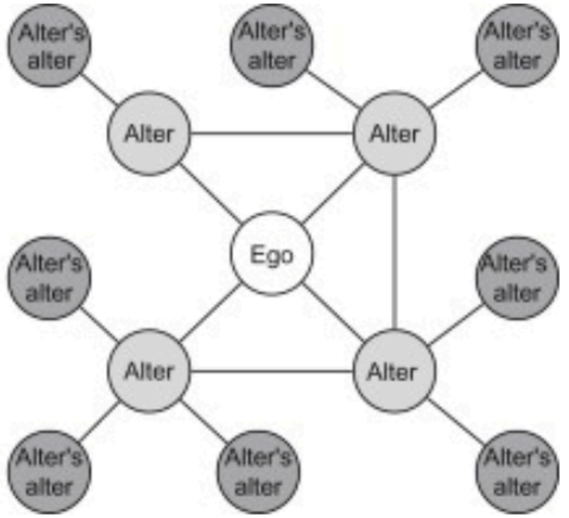
Milewicz et al. 2018

Ego networks

Usually name generator, name interpreter, and sometimes name interrelater.







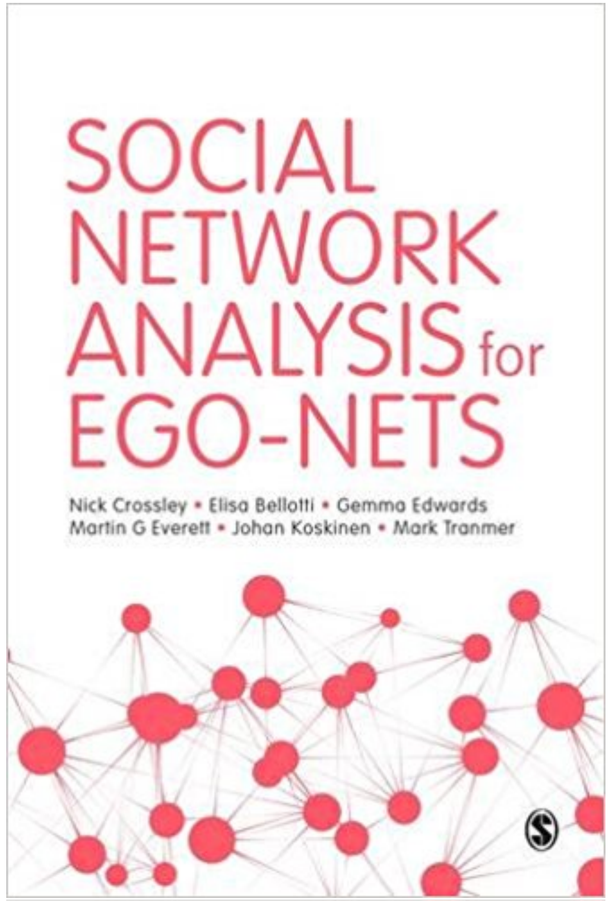
- can analyse larger networks (sampling)
- compatible with traditional methods (and designs)
- can study intersecting social circles (focal nodes)

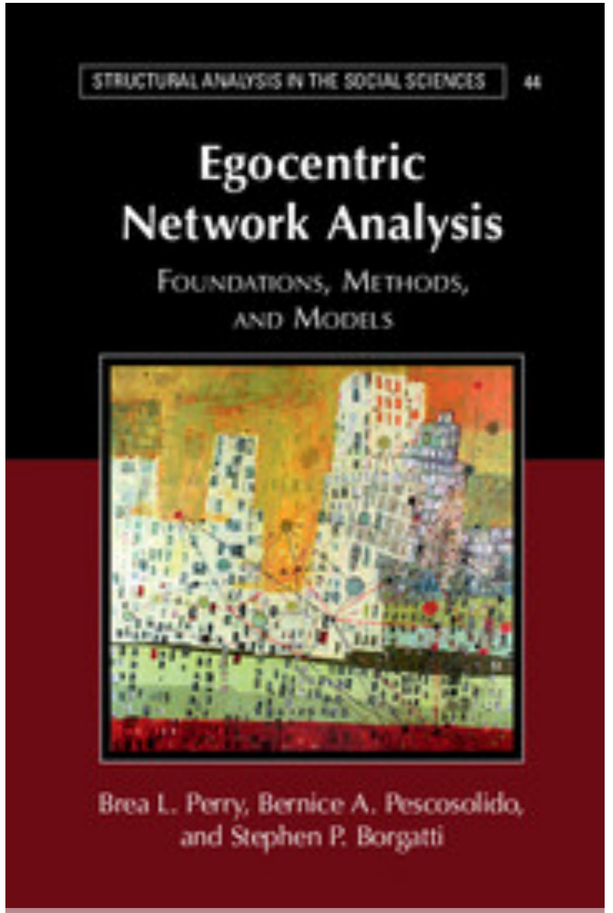
- demands traditional assumptions (exchangability)
- missing structural data (betweenness)
- recall often biased toward longer-term interactions (Freeman, Romney, & Freeman 1987)
- recall inaccurate when reporting perceptions of relationships between third parties (McEvily 2014)

Example: US General Social Survey

- Name generator: Asks respondents to report names of all people with whom they discussed important matters in last six months
- Name interpreter: Collects information on characteristics of first five people named (Burt 1984, Marsden 1987)
- Name generators a commonly used method, but some important methodological issues (e.g. adams and Moody 2007, Campbell and Lee 1991, Ferligoj and Hlebec 1999, Marsden 1993, 2003, Matzat and Snijders 2010)
- E.g. McPherson et al. 2008 found personal networks of Americans had shrunk significantly between 1985 and 2004, but it turned out that this was an artefact and actually due to:
  - Interviewer fatigue and other interviewer effects (e.g. uneven prompting, Bearman and Parigi 2004)
  - Respondents fatigue (e.g. when later in the survey, Paik and Sanchagrin 2013)
  - Respondents learning (e.g. panel conditioning, Warren and Halpern-Manners 2012)

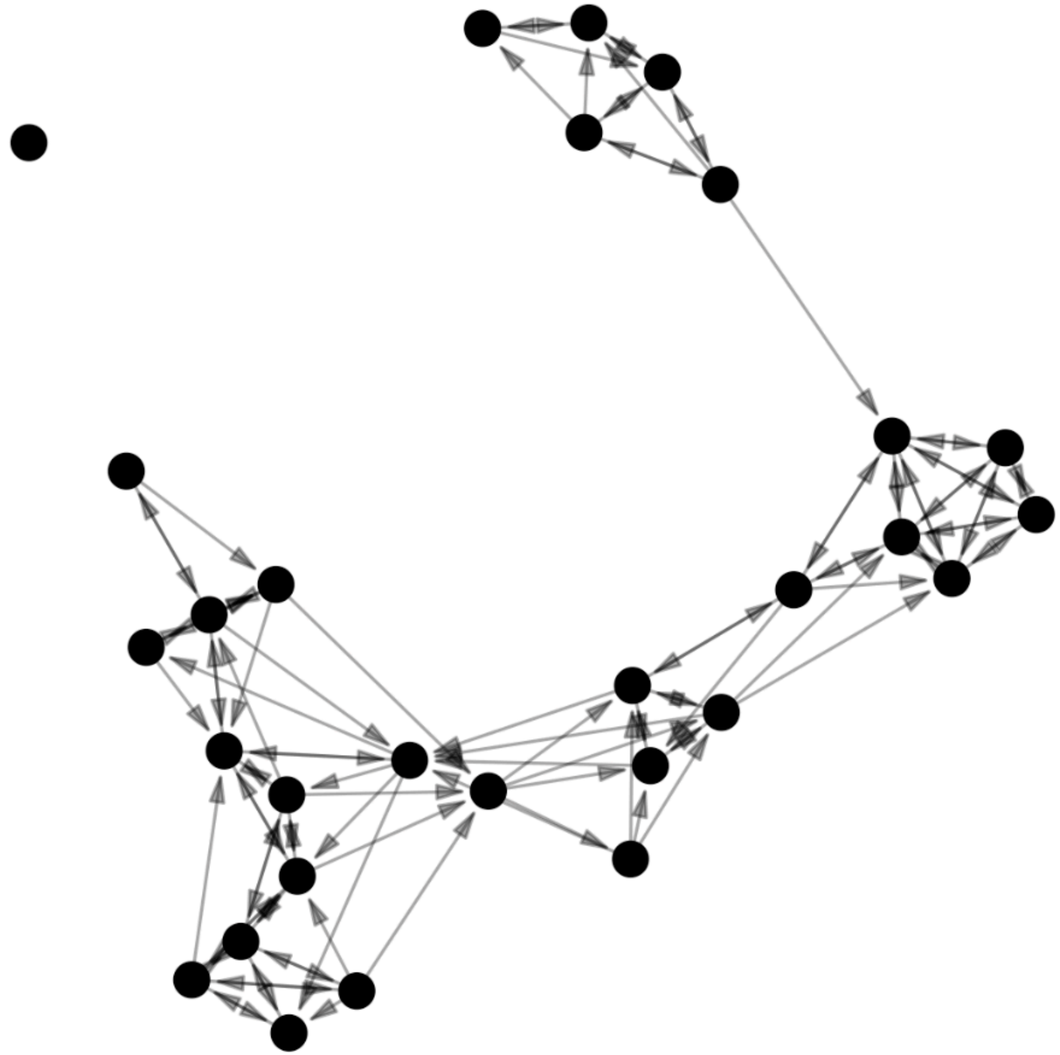
For more on ego networks, see:





Whole networks

- Respondents presented with a roster (complete list of individuals in a population of interest; Marsden 1990)
  - Reduces recall bias stemming from unreliably recalling interaction partners' names
  - Reduces perception bias cos reports cross-checked, or not even requested (though see Boda et al 2020)
  - Only possibly for small networks or networks where relational data publicly recorded
  - Large rosters can still become unreliable (Pustejovsky & Spillane 2009)



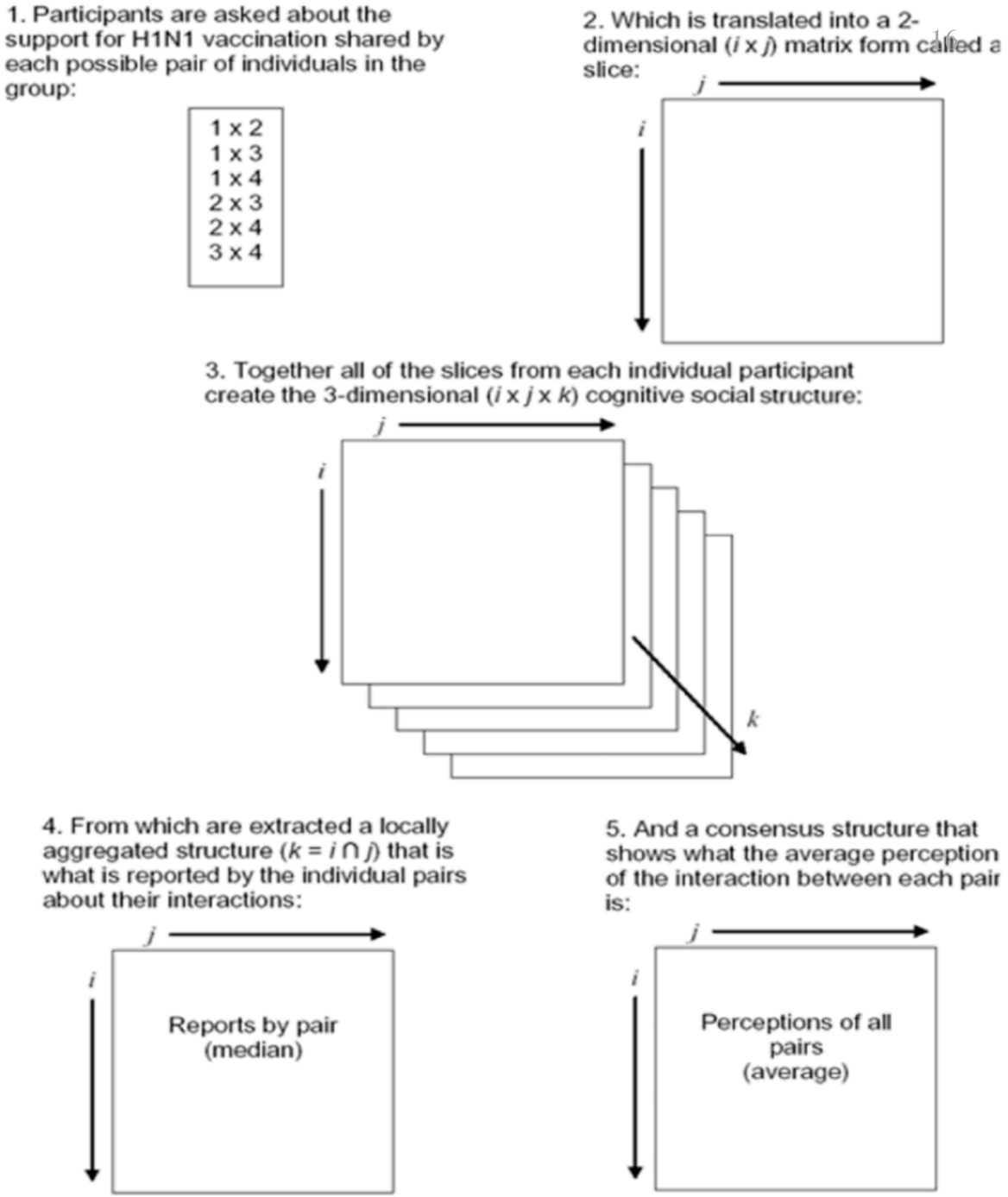
Archival alternative

Hollway and Koskinen 2016

Cognitive social structure (CSS)

- Asks respondents to report on the structure of relations of *others* in the network from their point of view
- Then similarities and discrepancies between the network as reported by individuals in the network is analysed
- Or impacts of perceived social structure on outcome is considered, e.g. perceived influence/popularity

See Krackhardt 1987; Knoke 1998; Knoke et al. 2019



Ego's Perception of Relationship Frequency and Valence

P01
Studies in Social Relationships and Health

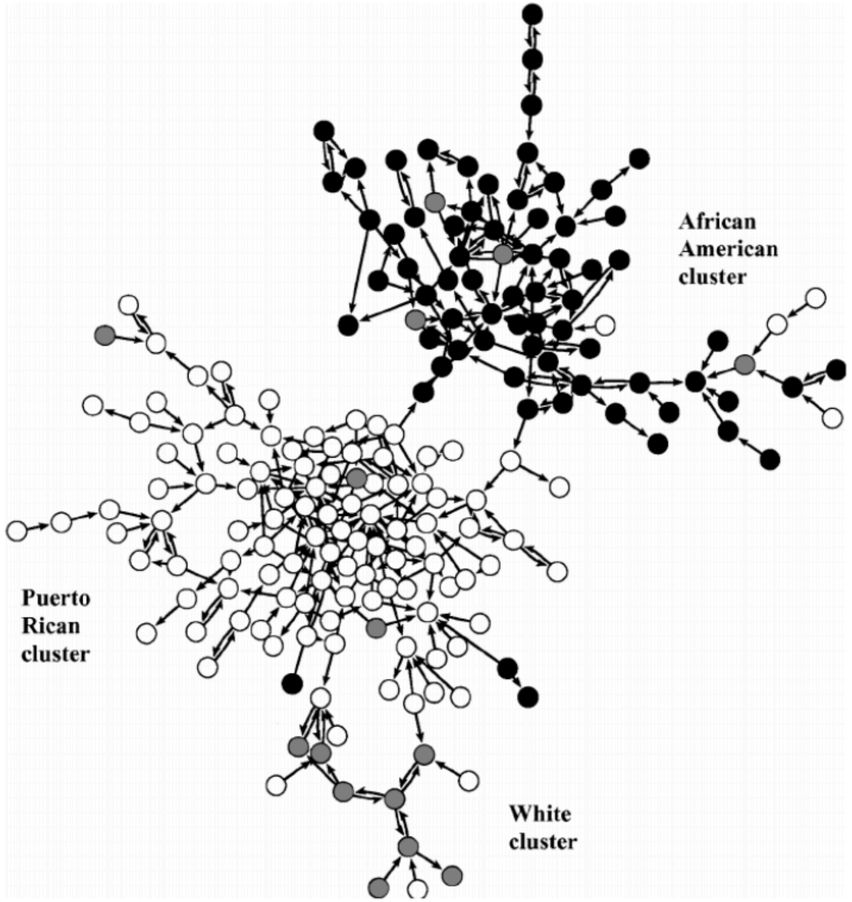
Y10 Personal Network Structure

Case-ID	Row #	Name	F	V	F	V	F	V	F	V	F	V	F	V	F	V	F	V	F	V
	1																			
	2																			
	3																			
	4																			
	5																			
	6																			
	7																			
	8																			
	9																			
	10																			

Frequency of Contact:
1) Every day
2) Several times a week
3) Once a week
4) Once every two weeks
5) Once a month
6) A couple of times a year
7) Once a year
8) Less than once a year
9) Have never spoken

Valence
-3) Dislike each other
-2)
-1)
0) Neither like or dislike
1)
2)
3) Like each other

Snowball sampling



Needle sharing among IV drug users in Hartford, CT

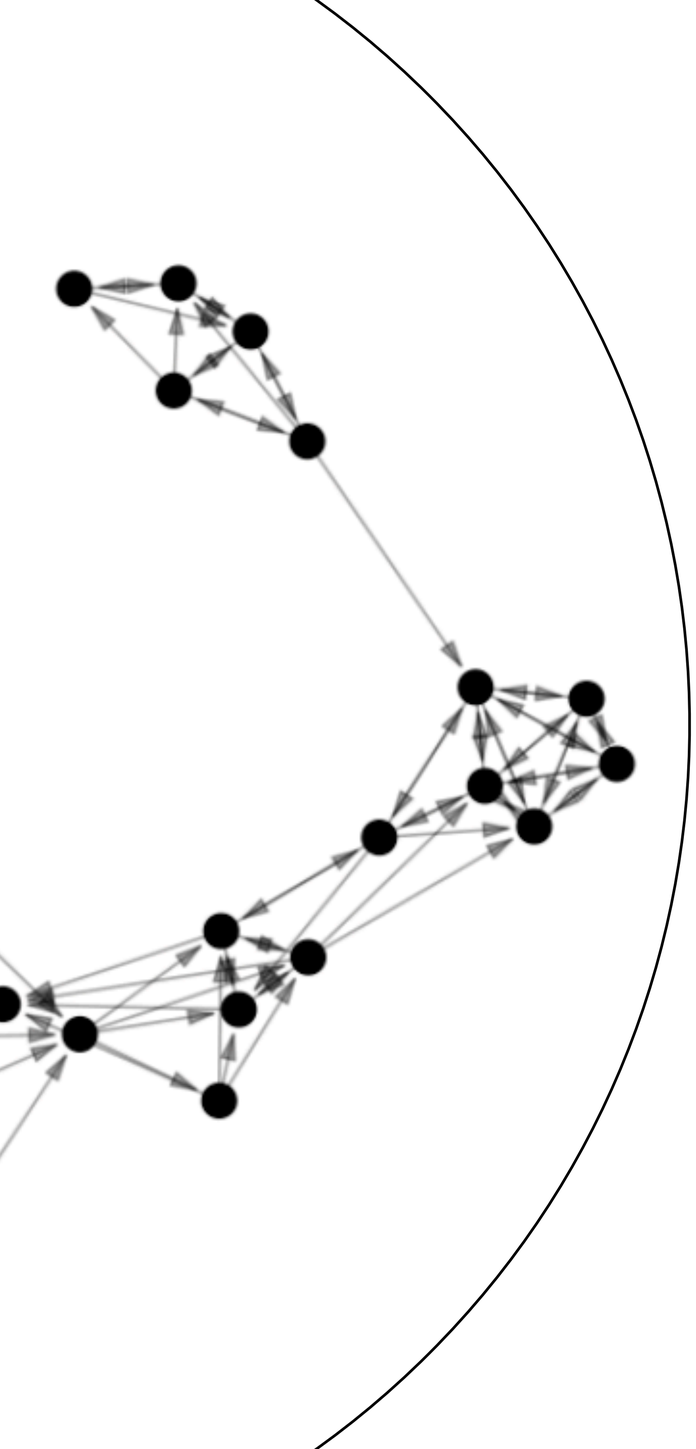
some seed nodes recommend others, who recommend more, usw; a non-probability sampling method that concentrates on most accessible part of the population

- cheap, simple, and cost-efficient
- easy way into hard-to-reach (small, covert) groups

But

- oversamples most public and well connected (more often seeds and lie on more recruitment paths) and may miss isolated/weakly connected individuals/groups
- biases towards particular network structures (like high degree) and not representative

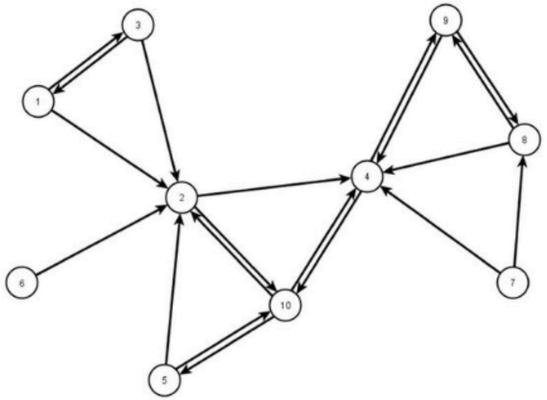
Respondent-driven sampling (RDS) aims to mitigate some of these concerns by weighting the sample to compensate for non-random recruitment patterns



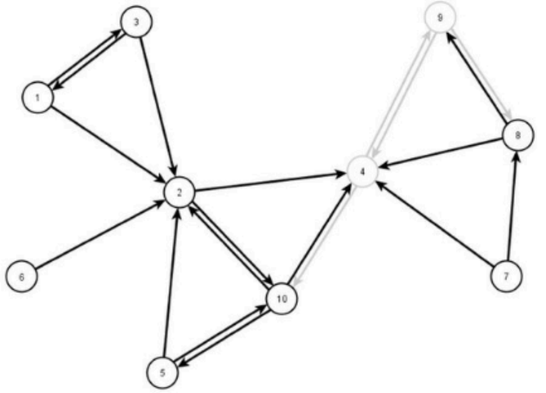
Boundaries

- Relational approach (i.e. connected):
  - e.g. all “relations” connected socially to main/seed individuals
- Event-based approach (i.e. attendance):
  - e.g. all “regulars” that go to the beach each day for 3 days
- Positional approach (i.e. characteristics):
  - e.g. all “employees” employed by an organization

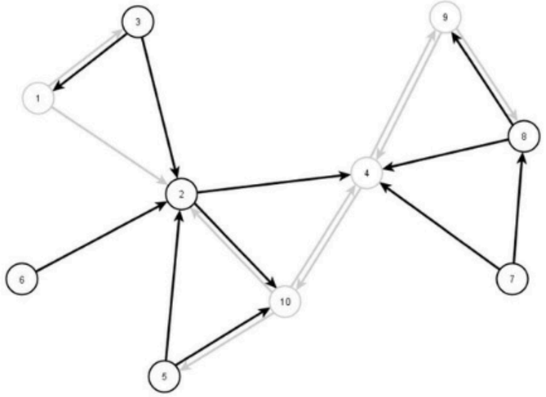
100% participation



80% participation



60% participation



Missing data

- Unfortunately, even a small fraction of missing observations can be problematic
  - A single non-response in a large survey is 1 missing observation
  - But a single non-response in a network survey is $n - 1$ missing observations
- It can significantly bias network structure
  - What do you do with missing targets?
  - Can skip how communities connected at the margin
  - Moreover, missing often not at random…

De la Haye 2017

Ethical concerns

- Consider study ethics from initial research design
  - IRB reviews
  - Report in publications
- Study participation needs
  - voluntary, informed consent
  - minimal intrusiveness
  - appropriate protections of privacy
- One of the main advantages of networks is also a source of greatest ethical concern...
  - *what is it?*

Lesson #1:
How data collected
affects how data can be analysed

Demanding data

- Networks is demanding of data
  - But true of all attempts at providing persuasive evidence
- Where data comes from crucial because:
  - how meaningful your descriptive or inferential conclusions depends on tie-data being *meaningful*
- **Most important lesson**: always ask what a tie *means*
  - i.e. how can we interpret structures relationally

Tie content

Actions

- talks to
- sells to
- gives aid/advice to
- sleeps with

Cognitive-affective

- likes
- knows
- despises
- recognises

Role-based

- kinship: brother of, daughter of
- social: friend of, competitor of
- organisational: boss of, teacher of

Tie content

Actions

- talks to
- sells to
- gives aid/advice to
- sleeps with

Cognitive-affective

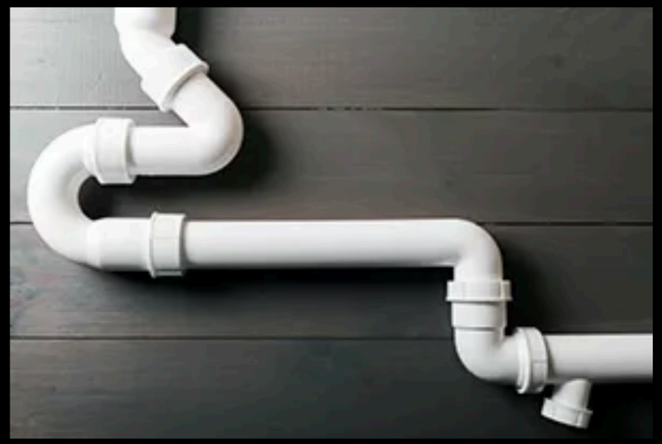
- likes
- knows
- despises
- recognises

Role-based

- kinship: brother of, daughter of
- social: friend of, competitor of
- organisational: boss of, teacher of

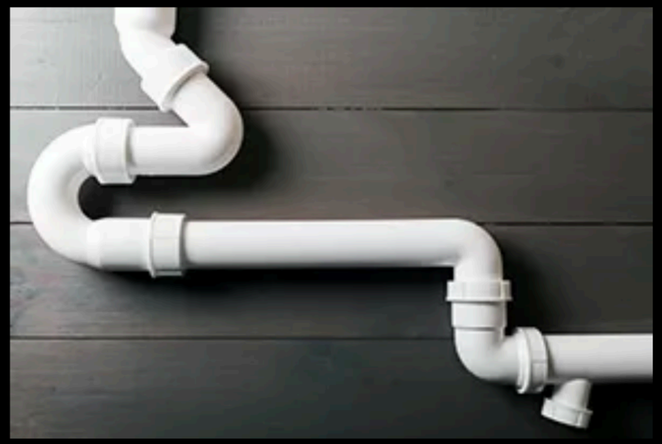
Ask: what is *at stake*?
What is the tie *doing* or *for*?

Pipes



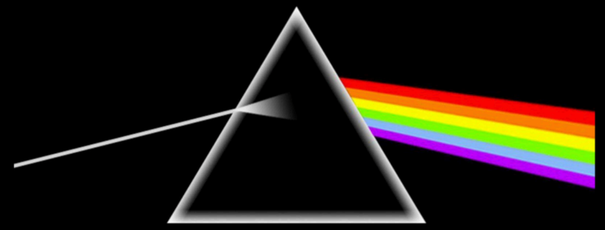
- network ties often seen as “plumbing” through which “stuff” flows
- stuff can be ideas, “capital”, etc
- micro questions about position, inequality, etc
- macro questions about network intervention, resilience, etc

Pipes



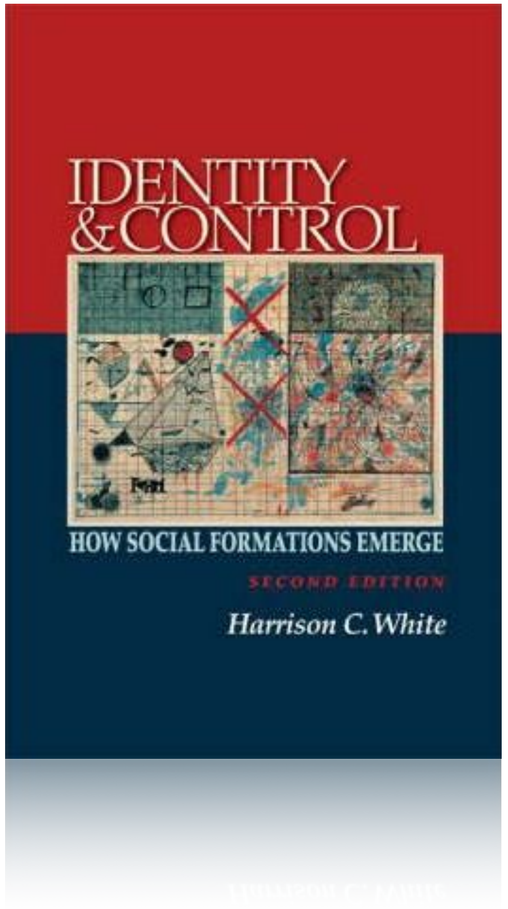
- network ties often seen as “plumbing” through which “stuff” flows
- stuff can be ideas, “capital”, etc
- micro questions about position, inequality, etc
- macro questions about network intervention, resilience, etc

Prisms



- network ties also serve as informational “cues”
- cues can be to identity, preferences, etc
- micro questions about how local networks affect perception
- macro questions about emergent cultures from certain network topologies

Networks and Culture



- *structure* long linked with *culture* (e.g. Berger and Luckmann, Bourdieu, Douglas)
- 1970s "breakthrough" established social networks as method of *structural* analysis distinct from *cultural* analysis and Parsonian normative theory (Blau, White et al 1976)
- today tentative reconciliation in different directions:
  - *ties affect culture* (Erickson, Krackhardt, Kilduff, Carley, DiMaggio)
  - *culture affects ties* (symbolic interactionalism/cultural sociology of Emirbayer and Goodwin, Lizardo, Daisy)
  - a *duality* understanding (Breiger 1974, White 1992, Mische, Fuhse)

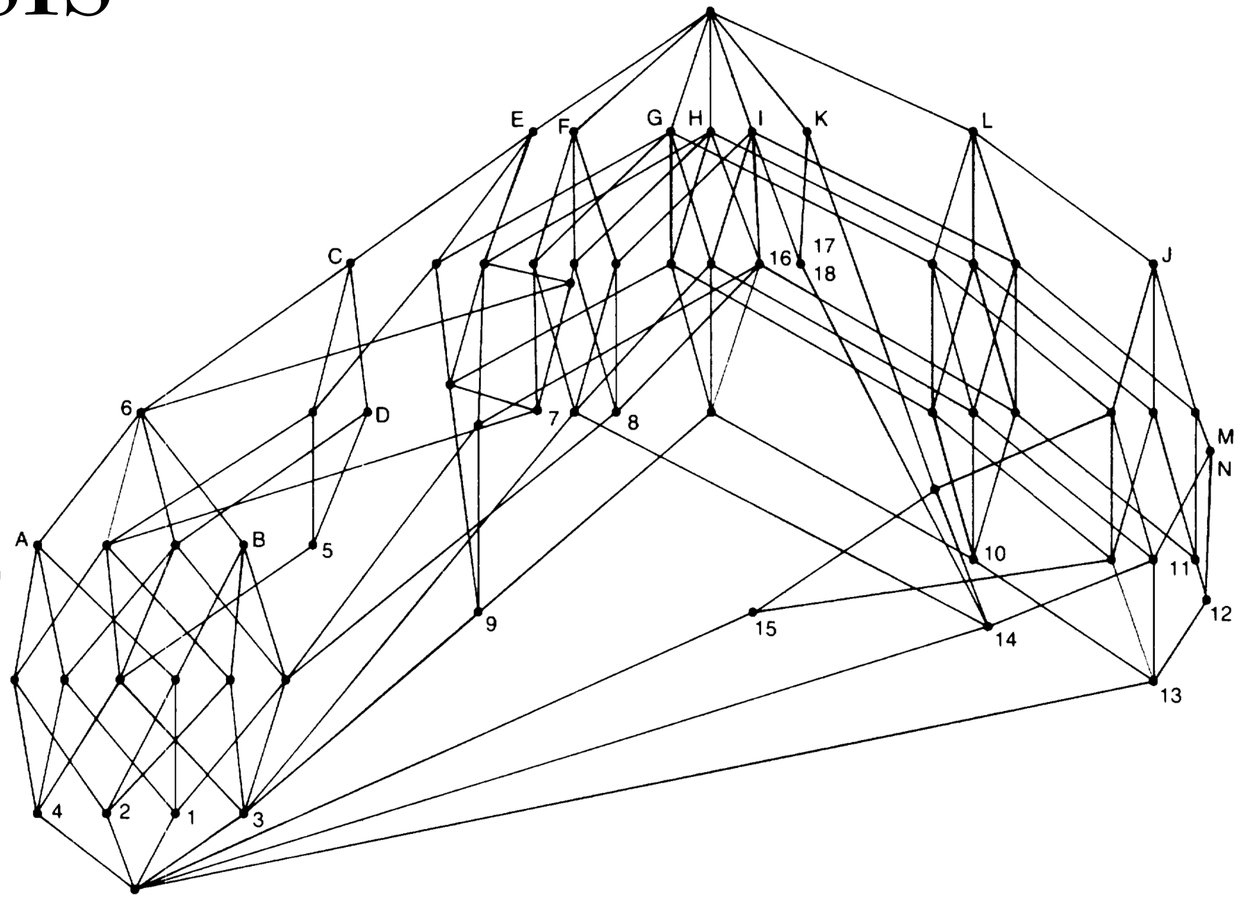
Duality

- Sociology of culture explores duality of persons and groups, persons and cultural elements, etc.
  - Dualism: bifurcating to separate things into distinct categories
  - Duality: recognition that thing has another, inseperable side to it
- I.e. we can understand a field of political action both from the perspective of politicians expressing beliefs but also beliefs affiliated to by politicians

Breiger 1974, Mohr and Duquenne 1997, Mische and Pattison 2000, Lee and Martin 2018

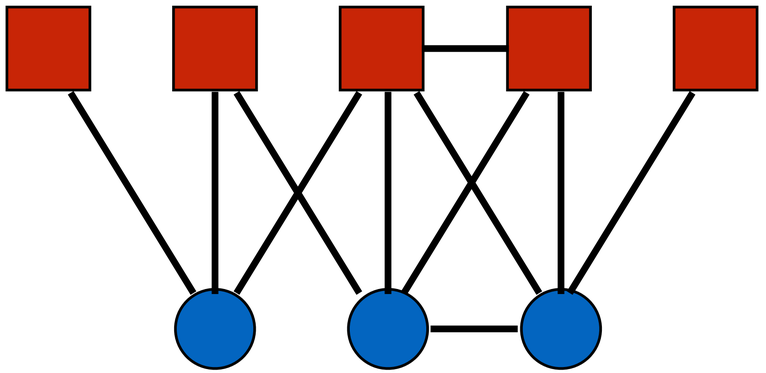
Formal concept analysis

- Takes as input a two-mode network, e.g. individuals (#) + events (letters) or documents + topics, etc
- Galois lattice represents the partially ordered subsets of elements from the perspective of both modes, and can be visualised in a Hasse diagram
- Reading down events included actors labelled at or below; EFGHIKL primary and contain most, CJ secondary (contained in E&L), ABDMN tertiary
- Reading up actors participated in all events labelled at or above; 1,2,3,4,13,14,15 primary, rest secondary except for 16 (tertiary)
- ABCDE and JKLMN share no common actor; FGHI are 'bridging' events, with at least one actor from each, and 8 and 16 are in between



Freeman and White 1993, Ganter and Wille 1996

Multilevel Meaning



Socio-semantic networks: Roth et al

Socio-ecological networks: Bodin et al

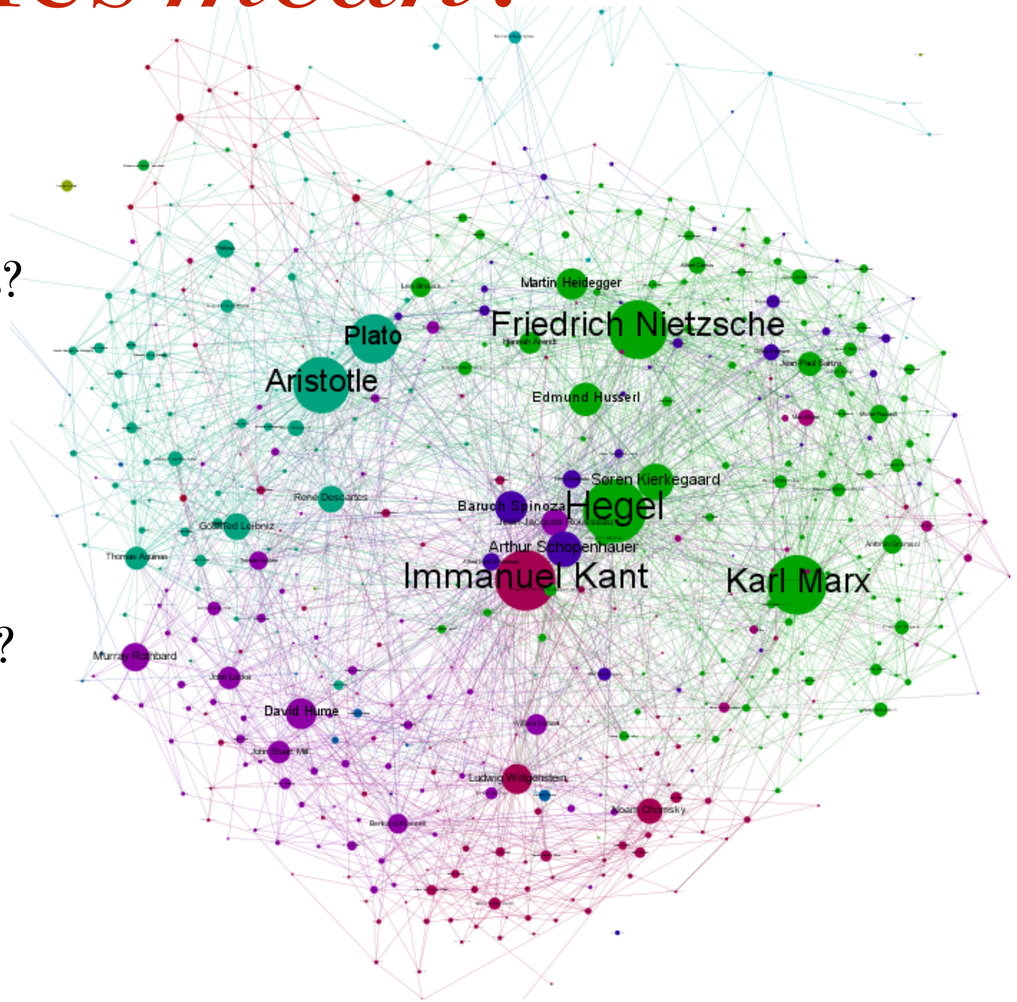
Socio-organizational networks: Hollway et al

Socio-political networks: Knoke et al

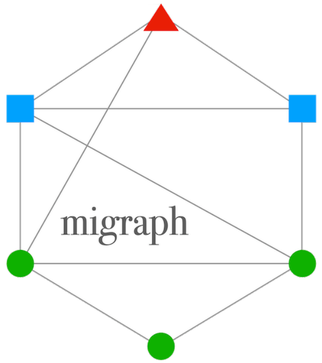
Lesson #2:
Meaning matters

What do you think these ties *mean*?

- A network of prior knowing between class members?
- A network of philosophers on Wikipedia influenced by other philosophers?
- A network of kinship in a royal family over time?
- A network of citations between articles on a topic for the past five years?
- A network of co-citations between articles on a topic for the past five years?
- A network of hyperlinks between NGO websites?
- A network of keyword co-occurrences in texts?
- A network of retweets (sic) between accounts/posts?



Getting started



- Download and install R (click download R and then select closest mirror): https://www.r-project.org/

- Download, install and open RStudio (you don't need to open R): https://www.rstudio.com/products/rstudio/#Desktop

- Install migraph (you don't need to install any other packages) by typing `install.packages("migraph")` or use point+click