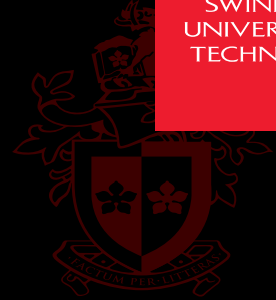


SNA study design and data structures

SWIN
BUR
NE

SWINBURNE
UNIVERSITY OF
TECHNOLOGY



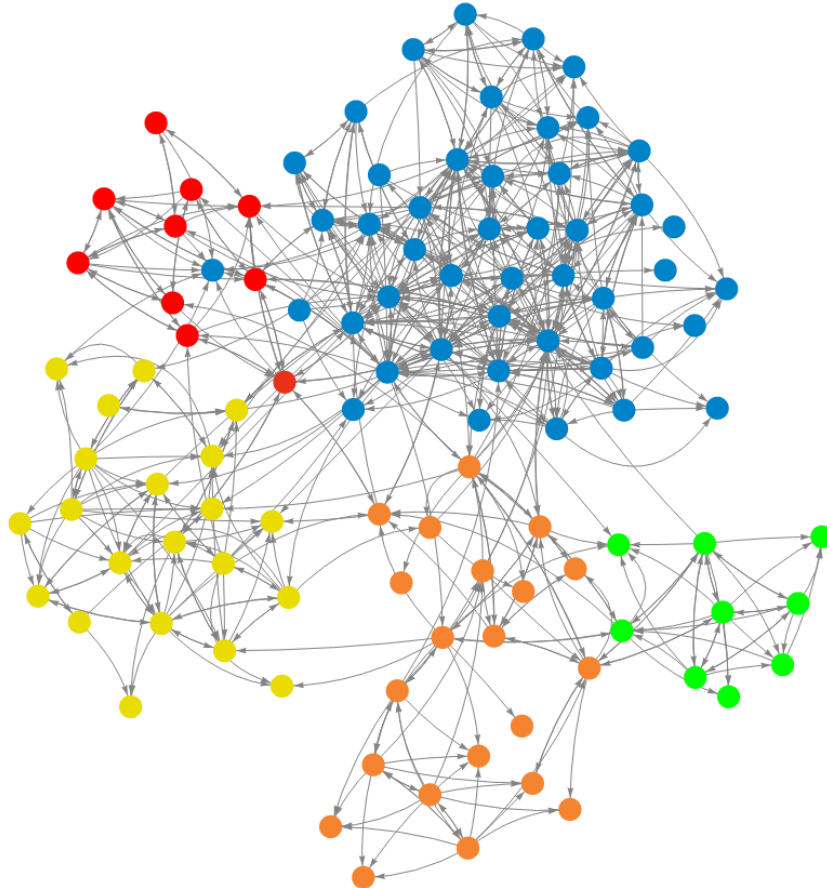
Basic terminology

- *Actors*
- *Ties*
- *Social Network*
- *Graph*
- *Nodes*
- *Edges, arcs*
- *Directed, undirected*
- *Adjacency matrix; sociomatrix*
- *Actor attributes*

1. Whole network designs

- A *whole network design*, also called a *full network* study, requires a single set of actors within a well-defined **network boundary**.
 - The data includes the ties that are present among all actors, sometimes referred to as a *network census*.

1. Whole network designs



The choice of boundary ought to be theoretically justified

Should we include clients?
Suppliers?



1. Whole network designs

- Through a survey, actors may be asked to nominate network partners.
 - The nomination of network partners is often done through a *name generator* survey item
 - Individual actors are also typically surveyed about attribute variables.
 - Electronic and digital data collection also possible.
- *Unipartite* or *one-mode* design, because only one type of node.

Three types of attribute variables

1. Binary ("0" or "1")

- Female/Male
- Success/Fail

2. Categorical

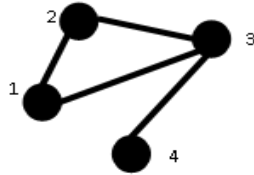
- Work teams within a company
- Political party

3. Continuous

- Age, attitudes

Name	Female	Attitudes	Performance	Department
Annie	1	7	95	1
Erin	0	7	90	4
Faye	0	6	85	2
Brett	0	7	100	4
Lynette	1	5	80	1
Constance	0	7	80	4
Kristi	1	6	70	4
Santos	0	6	70	2
Alexis	1	4	90	3
Rickey	1	5	85	4
Amelia	1	5	86	4
Luis	0	6	87	2
Violet	1	7	100	4
Inez	0	7	100	4
Alfonso	0	7	95	1
Holly	1	7	100	4
Wilson	0	5	80	3
Nancy	1	5	80	2
Mercedes	0	6	70	4
Katie	1	3	99	2
Marvin	0	7	100	1

Undirected whole network data



(a)

	1	2	3	4
1	0	1	1	0
2	1	0	1	0
3	1	1	0	1
4	0	0	1	0

(b)

0	1	1	0
1	0	1	0
1	1	0	1
0	0	1	0

(c)

1 2
1 3
2 3
3 4

(d)

(a) Graph/sociogram

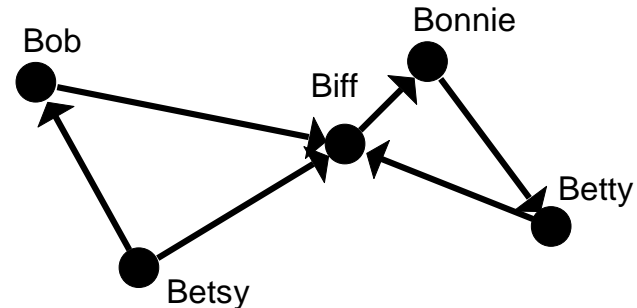
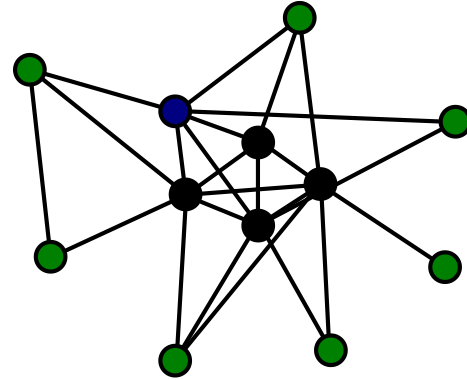
(b) Adjacency matrix with row and column headings

(c) Adjacency matrix

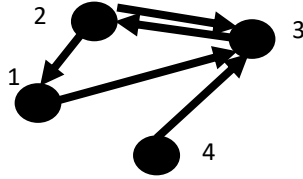
(d) Edge list

Directed vs. Undirected

- Undirected relations
 - Attended meeting with
 - Communicates daily with
- Directed relations
 - Lent money to
 - Asks advice from
 - Is friends with??



Directed whole network data



(a)

0	0	1	0
1	0	1	0
0	1	0	0
0	0	1	0

(c)

	1	2	3	4
1	0	0	1	0
2	1	0	1	0
3	0	1	0	0
4	0	0	1	0

(b)

1 3
2 1
2 3
3 2
4 3

(d)

(a) Graph

(b) Adjacency matrix with row and column headings

(c) Adjacency matrix

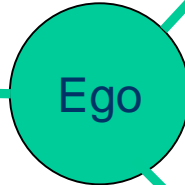
(d) Edge list

2. Egocentric network studies

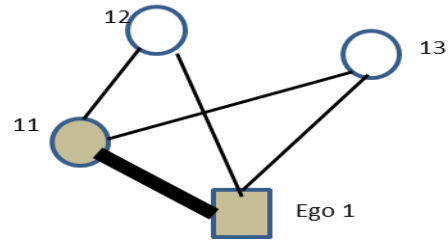
- An ego network is a network centred on a single actor (e.g. a person), referred to as the 'ego', together with
 - his/her 'alters'
 - usually alter to alter relationships
 - Often attribute data on ego and/or alters
 - All information about alters and alter-alter ties are based on reports by ego.

Who does ego vote for?

Information
Social influence



With whom do you
discuss politics?



Ego-ID	Sex
1	1
2	2
3	1

Alter-ID	Ego-ID	Sex-A	Close
11	1	1	1
12	1	2	0
13	1	2	0
21	2	2	1
22	2	2	1
23	2	2	0
31	3	1	1
32	3	1	1
33	3	2	1

0	1	1
1	0	0
1	0	0

Ego file

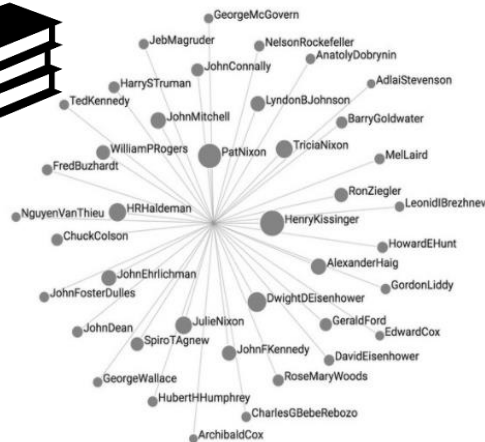
Alter file

Alter-alter file for Ego1

Initials

Initials	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1																				
2																				
3																				
4																				
5																				
6																				
7																				
8																				
9																				
10																				
11																				
12																				
13																				
14																				
15																				
16																				
17																				
18																				
19																				
20																				
Gender	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M
Initials	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F


Richard Nixon's ego network



Richard Nixon's ego network, based on his memoirs.




Network Canvas




ABOUT THE PROJECTDOCUMENTATIONFEEDBACK

DOWNLOAD

 We are proud to announce the stable release has arrived!
[Click here to download](#), or [read migration instructions here!](#)

Software to simplify complex network data collection.



Network Canvas is a **free and open-source** set of software for surveying personal networks, designed around the needs of both researchers and participants.

Strengths and weaknesses of whole network and egonet designs

1. Whole network designs

Strengths

- Capture all connectivity within the network boundary – we are able to draw conclusions about the whole social system
- Social influence can be better assessed, as we are not reliant on ego's reports of alter's attributes.
- Network structure can be properly assessed because the data does not have to be inferred from sampling procedures

1. Whole network designs

Issues

- a whole network design typically requires a high response rate.
 - So may not be useful for large scale studies, unless the data can be collected automatically, perhaps electronically.
 - Data collection can be very resource-intensive
- cross-sectional whole network designs can assess associations between ties and attributes but cannot differentiate selection and influence, or other causal effects
- Often a case study (how to generalize?) – e.g. our organization is not a representative sample of all organizations

2. Egocentric network studies

Strengths

- Participants can be sampled in standard ways from a population.
- Egonets can be treated as a sample of independent observations, assuming that alters are not themselves respondents and are not shared across egonets.
- As a result, once summary measures about the egonet are calculated, these measures can then be analyzed using standard statistical approaches because observations are independent between egos.

2. Egocentric network studies

Issues

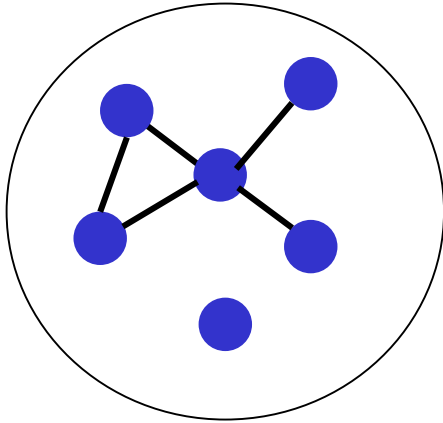
- assumes that respondents can reliably provide information about alters and alter-alter ties.
 - may be problematic when we need good measures of behaviors or attitudes for both partners in a network tie.
- provides no observations of network ties beyond each egonet, and so is limited in assessing network connectivity.

Other types of network research designs

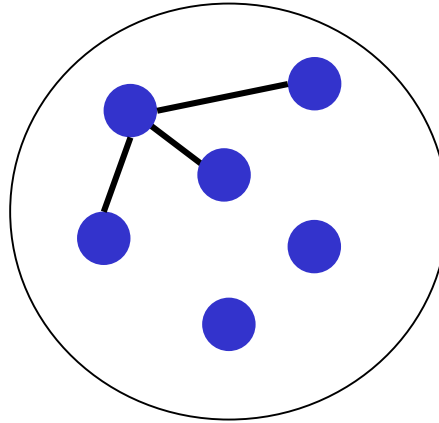
3. Multiple network designs

- Studies that involve several different whole networks.
 - network data is collected from several collective units (schools, classes, teams etc.) with each unit treated as creating a boundary and thereby a separate whole network.
 - The networks are assumed to be distinct without ties across boundaries to other units.
- The idea behind a multiple network study is to compare network effects across units with the goal of identifying major effects that apply to all units.

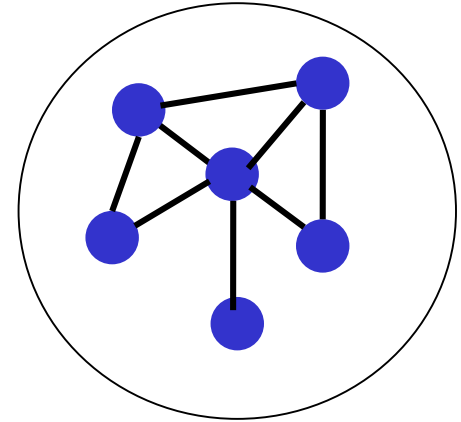
Team 1



Team 2



Team 3



Network Density = 0.33
Performance = 8

Network Density = 0.2
Performance = 2

Network Density = 0.5
Performance = 5

3. Multiple network designs

Strengths

- inferences can be made across many units, and not simply rely on a case study of one or a few

Issues

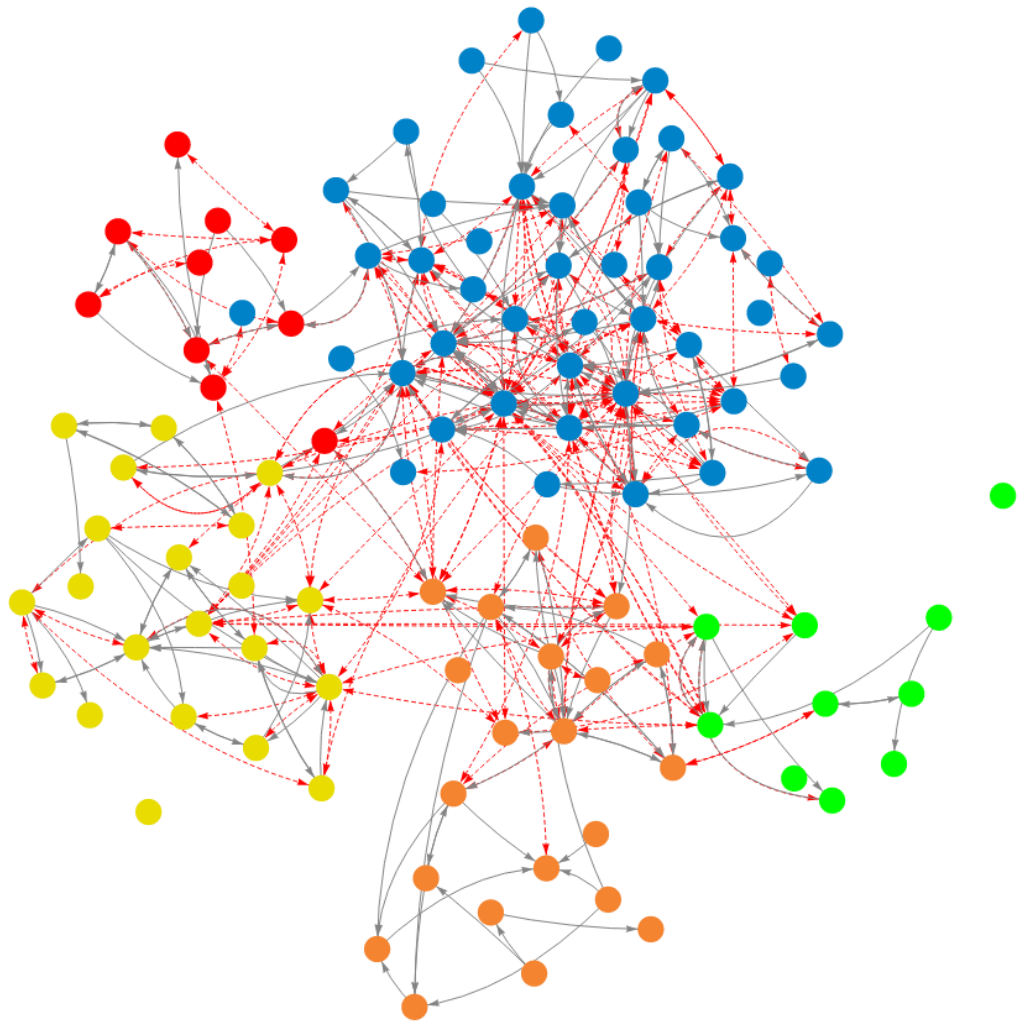
- assumption is that the units are not connected by the types of tie under examination (in other words, that these networks can be treated as distinct).
 - It is not always obvious how well this assumption is met, so care needs to be taken in sampling and selecting units.
- Depending on number of units, data may be more costly and difficult to obtain.

4. Multiplex network studies

- several different types of relational ties are measured on the same set of nodes.
 - a simple extension of the whole network data structure, except that there are now multiple adjacency matrices or edge lists, one for each type of relational tie.
 - may be entered into k different files (for k types of relational ties),
 - or in some cases, the matrix can be “stacked” into one $k \times (n \times n)$ matrix in the one file.

E.g. advice

E.g. friendship



Network sampling

7. Network sampling designs

- traditional random sampling is **not** sensible for a network study (except for egonet studies).
 - Dependencies among individuals are lost by random sampling
 - an understanding of network connectivity is better obtained with some form of snowball sampling or link tracing design.
- Two broad motivations
 - to obtain information about network structures in larger communities
 - to obtain information about individuals in populations that are “hidden” or “hard to reach” (e.g. drug users).

7. Network sampling designs

- ***Strengths and issues***
 - Network sampling can provide direct evidence about network effects in a larger context not well suited to whole network studies. But data collection demands can be substantial.

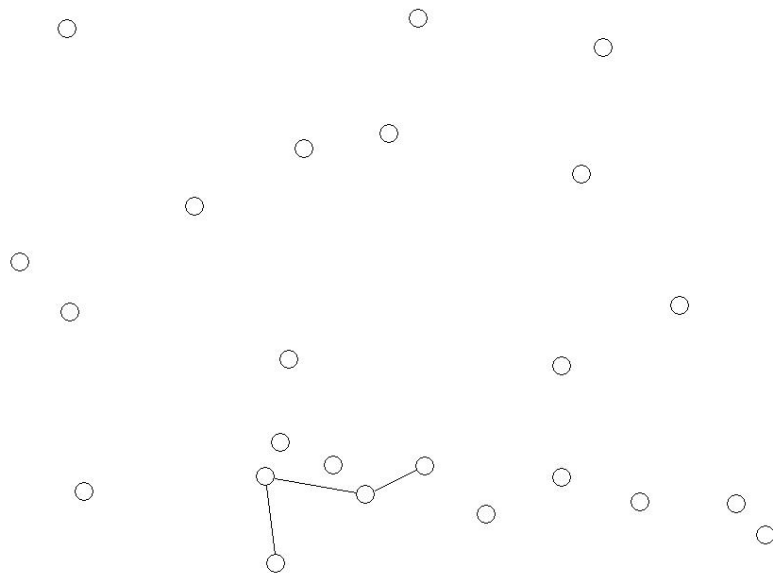
7a. Snowball sampling

- the preferred approach for investigating network structure using sampled data
- A snowball sample is obtained by starting from an initial set of respondents (*the seed set*), determining their network partners (*wave 1*), determining the new network partners of wave 1 respondents (*wave 2*) and so on until stopping at an agreed number of waves.

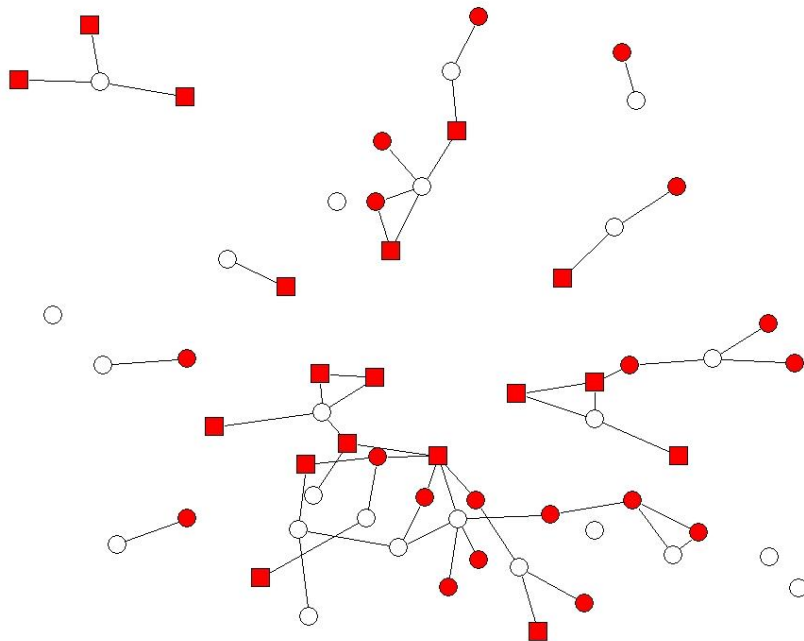
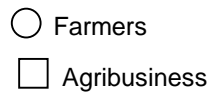
Network sampling

○ Farmers
□ Agribusiness

See set (wave 0)

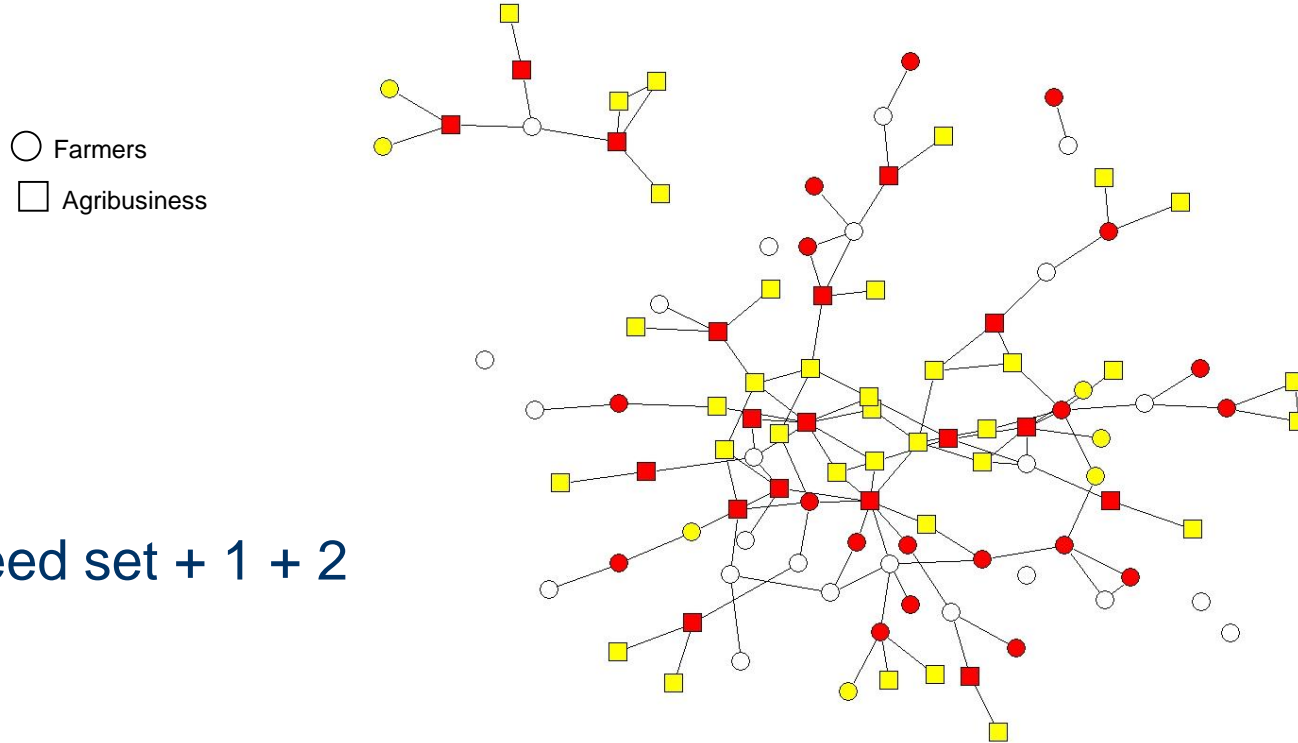


Network sampling



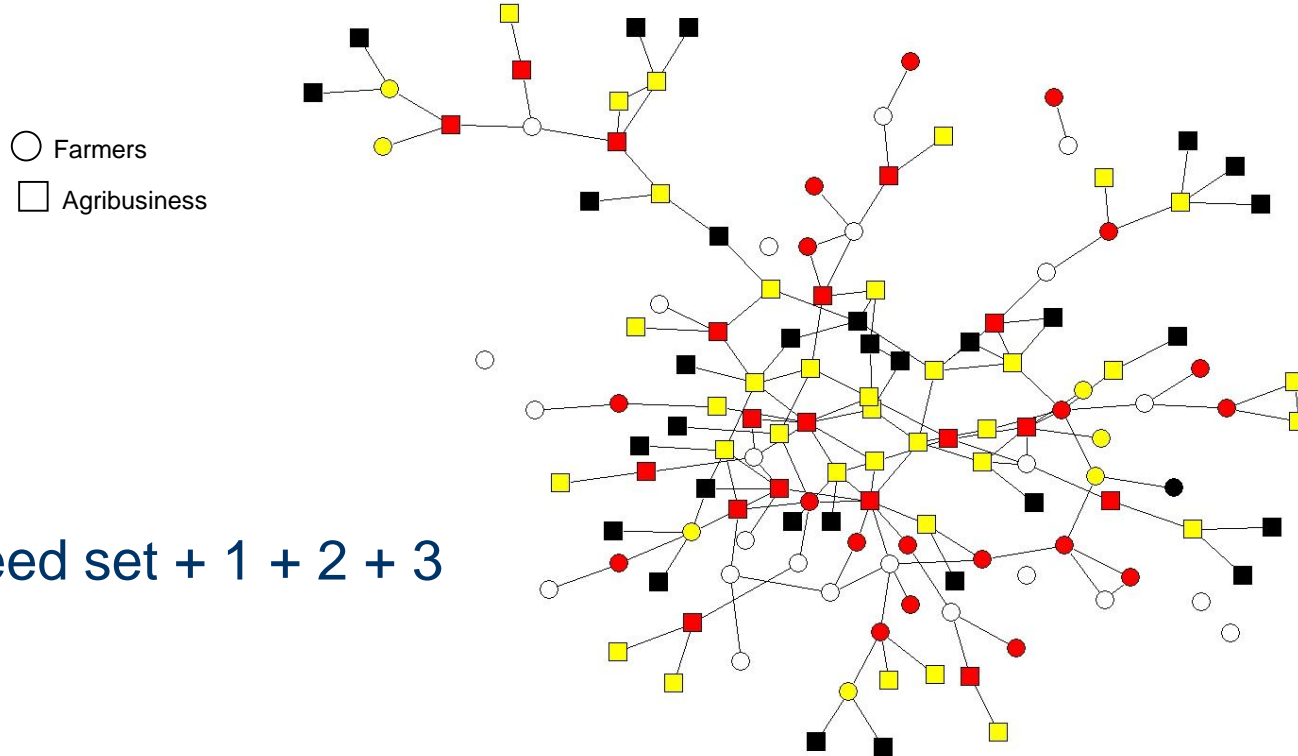
Seed set + 1

Network sampling



Seed set + 1 + 2

Network sampling

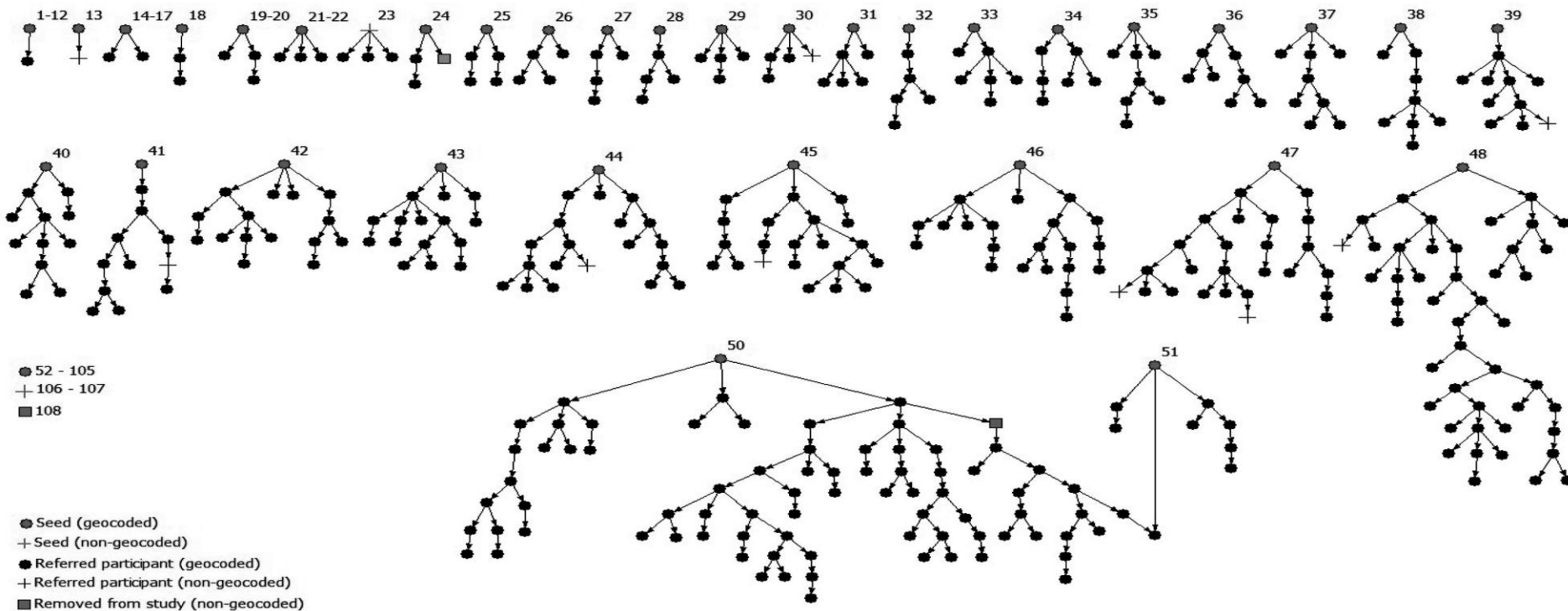


7b. Respondent driven sampling

- used to identify actor attributes (e.g. disease status) of a hard-to-reach population,
 - not so relevant to investigations of network structure
 - More common in epidemiological research
- a *chain referral* or *link tracing* method, whereby respondents themselves try to recruit one or more of their network partners
 - some chains of referrals may be very long, but others very short when some respondents do not recruit or recruit only a few partners.

7b. Respondent driven sampling

- Method:
 - a small number of participants recruited as a seed set;
 - all participants offered an additional financial reward for recruiting network partners into the study;
 - each participant given recruitment coupons so that they were financially rewarded if they were passed to peers from the population who then turned up for interview.



Young et al., (2014). Spatial, temporal and relational patterns in respondent-driven sampling: evidence from a social network of rural drug users. *Journal of epidemiology and community health*, 68(8), 792-798.

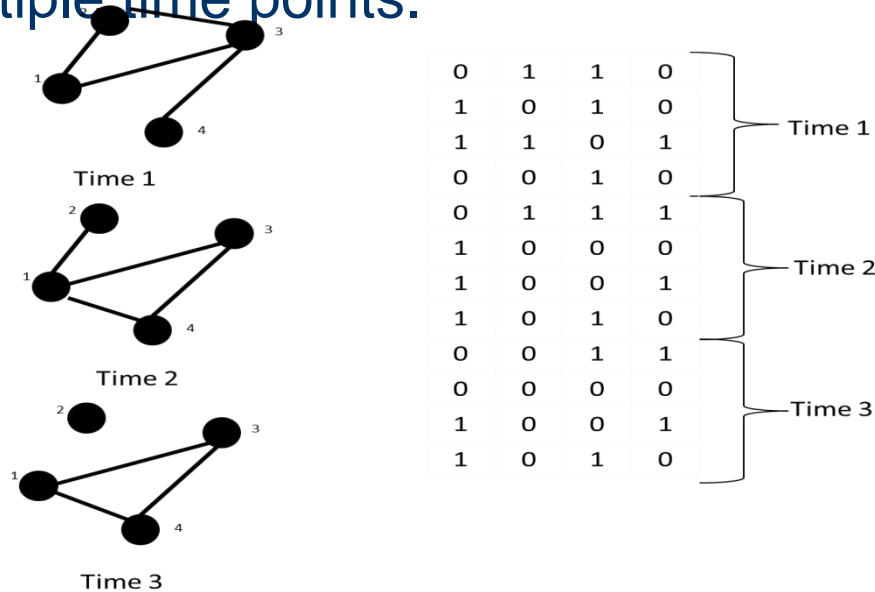
Longitudinal network designs

8. Longitudinal designs

- Egonet, whole network and bipartite network designs can be readily adapted to a panel approach simply by collecting data at multiple time points.
 - inferences differentiating selection from influence can be made
- Event-based designs are also becoming popular

8a. Panel network data

Whole network data measured within the one network boundary at multiple time points.

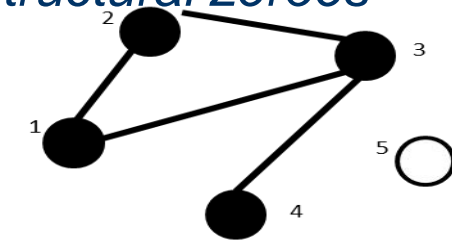


May also have actor attributes measured at multiple time points:
Coevolution design

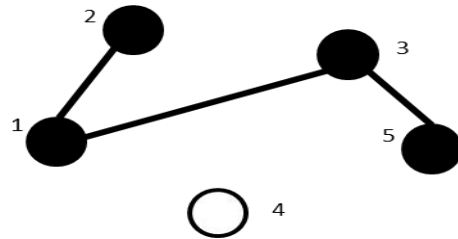
8a. Panel network data

But node set may change at different time points, or respondents may only complete survey at some time points.

- *Structural zeroes*



Time 1



Time 2

0	1	1	0	0
1	0	1	0	0
1	1	0	1	0
0	0	1	0	0
0	0	0	0	0
0	1	1	0	0
1	0	0	0	0
1	0	0	0	1
0	0	0	0	0
0	0	1	0	0

Time 1

Time 2

8b: Relational event design

- A network boundary is determined to define actors.
- Data is a sequence of **transactions** across time.
 - Next transaction (event) predicted from the past patterns of events.
- well suited to email and other electronically collected data, where it is simple to collect the timing of the event (e.g., an email from i to j at time t).
- The sequence of events can be recorded as an edge list, where the first edge is the caller and receiver for the first transaction and so on.
 - The precise start and end time of a transaction can be entered as a third and fourth column.

Relational Event Modelling example

- <http://www.melnet.org.au/events>
- Relational Event Modelling Workshop
- Code, data and slides for running REMs on data from the movie *Frozen*