
**What are network data:** social network data consist of at least one structural variable measured on a set of actors.

- **Structural variables:** measure ties of a specific kind between pairs of actors.
- **Composition variables:** measure actor attributes.
- **Mode:** a distinct set of entities on which the structural variables are measured.

**Boundary specification and sampling:** a researcher must identify the population and figure out how to sample when necessary.

- **Population:** who are the relevant actors? It is assumed that “we can obtain relevant information on all substantively important actors” and these actors “consist of all social units on which we have measurements”
- **Sampling:** “a sample of actors might be taken” when it is “not possible to take measurement on all the actors”

**Types of network:** networks are categorized by the nature of the sets of actors and the properties of the ties among them.

- **One-mode network:** a single set of actors + one or more types of relations between pairs of the actors + actor attributes
- **Two-mode network:**
  - Dyadic two-mode network: two sets of actors + one or more types of relations between actors in the two sets;
  - Affiliation network: one set of actors and one set of events + attendance/membership + attributes of the actors and the events
- **Ego-centered and special dyadic networks:** couples; mothers-children; ego-centered network

**Network data, measurement and collection:**

- **Measurement:** “social network data consist of one (or more) relations measured among a set of actors”
  - Unit of observation: actor/dyad/triad/subset of actors/network
  - Modeling unit: actor/dyad/triad/subset of actors/network
  - Relational quantification: directional vs. non-directional; dichotomous vs. valued
- **Collection:** techniques used to gather network data
  - Questionnaire: roster vs. free recall; free vs. fixed choice; ratings vs. complete ranking
  - Interview: when questionnaires are not feasible; ego-centered network
Week 4. Data Collection & Social Cognition

- **Observation**: small group of people; when questionnaire and interview are not feasible; affiliation network data
- **Archival records**: longitudinal relations and ties existing in the past
- **Other**: special network designs
  - **Cognitive social structure**: “respondents give information on their perceptions of other actors’ network ties”
  - **Experimental**: selected actors (and specified pairs)
  - **Ego-centered**: egos and alters
  - **Small world**: the length of the chain and the characteristics of the actors
  - **Diary**: personal network
- **Longitudinal data collection**: “how ties in a network change over time”
- **Measurement validity, reliability, accuracy, error**: true structure vs. observed structure
  - **Accuracy**: the accuracy of the verbal report; long term pattern
  - **Validity**: construct validity
  - **Reliability**: test-retest comparison; comparison of alternative question formats; reciprocity of sociometric choices
  - **Measurement error**: error in fixed choice data collection design

**Data sets found in these pages**

- **Krackhardt’s High-tech Managers**: one-mode (21 actors); three relations; four attributes; questionnaire
- **Padgett’s Florentine Families**: one-mode (16 actors); two relations; three attributes; archival
- **Freeman’s EIES Network**: one-mode (32 actors); two relations; two attributes; archival?
- **Countries Trade Data**: one-mode (24 actors); five relations; four attributes; archival
- **Galaskiewicz’s CEOs and Clubs Network**: two-mode affiliation network (26 CEOs-15 events); several attributes; interview + archival


**Issue under study**: “the problem of specifying system boundaries”

- “The selection of actors or nodes for the network”
- “The choice of types of relationships among the actors”
- Errors in the system boundary definition will result in “a fundamental misrepresentation of the process under study”

Terence Meehan & Ning Sa
Approaches to boundary definition: two basic ways

- **Realist approach**: setting network boundaries by definition
  - Works better for formally constituted groups
- **Nominalist approach**: the investigator draws the boundary for his/her own purposes

Definitional foci for the inclusion of actors:

- **Attribute based**
  - Positional approach: presence/absence of some attribute
  - Reputational approach: judgments of knowledgeable informants
- **Relationship based**: whether the actors participate in a specific social relationship
  - Snowball sampling
- **Event/activity based**: whether the actors participate in a specific event or activity
- **Multiple foci**: using two or more of the three foci

Illustrative boundary specification strategies: 8 boundary specification approaches

- **Realist x Attributes (I)**: attendance at a particular high school
  - small, tightly bounded group
- **Nominalist x Attributes (II)**: Directors of xxx largest corporations in year xxxx
  - large networks
- **Realist x Relation (III)**: primary group
- **Nominalist x Relation (IV)**: small world
- **Realist x Activity (V): Street corner society**
  - Alternative to strategy I
- **Nominalist x Activity (VI)**: “invisible college”
- **Realist x Multiple foci (VII)**: “Marxian concept of class for itself”
- **Nominalist x Multiple foci (VIII)**: “American political elite”
- **Central difficulty with VII and VIII**: “they consume many theoretical degrees of freedom”

On inclusion rules for relations:

- **Partial system fallacy**: a set of relationships is analyzed without considering the entire set of actors
- **Issues in multiple network studies**
  - Potential generator
  - Absent ties

Boundary specification for activities:

- The analyst is “intrinsically interested in the events”
- The selection is an “intermediate step” in describing the structure
Week 4. Data Collection & Social Cognition


**Intro:** Network data in the field of interorganizational studies is increasingly used and useful. They help measure the environment in which organizations work. Reliability of data is an open question.

- One strategy to gather data is through secondary data. For example, looking into interlocking directorates.
- Another strategy is direct researcher observation, though that had not been used for interorganizational studies at the time of this paper (1993).
- A third strategy is through surveys and questionnaires, but studies have shown people have difficulty remembering who they have interacted with unless the interaction is routine.

To look at the reliability of data, the authors look at mutual confirmed relationships from a large dataset.

**Dataset:** Data were gathered in 1989 (for a separate study) from 31-44 mental health service providers in 4 US metro areas. Key informant was identified and interviewed. The informant then filled in a questionnaire.

**Measurement:** Two questions – answered on a Likert scale of 0 to 4 – from the original survey were used to explore accuracy of the data:

- “To what extent does your agency receive information for coordination, control, planning, or evaluation purposes from this agency?”
- “How well coordinated are the activities of your agency with those of this other agency?”

Symmetrical relations are assumed on these questions and then correlations were calculated.

**Reliability using organizational agents:** Confirmed presence or absence of links is used to estimate the reliability or quality of the network data. 2 aspects will be considered: knowledge of who informants have communicated with and their ability to correctly recall the intensity of the communication.

The correlations were 0.44 for the first questions and 0.43 for the second question. Perhaps this is partly because symmetry is not a good assumption. Or because the data are not reliable. (this section confused me and therefore is not well summarized here).

Overall percent agreement on an organization by organization basis was about 70%. Network data is relatively reliable.

**Systematic error in cognitive network data:** Does the presence of strong relations between organizations cause informants to be more aware of relationships? If so, this would introduce systematic errors in network data. The analysis confirms the presence of such error, but it appears inconsequential. But the stronger the relationship, the more likely it is to be confirmed.
Article comprehensively reviews findings about network measurement as of 2005. It can be used as a reference.

Network study designs:

- “Whole-network” studies consist minimally of one set of objects linked by one set of relationships on one occasion. Wasserman and Faust call this a “one-mode” data set. Expansions of this design include multiple occasions, multiple sets of objects (e.g. two-mode data), multiple relations (e.g. collaboration, advising and friendship) and/or longitudinal questions. This chapter won’t deal with object attributes, but it is common to collect these.
- Cognitive social structure design (CSS) includes measurements of relationships obtained from multiple sources.
- Egocentric designs consider a focal object (ego) and the actors it is linked to (alters).

Setting network boundaries:

3 basic strategies:

- positional approach (e.g. employment by an organization)
- event-based approach (e.g. people going somewhere 3 times or more)
- relational approach
  - expanding selection
- k-core concept, with researcher varying k.
- ego-centric studies use the “name-generator” questions to set boundaries.

Survey and questionnaire methods:

- Surveys often used, but best way varies depending on study design
  - global questions (like asking who is in your network) are not good
- Name generator instruments for egocentric networks
  - name generators identify respondent’s alters
    - single name
    - multiple name
  - name interpreters obtain info on the alters and their relationships
  - comparing name generators
    - various criteria” specific social exchanges or affective criteria
  - recall, recognition and forgetting
    - recall improves with closer ties
    - possible solution is to use recognition (from a predetermined list) rather than recall
Test-retest studies
- more than 75% of first-occasion alters are cited on second occasion.

Patterns in the free recall of persons
- people tend to remember others in clusters of social relations

Meaning and interpretation of name generators
- interpretation of relations (e.g. “friendship”) can differ between researcher and respondents.

Interview context effects
- context influences the interpretation of a name generator

Interviewer effects
- interviewer affects survey quite a bit
- training is necessary

Name interpreters
- most studies follow up with these
- respondents often become bored and so researcher must limit questions

Additional instruments for egocentric networks
- developed to elicit weak ties, for example
  - instruments for measuring extensive network size
    - “summation” method suggests that mean network size is between 280 and 290
    - “scale up” methods suggest mean of around 1700
    - “reverse small world (RSW) method suggests mean of 129
- position generators
  - measures linkages to specific locations directly (rather than the standard way of identifying particular alters and later identifying their social locations using name interpreters)
  - e.g. “do you have relationships with people having any of these 15 occupations?”
  - identifies weak ties

Resource generator
- to measure “social capital” – “resources owned by members of an individual’s social network, which may become available to the individual”
- measures whether someone is in personal contact with anyone having a specific possession or capability

CSS data
- These are judgments by several perceivers about each dyadic relationship in a network
  - can then look at:
    - a single observer’s “slice” of judgments
    - a “locally aggregated structure” of judgments by the two actors involved
    - a “consensus structure” based on all judgments about a given dyad

Informant biases in network perception
Week 4. Data Collection & Social Cognition

- “ego bias”: people think they’re more important/central to the network than the consensus perception
  - but “fairly reliable”

Informant Accuracy and Competence

- Correspondence between reports and observations
  - major source of inaccuracy lies in the different response sets or thresholds that respondents use when making citations
  - respondents also tend to “group” relations rather than see them dyadically.
  - observing social ties is also difficult
- Studies of informant competence
  - Some informants are better than others: “competence” varies
    - centrally positioned informants tend to be more competent
- Prospective uses of informants
  - different strategies for eliciting CSS-type data from informants

Archival network data

- Data collected for other reasons are useful for researchers and are less expensive and time consuming.
  - validity of such data rests on the correspondence between measured connections and conceptual ties of research interest
    - Sometimes they’re close
    - Sometimes they’re not
    - example of research journal citations, which are problematic
    - computer-mediated systems are a potential source of data (but they leave out face-to-face interaction).

Observation

- Used less and less often, due to difficulty

Conclusion

Notable advances in network measurement have occurred since 1990, especially for survey and questionnaire data. Instruments for measuring egocentric networks are now much better understood, and much has been learned about cognitive processes and biases involved in answering questions about social relationships.

Important questions of validity and reliability for survey/questionnaire data remain. The number and range of network studies that draw on archival materials has risen. Given the opportunities that archival sources present, it is important to scrutinize the quality of such data as closely as data from self-reports. Assessments of data quality, regardless of source, will be facilitated if researchers clearly articulate their concepts of the “true scores” they seek to capture with empirical indicators of network ties.

Terence Meehan & Ning Sa

Central argument: structural analyses can be improved by bringing in individual perceptions. In other words, to some extent, social structure in the eye of the beholder. Example of the internal market for reputation in organizations is given. Were individual perceptions more important than an objectively measured social structure in determining these reputations? (the answer is yes)

**Theoretical framework:**

- balance theory says that someone who is perceived as a friend of a person with a good reputation will also have a good reputation.
- The evaluation of reputation in a company (i.e. the “pricing” of the individual) is difficult, so people look for signals, especially ties to people with good reputations
- Hypothesis 1 is that an observer’s perception of an individual’s performance will be significantly influenced by the degree to which the observer perceives that individual to have a prominent friend.
  - We assume people within the company are jockeying for high reputation by publicizing their links to prominent others.
    - perceptual measures will be more important in this process than objective measures.
    - Each member has a “cognitive map” of the social structure based on his/her perceptions.
- Hypothesis 2 is that perceptual measures will be more important in this process than objective measures.

**Methods:**

- Site
  - small entrepreneurial firm on 1 floor where all employees saw each other regularly.
  - 28 men and 8 women
  - questionnaire completed by 92% of employees
- Measures
  - Network indexes: friendship and advice networks
    - asked each person about the perceptions of every other persons’ friendship and advice networks
    - Every person in the firm was listed as choices
    - Actual network structure was found by only using ties that two people mutually self reported (for friendship) For advice, mutuality wasn’t required.

Terence Meehan & Ning Sa
Week 4. Data Collection & Social Cognition

- Independent variable: friend’s prominence matrix
  - focus on each person’s most prominent friend
  - measured each friend’s prominence in four ways
    - perceived vs. actual network measures
    - network measure of prominence vs. organizational chart

- Dependent variable: performance reputation matrix
  - Each person asked to rate every other person on 7-point performance scale

- 1st control variable: job performance matrix
  - Managers evaluated subordinates on 7-point scale

- 2nd control variable: formal status matrix
  - controlled for member being and owner-manager, a manager or a non-manager.

Analysis and results:

- used Multiple Regression Quadratic Assignment Procedure
- performance reputation was significantly correlated with all 4 independent variables
- when including control variables, being perceived to have a prominent friend helped boost an individual’s reputation as a high performer, but actually having such a friend had no significant effect. Actually being a high performer (as judged by supervisor) also helped boost an individual’s reputation.
- Hypotheses 1 and 2 are supported.