Wasserman & Faust,  
Chapter 2 Social Network Data: Collection and Analysis

2.1 Introduction: What are Network Data?
- Social network data measures at least one structural variable in a set of actors
- Concerns and theories focus on identifying structural variables and measurement techniques

2.1.1 Structural and Composition Variables
- **Structural Variables** – variables measured on pairs of actors, cornerstone of social network data sets (ex. transactions among corporations, friendships between people, trade between nations)
- **Composition Variables** – *actor attribute* variables; measurements of actor attributes that are of the standard social and behavioral science variety, and defined at the level of the individual (ex. gender, ethnicity for people, geographical location)

2.1.2 Modes
- Mode – distinct set of entities on which the structural variables are measured: one-mode, two-mode, etc.

2.1.3 Affiliation Variables
- Affiliation variables – variables that are part of affiliation networks
  - Affiliation networks – special social networks that arise in two-mode networks when there are two modes, one of them an event (ex. clubs or volunteer organizations), one set of actors
- Example: Considering a set of actors, and three elite clubs in some city, we define an affiliation variable for each of these three clubs. Each of these variables gives us a subset of actors who belong to one of the clubs.

2.2 Boundary Specification and Sampling

2.2.1 What is your population?
- **Boundary** – allowing a researcher to describe and identify the populations of a study
  - Defined based on frequency of interaction and intensity of ties among members as contrasted with non-members
- **Approaches to boundary specification** - while the *realist approach* defines boundaries as actors in the data set perceive themselves, the *nominalist approach* defines boundaries through the theoretical research concerns
- A set of actors consist of all social units to which there are measurements (either structural variables, or structural and compositional variables)
- Small populations have clearly defined actor set boundaries (ex. classrooms, offices, social club, and villages); large populations have less well-defined boundaries (ex. interorganizational networks in a community)
Snowball sampling and random nets – special sampling techniques when the boundary is unknown

2.3 Types of Networks

- **Number of modes** refers to the number of distinct kinds of social entities in the network
- Networks categorized by how many modes the network has, and by whether affiliational variables are present

2.3.1 One-Mode Networks

**Actors.**
- Actors may be people, subgroups, organizations, collectives/aggregates (communities, nation-states)
  - Subgroups usually consist of people
  - Collectives/aggregates usually consist of organizations and subgroups

**Relations.**
- Relations are usually viewed as representing specific connections, or “relational contents”
- The kinds of relations may be:
  - Individual evaluations: friendship, liking, respect, etc
  - Transactions or transfer of material resources: lending/borrowing, buying/selling
  - Transfer of non-material resources: communications, sending/receiving information
  - Interactions
  - Movement: physical (from place-to-place), social (between occupations or statuses)
  - Formal roles
  - Kinship: marriage, descent

**Actor attributes.**
- The characteristics of the actors constituting the network can be measured

2.3.2 Two-Mode Networks

**Two Sets of Actors**
- Actors can be of the general types as those in one-mode networks
- Relations are measured in at least one way between actors in the two sets
  - Ex. Collection of corporate headquarters and non-profit groups in the Minneapolis/St. Paul measuring flow of donations from corporations to non-profit groups (unidirectional flow)

**One Set of Actors – One Set of Events**
- **Affiliation Networks** (or Membership Networks), arise when one set of actors is measured with respect to attendance at, or affiliation with, a set of events or activities; the first mode in an affiliation network is a set of actors, and the second is a set of events which affiliates the actors
- **Actors** types can be exactly the same as those in one-mode and two-mode networks
Actors must be affiliated with at least one event

- **Events** are defined on the basis of membership, attendance, or socializing in a group, etc.
  - Nature of events depend on the types of actors involved
- **Attributes.** Actor attribute variables are of the same types as those for one-mode and two-mode networks
  - Two set of attribute variables can be found in actors and events

### 2.3.3 Ego-centered and Special Dyadic Networks

- **Special Dyadic** – non-network relational data sampled from a larger population centering on the interaction between pairs; ex. husband-wife, father-son
  - An actor may relate to a limited number of “special” other actors; this design can constrain interactions among actors so that all people cannot interact with all others
- **Ego-centered Network** consists of a focal actor, *ego*, as sets of alters with ties to the ego and measurements on the ties
  - Ex. Each respondent reports on a set of alters to whom they are tied, and on the ties among these alters (*Personal network data*)

### 2.4 Network Data, Measurement, and Collection

#### 2.4.1 Measurement

- Social network data is different from standard social and behavioral sciences in that its data consist of at least one relation measured among a set of actors
- Presence of relations has implications among many measurements such as the unit of observations (actor, pair of actors, relational tie, or event), modeling unit (the actor, dyad, triad, subset of actors, or network), and the quantification of relations (directional vs. nondirectional)
  - **Modeling unit** – level of network analysis being studied

**Unit of observation** is the entity on which measurements are taken and an actor from whom information about ties is elicited.

**Modeling Unit**

- Levels at which network data can be modeled or summarized are Actor, Dyad, Triad, Subgroup, Set of actors or network

**Rational quantification** refers to measurements and whether the relation is directional or nondirectional, and whether it is dichotomous or valued
  - Directional – relational tie has an origin and destination
  - Nondirectional – relation has no direction,
  - Dichotomous – relation is coded as either present or absent
  - Valued – relation has values such as strength, intensity, or frequency

#### 2.4.2 Collection

- Social network data can be collected through the following techniques: Questionnaires, Interviews, Observations, Archival records, Experiments, and Other techniques such as ego-centered, small world, and diaries/journals
**Questionnaire**

- Most commonly used collection method
- Usually contains questions used to identify relations between actors
  - Three types of question formats: Roster vs. free recall, Free vs. fixed choice, and Ratings vs. complete rankings
- **Roster vs. Free Recall** – issue of whether questionnaires should be presented with a complete list, roster, or respondents be allowed to generate lists of names, free recall
  - Roster can only be used when researcher knows entire members of the set prior to data gathering (Ex. Friends in a class)
  - *Free recall* (a format where respondents generate the lists of names) can be used when entire members of the set may not be presented (Ex. Actors are asked to name other actors, but were not presented with a roster in studies such as Friends in two junior high schools or community elites)
- **Free vs. Fixed Choice**
  - On how many nominations respondents can provide,
    - *Free choice* – actors are not given constraints
    - *Fixed choice* – actors are given constraints
- **Ratings vs. Complete Ranking** – used to reflect intensity of strength of ties
  - Ratings require respondents to assign a value or rating to each tie
  - Complete rankings require respondents to rank their ties to all other actors
  - Full rank-orders and rating scales with multiple responses generate valued relations; *dichotomous* and *directional*

**Interview**

- Used to gather network data when questionnaires are not feasible (Ex. CEO interviews in Minneapolis/St. Paul)

**Observation**

- Used to gather network data in field research, usually relations among relatively small groups of people who are engaged in face-to-face interactions
- Useful with people who are not able to respond to questionnaires or interviews
- Useful for collecting affiliation data for attendance at events

**Archival records**

- Measures ties through examining measurements from past recorded interactions
  - Ex. Patterns of citations among scholars examining “who cites whom” to study diffusion of a scientific innovation

**Other**

- *Cognitive social structure* is a design where respondents are asked about perceptions of network ties and perceived relations are measured (Ex. Fast food restaurant perceptions)
  - This design can collect more information than general sociometric designs as the respondent reports not only on their own ties, but ties belonging to other actors
- **Experimental**
  - Method 1 – select a group of actors, observe their interactions in an
experimentally controlled situations, then record interactions between pairs of actors

- Method 2 – select a group of actors, specify which actors can interact with each other during the experiment, then record interactions between only those specified pairs of actors (Ex. Group problem-solving experiments)

- *Ego-centered* – respondent is set up as ego with data measured among the ties from the ego to the alters (Ex. Survey about the people with whom you discussed matters important to you)

- *Small world* is an attempt to determine how many actors a respondent is removed from a target individual based on acquaintanceship
  - Can be used to compare demographic characteristics and chains
  - Reverse small world focuses on ties from a specific respondent to a variety of hypothetical targets

- *Diary* – respondents are asked to keep a daily record of whom they interact
  - Variance of ego-centered
  - Data sets include information on the relation type and characteristics of the alters

### 2.4.3 Longitudinal Data Collection

- Focuses on how ties in a network change over time and how well the past can predict the future using methods previously discussed (questionnaires, interview, observation, etc.)
  - Commonly used to examine friendships over time
  - Ex. Interaction among fraternity members over time (P.55)

### 2.4.4 Measurement Validity, Reliability, Accuracy, Error

- *True structure* – social structure referring to a relatively prolonged and stable pattern of interpersonal relations

- *Observed structure* – measured network data that might contain error

**“Accuracy”**

- Issue of *informant accuracy* – information collected using verbal reports and information collected through observation
  - People are not good at accurately reporting on their interactions in particular situations

- “True” structures are of most interest and network studies should study focusing on long-term patterns, not particular interactions of individuals
  - Issue comes up when looking at interactions among organizations being reported on by members with imperfect information about the organization

**Validity**

- A concept of a measure is *valid* to the extent that it measures what it is intended to measure

- *Construct validity*, a more formal construct, arises when measures of a concept behave as expected in theoretical predictions and can be studied by examining how these measures behave in a range of theoretical propositions
Reliability

- A measure of a variable or concept is *reliable* if repeated measurements give the same estimates of the variable
  - Three approaches have been used to assess the reliability of social network data: test-retest comparison, comparison of alternative question formats, and reciprocity of sociometric choices
- “True” value of a variable must be assumed to not change over time for test-retest comparison to be appropriate
- Reliability can be assessed at different levels of analysis
- Sociometric questions using ratings or full rank orders are more reliable than fixed choice designs
- Sociometric questions about more intense or intimate relations have higher rates of reciprocation than sociometric questions about less intense or intimate relations (Marsden, et al.)
- Reliability of aggregate measures (ex. popularity) is higher than the reliability of “choices” made by individual actors (Burt, et al.)

Measurement error

- Occurs when there is a discrepancy between the “true” score or value of a concept and the observed (measured) value of that concept
- *Measurement error* – the difference between the true and observed values
- Levels of analysis must be kept in mind when determining the implications of measurement error
The Problem of Informant Accuracy: The Validity of Retrospective Data

Bernard, et al.

1. Introduction
   - Focusing on “fugitive problem” of informant accuracy in reporting past events, behavior, and circumstances
   - A theory or prediction would provide with a basis for assessing the degree of validity of informant

2. Literature Review on Informant Accuracy
   - Three substantive areas where informant accuracy has been studied: Recall of child care behavior, health seeking behavior, and communication and social interactions

   **Child Care**
   - Accuracy of mothers’ recall of caring for their children was tested
   - Recall was inaccurate by about a third to three quarters, usually as a result of underreporting
   - ‘Trained field observers’ to judge levels of disagreement between respondents and professionals were used and compared with the reports of the children themselves: greater disagreement between them

   **Health Care**
   - Comparison of data obtained from interviews with one from objective records on pediatric history and other kinds of health behavior
   - Underreporting found in interviews, increased as time passed
     - “The best documented phenomenon of underreporting of health events as well as of a wide variety of other types of events and behaviors, is the decrease in the reporting of events as time elapses” (p 498)

   **Communications and Social Interactions**
   - Tested recall of social network or communication contacts through series of seven experiments by Bernard and company
     - Concluded that “what people say about their communication bears no useful resemblance to their behavior”; individual differences in accuracy could not be accounted for by actor attributes; “the error is so great that statistical and numerical techniques for washing data collected by recall instruments cannot solve the problem” (p 499)
   - Study of telephone habits by Hyett
     - People who made very few calls tended to overreport, those who made a great many calls tended to underreport

   **Some Isolated Studies**
   - Young & Young
     - For publicly available information there was both agreement and accuracy
For questions where the answers were not public there was very little agreement and high inaccuracy

**Summary of the Literature**

- On average, about half of what informants report is probably incorrect in some way
- Social scientists are accustomed to low correlations between variables representing different concepts
  - Weak relationship between two distinct concepts is to be expected
  - Weak relationship between a concept and the accurate measurement of the concept is unacceptable

3. **Conceptual Variables: A Way Out?**

- In the absence of either a well-defined theory or a clearly described dependent variable it becomes meaningless to discuss the validity or accuracy of such conceptual variables
  - A well-defined theory is needed to seriously discuss validity and separate one’s assessment of the validity of the operational measure from the more general question of the validity of the theory itself
- Given that people are telling the truth and practicing deceit in responses, it is hard to imagine how to test the validity of attitudinal or other internal-state responses against anything

4. **What Can Be Done?**

- Sudman & Bradburn
  - “Memory effects in surveys can be described by a function that is the product of effects due to omissions and telescoping”
  - Events in the past are likely to be recalled as being more recent than they actually are
- Models to compare various data collection techniques:
  - “aided recall” – presented with fixed alternatives
  - “record assisted recall” – keep an accurate daily diary
  - “bounded recall” – interviewed to set a baseline at least once
- D’Andrade’s and Shweder’s work
  - Idea that memory is subject to systematic distortion due to cultural training
  - Reporting cultural norms rather than dredging up actual events, circumstances, behaviors, or personality traits
- Accepted fact that informants are inaccurate as memory decays exponentially over time
- It is worthwhile to account for variation of accuracy both between and within instruments
- “After making empirical and/or theoretical corrections”, it is important to know not only which informants are likely to be at least accurate is important but also whether any informant is sufficiently accurate to rely on
The Social Structural Basis of the Organization of Persons in Memory
Devon D. Brewer

- This study describes the similarities in cognitive structures that individuals within different organizations have when thinking about their organization’s networks
  - Three communities studied: a graduate academic program, a Taiwanese-American religious fellowship, and a university’s public affairs department
  - All members of these networks were asked to list from memory every member of their network
  - Regardless of the type of organization, the respondents recalled members of the network as they adhered to perceived structure

- The method used by psychologists is to study semantic cognitive structure, learning how words are related to each other in the mind as they are freely recalled by respondents and understand how people are linked in the mind

- Three patterns of free recall that have been identified:
  - Association - adjacently recalled persons are connected by some type of relationship, thus reflecting a cognitive structure within the recaller’s mind
  - Serial order - the order in which people are recalled reflects their salience in the mind, the first person thought of likely has much more meaning than the last
  - Frequency - how often a person is recalled indicates that person’s salience within the recaller’s mind

Study 1: Graduate Academic Program

- 15 graduate students were asked to list all the graduate students in the program
- People within the same cohort (subject year) were recalled adjacently much more often and with much less time between responses than with people who were from different cohorts
  - Figure 1 shows how closely students were listed when the respondents freely recalled them in a list
  - The number indicates the recalled persons’ year in the program
  - The closer two persons are, the more they were recalled consecutively
Respondent would recall students in closer cohorts than more distant ones

While this study showed that the cohort structure was critical for how the network was organized in memory, it could not show which aspect of the cohort structure was most important; the basic organization of a university setting, or the tendency for social interaction to stay within a cohort.

**Study 2: Taiwanese-American Religious Fellowship**

- 25 people within a religious fellowship were asked to recall all persons within the entire fellowship.
- People recalled adjacently were much more likely to be socially close than chance would predict.
  - The adjacency of recalled names could also not be explained by other social structures, such as kinship or gender.
  - Two respondents attempted to list everyone in alphabetical order, yet their response pattern clearly showed a bias for social proximity
- Figure 2 below shows one subject’s recall demonstrating the effect of social proximity on associative patterning
  - Adjacently recalled pairs of persons are joined quite high
  - Subsequences of recalled persons correspond to the overall social proximity group structure
  - Persons listed together were perceived to interact with one another a lot
Response time was significant when the respondents were listing the members of the organizations.
  
  - When two adjacently recalled people were highly socially connected, the respondent took very little time between recalling them.
  - As social proximity decreases, inter response time rises (Figure 3).

The order that persons were recalled by respondents revealed social network influences.
  
  - Persons socially closer to the respondent were recalled earlier, as were people who are highly visible within the organization.
Study 3: Department in a Formal Organization

- 13 employees of a public affairs department were asked to recall the names of their coworkers.
- Work proximity was the defining structure of recall, and those people who worked closely together were often recalled adjacently.
  - Factors such as status within the organization, spatial locations within the office, and work sections within the department could not account for the clustering of recalled people.
  - Figure 4 shows how recalled persons are clustered and the structure between work proximity and adjacency in recall.

![Figure 4](image)

*Figure 4. Department of a formal organization: two dimensional principal components analysis representation of persons in terms of work proximity (circles) and adjacency in recall (triangles) from aggregated data (p. 387)*

- The circles represent persons in terms of work proximity between people.
- The linked triangles are those same people in terms of adjacency in recall as they were recalled by respondents, the closer the triangles are, the closer they were listed to each other by the respondents.
- Lines connect each person’s position for work proximity to one for adjacency; short lines illustrate similarity between work proximity and adjacency in recall.

- Figure 5 shows one subject’s recall illustrating the effect of work proximity on associative patterning (as described in the earlier study for the religious fellowship).
  - Adjacently recalled pairs of persons are joined quite high.
  - Subsequences of recalled persons correspond to the overall work proximity group structure.
  - Persons together in recall work with each other closely.
- Response time between adjacently recalled people relied on their social proximity, and the order that people were recalled in was based on proximity to the respondent as well as visibility (persons of higher status).
Conclusions:

- In terms of associative pattern, the fundamental cognitive structure of persons in a community is from on the community’s social structure, rather than any other type of structure such as kinship, physical closeness.

- In terms of frequency and serial order pattern, there are only a few variables that dictate the ‘salience’ of person (how early on they are recalled) which include their social proximity to the respondent, their status within the social network, or their visibility within the network.

- The pattern of recall was the same among all respondents from all three different types of organizations, and across time (the third study was repeated after 3 weeks with the same result).

- Regarding why this cognitive structure dominates the way we think about networks, there are two (non-competing) possibilities:
  - Interactions among people are observed and those associations are committed to memory, thus making them conveniently easy to recall in this type of study.
  - In evolutionary terms, accurate knowledge of social networks is so important and advantageous that humans think about each other primarily through the lens of social structure, it is ingrained in us to think this way.
COGNITIVE STRUCTURE AND INFORMANT ACCURACY
Linton Freeman, A. Kimball Romney, Sue C. Freeman

Main Points to Consider

- Principles of Memory Organization taken from cognitive psychology can help to explain informant error and to predict the types of informant error that may occur during data collection.
- The “best” informants (regular attendees at a series of events, for instance) are best used to reveal long-range stable patterns of events.
- The “worst” informants (irregular attendees at the same series of events, for instance) are best used to reveal details of a particular event.

The authors cite Bernard, Kilworth and Sailor (BKS) in stating that about half of what informants recall about a set of social interactions may be wrong in two ways: (1) people forget about particular interactions and (2) people recall interactions that never really occurred. But the focus of most social science is on long term, repeating patterns of behavior. By the end of the article, the authors assert that the error contained in responses can be revealed and understood so that it is not such a significant problem.

Through studies of subjects who attended a series of colloquia, two of the authors (Freeman and Romney) found in a previous study that participants’ recall of one session was full of errors. The errors, however, were biased toward revealing the long-range pattern of attendance: when questioned about the attendance at a particular “target” session, informants forgot to recall those who attended the colloquia infrequently, and tended to falsely recall those who attended frequently. It turns out that the responses containing errors were a better indicator of the long term pattern of attendance at the colloquia than the actual attendance at the particular target session.

The study described in this article is a replication of the BKS study. It focuses on two questions:
1. how can the preliminary study (by BKS) be expanded to better measure long-term patterns, and
2. how can we better explain informants’ bias toward long-term patterns?

So, the study unfolds thus: data is collected on attendance at a series of colloquia over an entire school term. This data addresses more adequately the precision of the long term pattern of attendance. Hypotheses are formulated after a consideration of some important principles from cognitive psychology (described later) and the analysis tests these hypotheses.
THE DATA
Attendance records were kept on a series on 9 colloquia throughout the semester. 2 observers recorded attendance, seating arrangements and arrival and departure times. These data generated two data sets: The social structure of the group was constructed from the total attendance data over 9 sessions, and the 9th (final) attendance data was designated as the target session on which respondents would be questioned. Interviews were conducted by computer, allowing the researchers to collect keystroke data (indicating the order of the recall and the time required by the informant to recall the information).

Attendance at the sessions is straight forward (no subgroups, cliques or clusters) and differentiated only by individual attendance rates. The results of this study closely matched those of the BKS study in terms of recall error: about half of the recall data contained error; mostly that informants forgot to name a co-attendee (115 of 272 opportunities for correct responses). Less often (26/272), informants named a co-attendee who was not present at the target session. These data also indicate a bias toward long-term attendance patterns: the correlation between attendance frequency and the number of times named by an informant was .85. These findings corroborate the findings of both previous studies (BKS and by Freeman and Romney).

IMPORTANT PRINCIPLES FROM COGNITIVE PSYCHOLOGY USED FOR HYPOTHESIS FORMULATION
- Human memory is organized.
- The organization embodied in a mental structure is revealed in free recall.
- The organization of memory is based on experience.
- The tendency of a person to recall an element that occurred in an event depends on two factors: (a) the amount of elaboration of the person’s mental structure, and (b) the degree to which the element is typical in the kind of events being examined.
- The tendency of a person to falsely recall and element that did not actually occur in an event depends on two factors: (a) the amount of elaboration of the person’s mental structure, and (b) the degree to which the element is typical in the kind of events being examined.

IMPLICATIONS FOR ACCURACY
The researchers formulated an indicator for experience with the colloquia based on attendees’ social standing (capital) in the academic program and their office space assignments. They then generated a series of assumptions about the extent of experience and how these participants would recall attendees and subsequently generate error. Those with high levels of experience (in-group) should: be forgotten be others less, forget others less often, and generate more false recalls. Those with less experience (out-group) should: be forgotten more often and generate more “forgets” in the data. Finally, experienced attendees should display better organization of cognitive structures than less-experienced attendees.
The results of the experiment are consistent with hypotheses drawn from cognitive psychology literature: “the notions of cognitive structuring provide a key to understanding the kinds of biases that intrude when people are asked to recall the details of a particular social event”. In-group members provide information about a particular event the error of which is biased toward long-term patterns. Out-group members provide more accurate information about an event, but there is a caveat.

The hypothesis was validated further by examining the responses of the most experienced of the in-group members and the responses of the least experienced out-group members. Up to a certain point, the in-group members provided information about the long-term attendance patterns with .85 accuracy. As more responses were examined, the accuracy began to decrease. Up to a certain point, the out-group member provided information about the attendance at the particular event that contained only 5% errors. Again, as more data from more informants is added, the accuracy decreases. Unfortunately, there is no rule about when to stop adding data. In a situation where the actual attendance at an event were unknown, this would render this strategy almost useless.
THE BOUNDARY SPECIFICATION PROBLEM IN NETWORK ANALYSIS
Edward O. Laumann, Peter V. Marsden, and David Prensky

Main Points to Consider

• Boundaries may be defined through rules of inclusion: for the selection of actors to be included and to the choice of types of relations among those actors to be studied.
• The stakes are high: errors in defining system boundaries can result in gross misrepresentations of the network/system under study.
• There are no hard and fast rules – only that boundary specification needs careful thought and attention in network analyses.
• The article does not focus on egocentric approaches to network generation, rather on sociocentric or structural approaches.

APPROACHES TO BOUNDARY DEFINITION
Two basic approaches:
1. **Realist**: The network exists as an artifact of some subjective meaningfulness (conscious experience) of the actors within it. Boundaries are defined by rules or definition.
2. **Nominalist**: inclusion in the network is determined and imposed by the analyst. The actors may not realize the reality of the analyst’s assumptions, and awareness of these boundaries becomes an empirical question.

DEFINITIONAL FOCI FOR THE INCLUSION OF ACTORS
Inclusion of actors is defined by one of three criteria:
• Actors
  o Most commonly used definition
  o Focuses on attributes of actors
  o Two common tactics:
    ▪ Positional Approach: refers to the presence or absence of some attribute, commonly a position in a formal group
    ▪ Reputational Approach: uses the judgments of knowledgeable actors to determine inclusion.
  o Relations and participation in events are left free to vary.
• Relations
  o Includes actors participating in a relationship of some specified type.
  o Includes “Snowball Sampling”: the network connections of a small set of actors is traced until some termination criteria are met.
  o The design of the study constrains relational features, but makes attributes or participation in events empirical questions.
Events
  - Focuses on a behavioral definition for inclusion in some group
  - Eaves attributes and relations of those participating in an event or activity as empirical questions
  - A combination of two of the above definitional approaches.

Cross-tabulating the boundary definition approaches with the definitions for inclusion of actors results in 8 boundary specification strategies.

<table>
<thead>
<tr>
<th>Attributes of Nodes</th>
<th>Relations</th>
<th>Participation in Event</th>
<th>Multiple Foci</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Realist</strong></td>
<td>Study of small, tightly bounded groups</td>
<td>Utilized in very few studies. “Primary group” concept used to study and understand network behavior, not as method for network definition.</td>
<td>Primary alternative to Realist/Attributes approach. Participation in series of events reveals cliques; influential actors; nodal elements.</td>
</tr>
<tr>
<td><strong>Nominalist</strong></td>
<td>Applied to expansive groups (100's or 1000's). Actors are often organizations.</td>
<td>Studies of the “small world” problem. Relations are traced from starters to targets.</td>
<td>Used infrequently. Invisible college study cites interest of scientists in a research area irrespective of disciplinary focus.</td>
</tr>
</tbody>
</table>

Some combinations can cross over realist/nominalist definitions. Kadushin included actors who shared some political or social interest along with only indirect relational connectedness.

Combining definition strategies uses theoretical degrees of freedom (leaves fewer empirical questions). If we choose actors based on relations, then we aren’t interested in knowing that the network is “dense” or “well-connected”.

**ON INCLUSION RULES FOR RELATIONS**
No common definition (typology) for social relations makes constructing rules about inclusion tricky. Even if we had good definitions, selecting the types to be analyzed is still left to the researcher. The literature provides evidence that little objective thought is given to this issue; rather appeals for common sense are made. This may lead to...

**Partial System Fallacy** “This is present wherever a set of relationships connecting a subset of the actors to which the relations are relevant is analyzed without prior attention to the entire set of actors.” For example, consider a study of financial and information exchanges among a set of agencies. One might impose some restriction on the geographical relationship of actors for network inclusion at the expense of some other important relation, like the transfer of funds or transfer of information. The 'partial' system that results would
be a gross misrepresentation of the actual network in terms of transfers of finances or information.

**BOUNDARY SPECIFICATION FOR ACTIVITIES**

Be careful to explain the theoretical underpinnings for the choice of event as the subject for network inclusion.

**CONCLUSION**

There is no sense in which social networks must “naturally” correspond to social systems. Thought and attention must be lent to the definitions by which actors and their relations are included in social network models.
Main Points to Take Home

- People in an organization attempt to perceive social networks.
- Perceptions, often with judgment attached, influence decisions and behavior of others.
- These links between individualist and structural approaches are important in network analysis.
- “...the individual must be brought back in to acknowledge and account for the micro-foundations of structural research.”¹ In fact, much of structural theory relies on social psychological underpinnings.

The Theory

Balance Theory (referring to reputation in an organization): the friend of a positively valued other is also positively perceived. This represents a strain towards cognitive balance. Knowing this, people often choose to make these positively perceived connections known to others (basking in reflected glory). Previous work has not studied this effect in the context of performance.

A market for reputation: the more positive one’s reputation, the more valuable that person is within this organizational market. Pricing is then a cognitive process involving status (title) and social ties as indicators of high quality. So,

Hypothesis 1: The prominence of an individual’s most prominent friend will influence the individual’s performance reputation in an organization, controlling for the individual’s formal status and job performance.

Important: perceptions about relations are more important in forming opinions than objective (actual) network positions. In fact, individuals construct ‘mental maps’ of friendship networks and use them to navigate social relations.

Hypothesis 2: Measures of perceived network relations will lead to better predictions of performance reputation than will measures of actual network relations.

¹ Much of the literature published prior to this article about structural analysis had rejected the premise that individual psychology played an important role in structural phenomena.
**METHODS**

**Site**
A small Silicon Valley tech company. The company is wholly owned by its three top managers, each holding equal shares. Respondents were given $3 to complete a lengthy survey.

**Measures**
Each person was asked to map their personal friendships and to map the personal friendships of every other employee in the firm. The researchers constructed the actual social system by only including friendship links that were verified by both parties. Figure 1 shows an example of a respondent's perceived friendship map, while Figure 2 shows the actual social structure of the firm. These figures illustrate how discrepant from reality a person's perceptions can be.

![Figure 1: An Employee's Cognitive Map of the Friendship Relations at Silicon Systems](image)

*The sociogram was drawn by KnackPlot (Krackhardt, Lundberg, & O’Rourke, 1993)*.

Advice networks were constructed using row dominance, that is, advice connections were recorded if a respondent reported that the link existed regardless of the other's reciprocity. Each respondent also filled out perceived advice maps for each other employee of the firm.

**Prominence matrices**
Prominence was determined in four ways, shown in Table 1. It is important in supporting the theory being tested that measures of prominence be directly observable. So, which actors have a high indegree centrality relative to advice seekers? The first measure of the independent variable was to identify, for each actor, the friend with the highest indegree centrality of advice seekers.
The second measure of the independent variables was the actual indegree centrality of the friend's advice and friendship networks. The third and fourth measures of the independent variable came from formal status indicators existent in the company's organizational chart.
The highest status of a friend was recorded from both the perceived and the real friendship networks as the third and fourth independent variables.

**Dependent variable: Performance reputations:** Each respondent provided information about the perceived performance of each other employee.

**First control variable: Supervisor's performance rating:** Collected from company records.

**Second Control Variable: Formal Status Matrix:** Little formal hierarchy, so this status was supplied by the individual's title within the organization.

**ANALYSIS**

OLS cannot be used here for danger of bias; Multiple Regression Quadratic Assignment Procedure (MRQAP) was used: first, OLS estimates of the parameters are calculated. Then the rows and columns of the dependent variable are permuted to give a new, mixed up matrix. OLS is performed on the new matrix, $R^2$'s and betas are stored. "The matrix is permuted again, and the procedure is repeated an arbitrarily large number of times. The distribution of the stored betas and $R^2$'s for each of the independent variables under each of the permuted regressions becomes the reference distribution against which the observed original values are compared. If fewer than 5% of the betas derived from the permuted regressions are larger than the observed beta, the beta is considered significant at the .05 level. This method is robust against any existing correlation in the rows and columns of the dyadic data because this test is a conditional nonparametric test.

**RESULTS**

**TABLE 2**  
Means, Standard Deviations, and Correlations$^a$

<table>
<thead>
<tr>
<th>Variables</th>
<th>Means</th>
<th>s.d.</th>
<th>1</th>
<th>2a</th>
<th>2b</th>
<th>2c</th>
<th>2d</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Performance reputation</td>
<td>4.93</td>
<td>1.39</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Friend's prominence</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a. Perceived friend's indegree centrality</td>
<td>6.70</td>
<td>8.86</td>
<td>.23</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b. Actual friend's indegree centrality</td>
<td>7.87</td>
<td>7.00</td>
<td>.26</td>
<td>.22</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c. Perceived friend's status</td>
<td>1.40</td>
<td>0.73</td>
<td>.28</td>
<td>.66</td>
<td>.30</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>d. Actual friend's status</td>
<td>1.55</td>
<td>0.83</td>
<td>.28</td>
<td>.15*</td>
<td>.83</td>
<td>.30</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Job performance</td>
<td>4.91</td>
<td>1.15</td>
<td>.33</td>
<td>.14*</td>
<td>.31*</td>
<td>.28</td>
<td>.49</td>
<td></td>
</tr>
<tr>
<td>4. Formal status</td>
<td>1.31</td>
<td>0.62</td>
<td>.30</td>
<td>.17</td>
<td>.54</td>
<td>.38</td>
<td>.65</td>
<td>.47</td>
</tr>
</tbody>
</table>

$^a$ All correlations are significant at $p < .01$, except for those with an asterisk.

Table 2 shows that the dependent variable, performance reputation, was highly correlated with all four measurements of the independent variable. The researchers performed simple regressions to determine if significant correlations remain significant when controlling for other variables.
Friendship with prominent others does boost performance reputation, but the effect depends on how friendship links were assessed. Models 2, 3, and 4 together show that only perceived measure of friendship has an effect on performance reputation. Models 5, 6 and 7 use indegree centrality as the measure of actual and perceived friendship and the results repeat the pattern of models 2, 3, and 4. Actually having friends in high places does little to boost your performance reputation, the relationships have to be perceived by others.

TABLE 3
Results of Multiple Regression Analysis*

<table>
<thead>
<tr>
<th>Variables</th>
<th>Status Models</th>
<th>Centrality Models</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Friend's prominence</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Perceived friend's</td>
<td></td>
<td></td>
</tr>
<tr>
<td>indegree centrality</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Actual friend's</td>
<td></td>
<td></td>
</tr>
<tr>
<td>indegree centrality</td>
<td>.028**</td>
<td>.025**</td>
</tr>
<tr>
<td>Perceived friend's</td>
<td></td>
<td></td>
</tr>
<tr>
<td>status</td>
<td>.024</td>
<td>.010</td>
</tr>
<tr>
<td>Actual friend's</td>
<td></td>
<td></td>
</tr>
<tr>
<td>status</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Job performance</td>
<td>.296**</td>
<td>.277*</td>
</tr>
<tr>
<td>Formal status</td>
<td>.407</td>
<td>.326</td>
</tr>
<tr>
<td>Intercept</td>
<td>2.934**</td>
<td>2.961**</td>
</tr>
<tr>
<td>R²</td>
<td>.136</td>
<td>.139</td>
</tr>
</tbody>
</table>

* Beta coefficients are unstandardized. Their significance was determined by the Multiple Regression Quadratic Assignment Procedure (MRQAP, Krackhardt, 1993).
* p < .05, one-tailed test
** p < .01, one-tailed test