Social and Information Networks

Module 1 - Introduction

Reference Book:

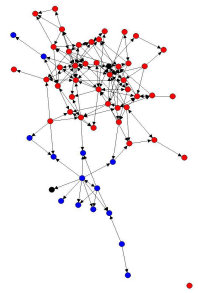
Wasserman Stanley, and Katherine Faust. (2009). Social Network Analysis: Methods and Applications, Structural Analysis in the Social Sciences.

Introduction

- Social Media
- Social Network

Social Network

- A social network is a **social structure made up of a set of social actors** (such as individuals or organizations), sets of dyadic ties, and other social interactions between actors.
- An articulation of social relation
- A set of actors (nodes, points, vertices)
  - Individuals (e.g., persons, animals)
  - Collectivities (e.g., groups, firms, industries, nations, web pages)
- A set of ties (links, lines, edges, arcs) of a given type that connect pairs of actors
- Set of ties of a given type constitutes a social relation
- Different relations (e.g., advice, friendship) have different structures & consequences



Social Network (Cont'd)

- Ties / edges / arcs / lines / links
  - connect pair of actors
  - types of social relations
    - friendship
    - acquaintance
    - kinship
    - advice
    - hindrance
  - allow different kind of flows
    - messages
    - money
    - disease

Social Network(cont'd)

Example Relations **among Persons:**

- Kinship
  - mother of, wife of
- Other role-based
  - boss of, teacher of
  - friend of
- Cognitive/perceptual
  - knows
  - aware of what they know
- Affective
  - likes, trusts
- Interactions
  - gives advice, talks to, fights with, has coffee with
- Affiliations
  - belong to same clubs
  - is physically near

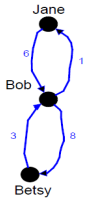
Social Network(cont'd)

Example Relations **among Organizations:**

- As corporate entities
  - Buy from / sell to, leases to, outsources to
  - Owns shares of, subsidiary of
  - Joint ventures, cooperate sales agreements, alliances
  - Regulates
- Through their members
  - Personnel flows
  - Personal friendships
  - Co-memberships

Social Network(cont'd)

- We can attach values to ties, representing quantitative attributes
  - **Strength** of relationship
  - **Information capacity** of tie
  - **Rate of flow** or traffic across tie
  - **Distances** between nodes
  - **Probabilities** of passing on information
  - **Frequency** of interaction



Social Network(cont'd)

- Actors and their actions are viewed as interdependent rather than independent, autonomous units.
- Relational ties (linkages) between actors are channels for transfer or "flow" of resources (either material or nonmaterial).

Social Network Analysis (SNA)

- SNA is a technique which studies
  - Relationships between people and groups
  - How those relationships arise
  - Consequences of the relationships
- A method to analyze the *connections* across individuals or groups or institutions.
- It allows us to examine how actors or institutions are *interrelated*.
- Helps us understand the linkages among social entities and the implications of these linkages.

Social Network Analysis

- Advantage of social network analysis

  - It focuses on interaction (rather than on individual behavior [EgoCentric Analysis]) among large groups.

  - Allows us to examine how the configuration of networks influences individuals , groups, organizations or systems function.

History of Social Network Analysis

- In the 19th century, Durkheim wrote on "social facts"—phenomena created by the interactions of individuals -> a reality that is independent of any individual actor.

  → Sociology is the study of social facts

History of Social Network Analysis

- At the turn of the 20th century, Simmel was one of the first scholars to think in relatively explicit social network terms. He examined how third parties could affect the relationship between two individuals—and he examined how organizational structures or bureaucracies were needed to coordinate interactions in large groups.

History

- One of the first examples of empirical network research can be found in 1922, in Almack's "The Influence of Intelligence on the Selection of Associates." Almack asked children in a California elementary school to identify the classmates with whom they wanted as playmates. He then correlated the IQ's of the choosers and the chosen, and examined the hypothesis and the choice happened to be **homophilous** (tendency of individuals to associate and bond with similar others**).**

History

- In 1928, Bott took an approach to examine the behavior of preschool children in Toronto. She identified five types of interaction: talking to one another, interfering with one another, watching one another, imitating one another, or cooperating with one another. She then used "focal sampling", observing one child each day.

History

- Bott's work also was a harbinger of the network research, in that she organized her data into matrices, and discussed her results in terms of the linkages between individuals.

History

- In "The Companionships of Preschool Children", Hagman (1933) observed interaction throughout the term, and interviewed children to measure their recollections of their interactions earlier in the term.

(University of Iowa Studies in Child Welfare)

History

- In 1933, the New York Times reported on the new science of "psychological geography" which "aims to chart the emotional currents, cross-currents and under-currents of human relationships in a community".
- Jacob Moreno analyzed the interconnections across 500 girls in the State Training School for Girls, and the interconnections of students within two schools.
- Moreno concluded that many **relationships were non-reciprocal**—and that many individuals were **isolated**.
- Moreno's quantitative method to map relationships is called "**sociometry**".

Social Network(cont'd)

Graph

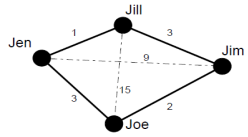
- A graph is a model for a social network
  - G ={N, L} ; N is set of nodes(actors) and L is set of lines(ties)
- Example Ties: Adjacency Matrices

Friendship

	Jim	Jill	Jen	Joe
Jim	-	1	0	1
Jill	1	-	1	0
Jen	0	1	-	1
Joe	1	0	1	-

Proximity

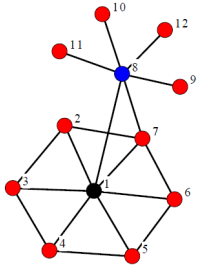
	Jim	Jill	Jen	Joe
Jim	-	3	9	2
Jill	3	-	1	15
Jen	9	1	-	3
Joe	2	15	3	-



Social Network(cont'd)

Length , Distance

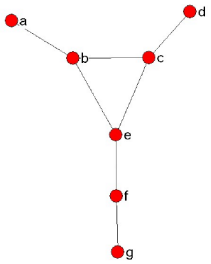
- Length of a path is number of links
- Distance between two nodes is length of shortest path (aka geodesic)



Social Network(cont'd)

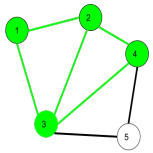
- Geodesic Distance Matrix

	a	b	c	d	e	f	g
a	0	1	2	3	2	3	4
b	1	0	1	2	1	2	3
c	2	1	0	1	1	2	3
d	3	2	1	0	2	3	4
e	2	1	1	2	0	1	2
f	3	2	2	3	1	0	1
g	4	3	3	4	2	1	0



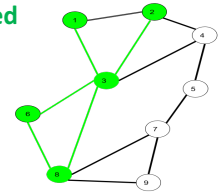
Social Network(cont'd)

- **Walk:**
  A walk is a sequence of vertices and edges of a graph.
  - If we traverse a graph then we get a walk.
    Vertex can be repeated
    Edges can be repeated
- Example walk:
  1->2->3->4->2->1->3
- Walk can be open or closed.
- Walk can repeat edges or vertices.



Social Network(cont'd)

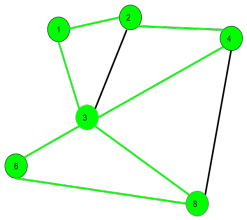
- **Trail:**
  A Walk in which no edge is repeated is a trail
  Vertex can be repeated
  Edges cannot be repeated
- Example Trail:
  1->3->8->6->3->2



Social Network(cont'd)

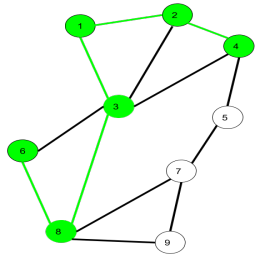
- **Circuit** (A closed trail) **:**
  Traversing a graph such that no edge is repeated but vertex can be repeated and it is also closed.
  Vertex can be repeated
  Edge cannot be repeated
- Example Circuit:
  1->2->4->3->6->8->3->1

Circuit is a closed trail.
These can have repeated vertices only.



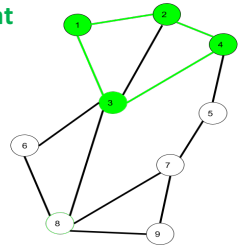
Social Network(cont'd)

- **Path:**
  It is a trail in which neither vertices nor edges are repeated i.e.
  - Traverse a graph such that we do not repeat a vertex and nor we repeat an edge.
    Vertex cannot be repeated
    Edge cannot be repeated
- Example Path:
  6->8->3->1->2->4



Social Network(cont'd)

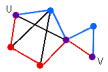
- **Cycle(**closed path) **:**
  Traversing a graph such that we do not repeat a vertex and an edge
  – but the starting and ending vertex must be same
  Vertex cannot repeat
  Edge cannot repeat
- Example Cycle:
  1->2->4->3->1



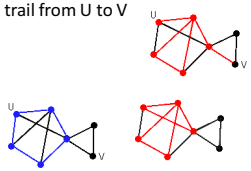
Path, Walk and Trail

- Walk: An alternating sequence of vertices and connecting edges.
- Path is a walk that does not include any vertex twice.
- A trail is a walk that does not pass over the same edge twice. A trail might visit the same vertex twice, but only if it comes and goes from a different edge each time.
- A cycle is a path that begins and ends on the same vertex.
- A circuit is a trail that begins and ends on the same vertex.

- Two paths from U to V



- A trail from U to V



Fundamental Concepts in Network Analysis

ACTOR:

- The social entities are referred to as actors
- Actors are discrete individual, corporate, or collective social units
- Examples:
  - People in a group
  - Departments within a corporation
  - Public service agencies in a city
  - Nation-states in the world
- The term "actor" does not imply that these entities necessarily have the ability to "act."
- Most social network applications focus on collections of actors that are of the same type. We call such collections one-mode networks.

Fundamental Concepts (cont'd)

Relational Tie:

- Actors are linked to one another by social ties
- The defining feature of a tie is that it establishes a linkage between a pair of actors.
- Examples:
  - Evaluation of one person by another (friendship, liking, or respect)
  - Transfers of material resources (business transactions, lending or borrowing things)
  - Association or affiliation (jointly attending a social event, or belonging to the same social club)
  - Behavioural interaction (talking together, sending messages)
  - Movement between places or status (migration, social or physical mobility)
  - Physical connection (a road, river, or bridge connecting two points)
  - Formal relations (authority)
  - Biological relationship (kinship or descent)

Fundamental Concepts (cont'd)

Dyad :

- A dyad consists of a pair of actors and the (possible) tie(s) between them.
- Many kinds of network analysis are concerned with understanding ties among pairs taking dyad as the unit of analysis.
- Dyadic analyses focus on the properties of pairwise relationships, such as whether ties are reciprocated or not.

Fundamental Concepts (cont'd)

Triad:

- A subset of three actors and the (possible) tie(s) among them
- Example:
  - Is Triad transitive ? (if actor i "likes" actor j, and actor j in turn "likes" actor k, then actor i will also "like" actor k)

Fundamental Concepts (cont'd)

Subgroup :

- A subgroup of actors is any subset of actors and all ties among them.

Fundamental Concepts (cont'd)

Group:

- Collection of all actors on which ties are to be measured
- Consists of a finite set of actors for conceptual, theoretical, or empirical reasons, on which network measurements are made.
- Actors in the group belong together in a more or less bounded set.

Fundamental Concepts (cont'd)

Relation:

- The collection of ties of a specific kind among members of a group is called a relation.
- The collection of ties of a given kind measured on pairs of actors from a specified actor set.
- Example
  - The set of friendships among pairs of children in a classroom
  - The set of formal diplomatic ties maintained by pairs of nations in the world.

Fundamental Concepts (cont'd)

Social Network:

- A social network consists of a finite set or sets of actors and the relation or relations defined on them.

Social Network Data

- Two types of variables in a network data set:
  - **Structural**
  - **Composition**
- Structural variables are measured on pairs of actors (subsets of actors of size 2) and are the corner-stone of social network data sets.
- Structural variables measure ties of a specific kind between pairs of actors.
- Examples:
  - measure business transactions between corporations
  - friendships between people
  - trade between nations

Social Network Data (cont'd)

- Social network data
  - consist of at least one **structural variable** measured on a set of actors.
- A specific network study usually determines which variables to measure, and which techniques are most appropriate for their measurement.
- Example: **Studying economic transactions between countries**
  - Cannot use an observational techniques; probably use archival records to obtain information on such transactions.
  - **Studying friendships among people**
    - Cannot use archival or historical records ; can, use questionnaires or interviews
- The nature of the study determines whether the entire set of actors can be surveyed or whether a **sample of the actors** must be taken.

Social Network Data (cont'd)

- Composition variables are measurements of actor attributes.
- Composition variables are of the standard social and behavioural science variety, and are defined at the level of individual actors.
- Examples:
  - gender, race, ethnicity of people
  - geographical location, after-tax profits, or number of employees of corporations

Social Network Data (cont'd)

Mode:

- Refers to a distinct set of entities on which the structural variables are measured
- Structural variables measured on a single set of actors give rise to one-mode network
- Example:
  - Friendships among residents of a neighborhood
- One-mode network -> all actors come from one set
- Structural variables can be measured on two (or even more) sets of entities.
  - Example: Study actors from two different sets, one set consisting of corporations and a second set consisting of non-profit organizations.
    - Helps us measure the flows of financial support from corporations to non-profit actors.

Social Network Data (cont'd)

- A network data set containing two sets of actors is referred to as a two-mode network
- A two-mode network data set contains measurements on which one of the sets have ties to actors in the other set
- Actors in one of the sets are "senders", while those in the other are "receivers"
- ***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

Social Network Data (cont'd)

Affiliation Networks:

- A special type of two-mode network in social network studies
- Affiliation networks are two-mode, but have only one set of actors
- The second mode in an affiliation network is a set of events (such as clubs or voluntary organizations) to which the actors belong
- In an affiliation network data the two modes are the actors and the events

Social Network Data (cont'd)

Affiliation Networks:

- In an affiliation network data the two modes are the actors and the events
- Events are not defined on pairs of actors, but on subsets of actors
- These subsets can be of any size
- An affiliation variable is defined on a specific subset of actors
- A subset of actors affiliated with an affiliation variable is that collection of actors who participate in a specific event
  - Example: Actors belonging to a given club

Social Network Data (cont'd)

- Affiliation Network : Example

  Consider a set of actors, and three elite clubs in some city.
  - One can define an affiliation variable for each of these three clubs.
  - Each of these variables gives us a subset of actors - those actors belonging to one of the clubs.
  - When events are clubs, boards of directors of corporations, or committees, the membership lists or rosters give the actors affiliated with each event.
- Example Events:
  - Informal social occasions, such as parties or gatherings
    - observations or attendance or interactions among people provide the affiliations of the actors.

Social Network Data (cont'd)

Boundary specification and sampling:

- The boundary of a set of actors allows a researcher to describe and identify the population under study
- Researchers often define actor set boundaries based on
  - relative frequency of interaction
  - intensity of ties among members as contrasted with non-members
- Two different approaches to boundary specification [Laumann, Marsden and Prensky (1989)]
  - Realist approach
  - Nominalist approach
- Realist approach :
  - focuses on actor set boundaries and membership as perceived by the actors themselves
  - Example: A **street-corner gang** is acknowledged as a social entity by its members and the membership of the gang is the collection of people the members acknowledge as belonging to the gang

Social Network Data (cont'd)

- Nominalist approach:
  - based on the theoretical concerns of the researcher
  - Example: A researcher might be interested in studying the flow of computer messages among researchers in a scientific specialty.
  - In such a study, the list of actors might be the collection of people who published papers on the topic in the previous five years
  - This list is constructed for the analytical purposes of the researcher, even though the scientists themselves might not perceive the list of people as constituting a distinctive social entity

Social Network Data (cont'd)

Many network studies focus on defined actor set boundaries.

Clearly defined actor set boundaries

- E.g: classrooms, offices, social clubs, villages, occasionally/artificially created and manipulated laboratory groups

Less well-defined boundaries

- E.g: Interorganizational networks in a community
  Interorganizational networks across an entire nation
- Population: who are the relevant actors? It is assumed that "we can obtain relevant information on all substantively important actors"
  - These actors "consist of all social units on which we have measurements (either structural variables, or structural and compositional variables)."

Social Network Data (cont'd)

- When the boundary is unknown, special sampling techniques such as **snowball sampling and random nets** can be used to define actor set boundaries.
- Not all times it is possible to take measurements on all the actors in the relevant actor set.
- Sampling: "A sample of actors might be taken" when it is not possible to take measurement on all the actors in the relevant actor set.
- A sample of actors may be taken from the set, and inferences are made about the "population" of actors from the sample.

Social Network Data (cont'd)

- The sampling mechanism is known, and the sample is a good probability sample (with known selection probabilities).
- The sample is viewed as the representative of the larger, theoretically interesting population and uses the sampled actors and data to make inferences about the population.
- Example: In a study of major corporate actors in a national economy, a sample of corporations may be taken in order to keep the size of the problem manageable
  - Reduces time and resource consumption

Social Network Data (cont'd)

Network Sampling :

- Primary focus is on estimation of network properties
  - average number of ties per actor
  - degree of reciprocity present
  - level of transitivity
  - density of the relation under study
  - frequencies of ties between subgroups of actors based on the sampled units

Social Network Data (cont'd)

Snowball network sampling

- A snowball network sample begins when the actors in a set of sampled respondents, report on the actors to whom they have ties of a specific kind
  - These nominated actors constitute the "first order" zone of the network.
  - The researcher then will sample all the actors in this zone, and gather all the additional actors (those nominated by the actors in the first-order zone who are not among the original respondents).
  - These additional actors constitute the "second-order" zone
  - This snow balling proceeds through several zones.
  - Subject to biases
  - E.g: Drug Use, Recruitment

Social Network Data (cont'd)

- Chain methods : Methods designed to trace ties through a network from a source to an end
- Chain methods include snowball sampling and small world technique (uses small-world network in which most nodes are not neighbours of one another)
  E.g: Strangers being linked by a short chain of acquaintances

- In some network sampling situations, it is not clear what the relevant sampling unit should be.
  - Sample actors, pairs of actors, triples of actors, or perhaps even subsets of actors …
- In other situations, one might sample actors, and have them report on their ties and the ties that might exist among the actors they choose or nominate.

Social Network Data (cont'd)

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

- Mode of a network: number of sets of entities on which structural variables are measured
- 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: mother-children

Social Network Data (cont'd)

- The number of modes in a network refers to the number of distinct kinds of social entities in the network.
- Example: A study in which subjects respond to a set of stimuli (such as questionnaire items) gives rise to two modes: the subjects and the stimulus items.
  - In the standard sociometric data design, a number of actors are presented with a list of the names of other people in the actor set, and asked to rate each other person in terms of how much they "like" that person.
  - In a non-network context one could view these data as two-mode : the people as respondents are the first mode, and the names of the people as stimulus (questionnaire) items are the second mode. However, as a social network, these data contain only a single set of actors, and thus, in our terminology, it is a one-mode network in which the relation of friendship is measured on a single set of people.

Social Network Data (cont'd)

One-Mode Networks:

- Involves measurements on a single set of actors
- Actors: The actors can be of a variety of types
  - People
  - Subgroups
  - Organizations
  - Collectives/Aggregates:
    - Communities
    - Nation-states
- Relations: measured on the single set of actors, viewed as representing specific substantive connections, or "relational contents"

Social Network Data (cont'd)

One-Mode Networks:

- Kinds of relations :
  - Individual evaluations: friendship, liking, respect
    - Measurements of positive or negative affect of one person for another.
    - These relations are labeled sentiment according to early sociometricians
    - Historically the most studied
  - Transactions or transfer of material resources: lending or borrowing; buying or selling
    - Include business transactions, imports and exports of goods
    - Specific forms of social support, such as lending and borrowing
    - Contacts made by one actor of another in order to secure valuable resources, and transfer of goods (exchange of gifts, borrowing or lending items, and sales or purchases)
  - Transfer of non-material resources: communications, sending/receiving information
    - Frequent communications between actors, where ties represent messages transmitted or information received.
    - Ties involve sending or receiving messages, giving or receiving advice, passing on gossip, and providing novel information.
    - Information about innovations is frequently diffused over such communication channels

Social Network Data (cont'd)

- Kinds of relations :
  - Interactions
    - Involve the interaction of actors or their presence in the same place at the same time.
    - Examples : sitting next to each other, attending the same party, visiting a person's home, hitting, conversing
  - Movement: physical (migration from place-to-place), social (movement between occupations or statuses)
    - Individuals moving between communities can be counted
    - Workers changing jobs or people changing statuses
  - Formal roles
    - Dictated by power and authority
    - Represent authority of one actor over others
    - Example : boss/employee, teacher/student, doctor/patient
  - Kinship: marriage, descent
    - Ties can be based on marriage or descent relationships

Social Network Data (cont'd)

One-Mode Networks:

- Actor Attributes: Social network data sets can contain measurements on the characteristics of the actors
  - Measurements constitute the composition of the social network
  - People can be queried about their age, gender, race, socio-economic status, place of residence, grade in school
  - For corporate actors, one can measure their profitability, revenues, geo graphical location, purpose of business

Social Network Data (cont'd)

Two-Mode Networks:

- Involves measurements on two sets of actors, or on a set of actors and a set of events
- Two Sets of Actors:
  - Dyadic two-mode networks : Relations measure ties between the actors in one set and actors in a second set
    - Relations are functions of dyads in which the first actor and the second actor in the dyad are from different sets

Social Network Data (cont'd)

Two-Mode Networks:

- One Set of Actors and One Set of Events (Affiliation Network)
  - Arises when one set of actors is measured with respect to attendance at or affiliation with, a set of events or activities.
  - First mode is a set of actors and the second is a set of events which affiliates the actors
  - Example: A set of women attended a variety of social functions, and the attendance was recorded over a period of several months. Each social function can be viewed as a variable and a binary measurement made as to whether a specific actor attended the specific function. These variables are termed as affiliational.
  - Actors : First set of actors and a second set of events or activities to which the actors in the first set attend or belong.
  - Actor types are similar to those in one-mode network
  - Actors must be affiliated with one or more events

Social Network Data (cont'd)

Two-Mode Networks:

- Actors: Contains two sets of actors
  - Two sets of actors may be of different types
- Relations: At least one relation is measured between actors in the two sets
  - In an extensive two-mode network data set, relations can also be defined on actors within a set.
- Example: A collection of corporations (actor set 1) and the non-profit organizations (actor set 2) which rely on contributions from the public sector for their operating budgets
  - Primary relation was the flow of donations from the corporations to the non-profit organizations
  - Relation is unidirectional since it flows from actors in one set to actors in the other set, but not the reverse.
  - Other relations : Shared country club memberships among the chief executive officers, interlocking boards of directors

Social Network Data (cont'd)

Two-Mode Networks:

- One Set of Actors and One Set of Events (Affiliation Network)
  - Events: actors (the first mode) are related to each other through their joint affiliation with events (the second mode)
  - Events defined on the basis of membership in clubs or voluntary organizations, attendance at social events, sitting on a board of directors or socializing in a small group
  - Nature of the events, which affiliate the actors, depends on the type of actors involved
    - People may attend social functions or belong to athletic clubs
    - Subgroups of people may attend various committee meetings (for example, departments at a major university send representatives to college committee meetings)
    - organizations may be represented on various boards of directors in a community
    - countries might belong to treaty organizations

Social Network Data (cont'd)

Two-Mode Networks:

- One Set of Actors and One Set of Events (Affiliation Network)
  - Attributes: Same types as in one-mode and two-mode networks
  - Events may have characteristics associated with them which can be measured and included in the network data set
  - Example: Clubs will be of a particular size or located in a specific geographical area. Events usually occur at discrete points in time, as well as in particular geographical places. Thus, there can be two sets of attribute variables in an affiliation network data set: attributes of the actors, and attributes of the events.

Social Network Data (cont'd)

- Special Dyadic Networks:
  - Some data collection designs gather structural information on some pairs but not others.
  - Example: Study of Student's Project Team. Each partner in the team can interact with the other but with no other student during project review sessions. Interactions during these sessions my be recorded.
  - When interest centers on a collection of pairs (Father-Son, Mother-Child), one frequently samples from a large population of such pairs.
    - This non-network relational data is referred as special dyadic designs.
  - An actor may also relate to a limited number of "special" other actors. Example: Mothers interacting with their own children
    - Mothers only interact with their own children, and children only interact with their own mother
    - Partners for one person are different from the partners for another (The design of the experiment constrains the interactions among the set of people so that all people cannot, theoretically, interact with all the others)

Social Network Data (cont'd)

- Ego-centered network:
  - Consists of a focal actor, termed ego, a set of alters who have ties to ego and measurements on the ties among these alters.
  - Example: When studying people, one samples respondents, and each respondent reports on a set of alters to whom they are tied and on the ties among these alters (Personal network data).
  - These data are relational, but limited, since ties from each actor are measured only to some alters.
  - Widely used by anthropologists to study the social environment surrounding individuals or families
  - Used in the study of "Social Support" (used to refer to social relationships that aid the health or well-being of an individual)
    - E.g: Clinical and community psychology/sociology
  - A Social Survey - "Looking back over the last six months - who are the people with whom you discussed matters important to you? "

Social Network Data (cont'd)

Network data, measurement and collection:

- Topics: Issues concerning
  - measurement of network data
  - collection of network data
  - accuracy
  - validity
  - error associated with these data
  - particular design considerations that can arise in network
- Measurement: "social network data consist of one (or more) relations measured among a set of actors"
  - These relations has implications for a number of measurement issues
    - 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

Social Network Data (cont'd)

Network data, measurement and collection:

- Issues in Measurement:
  1. Unit of observation : actor/dyad/triad/subset of actors/network
     - The entity on which measurements are taken
     - Network Data can be observed at different levels
     - Network data often are gathered at a level that is different from the level at which they are modeled
     - Collected by observing, interviewing, or questioning individual actors about the ties from these actors to other actors in the set
  2. Modeling unit: actor/dyad/triad/subset of actors/network
     - Refers to the level at which network data are modeled/summarized
     - Levels are
       - Actor (Example: The number of "choices" that an individual actor receives from others in the network).
       - Dyad (Example, if one person "chooses" another as a friend, is the "choice" returned by the second person?).
       - Triad
       - Subgroup (Example: Study whether there are subsets of actors in the network who interact frequently with each other)
       - Set of actors or network(Example: Proportion of ties that are present in the network)

Social Network Data (cont'd)

Network data, measurement and collection:

- Issues in Measurement:
  3. Relational quantification: directional vs. non-directional; dichotomous vs. valued
     - Refers to the properties of relations that are important for understanding
     - Directional relation: The relational tie between a pair of actors has an origin and a destination(tie is directed from one actor in the pair to the other actor in a pair)

       Example: One country exports manufactured goods to a second country; the first country is the source of the manufactured goods, and the second country is the destination
     - Non-directional relation: Tie between a pair of actors does not have a direction.

       Example: Tie is present between two countries if they share a border

Social Network Data (cont'd)

Network data, measurement and collection:

Data 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

Social Network Data (cont'd)

Data Collection:

- 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 actor's 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

Social Network Data (cont'd)

Data Collection: Techniques used to gather network data

Questionnaire:

- Most commonly used data collection method
- Contains questions about the respondent's ties to the other actors
- Most useful when the actors are people, and the relation(s) that are being studied are ones that the respondent can report on.
- Example: People can report on who they like, respect, or go to for advice.
- Questionnaires can also be used when the actor in a study is a collective entity, such as a corporation, but an individual person representing the collective, reports on the collective's ties.
- Example: "ABC" askes officers in-charge of corporate whether or not the corporation had made a donation to a non- profit agency.

Social Network Data (cont'd)

Data Collection - Questionnaire:

- Three different question formats used in a questionnaire :
  - Roster vs. Free recall
  - Free vs. Fixed choice
  - Ratings vs. Complete rankings

Roster:

- Should each actor be presented with a complete list, or roster, of the other actors in the actor set?
  - Rosters can be constructed only when the researcher knows the members in the set, prior to data gathering
  - Example: Friendships among members of a university class ["trust as a friend", "know well", "acquaintance", "associate name with face", and "do not know"]

Social Network Data (cont'd)

Data Collection - Questionnaire:

Free Recall:

- The researcher does not present a complete list of the actors in the network to the respondent on the questionnaire.
- Ask respondents to "name those people with whom you would like to get advice from "
  - This format of generating the list of names, is called free recall
- Example: Study of friendships in two high schools. Students were asked to list their best friends, but were not presented with a roster.

Social Network Data (cont'd)

Data Collection - Questionnaire:

Fixed Choice:

- If actors are told how many other actors to nominate on a questionnaire

  then each person has a fixed number of "choices" to make.
- **Each actor has a fixed maximum number of ties to the other actors in the set of actors.**
- Example 1: Name a specific number of "best friends"
- Example 2: Study of diffusion of a medical innovation among physicians, interviewed all physicians in a community. Each doctor interviewed was asked three sociometric questions:

  (i) "To whom did he most often turn for advice and information?"

  (ii) "With whom did he most often discuss his cases in the course of an ordinary week?"

  (iii) "Who were the friends, among his colleagues, whom he saw most often socially?"

Social Network Data (cont'd)

Data Collection - Questionnaire:

Free Choice:

- Actors are not given any constraints on how many nominations to make.
- Example: Study of ties among people in an "invisible college" of users of a computer program at a variety of universities
- Each individual was asked to denote for each member of the user group whether or not they:
  - Had an office next to each other
  - Attended the same school at the same time
  - Shared an office
  - Lived in the same living apartment
  - Were at the same school at the same time
  - Were in the same academic department at the same time
  - No constraint on the number of people that an individual respondent can choose on these relations.

Social Network Data (cont'd)

Data Collection - Questionnaire:

Rating:

- In some network designs, actors are asked to rate or rank order all the other actors in the set for each measured relation.
- Measurements reflect the intensity or strength of ties.
- Ratings require each respondent to assign a value or rating to each tie.

Complete ranking: requires each respondent to rank their ties to all other actors.

- Full rank-orders and rating scales produce valued relations.

Social Network Data (cont'd)

Data Collection - Interview:

- Either face-to-face or over the telephone
- Occasionally used to gather network data in instances where questionnaires are not feasible
- Interviews have been used to gather data from respondents in ego-centered networks

Social Network Data (cont'd)

Data Collection - Observation:

- Observing interactions among actors.
- Widely used in field research to study relatively small groups of people who have face-to-face interactions.
- Since data are collected by observing interactions, without requiring verbal responses from the people, this method is quite useful with people who are not able to respond to questionnaires or interviews.
- Also used in the study of interactions among non-human primates
  - Example: Observing a colony of monkeys, and recording which monkeys visited a river together
- Useful for collecting affiliation network data
  - Example: Researcher can record who attends each of a number of social events.

Social Network Data (cont'd)

Data Collection - Archival Records:

- Researchers measure ties by examining measurements taken from records of interactions.
- Records can take many forms, such as
  - measurements on past political interactions among nations
  - previously published citations of one scholar by another
- Can be used to reconstruct ties that existed in the past.

Social Network Data (cont'd)

Data Collection – Other Special Network Design Methods:

- Cognitive social structure design (extension of sociometric data to include actor perceptions of the network)
- Experimental studies (Network data are collected under controlled situations)
- Ego-Centered Network
- Small Word Network

Social Network Data (cont'd)

Data Collection – Other Special Network Design Methods:

Cognitive social structure design (extension of sociometric data to include actor perceptions of the network)

- In a standard sociometric questionnaire, one asks respondents about their own ties
- A variation of this design is to ask respondents to give information on their perceptions of other actor's' network ties.
- They measure perceived relations
- Gives considerably more information than the usual sociometric design, since actors report not only on their own ties, but also on their perceptions of ties among all pairs of actors.
- Example: Study of turnover in several fast food restaurants
  - They were interested in the employees' perceptions of friendships among all other employees in the restaurant . They had to gather information from each person not only about their own friendships, but also about their perceptions of the friendships among all other pairs of employees.

Social Network Data (cont'd)

Data Collection – Other Special Network Design Methods:

Experimental:

- Method 1 :
- One can choose a set of actors and observe their interactions in an experimentally controlled situation
- Researcher records interactions or communications between pairs of actors
- Ties may be observed between all pairs of actors
- Method 2:
- Choose a set of actors but also specify which pairs of actors are permitted to communicate with each other during the experiment
- Record the frequency or content of communications between those pairs of actors who are permitted to interact
- Example: Group problem-solving experiments

Social Network Data (cont'd)

Data Collection – Other Special Network Design Methods:

- Ego-centered network:
  - Consists of a focal actor, termed ego, a set of alters who have ties to ego and measurements on the ties among these alters.
  - Example: When studying people, one samples respondents, and each respondent reports on a set of alters to whom they are tied and on the ties among these alters (Personal network data).
  - These data are relational, but limited, since ties from each actor are measured only to some alters.
  - Widely used by anthropologists to study the social environment surrounding individuals or families
  - Used in the study of "Social Support" (used to refer to social relationships that aid the health or well-being of an individual)
    - E.g: Clinical and community psychology/sociology
  - A Social Survey - "Looking back over the last six months - who are the people with whom you discussed matters important to you? "

Social Network Data (cont'd)

Data Collection – Other Special Network Design Methods:

Small World Network Design:

- The small world method consists of presenting each of the person in a "starting population" with the description of a given "target person" his name, address, occupation, and other selected information.
  - The task of a starter is to advance a booklet towards the target person by sending the booklet to a personal acquaintance whom he/she considers more likely than himself/herself to know the target.
  - Each person in turn advances the booklet in this manner until the chain reaches the target
- The intermediaries are asked to return a postcard to the researcher reporting some basic demographic characteristics.
- The researcher can then compare characteristics of successful and unsuccessful chains.

Social Network Data (cont'd)

Data Collection – Other Special Network Design Methods:

- Reverse Small World Network Design
- Study focuses on the ties from a specific respondent to a variety of hypothetical targets
- Individuals are asked to imagine that they will pass something to someone who is to eventually reach a target person they do not know.
- They are instructed to think of someone they know, who might be a first link in a chain to the target person
- The respondent is given a list of possible targets who are located geographically and socially in different parts of the society.
- In this way the reverse small world method clearly maps the outgoing network of the people who complete the questionnaire.
- Possibility of biases using the small world technique
  - These biases arise as response rates are typically much lower with this form of network data collection.

Social Network Data (cont'd)

Data Collection – Other Special Network Design Methods:

Diary:

- Ask each respondent to keep a continuous record of the other people with whom they used in the study of personal networks among people
- Social support researchers sometimes ask respondents to keep daily records of all people with whom they come into contact
- These data sets frequently include information on the type of relation and characteristics of the alters in each ego-centered network

Social Network Data (cont'd)

Longitudinal Data Collection :

- Researcher is interested in how ties in a network change over time
- Researcher measures one or more relations at fixed intervals of time
- Study how stable ties are and whether such ties ever reach an equilibrium state
- Two research questions arise when we study longitudinal network data
  - How the process has changed over time?
  - How well the past, or the history of the process, can predict the future?

Social Network Data (cont'd)

Longitudinal Data Collection :

- Data can be collected using any of the previously discussed methods
- Example: Study of friendship/Study of communications throughout a network over time
  - Study of twenty men housed together, for a period of fifteen weeks, who were initially unknown to each other. Each person was asked to rank each of his fellow fraternity members on the basis of positive feeling. Rankings were recorded each week.
- Example: Measurements of computer mail interactions, over the course of an eighteen month period, among a set of quantitative researchers studying social networks.

Social Network Data (cont'd)

Measurement Accuracy, Validity, Reliability, Error:

- Social network research is concerned with studying patterns of social structure.
- The structure refers to a relatively prolonged and stable pattern of interpersonal relations.
- The pattern is expected to be the true structure, in contrast to the observed structure contained in the measured network data, which might contain error.
- Since social network data are often collected by having people report on their own interactions, the accuracy of these self-report data is also a concern.

Social Network Data (cont'd)

Accuracy:

- Sociometric data are collected by having people report on their interactions with other people.
- Example: A researcher might ask each actor to report "With whom did you talk last week?", or "What other people were at the party with you last Saturday?"
- The respondent is asked to recall his or her interactions.
- Issue 1: Relationship between information collected using verbal reports and information collected by observing the actor's interactions.
- Research concluded that "Half of what people report about their own interactions is incorrect in one way or another"
  - People are not very good at reporting on their interactions in particular situations.

Social Network Data (cont'd)

Accuracy:

- "True" structure of the network, relatively stable patterns of interaction, are of most interest.
- Long term patterns are not supposed to be studied on interactions of individuals
  - What people report about their interactions is in fact related to the long-range social structure, rather than to particular instances.
- Verbal reports (recall of interactions) should be understood using principles of memory and cognition.
- Issue 2: When the actors in the network are organizations (for example corporations) but information on ties is collected from individuals as representatives of the organization.
  - Example: Measurement of donations from corporations to non-profit agencies by interviewing the officer in charge of corporate giving the fund.
  - One must be able to assume that the individual who is interviewed in fact has knowledge of the information being sought.

Social Network Data (cont'd)

Validity :

- A measure of a concept is valid to the extent that it actually measures what it is intended to measure.
- A researcher assumes that the measurements of a concept are indeed valid.
- Example: One might assume that asking people "Which people in this group are your friends ? has face validity as a measure of friendship.
  - Answer to the question gives a set of actors who are related to the respondent through friendship ties
- Validity of a measure of a concept is seldom tested in a rigorous way.
- Construct validity (formal notation), arises when measures of concepts behave as expected in theoretical predictions.
- Construct validity of social network measures can be studied by examining how these measures behave in a range of theoretical propositions.
- Example: The sociometric measure, "Number of choices" received by an actor, were related to a number of actor characteristics such as leadership and effectiveness, thus demonstrating the construct validity of those sociometric measures.

Social Network Data (cont'd)

Reliability :

- A measure of a variable or concept is reliable if repeated measurements give the same estimates of the variable.
- In psychometric test-theoretic framework the reliability of a measure can be assessed by comparing measurements taken at two points in time (test-retest reliability), or by comparing measurements based on subsets of test items.
- Three approaches to assess reliability of social network data :
  - test-retest comparison
  - comparison of alternative question formats
  - reciprocity of sociometric choices
- In test-retest assessment, one must assume that the "true" value of a variable has not changed over time.
  - Assumption is inappropriate for social network properties, since social phenomena cannot be assumed to remain in stasis.
  - Assessing reliability of social network measurements using the test-retest approach is therefore problematic

Social Network Data (cont'd)

Reliability :

- Can be assessed at different levels
  - study the reliability of the "choices" made by individual actors/measures aggregated over a number of individual.

- **Sociometric questions** using ratings or full rank orders are more reliable (have higher test-retest reliability) than fixed choice designs in which just a few responses are allowed.
- Responses to sociometric questions about more intense relations have higher rates of reciprocation than sociometric questions about less intense relations.

- Reliability of aggregate measures (such as popularity) is higher than the reliability of "choices" made by individual actors

Social Network Data (cont'd)

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.
- Observations or measurements of a concept are an additive combination of the "true" score plus error (or noise).
- Error - Difference between the true and observed values, is referred to as measurement error.
- In a fixed choice design, the respondent is instructed to nominate or name some fixed number of others for each relation.
  - Example: Each person may be asked to "List your three best friends" .This design introduces error since it is quite unlikely that all people have exactly three best friends.