Social Network Analysis

Social network analysis (SNA) is both a theoretical perspective and a set of methods. In terms of theory, SNA extends and complements traditional social science by focusing on the causes and consequences of relations between people and among sets of people rather than on the features of individuals. In terms of method, SNA focuses on the measurement of relationships between people. By quantifying the relationships between people, network analysts can apply models and techniques that are commonly used across the social and natural sciences.

Two distinct approaches to SNA arose from two distinct historical traditions. The sociocentric (whole) network approach comes from sociology and was heavily influenced by the work of Georg Simmel. Sociocentric network analysis involves the quantification of relationships between people within a defined group – a classroom of children, a board of directors, the residents of a village or town, the trading partners in a bloc of nations. By representing relationships as numbers, many powerful mathematical and statistical analyses can be applied. Sociocentric network analysis begins with the assumption that members of a group interact more than would a randomly selected group of similar size. The focus is on measuring the structural patterns of those interactions and how those patterns explain outcomes, like the concentration of power or other resources, within the group. Sociocentric network analysts are interested in identifying structural patterns in cases that can be generalized, and in this sense they are like physicists or economists who are interested in modeling behavior.

The egocentric (personal) network approach arose from anthropology and traces its roots to A. R. Radcliffe-Brown, among others. This form of SNA is almost always about people rather than about groups. An egocentric network comprises the people (what social network experts call alters) that a person (referred to as ego) knows. An egocentric network thus may have as its members spouses, children, cousins, co-workers, church members, book club members, or just plain friends. So the personal network of an elementary school teacher may contain her husband, her son and daughter, all of their friends and relatives, her own friends and relatives, her co-workers, students, parents of students and members of her church. But she may have more family relations than, for example, the CEO of a large company who has less time to maintain those relationships.

Egocentric SNA is concerned with making generalizations about the features of personal networks that explain things like longevity, consumer and voting behavior, coping with difficult life situations, economic success or failure, and so on. With its focus on individuals, the egocentric network approach has been more germane to studies of community than the sociocentric network approach. It is also possible to treat organizations, classrooms, communities or even nations as the ego in an egocentric network study.

Historical Development

The roots of social network analysis are to be found in the work of German sociologists at the turn of the 20th century. While other theorists focused on describing social phenomena like war, economics and religion, Georg Simmel and others sought to construct a theory that explained how these social phenomena came about. Simmel's writing on the fundamental difference between the interactions in dyads (two people) and triads (three people), and his notion of urban systems being composed of intersecting networks and circles, was the basis for his 'formal sociology,' the precursor to social network analysis.

Jacob Moreno was the first to operationalize a social network (1934). Moreno devised a system for representing a social network as a combination of points and lines. Using this system, he illustrated how different network configurations might affect a given network member. Moreno coined the term 'sociometrics' to describe these configurations.

Building on Moreno's work, Dorwin Cartwright and Frank Harary recognized the potential of applying the concepts of the existing field of graph theory (a set of propositions dictating how points and lines in a graph are connected) to the sociogram. By adding direction to the lines and the possibility for the relationship to be positive or negative, Cartwright and Harary were able to show much more complexity in the patterns of social relations. Analyses based in graph theory are now fundamental tools of social network analysis.

Two separate traditions can be traced to the ideas of the anthropologist Radcliffe-Brown, beginning in the late 1930s. The first is the work of a group at Harvard University on ways to find subgroups of people in larger groups. Independently of Moreno, they developed a system for illustrating social networks and evolved a set of rules for representing various relationships. They were also the first to represent social networks as matrices (blocks of numbers where each number represents the relationship between the column and row it intersects), paving the way for the development of matrix-based clique-finding algorithms.

More important for community studies was the influence that Radcliffe-Brown had on a group of anthropologists at the University of Manchester, including John Barnes, Clyde Mitchell and Elizabeth Bott. Taking their cue from Radcliffe-Brown's notion of societies being built up from a 'web of relations,' the Manchester anthropologists conceived of society as a tapestry woven from the social networks of individuals. Rather than focusing on the whole society, they studied the networks of relations surrounding individuals. This concept was tested by Bott in her work with English families, by Barnes in his work with Norwegian fishermen, and by Mitchell in his work with rural migrants to towns of what was then Northern Rhodesia (Zambia today).

These two branches of social networks – one grounded in sociology, the other in anthropology – developed more or less independently until the 1970s. Then, the increased availability of computers made possible the development of sophisticated measures of network structure and the analysis of network data from studies of large groups. Since 1978, the International Network for Social Network Analysis and the journal, *Social Networks,* have brought network analysts from across the social sciences together, as has the Sunbelt Social Network Conference since 1981.

Sociocentric network methods and analysis

The basis for sociocentric network analysis is a matrix where the rows and columns represent the members of the group being studied, and each cell of the matrix contains a measurement of some tie between those members. The diagonal of the matrix is the intersection of each member (person, organization, nation) with itself, and is often ignored in analytical routines.

Some groups leave material traces of their interactions – traces that can be collected unobtrusively. John Padgett made matrix representing the interactions among the set of elite families in Renaissance Italy using their marriage records. Many studies of the structure of academic communities are based on the number of times pairs of people in a group have published together.

Mutual affiliation also produces sociocentric matrices that are appropriate for SNA. In the 1940s, Allison Davis and others recorded the number of times a pair of socialites attended an

event together. This approach of recording mutual attendance at events has since been used to analyze interlocking corporate directorates, where each cell in the matrix is the number of boards of directors on which each pair of directors sits. H. Russell Bernard and Peter Killworth conducted a series of studies comparing records of human interaction (like the tape recorded communications of ham radio operators and the paper records of teletype communications among the deaf) with informants' reports and found that people are quite inaccurate in reporting the amount of interaction they have with others in a group. Several researchers (Linton Freeman and A. Kimball Romney, among others) have since found that people answer the question "who do you communicate with?" with what they understand to be generally true.

Asking people about their interaction with others –their communications, their exchange of advice and other resources – remains the source of most sociocentric network data. When groups are small (up to 150, but usually 20-60) the researcher can list the members' names and ask each person how well they know each other person (on a scale of 0 to 5, for example), or how often they interact with each other person (for example, once a week, once a month or never). For example, a researcher may present all students ina class with a list of all the students in the class and ask them to rate how well they know each one. In large communities (more than 150) this is not possible as the list of mebers the respondent must comment on is too large. Instead the researcher might ask respondents to offer the names of the five or ten people within the large community with whom they exchange a particular resource or to whom they feel close.

All of these methods produce a member-by-member matrix where each cell represents the strength of relationship between members of the group. For very large groups (one with a thousand members, for example) the sociocentric network approach has traditionally been of limited value because software for storing these matrices was unavailable. This limitation is being overcome, however by advances in computer technology and software design.

Analyses of socoiocentric network data focus on structural properties of relations, rather than the individual relationships themselves. There are two broad categories of structural analyses – graph based and statistics-based. Graph based analyses are derived from graph theory where the focus is on the existence of a relationship between two network members – a tie between two network nodes, in the jargon of the field – rather than on the strength of the relationship. Some measures (called clique-finders) are designed to find subnetworks, like groups of friends in an elementary school classroom or affiliation groups in an organization. Measures of centrality show the extent to which relationships are concentrated in a few people, like an office manager in an academic department. Measures of structural equivalence classify group members by similarities in the pattern of their ties to others in the network. None of these analyses are available in conventional statistical analysis packages, so specialized network analysis software has developed and is now widely available.

Statistics-based analyses rely on the concept of variance and statistical distributions of means to describe structure. In this approach, a matrix of similarity between members of a group is created (using correlation coefficients, for example), and a variety of statistical analyses are applied. Multidimensional scaling is useful for displaying data and for discovering the underlying forces that bind people together. Cluster analysis is useful for finding subgroups within networks, taking into account the strength of the relationship between members. Many of these procedures are available in the major statistical packages such as SAS and SPSS.

Network visualization is a third type of analysis and can be either graph based or statistics based. Many network researchers find that the summary measures described above are useful because they restrict the view of the network to specific aspects. Just like an astronomer

may want to focus only on the infrared spectrum or gamma rays, the network researcher often wants a clear picture of only one aspect of structure. For this they use the various measures described. At other times network researchers want to explore the structural richness of their data. The best way to do this is with network visualization.

Network visualization software lets the researcher see all the connections within a network simultaneously and display network members in different colors, according to characteristics like gender, race, and age. One can quickly see divisions and sub-groupings within networks in this way. Visualization methods are most useful with relatively small networks.

Egocentric methods and analysis

Unlike sociocentric network analyses where the focus is on the pattern of relations within a socially defined group, egocentric network analysis focuses on the networks of individuals. From this perspective, each person has their own network of relationships that cut across many groups and that contribute to their behaviors and attitudes. Egocentric network research typically does not focus on network structure or pure models of behavior. Its strength is in its ability to capture the diversity of the social environment and to apply standard survey sampling techniques, which in turn allows results to be generalized. In 1986, for example, through the efforts of Ronald Burt, the General Social Survey (an annual face-to-face survey of a representative sample of about 1600 Americans) added a social network component based on a respondent's list of people with whom he or she discussed important matters.

Egocentric network researchers are interested in the list of a person's network members – called "alters" in the jargon of the field – and also in the relations among all pairs of those alters, Of course, researchers cannot interview each respondent's alters (his or her mother, co-workers, friends, etc.) and must rely instead on respondents to report their relationships with their alters. These reports are elicited from the respondent using one or more network generators.
In studies of social support, respondents are typically asked to name a small number of alters (three, five, ten) on whom they rely for advice or material help. Respondents may be asked to think of five people they talk to about important matters, or three people they talk to about health care decisions. In studies of support that involve weak ties (acquaintances, for example, rather than relatives or close friends or co-workers)respondents may be asked to list 50 people they know.

The method for sampling respondents varies greatly depending on the type of study. A balance must be achieved between the number of respondents, the number of alters they will be asked about, the amount of information about each alter elicited, and the method of data collection (face-to-face, mail or telephone). Some network studies have only a handful of respondents, while others have thousands.

Typically, egocentric network researchers ask respondents to rank the strength of their relationship with each alter, such as on a scale of 1 to 5. Other commonly-asked questions are the gender, age and race of the alter. It is also common to have respondents indicate how they know each of their alters – as family members, friends or acquaintances, co-workers, etc. Specific questions may be asked depending on the purpose of the study. For example a study of social support may include a question about the amount of assistance the alter provides the respondent with child-care. A marketing study may ask whether the alter has ever recommended a brand of laundry detergent. The researcher must consider carefully the respondent's ability to

answer questions about their alters. For example, respondents may not be able to report reliably on an alter's political affiliation or the alter's attitudes on child rearing.

Most analyses of egocentric network data summarize the composition of the network as a set of variables that become attributes of the respondent. Along with the age, education and income level of a respondent, the researcher may have the average age of their alters, the average strength of their ties with alters, the percent of their network that are family or co-workers, or the percent of their network from which they can borrow money or get a ride to the doctor. These measures may, in turn, be used as independent variables to predict things like scores on a depression scale.

Some egocentric network researchers try to measure structure within each respondent's network. To do this they must get respondents to report not only on their relationship with each alter, but also on the relationships of all pairs of alters. Typically, researchers ask only about the existence of a tie, and possibly its strength and avoid questions about asymmetric ties because these are unlikely to be well understood by respondents. For example, respondents are not likely to know if two of their co-workers know each other unequally. Even limiting these reports to symmetric ties, the number of tie evaluations grows geometrically as alters are added. For a network of 10 alters a respondent must report on 45 ties. For a network of 50 alters they must report on 1,225 ties. There is, of course, some question as to whether respondents can report accurately on the existence of ties between so many pairs of alters.

Structural data such as these may be analyzed with many of the measures described in the previous section on sociocentric networks. As with most egocentric network analyses, the structural characteristics may be summarized to the respondent level and used as independent or dependent variables. For example, a researcher may want to explain variability in levels of social network density or centrality using the respondent's age or race. Given the difficulty of collecting these data, studies such as these are rare.

"Small world" studies, developed by Stanley Milgram in the 1960s and 1970s, represent an interesting and less conventional use of egocentric network data. In these studies, respondents in various parts of the United States are told the name, occupation, and city of residence of some target person and are asked to mail a packet of papers to the that person if and only if they know the target personally. If respondents do not know the target personally, they are asked to send the packet to someone who they do know and who they believe has a chance of knowing the target. Tracking the path of the packets provides information on how people know each other and on the average number of links between pairs of randomly chosen people in a large society like the United States. Egocentric networks are also studied in research on estimates of personal network size and in research on the size of hard-to-count populations (the homeless, rape victims, IV drug users, etc.).

Social Networks and Studies of Community

While there have been some notable sociocentric analyses of communities, most of the research in this area has used the egocentric network approach. One of the earliest community studies to use the egocentric network approach was Barnes's examination of community integration in a Norwegian village. Barnes selected the community because it appeared reasonably small and stable. This let him see how typical community institutions, such as governing bodies and social groups, related to the composition of egocentric networks. Ethnographic studies like Barnes' are appropriate for small communities, but not for whole cities. In the Detroit Area Study, Edward Laumann used standard survey methods to collect data

on over a thousand respondents, each one naming up to three network alters. Laumann used multidimensional scaling to understand the underlying principles that organized hierarchy within occupations and later to study social position and friendship. Barry Wellman collected narratives from a sample of respondents in Toronto to determine how people retained a sense of community despite living in a highly urbanized area. This combined ethnographic and survey methods in a single study, a strategy that Wellman has continued to use in studies of social support and of virtual communities that form on the Internet.

In the Northern California Study, Claude Fischer elicited up to 14 alters from more than a thousand respondents. The detail from this study provided rich insights into the composition of egocentric networks and how they vary by respondent characteristics. Fischer was able to relate respondent characteristics to characteristics of different types of communities.

Egocentric networks have been used in studies of many communities, including the mentally ill; children in classrooms, in schools, and in whole school systems; ethnic groups embedded in larger populations; and people adjusting to disasters and wars. By adjusting the respondent selection criteria, the number of network alters elicited and the information about each alter, the egocentric approach can be applied to a variety of communities.

While sociocentric SNA is less prevalent in the study of communities, it has been used to great effect in studies of the diffusion of innovations. In the 1950s, James Coleman, Elihu Katz, and Herbert Mendel did a pioneering study of the adoption of tetracycline (a new drug at the time) among physicians in Peoria, Illinois. The study showed how adoption of the drug moved through the community and which attributes of various network members best explained the pattern of diffusion. While individual relationships are key in the diffusion of an innovation in its early stages, network roles play a more important part. Everett Rogers used a similar approach to describe agricultural innovation in Korean villages.

In a classic community study using the sociocentric approach, Allison Davis, Burleigh Gardner and Mary Gardner focused on class and racism in 1941 in Natchez, Mississippi. Using lists of attendees at formal events and parties, they were able to discern subgroups among a seemingly homogenous group of elite women. These network subgroups, difficult to observe ethnographically, helped to explain the inner workings of a caste system, and interactions between Blacks and Whites.

Much like Barne's study in Norway, W. Lloyd Warner's Yankee City studies of Newburyport, Massachusetts aimed to show how individuals were integrated into the community via informal social groupings. Through ethnographic observation and interviewing, Warner and his colleagues discovered a set of socially recognized informal groupings, such as people who hang out together. This led to the formal concept of the clique and made it possible to understand the relationship between informal groups and formal institutions.