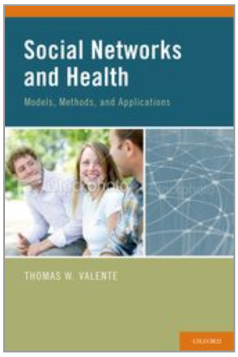

Social Networks and Health: Models, Methods, and Applications
Thomas W. Valente

Print publication date: 2010
Print ISBN-13: 9780195301014
Published to Oxford Scholarship Online: May 2010
DOI: 10.1093/acprof:oso/9780195301014.001.0001

Ego- and Personal-Network Effects

Thomas W. Valente

DOI:10.1093/acprof:oso/9780195301014.003.0004

Abstract and Keywords

This chapter shows how researchers can collect network data from randomly drawn samples. The data only represent the respondent personal network, usually from the respondent's perspective. The chapter then details the various measures derived from personal network data and the research questions and hypotheses that have been tested. Personal network exposure, concurrency, and tie strength hypotheses are discussed. A comparison between ego-centric and sociometric data is presented. The chapter discussed how to convert egocentric data to a dyadic dataset thus facilitating analysis and testing of certain hypotheses. Although ego centric data are somewhat limited, they still provide powerful measures of interpersonal influence that are strongly predictive of behavior. The chapter closed with a discussion of the application of snowball sampling including network recruitment.

Keywords: egocentric, personal networks, dyadic data, network exposure

This chapter shows how researchers can collect network data from randomly drawn samples. The data only represent the respondent's personal network, usually from the respondent's perspective. The chapter then details the various measures derived from personal network data and the research questions and hypotheses that have been tested. A comparison between egocentric and sociometric data is presented. Although egocentric data are somewhat limited, they still provide powerful measures of interpersonal influence that are strongly predictive of behavior.

The network perspective emphasizes that individual attitudes, beliefs, and behaviors are often a function of the attitudes, beliefs, and behaviors of their friends, family, colleagues, and associates. Personal network exposure is the number or proportion of ties holding a particular belief or engaging in a particular behavior. Generally, network exposure is associated with adoption, and the degree of exposure required for adoption is a personal network threshold. Most of the evidence for network exposure and threshold effects comes from egocentric data, in which data on a person's social network are gathered by asking the focal individual and not necessarily interviewing his or her network contacts. Analysis of egocentric data is usually done with standard attribute-based statistical programs such as SAS, SPSS, and STATA.

(p.62) Social networks are important influences on behavior for many reasons. First, social contacts provide information about opportunities, resources, products, and about everything people want or need. Word-of-mouth communication is frequently cited as one of the most frequent channels people report on how they first heard about something or what they know about it (Van den Bulte & Wuyts, 2007). Information about a job or service travels readily through interpersonal channels. Second, social networks also provide resources typically referred to as *social capital.* The resources available in networks often consist of how-to information such as how to perform a work-related task or get something done.

Social networks also provide role models for behaviors. It is easier for people to adopt a new behavior once someone they know has done so because they see how it is done. Role modeling is an important component of vicarious learning and building the self-efficacy needed to engage in new behaviors

(Bandura, 1986). Finally, social networks can provide the support needed to continue adopting a new behavior even when it becomes difficult or challenging to do so. Thus, measuring social networks provides an important tool needed to understand human behavior. Personal networks can be measured using egocentric techniques and sociometric ones. This chapter explains the egocentric techniques.

Burt (1984) proposed a set of egocentric questions that were included in the General Social Survey (GSS), fielded in 1985 by the National Opinion Research Corporation. The survey consisted of a random digit dial sample of U.S. households and included for the first time a measure of social networks. After extensive pilot testing, Burt proposed the survey items reproduced in Appendix C. These questions were designed to measure Americans' close personal networks derived from responses to the name generator "Who do you talk to about important matters?" There are other questions one might use to generate egocentric names such as "Who do you talk with most frequently?" or "Who are your closest friends?" depending on the specific research questions and settings. The respondent only need provide the first names, nicknames, or initials since the researcher will not attempt to contact the persons named. The named persons are often referred to as *alters*.

Once the names are generated, the researcher asks a series of questions about each person named. For example, researchers generally measure sociodemographic characteristics of each alter, such as their gender, ethnicity, age, marital status, and how the respondent, ego, is related to each alter (e.g., family, friend, or colleague). The researcher can then assess substantive issues specific to the research, such as whether the person supports a political candidate, is a smoker, uses substances, practices safe sex, supports gun control legislation, or other research issue. Once the personal networks have been measured, the researcher can characterize people's immediate close **(p.63)** social networks and determine whether the network characteristics are associated with substantive phenomenon.

For example, the GSS 1985 data were used to characterize Americans' core social networks (Marsden, 1987). Marsden (1987) showed that Americans generally had 3.0 (standard deviation [SD] = 1.7) close contacts and these contacts were quite homogeneous in the sense that people associated with others of their same ethnicity, age, and education levels. Marsden also showed that urban Americans had more heterogeneous networks than rural ones.

When constructing an egocentric survey, it is typical to measure the following characteristics:

1. Strength of relationship (e.g., closeness, acquaintance, stranger; how long known)
2. Frequency of interaction (e.g., how often talked to, how often consulted)
3. Type of relation (e.g., family, friend, coworker)
4. Socioeconomic characteristics (e.g., educational attainment, wealth, income)
5. Demographic characteristics (e.g., age, residential location)
6. Substantive characteristics (e.g., smoke, practice safe sex, practice family planning, support a candidate)
7. Content of communication (e.g., discuss politics, health, child rearing) or risk behavior (e.g., unprotected sex, share syringes)

Table 4–1 shows the kinds of measures derived from egocentric data. Typically, these measures are of one of two types: compositional measures and variance ones. Compositional measures are those derived by counting or taking the average of egocentric network variables. For example, the number or proportion of males in the personal network is a compositional variable.

Table 4–1. Egocentric Network Measures

Level of Measure	Example	Composition	Variance (Heterogeneity)	Population-Level Variance
Binary (0/1)	Smoking	Percent in positive category	IQV	
Nominal	Ethnicity	Percent in reference category	IQV	
Ordinal	Education	Average	SD	Mean of SD
Interval	Age	Average	SD	Mean of SD

SD, standard deviation.

(p.64) For a behavioral example, the number of smokers in the personal network is a compositional variable. Variance measures are those derived by calculating the variance or SD of the egocentric network variables. For example, the SD of the age of the alters is a variance measure.

All of the measures taken on personal networks can be collapsed into compositional and variance measures. For example, the number or percentage of females and the IQV of the gender variable can be calculated. Each individual in the data can be characterized as the degree his or her personal network is female and the sample characterized as the extent to which the percent female varies. For relation type, the proportion or number of the personal networks who are family members and the diversity of that proportion can be calculated.

For binary variables, the proportion and variance will be highly correlated (although not linearly) since the variance is greatest when the networks are similarly distributed across the categories. For example, suppose a person has four friends—two male and two female. The personal network composition is 50% female with a maximum IQV. For other respondents with a larger percentage who are female, IQV decreases. For nominal, ordinal, and interval variables, the compositional and variance measures are independent. Ethnicity, for example, might be calculated as the percentage of the personal network who are Hispanic/Latino and the variance measure indicates whether the distribution of alters across other ethnicities is evenly distributed (high variance) or skewed to one or few other categories (low variance).

Likewise with interval-level variables such as the network alters' age, researchers calculate the average age of the network and its SD. These network variables can be used to explain individual behavior. For example, it might be hypothesized that adolescents with older and more varied (on age) personal networks are at greater risk for substance use or other risky behaviors. Adolescents with older friends, and some much older friends, are probably exposed to more risk behavior such as smoking friends or having friends who engage in unprotected sex.

Table 4–1 also has a column labeled population-level variance. This column indicates that the mean of the SD is a measure of the overall population variance. That is, the average of the personal networks' SD indicates how much the population network varies. A large average SD indicates that overall people are quite varied in the ages of their friends, whereas a small average SD indicates that people named others who all had similar ages.

In sum, the egocentric network nominations are used to characterize individual's personal networks. There are seven types of network questions asked to gauge these personal networks. Compositional and variance measures from these seven types are derived from these variables, and they can be used to describe personal networks of respondents and determine whether **(p.65)** personal networks are associated with behaviors. In addition to the compositional variables, there are several standard measures calculated from the personal network data. Typical egocentric variables are size, personal network exposure, tie strength, concurrency, density, and constraint.

Measures

Size

A basic yet critical variable in network analysis is individual network size. Some people have small networks, while others have quite large ones. Although techniques exist for measuring a person's network size (Bernard et al., 1987), in egocentric research, size is simply a count of the number of names or nicknames provided in response to the name generator. Often size will vary from zero to 5 or 6 since name generators often limit the number of persons named, because each name will also require additional information to be recorded. Providing more names lengthens the survey and naming too many names can make a survey prohibitively long.

Some researchers have suggested allowing the respondent to name as many persons as he or she can or at least increasing the limit to 15 or 20 to measure size better. It may also be desirable to increase the size and only ask one substantive question for all alters named, and then administer the full battery of network questions to the first five named. This allows the researcher to characterize the close personal networks in detail and also understand the attitudes or behaviors of a larger number of the respondent's network, including the behavior of weak ties. It has also been suggested to allow the respondent to name many alters and then ask the network questions for a random sample subset of those persons named. This later technique enables the research to study network effects beyond the close (strong) ties.

Personal Network Exposure

Personal network exposure is the degree to which a focal individual's alters engage in a particular behavior. Egocentric data can also be used to show that people who engage in certain behaviors are more likely to have close personal network associates who also engage in those behaviors. For example, in a study of factors associated with the adoption of contraceptive methods in Bolivia, people who reported using a current contraceptive method had 63% of their personal networks who also used contraception, while those who did not use reported only 37% of their personal network used contraception ($p <.001$, $N = 5{,}691$). This difference persisted but diminished when the respondent's partner was removed from the calculation (54.4% versus **(p.66)** 38.3%, respectively; $p <.001$, $N = 4{,}156$). The overall rate of contraceptive use (both traditional and modern methods) in these data was 55.7%. Thus, those who use contraception to limit their family size are more likely to have people in their network who also use contraception. This finding is not surprising given that people usually associate with others like themselves (the homophily principle) and that information about contraceptive availability and use would pass through social networks. Egocentric data can tell us something about how networks influence individual decisions. (Commands for calculating network exposure using egocentric data in STATA are available from the author.)

Personal network exposure is a fundamental and critical variable to be calculated in network research. It captures social influence by measuring the extent one's network engages in a behavior. There is often an assumption that the networks influence the respondent given that the respondent knows these alters engage in the behavior. This social influence and the assumption that people are influenced by their peers can be tested in at least two ways. First, social influence can be tested by asking the respondent to indicate whether he or she was influenced by each alter. For example, after eliciting network nominations to measure social influence on smoking the researcher can ask, “Did this person (name) offer you a cigarette?” Second, social influence can be tested by weighting exposure by tie strength or communication frequency. In this case, the behavior of the tie is multiplied by the frequency of communication or contact.

Tie Strength

A classic finding in the network field is the importance of weak ties and bridges for connecting different groups so that information and behavior can spread throughout a community or population (Granovetter, 1973). Strong ties, however, are important for disease transmission and behavioral adoption at the individual level. Strong ties are important because people are more likely to be influenced by those they are close to and have multiplex relations with than those to whom they are weakly connected. These weak ties are important at the global level but much less so as an individual makes a behavior change decision. It should also be noted that weak ties may be more important for information spread rather than behavior change since it is easy for individuals to accept or learn information, whereas behavior change is a more complex and cognitively challenging process. Weak ties are effective at transmitting information but less so for transmitting behavioral influence.

Valente and Vlahov (2001) conducted a study to test the importance of strong ties for social influence in the evaluation of the Baltimore Syringe **(p.67)** Exchange Program (SEP). In that study, participants were asked to provide the initials or nicknames of their five closest friends. For each friend named, participants indicated whether they engaged in any risk behaviors with each friend such as drinking alcohol, having sex, doing drugs, or sharing syringes. The last point was of particular concern because the needle exchange program is designed to reduce syringe sharing to reduce the spread of HIV. In another part of the survey, we also asked whether they had shared syringes with anyone in the past 3 months.

As Figure 4–1 shows, the amount of syringe sharing was about 30% at baseline and decreased as participants returned to the SEP and completed subsequent interviews. The decrease in syringe sharing is evidence of program effects for those who repeatedly return to the SEP and received behavior education. Figure 4–1 also shows the amount of syringe sharing with friends by the rank order of the friends named. The assumption is that friends nominated first are the closest friend, and the one named second, the next closest, and so on. The data show that syringe sharing—risk behavior—was more likely with friends named first or second rather than those named third, fourth, or fifth. This makes sense as the Baltimore injection drug users established bonds with these close friends and the trust relationship is solidified with syringe-sharing behavior. Further, these close relationships may also involve sexual relationships; thus, the syringe sharing is only one form of risk behavior for the dyad. Risk behavior among strong ties occurs in other studies as well. For example, studies have shown that adolescents are more likely to smoke if their best friend smokes (Alexander et al., 2001; Urberg et al., 1997).

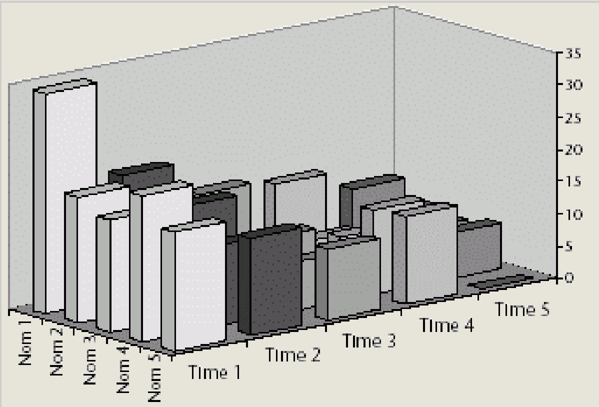

Figure 4–1. Reported syringe sharing among injection drug users by the order nominated (Nom 1 = first named) and survey wave (Time 1 = baseline). Injection drug users reported more syringe sharing with those first named, presumably their closest friends.

(p.68) Serial monogamy or serial risk taking may seem a reasonable strategy to protect oneself from disease, particularly sexually transmitted diseases, but there is still some risk associated with the behavior. This is particularly true for injecting drug users, who may not have stable residences. In the Baltimore SEP, we found that only 29% of the close friends named at baseline were named 6 months later. This rapid turnover in networks means that the rational strategy of engaging in risk behavior only with closest friends loses its protective effect when those close friends change often. This keeps HIV spreading in the population.

Concurrency

Serial monogamy of course is not always the rule. Kretzschmar and Morris (1996; Morris & Kretzschmar, 1997) hypothesized that many people might engage in concurrent relationships (Figure 4–2). Concurrency occurs when a person is engaged in sexual or other risk behavior with multiple others within the same time frame. For example, node 1 in Figure 4–2, who has three serial sexual partners, offers less opportunity for disease spread in the community than node 2, who has three sexual partners but where the second partner overlaps with the first and third.

Concurrency is measured by asking respondents to provide information on the duration of the sexual relationships with all of their sexual partners. This is not overly burdensome, and many studies have shown this is possible (Morris, 2004). The researcher then calculates the relationship intervals and the degree of overlap between alters. Each person can then be characterized to the extent their sexual relationships overlap. Moreover, the extent of concurrency for the sample, and hence in the community, can be calculated.

There are other egocentric measures in addition to size, network exposure, tie strength, and concurrency. It is common, for example, to ask the respondent to indicate whether his or her friends know one another and how well they know one another.

(p.69)

Density and Constraint

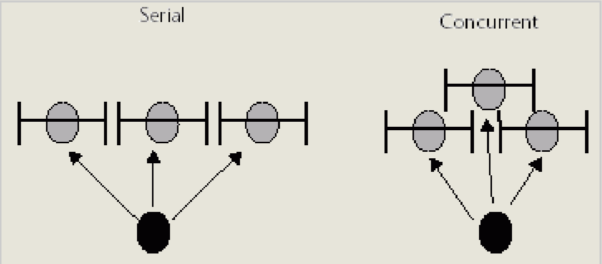

Figure 4–2. Sexual relations can be serial or concurrent. Serial relations occur in sequence with no overlap whereas concurrent ones contain overlap.

The survey example in Appendix C asks respondents to indicate which of their friends know one another. These data can then be used to construct a personal network density variable that reflects the extent a person's closest contacts are connected to one another. Figure 4–3 shows two personal networks, one that has some interconnectivity among alters (dense), and the other with alters that connect to others outside ego's immediate network (radial). Dense personal networks provide reinforcement for prevailing norms and practices and can provide protection from outside sources of influence or risk. Conversely, radial personal networks provide access to more information or influence that may be circulating in the network. Radial networks can be advantageous or disadvantageous depending on the topic or behavior being studied and its prevalence in the community or network.

Burt (1992) introduced a personal network measure called *constraint* that extends the personal network density measure by extracting more information from the pattern of relationships among the alters. Personal network density calculates the degree of connectivity among the alters nominated by ego. Constraint, on the other hand, measures the connections between alters from each alter's perspective. A constrained ego is one which the alters are connected

(p.70) to each other and so ego's actions and perceptions are controlled by his or her personal network. In an un-constrained personal network, the alters are not connected to one another and so the personal network cannot collude to keep information from ego. Constraint is calculated as follows (Burt 1992, p. 55): 4–1

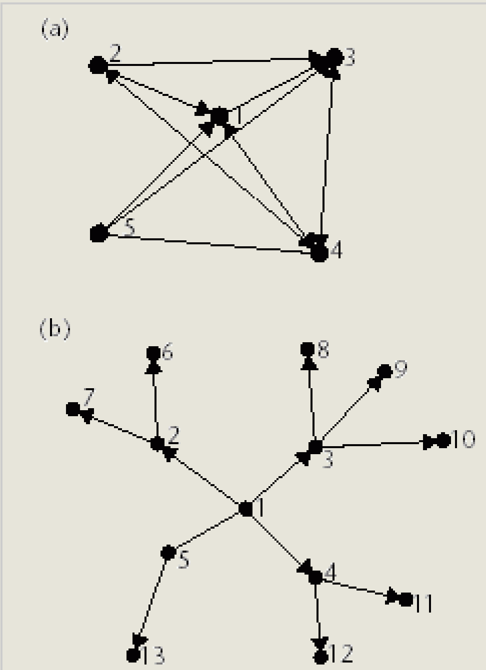

Figure 4–3. Personal networks can be dense in which the focal person's friends know one another (a) or radial in which they are not connected. (b) Burt (1992) defined radial networks as those with low constraint.

$$C_i = \left(p_{ij} + \sum_q p_{iq} p_{qi}\right)^2, q \neq i, j$$

The calculation for constraint entails summing the degree to which each of the alters is connected to others in the personal network. Conceptually, the measures (density and constraint) are similar, but the key distinction is that constraint uses more of the information available in the personal network. Burt equates low constraint with occupying a position of structural holes. People with low constraint can access their networks better and span structural holes in the networks. Burt (2005) has shown that low constraint is associated with professional achievement and better performance.

Statistical Analysis

The measures and analyses mentioned thus far in this chapter are derived from data that could be collected randomly; that is, the respondents are not connected in a preexisting social network. Statistical analysis of associations between network variables and outcomes do not need to control for nonindependence of observations other than what would normally apply to the data. In a study of reproductive health practices in Bolivia, respondents were randomly selected households in the seven largest cities. Evaluation of communication campaign designed to promote reproductive health services was conducted using normal inferential statistical analysis.

In addition, egocentric network data were collected by asking respondents to name up to five people with whom they discuss personal issues. A series of questions were then asked about those named, including whether the alters knew one another, their gender, language spoken at home, frequency of communication, and whether the respondent thought each practiced family planning. For all of these questions, compositional and variance variables can be created from the data and treated as individual-level variables in statistical analysis.

Dyadic Data

As mentioned in Chapter 3, network data are often stored or converted to a dyadic format. *Dyadic data* refer to observations in which each case consists of the respondent and one network alter. In other words, in dyadic data, each case is a relationship pair. Dyads are the respondent (ego) alter pair, and dyadic data are the data associated with the respondents and alters. Sociometric data are often stored in dyadic format, and this is sometimes **(p.71)** referred to as a *link-list format*. Egocentric data are often converted to dyadic format to facilitate statistical analysis. In addition, once in dyadic format, it is possible to calculate network exposure by merging on the alter ID and reshaping the data back to the original wide format (Box 4–1).

Box 4–1. Calculating Exposure

It is possible to calculate network exposure within statistical analysis programs such as SAS, SPSS, or STATA without relying on matrix manipulation or specialty network programs. To calculate exposure, the researcher converts a standard dataset to dyadic, or “long” in STATA terminology. The researcher then merges the dyadic data on the alter ID number (or name) with a dataset containing the behavior or attribute of interest. Now the researcher has a dyadic dataset containing the ego ID, the alter ID, and the alter's attribute. Converting the dyadic data back to a “wide” dataset and calculating the average of the alter behavior provides a network exposure score.

```
/* start with the original wide version of the dataset
and reduce to the unique subject ID the attribute of
interest and alter nominations */

use c:\data

keep net_id attrib nom1 – nom5

reshape long nom, i(net_id) j(alter 1-5)

sort net_id
```

```
drop if nom==.
save c:\dyadic, replace
use c:\data
keep net_id attrib
ren net_id nom
ren attrib alter_attrib
sort nom
save c:\ego_as_alter, replace
use c:\dyadic, replace
sort nom
merge nom using c:\ego_as_alter
sort net_id
drop if _merge!=3
drop _merge
save c:\dyadic_alter_beh, replace
reshape wide nom alter_attrib, i(net_id) j(alter)
save c:\ego_with_alter_behavs, replace
egen attrib_expos=rmean(alter_attrib1 alter_attrib2
alter_attrib3 alter_attrib4 alter_attrib5)
```

(p.72) If the data are converted to dyadic format, the observations are no longer independent but are clustered on the respondent. Some respondents may have provided information on one alter while others provided information on four or five alters. Consequently, the analysis needs to control for clustering on the respondent ID. This is not difficult and usually entails specifying a multilevel or hierarchal model to explicitly account for the clustering. Sociometric studies, on the other hand, use data from a saturated sample such as a school or organization and so the statistical issues involved in estimating network effects are more complex (see Chapter 9).

Converting a regular attribute dataset to dyadic has certain advantages (Chapter 3 discussed how to do this). Because each case is the respondent and the data on the alter he or she named, analysis can be conducted to test how relationship characteristics are associated with behavior. For example, one might ask survey respondents to name their five closest friends and then ask for their gender, age, religious beliefs, and smoking behavior. The researcher can test whether people report smoking more with same-gender friends versus non–same-gender friendships.

It must be emphasized that dyadic data are not independent, that each case is not randomly selected from the population; rather, the cases are clustered on the respondent. In many cases, there may be multiple levels of clustering such that the cases are clustered by survey wave (e.g., baseline and follow-up) and the respondents. Researchers should use hierarchical linear models or random effects models or some other technique that controls for the nonindependence of the data.

Personal Network versus Sociometric Variables

Personal network exposure using egocentric data is usually easy to calculate in a statistical package, but this variable can be limiting when the researcher suspects that many respondents may not accurately know their alters' opinions or behaviors. These misperceptions can arise for many reasons. First, the respondent may not know his or her alters' attitudes and behaviors. Second, the respondent may purposively provide inaccurate information on his or her alters' attitudes and behaviors. Third, the respondent may misperceive the alters' opinions and behaviors because he or she wants them to be congruent with his own (reducing cognitive dissonance). Fourth, ego may misperceive the alters' opinions and behaviors because the alters tell ego things to pretend they hold a different opinion than they actually do. For example, people may say they support a political candidate in order to avoid an argument over political beliefs.

For example, personal network exposure was calculated on contraceptive use for women from voluntary associations in Yaoundé Cameron (Valente **(p.73)** et al., 1997). Women were asked to indicate whether they had heard of family planning

(FP) and any methods available to delay having children and to name their closest friends in their group. For each woman, knowledge of FP methods and use of a method (if any) was calculated. The methods her friends knew and used was also calculated (network exposure). Women were more likely to be aware of the same methods their friends knew and more likely to use the FP method their friends used.

Participants were also asked if they knew whether their friends used modern or traditional FP methods. Since the friends' reports of method use were also recorded, we could calculate whether the participant was correct or incorrect in her assessment. Finally, whether the respondent thought those friends encouraged them to use methods was also recorded. The study showed that perceiving friends to have encouraged use was the variable most strongly associated with method use, regardless of whether one's friends use, and regardless of whether one was correct or incorrect in her assessment of friends' use. This study highlighted the importance of perceived peer influence and that people may to some extent justify their actions by believing that others support and encourage those actions.

These analyses were possible because we recorded the participants' behaviors, asked who their friends were, and had those friends' self-reports. It is desirable to have alters' self-reports to calculate exposure to behaviors rather than relying solely on the participants' perceptions (Ianotti & Bush, 1992; Rice et al., 2003). And as will be illustrated, the basic diffusion model is one in which behaviors flow through networks such that network exposure is the fundamental building block for understanding networked diffusion and communicable disease spread. The egocentric results reported here, however, complicate analysis of personal network effects.

They indicate that tie strength may matter for some behaviors, meaning we possibly need to weight the exposure calculations by tie strength (Bauman et al., 2007). Weighting by tie strength is computationally easy, but substantively difficult because there are so many candidate weights to choose from. For example, one could weight ties by the frequency of contact or perceived emotional closeness. One could also weight on similarity of personal attributes such as stronger influence by those of the same gender, age, socioeconomic status, or any other attribute. There is a nearly infinite number of ways to include tie strength in exposure calculations.

Researchers have also proposed a threshold model in which people vary in the degree they are influenced by others in their network (Valente, 1996). Researchers may have incorrectly concluded there are threshold effects because there was an incorrect specification of the exposure that did not include correct weights for tie strength. Thus, there is tension between the proper exposure specification and threshold effects. If network exposure **(p.74)** is not significantly associated with behavioral adoption, it may be due to threshold effects or inadequate modeling of exposure weights.

Snowball/Sequenced Data

Egocentric data provide a view of the respondent's network from the respondent's perspective. Once the names are generated, however, the researcher can create study designs in which the members of the personal network are also interviewed. This is referred to as snowball sampling and occurs when the researcher asks the respondent to indicate their friends (or other network members) and asks the respondent to recruit their friends. Snowball study designs are used in two types of studies: (1) when the researcher wishes to track behavior or communications among network members and (2) when using initial index cases to recruit their network members into an intervention or study.

There are two kinds of "snowballs"—one in which the interview attempts to contact and interview every one of the respondent's alters (all of their network) and the other in which some specified subsample of the personal network is contacted. When all members of the personal network are contacted, this is a "true snowball." The sample sizes in a snowball can grow quite rapidly. For example, 10 index cases can provide 10 names yielding a pool of 100 subjects who can also provide 10 names, resulting in 1,000 subjects in only two steps. Of course, the name generator or network definition might involve a concept that generates only few connections and this can slow the growth of the snowball. For example, the study might measure only current sexual contacts, so 10 indexes who have two sexual partners yields only a pool of 20 new subjects.

The second type of snowball entails interviewing some subset of each respondent's personal network; this sample can be randomly chosen in an effort to make estimates regarding a population parameter. The sample can also be defined as those who are closest to the respondents, thus capturing strong rather than weak ones. Most often, the researcher is interested in a specific type of tie and will follow up only with those. For example, the researcher might want to interview all the drug users in a community and so elects to have indexes provide the names of his or her friends who use drugs. This provides entrée into the drug-using community, and further interviews will yield the structure of the drug using network.

Klovdhahl (1989) suggested eliciting a high number of alters and then taking a random sample of them to see how the network grows. Because the alters are randomly chosen, study results provide valid estimates of network properties attributable to the network. After one or two generations, **(p. 75)** researchers will have a large pool of subjects in the study who will know one another and hence be linked into a network. If the researcher starts with a set of randomly chosen index cases, then the network results provide a potentially valid parameter estimate of network structure.

Snowball sampling can be used to verify respondent assessments of their network alters' behaviors. For example, a snowball study of adolescent smoking can be conducted in which the ego and alter are both asked about their smoking behavior and that of their peers. If these self-report measures are validated with a biomarker (say cotinine), the researcher can then determine whether respondents are accurate in their assessments of their network partners' behaviors. It may also be possible to use snowball sampling to follow the trail of a rumor or piece of gossip or information. For example, if respondents report that they first heard a news item from people in their network, these alters can be interviewed and a trace of the information followed. These applications, however, are somewhat limited, and consequently few examples of snowball studies exist. More frequently, however, snowball sampling has been used to recruit subjects into studies or interventions.

Networks for Recruitment

Snowball techniques can be particularly useful for recruiting people into studies or health promotion and disease prevention programs. Network recruitment occurs when individuals are identified via outreach, clinical, service, or other methods, and these index cases are instructed to identify network partners that can also participate in the study or receive some program. For example, many studies have been conducted among drug users in which people receiving treatment or participating in a study are invited to bring their friends or substance using partners to the study setting so they can also receive treatment.

The idea of using social networks to identify people at risk for disease is not new and dates back to the use of contact tracing by local health departments. When individuals were diagnosed with a sexually transmitted disease (STD), workers at the health department would ask that person to name those they had had sexual relations with over the past 3 or 6 months and to provide their addresses. The workers would then contact those sexual partners and inform them that they are at increased risk for an STD and should be tested. These sexual network contacts were then recruited for STD testing and treatment and also asked to indicate their sexual networks if they had an STD.

More recently, interventions have been conducted in which individuals are invited to bring their network partners, friends, or close associates into a clinical or outreach setting to be given health promotion materials/interventions. For example, Valente and others (2009) set out to determine if a clinic **(p. 76)**

population of HIV-positive patients could be used to create a cohort of high-risk individuals who might be appropriate to receive an HIV vaccine. Figure 4–4 displays network linkages among indexes and alters in which at least one alter was enrolled in the study (794 links, 59.2%). Squares represent the 59 index patients; triangles, the 62 enrolled alters; and circles, named but not enrolled alters. Links are coded by willingness to invite alters to participate in vaccine preparedness activities: solid arrows represent willingness; dashed arrows, not willing; dash-dotted arrows, absence of a response; and dash-dot-dot, enrolled but not named first-degree alter. The study showed that these indexes could enroll a cohort with the desired characteristics, but the networks were quite diverse.

Figure 4–4. Networks of HIV-positive patients in which at least one alter was enrolled in an HIV vaccine preparedness study. Indexes are squares, enrolled alters are triangles, and nonenrolled alters are circles. Links are coded by willingness to invite the alter to participate in vaccine preparedness activities: solid arrows represent willingness, dashed arrows, not willing; dash-dotted arrows, absence of a response; and dash-dot-dot, enrolled but not named first-degree alters.

The rest of this book will cover sociometric data and techniques used to analyze data from complete networks. It is wise to remember, however, that the complete network is composed of egocentric networks and that the sociometric data can be treated as egocentric. The building blocks of the sociometric data are the individual egocentric networks.

(p.77) Summary

This chapter presented research on egocentric network data. Egocentric data are collected by asking people to name their closest friends or people they are connected to in some way. Egocentric data do not provide a connected network that can be mapped; rather they provide data that can be used to characterize each person's personal network environment. Egocentric data have been used to characterize the personal networks of populations. For example, data from the 1985 GSS were used to show that Americans have ties mostly to people like themselves (homogeneous) and that networks vary by geography (urban respondents had more heterogeneous networks than rural ones).

Many hypotheses regarding egocentric variables and behavior have been proposed and tested. Researchers have shown that behaviors are influenced by tie strength. For example, injection drug users were more likely to engage in risky behavior with close friends than with those less close and adolescents were more likely to smoke if their best friends smoke. Researchers have also shown that degree of concurrent sexual relationships in a community affects the prevalence of HIV. The chapter discussed how to convert egocentric data to a dyadic dataset, thus facilitating analysis and testing of certain hypotheses. The chapter closed with a discussion of the application of snowball sampling including network recruitment. **(p.78)**