Social Networks, 3 (1982) 225–249

Hypernetwork Sampling: Duality and Differentiation Among Voluntary Organizations*

J. Miller McPherson
*University of South Carolina***

One of the recurrent problems in the study of social organization is that survey researchers are forced to study individuals, while they are often interested in hypotheses about organizations or systems of organizations. This paper offers some methods which allow survey researchers to address hypotheses about the relations generated among individuals by organizations, the organizations themselves, and the relations among organizations. Estimators are developed for the number and size distribution of organizations in communities, the density of relations among individuals generated by organizations, the number and density of interorganizational linkages, and the amount of membership overlap among organizations. This approach combines Granovetter's idea of network sampling with Breiger's notion of the duality of persons and groups to produce a quantitative approach to the study of voluntary organization. We provide an illustration of the methods we develop by testing two of Blau's structural hypotheses on data concerning voluntary organizations. The hypotheses that structural differentiation increases with system size and that this increase is at a decreasing rate are strongly supported. The paper closes with some suggestions for further elaborations of the hypernetwork approach.

Introduction

Since the watershed years of the early community studies such as those of Warner and Lunt (1941) and the Lynds (1929, 1937), two separate research traditions have existed in the voluntary association literature: organizational case studies and surveys of individuals. These two traditions have never been

*The author is indebted to anonymous readers and the Structuralist Group (Professors L. Smith-Lovin, Bruce Mayhew, Patrick Nolan, Ken Shin and John Skvoretz) at the University of South Carolina for their comments. Assistance from Edward Nevius of the Department of Mathematics, University of Nebraska in early stages of this paper is gratefully acknowledged. Any remaining defects are the responsibility of the author. Data were provided by the Nebraska Annual Social Indicators Survey, Bureau of Sociological Research, University of Nebraska-Lincoln.

**Department of Sociology, University of South Carolina, Columbia, SC 29208, U.S.A.

well integrated since the discipline moved away from studying whole community systems. An example of the gap between the two is the survey literature's endless argument over which ethnic category has the higher participation rate (Wright and Hyman 1958; Babchuk and Thompson 1962; Orum 1966; Olsen 1970; Williams *et al.* 1973; Antunes and Gaitz 1975), while the case study literature focuses on elaborate descriptions of specific organizational systems (Zald and Denton 1963; Zald 1970; Young and Finch 1977).

The two streams of research have entirely different emphases and metatheoretical assumptions: the abstract empiricism of survey research on voluntary affiliation mixes poorly with the grounded theory of the case studies of organizations. Yet, there are some limited areas in which the two traditions touch each other. Clearly, the memberships studied in the aggregate by the survey researchers come together to form the organizations studied in the concrete by the case studies. The characteristics of the individuals who appear in probability samples must also appear in the organizations to which they belong. This fact raises the question of what the survey data on individuals can tell us about the system of organizations. Our basic goal in this paper is to develop some simple mathematical relationships between the two streams of research; it is shown how it is possible to study certain characteristics of organizations and the social networks that they generate with survey data on individuals.

We have no reliable information on even the average number of voluntary organizations in our cities, much less any systematic information on the social networks which create, and are affected by, voluntary associations. The published figures are based on either complete enumerations from small towns (Laskin 1961), or incomplete censuses in larger cities (Johnson 1977); estimates range from about nine to over sixty organizations per thousand persons (cf. Babchuk and Edward 1965). Likewise, we have almost no systematic information on the sizes of these organizations. Organizational size is an extremely critical variable because it determines a number of characteristics of the organization such as internal differentiation, salience in the community, and other important features (McPherson and O'Donnell 1980), as well as setting the stage for acquaintance formation within the organization.

Our approach in this paper is to view the members of face-to-face organizations as groups with heightened probability of contact. Acquaintance networks have been studied in the past through the small-world technique (Milgram 1967; White 1970), through the study of egocentric networks (Burt 1980), and through various modified survey techniques (Lin *et al.* 1981). We view voluntary organizations as settings in which potential acquaintances are formed through common membership. As Mark Granovetter points out in his seminal paper on social networks (1973:1375), formal organizations are a common source of weak ties, through which much important information about jobs and other opportunities passes. Thus, the links among members of voluntary organizations are an important and neglected topic for research.

Common membership in an organization is a structural feature which increases the probability of acquaintance formation, but it can also be useful and important *sui generis.* Common membership in organizations has often been recognized by sociologists as a critically important social characteristic (e.g. Baltzell 1966). In fact, the notion of intersecting social circles is one of the earliest structural concepts to appear in sociological theory (Simmel 1902, 1922). We believe that it is unfortunate that the methodological individualism (Webster, Jr. 1973) produced by an emphasis on survey research has led research on voluntary associations away from these key ideas. We propose to develop methods for survey research which move back in the direction of the study of structural properties.

Thus, we are interested in the potential relationships among individuals created by organizations, the characteristics of organizations, and the relationships among organizations. We use techniques borrowed from a tradition of the study of structural characteristics with survey data. This literature began with a seminal paper by James Coleman (1958), who named his technique relational analysis. Ove Frank (1971) showed that statistical sampling theory could be applied to graphs and networks. Most recently, Mark Granovetter (1976) indicated how network sampling was applicable to sociological problems. We extend this tradition to study properties of networks which have two levels—the inter-individual level and the inter-organizational level. We are interested in some very simple characteristics of both levels, such as the density of potential contacts among individuals generated by common memberships in organizations, the size and number of the organizations, the number and density of connections among organizations, and the amount of membership overlap among organizations.

The basic conceptual device we use has been called a dual network (Breiger 1974), and is known in the mathematical literature as a hypernetwork (Berge 1976). These hypernetworks are generated whenever there exists a set of elements (in our case, individuals) linked together by a set of relations (common membership in organizations), when that relation also defines a new set of elements (organizations) which are linked by relations defined by the first set of elements (individuals). The hypernetwork can be viewed as though it consists of individuals linked by common membership in organizations, or of organizations linked by individuals who belong to several organizations at once. This imagery allows us to investigate properties of the system on both the individual and organizational levels, using information only from the individual level.

As an example of such cross-level inference, we take the case in which we know only the number of organizations to which an individual belongs. Since this individual is common to each of those organizations, he or she represents a link between them—a shared resource, a flow of information, a path of communication, and so forth. A person with three memberships thus produces three links among three organizations. Thus, if we simply know the number of memberships for each of the individuals in a representative sample of people from a community, we can estimate the total number of links among organizations in the system without any direct information on the organizations themselves. In this way, then, we can begin to get insights into the system of organizations from information gathered only from individuals.

The model

Breiger (1974) has discussed the notion of a binary (zero–one) adjacency matrix of persons to groups. This matrix (called the adjacency matrix of a hypernetwork, or dual network) completely specifies the membership structure of a system in which individuals may be members of multiple groups. Figure 1 presents such a matrix (*M*) for a hypothetical set of *N* individuals and *K* groups. Each row in the matrix represents an individual; the columns are organizations. Thus, individual 1 belongs to organizations 1, 2, and *K*, while organization 3 has among its members individuals 2 and *N*. The power of this representation lies in the fact that the matrix may be manipulated to examine the individual level, the organizational level, or the inter-organizational level. We will examine all of these levels.

Figure 1. *The matrix M.**

$$
\begin{array}{r c l}
 & \left[\begin{array}{cccccccc}
1 & 1 & 0 & \bullet & \bullet & \bullet & \bullet & 1 \\
0 & 1 & 1 & \bullet & \bullet & \bullet & \bullet & \bullet \\
\bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet \\
\bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet \\
\bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet \\
\bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet \\
1 & 0 & 1 & \bullet & \bullet & \bullet & \bullet &
\end{array}\right] & \begin{array}{c} m_1 \\ m_2 \\ \bullet \\ \bullet \\ \bullet \\ \bullet \\ m_n \end{array} \\
\text{Column Sums} & \begin{array}{cccccccc} s_1 & s_2 & \bullet & \bullet & \bullet & \bullet & \bullet & s_k \end{array} &
\end{array}
$$

N Individuals

*Each row represents an individual. Each column represents an organization. The row-sums are the total number of affiliations for the given individual. The column sums are the organizational sizes.

The total number of organizations to which an individual belongs is the sum of the row corresponding to that individual. This sum represents the number of connections between that individual and the organizations in the system. This quantity is the most studied variable in the voluntary association literature; it has been correlated with everything from social class (Bell and Force 1956) to self-esteem (Aberback 1969). It is easy to see why this variable is so useful; it summarizes the integration of the individual into the system of voluntary organizations. However, a very important limitation of this measure is that it summarizes the connection of the individual to the organizations only in very general terms. Clearly, when we add across the ith row to get the number of organizational memberships for the ith person, we lose the information concerning which organizations are involved. This fact has long been a great problem in the literature: the measure of affiliation aggregates across organizations of unknown size, with unknown other members. For our purposes, the critical characteristic of these organizations is size,[1] since the size of an organization determines the probability that it will appear in a survey sample of individuals. As we will see, size also determines a number of other structural characteristics.

Structural characteristics

Define $m_{i.}$ as the number of memberships possessed by the ith individual (the ith row-sum of $\boldsymbol{M}$). The total number of memberships in the system will be the simple sum of $m_{i.}$ across all N individuals:

$$T = \sum_{i=1}^{N} m_{i.} \tag{1}$$

Clearly, T is a population parameter; a survey researcher will be interested in estimating T from a sample of individuals. For a simple random sample, such an unbiased estimator would be

$$\hat{T} = \frac{N}{n} \sum_{i=1}^{n} m_{i.} \tag{2}$$

where n is the number of cases in the sample. Now, T also defines a new population—a population of memberships. T can be either smaller or larger than N, depending upon the rate of affiliation in the population of individuals. The relationship between T and the mean rate of affiliation $\bar{m}$ is

$$T = N\bar{m} \tag{3}$$

since $\bar{m}$ is the sum of $m_{i.}$ divided by n. Thus, to estimate the total volume of memberships in the system, we need to know only the mean rate of affiliation and the numbers of individuals in the sample and population. Note that T is always larger than N when $\bar{m} > 1$.

It is important to note that the distribution of any given variable can be quite different in the membership population than in the individual popula-

[1] Another very important feature is the homogeneity of the members of the organization. If a variable correlated with affiliation, such as social class, is also the basis for selection of members into organizations, then organizations composed of high-status individuals will be located in very densely connected regions of the inter-organizational system. That is, segregation magnifies the centrality of high-status organizations. Such organizations will also be linked directly to other high-status organizations, since high-status individuals tend to have multiple memberships. In this way, exclusivity in organizational composition may operate to mirror, or even magnify, the distribution of inequality by social class in society (McPherson 1981a, 1981b).

tion, since individuals are replicated in the membership population as many times as they have memberships, while people with no memberships do not appear at all. The massive amount of substantive research on the correlates of affiliation takes on a new dimension when viewed in this light; any variable correlated with affiliation will shape the population of memberships. If Blacks have a higher rate of affiliation than Anglos, then they will contribute a disproportionate number of memberships to the system. Similarly, since the strongest correlate of affiliation in the United States is socio-economic status, the voluntary sector in this country has a disproportionate number of high-status individuals, when compared with a country like Mexico or Italy, where affiliation is less highly correlated with status (McPherson 1981b). These differences in the voluntary sector may be very important in understanding large-scale social movements (e.g. McCarthy and Zald 1977), since the voluntary sector often is a resource base for social movements (Gove and Costner 1969).

Now, the T memberships in the system are partitioned into K organizations. That is, the sizes of the K organizations sum to T, the total number of memberships. Knowing nothing about these sizes, we expect that T and K will be positively correlated. That is, we expect the size of the system to be related to the number of organizations in the system. This hypothesis is easily deduced from the extensive literature on size and structural differentiation (Spencer 1877; Simmel 1902; Hawley *et al.* 1965, Blau 1970; Kasarda 1974). We can test this hypothesis (and other structural hypotheses) for voluntary organizations if we can estimate K and T from data. Since there is no existing probability sample of voluntary organizations (for reasons which will become clear), we must try to develop estimators for T and K from data on individuals.

The key insight necessary to understand our estimator of K (the number of organizations) is that the larger the organization, the more likely one or more of its members will be picked in a sample of individuals. An organization of size 200 is twice as likely to be named by a respondent in a random sample of individuals than another organization of size 100. The larger organization has 200 'chances' to be detected in a random sample, while the smaller organization has only 100 such chances.

Since larger organizations will appear more often than small organizations in a survey of individuals, methods of estimating the number of organizations which do not take this fact into account must be ruled out. For instance, one such approach might be to record the rate of new organizations appearing in samples of increasing size, in order to estimate K for the system. A careful analysis reveals a serious defect in this method: only very large organizations are likely to appear more than once in a sample of individuals. That is, almost all organizations will be 'new' in systems of any appreciable size. For a simple random sample of n individuals from a population of size N, an organization of size S will be expected to appear $n(S/N)$ times, since S of the N people in the system will report this organization, should they be among the n sample members. In fact, the number of reports of this organization appearing in the sample will be distributed binomially, with S/N defining p, the probability of a success. Thus, the multiple appearance of an organization depends directly on the ratio of its size to the size of the system. If this ratio is small, as it will be for systems of any substantial size, then the likelihood is that most organizations will appear only once, for any reasonable sample size.

For example, an organization of size 50 in a system of size 50 000 has a probability of one in one thousand of being mentioned by any given person. Using the Poisson approximation for small p and large N (Feller 1957), we find that the probability that this organization will appear in a sample of 100 memberships more than once is less than 0.005. Since the average face-to-face organization in our data is much smaller than 50 members, the multiple reporting of a given organization will be a rare event. Thus, this possible method of estimating the number of organizations from their multiple appearance in a sample would work only when either the organizations are very large, the system is very small, or the sample exhausts most of the population. Note also that this method would depend upon the size distribution of organizations in the system; if most organizations were small, the sample size necessary to determine the number of organizations will approach the population size.

Therefore, since the size of the organization dominates its probability of appearing in a sample of individuals, we must construct a weighting system which takes this fact into account. We can construct such weights if we get each respondent to report the size of the organizations to which (s)he belongs. In order to increase the accuracy of these reports and to ensure that common membership will be meaningful, we confine ourselves to face-to-face organizations in which the respondent is most likely to know the actual size. The fact that we rely upon individual reports of the sizes of organizations rather than a direct measure of size induces error in the weights to be attached to each observation. We analyze the effects of these errors in Appendix B.

Now, our problem is to estimate K when we do not have information on all of the individuals in the system. If we sample only n/N of the individuals (where n is the number of individuals in the sample, and N is the number in the population), an organization of size 100 will tend to appear $100(n/N)$ times. Since we expect an organization of size S in a sample of n persons from a system of size N to appear $S(n/N)$ times, the weight we assign to each report of an organization with size S is the reciprocal of this quantity, $(1/S) \times (N/n)$. Thus, to estimate the population frequency of organizations of size S, here denoted as $f(S)$, we sum the observed sample frequency of organizations of size S, weighted by the above quantity:

$$\hat{f}(S) = \sum_{i=1}^{n} m_{iS}(1/S)(N/n) \tag{4}$$

where $\hat{f}(S)$ is the estimator of the population frequency of organization $f(S)$, m_{iS} is the observed number of organizations of size S reported by the ith individual, and $(1/S)(N/n)$ is the weight which we developed earlier. Since we are interested in the total number of organizations K, we can simply sum (4) across all observed sizes to get an estimate of the total number of organizations of all sizes:

$$\hat{K} = \sum_{j=1}^{S_L} \hat{f}(S_j) \tag{5}$$

where K is the estimated total number of organizations, S_L is the size of the largest organization, and $\hat{f}(S_j)$ is the estimated population frequency of organizations of size S_j. In order to make our notation unambiguous mathematically, we may write eqn. (5) as

$$\hat{K} = \frac{N}{n} \sum_{i=1}^{n} \sum_{j=1}^{S_L} m_{ij}/S_j \tag{5a}$$

where m_{ij} refers to the number of organizations of size S_j to which individual i belongs. $\hat{K}$ is an unbiased estimator of K in the absence of measurement error in S. Appendix B of this paper analyses some consequences of measurement error in S.

To estimate other features of the size distribution of organizations we simply use the estimated frequencies from eqn. (4). For instance, the estimated mean size of organizations $\hat{\bar{S}}$ will be

$$\hat{\bar{S}} = \frac{\sum_{i=1}^{n} \sum_{j=1}^{S_L} m_{ij}}{\sum_{i=1}^{n} \sum_{j=1}^{S_L} m_{ij}/S_j} \tag{6}$$

An alternative form of (6), is

$$\hat{\bar{S}} = \hat{T}/\hat{K} \tag{7}$$

which shows that the average size of all organizations is simply the total volume of memberships divided by the number of organizations. $\hat{\bar{S}}$ from eqn. (7) is a consistent estimator of the mean size of organizations in the population whose bias is negligible for large samples (Cochran 1963:29–31). Thus, we have obtained results for estimating the number of organizations in the system, their size distribution, and their average size. We now turn to the problem of overlap.

There is practically an infinite number of ways to conceive of the relationships among organizations, ranging from legal, contractual arrangements (Zald 1970) to the much more slippery notion of 'consensual solidarity' (Turk 1977). We avoid the inherent problems of many other conceptions of inter-organizational relationships by tying our definition to multiple memberships.[2]

[2] A reader has pointed out that another potentially important way in which organizations may be related is through friendship ties between acquaintances who belong to different organizations. Since detailed study of acquaintance networks is still in the beginning stages (cf. Burt 1980; Lin *et al.* 1981; and others), we await more complete information on these links than we currently possess. We show in this paper what can be done with existing survey data on individuals and their memberships.

If one person belongs to two different organizations, then the two organizations have an overlapping membership. Thus, the number of overlapping memberships in the system is simply the number of memberships possessed by those individuals who have more than one membership. To show this symbolically, we need to define a new quantity, P_1, the proportion of individuals in the system with exactly one membership. Then an unbiased estimator of the number of overlapping memberships in the system, O, is

$$\hat{O} = \hat{T} - N\hat{P}_1 \tag{8}$$

Since $T = N\overline{m}$,

$$O = N\overline{m} - NP_1 = N(\overline{m} - P_1) \tag{9}$$

which has a lower limit of zero (when either there are no memberships at all, or when all persons with memberships have only one), and an upper limit of K (when everyone belonging to any organizations belongs to them all). If we are interested in the proportion of overlapping memberships in the system, we simply divide (9) by T to form

$$P_O = \frac{N(\overline{m} - P_1)}{T} = \frac{\overline{m} - P_1}{\overline{m}} \tag{10}$$

which is a positive function of $\overline{m}$; a consistent estimator of P_O may be formed from sample estimates of $\overline{m}$ and P_1.

Thus, the amount of overlap in the system is a simple function of the mean rate of affiliation. Since the mean rate of affiliation is the parameter most often investigated in survey research (i.e. it varies with race, sex, age, socio-economic status, city size, and a host of other variables), we have opened the door to an explicit consideration of structural features of the system based upon the accumulated literature of survey results. For instance, it should now be possible to investigate whether organizations in small towns are more likely to overlap than those in large towns, as would be suggested by the sizeable literature on urbanism (cf. Wirth 1938; Milgram 1970). We give an example of such a comparison in a later section.

Overlap among organizations generates links among those organizations, since the same person will appear in different organizational settings. An individual with $m_{i.}$ memberships generates $m_{i.}(m_{i.} - 1)/2$ links between his organizations, since the number of ways of choosing two objects from among $m_{i.}$ objects is $\binom{m_i}{2}$. For instance, a person with three memberships generates three links among these organizations; a person with four generates six links, and so on. Summing across all individuals in the system, we have

$$L = \sum_{i=1}^{N} m_{i.}(m_{i.} - 1)/2 \tag{11}$$

This statistic is the number of links among organizations generated by the individuals in the sample. To estimate the number in the population, we have

$$\hat{L} = \frac{N}{n} \sum_{i=1}^{n} m_{i.}(m_{i.} - 1)/2 \tag{12}$$

which is an unbiased estimator of L, the number of links among organizations in the community. Note that this number does not need to take into account the number of organizations, or their sizes, since it depends only on the row sums of $\boldsymbol{M}$. An alternative formula for $\hat{L}$ from McPherson (1981a) which shows its relationship to research on mean rates of affiliation is

$$\hat{L} = \frac{N}{2}(\sigma_m^2 + \overline{m}^2 - \overline{m}) \tag{13}$$

which is a positive function of $\overline{m}$ when $\overline{m} > 1.0$.

While eqn. (13) gives the number of links among organizations, we still need to discuss the density of these links. The usual way of conceiving of density is to take the ratio of the number of observed links to the number logically possible. The number of observed links may be estimated by equations (12) and (13). The number logically possible is:

$$L_u = NK(K - 1)/2 \tag{14}$$

where L_u is the number logically possible, N is the number of persons in the system, and K is the number of organizations. L_u refers to the case in which each of N persons belongs to all K organizations.[3] Thus, we now may construct a measure of the density of inter-organizational links in the hypernetwork:

$$\hat{D}_o = \frac{\hat{L}}{\hat{L}_u} = \frac{N}{n} \sum_{i=1}^{n} \frac{m_{i.}(m_{i.} - 1)}{2} \bigg/ \frac{N\hat{K}(\hat{K} - 1)}{2} \tag{15}$$

Equation (15) shows that inter-organizational density of the hypernetwork is a direct function of the number of linkages (and thus the mean rate of affiliation), and an inverse function of the number of organizations. The careful reader will note that N cancels out of both the numerator and denominator of (15). The numerator is now simply the mean number of links among organizations generated per individual, while the denominator becomes the total number of links among organizations generated by an hypothetical single individual who belonged to all organizations. Thus, the inter-organizational density measure compares the average individual's links to the theoretical maximum for the individual in the system.

On of the dominant themes in the study of voluntary affiliation has been that individuals are affected by the other members of the organizations to which they belong. Since our hypernetwork model explicitly shows the relationship between the individual and organizational levels, it allows us to construct a measure of the individual-level density of the connections generated by common membership in organizations.

The logical upper limit to the number of connections among N individuals with K organizations is

[3] We should emphasize that we are dealing with networks which allow multiple connections between points. One reader of this paper has suggested that we call this type of density 'multistranded', in the sense that the density coefficient represents the proportion of possible multiple connections which exist, given the number of organizations in the community.

$$I_u = KN(N-1)/2 \tag{16}$$

where each of the N persons belongs to all K organizations.

To estimate I_u, we take N as known and estimate with

$$\hat{I}_u = \hat{K}N(N-1)/2 \tag{16a}$$

which is an unbiased estimator of I_u.

Now, taking a simplistic view of the operation of voluntary organizations, each person in an organization of size S_j is linked to all other persons in that organization.[4] Thus, each organization of size S_j generates $S_j(S_j - 1)/2$ links. Since there are $f(S_j)$ of these organizations in the system, we can sum across all sizes to get

$$\hat{I} = \sum_{j=1}^{n} \hat{f}(S_j)S_j(S_j - 1)/2 \tag{17}$$

which expresses the number of links in the system in terms of the estimated frequencies of the organizational sizes, and the sizes themselves. If we take S_j as known, $\hat{I}$ is an unbiased estimator of I. Thus, we may construct a consistent measure of individual-level density:

$$\hat{D}_I = \hat{I}/\hat{I}_u \tag{18}$$

Contained in this definition of density is a powerful image of the ability of social structure to magnify the effects of one-to-one relationships among individuals. The usual definition of density in social networks is simply the ratio of the observed number of one-to-one relationships to the number possible. This usual definition contains the implicit assumption that social space is unidimensional; there is only one kind of tie, and it can occur only once between two individuals. The hypernetwork model explicitly acknowledges that there is a multiplicity of possible ties among the individuals in a structured social system; people move from group to group, and situation to situation. Our measure of individual-level hypernetwork density is the observed proportion of all possible such ties among individuals. High values of individual-level density imply that the individuals in the system are more likely to encounter each other repeatedly in different settings. Low values of this density suggest that two randomly chosen individuals are very unlikely to encounter each other in an organization. Thus, individual-level hyper-

[4] This simplistic view is more likely for the face-to-face organizations we study here. Individual reports of organizational size are subject to measurement error (see Appendix B). The measurement error in size is important for this paper in several ways. Firstly, the error in size reports is correlated with the size of the organization; a respondent is likely to induce more error in absolute numbers when reporting the size of a large organization. Secondly, members may respond to queries about size in terms of typical attendance at meetings, rather than in terms of actual membership. This tendency is not a disadvantage, since we are interested in face-to-face organizational settings. A preliminary analysis of differences in reported size among ordinary members versus officers (who would be more likely to respond in terms of formal membership) suggests no serious differences between the two (see McPherson and Smith-Lovin 1981, McPherson and O'Donnell 1980). Most problems related to estimating size could be resolved by using our methods to generate a probability sample of organizations, which could be contacted directly.

network density measures the integration of individuals in the system through the organizations; dense systems should promote heightened surveillance, the merging of common interests, similarities in experience, and so forth. Sparse systems would imply the reverse.

An illustration

The data we use come from the *Nebraska Annual Social Indicators Survey 1977*. Detailed information of the survey itself may be obtained from Johnson (1977). In brief, 1200 of the interviews were a simple random sample of all households in the state with a telephone. An additional 677 respondents were chosen through a multi-stage stratified area probability sample of households for personal interviews. The respondents were asked the name of each organization to which they belonged, along with detailed information on such variables as the length of time they had belonged, the hours of time spent in organizational activities, the size of the organization, and so forth. The respondents were prompted to remember their affiliations through aided recall techniques (Johnson 1977). The types of organizations included for this analysis were church-related, sports, youth, social, neighborhood improvement, women's, hobby, and community service groups. These types of groups were aggregated from a detailed list of unique groups and types consisting of over 400 categories. Every effort was made to include only groups which would be likely to provide face-to-face interaction among members. Examples of organization types excluded were professional groups, labor unions and welfare organizations.

The mean number of such memberships for the entire sample was 0.937. The distribution is positively skewed, with a variance of 1.373. The relationship between the various parameters and town size is shown numerically in Table 1. As an overall scan of the Table shows, the mean affiliation rate, which has been the principal measure in the literature, is basically unrelated to town size. Mean affiliation shows a slight peak at town size approximating 7500, and declines gently with increasing town size thereafter. Thus, the individual characteristic most often studied in the survey literature on voluntary associations does not vary much over different system sizes; individuals tend to belong to the same number of organizations regardless of town size. However, the organizational characteristics such as number, average size, and density of connections vary dramatically with system size, as we will now show.

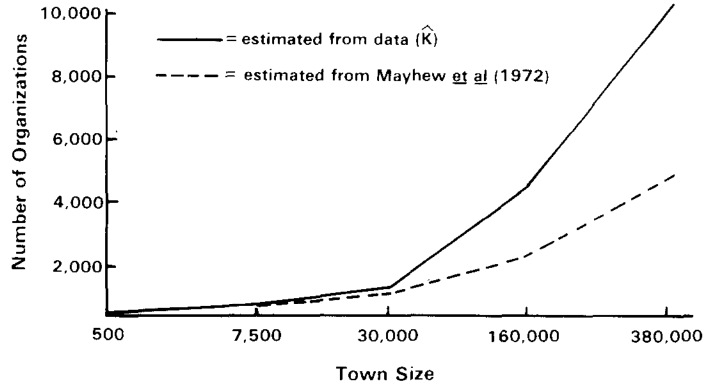
Figure 2 presents the relationship between town size and the estimated number of organizations. This Figure shows that K is a strongly increasing function of system size. For towns in the range of 1000 people or less, we find about 30 organizations; when town size exceeds 100 000, the number of organizations is over 5000. Thus, the size–structural differentiation hypothesis is strongly supported for these data. Larger systems (large N) generate larger total volumes of memberships (large T), which are differ-

Table 1. *The parameters of affiliation for varying town size**

	State	Omaha	Lincoln	30 000	7500	500
N	1 559 358	379 403	157 945	30 000	7500	500
n	1603	454	189	280	430	250
$\hat{T}$	1 462 677	318 319	137 096	26 250	7762	502
	(43 472)	(21 568)	(13 755)	(1930)	(462)	(26)
$\hat{\bar{m}}$	0.938	0.839	0.868	0.875	1.035	1.004
	(0.028)	(0.057)	(0.087)	(0.065)	(0.063)	(0.073)
$\hat{L}$	1 085 313	252 303	103 612	15 870	6608	337
	(73 813)	(37 584)	(20 724)	(2421)	(904)	(42)
$\hat{L}_u$	3.532×10^{15}	1.923×10^{13}	2.042×10^{12}	1.732×10^{10}	5.470×10^{8}	2.127×10^{5}
$\hat{D}_o$	3.073×10^{-10}	1.312×10^{-8}	5.073×10^{-8}	9.160×10^{-7}	1.208×10^{-5}	1.585×10^{-3}
$\hat{K}$	67 301	10 069	5085	1076	382	30
	(2662)	(903)	(648)	(120)	(32)	(1.8)
$\hat{\bar{S}}$	21.73	31.61	26.96	24.41	20.30	16.92
	(0.67)	(2.39)	(2.42)	(2.19)	(1.19)	(0.84)
$\hat{I}_u$	8.183×10^{16}	7.247×10^{14}	6.343×10^{13}	4.838×10^{11}	1.075×10^{10}	3.701×10^{6}
	(2.24×10^{15})	(6.50×10^{13})	(8.08×10^{12})	(5.39×10^{10})	(8.88×10^{8})	(2.25×10^{5})
$\hat{I}$	4.675×10^{8}	1.601×10^{8}	8.263×10^{7}	9.681×10^{6}	1.555×10^{6}	8.732×10^{4}
	(4.57×10^{7})	(2.55×10^{9})	(2.24×10^{7})	(2.35×10^{6})	(3.28×10^{5})	(2.13×10^{4})
$\hat{D}_I$	5.713×10^{-9}	2.209×10^{-7}	1.302×10^{-6}	2.001×10^{-5}	1.445×10^{-4}	2.359×10^{-2}
	(2.60×10^{-11})	(1.04×10^{-9})	(1.29×10^{-8})	(1.92×10^{-7})	(1.67×10^{-6})	(3.48×10^{-4})
$\hat{P}_O$	0.732	0.698	0.712	0.715	0.756	0.750

*Data from *Nebraska Annual Social Indicators Survey 1977.* Standard errors, where available, are in parentheses. Symbols are as follows: N = size of system, n = size of sample, $\hat{T}$ = number of memberships, $\hat{\bar{m}}$ = mean affiliation rate, $\hat{L}$ = number of links among organizations, $\hat{L}_u$ = number of possible links among organizations, $\hat{D}_o$ = density of organizational links, $\hat{K}$ = number of organizations, $\hat{\bar{S}}$ = mean size of organizations, $\hat{I}_u$ = number of possible links among individuals, $\hat{I}$ = number of individual links, $\hat{D}_I$ = density of individual links, $\hat{P}_O$ = proportion of overlapping memberships. (See text and appendices for discussion of estimators.)

Figure 2. *The relationship between town size and the number of organizations (Nebraska 1977)*



entiated into a greater number of subcomponents (large K). Blau's (1970) hypothesis derived for business organizations is strongly supported here for voluntary organizations as subcomponents of communities.

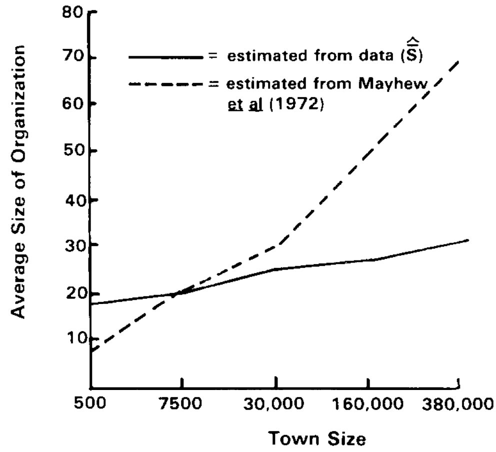
It is useful to compare the number of organizations generated by our estimate with that generated by the random baseline of Mayhew *et al.* (1972). Mayhew's concern was with distributing partitions of the total size of the system and taking the average number of such partitions over all possible structures. The function generated by Mayhew's technique is plotted for our data as a broken lines in Fig. 2.[5] As is obvious from the Figure, the random baseline generator predicts substantially fewer organizations for the larger systems than we observe.

Clearly, some process is operating to generate more organizations, given the total volume of memberships, than random expectation would lead us to predict. We believe that the key lies in the fact that we consider only organizations in which face-to-face interaction is most likely to occur, such as church-related, sports groups, and the like. There is probably an upper limit on the size of such organizations. For instance, the size of the Tuesday Evening Square Dance Club is limited by the size of the facilities available to house the meetings. This constraint will operate not so much to set an absolute ceiling in general, but will tend to suppress the relationship between system size and average organization size that we might otherwise observe.

This impression is mirrored in the results of Fig. 3, which shows that average organization size increases dramatically with increasing system size. The organizations in towns around size 500 tend to exceed the predicted values, perhaps because there are lower limits on the viable size of voluntary

[5] The published functional relationship between size and differentiation in Mayhew *et al.* (1972) is $\ell n(\text{differentiation}) = -0.043 + 0.680\,767\ \ell n(\text{size})$. This function was based upon an algorithm which generated systems up to size seventy. A subsequent improvement in the algorithm allows the generation of much larger systems. The functional relationship for the extended range of system size is $\ell n(\text{differentiation}) = 0.6667\ \ell n(\text{size})$, with a coefficient of determination of approximately 1.0. This functional relationship is used to generate the estimates for Figs. 2 and 3.

Figure 3. *The relationship between town size and the average size of voluntary organizations.*



organizations which are distinct enough to be named in a survey interview. However, the increase in expected size of organizations under the baseline model is far greater than we observe in our data. Thus, while the general size–differentiation relationship is strongly supported by these data, the specific form of the function differs from the baseline generator of Mayhew *et al.* (1972) for these restricted types of organizations.

As discussed earlier, there is a dramatic relationship between town size and the number of organizations. However, because of the increase in size of these organizations, the number of organizations per individual actually declines with increasing city size. The decline is from approximately five organizations per hundred individuals to less than three organizations per hundred. Thus, the opportunity structure for organizational participation changes in character with increasing city size; smaller towns tend to have proportionately more organizations, but these organizations are smaller than their more urban counterparts.

As Table 1 shows, both individual and organizational density show dramatic monotonic decreasing relationships to town size. Both densities decrease by a factor of approximately 10 000 times over the system size range 500–380 000. The fact that these densities are extremely small to begin with, and decrease markedly with increases in system size, points up the dramatic sparseness of the network spanning social life in general. Note, however, that these networks, sparse as they may be, still can operate with stunning efficiency in passing messages through the system, as the small-world literature (Milgram 1967) has demonstrated.

The drastic decline in individual-level density with system size accords well with our common-sense notions of small town life; just as one may be more likely to meet one's neighbor at the grocery store in a village, so too is one more likely to belong to the same church group and the same local

PTA. These results also accord with (possibly) more sophisticated sociological theories of the greater impersonality of urban life (e.g. Wirth 1938, Milgram 1970), and with the carefully reasoned structural paradigm of Blau (1977), who argues that smaller systems will be more integrated at the level of social associations.

The decline in organizational-level density is less intuitively obvious. If the increased number of people in the system had been offset by a proportional increase in the size of the organizations in the system, then organizational-level density could have remained constant. This fact is true because the size of the system (N) cancels out of eqn. (7) in the numerator and the denominator, leaving the number of organizations (K) in the denominator to dominate the expression. Since K increases dramatically with system size, the denominator of (7) increases strongly with N, decreasing the density. Thus, the fact that the size increases in organizations fail to keep pace with the size increases in the system[6] forces the number of organizations to increase, which lowers their density of interconnection.

Notwithstanding the dramatic decline in density with increasing city size, the mean number of links per organization (obtained by dividing L by K) tends to increase with city size, because of the increase in organizational size.[7] Thus, we would expect that organizations in urban areas would tend to have access to and/or information about the resources and activities of a larger number of other organizations than would organizations in smaller cities. These urban organizations would at the same time be less connected to the totality of all organizations in the system, because they are located in a less dense (hyper)network. We suspect that this fact implies that a given urban organization is less likely to attain dominance (in the sense of centrality, or minimum network distance) in larger systems because of the greater number of organizations and the lower density. Tests of this possibility must await more detailed information on the distribution of linkages among organizations.

Discussion

We have developed estimators for macro-level characteristics of voluntary affiliation. These characteristics include T, the total volume of memberships; K, the number of organizations; $\bar{S}$, the average size of organizations; L, the number of links among organizations; L_u, the number of links possible among organizations; I, the number of links among individuals; I_u, the num-

[6] Note that this increase of size in organization at a decreasing rate is strictly in accord with Blau's (1970) generalization (1), which predicts that increasing size produces structural differentiation at decelerating rates.

[7] A reader has pointed out that the increase in organization size in larger towns may mean that members are acquainted with a smaller proportion of their co-members. If so, the measures of density will tend to vary even more dramatically with town size, since we would be overestimating the true density for larger towns. We need to remember, however, that we are dealing only with face-to-face organizations here. Even in the largest city, the mean size is only slightly above thirty members — a not unreasonably large number to have at least nodding acquaintance with.

ber of links possible among individuals; D_O, the density of links among organizations; and D_I, the density of links among individuals. Applying these estimators to data from a probability sample of individuals in towns of varying size in the state of Nebraska, we found that: (1) as city size increases, mean organization size increases, (2) as city size increases, inter-organizational and inter-individual density decrease, (3) mean affiliation shows no monotonic relationship to city size, (4) the number of organizations increases with city size while (5) the number of organizations per person decreases with city size, (6) the number of inter-organizational links increases with city size, and (7) the number of inter-organizational links per organization increases with city size.

Our results show that the mean affiliation rate is not strongly related to the structural parameters in these data, since it does not systematically vary with town size. Equations (3) and (10) show that differences in means can produce substantial differences in structure across systems, *ceteris paribus*. One of the most important features of *ceteris* that must be *paribus* is the size of the system, as our analysis shows. A favorite theme of recent research in the area of voluntary affiliation is the interpretation of ethnic differences in mean affiliation. We would argue that these much-researched differences in mean affiliation do not necessarily imply anything about structural differences in the various communities without a careful analysis such as the one begun in this paper.

Finally, it is important to note that contemporary theories about mean differences in affiliation, such as the 'ethnic community' and 'compensatory' theories, offer us little or no insight into the structural properties of the systems. We have no idea how ethnically exclusive or culturally homogeneous these organizations are, nor any suggestion of how these differences in affiliation are apportioned among organizations in the community. We suspect that the ethnic community theory might imply greater density of linkages and/or shorter network distances among individuals in the minority but these possibilities have not been researched. We hope that the results of our present efforts will be to sensitize researchers to the potential of the hypernetwork formulation.

We close with some suggestions for other applications of hypernetworks. A very straightforward extension of these techniques to the analysis of interlocking directorates is possible; this literature has usually been confined to the study of only the very largest organizations (Mariolis 1975; Burt *et al.* 1980). The use of sampling techniques developed in this paper would allow the study of more general systems of interlocks among corporations. The duality of the linkages among corporations created by individuals, and the links among individuals created by corporations may provide fertile ground for theoretical development of the corporate interlock literature.

A totally different area of research which may benefit from hypernetwork imagery is the literature on behavior settings (Barker and Wright 1971). This area of research has developed very powerful images of human behavior as it takes place in geographically and temporally localized settings. What has been missing from these studies is any means of connecting these settings with one another; the people who move from setting to setting are lost in the shuffle. Hypernetworks can provide a much needed way of thinking about the connections between settings produced by sequential activities of individuals. and the connections among individuals produced by these settings.

Other promising applications of hypernetworks are in such diverse areas as the status attainment literature (multiple jobs held by individuals as links among occupations, or, the use of the family as the unit of analysis—jobs held by multiple members of a family create connections among economic organizations) and the scientific citation literature (multiple authorships linking individual researchers while the researchers link pieces of research). We believe that those, and many other applications, will emerge as the hypernetwork model develops.

References

Aberback, J. K.
1969 "Alienation and political behavior". *American Political Science Review 63*:86–99.

Antunes, G. and C. Gaitz
1975 "Ethnicity and participation: a study of Mexican Americans, blacks and whites". *American Journal of Sociology 80*:1192–1211.

Babchuk, N. and J. N. Edwards
1965 "Voluntary associations and the integration hypothesis". *Sociological Inquiry 35*(Spring): 149–162.

Babchuk, N. and R. V. Thompson
1962 "The voluntary associations of Negroes". *American Sociological Review 27*:647–655.

Baltzell, E. D.
1966 " 'Who's Who in America' and 'The Social Register' ". Pp. 266–275 in R. Bendix and S. M. Lipset (eds.), *Class, Status, and Power: Social Stratification in Comparative Perspective*. New York: The Free Press.

Barker, R. G. and H. F. Wright
1971 *The Midwest and Its Children: The Psychological Ecology of an American Town*. Hamden, CT: Anchor Books.

Bell, W. and M. T. Force
1956 "Social structures and participation in different types of formal association". *Social Forces 34*:345–350.

Berge, C.
1976 *Graphs and Hypergraphs*. Amsterdam: North-Holland.

Blalock, H. M.
1964 "Some implications of random measurement error for causal inferences". *American Journal of Sociology 71*:836–854.

Blau, P. M.
1970 "A formal theory of differentiation in organizations". *American Sociological Review 35*: 201–218.
1977 "A macrosociological theory of social structure". *The American Journal of Sociology 83*: 26–54.

Breiger, R. L.
1974 "The duality of persons and groups". *Social Forces 53*:181–189.

Burt, R. S.
1980 "Models of network structure". Pp. 79–141 in *Annual Review of Sociology*. Vol. 6. Palo Alto: Annual Reviews.

Burt, R. S., P. Christman and C. Kilburn, Jr.
1980 "Testing a structural theory of corporate cooptation". *American Sociological Review 45*: 821–841.
Cochran, W. G.
1963 *Sampling Techniques.* New York: Wiley.
Coleman. J. S.
1958 "Relational analysis: the study of social organizations with survey methods". *Human Organization 17*:28–37.
Feller, W.
1957 *An Introduction to Probability Theory and Its Applications.* Vol. 1. New York: Wiley.
Frank, O.
1971 *Statistical Inference in Graphs.* Stockholm: FOA Reproductions.
Gove, W. R. and H. L. Costner
1969 "Organizing the poor: an evaluation of a research strategy". *Social Science Quarterly 50*: 643–656.
Granovetter, M.
1973 "The strength of weak ties". *American Journal of Sociology 78*:1360–1380.
1976 "Network sampling: some first steps". *American Journal of Sociology 81*:1287–1303.
Hawley, A. H., W. Boland and M. Boland
1965 "Population size and administration in institutions of higher education". *American Sociological Review 36*:252–255.
Johnson, D. R.
1977 *Nebraska Annual Social Indicators Survey 1977, Report #5, Designs, Procedures, Instruments and Forms for the 1977 NASIS.* Lincoln, Nebraska: Bureau of Sociological Research, University of Nebraska.
Jöreskog, K.
1973 "A general method for estimating a linear structural equation". Chap. 5 in A. S. Goldberger and O. D. Duncan (eds.), *Structural Equation Methods in the Social Sciences.* New York: Academic Press.
Kasarda, J. W.
1974 "Structural implications of system size: a three level analysis". *American Sociological Review 39*:19–28.
Laskin, R.
1961 *Organizations in a Saskatchewan Town.* Saskatchewan, Canada: University of Saskatchewan Press.
Lin, N., J. C. Vaughn and W. M. Ensel
1981 "Social resources and occupational status attainment". *Social Forces 59*:1163–1181.
Lynd, R. S. and H. M. Lynd
1929 *Middletown.* New York: Harcourt.
1937 *Middletown in Transition.* New York: Harcourt.
McCarthy, J. W. and M. N. Zald
1977 "Resource mobilization and social movements: a partial theory". *American Journal of Sociology 82*:1212–1241.
McPherson, J. M.
1981a "Voluntary affiliation: a structural approach". In P. M. Blau and R. K. Merton (eds.), *Continuities in Structural Inquiry.* London: Sage Press.
1981b "A dynamic model of voluntary affiliation". *Social Forces 59*:705–728.
McPherson, J. M. and D. O'Donnell
1980 "The size of voluntary associations". Paper read at the annual meetings of the American Sociological Association, New York.
McPherson, J. M. and D. L. Smith-Lovin
1981 "Women and weak ties: sex differences in the size of voluntary organizations". *American Journal of Sociology*, in press.
Mariolis, P.
1975 "Interlocking directorates and the control of corporations: the theory of bank control". *Social Science Quarterly 56*:425–439.

Mayhew, B. H. and R. L. Levinger
1976 "Size and the density of interaction in social aggregates". *American Journal of Sociology 82*:86–110.
Mayhew, B. H., J. M. McPherson, R. W. Levinger and T. W. James
1972 "System size and structural differentiation in formal organizations: a baseline generator". *American Sociological Review 37*:629–633.
Milgram, S.
1967 "The small world problem". *Psychology Today 1*:61–67.
1970 "The experience of living in cities". *Science 167*:1461–1468.
Olsen, M. E.
1970 "The social and political participation of Negroes". *American Journal of Sociology 8*: 1–46.
Orum, A. M.
1966 "A reappraisal of the social and political participation of Negroes". *American Journal of Sociology 72*:32–46.
Siegal, P. M. and R. W. Hodge
1968 "A causal approach to the study of measurement error". Pp. 28–59 in H. M. Blalock and A. B. Blalock (eds.), *Methodology in Social Research.* New York: McGraw-Hill.
Simmel, G.
1902 "The number of members as determining the sociological form of groups. I". *American Journal of Sociology 8*:1–46.
1922 *Conflict and the Web of Group Affiliations* (translated by K. H. Wolff and R. Bendix, 1955). New York: The Free Press.
Spencer, H.
1877 *The Principles of Sociology.* New York: Appleton.
Turk, H.
1977 *Organizations in Modern Life.* San Francisco: Jossey-Bass.
Warner, W. L. and P. S. Lunt
1941 *The Social Life of a Modern Community.* New Haven: Yale University Press.
Webster, Jr., M.
1973 "Psychological reductionism, methodological individualism, and large scale problems". *American Sociological Review 38*:258–273.
White, H. C.
1970 Search parameters for the small world problem". *Social Forces 49*:259–264.
White, H. C., S. A. Boorman and R. L. Breiger
1976 "Social structure from multiple networks. I. Blockmodels of roles and positions". *American Journal of Sociology 81*:730–780.
Williams, J. A., N. Babchuk and D. R. Johnson
1973 "Voluntary associations and minority status: a comparative analysis of Anglo, black, and Mexican Americans". *American Sociological Review 28*:637–646.
Wirth, L.
1938 "Urbanism as a way of life". *American Journal of Sociology 44*:1–24.
Wright, C. K. and H. H. Hyman
1958 "Voluntary association memberships of American adults: evidence from national sample surveys". *American Sociological Review 23*:287–296.
Yakowitz, S. J.
1977 *Computational Probability and Simulation.* Reading, MA: Addison-Wesley.
Young, D. R. and S. J. Finch
1977 *Foster Care and Non-Profit Agencies.* Lexington, MA: D.C. Heath.
Zald, M. N.
1970 *Organizational Change.* Chicago: The University of Chicago Press.
Zald, M. N. and P. Denton
1963 "From evangelism to general service: the transformation of the YMCA". *Administrative Science Quarterly 8*:214–234.

Appendix A

Standard errors for estimators

In this Appendix, we develop standard errors for some of the parameters discussed in the body of the text. As we shall see, a key issue in developing standard errors for our parameters is the extent to which the size of each organization is measured with error. Two of our estimators, $\hat{T}$ and $\hat{L}$, are completely unaffected by any error in the measurement of size. We will discuss these estimators first.

Let $m_{i.}$ be the number of organizations belonged to by individual $i\cdot$. Then,

$$\hat{T} = \frac{N}{n} \sum_{i=1}^{n} m_{i.} \tag{A-1}$$

is an unbiased estimator of the total volume of memberships, whose standard error is

$$\text{S.E.}(\hat{T}) = N \left[\frac{\text{var}(m_{i.})}{n} \left(1 - \frac{n}{N} \right) \right]^{1/2} \tag{A-2}$$

Now, the estimator for L, the total number of lines between organizations, is

$$\hat{L} = \frac{N}{n} \sum_{i=1}^{n} m_{i.}(m_{i.} - 1)/2 \tag{A-3}$$

which is an unbiased estimator of L. We define $P_i = m_{i.}(m_{i.} - 1)/2$. Then the standard error of L is

$$\text{S.E.}(\hat{L}) = N \left[\frac{\text{var}(P)}{n} \left(1 - \frac{n}{N} \right) \right]^{1/2} \tag{A-4}$$

The following estimators and standard errors are affected by measurement error in size.

We define a basic quantity, the sum of the reciprocals of the sizes of a particular individual's organization:

$$C_i = \sum_{j=1}^{S_L} m_{ij}/S_j \tag{A-5}$$

The variance of C will be

$$\text{var}(C) = \sum_{i=1}^{n} (C_i - \bar{C})^2 / (n - 1) \tag{A-6}$$

where $\bar{C} = E(C)$ = mean of C for all individuals. Now, from (5a) in the text, and (A-5), we see that the estimator for K, the total number of organizations in the population, may be rewritten as

$$\hat{K} = \frac{N}{n} \sum_{i=1}^{n} C_i \tag{A-7}$$

The standard error of $\hat{K}$ is (Cochran 1963)

$$\text{S.E.}(\hat{K}) = N\left[\frac{\text{var}(C)}{n}\left(1 - \frac{n}{N}\right)\right]^{1/2} \tag{A-8}$$

$\hat{K}$ is unbiased and consistent in the absence of measurement error in S (see Appendix B). The estimator of the mean size of organizations in the system is

$$\hat{\bar{S}} = \hat{T}/\hat{K} = \sum_{i=1}^{n} m_{i.} / \sum_{i=1}^{n} C_i \tag{A-9}$$

$\hat{\bar{S}}$ is a consistent estimator of the mean size of organizations, whose bias is negligible for large samples (Cochran 1963) in the absence of measurement error in S. The standard error of $\hat{\bar{S}}$ is

$$\text{S.E.}(\hat{\bar{S}}) = \frac{1}{\bar{\bar{C}}}\left[\frac{1 - n/N}{n} \sum_{i=1}^{n} \frac{(m_{i.} - \hat{\bar{S}}C_i)^2}{n-1}\right]^{1/2} \tag{A-10}$$

Now,

$$\hat{I}_u = \frac{N(N-1)}{2}\hat{K} \tag{A-11}$$

is an unbiased estimator of I_u, the upper limit on the number of individual connections. The standard error of I_u is

$$\text{S.E.}(\hat{I}_u) = \frac{N(N-1)}{2}\,\text{S.E.}(\hat{K}) \tag{A-12}$$

We define

$$q_i = \sum_{j=1}^{S_L} m_{ij}(S_j - 1)/2 \tag{A-13}$$

then,

$$\hat{I} = \frac{N}{n}\sum_{i=1}^{n} q_i \tag{A-14}$$

is an unbiased estimator of I, the number of links at the individual level, in the absence of measurement error in S. The standard error of I is

$$\text{S.E.}(\hat{I}) = N\left[\frac{\text{var}(q)}{n}\left(1 - \frac{n}{N}\right)\right]^{1/2} \tag{A-15}$$

Now, D_I, the density of links among individuals, is

$$\hat{D}_I = \hat{I}/\hat{I}_u \tag{A-16}$$

which is a consistent estimator of D_I in the absence of measurement error in S. The standard error of D_I is

$$\text{S.E.}(\hat{D}_I) = \frac{2}{N(N-1)}\left\{\frac{1 - n/N}{n}\sum_{i=1}^{n}\left[q_i - \frac{\Sigma q_i}{\Sigma C_i}(C_i)\right]^2 \Big/ (n-1)\right\}^{1/2} \tag{A-17}$$

Appendix B

Some simulation results for estimators of size and number of organizations when size contains measurement error

Since we are weighting each observation by a function of organization size, it is important to understand how our estimates will be affected by measurement error in size. Below we report results of a pilot study which investigates how the estimators for size and number are affected by measurement error. We confine ourselves in this report to the estimators of size and number for two basic reasons. Firstly, size and number are the most important aspects of the system. Without a knowledge of the size and number of organizations, there is little chance of any important structural insights into the system. In terms of the basic hypernetwork model, the number of organizations is the number of columns in the adjacency matrix; the sizes are the column sums. Secondly, size and number are characteristics of the organizations themselves; they do not depend upon individual characteristics, such as the number of affiliations. Because of this fact, we may focus upon a pool of memberships without having to assemble these memberships into individuals, which is a difficult task because the memberships in organizations of different size are not randomly distributed among individuals (McPherson and O'Donnell 1980).

For the simulation, we produce a hypothetical population of 100 000 memberships, distributed among 2706 organizations. This combination generates an average organizational size of 33.6 members. The organizations range in size from 10 to 100; from the individual-level point of view, the sample will appear to have a mean organization size of approximately 55, since larger organizations will be replicated more often in the sample of memberships than smaller organizations.

We use a linear congruential random number generator (Yakowitz 1977) with random shuffling of digits, implemented on an AMDAHL 470/V6 in IBM VS BASIC language, to produce each sample of memberships.

For each membership, a random digit from 0 to 9 is generated which determines the size of the organization attached to the membership; digit '0' produces a membership with reported size 10, digit '1' produces size of 20, and so forth. Thus, roughly an equal number of observed memberships is produced at each size value. Note that while this procedure produces a rectangular distribution of size in the membership sample, the actual distribution of size in our hypothetical population of organizations is heavily skewed, since large organizations are strongly over-reported in the sample. We generate one thousand samples from this population at each sample size, and each value of measurement error. To introduce measurement error, we simply add a random component to each observed size, consisting of a percentage of the true value, the sign of which is chosen randomly. We express the random component as a percentage of the true size of the organization since it is reasonable to suppose that individuals in a small organization will be likely to make smaller absolute errors than people in large organizations. For our indicator of amount of measurement error, we use the ratio of true-score variance to observed variance—the reliability coefficient r.

The basic finding in Table B-1 is that increasing amounts of error in the measurement of size biases upward the estimates of the number of organiza-

Table B-1. *A simulation study of estimates of the number of organizations and their average size for a hypothetical system of 100 000 memberships**
(T = 100 000, K = 2976, $\bar{S}$ = 33.6)

n	r	$E(\hat{K})$	S.D.$(\hat{K})$	E[S.E.$(\hat{K})$]	$E(\hat{\bar{S}})$	S.D.$(\hat{\bar{S}})$
50	1.00	2963.0	38.8	38.0	33.7	0.44
	0.95	2982.7	40.2	38.3	33.5	0.45
	0.90	3039.0	40.9	39.7	32.9	0.44
	0.80	3086.0	41.7	41.5	32.4	0.43
	0.70	3183.0	43.9	43.6	31.4	0.43
	0.60	3278.9	44.8	45.9	30.5	0.42
100	1.00	2976.4	25.7	27.1	33.5	0.29
	0.95	3004.3	26.0	27.8	33.2	0.29
	0.90	3047.2	27.6	28.6	32.8	0.30
	0.80	3145.1	29.9	29.9	31.8	0.30
	0.70	3301.9	32.7	33.0	30.3	0.30
	0.60	3396.3	33.9	35.0	29.4	0.29
200	1.00	2979.2	17.4	19.2	33.6	0.20
	0.95	3024.2	19.2	19.6	33.1	0.21
	0.90	3059.9	17.3	20.2	32.7	0.19
	0.80	3175.0	20.6	21.9	31.5	0.20
	0.70	3282.5	21.9	23.2	30.5	0.20
	0.60	3388.0	22.2	24.6	29.5	0.19
500	1.00	2986.9	12.9	12.2	33.5	0.14
	0.95	3010.0	13.1	12.4	33.2	0.14
	0.90	3113.0	13.1	13.2	32.1	0.14
	0.80	3202.4	13.8	13.9	31.2	0.13
	0.70	3280.5	14.5	14.6	30.5	0.13
	0.60	3411.4	14.6	15.6	29.3	0.13

*1000 samples at each sample size n. $E(\hat{K})$ is the mean of the estimated sizes across the 1000 samples, S.D.$(\hat{K})$ is the standard deviation of the estimate mean size across 1000 samples, E[S.E.$(\hat{K})$] is the mean standard error of estimated K across samples, $E(\hat{\bar{S}})$ is the mean of the size estimates across samples, S.D.$(\hat{\bar{S}})$ is the observed st^ndard deviation of the estimated mean size across samples, r is the ratio of true variance in size to observed variance.

tions K and biases downward the estimates of mean organization size $\bar{S}$. When there is no measurement error (r = 1.0), the estimators show no bias (a maximum deviation from the true value of less than 0.5%), and their analytically derived standard errors match well with their observed variability across one thousand samples. Thus, when there is no measurement error, we see that the simulation demonstrates the results in Appendix A; the estimators are unbiased, and their standard errors reflect their actual variability.

As the measurement unreliability increases at all sample sizes, we see a monotonic increase in upward bias in K, and a corresponding downward bias in $\bar{S}$. The analytically derived standard error of K appears to behave quite well under conditions of increased variability in the estimator about its mean. If anything, the standard error appears to become slightly more conservative with measurement error, producing a slightly wider confidence interval at higher levels of error.

The maximum bias we find in $\hat{K}$ and $\hat{\bar{S}}$ is approximately 12%, for a reliability of 0.6. We believe that these results suggest that our estimators are actually fairly robust in the presence of measurement error, although they are clearly biased. It is important to realize that the level of measurement error necessary to produce a reliability as low as 0.6 means that all *individuals are over- or underestimating their organizations by nearly 50%.* We believe that this level of inaccuracy is quite unlikely for the types of organizations in which we are interested.

In order to put this error-induced bias in $\bar{S}$ and K into proper perspective, let us consider a statistical alternative. An apparently reasonable approach to measuring the mean size of organizations would be to take the unweighted mean of the sizes reported for each membership. For our case, this mean would be approximately 55, regardless of the amount of measurement error in size. Note that this (unweighted) mean is *biased over 60%*, regardless of sample size or measurement error. In cases of measurement error, the true mean size of organizations will fall between the unweighted and weighted means. Ordinarily, the true mean will fall much closer to the weighted mean. In the worst case, the unweighted and weighted means serve as upper and lower limits on the mean size of organizations in the system.

Finally, it is important to point out that bias in an estimator in the presence of random measurement error is not unusual. For example, it has long been known that random measurement error in the independent variable of a regression analysis biases the regression coefficient (Blalock 1964). In fact, the knowledge that measurement error induces bias served as a stimulus for the development of models which adjust for such measurement error, such as the LISREL models of Jöreskog (1973), and the causal approach of Siegal and Hodge (1968). In conclusion, we note that the issue of measurement error in size can only be rigorously investigated in the context of a formal model. We believe that this paper is a step in the direction of developing such a model.