A method is described for interviewing a random sample of persons drawn from a large population so as to describe role-sets defining statuses in the population social structure. The key to the method is a connection between the concept of an actor's network position in social structure and combinations of attributes that define statuses in the social structure. With data obtained in a survey interview with a randomly selected respondent, it is possible to describe the relational pattern defining his "ersatz network position" in the population social structure from which he has been drawn. Given ersatz network positions for a representative sample, it is possible to test hypotheses concerning satus/role-sets in terms of which the population is stratified.

Studying Status/Role-Sets as Ersatz Network Positions in Mass Surveys

RONALD S. BURT
University of California—Berkeley

The single factor most restricting structural theory employing networks concepts to small systems of actors is the realpolitik of data processing. Network concepts typically, although not always, call for data on relations among all actors in a system. In order to meet this need, the typical network study involves fewer than 100 distinct actors; children in a classroom, employees in a small bureaucracy, a small number of large corporations within a geographical region, a small number of nations in some type of exchange system.[1] Unfortunately, the number of relations to be estimated for each of the networks within a system increases exponentially with system size. Where there are N(N−1) relations to be estimated within a network among

AUTHOR'S NOTE: *This article is part of the ongoing work under the Project in Structural Analysis at the Survey Research Center, University of California at Berkeley. It is a by-product of support from the National Science Foundation*

SOCIOLOGICAL METHODS & RESEARCH, Vol. 9 No. 3, February 1981 313-337

N actors, there are 9,900 relations to be estimated in a network among 100 persons, 249,500 relations to be estimated in a network among 500 persons, 6,247,500 relations to be estimated in a network among 2500 persons, and so on. Even a vague familiarity with data processing is sufficient to know that it would be difficult to obtain such extensive data from respondents and no easy task to analyze it.

But who lives in an area occupied by only 2500 persons? In seeking network data on a single network in such a system, a system the size of a small rural hamlet, we have outstripped our data analysis capabilities as well as the patience of the typical survey respondent. If a rural hamlet is too large for analysis, what about cities, states, regions, or the nation? It is in these larger systems that the typical citizen resides.

System size is not the only problem. Structural theory makes statements about perceptions and behaviors in terms of the network context of actors. This context is lost for a random sample of actors. Within the hypothetical hamlet of 2500 persons, 100 could be randomly selected to be interviewed concerning their relations to other persons. But there is no method of knowing how those respondents are connected within the system. In a random sample of N* respondents drawn from a system of N actors, information can be obtained on relations among the respondents and to others in the system—a total of N(N*–1) relations. Information on relations among the noninterviewed actors and relations from the noninterviewed actors to the sampled respondents is not obtained—a total of N(N–1) – NN* relations. The network data ignored in a random survey of k% of a population is roughly (1–k)% so that a random sample of 10% of a population ignores 90% of its network data, a random sample of 25% ignores 75%, and so on.[2] These lost data are

(SOC77-22938, SOC79-25728) and has received direct support from the Pacific Institute for Research and Evaluation under a grant from the National Institute on Drug Abuse (1-R01DA02566-01). The article is abbreviated from a working paper written during a leave of absence spent in the Department of Sociology, State University of New York at Albany. The discussion has benefited from comments by H. R. Bernard, P. M. Blau, M. Granovetter, D. R. Heise, N. Lin and J. A. Wiley, all of whom are absolved of the uses to which I have put their sage advice.

significant. To what extent do the nonsampled persons reciprocate relations directed to them from the sampled respondents? Are respondents the object of strong relations from the system as a whole or are they relatively isolated within the system? How are the nonsampled persons interconnected, apart from the respondents? Answers to these questions define the network context of the random sample, but the typical survey research design obliterates that context from our view.

In theory, new research designs could be proposed so that network concepts could inform survey data. Beginning with accolades for the theoretical significance of network concepts, for example, Granovetter (1976: 1287) writes that most discussions of network concepts "have had practical application only to small groups." Striving to bring such discussions "more squarely into the mainstream of sociological research," he builds on work estimating the total number of persons and relations in a large population (Frank, 1975) to propose a sample estimate of network density in a large population; density being a scalar estimate of overall interaction in the population. As Granovetter takes pains to point out, these are clearly "some first steps." But even though density is among the more primitive of network concepts, there are practical problems in data collection that make even density estimates difficult (Morgan and Rytina, 1977).

Of course, there is always the short-term solution of gathering what network data a survey research design permits and making the severely limited inferences such data allow. Concomitant with the increasing popularity of the term "network" for example, an increasing number of articles have appeared as "network studies," in which the author refers to a survey respondent's sociometric citations as a network. To the extent that "networks" have an effect on some dependent variable under study, the number of citations a respondent makes is expected to strongly predict the dependent variable. This measure of ego-network range is perhaps the least presumptuous of all network concepts, yet even it has been useful in research; particularly research on social stress. Researchers accustomed to the subtlety and depth of structural theory applied to typical network data can certainly frown on the ostensible naiveté of research in which network structure is reduced to frequency counts of contacts. But with the need to make inferences about persons in large populations, and therefore a need to use traditional survey research designs ruling out designs in which typical network data are obtained, how can more sophisticated network concepts be invoked?

One kind of answer to this question is the use of what I shall term "ersatz network positions" in survey data. A connection exists between the usual concept of an actor's network position in social structure and combinations of attributes defining statuses in that social structure. Survey data on a randomly sampled respondent can be used to describe the relational pattern defining his "ersatz" network position in the social structure from which he has been drawn. In proposing the idea of an ersatz network position, I draw on the data collection strategy developed by Edward Laumann and Claude Fisher, with their respective colleagues, for studying respondent ego-networks (Laumann, 1966, 1973; McCallister and Fischer, 1978a, 1978b; Fischer, 1981), the macrolevel conception of parameters defining positions in social structure developed by Peter Blau (Blau, 1974, 1977a, 1977b), and my own work on network models of status/role-sets in social structure (Burt, 1975, 1976, 1977, 1980, 1981). The next section defines status/role-sets as network positions. The connection between positions and attributes is then described, and, in the third section, I use that connection to propose ersatz network positions.

THE STATUS/ROLE-SET AS A NETWORK POSITION

The classic status/role-set duality is captured in network models as a network position jointly occupied by structurally equivalent actors; their pattern of relations defining the role-set and the rights and obligations of performing those relations defining the status. The key concepts in this representation are position, distance and equivalence (see Burt, 1980: 100-109, for detailed review). Since the extension of these concepts from a single network to multiple networks is obvious (compare Burt, 1976, 1977, and the review in Burt, 1980), I shall focus on single network systems throughout this discussion. The network of relations among N actors is given by an (N,N) matrix Z_{NN}, in which element z_{ji} is the strength of J's relation to actor I. Actor J's network position in the system of N actors is given by his relations to and from actors. This 2N vector of relations, Z_j, is given by the elements in row and column J of the matrix Z_{NN} as J's relational pattern:

$$Z_j = (z_{j1}, z_{j2}, \ldots z_{jN}, z_{1j}, z_{2j}, \ldots z_{Nj}). \quad [1]$$

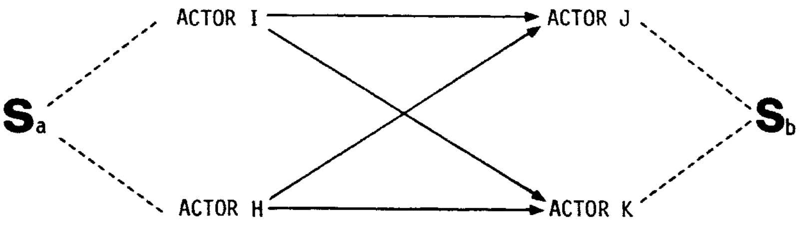
A set of actors jointly occupy a position to the extent that they have identical relational patterns within the system. Two actors are separated by high distance to the extent that they have very different relational patterns and separated by zero distance to the extent that they have the same relational pattern. This distance between actors J and I can be estimated by comparing each corresponding relation involving them:

$$d_{ij} = d_{ji} = (\Sigma_k^N [(z_{ik} - z_{jk})^2 + (z_{ki} - z_{kj})^2])^{1/2},$$
$$= ((Z_i - Z_j)(Z_i - Z_j)')^{1/2}. \quad [2]$$

To the extent that they are separated by zero distance, actors J and I are structurally equivalent and jointly occupy a single network position. Social structure in the system can now be described in terms of role relations among jointly occupied statuses. The (N,N) matrix of relations among actors, Z_{NN}, can be used to compute an (M,M) matrix of role relations among M statuses jointly occupied by structurally equivalent actors, Z_{MM}, where element z_{ab} is the mean relation, or density, from occupants of status S_a to occupants of status S_b;

$$z_{ab} = (\Sigma_h \Sigma_k z_{hk})/(n_a n_b), \quad [3]$$

where summation is across all n_a actors H occupying S_a and all n_b actors K occupying S_b. The matrix Z_{MM} is a density table.



$$z_{ab} = (z_{ij} + z_{ik} + z_{hj} + z_{hk})/4$$

Figure 1: The Relation from Status S_a to Status S_b in Terms of Relations Between Individual Occupants (Occupancy Given by Dashed Lines and Relations Given by Arrows, z_{ab} Defined in Equation 3)

Figure 1 illustrates the logic of this operation in a manner that will be useful. There are two mapping operations in computing an interstatus role relation as a density of relations between individuals. It is assumed that each indvidual actor occupies only a single status; actors I and H occupy status S_a in Figure 1. This means that the relation from one actor to another corresponds to one and only one role relation between statuses. In consequence of these assumptions, the role relation from S_a to S_b is equally reflected in any relation from an occupant of S_a to an occupant of S_b so the four estimated interactor relations in Figure 1 are merely variations on the interstatus relation and that relation is assumed to be their expected value, i.e., $z_{ab} = (z_{ij} + z_{ik} + z_{hj} + z_{hk})/4$. Of course, since occupants of each status are structurally equivalent, the four interactor relations will be very similar in magnitude by definition. The role-set defining status S_a is now given by the 2M relations in row and column A of the desity table Z_{MM}:

$$Z_a = (z_{a1}, z_{a2}, \ldots z_{aM}, z_{1a}, z_{2a}, \ldots z_{Ma}). \quad [4]$$

Given role-sets defining statuses in a system and the relational patterns defining the network positions of individual actors, a variety of hypotheses can be tested regarding actor perceptions and behavior, e.g., see Burt, 1981 for review and discussion.

Unfortunately, data on role-sets defining statuses in the social structure of large systems cannot be obtained in the typical survey research design. Such designs are not suited to estimating relations among all actors in a system. Without these elements for Z_{NN}, routine network analysis strategies for locating statuses as jointly occupied positions and assigning individual actors to specific statuses cannot be invoked.

THE ASYMMETRIC CONNECTION BETWEEN ACTOR ATTRIBUTES AND STATUSES

It is well known that actors jointly occupying a status tend to be homophilous in regard to one or more attributes such as race, sex, beliefs, occupation, religion, political affiliations, and so on. I have elsewhere discussed the social homophily of status occupants in detail (Burt, 1981: chs. 5, 6) but take it as given for the purposes here.

Some persons have suggested that actor attributes can be used directly as a surrogate for defining structural equivalence—that is, actors homophilous on key attributes can be treated as jointly occupying a single status. With various colleagues, Laumann (1966, 1973) has made major advances in developing this line of research so as to describe relations between racial, religious, and particularly occupational statuses in the United States. While not backing his arguments with the same extent of data, Blau (1974, 1977a, 1977b) has more systematically developed the analysis of actor attributes as "parameters" of social structure. In an analogue to the stratification space defined by distances between positions in network analysis (d_{ij} in equation 2), Blau (1977a: 30) suggests that:

> the structure of societies and communities are delineated by parameters. Structural parameters are the axes in the multidimensional space of social positions among which the population is distributed. They are attributes of people that underlie the

> distinctions they themselves generally make in their social relations, such as age, race, education, and socioeconomic status. . . . In short, a parameter is a variable that characterizes individuals and differentiates their role relations and social positions.

Blau then derives hypotheses concerning intergroup relations in stratified society from data on relations between persons with different attributes; hypotheses testable with data obtained in a typical survey research design.

This use of actor attributes as a surrogate definition of positions in social structure, a surrogate for direct investigation of interactor relations, is limited by the asymmetric connection between actor attribute homophily and structural equivalence. While all status occupants can be expected to share some common parameters in Blau's sense, all persons sharing common parameters need not be structurally equivalent. For example, consider a classroom of students stratified across four very cohesive cliques, each constituting a set of structurally equivalent children; two cliques composed of boys and two composed of girls. The four cliques are homophilous with regard to sex, but each boy is not in a clique with each other boy and each girl is not in a clique with each other girl. More generally, hypotheses derived from actor attribute homogeneity as a surrogate for actor structural equivalence will always have an initial empirical link in the chain of deductions leading to hypotheses. If attributes are incorrectly selected as parameters on which actors are homophilous, incorrect in the sense that parameters are selected which do not accurately differentiate jointly occupied statuses, then empirical tests of structural hypotheses stated in terms of the selected attributes are meaningless.[3] Accordingly, Blau (1977a: 30) requires that actor attributes only be considered as parameters when they differentiate:

> People can be classified on the basis of innumerable attributes, any of which may be a parameter. But if a classification made by an investigator does not influence social relations at all, or exerts only idiosyncratic influence on the personal relations of some individuals, it is not meaningful to consider it indicative of social positions. Hence, the double criterion of a parameter circum-

> scribing social positions is that it is an attribute by which a population is classified and that the social relations among persons similarly classified differ on the average from the relations between persons in widely different categories.

This is a necessary—but not sufficient—condition for an attribute being a parameter in the sense of defining a position jointly occupied by structurally equivalent actors. In the four-clique example above, sex as a parameter does classify and differentiate children in the classroom. However, it also obscures classroom differentiation. Separating boys and girls as two groups would show higher within group relations than between group relations, since each group would contain two cohesive cliques and no clique would cut across the two groups. But each group homophilous on sex is differentiated into two completely separate cliques that are capriciously combined by the sex parameter.

Suppose that in addition to knowing the sex of children in the hypothetically cliqued classroom, data are also available on their involvement in the Little League and their intelligence as IQ scores. The following distribution of attributes across the four cliques is observed: Clique A is composed of *boys* with *high* involvement in the Little League and *low* IQ scores. Clique B is composed of *boys* with *low* involvement in the Little League and *low* IQ scores. Clique C is composed of *girls* with *low* involvement in the Little League and *low* IQ scores. Clique D is composed of *girls* with *low* involvement in the Little League and *high* IQ scores.

With these three variables; sex, involvement in the Little League, and IQ score, actor attributes can be used to define structural equivalence. The first male clique is composed of sports enthusiasts while the second is composed of boys not involved in sports. The first female clique is composed of unintelligent girls while the second is composed of intelligent girls. Note that any one of the variables alone would differentiate the children in the sense that there would be zero interaction across groups and some interaction within groups—but it would also group together separate cliques. Intelligence erroneously groups together the first three cliques and involvement in sports erroneously groups together the last three cliques. Similarly, no two variables together are able to accurately differentiate cliques, even though any two variables taken together would distinguish separate cliques. It is only when all three variables are considered simultaneously that they are able to accurately capture the four classroom cliques.

This illustrates what will be a useful concept: Refer to the set of structural parameters sufficient to accurately distinguish the status/role-sets in a system as the system's "parameter set." Each parameter in the set need not define structurally equivalent actors. However, some combination of values on the parameters in the set must define each status jointly occupied by structurally equivalent actors. Sex, involvement in sports and intelligence constitute the example classroom's parameter set. Any one of the three variables is unable to accurately distinguish the four cliques, however, each clique has a unique pattern of values on the three variables. At the same time, all combinations of parameter values need not define a status. In the four-clique classroom, for example, there are no groups composed of intelligent boys, nor any groups composed of sports-minded girls. The key to using actor attributes as structural parameters defining jointly occupied statuses is to detect which combinations of values in a system's parameter set actually define statuses. This requires some initial information on relations as they are associated with actor attributes in a system's parameter set.

THE STATUS/ROLE-SET AS AN ERSATZ NETWORK POSITION

Two types of information are required in order to construct a survey instrument for eliciting data sufficient to define what I shall term ersatz network positions; an enumeration of the variables in the parameter set for the population from which respondents will be sampled and an enumeration of types of relations that define networks in the population. For the purposes here, I focus on relations within a single network and let Q refer to the number of possible different combinations of values on variables in the population's parameter set, i.e., Q equals the product across variables in the set of the number of values on each variable. The purpose of the survey instrument will be to obtain data from each respondent on the attributes of persons to whom he goes for each type of relation to be studied. Such data have been obtained successfully in several surveys to date, e.g., Laumann, 1966, 1973; Wellman, 1979, but the most extensive study to date is the Northern California Community Study conducted by Claude Fischer (1981) and his colleagues. Each respondent was asked to name persons who were contacts for 10 important types of social exchange, ranging from discussing personal worries, to borrowing a large sum of money, to discussing hobbies. On average, respondents named 19 different persons across all 10 types of exchange relations; some persons being the source of only a single type of exchange while others were the source of multiple types. Given a list of persons named as the object of relations from a respondent, it is a simple matter to obtain attribute data on each person named where attributes are taken from the parameter set for the population under study.[4] Focusing on a single network for the purposes here, let r_j be the number of persons to whom respondent J gives a sociometric citation during his interview. Also obtained during the interview are parameter attributes for each person cited and for the respondent himself.

Theoretically, some combination of attributes in the population parameter set defines each jointly occupied status in the population. Instead of locating statuses directly by analyzing relations among individual actors, as is usually done in network analysis, it should be possible to locate them indirectly by detecting those combinations of attributes which are associated with structurally equivalent respondents.

RELATIONS BETWEEN COMBINATIONS OF ATTRIBUTES

Consider a population with a parameter set composed of two variables, each with some number of values for a total of Q* different possible combinations in the population (Q* equals the number of values of variable one, times the number of values of variable two). The extent to which respondent J goes to persons with attribute I on the first variable in the parameter set and attribute T on the second variable (goes to persons with attribute combination IT in other words), is given as:[5]

$$z_{J,IT} = f_{J,IT} / r_J, \qquad [5]$$

where r_J is the number of citations J makes (his network range) and $f_{J,IT}$ is the number of those citations he directs to persons with the attribute combination IT. The relation $z_{J,IT}$ is the proportion of respondent J's citations directed to persons who have attributes I and T.[6] There will be Q* possible relations defined for respondent J by equation 5; one relation from him to each of the Q* different combinations of attributes in the population's parameter set.

There are two further features of equation 5 that merit special mention. First, since each respondent and each cited person corresponds to a single combination of attributes, respondent J's relations sum to 1 across the Q* combinations of attributes in the parameter set: $1 = \Sigma_q z_{J,q}$. The fact that each person maps into a single combination of attributes also limits the number of the Q* possible attribute combinations actually observed. The number of attribute combinations to which respondent J has nonzero relations has a maximum equal to the number of persons he cites. The actual number of attribute combinations to which he has nonzero relations will be lower than this maximum to the extent that he cites some persons with the same attributes.

Let Q equal the number of different attribute combinations observed in a sample. This would include respondent attributes as well as those of persons cited. The maximum value of Q would be the sample size plus the total number of different persons cited by respondents in the sample. In the typical network study, for example, Q would equal N, the order of the network under study, Z_{NN}. In practice, of course, Q will be much smaller than this maximum. It would only reach its maximum if every respondent and every person cited by any respondent, had a unique combination of attributes. Since relations tend to occur among socially homophilous persons, the likelihood of Q reaching its maximum seems low. At the other extreme, if all respondents have the same attributes and all of their citations are to persons with those attributes, then Q equals its minimum of one; indicating that the respondents are members of a completely homophilous system.

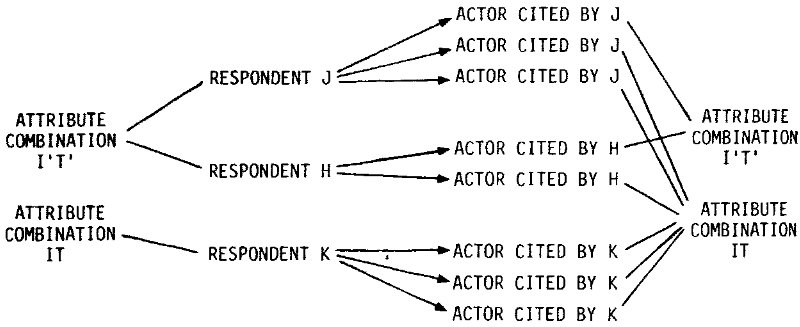
The respondent to attribute combination relations defined by 5 can be used to generate a Q by Q matrix of attribute combination to attribute combination relations. First, relations can be summed across respondents with identical attributes:

$$f_{(i't',it)} = \Sigma_j (z_{j,it})(\delta_{j,i't'}), \qquad [6]$$

where $\delta_{j,i't'}$ is a dummy variable equal to zero unless respondent J has attribute combination I'T', whereupon it equals one. Element $f_{(i't',it)}$ ranges from zero (if no respondent with attribute combination I'T' cites any person with attribute combination IT) up to a maximum equal to the number of respondents who have attribute combination I'T' (which occurs if every respondent with that combination of attributes only cites persons with attribute combination IT). Then dividing equation 6 by its maximum, $n_{i't'}$, the number of respondents with attribute combination I'T' yields, as the mean proportion of their citations, persons with attribute combination I'T' direct toward persons with combination IT:[7]

$$z_{(i't',it)} = f_{(i't',it)}/n_{i't'}. \qquad [7]$$

Figure 2 illustrates the logic of the above operation as a modification of the typical procedure illustrated in Figure 1. Each respondent and each person cited by a respondent, maps into a single combination of attributes; respondents J and H have attribute combination I'T' in Figure 2. According to the defini-



$$z_{(i't',it)} = (2/3 + 1/2)/2$$

$$Z_{QQ} = \begin{bmatrix} .42 & .58 \\ .00 & 1.0 \end{bmatrix} \begin{matrix} I'T' \\ IT \end{matrix}$$

Figure 2: The Relation from Attribute Combination I′T′ to Attribute Combination IT in Terms of Citations by Survey Respondents to Unknown Actors (Actors Attributes Indicated by Solid Lines and Citations Given by Arrows, $z_{(i't',it)}$ Defined in Equation 7)

tion of a system's parameter set, actors with identical attributes within the set are structurally equivalent. Therefore, a respondent's citations to persons with identical attributes can be aggregated in 5 as citations to structurally equivalent actors. Similarly, relations from respondents with identical attributes can be aggregated in 7 as relations from structurally equivalent actors. In Figure 2, respondent J cites 3 persons, 2 of whom have attribute combination IT, so that the proportion of his citations given to persons with that combination is 2/3. Similarly, respondent H gives 1 of his 2 citations to persons with that attribute combination. Respondents J and H have attribute combination I′T′, so that the average of their individual relations to persons with attribute combination IT gives the mean proportion of citations persons with attribute comination I′T′ direct toward persons with combination IT; .58 = (2/3 + 1/2)/2. The matrix of relations among attribute combinations, Z_{QQ}, is given in the diagram.

RESPONDENT ERSATZ NETWORK POSITIONS

Respondent J corresponds to a single row and column of the matrix Z_{QQ} as defined by equation 7; the row and column containing his combination of attributes in the parameter set. I refer to respondent J's ersatz network position as his pattern of relations with each of the Q attribute combinations; Q relations from him to persons with each combination of attributes, and Q relations to persons with his attributes from respondents generally;

$$Z_j = (z_{j1}, z_{j2}, \ \ldots \ z_{jQ}, z_{1j}, z_{2j}, \ \ldots \ z_{Qj}), \qquad [8]$$

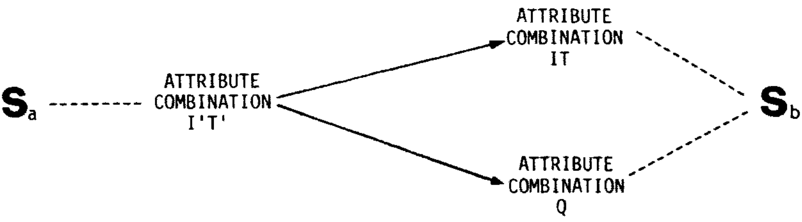
where relations from the respondent are given by the $z_{j,q}$ in 5 and relations to him (as a person with attribute combination IT) are given by the z_{qj} in 7. A respondent's ersatz network position corresponds to the idea of a position defined for typical network data by equation 1, except that instead of containing relations to and from other actors, the ersatz relational pattern consists of relations to and from attribute combinations in a parameter set. Given a relational pattern for each respondent, the distance between respondents I and J can be computed from equation 2. This provides a check on the assumption that actors with the same attribute combination are structurally equivalent. The (n_{it}, n_{it}) covariance matrix computed from the N distances to each of the n_{it} respondents having attribute combination IT should have a rank of one where N is the number of respondents interviewed (see Burt, 1980: 107-109, for references and detailed discussion). This test only considers differences in the relations respondents direct toward other actors since differences in relations to respondents with the same attribute combination are identical by definition in 7. If the hypothesis of structural equivalence is rejected, then the parameter set has not been correctly specified. In order to accurately differentiate statuses in the population, either more variables are needed or more distinctions are needed on the variables already specified.[8]

STATUS ROLE-SETS

A status in the population social structure is jointly occupied by structurally equivalent actors; actors with similar relations to every other status and similar relations from those statuses. In terms of the parameter set, statuses are defined as structurally equivalent combinations of attributes; attribute combinations that are similarly the object of relations from respondents with each combination of attributes and are found in respondents who have similar relations to each attribute combination.

Consider the interattribute relations defined by equation 7. The extent to which attribute combinations I′T′ and IT are similarly the object of relations is given by the sum of squared differences in their respective column elements of Z_{QQ}; $d_c^2 = \Sigma_q(z_{q,I'T'} - z_{q,IT})^2$. The extent to which respondents with these two combinations of attributes have the same tendency to cite persons with each different attribute combination is given by the sum of squared differences in corresponding elements of rows in the matrix; $d_r^2 = \Sigma_q(z_{I'T',q} - z_{IT,q})^2$. A status can now be defined as a set of attribute combinations separated by zero distance; i.e., for all pairs of attribute combinations I′T′ and IT proposed as defining the same status; $0 = d_{(I'T',IT)} = d_{(IT,I'T')} = (d_c^2 + d_r^2)^{1/2}$.

In fact, relations are likely to be measured with error so that some search procedure in the form of a computer algorithm will be needed in order to locate structurally equivalent attribute combinations. If Q is sufficiently small, standard network analysis algorithms can be employed. I have reviewed these elsewhere (Burt, 1978, 1980). Also, the matrix F defined by 6 can be analyzed as a Q by Q frequency table in order to determine which



$$z_{ab} = \frac{z_{(i't',it)} n_{it} n_{i't'}}{n_a n_b} + \frac{z_{(i't',q)} n_q n_{i't'}}{n_a n_b}$$

$$n_a = n_{i't'}$$

$$n_b = n_{it} + n_q$$

Figure 3: The Relation from Status S_a to Status S_b in Terms of Relations Between Attribute Combinations (Occupancy Given by Dashed Lines and Inter-Attribute Relations Given by Arrows, $z_{(i't',it)}$ Illustrated in Figure 2, z_{ab} Defined in Equation 9)

categories can be collapsed into one another, e.g., Duncan, 1975; Goodman, 1979. When Q is very large, an iterative data analysis would be necessary (see Comments section).

Fortunately, statuses proposed by any algorithm can be statistically tested. Distances from each of the N respondents to the ersatz network position of each of the n_a respondents proposed as occupying S_a can be computed via equations 2 and 8. Under the hypothesis that these n_a actors are structurally equivalent as occupants of S_a, the (n_a, n_a) covariance matrix among distances to the n_a occupants will have a rank of one, e.g., Burt, 1980: 107-109). Since the distances being correlated are computed for individual respondents, this test simultaneously assesses the extent to which status occupants with identical attributes are structurally equivalent and the extent to which occupants with different attributes are structurally equivalent.

An (M, M) matrix of ersatz role relations can now be computed where M is the number of different statuses defined by the matrix Z_{QQ} and element (a, b) is the mean relation from occupants of status S_a to occupants of status S_b:

$$z_{ab} = \Sigma_{i't'} \Sigma_{it} (z_{(i't',it)}) (n_{i't'} n_{it}) / n_a n_b, \quad [9]$$

where $z_{(i't',it)}$ is defined by [7], $n_a = \Sigma_{i't'} (n_{i't'})$, $n_b = \Sigma_{it}(n_{it})$, and summation is across all attribute combinations I′T′ that define status S_a and all attribute combinations IT that define S_b. Where $n_{i't'}$ and n_{it}, respectively, are the numbers of respondents with attribute combinations I′T′ and IT, n_a and n_b, respectively, are the numbers of respondents occupying S_a and S_b. The density in 9 corresponds to that in 3, in the sense that both are average relations between sets of structurally equivalent actors. Building on Figure 2, Figure 3 illustrates the logic of the operation in 9 so as to be comparable to Figure 1 illustrating equation 3. Figures 1 and 3 differ in two ways. Figure 1 show actors being mapped into unique statuses—each actor mapping into a single status, while Figure 3 shows actor attribute combinations being mapped—each attribute combination mapping into a single status. Second, Figure 1 shows each pair of actors being weighted equally in determining the interstatus relation; that weight being $1/n_a n_b$. It would not be appropriate to weigh attribute combinations equally in determining interstatus relations, since the combinations need not occur with equal frequency in the population. Therefore, equation 9 weighs the pair of attribute combinations I′T′ and IT by the extent to which they are observed in the sample; the extent to which all respondents in status S_a have attributes I′T′ ($n_{i't'} / n_a$), and the extent to which all respondents in status S_b have attributes IT (n_{it} / n_b). Given the citations made by occupants of status S_a, z_{ab} in equation 9 is the mean proportion they give to occupants of S_b. The role-set defining S_a as an ersatz network position is now given by the 2M relations in row and column A of the density table defined by 9 as is the case in a typical network analysis (equation 4). With these data on status/role-sets for population social structure and data on the relational patterns for actors representative of the population (equation 8), the wide range of hypotheses regarding network positions can be used to predict respondent attitudes and behaviors.

COMMENTS

I have argued that, in theory, data obtained in a standard survey research design can be used to describe the network positions of each respondent in the social structure of a large population from which each is randomly drawn. Of course, these are ersatz representations of the positions typically captured by models of network structure. Relations to and from a respondent are not estimated with equal precision. As captured by 5, relations from a respondent to status occupants in the population are based on ties involving him personally. However, relations to him from status occupants are aggregate relations to his combination of attributes rather than to him personally (equation 7). Respondents with identical attributes are therefore the object of identical relations by fiat. This aggregate quality in 8 as a representation of a respondent's relational pattern makes it an ersatz representation of his pattern as it exists (equation 1). Its ersatz character notwithstanding, the availability of relational patterns defining respondent positions and statuses in large population social structure means that structural theory stated in terms of network concepts can be used to inform standard survey data.

Of course, theory and practice are not the same thing. I do not expect ersatz network positions to be used in practice in the same manner that I have introduced them in theory. In practice, a good deal of iterative data analysis will be required in order to specify the parameter set.

I have assumed that the Q attribute combinations specified as the parameter set for a population accurately distinguish statuses in the sense that each status in the population social structure corresponds to at least one of the Q attribute combinations. In practice, the variables defining these Q attribute combinations will be selected and coded according to hunches an investigator has about attributes likely to be associated with jointly occupied statuses in the population under study. Since these hunches can be wrong, it is quite possible that a system's a priori specified parameter set is misspecified.

There is a serious problem with underspecifying the parameter set, i.e., failing to consider one or more attributes that actually do stratify the population. Some statuses in the final model would then refer to structurally nonequivalent actors. These actors stratified by deleted parameters would appear to be structurally equivalent (see note 8).

However, there is no problem with overspecifying the parameter set, i.e., including more attributes in the set than are actually necessary to distinguish population statuses. When in doubt about the importance of an attribute as what Blau terms a structural parameter, therefore, it is wisest to err on the side of including too many attributes, rather than too few. These additional parameters will merely be deleted from the final model in much the same manner that insignificant predictors are deleted from a regression equation. If an attribute is negligible, it will not affect the structural equivalence of respondents; respondents homophilous on the attribute will be as likely to be as structurally equivalent as are respondents heterophilous on the attribute. It might appear that overspecifying the parameter set would lead to problems in data processing. A necessary, but not sufficient, condition for Q increasing is an increasing number of distinctions in the parameter set. As more combinations of attributes become possible, more combinations could be observed in fact. Since the matrix of interattribute relations defined by 7 is a Q by Q matrix, overspecifying the parameter set could result in the matrix F (equation 6) and the matrix Z (equation 7) being too large to analyze with routine computer packages.

This potential problem is easily circumvented by iteratively analyzing the ersatz relational patterns in order to discover those combinations of attributes actually defining statuses. The ersatz network positions of respondents can be computed, for each respondent (equation 8) where Q refers to all observed combinations of attributes in the possibly overspecified parameter set. The extent to which specific combinations are actually stratifying the population can now be tested by testing the structural equivalence of respondents homophilous on the combination. For example, suppose race is assumed to be the key variable stratifying a population under study where race is coded into four attribute categories; Black, Asian, Chicano and White. Under the hypothesis that race defines statuses in the population, respondents homophilous on the variable should be structurally equivalent. Consider the Asian respondents. The hypothesis that all Asians are structurally equivalent can be statistically tested by computing the rank of the covariance matrix among distances to them from each respondent (see "respondent ersatz positions," in the text). If this hypothesis is rejected, then Asians are themselves stratified in terms of further attributes.

Either of two strategies could be adopted in order to correctly respecify the attribute category "Asian." The investigator could guess at the missing attributes and then test the hypothesis that Asians homophilous on these further attributes are structurally equivalent. Alternatively, distances among the ersatz network positions of the Asians could be computed and subjected to an exploratory network analysis in order to locate structurally equivalent Asians. I have reviewed this work elsewhere (Burt, 1978, 1980). Given groups of structurally equivalent Asians, discriminant function analysis or one-way analysis of variance could be used to locate attributes most homophilous among structurally equivalent Asians (within groups) and most heterophilous between structurally nonequivalent Asians (across groups). Once a set of attributes has been located as the basis for stratification among the Asians, the hypothesis of structural equivalence can be tested for each set of Asians homophilous in regard to each attribute combination. The same procedure could be repeated for the other racial groups. Of course, any attribute combination could be used as an initial parameter set, depending on one's initial hunches; some systems obviously being stratified by race, some being stratified by occupation, some being stratified by race and occupation jointly, and so on. By successively testing the structural equivalence of respondents homophilous on attributes in a tentative parameter set, attribute combinations actually stratifying the population can be uncovered.

In other words, the parameter set in terms of which erstaz network positions are estimated need not be known prior to conducting a survey (although all possible attributes in the set must be known) and there need not be a single parameter set distinguishing statuses in all subgroups of the population, e.g., Blacks, Asian, Chicanos and Whites. Statuses in one subgroup might be defined by attributes different than those defining statuses in another subgroup. The parameter set for the population from which a specific sample has been drawn can be uncovered by iterative data analysis; first hypothesizing a specific parameter set, then testing the structural equivalence of respondents homophilous in regard to attribute combination distinguished in the set, then respecifying the initial parameter set so that it accurately defines structurally equivalent respondents.

NOTES

1. The phrase "distinct actors" is important here. If individual actors are aggregated in some manner so that the number of distinct actors being studied is small, then network studies can be conducted on large populations such as actors in a community (Laumann, 1966; Laumann and Pappi, 1976: Ch. 12; Burt et al., 1980), manufacturing establishments in the American economy (Pfeffer, 1972; Burt et al., 1980), or persons and corporations in the United States during the last century (Burt, 1975; Burt and Lin, 1977). This aggregation, of course, eliminates the possibility of analyzing networks at the level of individual actors.

2. Specifically, the proportion of a network ignored by interviewing k% of the actors in the network is; $1 - kN^2/N(N-1)$. As system size increases, the term $N^2/N(N-1)$ approaches 1, so the proportion ignored is roughly $1-k$. To be sure, most of the relations in this network will be absent; a small number of residents are actually connected to any one resident (see Wellman, 1979). The problem is knowing where those few actual relations occur in the network as a whole.

3. It might seem that this argument can be turned back on theory based on structural equivalence, since, a priori to observing a system, one does not know who is structurally equivalent to whom. Once relations are observed, however, there are explicit conditions under which a set of actors are structurally equivalent occupants of a single position.

At this point, falsifiable hypotheses are stated in terms of role-sets defining statuses. In contrast, the observation that a set of actors is homophilous in terms of one or more attributes as a surrogate for structural equivalence demonstrates no more than the fact that actors in the set share attributes. They may or may not be structurally equivalent in addition to sharing those attributes. Thus, theory derived from attribute homophily is not deductive theory concerned with positions as statuses in social structure. It is deductive theory concerned with attribute homophily which can be extended with uncertainty (and thereby transformed into inductive theory) to hypotheses regarding positions in social structure.

4. In order to obtain more complete information on the link between the respondent and the named persons, e.g., how did you meet this person, how long have you known this person, how often do you get together with this person, and so on, Fischer focuses on three to five of the persons cited by a respondent. The information required here is less extensive on each cited person, but is required for all persons cited.

5. I wish to make a special acknowledgement to James A. Wiley here for suggesting the ratio in equation 5 in lieu of an inelegant formulation I had used, based on the geometric mean.

6. At this point, multiple networks require special mention. The relation from respondent J to persons with attribute combination IT across multiple networks can be given as: $z_{j,it} = (\Sigma_k f_{j,it,k})/(\Sigma_k r_{jk})$, where summation is across the K different networks (types of relations) in which citations are obtained, r_{jk} is the number of persons respondent J cites within network K, and $f_{i,it,k}$ is the number of those persons who have attribute combination IT. As is typically done in multiple network models, each network is given equal weight. This representation further gives equal weight to the number of persons cited with attribute combination IT and the number of different types of relations directed at that combination. If one person with attributes I and T is cited in regard to each type of relation (K citations), for example, the above relation would be the same as if K persons with attributes I and T were cited in regard to only one type of relation. Alternatively, it seems reasonable to estimate $z_{j,it}$ across multiple networks as the number of persons J cites who have attributes I and T divided by the total number of persons he cites. This measure has an uncomplicated substantive meaning as the proportion of J's ego network that has the attributes I and T. The above representation seems preferable, however, because a person who is a source of multiple types of exchanges should be a more significant part of J's network than a person who is a source for only one type of exchange (see Burt, 1980: 89-90, 95-96, for references).

7. Although equation 7 divides the frequency of citations made by the maximum possible, it is not—strictly speaking—a density measure, since it cannot take into account the number of persons who could have possibly been cited. Equation 7 is quite correct as a proportion measure of relationship, but provides a biased density measure. If there are few actors in the population with attribute combination IT, then there can be few citations to that attribute combination from any one respondent so 7 will be low. In this case, a density measure of the same relation would adjust the mean upward by taking into account the few actors available to be cited. Similarly, 7 overestimates the density of citations to an attribute combination that is often observed in the population.

8. On the other hand, failure to reject the hypothesis of structural equivalence does not verify the parameter set as specified. If a significant attribute is deleted from the set, then respondents who differ only on the deleted attribute would appear to be structurally equivalent, when, in fact, they are not. As is the case when structural equivalence is tested in typical network studies, the equivalence of ersatz network positions can never be proven. Rather, it can only be rejected for a given set of relational patterns.

REFERENCES

BLAU, P. M. (1977a) "A macrosociological theory of social structure." Amer. J. of Sociology 83: 26-54.

——— (1977b) Inequality and Heterogeneity. New York: Macmillan.

——— (1974) "Parameters of social structure." Amer. Soc. Rev. 39: 615-635.

BURT, R. S. (1981) Toward a Structural Theory of Action: Network Models of Stratification, Perception and Action with a System of Actors. New York: Academic.

——— (1980) "Models of network structure." Annual Rev. of Sociology 6: 79-141.

——— (1978) "Cohesion versus structural equivalence as a basis for network subgroups." Soc. Methods & Research 7: 189-212.

——— (1977) "Positions in multiple network systems, part one: a general conception of stratification and prestige in a system of actors cast as a social topology." Social Forces 56: 106-131.

——— (1976) "Positions in networks." Social Forces 55: 93-122.

——— (1975) "Corporate society: a time series analysis of network structure." Social Sci. Research 4: 271-328.

——— and N. LIN (1977) "Relational time series from archival records," pp. 224-254 in D. R. Heise (ed.) Soc. Methodology 1977. San Francisco: Jossey-Bass.

BURT, R. S., K. L. LIEBEN, and M. G. FISCHER (1980) "Network power structures from informant perceptions." Human Organization 39: 121-133.

BURT, R. S., K. P. CHRISTMAN, and H. C. KILBURN (1980) "Testing a structural theory of corporate cooptation: interorganizational directorate ties as a strategy for avoiding market constraints on profits." Amer. Soc. Rev. 45: 821-841.

DUNCAN, O. D. (1975) "Partitioning polytomous variables in a multiway contingency analysis." Social Sci. Research 4: 167-182.

FISCHER, C. S. (1981) Private Worlds. Chicago: Univ. of Chicago Press.

FRANK, O. (1975) "Survey sampling in graphs." Research Report 1975-11, Department of Statistics, University of Lund.

GOODMAN, L. A. (1979) "A brief guide to the causal analysis of data from surveys." Amer. J. of Sociology 84: 1078-1095.

GRANOVETTER, M. (1976) "Network sampling: some first steps." Amer. J. of Sociology 81: 1287-1303.

LAUMANN, E. O. (1973) Bonds of Pluralism. New York: John Wiley.

——— (1966) Prestige and Association in an Urban Community. Indianapolis: Bobbs-Merrill.

——— and F. U. PAPPI (1976) Networks of Collective Action: A Perspective on Community Influence Systems. New York: Academic.

McCALLISTER, L. and C. S. FISCHER (1978a) "Studying egocentric networks by mass survey." Working Paper 284, Institute of Urban and Regional Development, University of California, Berkeley.

——— (1978b) "A procedure for surveying personal networks." Soc. Methods & Research 7: 131-148.

MORGAN, D. L. and S. RYTINA (1977) "Comment on 'network sampling: some first steps' by Mark Granovetter." Amer. J. of Soc. 83: 722-727.

PFEFFER, J. (1972) "Merger as a response to interorganizational interdependence." Admin. Sci. Q. 17: 382-394.

WELLMAN, B. (1979) "The community question: the intimate networks of East Yorkers." Amer. J. of Sociology 84: 1201-1231.

Ronald S. Burt is an assistant professor in the Sociology Department and a principal investigator in the Survey Research Center at the University of California, Berkeley. His interest is the general development of structural sociology.