SAMPLING AND SOCIAL NETWORKS:

RELATIONAL MEASURES IN

LARGE POPULATIONS

Steve Rytina

University of Michigan

December 1977

CRSO Working Paper #166

Copies available through:
Center for Research on
Social Organization
University of Michigan
330 Packard Street
Ann Arbor, Michigan 48109

"Survey research methods have often led to the neglect of social structure and the relations among individuals." With this sentence, Coleman (1958, p. 28) opened his attack on the atomistic methodology of survey research. Noting that as survey techniques developed, methodology had remained rooted in the individual as a unit of analysis, he observed that, "As a result, the kind of substantive problems on which (survey) research focused tended to be problems of 'aggregate psychology', that is, *within*-individual problems, and never problems concerned with relations among people." And so it remains today: the great bulk of survey-based research consists of the analysis of individual attributes.

Coleman's agenda has hardly been completed, but recent developments in the subdiscipline concerned with social networks provide the materials for attacking relational issues with surveys. Granovetter (1976) presented the methodological foundations for the investigation of large networks with the familiar methodology of the random sample survey. He showed that it was possible to obtain estimates of sociometric density, which can be defined as the probability that two randomly chosen people have a tie to each other.

The concept of density arises when all possible ties that might occur in a population are considered. With N people (or nodes) N(N-1) asymmetric ties are possible. (An asymmetric tie is one such that if I am tied to you, you need not be tied to me.) The number of symmetric ties is half that, or N(N-1)/2, which is also the number of pairs in the population. The density is the number of existing ties divided by the number of possible ties (for an extended discussion of the concept, see Niemeijer, 1973). A network with

many ties will have a high density, while one where ties are sparse will have a low density. Density can, therefore, be interpreted as the sociometric saturation of a network.

Measuring density is a substantial departure from previous efforts to investigate networks in large populations. Using Granovetter's technique, one could measure a relational property of a group. Previous techniques allowed researchers to investigate long chains of acquaintance and access (Milgram , 1967), or the immediate surroundings of a random sample of respondents (Laumann, 1973). But with Granovetter's technique, it became possible to think of networks as they extended within and between large categories in the population. Granovetter's own previous work had shown that most people have acquaintances or "weak ties" that exist far beyond their everyday haunts (Granovetter;1973, 1974). At last it could be investigated: Are some groups more close-knit than others? Are some groups disconnected from the rest of the population with whom they share a geographic region? Do social ties reflect the group boundaries in a population? Or, indeed, are the members of social categories so intermingled that, in the sense of tie segregation and differentiation, the categories hardly define groups at all?

These are exciting questions that hardly exhaust the possibilities for relational survey methodology. In the first part of this paper, further possibilities will be suggested. Some, which have been presented elsewhere, will be alluded to briefly. More attention will be paid to new issues, especially the possibility of employing relational measures where categorical attributes (or demographics) have been previously employed. Once we have expanded the pool of potential applications beyond what Granovetter discussed, we will turn to methodology proper. In the second section,we will compare Granovetter's subgraph sampling to an alternative which, we shall see, is often superior: direct questioning. In the final section, after establishing what is feasible, I shall outline a unified interpretation of survey relational data.

Relational Issues: Macro and Micro

A new set of theoretical issues is raised by considering aggregated social ties in large populations. David Morgan and I, while considering the application of Granovetter's method, came across a set of arithmetical relations between the typical amount of contact enjoyed by a group member, the relative size of the group, and the degree of segregation of ties (Rytina and Morgan, 1977b). The segregation might reflect a propensity of the group's members to select each other over others, or the unwillingness of others to associate with members of the group. Examples of outnumbered groups whose members probably direct the bulk of their ties inward to the group range from power elites to ghettoized, oppressed ethnic populations. Morgan and I found that such groups were objectively likely to be immersed in a sea of non-members, even while the density within their own group was substantially higher than the density in the surrounding world.

Blau (1977a, b) presents a similar argument and extends well beyond it to outline what he calls a primitive theory of social structure. Blau provides a new agenda for the investigation of social structure based, in part, on the evaluation of the quantity of social contacts within and between different social categories. Blau does not commit himself to an operational interpretation, but his analysis is unchanged by substitution of the within and between group densities that Granovetter's methodology has suggested.

When these notions are carried over to the individuals that make up groups, a whole new line of interpretation opens up. Just as groups could display variable degrees of nettedness, individuals could be connected to groups in varying degree. This leads to a reconsideration of the way membership in collectivities is typically measured.

It is very common in survey research to measure group membership by determining respondents' categorical attributes. Some of the attributes employed this way are among the most common measures of survey research, the demographics. An example of a demographic attribute that is often thought of as an index of group membership is race. Further examples include income, occupation, and education, which are often employed as indicators of social class. (Since the advent of the Blau-Duncan (1967) paradigm, sociologists have been more sensitive about applying these continuous attributes as measures of membership in discontinuous social groupings, but such usage still persists.) More obvious examples include area of residence, which is indexed by address, and ethnicity, indexed by the birthplaces of ancestors; all of these examples are attributes that are used as measures of the social categories to which people belong.

No one needs to be told that these measures often (statistically) explain more about other variables than anything else available. That is one reason why everyone who works with survey data is told sooner or later that they should "control" for the demographics. Whatever else can be said about these measures, they work. Therefore they deserve to be thought of seriously.

Sometimes these attributes actually refer to properties of the individuals in isolation from their social setting. There are, for example, those who believe that men and women differ in their hormonal makeup and therefore in their innate tendency toward aggression (Goldberg,1973). In a study of attitudes toward agression, men and women might be statistically distinguished in an attempt to control for the effects of individual biology. Fortunately, biological determinism is not the only rationale for interpreting group differences.

The more common view in sociology is that many categorical attributes are measures of membership in social groups. To be black, for example, is to have mostly other blacks as friends and relatives, probably to reside in a mostly black residential area, and to participate in mostly black associations. There are other implications of membership in this group. As an oppressed group, the social ties of its members to members of the dominant racial group are often ties of subordination. The meaning and impact of group membership is the result of the social relations that typify membership: both the relations to other group members, and the relations with members of other groups.

Both concerns are combined when the allocation of social relations is considered. An individual's relational life may in varying degree be contained within the boundaries of a group. This gives rise to a variable conception of group membership: the extent to which the individual is embedded within the group. Individuals are more embedded when more of their ties are to group members. For example, a concrete black individual is more embedded in the category of Afro-Americans to the extent that more of his/her ties are with other blacks, and to the extent that fewer of his/her ties are to non-blacks.

This notion — degree of embeddedness — contrasts with the idea of structural equivalence that is a central feature of the work of White and his co-workers (Lorain and White, 1971). Our concern is the degree to which categories defined by the possession of attributes act as sociometric attractors to their members, and not with the extent to which category members display similar patterns of relationships.

White's notion is deeper but requires more elaborate data, more elaborate methods, and is not obviously suitable for sample survey application (White, et. al. 1976; Boorman and White, 1976). Some parallels are present, and they merit discussion.

Group membership, or shared identity, could arise from two processes; one process is frequent contact with similar others, and the other process is similar reactions from dissimilar others. These may be illustrated with the example of sex. Females are not strikingly segregated from males in American society. Typical patterns of contact probably reflect the biological sex ratio of (approximately) one to one; that is, most people, male and female, probably know and associate with males about as much as females.

Specialization of certain types of relations is undoubtedly present. For example, most people have most of their physical sex with members of the opposite sex. In contrast, certain kinds of conversations (or more abstractly — social exchanges) are confined within sex groups. Expectations about the "appropriate" allocation of social relations between same sex and opposite sex pairings comprise the sex role structure of the society.

The impact of expectations can extend into concrete patterns of relations. Kanter (1977) has described in detail how females in mostly male occupational settings are affected by the predominant sexual expectations of the larger society. Females in such settings may have patterns of contact with males that are superficially similar to male patterns of contact, but the contents of the contacts are substantially different.

The sexual identities of such women are not simply a result of their patterns of contact but also of the ways those contacts react to them. The identities are formed not by a sharing of experience in contacts with each other, but through a shared experience imposed by the "typical" male response.

In White's terms, women in such settings enjoy (or suffer from) structural equivalence. We have used the term "structural equivalence" more informally than White has, so that we may distinguish the present approach from his.

White would have us search through the patterning of multiple relations to uncover persons or social positions with similar patterns. Such patterns could be related to attributes. To the extent that the possession of an attribute — race and sex being prime examples — imposes limitations on that patterning, there exists a racial or sexual role structure that imposes experiences on certain categories of people. Even if the people were not aware of their structural equivalence, and even if they were not in contact to develop shared understandings of their experiences, the similarity of their condition would cause many similarities in behavior, ideology, and so forth.

The present approach goes after something simpler and more operationally accessible. Attention is directed away from multiple relations to more grossly specified ties of sociation such as acquaintance and friendship. The issue is the allocation of such ties of sociation within and between different nominal categories. The concentration ot dispersion of such ties is a fundamental property of such categories as social entities.

Of course, some things are lost by this simplification. The deeper structure of social networks as patternings of relationships — not just around ego, but beyond into the social distance — is the first casualty. A closely related conventional notion, that group identity is the product of the reactions of others and not of interactions among group members, is also given short shrift. But something is gained. These methods can be easily applied by researchers versed in current survey technology.

Even though the degree of embededness is less deep than the patterning of relations sought by White, it still encompasses an interesting range of possibilities. People may be embedded in the groups to which their nominal attributes assign them, but they need not be. In the relational sense, they may be embedded in no groups or equally in many groups. These possibilities

arise when membership is treated as a variable matter of degree.

The categorical composition of social ties is a general, if clumsy, expression that encompasses this range of possibilities. From the standpoint of an individual, the world is divided into social categories. For a given type of tie, the numbers of tiesto each of those categories is the categorical composition. Abstractly, the categorical composition is a vector. For a given individual, type of tie, and category set, the vector has an entry for each category in the set. The magnitude of each entry is the number of ties the individual has to members of that category.

The variable degree of membership in a categorical group is indexed by the vector entry that refers to the group. For example, to examine membership in the category 'black', one looks at the number of ties to 'blacks'. The vector entry may be manipulated in a number of ways. It might be be divided by the size of the target group to index the improbability of knowing that many group members. Or it might be compared to the sum of the ties to other groups to determine the proportion of ties contained within the nominal group. In either case, the problem of membership centers attention on the single entry corresponding to a single group, and throws away much of the other information in the vector that refers to other groups.

One use of the information in the complete vector is to calculate the within and between group densities, whose theoretical importance was mentioned above. The within group density focuses on the same vector entry that indexes membership. Summing that entry across all members of the group gives the total number of ties between group members. (If the tie is symmetric, then every instance has been double-counted, and the sum should be halved.) Density is simply this sum divided by the possibilities, or N(N-1) where N is the size of the group. Between group densities are calculated in an analogous way, summing all the ties from group A to group B, and then dividing by N_aN_b, the product of the group sizes.[2]

These two uses of categorical compositions suggest that interpretations at both the individual and the aggregate level will be simultaneously available in a survey application. The individual categorical composition is a measure of attachment to groups. With a sample of such observations, the impact of attachment could be analysed. At the same time, with the same data, the densities of inter and intra group contacts could be estimated. These densities have the macro interpretations suggested by Rytina and Morgan and by Blau. The densities can be estimated from a sample becuase a density is a sort of average, over pairs of persons rather than persons as individuals, and, like any population average, a density can be estimated by the mean of the sample observations. In practise, a survey methodology for the estimation of categorical compositions will serve double duty, since the survey data may be used to estimate group level and individuallevel parameters at the same time. The twin uses of these measures will be a recurrent theme in the next section, where two methods of measurement are compared.

Measuring Categorical Compositions

In this section, two methods for assessing categorical compositions will be compared: the method of direct questions, and the method of random subgraph samples. Each method shall be defined in a moment. They will be compared in terms of costs, reliability, and applicability. Costs are conceived as an increasing function of the number of questions asked and of sample size. Reliability will be investigated with models of sampling error, which will themselves have implications for sample size and therefore costs. Applicability will be considered more informally in terms of the types of categories, ties, and populations where each method seems most suitable.

Because the methods are quite different, they lead to different problems. That will make it hard to discuss them in parallel. To make the issues as clear as possible, the presentation will open with sampling models, where direct comparisons are easy. This will require the adoption of certain qualifying assumptions. In particular, I shall assume, except when otherwise noted, that the tie in question is acquaintance. And I shall also assume that respondents are able to answer, without error, any question that might be asked. <u>This latter assumption is obviously contrary to fact</u>. After the sampling models are presented, I shall clarify the implications of relaxing this assumption.

The most straightforward approach to the measurement of categorical compositions is the direct question method. A sample is drawn and each respondent asked: "To how many members of category X do you have tie Y?" Each respondent would need to answer C*T questions where C is the number of categories and T is the number of types of ties.

Three drawbacks to this approach are readily apparent. First, the categories employed must be salient and consensually understood by all respondents.[3] Therefore one could not make use of sociological constructions like the Duncan SEI but would have to limit oneself . to categories in common usage among respondents. Second,one would have to pre-select categories when constructing the questionnaire, rather than classifying choices after the data was collected. Third, there is no obvious way of knowing how accurate people's choices will be.

An alternative method, randomly sampled subgraphs, was presented by Granovetter (1976). Granovetter intended his technique for investigations of density of acquaintance in single populations, but in principle, the method could be applied to any type of tie and to subgroups within populations (see Granovetter, 1977). As was noted above, the determination of density values and of categorical compositions is much the same problem.

To employ Granovetter's technique, the surveyor begins by drawing a sample of names. In addition, it would probably be useful to have pictures to go with each name, since a name like "John Smith" does not identify a unique individual. Each member of the sample is then asked about every other member of the sample. Granovetter implicitly limited questioning to the presence or absence of acquaintance. But there is no reason (except interview time.But see below for the probable rarety of claimed ties.) not to employ follow up questions to determine the strength and content of claimed ties. In effect, the surveyor can obtain complete information on a sampled subgraph of the population graph. Using the tools of statistical inference, hypotheses could be evaluated for the population graph (or network) from which the sample was drawn.

At first sight, this method is very attractive. Since it requires an interview with each member of the list of possible choices, which is also the sample, one could easily and conveniently obtain information about each respondent's tie objects. One would then know, for the sampled pairs, how many doctors had a tie to how many lawyers, or how many people

with SEIs above 70 have a given relation with other sample members having SEIs of less than 70. More generally, one could calculate an estimate for the density of ties within and between any categories that might be used to classify members of the sample. The practical question is the sample size necessary for statistical inference.

To discuss issues of sample size and inference, a statistical model is needed. I will begin with a simplified situation considered by Granovetter. Only a single population, not subdivided into categories, is assumed. It is also assumed, as mentioned above, that the tie is acquaintance and that respondents give accurate answers. Under these assumptions, the observed density in the subgraph is an unbiased estimator of the density in the population (Granovetter,

The quality of the estimate depends on the sampling variance of the estimator. Granovetter presented an expression for this variance:

$$(N-n)n(n-1)(n-2)s^2(a)/\binom{n}{2}^2(N-1)(N-2)(N-3)$$
$$+\ (N-n)(N-n-1)n(n-1)s^2(C)/\binom{n}{2}^2 2(N-2)(N-3)$$

which he obtained from Frank (1971). [The notation is Frank's, and I shall follow it throughout.] N is the size of the population; n is the size of the sample; $s^2(a)$ is the variance of the number of ties that each person has to the other members of the population; and $s^2(C)$ is the variance of ties in the graph. Since a tie is either present or absent, $s^2(C)$ is a binomial variance and could be written $\overline{P}(1-\overline{P})$ where $\overline{P}$ is variously the density, the chance that a given randomly selected person has a tie to a given other randomly selected person, or any appropriately calculated average of the numbers of ties in the population.

Although this expression is exact, it is not very intuitive. Frank's derivation is quite complex and is based on the consideration of 15 contributing possibilities. Neither the expression nor its derivation make clear the contribution of the individual level measurements (which are our relational measures) to the total variance. The expression is not well suited for consideration of disaggregation to the individual level measurements, or for considerations of reaggregation to the group level averages. Therefore I shall consider an alternative large sample derivation that is better suited for the issues at hand. (For full details of the derivation, see Appendix.)

At the individual level, subgraph sampling can be modeled as a Bernoulli or coin-toss experiment. Imagine that some person, i, has T_i ties to other members of the population. The chance that they have a tie to any randomly chosen individual is T_i/N, which I shall call P_i. Since each sampled individual is presented with n-1 randomly chosen persons, the observed number of ties is distributed as a binomial with probability P_i and n-1 trials. Therefore, the individual level observations have expected value $(n-1)P_i$ and variance $(n-1)P_i(1-P_i)$.[4]

Frank and Granovetter were concerned with estimating the total number of ties, and the density, respectively. (Recall that these are simple transformations of each other when population size is known.) To obtain the total number of observed ties, we must add up all the individual level observations. The variance of this sum is the desired sampling variance. To make our results comparable to Granovetter's, we assume one-way observation and symmetry. That is, we observe each tie from only one end, so that if we ask Smith if he is tied to Jones, we do not also ask Jones if she is tied to Smith.[5] When we take the variance of the density estimator $\left(=\frac{\text{number ties observed}}{\text{number of observations}}\right)$, we find that it approximately equals: $4s^2(a)/nN^2 + 2s^2(C)/n^2$. The notation is as before where $s^2(a)$ is the variance of the T_i, and $s^2(C)$ is $\overline{P}(1-\overline{P})$, and $\overline{P}$ is the average value of P_i for the population. This expression, suitably enough, is the variance expression that Granovetter obtained from Frank, except that N has been substituted for N-1, N-n-1, etc., and n has been substituted for n-1, n-2, etc. It is the large population result.

Not only is this expression less cumbersome than the Frank/Granovetter expression, it also enables us to decompose the error into two components. The first component, $4s^2(a)/nN^2$, is the sampling variance that would result if each person's T_i were measured exactly, instead of estimated from a sample of possible tie others. Recall that $s^2(a)$ is the population variance of the T_i. T_i/N is the individuals P_i. So $s^2(a)/N^2$ is the population variance of the P_i.The sampling variance for an estimate of the mean P_i based on n observations is $s^2(a)/nN^2$. The '4' in the numerator results from the symmetry of ties. For symmetrical ties, the density estimate is $N\overline{T}/(N(N-1)/2)$ or, inlarge populations, $\overline{T}/(N/2)$. The '4' comes from squaring the denominator, $1/(N/2)$. For a sample of size n, the first component is the variance of the density estimate with error free measurement of the T_i for all members of the sample.

The second component, $2s^2(C)/n^2$, is the sum of the measurement errors for each individual's T_i. Each of the measurements, as was pointed out, is a Bernoulli. There are n respondents, each asked about n/2 ties. The variance of the estimated probability of a tie is $\overline{P}(1-\overline{P})$, for a given observation. With n(n/2) observations, the variance is $\overline{P}(1-\overline{P})/n(n/2)$ or $2s^2(C)/n^2$, where we recall that $\overline{P}(1-\overline{P})$ is $s^2(C)$. The only surprise is that in the aggregate we may replace the P_i that govern the actual observation with $\overline{P}$ which governs the average observation.

This decomposition of the sources of measurement error for the random subgraph procedure has important implications for evaluating the utility of the method. The first component is unavoidable with any method of estimating or measuring individuals' T_is. It represents a commonplace of sampling: that the variance of an estimator is an increasing function of the population variance and a decreasing function of the sample size. But the second component arises when we sum up the measurement error that occurs at the individual level. It is unique to the random subgraph procedure. With this procedure, the value(s) derived from a single interview are estimates themselves.

When we compare the random subgraph procedure to any other procedure, we may do so in terms of the error that occurs at the point of interview. Any method that gets better estimates of the individual T_is will, when averaged over the sample, result in lower variance density estimators. From our model of the random subgraph procedure, we have an analytic formulation of the individual level measurement variances; these variances are the variances associated with the coin flips or binomial experiments.[6]

It is now possible to translate these results into examples to see the practical implications. To do this, I need several assumptions. First, I will consider three imaginary populations: cities of 10,000, 100,000, and 7,000,000. Next, assumptions must be made about the average number of acquaintance ties. Following Granovetter's lead, I shall assume that 100 is a reasonable lower bound, and that 1,000 is a reasonable upper bound. Finally, I shall fix n at 500 and n/2 at 250. The latter number is suggested by Granovetter as a reasonable upper bound for the number of tie objects a respondent will tolerate being asked about.[7]

In Table 1, we list the results of interest for subgraph sampling. Across the top we have the analytic formulas, while each row corresponds to a different size of city. The parenthesized results are based on the assumption that average number of ties is 1,000.

Insert Table 1 about here.

It is evident from the third column that estimates of the individual level P_i are really quite good. But column four tells another story. Estimates of T_i are quite bad. This is because $T_i = NP_i$, so that the variance of $\hat{T}_i = N^2$ (var $[\hat{P}_i]$).

This result should not seem counterintuitive. In a large city, even those with many acquaintances are unlikely to know any randomly chosen other person. That means that P_i will be a very small number. Column five shows that the expected number of acquaintances in a reasonably long list is also quite small. Since the expected value of both numbers is very close to zero, the observed values will not wander very far away. Hence, both will have very small variances. But in the large city, the tie objects offered to a sampled individual are a very small subset of the tie objects available in the city. Accordingly, any observed tie stands for hundreds and even thousands of ties not observed.

We have already implicitly compared the random subgraph method to the direct questioning method. If the latter results in perfect measurement of T_i, then the second component of (1) does not apply. In that event, the direct questioning method is better. It is, in fact, much better. To show this we will relax the assumption of no error in measurement, while still assuming perfect veracity for subgraph responses, and see how much error can be permitted direct questioning before it becomes inferior. Three forms for the error, case i,ii, and iii in Table 2, will be examined.

Insert Table 2 about here.

Case i: On the assumption that there is no systematic bias in answers, the problem is quite straightforward. For the random subgraph procedure, Table 1 gives the individual level variances. Direct questioning is better whenever the error variance that is associated is less than the results in column four of Table 1.

It is easy to see that this will almost always be the case in large populations. Let us look closely at our imaginary city of 100,000. Taking the square root of the values in column four, we find a standard deviation of 200 for the case of $\overline{T}_i = 100$, and one of 630 for the case of $\overline{T}_i = 1{,}000$. In the first case, we would do better if most respondents' answers were within 200 of the true value. In the second case, most would need to be within 630 of the true value. (This is a conservative way of putting it. If most were more accurate than ± 630, a few could be way off. But these calculations are intended to be representative. More sophisticated assumptions about the distribution of error would allow for more accurate results.)

Some conclusions are now apparent. Under the assumptions of case i, the direct questioning method becomes relatively superior when the average number of contacts falls and when city size rises. The assumptions for this city of 100,000 were picked as representative of a turning point. For ties with less than 100 average contacts, by this size city, the direct questioning method is almost surely superior. And even for high frequency ties, this is approximately the largest city where an advantage for the random subgraph method can be imagined.

This comparison can be made somewhat more transparent with the help of Table 3. On this graph, T_i, the number of ties possessed by respondent i, is graphed against the standard deviation of the estimates implied by the random subgraph model. The three curves represent populations of sizes 10,000, 100,000, and 7,000,000. The dotted line delimits the points below which the standard deviation is more than half of the quantity to be estimated. This (somewhat arbitrary) rule of thumb suggests when the random subgraph procedure is probably best avoided.

Insert Table 3 about here.

Case ii: It is certainly possible that people will give systematically inaccurate estimates. So in this instance, it is assumed that every respondent systematically inflates or deflates his/her answer by a constant factor of K. (The results of this discussion would generally hold if K were a random variable, except that $E(e_i^2) = \overline{K}^2 + \sigma_K^2$ would be substituted for K^2. I shall avoid such complications.) Now, if K were less than the standard error of the individual estimates in Table 1, then even with systematic bias, specifically if $K^2 < \sigma^2$, we get better estimates with direct questioning. But the same does not hold for the aggregate estimates.

A systematic bias will accumulate with repeated observations, which means that the unbiased property that we have allowed the random subgraph method would become important. With multiple observations the "K"s do not average out. Therefore, the K^2 factor reappears in its entirety in the expression for the variance of the aggregate estimate. Unless the K^2 is less than the number in column four divided by n, then in the sense of expected squared error, one should get better aggregate estimates from the random subgraph method. Only in very large populations would this accumulated bias be outweighed by the measurement errors of the random subgraph method.

But this superiority can be exaggerated. Most often, the interpretation of these estimates would be comparative, and not absolute; that is, one wouldn't care what the actual density is, but one would care which of two cities, or subgroups, or which of two aggregates possesses the greater or lesser density. These tests would usually take the form of t-tests for two group comparison, or more complicated analogs for multiple comparisons. And these tests would be indifferent to a constant bias. If the bias is everywhere the same, the test results would be exactly the same.

But if we can discount constant biases in this fashion, then the direct questioning method is superior. (A random bias that was the same for different groups would reduce the power of our tests. For purposes of comparisons, a random bias is a special case of case i.)

Case iii: Case iii is an unspecific representation of the worst case. The bias could be correlated in a systematic way with the hypotheses under investigation. Here there may be no hope. Without intelligent procedures to minimize such possibilities or to detect them as they arise, one could confuse true differences with mere difference in response bias.[8]

The problem of systematic bias must always be taken seriously. It seems most reasonable not to consider it a problem inherent to either method but as a difficulty that may be more or less severe depending on the particulars of an application. As applicability is discussed, the problem areas for each method will become more clear.

The tentative result for applicability is that the method of direct questioning is superior. It results in estimators with less variance. It requires smaller samples for "good" estimates. And it requires fewer questions per respondent. All of these advantages are greater when the expected number of positive answers (nP_i in our notation) is small. The method of direct questioning is to be preferred most when the tie(s) at issue are rare and/or when the population is large.

The method of sub-graph samples should be superior under complementary conditions. When the population is small and the number of ties per person is large (and nP_i is therefore large), the random sub-graph method will produce reasonable estimates at a reasonable cost. It has the considerable advantage of preserving the information about the actual identity[9] and full range of attributes of the tied others that it uncovers. And if

budgetary constraints allow, a researcher could questions both ends of a claimed tie. This could provide valuable control over error.

The most succinct comparison of applicability can be made metaphorically. The method of direct questioning is like a Swiss Army knife. It is nearly always somewhat inadequate, but it is nearly always useful. The random subgraph procedure is like a delicate surgical saw. It is an exremely fine tool for a limited set of circumstances but very impractical much of the time.

The most favorable circumstances for the use of the random subgraph procedure occur when nP_i is large. Most generally, that will be true for small populations. Of course, such small populations might be located in large populations but only when the intial focus of the investigation was a single, readily identifiable subset of the larger population -- for example if one set out to investigate the social relations of police, or blacks in a small city, or leaders of major community organisations.

Lots of money, and special kinds of a priori information would allow the use of random subgraph procedures on entire large populations. With enough money, sample size can be increased to the point where a reliable density estimate is obtained. (Granovetter,1976) But unless list length can be increased, the individual level estimates in large populations will become quite unreliable. Another use of massive funding would be the collection of a priori information, for example on the ethnicity and status respondents, and the preparation of multiple lists suited for each respondent. An Irish workingclass respondent might be shown a suitably drawn sub-sample of fairly similar others such that nP_i would be high. But expense aside, such sub-sampling would only provide decent estimates of sub-group densities and sub-group embeddedness of individuals. Any attempt to investigate ties to the unclassified "other" outside of the prespecified sub-group would founder on the miniscule size of nP_i in a large population. Presuppositions about the 'appropriate' list to show a respondent would dominate the results.

Direct questions, in contrast, allow for a search strategy little constrained by prior assumption. One can allow respondents to identify the major targets of their ties flexibly and then refine the information about these targets. The design need not constrain tie others to any presupposed pool. Ties to other locales and to other groups are fully admissable.

But the advantage of a surgical saw over a jack-knife is that the more refined tool gives more accurate results when correctly applied. While the cruder method of direct questions can be used in a wider variety of circumstances it's vulnerability to bias might lead to a butchered operationalisation. Actually, both methods are vulnerable, in particular ways,to response bias. I will take up the vulnerabilities of the subgrap method first.

With the method of random subgraphs, respondents are asked about names. They could bias results by failing to acknowledge ties or by claiming false ones. The failure to acknowledge a tie could reasonably be taken as the absence of the tie (unless we are studying ties where respondents had an interest in concealment, such as the ties that unite a black market). The likelihood that people will claim ties they do not possess seems more the threat. It seems probable that this is more likely to occur when the number of truly claimable ties declines. That is, a respondent, faced with a list of 250 names of which only two were actual acquaintances, might try to "help the interviewer out" by claiming additional, false, ties. The likelihood of such bias should increase as nP_i declines. Furthermore, the impact of false claims would increase as nP_i declined, since any absolute number of false claims would be a high proportion of total claims. This implies the usual conclusion: that the method of random sub-graphs is ill-advised when nP_i is small.

The method of direct questioning presents more subtle problems. Certainly some questions asked of some people will produce wildly erroneous results. We can minimize that to some degree by thinking through the questionnaire design. Questions that seem unlikely to produce accurate answers can be eliminated or interpreted with extreme caution. But there is no general answer to the problem. So rather than arguing in a complete vacuum, I will suggest strategies that seem potentially valuable. For the issue here is not only the accuracy of the results, but also the substance. It cannot be shown that the method of direct questioning produces universally valid data, but it can be demonstrated that it is well suited for the address of issues of considerable interest.

These issues take us back to our initial concerns. These concerns are the extent to which attributes (or the categories they define) structure and constrain the social contacts of the population. We are most interested in the groups to which people belong and in the groups to which they are closely tied, even though they may lack the nominal attributes that are typically employed as a measure of membership. We are less concerned with the distribution of people's ties among the many categories to which they do not enjoy a large number of ties.

These concerns coincide neatly with common-sense expectations about the accuracy of answers. The most accurate information will be obtained from direct questions when the answers are obvious to respondents. The most obvious answers would cover the most significant ties and the most salient groups. Generally, the most significant ties are both more intense and less common. Most people have fewer close friends than friends and fewer friends than acquaintances. Accordingly, people have more information about the objects of their more intense ties. Since there are fewer objects, there is less total information to remember and summarize. In a similar way, people are most aware of memberships in groups that are socially significant to them. Sociologists are probably most aware of the professional affiliations of tied others who are sociologists. Methodists are probably most aware of the denominational affiliations of other Methodists. In summary, direct questions should work best when applied to the section of the surrounding social world best known and most important to the respondent.

I can suggest two complementary speculations. Answers will be more inaccurate when the ties at issue have many others and when the groups are less salient. These considerations combine. It would require a great deal of recollection to determine the attributes of every person to whom one enjoys a common tie. This is most surely the case for attributes that are not seen as important. A Polish-surnamed American urbanite would probably be hardpressed to give an accurate count of the number of Irish-surnamed Protestants in his or her acquaintance net. In many instances, our putative respondent would not even be aware of the nominal memberships of such acquaintances. An important source of potential error is the failure of knowledge (or of recollection) when respondents attempt to identify attributes of barely known people.

This means that some information will not be obtainable from respondents because they simply do not possess it. This occurs especially with high-frequency ties. It has been shown that one can compare collectivities with respect to tie densities even when the individual answers are rather inaccurate. But too much is demanded by insistence that people subclassify their guesstimates into categories about which they actually have little information. A reasonable and honest respondent might be able to recount the attributes of acquaintances who possess highly salient attributes but would be forced to throw most of their acquaintances into an "other" category. So more faith should be placed in reported frequent contacts than in reported absences of contact. This means that one cannot hope to measure the complete attribute to attribute density matrix of, say, acquaintance from each to every other attribute. There are better prospects of learning the gross level of contact undifferentiated into subgroups

and of learning of the large bundles of weak ties and smaller bundles of intense ties that attach respondents to the categories most significant in their social world.The limitation must be accepted that respondents contact nets may include a large amount of undifferentiated "other" that stands for the absence of known contacts to certain categories.[10]

This has important implications for research strategies. One can search through respondents' social worlds for the categories to which they enjoy the greatest number of weak attachments. Particular attention should be paid to the attachments to categories to which the respondent belongs. One could probe further to see if the stronger and less common attachments follow these patterns. At some point in this narrowing investigation of a social net,one could begin to assign credence to the numerical values uncovered.[11]

The blind spot of these methods is the blind spots of the respondents. With direct questions, one commits oneself to the investigation of the social world as respondents see it. The cautious investigator may be forced to forgo certain objectivist interpretations of the data. But this should not be exaggerated. The underlying reality is objective. One should expect that as we get closer and closer to the world people know well, their guesstimates will more and more accurately reflect their social worlds. In particular, their close ties and important categories will be uncovered with some considerable accuracy. And this will allow the harvest of some of the promise of relational data, an investigation of group structure and its impact on the individuals embedded in those groups. The next section will present and elaborate some of the issues that could be addressed by such an investigation.

Relational Configurations, Group structure, and Group Impact

Methodological limits on the investigation of the categorical compositions of social networks have been discussed. But those limits, as presented here, do allow the study of the categories most close to a respondent as well as the complete compostitions of ties for intense and/or uncommon ties by the cheap and flexible method of direct questions. Because it was necesary to first establish the limitations on the methodology, it was not possible in the first section to discuss the interpretatons of the data that might be collected. Those interpretations are the subject of this section. It shall be shown that such data can be used to understand the relation of individuals to categories as well as the group-like features of categories in the aggregate. In short, the full motivation of relational survey procedures can finally be presented.

The individual level data would consist of the quantity of social ties to social categories. Usually these are categories thought to be groups. ~~We think that~~ this quantity of ties to a category can be usefully thought of as a measure of social constraint. It is a measure of the extent to which the contents and contexts of the individual's social life are typified by the qualities (lifestyles, values, norms, practices) identified with the group. To motivate this interpretation, I'll turn to some extremely basic sociological concerns.

A social tie represents a continuing pattern of interaction between individuals. Such interactions will be governed by obligations and expectations. Thus a social tie provides both the channel and the rules of social conduct. A social tie is the arena in which social influence takes place (Katz and Lazarsfeld,1955), beliefs and information are transmitted Coleman,Katz, and Menzel,1966'), and where values are shared and supported (Simmel,1955). From the standpoint of either party to the tie, the various interactions leave some residue. Focusing on the categorical quantity of ties, suggests that this residue will tend to amount to whatever makes a particular category (or group) distinct from other categories (or groups). The claim is that the quantity and intensity of the individual's social ties to a category (or group) is both a cause and an indicator of the impact of that category/group on the individual.

More content can be added to this vague but general principle with some notions from Stinchcombe (1976). He suggested that solidarity arises from the coextensiveness of social group and the facilities necessary to solving problems of the members of the group. Social ties are the channels along which such facilities may be sought and exchanged. This suggests that the degree of network binding of the individual to the group indicates the degree of access to problem-solving facilities contained within the group. Contrariwise, the lack of such binding may indicate an indifference to the group's facilities or access to equal or better facilities elsewhere. Therefore, the more embedded the indivfdual is within the group, the more valuable the group is likely to seem (and to be) and the higher is the individual's subjective sense of solidarity with that group.

Fireman and Gamson (1977) have suggested, in similar terms, that such solidarity may be a major determinant of the likelihood that the members of a group will mobilize for collective action. They argue, following Olson (1965), that it is often not economically rational for individuals to contribute voluntarily to efforts to obtain collective goods. But they argue that social ties and the subsequent solidary identification of the individual with the collective interests of the group can provide an alternative rationale for individual contributions. The contentious capacity of the group is then partially dependent on the typical amount of social attachment of the group's members. Contentious capacity is, of course, a group property. In this account it rests, in part, on the average connectedness of the group's members.

This is a powerful line of argument that may be generalized. Whatever the connectedness of a particular individual, the average connectedness of all members is a measure of the social cohesion the group enjoys. More cohesive groups should enjoy more intense, and more distinctive group lives and be more differentiated from other groups. Therefore, the effect of any given level of ties is affected by the average level of ties enjoyed by others. Whatever the sphere of impact examined, whether on political activity, lifestyle, normative adherence, or whatever, that impact should increase both with the average number of ties and with the number of ties enjoyed by a particular individual. A few ties to a highly cohesive group may have greater clear impact than many ties to a "group" that is not cohesive.

This interplay between the network properties of the group and its impact on the individual offers the most exciting possibility for survey relational data. It offers an opportunity to advance well beyond the nominal classification of respondents into social categories to a unified attempt to investigate the impact of the classifications generally on attitudes, behaviors and so forth. Suppose several respondents were found to be quite highly connected to members of labor unions. The expectation is that they would be strong backers of labor union political goals and so forth. But also discovering that the typical union member was not highly connected to labor union members, would change the expectation. Even the best connected people should then show less impact of their connection, and the impact of nominal membership should be rather slight.

It is because of this possibility of investigating the properties of groups that I insisted on the use of consensual categories that most respondents are aware of and use more or less as we do. Such categories can be investigated for the extent to which they delimit social networks, that is, the extent to the boundaries of the category bound the social interactions of the members. In an important sense, this would measure how well these categories function as groups.

Conclusion

I have focused on the use of direct questions to study the immediate social surroundings of individuals which seems the most secure application. But for close ties and/or highly salient groups, one could also use these methods to study the converse problem, the degree of social segregation that typifies the relations between different social categories. Perhaps the most unsatisfactory feature of this essay is the inability to succinctly suggest all the possibilities inherent in this method. We have stayed close to the original issue in hopes of making one point well rather than many points poorly. As a technique for investigating network concerns in large populations, it is addressed implicitly to a larger body of concerns than could actually be presented.

The entire approach is based on a single simplification of traditional network approaches. It is difficult and extremely costly to trace out the chains that networkers think of in large populations. It is even more difficult to relate such results to the generalizations derived from more traditional survey methodology. The sinlge abstraction to the sheer quantity of ties offers hope of illuminating network structures as well as refining the use of demographic categories in both theory and research. Certainly we should not forget our ignorance of what lies beyond the first step outward from the respondent. But the investigation of those immediate surroundings is compatible with our current research technology and offers considerable promise.

If the theoretical questions now being raised about the macro structure of social contact become central to our discipline, we shall want to be able to investigate them. The present methods are obvious tools for that task, but they are unfamiliar tools with unfamiliar difficulties. They are available, and they are workable, but they should be applied with care and caution.

Appendix

Approximating and Partitioning the Variance of the Density Estimator

At issue is the variance of the sum of the individual level extimators. For each individual, the true state of the world is described by

(1) $P_i = \frac{T_i}{N}$

where N is the size of the population and T_i is the number of ties enjoyed by person i. A sample of size n is drawn. With one-way questioning, each sample member is asked about $\frac{(n-1)}{2}$ others, and with two-way questioning, about (n - 1) others. To avoid uninteresting complications, I shall assume that n is an odd number so that $\frac{(n-1)}{2}$ is an integer.

I shall first consider two-way questioning. Each individual estimate is then governed by a binomial

(2) $\hat{P}_i \sim B(n-1, P_i)$

so that $E(\hat{P}_i) = P_i$, and $V(\hat{P}_i) = \frac{P_i Q_i}{(n-1)}$. The ultimate issue is obtaining aggregate estimates so I will examine the sum of the individual level estimates, $\sum_{i-1}^{n} \hat{P}_i$, seeking to obtain the variance of this sum.

first that the estimator is unbiased,as Frank and Granovetter have shown, it follows that

(3) $V(P_i) = E[\sum_{i-1}^{n}(P_i - \overline{P})]^2$

where $\overline{P} = \sum_{i-1}^{n} \frac{P_i}{N}$.

This may be written as

(4) $E(\sum_{i-1}^{n} \hat{P}_i - \overline{P})^2 + \sum\sum_{i \neq j}(\hat{P}_i - \overline{P})(\hat{P}_j - \overline{P})$.

The covariance (second) term may be reduced to

$$E\sum_{i\neq j}\sum(\hat{P}_i\hat{P}_j - \overline{PP})$$

which may be written

$$E\sum\hat{P}_i[(n-1)\hat{\overline{P}}_{-i} - \overline{PP}]$$

where $(n-1)\hat{\overline{P}}_{-i}$ is the sum of the $\hat{P}$s excluding $\hat{P}_i$. Each term of this sum is the covariance of an individual's estimate $\hat{P}_i$ with the sum of all the other individual estimates. It is intuitively obvious that all of these estimates are related, since all of the $\hat{P}_i$s are estimated with a single list. This sum can be evaluated with the methods and notation of Frank, but this would require the introduction of extensive notational and definitional apparatus. What shall be seen, soon enough, is that the sum is negligible in large populations.

The first term in (4) may be rewritten

$$(5)\quad E\sum_{i=1}^{n}(\hat{P}_i - P_i + P_i - \overline{P})^2 = E\sum_{i=1}^{n}(\hat{P}_i - P_i)^2 + (P_i - \overline{P})^2 + 2(\hat{P}_i - P_i)(P_i - \overline{P})$$

The expectation of the third term on the right is zero, so the formula simplifies to

$$(6)\quad E\sum_{i=1}^{n}[(\hat{P}_i - P_i)^2 + (P_i - \overline{P})^2]$$

Taking expectations, the latter term amounts to the population variance of the P_i; call this σ_p^2, which occurs n times. The former term is the sum of the expected variances of the $\hat{P}_i$. (6) then equals

$$(7)\quad n\sigma_p^2 + \frac{E\sum_{i=1}^{n}P_i(1 - P_i)}{(n-1)}$$

The second term can be written as $\frac{E\ P_i}{(n-1)} - \frac{E\ P_i^2}{(n-1)}$. Taking expectations and summing, the second term equals

$$(8)\quad \frac{n}{(n-1)}[\overline{P}(1-\overline{P}) - \sigma_p^2]$$

Replacing $\frac{n}{n-1}$ with 1, we see that (6) is approximately equal to

$$(9)\quad (n-1) \times \sigma_p^2 + \overline{P}(1-\overline{P})$$

To make this comparable to the Frank result, note that $s^2(a) = N^2\sigma_p^2$ and that $P(1-\overline{P})$ is $s^2(C)$. To make this comparable to Granovetter, it must also be noted that the density estimate is $\hat{\overline{P}} = \frac{\hat{P}_i}{n}$, so that the variance in

(9) should be divided by n^2. This gives the approximate result for two-way questioning.

$$(10)\quad V(\overline{P}_i) \approx [\frac{(n-1)}{n^2N^2}]\ s^2(a) + \frac{s^2(C)}{n^2}.$$

With one-way questioning we have the same number of respondents, but each is only asked about $\frac{(n-1)}{2}$ potential others. The denominator of the density estimator is now $\frac{(n-1)}{2}$. Furthermore, the n terms that contributed to the $s^2(C)$ term now have a denominator of $\frac{(n-1)}{2}$ instead of n-1. Approximating n-1 by n, it can be seen that

$$(11)\quad V(\hat{P}) \approx \frac{4s^2(a)}{nN^2} + \frac{2s^2(C)}{n^2}$$

The result that Granovetter derived from Frank was

$$(12)\quad V(\hat{\overline{P}}) = \frac{4(N-n)(n-2)s^2(a)}{n(n-1)(N-1)(N-2)(N-3)} + \frac{2(N-n)(N-n-1)s^2(C)}{n(n-1)\ (N-2)(N-3)}$$

It is evident that our approximate result is simply the Frank/Granovetter formula, simplified by the substitution of n for n-1 and n-2 and by the substitution of N for N-1, N-2, N-3, N-n, and N-n-1. Almost all of the difference between the approximate and the exact answer was the covariance term that was neglected earlier.

It is not very informative to examine an exact expression for the difference between the approximate formula and the exact formula. But inspection discloses when the two will be different. The most dubious substitution is that of N for N-n and N-n-1. When N, the population size, is small, then the sample size, n, may be a substantial fraction. In such instances, the formula (12) gives a smaller value for the variance of the estimator. For example, if n = 100 and N = 1,000, then the first term in (12) is 82.1 percent of the first term in (11). Under the same assumptions, the second term in (12) is 89.6 percent of the corresponding approximate term. The percentages are approximated by $100*[\frac{1-n}{N}]$ and $100*[\frac{1-n}{N}]^2$. Tables IV and V show the true values, the approximate values, and the percentage discrepancy for assorted values of N and n. the approximation is reasonably accurate once N and n are reasonably large.

Insert Table 4 and Table 5 about here

The approximation is slightly more convenient than the exact formula, but that savings is slight. The derivation of the approximation is more informative. In effect, we assumed that each individual level observation is independent of every other. This permitted a partition of the variance into sampling error of the P_i around the population value $\overline{P}$, and error in estimating the individual level P_i. We saw that neglecting the covariance of the $\hat{P}_i$ from a single sample made little practical difference once N and n are reasonably large.

Footnotes

[1] The variable degree of membership in a categorical group is indexed by the vector entry that refers to the group. At first thought, the most important entry is the largest. It might seem to reflect the most important group in respondent social world and would appear a natural choice for the group to which respondent most belongs. But complications intrude.

In a world of random contact, the most probable contacts are with the larger groups. In a random world, the expected number of ties is proportional to group size. A member of a small nominal group, such as sociologists, would be expected to have few of their contacts with other sociologists. Even if there were 20,000 sociologists in the population of the U.S., the random chance that a tie would be to a sociologist would be only one in 10,000. So even if I had 1,000 friends, in a random world I would expect to be friends with only .1 sociologist.

This is not a trivial problem, but it illustrates two points. The first is the extreme improbability under randomness that anyone knows more than a small number of similar but rare others. That I know many more than .1 sociologist indicates that my contact net strongly reflects that occupational attribute as does the social structure of my existence. One can expect the same to occur for other attributes. But the second point is that quantity must be compared to possibility. It is not the most frequent ties that count, but the most improbable. On a random model, probability is inversely proportional to category size. Weighting the vector of categorical compositions by category size is equivalent to dividing each entry by the size of the category to which it refers. The measure that results is the individual analog of density. It is the proportion of all possible links to the category's members that actualise in the individual's net.

It should be stressed that this adjustment is not the only possible adjustment, nor is it universally desirable. Other strategies for this problem exist in the literature. (cf . Fararo and Sunshine [1964]). It is suggested here both to make the issue apparent and because this adjustment makes the individual level observations consistent with density. When the individual level measures are so adjusted, density is the simple average of the individual measures.

[2] The calculation of densities only scratches the surface of potential applications of measures of ties within and between members of social categories. The most imaginative applications are well beyond the scope of the current essay. I pause briefly only to indicate the extreme flexibility of the concept of categorical composition to give more moment to the central concern of this essay, the measurement of such compositions.

One source of flexibility is the high abstraction of the concept of social tie. Many narrower, more focused concepts may be thought of in terms of social ties. In the vector terminology,one can imagine a collection of vectors, each corresponding to a particular type of tie. (Actually, each individual would have such a collection, so that a population would have a collection of collections.) There could be a vector of acquaintances, a vector of friends and a vector of significant others. Exchanges of political information, consumer advice or social disease could be so represented. The possibilities are almost endless.

Another source of flexibility is the varied notion of category. The social world can be meaningfully divided into a wide variety of category sets, mutually exclusive and jointly exhaustive classification of the population into disjoint sets. Race is such a category set (although there is a small remainder of other, unknown, infrequent or ambiguous that might trouble a logician if not a practical researcher). Sex is a category set. Race and sex crossed, yielding white-male, white-female, black-male, etc., is still another category set. The variety is almost endless. Issues such as overlap and interaction of category sets, the comparative impact of different category sets on separating social ties, and, most complicated of all, the interrelations of different category sets and types of ties could have important sociological interpretations.

A great variety of conventional sociological problems, as well as many problems not yet perceived, could be conceptualized and researched in these terms.This abstract reciatation is a distance from workable research strategies. But what I have tried to indicate is the enormous scope of such possibilities. There is a promise for the future which this essay only foreshadows.

[3] The categories need not be salient to all respondents. For example, One could inquire about categories to which the respondent belonged which might not be well known to all other members of the population. Thus, I could readily tell you how many of my acquaintances are sociologists, but I also know many people whose professional affiliation is a mystery to me. I could not give an accurate set of answers about the number of economists I know, but only about the number of people I know to be economists.

Under such conditions,one could not construct a complete density matrix showing the division of acquaintances among all of the categories of others. One could, however, investigate the issue of ties to self-category versus combined ties to all other categories. This "degree of sociometric closure", or the extent to which members' social ties are within the group, has a group level analog that is comparable to density.

It should be added that only "objective" or consensually used categories are suitable for the formation of density estimates which require knowledge of group size. Although it might be fruitful to investigate cognitive social maps — for example, the number of acquaintances seen as middle-class or politically liberal — these answers would be almost wholly subjective and not usable for aggregation across members of the groups so identified. Just because someone claims middle-class status for someone else does not insure that all the other someones will agree. So one could investigate whether people who strongly self-identify with some subjective category perceive their friends as similar. But this would not tell us whether the friends shared the subjective identification. These drawbacks all lead to important qualifications to the generality of the direct questioning method. But I shall not discuss them, indeed shall assume them unproblematic until after the sampling models have been constructed.

[4] Technically, this applies only in large populations where n and T_i are much less than N. The true distribution of the number of observed ties with N, n, and T_i fixed is hypergeometric but under large population conditions we may safely use the binomial (cf. Feller(1968))

[5] Granovetter could do this with little loss, since his tie of interest, acquaintance was considered symmetric for theoretical reasons. For the more general result with the complication of two-way questioning, consult the appendix.

[6] Granovetter's main concern was the necessary sample size implied by various assumptions about $S^2(a)$ and $S^2(C)$. A similar analysis could be carried out with these equations. As Granovetter points out, and as Frank details on p. 72 , $s^2(C)$ is fixed by the average number of ties, but $s^2(a)$ could vary between zero and a very large number. I will not concern myself with assumptions about $s^2(a)$, because our decomposition shows that its effect is the same, no matter what survey method is used. It is worth mentioning that Granovetter, in his "typical" case, usually assumed that the $s^2(a)$ component was a fairly small fraction of the total. He was not aware of the partition into the two sources of what he referred to as "sampling error", but in his examples, what is here called measurement error was usually by far the greater component.

[7] There are two assumptions here: one is sample size, and the other is list length. Granovetter went to some trouble to show that where list length seemed too long, one could increase accuracy by drawing multiple lists and showing each sample member only a subset of the entire sample. This would be expensive, and the present analysis indicates that in any population large enough to require multiple samples, the expected frequency of observations is too low to warrant use of his method.

[8] There is no obvious and general solution to the problem of bias in this method or in any other. Many of the most trusted and beloved survey results could easily and sometimes quite plausibly be modeled as the result of systematic bias. But I think that the burden of proof is best left to the skeptic. It seems reasonable to assume that questions ask what they seem to ask until it can be shown otherwise. This is not intended

to suggest that anyone should ask questions that people cannot reasonably answer and then claim a false precision for their guesses. But in the absence of strong evidence to the contrary I prefer to believe that guesses produce random responses more often than they result in systematic bias.

[9] This seems a good moment to emphasize the limited nature of the criticism of Granovetter. His specific attention was directed to the case where typical volume is high: the case of acquaintance. Furthermore, he had basis in past research for assuming that people's conscious awareness of their "weak ties" was limited. Finally, he claimed no interest in the individual level interpretation of his proposed measures. Taken at its face, his presentation was neither incorrect nor misleading. Our present effort does not vitiate his results but instead generalizes them for problems of subgroups, large populations, and small radius ties.

[10] This might be called the paradox of pluralism. A group might not be identified by others simply because the group label has very low saliency and not because the group is socially isolated. This is one reason for suggesting caution in the interpretation of the reported absence of ties when the ties are frequent and the groups are of low saliency.

For the central concern of this essay, the relational interpretation of membership in demographic categories, this problem is not very important. The focus is only on those categories in which our respondent might claim membership, either by possession of an attribute or by large numbers of social ties. Categories to which the respondent enjoys few ties are not of interest. But this will not always be the case. At some point these methods

might be applied to less narrow concerns. As was mentioned above, recent theoretical presentations by Blau and Rytina and Morgan could inspire concern with the peripheral categories in people's social worlds in addition to a concern with the central categories.

The present methods are by no means unsuitable, but they require some thought. For example, it is a comparatively simple matter to distinguish social isolation from social invisibility. Remember that one could inquire at both ends of the potential relationship, that is, one could ask Group A about its ties to Group B and ask the reverse question of members of Group B. Now if it is assumed that the tie is symmetric, one would have two completely independent estimates of its frequency. So if Group A is invisible to Group B but not vice versa, a major discrepancy will arise when Group A claims more ties to B than B will acknowledge. It might be the case that both groups are equally invisible. No discrepancy would arise in this instance, but each group would be claiming large numbers of ties to the undifferentiated "Other". But one could still untangle the true situation.

If invisibility is truly operative, there should be comparatively little differentiation with respect to more intense, smaller radius, more accurately reported ties. Therefore both Group A and Group B would show up in each other's intimate circles. If no (or little) social distance was observed at the intimate level, one should feel secure in ascribing the reported infrequency of weak ties to invisibility and not to isolation.

This example shows several things. The first is that these methods probably have wider application than the central essay describes. But application must not ignore the difference between awareness of contact with members of other groups, and the existence of such contact without awareness. This is a substantive issue. For some categories, such as ancestors' national origin, awareness may be more the exception than the rule. When a category is highly salient, we may expect quite accurate answers about even tiny numbers of weak ties. When the category is not so salient, it will be invisible in the far reaches of people's contact nets.

11. At some point in this narrowing search procedure, the method of direct questions converges to a method of itemised enumeration.Laumann has used this method to good effect when his interview scedule called for the identifacation of specific network alters whose attributes and relationships,as known to the respondent, were then investigated. For rare ties, this method is more direct and natural as indicated by the awkwardness of the question "And how many of your spouses are members of category X?" In such instances of extremely infrequent ties, it is simpler,and more conversationally natural, to establish shared names or identifiers for members of the respondents social network and ask about each of these tie others in turn.

As applied by Laumann, this method differs from what I have discussed in it's concentration of interview resources on a detailed investigation of the structure of the most immeadiate friendship net. His approach also fixed the size of the net investigated for each respondent as a feature of the instrument design.

Table 1

Variance of individual P_i and T_i estimates and expected number of observed ties per respondent for random subgraph estimators with different population sizes and tie densities

Population Size	Average Ties Per Capita	Tie Density	Variance of Individual P_i Estimates*	Variance of Individual T_i Estimates*	Expected Number of Ties Observed per Respondent
N	$\overline{T}_i$	$\overline{P}_i=\overline{T}_i/N$	$V(\hat{P}_i) = \frac{P_i(1-P_i)}{(n-1)/2}$	$V(\hat{T}_i) = N^2[V(\hat{P}_i)]$	$E(t_i)=\frac{n-1}{2} P_i$
10,000	100	.01	$3.976X10^{-5}$	3,976	2.49
	1000	.1	$3.614X10^{-4}$	36,140	24.9
100,000	100	.001	$4.012X10^{-6}$	40,120	.249
	1000	.01	$3.976X10^{-5}$	397,600	2.49
7,000,000	100	$1.4X10^{-5}$	$5.743X10^{-9}$	281,405	.000356
	1000	$1.4X10^{-4}$	$5.743X10^{-8}$	2,814,050	.00356

* for respondents with the average number of ties

Table 2

Error assumptions discussed in the text

case i: $E(e_i)=0$ all i, $V(e_i)= \sigma^2$

case ii: $E(e_i)=k$ all i, $V(e_i)= \sigma^2$

case iii: $E(e_i T_i) \neq 0$

Table 3

Standard deviations of individual level random subgraph estimators for different values of total number of ties and population size with a subgraph list of length 250.

	Total Number of Ties									
	10	50	100	150	200	300	400	500	700	1000
Population Size										
10,000	20	45	63	77	89	109	126	141	167	189
100,000	63	141	200	245	282	346	400	447	529	629
1,000,000	200	447	632	774	894	1095	1265	1414	1673	1999
7,000,000	489	1095	1549	2049	2190	2683	3098	3464	4098	4647

Graphical representation of Table 3

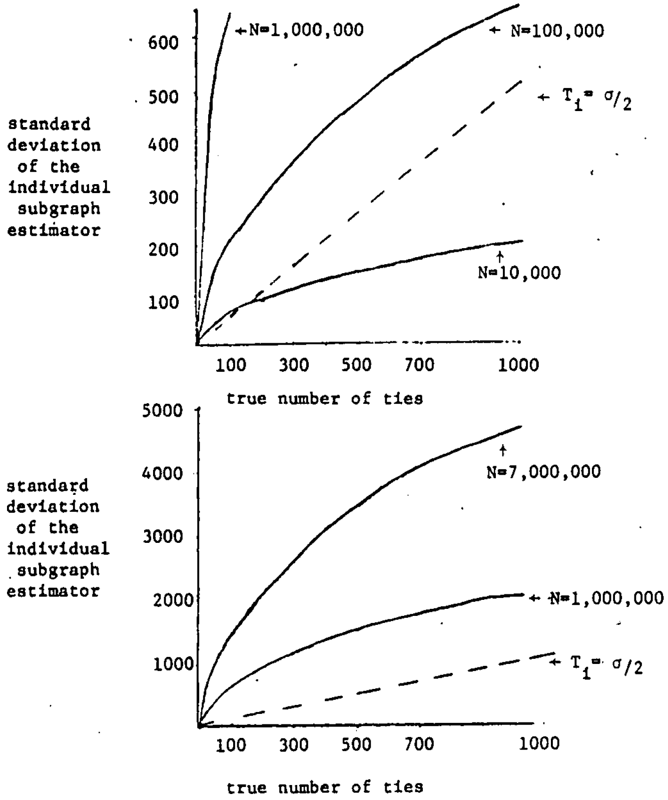

Table 4

The approximate variance associated with $s^2(C)$ as a percentage of the exact variance for different values of the population size and the sample size.

	Population size			
Sample Size	1,000	10,000	100,000	1,000,000
10	109.2%	111.0%	111.1%	111.1%
100	82.1%	99.0%	100.8%	101.0%
200	64.6%	96.6%	100.1%	100.5%
500	25.0%	90.4%	99.2%	100.1%

Table 5

The approximate variance associated with $s^2(a)$ as a percentage of the exact variance for different values of the population size and sample size.

	Population Size			
Sample Size	1,000	10,000	100,000	1,000,000
10	88.5%	88.8%	88.9%	88.9%
100	89.6%	98.0%	98.9%	99.0%
200	80.1%	97.6%	99.3%	99.5%
500	50.0%	94.8%	99.3%	99.8%

References

Blau, Peter M. 1977a. Inequality and Heterogeneity. New York: Free Press

____________. 1977b. "A Macrosociological Theory of Social Structure." American Journal of Sociology 83 (July) 26-54

Blau, Peter M. and Otis Dudley Duncan. 1967. The American Occupational Structure. New York: Wiley

Boorman, Scott A. and H.C. White. 1976. Social Structure from Multiple Networks. II. American Journal of Sociology 82 (May) 1384-1446

Coleman, James. 1968. "Relational Analysis: The Study of Social Organizations with Survey Methods. Human Organization 17 (Winter) 28-36

____________. 1961. The Adolescent Society. New York: Free Press

Coleman, James, Elihu Katz and Herbert Menzel. 1966. Medical Innovation. Indianapolis: Bobbs-Merrill

Duncan, Otis Dudley. 1961. "A Socioeconomic Index for all Occupations." in Occupations and Social Status. Albert J. Reiss(ed.) pp. 109-38. New York: Free Press

Feller, William. 1968. An Introduction to Probability Theory and It's Applications. vol. I.(3rd edition). New York: Wiley

Fireman, Bruce and William Gamson. [in press] "Utilitarian Logic in the Resource Mobilization Perspective." in J. McCorthy and M. Zald(ed.).

Fararo, T. J. and Morris Sunshine. 1964. A Study of a Biased Friendship Net. Syracuse, New York: Youth Development Center, Syracuse University

Frank, Ove. 1971. Statistical Inference in Graphs. Stockholm: Forsvarets Forskning-Ansalt.

Goldberg, Steven. 1973. The Inevitability of Patriarchy. New York: William Morrow

Granovetter, Mark. 1973. "The Strength of Weak Ties." American Journal of Sociology 78 (May) 1360-1380

__________. 1974. Getting a Job: A Study of Contacts and Careers. Cambridge, Massachusetts: Harvard University Press

__________. 1976. "Network Sampling: Some First Steps." American Journal of Sociology 81 (May) 1287-1303

__________. 1977. "Reply to Morgan and Rytina." American Journal of Sociology 83 (November) 729-31

Harary, Frank, R.Z. Norman and Dorwin Cartwright. 1965. Structural Models: An Introduction to the Theory of Directed Graphs. New York: Wiley

Katz, Elihu and Paul F. Lazarsfeld. Personal Influence. Glencoe, Illinois: Free Press

Laumann, Edward O. 1973. Bonds of Pluralism: The Form and Substance of Urban Social Networks. New York: Wiley

__________ and Franz Pappi. 1976. Networks of Collective Action. New York: Academic Press

Lorrain, Francoise and Harrison C. White. 1971. "Structural Equivalence of Individuals in Social Networks." Journal of Mathematical Sociology 1 (January) 49-80

Milgram, Stanley. 1967. "The Small World Problem." Psychology Today 1 (May) 62-67

Morgan, David and S. Rytina. "Comment on 'Network Sampling: Some First Steps.'" American Journal of Sociology 83 (November) 728-9

Niemiejer, R. 1973. "Some Applications of the Motion of Density to Network Analysis." The Hague: Mouton

Rytina, Steve and D. Morgan. 1977. "Social Networks in Large Populations." Mimeographed: Center for Research on Social Organization, The University of Michigan

Simmel, George. 1955. Conflict and the Web of Group Affiliations. Kurt Wolf and Reinhard Bendix(translators). New York: Free Press

White, H. C., S.A. Boorman and R.L. Breiger. 1976. "Social Structure from Multiple Networks, I. "American Journal of Sociology 81 (January) 730-80