Social Networks 4 (1982) 285-304
North-Holland Publishing Company

PREDICTING THE STRUCTURE OF A COMMUNICATIONS NETWORK FROM RECALLED DATA

A. Kimball ROMNEY and Katherine FAUST
University of California, Irvine *

In a series of papers on informant accuracy in social network data, Bernard, Killworth, and more recently, Sailer, have concluded that "what people say, despite their presumed good intentions, bears no useful resemblance to their behavior" (Bernard, Killworth, and Sailer 1982: 63). In this paper we reanalyze one of the data sets (the technical group) utilized by Bernard, Killworth and Sailer in arriving at their conclusions. Unlike Bernard et al. we find that the observed behavior data corresponds closely to the recalled data. Using different methods of analysis we find that the verbal recall data can be used to predict structural aspects of the observed data. Two major findings emerge from our analysis: first, the more similarly two people judge the communication pattern of others, the more they interact with each other, and, second, the more two people share accurate knowledge of others, the more they interact with each other. Implications of our findings for the assertions of Bernard, Killworth and Sailer are discussed.

1. Introduction

In a series of papers on informant accuracy in social network data, Bernard and Killworth (Killworth and Bernard 1976, 1979; Bernard and Killworth 1977) and more recently, Sailer (Bernard, Killworth and Sailer 1980, 1982), have concluded that "what people say, despite their presumed good intentions, bears no useful resemblance to their behavior" (Bernard, Killworth and Sailer 1982: 63). They have analyzed data sets covering an impressive array of social settings and utilizing a variety of sophisticated statistical methods. The data sets include communication among the following groups: (1) 94 deaf people on teletypes, (2) 44 amateur radio operators ("hams"), (3) 40 office persons, (4) 34 Technology Education Program people, (5) 58 fraternity brothers, and (6) 57 EIES (Electronic Information Exchange System) participants. In another related article, Burt and Bittner (1981) reanalyzed

* School of Social Sciences, University of California, Irvine, CA 92717, U.S.A.

the "hams" data and concluded that "the conclusions expressed by Bernard et al. (1980) regarding network subgroups are unwarranted" (Burt and Bittner 1981: 86). In response to Burt and Bittner, Bernard et al. said "the method by which data are traditionally obtained (some variant of "who do you interact with") does not yield any reliable information whatsoever about who people in the group actually interact with" (Bernard, Killworth and Sailer 1981: 89).

In this paper we propose to reanalyze one of the data sets used by Bernard, Killworth and Sailer, namely the "tech" data (Bernard, Killworth and Sailer 1980: 216–217). We were curious to see if we could find convincing and interpretable structural relationships between the observed behavioral data and the recalled cognitive ranking data. We knew that Bernard, Killworth and Sailer had analyzed the data using clique-finding, blockmodeling, and factor-analytic techniques and reported results as follows:

> "After defining a way to compare clique structures between behavioral and cognitive data, we found that there was no useful relationship between the two, and furthermore there was no significant difference in performance between any of the structure-finding algorithms." (Bernard, Killworth and Sailer 1980: 191)

Following our analysis of the data, we will discuss some possible implications for the conclusions of Bernard, Killworth and Sailer.

2. The data

> 2) Tech. These data are from a graduate program in technology education at West Virginia University. The program's faculty, graduate students, and secretaries are located in three buildings – two converted houses at the bottom of a hill, and a suite of offices on the hill in the main education building at the university. There are 37 people in the program; three of these are on full-time field assignment over 100 miles from the university.
>
> For one week a team of observers walked through the office spaces of the Tech program. They covered the same ground everly half hour, and noted all occurrences of persons in verbal contact. Any two persons in contact were scored. N-tuples were scored by counting each dyad. The same comments on obtrusiveness apply as for the Office data.

After a week of observation, each of the 34 persons on the main campus was handed a deck of cards containing the names of all other members of the group, and asked to rank the deck from most to least communication that week. The question was purposely left rather vague; amount, frequency, or importance of communication was not specified." (Bernard, Killworth and Sailer 1980: 194–195)

3. The analysis

In our analysis we will proceed as if we had collected the data and were testing the propositions about interaction mentioned in the abstract. An outline of the methods utilized will provide an overview of the analysis to follow. We first asked whether there was sufficient similarity between the choice data and the observed data to warrant further detailed analysis. The first section below demonstrates adequate significant similarity to continue. In the second section we treat the choice data as preference type data and ask how well they predict the observed behavior data. This approach produces a correlation of 0.74 between the two aggregated data sets. In the following section we eliminate subjects whose observed interactions were below a critical threshold and performed a proportional iterative normalization of the observed data of the reduced sample. The normalization allows us to study the structure of interaction separately from individual variability in amount of interaction. We then look at individual accuracy and find that there is a modest though consistent significant ability of individuals to rank behavior. Finally and most important we turn to the prediction of the behavior structure from the choice structure. Here we find that the subgrouping produced by the choice data almost exactly predicts the subgrouping of the behavior interaction structure (obtained only with the aid of normalization). From this correspondence of structures we are able to test the relevant propositions about interaction. To the extent that these tests are successful Bernard *et al.* are overly pessimistic about the use of recall data.

3.1. Overall similarity

The first question is whether the choice data and the observed data have any overall similarity, since if they do not, then it would be hopeless to mine the data for specific structural similarities. [1] The appropriate test to compare the two matrices is the Quadratic Assignment approach as described in Hubert and Schultz (1976) and more specifically as it applies to sociometric data in Hubert and Baker (1978). The approach is basically a non-parametric method that allows (among many other things) a comparison of two sociometric matrices. "A permutation distribution and an associated significance test are developed for the specific hypothesis of 'no conformity' reinterpreted as a random matching of the rows and (simultaneously) the columns of one sociometric matrix to the rows and columns of a second." (Hubert and Baker 1978: 31).

We applied the Quadratic Assignment Program to the two tech matrices, namely, the rank order recalled data and the observed behavior data. An approximate test based on the first two moments of the permutation distribution gave a Z of 11.2. We interpret this to mean that the two matrices are similar in some important respects. (Z scores for the other groups reported in Bernard, Killworth and Sailer 1980, are: Office, 10.1; Hams, 9.6; and Frat, 9.8). Bernard, Killworth and Sailer use the approach in the analysis of the EIES data and interpret their high Z scores by saying, "So the behavioral and recall matrices possess similar signals" (Bernard, Killworth and Sailer 1982; 60). The high degree of overall similarity encouraged us to continue the analysis.

3.2. Overall prediction of interactions

The marginal data (which summarizes each person's interaction) in the behavior matrix may be characterized in terms of either the number of people talked to or total interaction tallies. We can pose the question as to how well the amount of interaction, as indexed by one or other of the indicators mentioned, can be predicted by the scaled choice data. In order to answer this question we need an appropriate way to scale the rank order data to obtain a single combined ranking. We can do this by

1 In our understanding the behavior data should be symmetrical. However inspection of the tech behavior data (Bernard, Killworth and Sailer 1980: 215–217) revealed that there were some discrepancies. To correct this we substituted ones for zeros where there were differences to make the data symmetrical. However, we did not realize until our analyses were completed that the recall data had numbers on the diagonals where there should have been zeros. We re-ran a sample of our analyses on the corrected recall data and found that it did not make a noticeable difference in the results.

treating the choice data as logically equivalent to preference data.

One appropriate scaling technique for rank order preference data is Coombs' unfolding method (Coombs 1964; Chang and Carroll 1968a) as implemented in the program MDPREF (Carroll 1972). This program produces a "geometric configuration of stimulus points and subject vectors ... such that the projections of stimuli on each subject's vector corresponds optimally with the order of preference expressed by the subject" (Chang and Carroll 1968a: i). One major dimension of the ranking data on communication should correspond to the actually observed interaction frequencies. Figure 1 shows the results of fitting the number of people talked to as a vector to the stimulus configuration from MDPREF. In Fig. 1 each individual is identified by a number representing the number of people talked to. It can be seen that low interaction occurs in the upper left quadrant while larger numbers occur in the lower right quadrant. indicating that these individuals

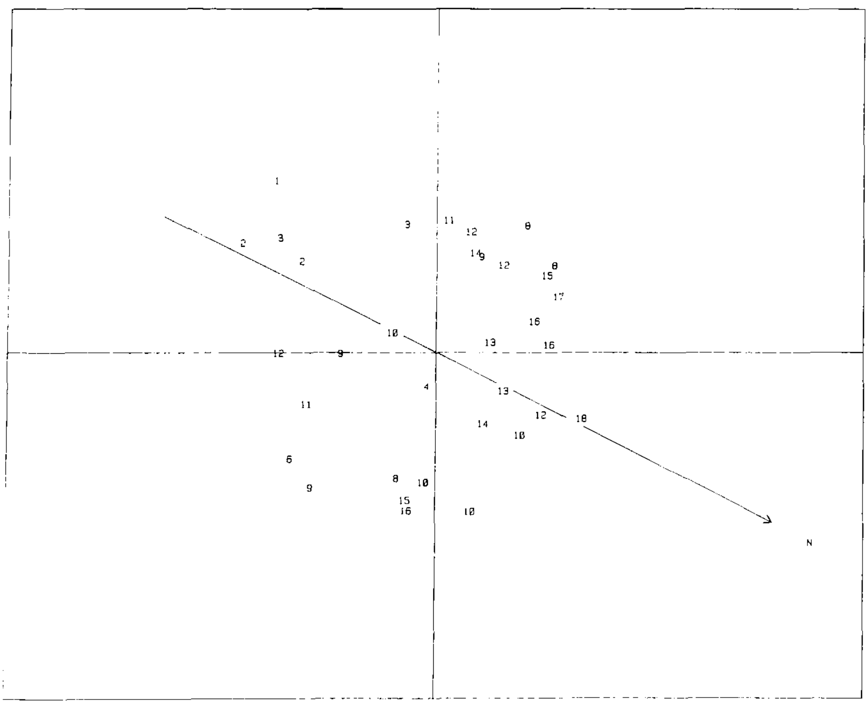

Figure 1

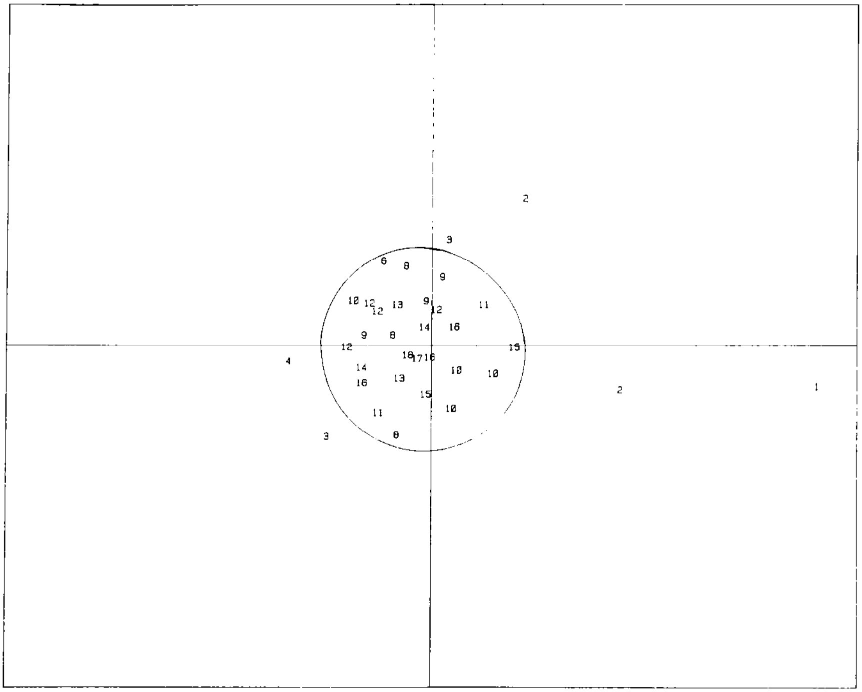

Figure 2

talked to more people. The line through the figure was fitted with a program called PROFIT (Chang and Carroll 1968b). The linear correlation between the vector represented by the line and the projections of the 34 individuals on the vector is 0.74. This may be interpreted as saying that the overall correlation between one dimension (represented by the vector) of the structure of the choice data and one aspect of the observed interaction data, namely the number of people talked to, is 0.74. We also performed the same analysis on total interaction tallies and the correlation is 0.66. The 0.95 confidence limits on a correlation of 0.74 with an n of 34 are 0.54 and 0.86. We conclude that there is a real, though modest, association between overall ranking data and marginal frequencies of observed data.

The above result is not really surprising and corresponds to the findings of Bernard, Killworth and Sailer who say

"Only one positive statement can be made about accuracy from our results. Although individual people did not know with whom *they* communicated, people *en masse* seemed to know certain broad facts about the communication pattern. Specifically, if we examine the aggregate of what everybody said about their communications with everybody, the resulting "most frequently communicated with members of the group" turns out to be correct. That is, the list of the top 6 most "popular" people is the same for both recall and behavioral data." (Bernard, Killworth and Sailer 1982: 62)

3.3. Eliminating outliers and normalizing data

The results above indicate some overall similarity between the observed behavior and the recalled rank data. In order to demonstrate specific structural similarities the data require some preliminary refining. Since the two data sets have such dissimilar distributional properties we also need appropriate transformations to make them more comparable.

The observational data in the form as given have enormous variability and confound communication interaction effects within the matrix with the differential marginal effects (see *e.g.* Mosteller 1968; Romney 1971; Romney, Klein and Kieffer 1973; Bishop, Feinberg and Holland 1975; Feinberg 1977). In addition, there are individuals who were observed interacting so few times that it really does not make sense to attempt to analyze their position in the group structure.

Table 1
Frequency distribution of number talked to for the 34 subjects observed behavior data.

Number of people talked to	Frequency	Number of people talked to	Frequency
1.	1	10.	4
2.	2	11.	2
3.	2	12.	4
4.	1	13.	2
5.	0	14.	2
6.	1	15.	2
7.	0	16.	3
8.	2	17.	1
9.	4	18.	1

Table 1 shows the frequency distribution of number of people talked to for all 34 subjects. It can be seen that the bulk of the subjects interacted or communicated with eight or more people. There are six clear-cut outliers who communicated with fewer people. The one subject who talked with 6 people is problematical. To illustrate the effects of these distributional problems, we scaled the raw behavior data using KYST (Kruskal, Young and Seery 1973). The results are shown in Fig. 2. Note that the low interactors are spread widely over the space while the subjects who interact with more people are clustered together. We have added a line around the high interactors to emphasize this. This effect is an artifact of the fact that the low interactors are distantly related both to the group and to each other simply because of low overall interaction. In addition to scattering the low interactors, this artifact also obscures the internal structure among the remaining subjects. We therefore eliminated the six subjects who interacted with four or fewer people from further analysis. Thus, the remainder of the paper will deal with a reduced sample of 28 subjects.

One of the main questions posed in this paper is whether the structure of the choice data is similar to the structure of the observed behavior data. In order to answer this question we need to be clear about what structure we are talking about and how to represent the structure in a way that allows direct comparison of the two structures. In the behavior data we are interested in the pattern of inter-communication among subjects separate from marginal frequencies that measure individual variability in amount of interaction. In order to unconfound or "take out" the effects of the marginals we do a proportional iterative fitting calculation that sets all marginals to a constant. In effect, it

Table 2
Mean individual correlations with 0.99 confidence limits with marginals and observed behavior.

Individual choice with following variable	Mean correlation [a]	0.99 Confidence limits[a]
1. Number of people talked to	0.37	0.30 to 0.45
2. Total interactions	0.33	0.24 to 0.41
3. Own behavior observations	0.47	0.41 to 0.53
4. Own normalized behavior	0.45	0.40 to 0.50

[a] Mean correlations and confidence limits calculated using Fisher's Z transformation (Fisher 1948: 197–210).

separates the pure interaction effects from the marginal effects. Several good sources are available for details of the method (see especially Bishop, Feinberg and Holland 1975: 97–102; Romney, Klein and Kieffer 1973). We substituted the figure of 0.05 for zero in our calculation. Further analysis of the structure of the behavior data is based on the normalized matrix. [2]

3.4. Individual accuracy on predicting marginals and own observed data

We have shown above that the overall MDPREF scaling of the choice data correlates 0.74 and 0.66 with number of people talked to and total interactions, respectively. We turn now to asking how well each individual can report, first, the interaction of others, and second, their own interactions. To measure the accuracy of each individual's report on the interaction of others, we correlate individual choice data (a rank order vector) with number of people talked to and with total interaction. To measure the accuracy of each individual's report of their own behavior we correlate individual choice data with the individual's original behavior interaction data and then with their own normalized behavior data. Thus we have for each of the 28 individuals four correlations (measures of accuracy) between rank order choice data and observed behavior data. The results are shown in Table 3. Table 2 shows the average correlations and the 0.99 confidence limits for the four sets of correlations.

As Bernard *et al.* note, there is some ambiguity as to the exact task that the subjects had in mind when they were "asked to rank the deck (i.e. the other subjects) from most to least communication that week. The question was purposely left rather vague" (Bernard, Killworth and Sailer 1980: 195). We do not know for sure whether the subjects were focusing upon overall communication, communications with self, or some combination. It is also posssible that different subjects had slightly different tasks in mind. With these cautions, we can attempt to interpret the figures in Table 2.

The first observation is that we can reject the hypothesis of chance accuracy with confidence despite the modest size of the correlations. The 0.99 confidence limits clearly exclude the zero point. Every one of

2 We want to thank Karl Reitz, an advanced graduate student at UCI for the normalization program, and Keiko Nakao for the programs to produce the figures.

Table 3
Individual data including number, total interactions, and correlations with marginals and observed behavior.

Subject	Number	Total	Correlation of rank data with			
			Number	Total	Raw observed	Normed observed
1	8	24	0.51	0.50	0.49	0.50
2	17	47	0.44	0.28	0.29	0.34
3	16	47	0.40	0.45	0.40	0.39
4	8	25	0.39	0.25	0.40	0.39
5	14	25	0.40	0.38	0.35	0.27
6	6	22	0.15	0.06	0.36	0.37
7	12	22	0.44	0.50	0.28	0.16
8	9	24	0.48	0.50	0.39	0.37
9	11	29	0.08	0.04	0.45	0.47
10	12	39	0.31	0.50	0.63	0.59
11	16	32	0.29	0.14	0.24	0.26
12	15	26	0.33	0.36	0.38	0.34
13	11	21	0.48	0.35	0.58	0.52
14	12	26	−0.05	0.01	0.48	0.51
15	12	46	0.50	0.57	0.67	0.60
16	9	28	−0.02	−0.03	0.47	0.48
17	10	32	0.41	0.50	0.60	0.53
18	18	79	0.54	0.39	0.42	0.42
19	10	18	0.33	0.14	0.44	0.44
20	10	21	0.42	0.23	0.34	0.35
21	13	50	0.40	0.44	0.63	0.57
22	16	34	0.40	0.42	0.60	0.54
23	9	13	0.10	0.02	0.53	0.54
24	13	27	0.48	0.41	0.42	0.44
25	10	23	0.55	0.49	0.50	0.47
26	14	28	0.41	0.33	0.60	0.57
27	8	17	0.55	0.38	0.53	0.53
28	15	30	0.49	0.37	0.50	0.44

the 28 correlations between choice and own behavior is positive. Even a simple binomial sign test shows that this result would occur by chance (assuming no accuracy) less than once in 100,000,000 times. The result is only a little less for number of people talked to, where 26 of 28 correlations are positive; a result that would occur by chance less than once in a million trials.

As to magnitude of the correlations, we can safely assume that the observed behavior marginal data are correlated with choice on an average between roughly 0.3 and 0.4. One's own behavior is correlated with choice roughly between 0.4 and 0.5. The question here is whether these individual correlations are too small to be of use in predicting structural regularities between verbal recall choice data and observed behavior data, as maintained by Bernard, Killworth and Sailer, or whether they contain useful and retrievable information. We turn now to an exploration of this question.

3.5. Prediction of behavior structure from choice structure

In order to talk about predicting structure, we need to be clear about just what we mean by structure and how to represent it. The structure of communication (or interaction, or friendship choices, or whatever) may be viewed as the pattern of interrelations among the subjects and may be represented in a spatial model in which individuals who communicate a lot with each other are placed close together in space and those who communicate less are placed further apart in space. Nonmetric multidimensional scaling provides an objective way of representing proximity data, such as the amount of communication, in Euclidean (or other) space.

Multidimensional scaling has several advantages for use in this context. First, it is objective in the sense that any of several different programs will provide the same constellation or pattern of points, i.e., the pattern of interpoint distances will be invariant under choice of program, computer, investigator, etc. Second, MDS will accept a variety of types of input data so long as the proximities reflect the desired content. This means that we can represent the behavior data in the same manner as we represent the rank-order choice data. Third, the output is metric and thus lends itself to a variety of statistical procedures. Fourth, the comparison of two or more structures is facilitated, so we can compare the structure of the behavior data to the structure of the choice data in a precise way. Fifth, the comparison of the structure to some outside variable is convenient and objective.

In this section we propose to show that the structure derived from the observed communication data corresponds very closely to a structure derived from the rank order choice data. We shall also demonstrate that one of the features or dimensions of the structure is associated with individual variation among the subjects on their overall knowledge, i.e., to show that knowledge is related to communication patterns. In order to make our analysis understandable, we need to make our theoretical assumptions explicit. We turn now to some informal theoretical speculations.

In most human groups in which the bulk of the members know each other, certain broad assumptions about their interactions may be made with some confidence. Whether we are talking about a college fraternity or sorority, a departmental faculty group, a face-to-face work group, or whatever, various regularities seem to emerge. For example, in whatever group we might observe, we would find that some members spend more of their time in the group, seem to have a bigger commitment to the group, and interact more with others of like interests in the same group. It may be assumed that part of what people in the same group talk to each other about are group concerns. More active members of the group talk to each other more often and exchange information, gossip, etc., more about matters relating to the group than do less active members. This leads to a situation in which there is a core of members who know more about what is going on in the group and who interact with each other more frequently than with others. They also share knowledge of the group more and, hence, if asked to rate group characteristics they should give answers more similar to each other than would members outside the central group.

There is no novelty in the above formulation. With respect to interpersonal attraction in a fraternity, Newcomb gave a rather complete and sensitive report of these dynamics over a quarter of a century ago (Newcomb 1956). He says, "Of all the objects about which we obtained responses, nothing compared in importance or in group relevance with the house members themselves. Very early they became differentiated in attraction status, so that it was easy to measure similarity, on the part of any pair of persons, in attraction towards the remaining members ... Thus the proposition could be tested that the greater the similarity between any two members in assigning General Liking scores to the other 15 members, the higher their attraction for each other." (Newcomb 1956: 582). Assuming that the general dynamics are similar over a wide range of interaction patterns, we can generalize the notion to the following general proposition: the more similarly two people judge the interaction pattern (whether it be knowledge, communication, friendship, etc.) of others, the more they interact with each other. There is a corollary to this proposition that may be stated as follows: the more two people share accurate knowledge of others, the more they interact with each other. The remainder of this section is devoted to the test of these two propositions.

The first proposition may be tested by comparing the structure of the observed communication behavior to the structure of the similarity of judgments of the subjects. The structure of the observed behavior is obtained by scaling the normalized 28 by 28 behavior data using KYST. The structure of the similarity of judgments of the subjects is obtained by forming a subject by subject matrix of correlations between subjects' rankings and then scaling the correlation matrix using KYST. To the extent that the general proposition is correct, and to the extent that both sets of data are reliable and valid, we would expect the two structures to be similar. The two structures may be compared by asking how well they could be accounted for by a single structure. The INDSCAL (Carroll and Chang 1970) model is a three-way multidimensional scaling technique that produces a common stimulus space, in this case a single representation of the interrelation of the 28 subjects. Unlike KYST, INDSCAL provides dimensions that are unique and invariant under rotation (Carroll and Wish 1974: 433). We performed these analyses and the single INDSCAL model in three dimensions accounts for 90 percent of the variance in the two KYST models used as input. The correlation between the computed scores (from INDSCAL) and the original data (from KYST) for the behavior data is 0.94, and for the ranking similarity data it is 0.95. Another measure of the similarity of these two matrices is given by the Quadratic Assignment Program which gives a Z of 10.6, clearly an indication of high similarity.

All of these measures are high and indicate that the model provides a good representation of the data and thus supports the general proposition that the more similarly two people rank the communication pattern of others, the more they interact with each other. In order to give a visual notion of just how similar the two structures are, we have mapped the two dimensional solutions into the same space. Figure 3 shows the results. We have oriented the two KYST structures to the solution provided by INDSCAL and drawn a line connecting each subject's ranking position to their corresponding behavior position. If the data were identical and perfect, then these positions for an individual would coincide. What is dramatic about the Figure is that the lines tend to be rather short and to remain in the same quadrant. We have drawn lines around the groups to highlight this. In fact the two KYST structures place all individuals but one into the same subgroup. This

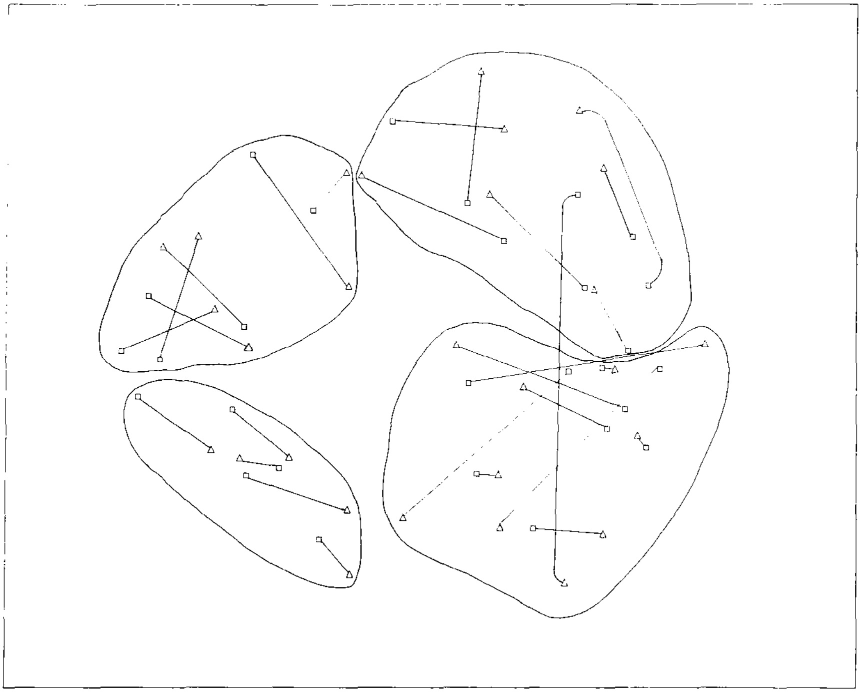

Figure 3

means that in 27 of 28 cases the structure derived from the rank-order choice similarities places individuals in the same subgroup as the observed behavior structure does. This seems to indicate to us that verbal data may be used to test propositions about observed behavior.

It remains to test the corollary proposition that states that people who share knowledge of the group should interact with each other more than with people with less knowledge. One index of knowledge of the group is the extent to which each individual's rank order choice is correlated with total interaction. We can think of this as an individual's knowledge of the marginal totals or total interaction. People with high knowledge share more knowledge than people with low knowledge. We can represent interaction with the joint INDSCAL picture. If we map individual knowledge scores and if the corollary proposition is correct, then similar scores should be found clustering in the picture. Results are shown in Fig. 4. It can be seen that the low scores tend to occur on

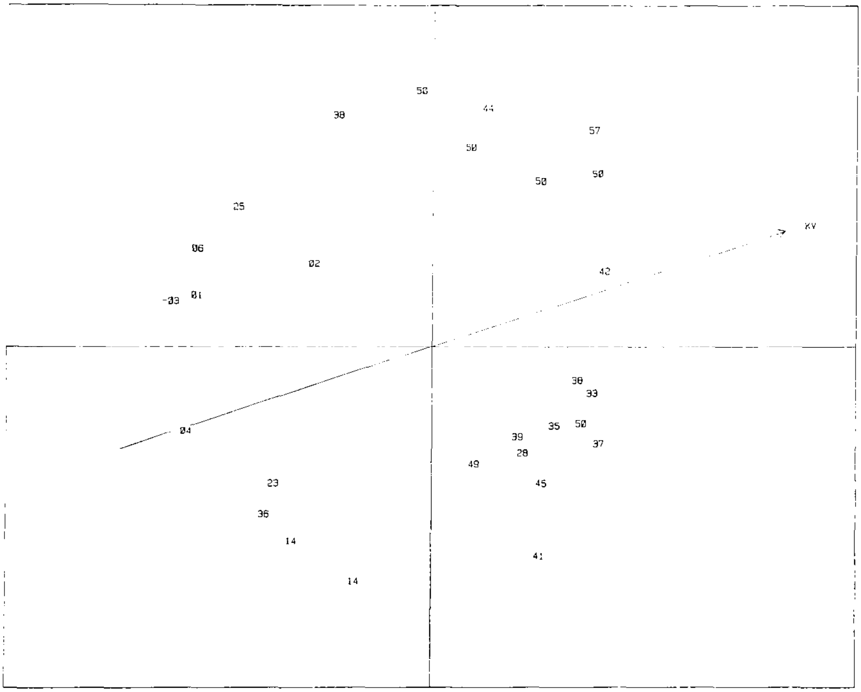

Figure 4

the left while high scores occur on the right. We can test for the strength of the association between a left–right dimension or vector and the high–low knowledge scores. PROFIT (Chang and Carroll 1968b) shows that there is a correlation of 0.83 between the fitted and actual projected scores associated with the vector. This means that there is a strong tendency for the high knowledge people to cluster on the right and, hence, interact together and for the low knowledge people to cluster on the left and, hence, interact with each other.

4. Discussion

In the latest of a series of papers dating back to 1975, Bernard, Killworth and Sailer summarize their findings and conclusions as follows:

> "However, one consistent and unavoidable conclusion has emerged from our studies of informant accuracy in network data: what people say, despite their presumed good intentions, bears no useful resemblance to their behavior. Many colleagues have found our conclusion objectionable. We must stress that we have never sought to test the accuracy of informants (in spite of the title of this paper); we have tested only the accuracy of social scientists' attempts to capture the structure of a communications network from recalled data." (Bernard, Killworth and Sailer 1982: 63)

It is our contention that in testing a reasonable proposition about human behavior we have been able to "capture the structure of a communications network from recalled data." We feel that the correspondence, individual by individual, of the four groups generally corresponding to the four quadrants in Fig. 3 is strong evidence in favor of the proposition that "the more similar two people judge the communication patterns of others, the more they interact with each other."

Bernard, Killworth and Sailer say, "we feel that it is vital in any field to have accurate (not just reliable) data. It is virtually impossible to develop a theory for any process unless one can obtain accurate data about that process" (Bernard, Killworth and Sailer 1982: 63). They have demonstrated to their satisfaction and the tech data, as well as a half dozen other data sets they present, do not contain sufficient accuracy to warrant basing any theory on it. We have demonstrated to our satisfaction that it lends strong support to theoretical propositions. This difference deserves a frank and honest discussion since it is undoubtedly based on differing assumptions and viewpoints. It also bears on some fundamental general issues in social science and the philosophy of science. Our speculations on the differences are offered below and are meant as tentative explorations rather than definitive.

Some of the obvious differences include the following: first, the difference is that Bernard, Killworth and Sailer were looking for error while we were looking for regularities; second, their expectations for the reliability and accuracy of any social science data were probably high while ours were low; third, we infer (although Bernard, Killworth and Sailer do not really come out and state it explicitly) that they feel the observational behavior is inherently more "accurate" or stable than the rankings, while our intuition is that the rankings are the more "accurate" and stable; fourth, Bernard, Killworth and Sailer focused upon the specific issue of how accurately people reported different aspects of their observed behavior (i.e., they asked, in effect, is recall a proxy for behavior?) while we broadened the question to ask how verbal behavior might relate in theory to observed behavior; and fifth, their methods focused more on individual measures while our main test took advantage of global aspects captured by normalization, aggregation, and multidimensional scaling. We will elaborate on each of these below.

The first difference mentioned above, that Bernard, Killworth and Sailer were looking for error while we were looking for regularity, is closely related to the second difference concerning expectations. We will therefore discuss them as one. The most extreme form of this difference would contrast a null model of pure chance association between recall and observed data (where even a small though statistically significant result might seem pleasing) to a null model of perfect association between recall and observed data (where any error would loom large and thus modest correlations would be discarded). Let us examine this with reference to the tech data. In Table 3, column 5, we find the data reporting the correlation for each of the 28 individuals in our final sample between their rank order recalled communication and their own observed behavior. This is one of the simplest tests of direct accuracy. If we assume a null model of chance association, we would expect that on the average half of the correlations would be negative and half positive. A simple binomial test shows that the probability of getting 28 positive correlations assuming such a model is about 0.000000004. Clearly, we have to reject such a null model. If we were to test a null model that says the correspondence is perfect, we would also have to reject it with an equally unrealistic probability value. What is the actual situation? Table 2 shows that after making the appropriate Z transformations (Fisher 1948), the average correlation is 0.47 with 0.99 confidence limits on the mean of values between 0.41 and 0.53. Does 0.47 reflect high accuracy or low accuracy? If one were expecting an accuracy of 0.9, then it would probably look pretty bad. If, on the other hand, one were expecting an accuracy of 0.2, then it would probably look pretty good. The issue is whether or not data of this accuracy or inaccuracy are useful for testing scientific propositions.

This problem is related to the problem of aggregation. In test construction, for example, it is recognized that average item intercorrelations can be rather low at the individual level but aggregate to highly reliable scales in the aggregate. For example, interitem correlations around 0.2 aggregate to a scale reliability of around 0.93 with 50 item scales. Clearly, there is no simple way to say whether 0.47 is accurate or inaccurate, but since the figure is an average of a series of individuals, we probably do not want to rule out the idea of finding any aggregate or global regularities.

We turn now to a consideration of the possible inherent superiority of the observed data *versus* the ranked recall data. Our own feeling is that the observed data have very severe sampling and distributional problems. We elected to eliminate, for example, subjects who were seen interacting with four or fewer other people. To attempt to predict much about the behavior of someone who is only observed to interact once in a week seems to us to be unreasonable. In a similar vein, to assume that just because the observers did not see two people interact during the week, that the people did not in fact communicate, seems to us to be completely unwarranted. If there were an ultimate reality of actual communication, we suspect that the rank order recall data might reflect it about as well as a sophisticated observational design.

The last two differences mentioned earlier are interrelated in such a way that we might best discuss them together. These have to do with testing some general theoretical ideas rather than with testing accuracy and with our methods of normalizing and scaling the aggregated data. We believe that the reason Bernard, Killworth and Sailer failed to find convincing structural similarities between the recall data and the observed data is that the makeup and distribution of the two data matrices are so radically different that most methods applied to the two sets of original data would fail to reveal the similarities. The Quadratic Assignment Program alerts us to the fact that there are striking similarities, but it does not specify what these similarities are. We feel that it was essential to remove the outliers (low interactors) from the observed data. It was then critical to remove the effects of the marginals by normalization. Only when these two steps were taken was it possible for multidimensional scaling properly to portay the underlying interaction pattern. The theory also led to the choice of person–person similarities in choice as the appropriate form for the recall data.

There is a final matter that we would like to speculate upon. It has to do with the idea that the "proof is in the pudding." By this we mean that if one accepts as convincing the evidence that the proposition is supported – i.e., that the more similar two people judge the communication pattern of others, the more they interact with others – then it follows that the data were sufficiently reliable and accurate to demonstrate support for the proposition. In other words, we have answered the question as to how accurate or inaccurate the data are. It is accurate enough to show that one can predict the observed behavior structure from recall data.

References

Bernard, H.R. and P.D. Killworth

1977 "Informant accuracy in social network data II". *Human Communication Research* 4:3–18.

Bernard, H.R., P.D. Killworth and L. Sailer

1980 "Informant accuracy in social network data IV: A comparison of clique-level structure in behavioral and cognitive network data". *Social Networks* 2: 191–218.

1981 "A note on inferences regarding networks subgroups: Response to Burt and Bittner". *Social Networks* 3: 89–92.

1982 "Informant accuracy in social-network data V. An experimental attempt to predict actual communication from recall data". *Social Science Research* 11: 30–66.

Bishop, Y.M., S.E. Feinberg and P.W. Holland

1975 *Discrete Multivariate Analysis*. Cambridge, MA: MIT Press.

Burt, R.S. and W.M. Bittner

1981 "A note on inferences regarding network subgroups". *Social Networks* 3: 71–88.

Carroll, J.D.

1972 "Individual differences and multidimensional scaling". In Shepard, R.N., A.K. Romney and S.B. Nerlove. *Multidimensional Scaling: Theory and Applications in the Behavioral Sciences*, Vol. 1. New York: Seminar Press.

Carroll, J.D. and J.J. Chang

1970 "Analysis of individual differences in multidimensional scaling via an *n*-way generalization of 'Eckart–Young' decomposition." *Psychometrika* 35: 283–319.

Carroll, J.D. and M. Wish

1974 "Multidimensional perceptual models and measurement methods". In E.C. Carterette and M.P. Friedman (eds) *Handbook of Perception*, Vol. 2. New York: Academic Press.

Chang, J.J. and J.D. Carroll

1968a "How to use MDPREF, a computer program for multidimensional analysis of preference data". Bell Laboratories (unpublished).

1968b "How to use PROFIT, a computer program for property fitting by optimizing nonlinear or linear correlation". Bell Laboratories (unpublished).

Coombs, C.H.

1964 *A Theory of Data*. New York: John Wiley.

Feinberg, S.E.

1977 *The Analysis of Cross-Classified Categorical Data*. Cambridge, MA; MIT Press.

Fisher, R.A.

1948 *Statistical Methods for Research Workers*.. New York: Hafner.

Hubert, L.J. and F.B. Baker

1978 "Evaluating the conformity of sociometric measurements". *Psychometrika* 43: 31–41.

Hubert, L.J. and J. Schultz

1976 "Quadratic assignment as a general data analysis strategy". *British Journal of Mathematical and Statistical Psychology* 29: 190–241.

Killworth, P.D. and H.R. Bernard
1976 "Informant accuracy in social network data". *Human Organization* 35: 269–296.
1979 "Informant accuracy in social network data III, or: A comparison of triadic structure in behavioral and cognitive data". *Social Networks* 2: 19–46.

Kruskal, J.B., F.W. Young and J.B. Seery
1973 "How to use KYST, a very flexible program to do multidimensional scaling and unfolding". Bell Laboratories (unpublished).

Mosteller, F.
1968 "Association and estimation in contingency tables". *Journal of the American Statistical Association* 63: 1–28.

Newcomb, T.M.
1956 "The prediction of interpersonal attraction". *The American Psychologist* 11: 575–586.

Romney, A.K.
1971 "Measuring endogamy". In P. Kay (ed.). *Explorations in Mathematical Anthropology*. Cambridge, MA: MIT Press.

Romney, A.K., R.E. Klein and M. Kieffer
1973 "A normalization procedure for correcting biased response data". *Social Science Research* 2: 307–320.