DOCUMENT RESUME

ED 235 515 CS 504 300

AUTHOR Barnett, George A.; Palmer, Mark T.
TITLE Longitudinal Network Analysis Using Multidimensional Scaling.
PUB DATE May 83
NOTE 46p.; Paper presented at the Annual Meeting of the International Communication Association (Dallas, TX, May 26-30, 1983).
PUB TYPE Viewpoints (120) -- Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Air Transportation; *Communication Research; Economics; Longitudinal Studies; *Multidimensional Scaling; *Network Analysis; *Research Methodology; *Social Networks
IDENTIFIERS *Galileo System

ABSTRACT

The Galileo System, a variant of metric multidimensional scaling, is used in this paper to analyze over-time changes in social networks. The paper first discusses the theoretical necessity for the use of this procedure and the methodological problems associated with its use. It then examines the air traffic network among 31 major cities in the United States over the 14-year period of 1968-1981, demonstrating how the proposed method provides insights into activities within the network and how exogenous factors such as the physical distances among the nodes, changes within the airline industry, and economic conditions affect the changing network structure. (Author)

LONGITUDINAL NETWORK ANALYSIS USING MULTIDIMENSIONAL SCALING

George A. Barnett
State Universiy of New York at Buffalo

Mark T. Palmer
University of Wisconsin--Madison

Presented to the Information Systems Division of the International Communication Association, Dallas, May 26-30, 1983.

LONGITUDINAL NETWORK ANALYSIS USING MULTIDIMENSIONAL SCALING

ABSTRACT

This paper proposes that a variant of metric multidimensional scaling, the Galileo System (tm) be used to analyze over-time changes in social networks. The paper discusses the theoretical necessity for the use of these procedures and the methodological problems associated with their use. Next, it examines the air traffic network among 31 major American cities over the 14 year period, 1968-1981. It demonstrates how the proposed method provides insights into activities within the network and how exogenous factors such as, the physical distances among the nodes, changes within the airline industry and economic conditions impact upon the changing network structure.

LONGITUDINAL NETWORK ANALYSIS USING MULTIDIMENSIONAL SCALING 1

INTRODUCTION

A social network may be precisely defined by a N x N matrix S, where N equals the number of nodes or interacting units in the network. 2 The value in each cell (sij) is some measured attribute of the relationship or link between nodes i and j. In communication research, the value is generally the frequency of communication often weighted by the perceived importance. While there exist a variety of techniques for analyzing this matrix, sociometry (Moreno, 1934), matrix manipulations (Forsyth & Katz, 1946; Festinger, 1949), network analysis (Pitts, 1979; Richards [NEGOPY], 1974; Breiger, et al.[CONCOR], 1975; Bernard & Killworth [CATIJ], 1973; Alba [COMPTL], 1973) and multidimensional scaling (Goldstein, et al., 1966; Jones & Young, 1972; Lankford, 1974), none of these methods is clearly superior for the analysis of sociometric data and all are incapable of precisely describing changes in networks over time. A variant of metric multidimensional scaling, the Galileo System(tm) (Woelfel & Fink, 1980), however, may be used to precisely analyze over-time changes in social networks (Gillham & Woelfel, 1977), and to provide insights into the nature of networks (Barnett, 1979). This paper will discuss the theoretical necessity for using these procedures for the analysis of network data and certain methodological problems associated with this approach. These problems include the specification of a mathematical transformation to change network data into the proper form for multidimensional scaling and under which conditions to apply one of many alternative rotational algorithms which describe how networks change. It will then examine the American air traffic network to demonstrate the utility of the method for longitudinal network analysis.

Although variants of network analysis have been applied to study social and organizational structure for nearly fifty years, little progress has been made in developing procedures to study change in networks. Changes in social networks may be caused by external factors such as technological innovations or information made available to the members of the system, or internal factors, such as the growth of an organization or the departure of a member from the system. The critical point is that social networks do change over time.

Rogers and Kincaid (1980) report few over time studies in their review of network analysis. Among the reported studies were, Lloyd-Kolkin's investigation of the evolution of 11 R & D organizations into an interconnected system over a nine month period; Stern's (1979) historical study of the NCAA; Freeman and Freeman's (1979) study of computer-based teleconferencing among network scholars; and, Morett-Lopez's (1979) research on network stability in Monterrey, Mexico's slums.

Burt and Lin (1977) developed a structural equation model to describe change in network structure over a 95 year period in the United States. Performing a content analysis of archival records (the front page of the New York Times), they formed sociomatrices based on the structural equivalence of actor categories averaged over a four year period. They reported that over time greater attention was paid to members of government agencies and less to individuals connected to political parties or business leaders.

Roberts and O'Reilly (1978) examined the communication networks within three interrelated high-technology Navy organizations at two points in time. The first was three months after the unit was established and the second was a year later. Using Richards's (1974) procedures, they found that the ratio of participants to isolates was roughly consistant, the limited change which occurred was in the direction from isolate to participant, rather than in reverse, and that network integration increased over time. The number of group members and groups were greater at time two. While group size remained relatively stable, the interconnections among the groups increased.

The dynamic nature of social networks has not been studies for two major reasons. One, the data generally gathered by social scientists has predominately been cross-sectional rather than longitudinal (Rogers & Kincaid, 1980). And, two, there have not been procedures to analyze over time network data. This paper presents such a methodology, the Galileo System (tm) of metric multidimensional scaling. It was designed to study the changes in distance matrices (like S) under a variety of theoretical constraints.

This change from analyzing networks as static structures to dynamic entities is theoretically important for communication. Perhaps no single variable is more central to the study of communication than time (Barnett, 1982a). Communication is universally defined as a process whereby information is exchanged among systems. However, while there has been considerable verbal theorizing about the communication process, little has been done to empirically demonstrate these processes. This may be due to the lack of precise procedures to measure the structural change which results from the exchange of information among social networks. As a result, there has been little advancement in theory about communication networks. The proposed procedures will make it possible to identify descrepancies between verbal theory and empirical observations. Thus, theories can be adjusted to account for these descrpancies. Thus, adoption of these procedures will allow for rapid growth of communication theory.

THEORY

Implicit to any theory of networks is the notion of "betweenness". That is, one node (a) lies between two others (b & c) such that information passed between nodes b and c almost always goes through node a. B and c rarely communicate directly. This would be the case if node a were a central switching facility (Schwartz, 1977) or a liaison in a social organization (Rogers & Agarwala-Rogers, 1976). In terms of the communication distances among the nodes, a is very close to b and c, but b and c are quite far from one another. These distances may be considered the inverse or reciprocal of the frequency of the use of link between the nodes.

The distance among the nodes may be represented by a matrix like the one below:

S

	a	b	c
a	0	1	1
b	1	0	9
c	1	9	0

The diagonal contains zeros because the distance between any node any itself is zero by definition.

If matrix S were converted to a spatial model by finding the eigenvector of its scalar products matrix S^TS, 3 one would find that the eigenroots (eigenvalues) or characteristic roots of S^TS would include one negative root. The reason for this is that the triangle formed from the links of the abc triad cannot exist in a two-dimensional Euclidean space. 4 The abc triangle has two very short legs (ab & ac) and one very long one (bc). As a result, the sum of the triangle's angles exceeds 180^o. Thus, this triad cannot be accurately described without a complex dimension (one with a negative root) to foreshorten the bc leg. 5

Network data need not be Euclidean, i.e., at least one of the characteristic roots of S^TS may be imaginary. The reason for this is that if any three nodes vary in centrality, the points must violate the rule of triangular inequalities. The exception is a completely and approximently equivalent interconnected network. Any three points (nodes) may be said to form an Euclidean triangle if and only if the sum of the square of any two of the distances among them does not exceed the third squared (Tversky, 1979). In the example above, bc must be less than or equal to $\sqrt{2}$, if this triad is to exist in an Euclidean space. For any set of N nodes in matrix S, those nodes will be represented in Euclidean configuration if and only if the triangluar inequalities rule is not violated for any triple of points. The result is a Riemann manifold represented by a coordinate system in which some of the dimensions are imaginary. They have negative eigenroots. The locations of the non-Euclidean relations among the points may be determined by equation 1.

$$d_{jk}^2 = d_{ij}^2 + d_{ik}^2 - 2d_{ij}d_{ik}\cos\theta \qquad 1.$$

In the case where, $\cos\theta \leq 1.0$, the relations may be considered Euclidean. Where, $\cos\theta > 1.0$, the relations among the three nodes may be considered non-Euclidean or Riemannian. It is from this latter case that complex eigenroots result (Woelfel & Barnett, 1982).

While multidimensional scaling has frequently been applied to analyze social networks (Goldstein, et al., 1966; Jones & Young, 1972; Lankford, 1974; Breiger, et al., 1975; Gillham & Woelfel, 1977; Freeman & Freeman, 1979; Romney & Faust, 1982), less than satisfactory results have been reported (Lankford, 1974; Breiger, et al., 1975). One reason for this may be the failure to take into account the imaginary dimensions.

Historically, psychometricians have treated the variance on these dimensions as error variance to be removed through the addition of an additive constant (Messick & Abelson, 1956) or adjusted away by some non-metric algorthim (Shepard, 1962a,b; Kruskal, 1964a,b). They assumed that social and psychological structures were Euclidean and that any departure from a positive semi-definate scalar products matrix (S^TS), one with only positive values in its eigenvector, was caused exclusively by measurement error. Thus, these dimensions were ignored and inadequate descriptions of sociometric data resulted. Additionally, the stated purpose for using multidimensional scaling was to identify some underlying structure, such as, the dimensions by which a group was differentiated. This resulted in the removal of true variance. The imaginary variance went first. However, since the underlying dimensions are only orthonormal reference vectors upon which no meaning may be directly attributed, all dimensions should be retained for any further analysis, including those with negative eigenroots (Barnett & Woelfel, 1979). Attribution of meaning to the dimensions may be made only by regressing an attribute vector through the multidimensional space.

Recently, however, psychometricians have become interested in multidimensional scaling in Riemann space (Pieszko, 1975; Lindman & Caell, 1978). One algorithm exists which allows for the analysis of all the dimensions in an multidimensional manifold including those with negative roots. It is known as Galileo (tm) (Woelfel, et al., 1976). The computer program takes ratio level measurements of discrepancies (distance or dissimilarities), such as matrix S, and converts it to an adjusted scalar products matrix following Torgerson (1958), (S^TS). It then finds that matrix's eigenroots and Cartesian coordinates (S^*) for all dimensions, real and imaginary, through Jacobi's method (Van de Geer, 1971).

One reason for performing network analysis has been clique or group identification. Two procedures may be perform this function, cluster analysis or multiple discriminant analysis (MDA). Once the Riemann space (S^*) has been obtained, the researcher may perform a cluster analysis to identify groupings within the space. An alternative technique, when groups identification is known or hypothesized, is MDA (Jones & Young, 1972). In this case, group membership may be considered the dependent variable and the dimensions (real and imaginary) the predictor variables.

Change in network structure may be examined by repeating the measurement phase and transforming the data for each point in time into multidimensional spaces. To compare several points in time (or several different groups at the same time), the spaces must be translated to a common origin and rotated to a least squares best fit which minimizes the departure from congruence among the spaces. Change in the position of the nodes may be calculated by subtracting the coordinate values across time. From these change scores trajectories of motion can be determined to describe the relative changes in the structure. With these measured velocities (the rate of change over time) and accelerations future network structure can be predicted accurately (Barnett, 1979; Barnett & Kincaid, 1983).

When no additional information about the relative stability of the nodes exists, the ordinary least square procedure may be applied. When knowledge about the nodes stability or that the position of certain ones have changed is known, alterative rotational algorithms exist (Woelfel, et al., 1979). The ordinary least squares procedures has the effect of overestimating some changes while underestimating others. This may lead to erroneous conclusions. The alternative rotational schemes use theoretical or "extra" information which simplifies the apparent motion. Since it is independent of the coordinate values, it may be treated as invariant under rotation and translation of the coordinates.

One alternative scheme rotates only the theoretical stable points to a least squares best fit and then incorporates the dynamic ones into the new coordinate system. This is similar to the procedure used in astronomy where the position of fixed stars are used to measure the motion of other stellar bodies. Another procedure weights the individual points, and then rotates to a weighted solution. One of these schemes should be used when manipulating the relational patterns of a nodes toward a subset of nodes. In that case, the manipulated nodes are considered dynamic and the unmanipulated ones are treated as theoretically stable reference points (Woelfel, et al., 1980; Barnett, 1980). The algorithims necessary to perform the rotations described here are unique to the Galileo (tm) computer program and make it possible to precisely study change in networks.

Previous research with Galileo has shown that the loadings on the imaginary dimensions are reliable both across groups and over time (Woelfel & Barnett, 1982). Also, theoretically valid predictions have been made using the imaginary dimensions. Woelfel and Barnett (1982) have shown that the dimensions with negative roots result when pair comparisons among three or more stimuli concepts are made from two or more semantic domains or when the stimuli are incongruent or produce a psychological state of inbalance. Krumhansl (1978) examined psychological non-Euclieanisms in geometric models. She found that violations of triangular inequalities resulted in similarity data when the scaled points varied greatly in their relative density. In spaces where the points were distributed homogenously, there was a greater tendency for the space to be Euclidean.

Barnett (1979) found that imaginary dimensions resulted in the analysis of social networks. Using the frequency of air traffic for the year ending June 30, 1978, among 16 American cities, he found that 40.2% of the total variance in S^* was accounted for by those characteristic roots of S^TS which were negative. A warp factor of 3.04 was obtained. Warp is the ratio of the sum of all the eigenroots (positive and negative) to the sum of the positive roots. Thus, it provides a convenient measure of the degree to which the space is non-Euclidean. A warp of 1.0 indicates an Euclidean space. An examination of the three dimensions (2 real and 1 imaginary) which accounted for the greatest proportion of the variance (70.8%) and would have been retained by a scree test (Barnett & Woelfel, 1979), suggested that the frequency of air traffic may be described as a star-type network with tendencies toward a tree-type configuration, although among the nodes at the center (hub) a mesh-type network was the best descriptive label (Schwartz, 1977). In order to travel by airplane from New Orleans to Phoenix or Seattle, the nodes at the periphery or the points of the star, one had to go through one of the central switching nodes, such as, Chicago. Also, the results suggested that Atlanta served as a tree node or an intermediate switching facility, taking passengers from New Orleans and Miami and rerouting their travel prior to reaching the more central nodes. Among the central nodes (Chicago, Cleveland, Dallas, Denver, New York, Los Angeles, San Francisco and Washington), each node had a direct link to each other.

This study clearly demonstrated the utility of using a multidimensional scaling algorithm which is not restricted to an Euclidean solution. That study provided only a static description and did not demonstrate change over time in social networks. This paper will focus on the longitudinal nature of the air traffic network.

The utility of any scientific methodology is ultimately its ability to precisely describe attributes of phenomena and to make accurate predictions of the values of these attributes at future points in time. These predictions are based upon and evaluated against the prevaling theories about the phenomena for which the methodology was developed. These descriptions should lead to parsimonious "law-like" relations between measured attributes of the phenomena and other variables which are theoretically related. Generally, these are in the form of mathematical functions.

Since the proposed procedures are designed for the study of change in networks, it is necessary to demonstrate that they provide a description of the change in simple "law like" functions. They should covary with those exogenous factors which predict change in network structure. Such factors might be the physical relations and similarity among the nodes, economic conditions, the diffusion of new communication technologies, population growth and mobility and changes within the network itself.

Up to this point, this paper has discussed the theoretical necessity of using a non-Euclidean multidimensional scaling algorithm to describe social or communication networks. It has been suggested that any new methodology's utility should be evaluated against theoretical criteria. This paper will empirically demonstrate these procedures using data on the frequency of air traffic between 1968 and 1981. Change in this network will be described by simple law like functions which will be analyzed with respect to certain theoretical criteria. In this manner the adoption of this methodology for the study of network change will be justified.

AN EMPIRICAL EXAMPLE

THE DATA

To demonstrate the utility of these procedures, data from the annual "Domestic Origin-Destination Survey of Airline Passenger Traffic" conducted by the U.S. Civil Aeronautics Board (CAB), in cooperation with the certified route air carriers and the Air

Transport Association of America, were analyzed. 6 A single survey is conducted continuously on the basis of a 10% sample. Flight coupons surrendered by passengers upon boarding are the source of the survey data. The universe consists of all coupons lifted by participating air carriers. Coupons are selected for analysis with ticket serial numbers ending in zero. These data are compiled by the CAB. They edit the data, remove inconsistancies, such as duplication of the same flight by different carriers, itineraries in which no destination is reported, single coupons in which the origin and destination are the same, and itineraries where the carrier(s) into and out of an intermediate point do not serve the city. Also removed from the data base are records which fail compute editing tests. In all, less than two per cent of the total reported number of flights are dropped from the survey.

Thirty-one cities (SMSA) with a population greater than one million were selected as the nodes. 7 They are listed in table 1. In 1980, these 31 cities had a cumulative population of 94,092,000 or 43.5% of the total U.S. population. The links in sociomatrix, S, were the number of passengers outbound plus inbound (nondirectional) between the cities. 8 Since nondirectional relations were used, S was symmetrical (sij=sji). The diagonal contained zeros. Fourteen separate sociomatrices were created, one for each year 1968 to 1981. This made it possible to examine the change in the air traffic network for this time frame.

TABLE 1 ABOUT HERE

These data were obtained on microfilm and were first converted to hard copy. To insure a minimum of coding error, both sij and sji were recorded. Then, they were checked for equivalence. Complete sociomatrices were entered into the computer for analysis. Again, sij and sji were compared and corrections made. In summary, the data consisted of 14 symmetrical sociomatrices containing the frequency of nondirectional passenger air traffic among 31 U.S. cities.

These data are not subject to the criticism of self-report network data (Bernard & Killworth, 1977). Rather than being reports of travel by individuals, they are objetive, coming from used airline tickets. Further, they are aggregate data (Rogers & Kincaid, 1980). The nodes (unit of analysis) in this study are cities, not individuals. Thus, the interaction among aggregates were examined. Danowski (1980) and Barnett (1982a) have argued that the process of aggregating to the group level filters out a significant amount of measurement error because random individual variation and the effects of other communication channels are randomized. The result is stable estimates of the state of the system which improves the ability to describe the underlying mathematical relations among the variables of interest. In this case, a 10% sample of air traffic is sufficiently large to assume that random perturbations contribute little to the description of the network.

TRANSFORMING MATRICES OF FREQUENCY TO COMMUNICATION DISTANCES

The first step in the analysis of these data is to transform the matrices of frequency of interaction to (S) to matrices of communication distance (S') to confirm with the Galileo (tm) framework. The goal of this operation is to assign the smallest value to the greatest frequency. The logic is that the greater the interaction between two nodes, the closer they are in a spatial network. The problem is what functional transformation to apply. Two candidates are the inverse and the reciprocal.

The frequencies may be subtracted from an arbitrarily large constant, k, where k is greater than the largest value of sij. s'ii equals zero. This function is presented as equation 2.

$$S' = K - S \qquad 2$$

In this case, S' is a linear function of S. It has the advantage of simplicity. One problem is what value to assign to k. According to Woelfel (personal correspondence), k should have theoretical significance. For example, k could be set equal to the maximum possible number of passengers flying among the nodes. If a "convenient" value, rather than a theoretical one, is chosen, the rank of matrix S becomes arbitrary and therefore no meaning can be attributed to the warp of the network space. This is not a major problem if network spaces are compared relative to one another rather than to some external criterion. However, as k becomes larger, the transformation has the effect of adding an "additive constant", which alters the dimensionality (rank) of the network space. This problem may be exasperated when the same constant is applied to several different sociomatrices. What may be the "most convenient" constant for one point in time may not be appropriate for another.

To analyze the air traffic network, k was set equal to the maximum frequency in the data plus one. That value was 318,673 or one greater than the frequency of traffic between Los Angeles and San Francisco in 1981. This is summarized in 2.1.

$$S' = 318,673 - S \qquad 2.1$$

An alternative transformation function is the reciprocal, presented in equation 3.

$$S' = S^{-1}(k) \qquad 3$$

k is a scaling constant. This function is nonlinear (hyperbolic). As $s_{ij} \to 0$, $s'_{ij} \to \infty$, and $s_{ij} \to \infty$, $s'_{ij} \to 0$. sij may obtain the value of 0 when there is no link between i and j. Thus, this function has the disadvantage of placing too much emphasis on very weak links. For example, in 1968, the frequency of traffic between San Diego and Ft. Lauderdale was only two. Where two nodes are not linked, S' is undefined. Therefore, this transformation cannot be applied in those instances. In this data set, all nodes are completely interconnected and only the weakly linked ones merit concern.

There is also the problem of what value to assign to k. In this case, an external theoretical criterion, the physical distances among the cities, was selected. k was set equal to the value required to set the trace of the time one (1968) network space equal to the trace of the space of the great circle distances (in kilometers) among the cities. This makes it possible to directly compare the network space to the physical.

Because of the weak link problem, a third alternative function was selected for analysis. It is the log of the reciprocal and it is presented as equation 4.

$$S' = \log S^{-1}(k) \qquad 4$$

This transformation foreshortens extreme values and linearizes the function. Prior research has shown that logarithmic transformations alter the rank of spatial manifolds (Woelfel & Barnett, 1982), producing essentially Euclidean spaces. This requires that the network space be compared relative to one another rather than against some absolute criterion. k was set equal to the value required for equivalent traces between the spaces produced by the physical distances between the cities and air traffic networks. That value was 14,638. The final transformation function is presented in 4.1.

$$S' = \log S^{-1}(14{,}638) \qquad 4.1$$

All analyses discussed in this paper will be based upon communications distances generated with equation 4.1. It was selected because the coefficients which resulted from its application were within a convenient or middle range (Stevens, 1951). The values were easy to work with and thus, accessable or interpretable to network scholars. Accessibility facilitates communication among scientists (Barnett, 1982b).

Equation 4.1 represents a compromise between the reciprocal and the inverse functions. The 1968 data were used to test the three function discussed above. The inverse was viewed as unacceptable because it resulted in an almost Euclidean solution (warp = 1.02). The San Diego, Ft. Lauderdale, New York City triad was clearly non-Euclidean. There were only 2 trips between San Diego and Ft. Lauderdale and 8,396 and 24,662 between New York and the other two nodes respectively.[9] This result occurred because k acted as an additive constant. Also, the values which resulted from this analysis were not easily used. The trace of the 1968 sociomatrix was 1.323 x 10^{10}. Additionally, the first two dimensions accounted for only 8.0% and 6.1% of the variance, with dimensions three to twenty-seven accounting for between 4.4% and 2.2%. If all the links among the nodes were of equal strength each dimension would account for 3.2% of the variance. While a large number of dimensions would be expected because all the nodes are interconnected, there was little differentiation among the eigenroots. This raised some concern about the utility of this function. The first dimension separated the peripherial nodes, San Diego and Ft. Lauderdale, from the more central nodes. There was little differentation among the remaining 29 nodes. Finally, an examination of the difference scores among all 14 spaces failed to reveal any apparent pattern.

The first dimension of the simple reciprocal transformation had the two nodes with the least contacts as bipolar, and the other nodes at the center, rather than differentiating them from the other nodes as with the inverse. While this result was desirable, others were not. All variance in the space occurred on the first and last (largest imaginary) dimension. They accounted for 1460.0% and -1360.5% of the variance respectively. The second dimension accounted for only 15%. Due to the extent of interconnection among the nodes, this result seems inappropriate, as did the warp which was 14.6.

For the theoretical reasons discussed above and these empirical results, the decision was made to base the description of the change in the air traffic network on equation 4.1. Its warp was 2.48. There was some differentation among the dimensions. Dimension 1 accounted for 123.1% of the variance, dimension 2, 31.5%, dimension 30, -28.3% and dimension 31, -79.77%. The resultant values were a convenient size. Finally, Ft. Lauderdale and San Diego were at the extremes and the other nodes were near the origin.

RESULTS

CHANGE IN CONNECTEDNESS OVER TIME

The 14 sociomatrices of airline traffic were transformed into multidimensional spaces and comparisons made using a rotation to a least squares best fit which minimized overall departure from congruence. In this way, global changes in the air traffic network were examined. Rather than presenting all 14 sets of coordinates and 13 comparisons among the coordinates, only summary indicators of the changing relations will be reported. One such indicator is the trace of the coordinates matrix. The trace or sum of the eigenroots may be taken to be a measure of the network's size. It may be used as an indicator of the network's connectedness. The smaller the value, the greater the connectedness. The traces' values for the 14 years are presented in table 2.

To describe the change in connectedness over time, these 14 values were plotted against time. A visual examination of the data revealed that the trace decreased rapidly during the first few data points, levelled off and increased slightly for the last two points. In other words, connectedness increased rapidly during the first few years, levelled off during the later years and decreased slightly at the end. This examination also suggested that the pattern of change could be described by a simple exponential decay function,

$$Y = a + be^{-kt} \qquad 5$$

where,

a = asymptote
b = Y at time zero
k = coefficient of decay

Since the last two values of the trace were greater than the ones which immediately preceeded them, an alternative function was suggested. It was a polynomial with an intercept, a negative linear component and a positive quadratic term. The later term would account for the reversal in the trend.

$$Y = a - b_1 t + b_2 t^2 \qquad 6$$

The data were fit to both functions. In the case of exponential decay, a = 43,838 b = 51,919 and, k = -.82, R^2 = .864.

For the polynomial, a = 63,041, b_1 = 3,806 and b_2 = 169, R^2 = .752. Both the linear and quadratic terms were statistically significant. F = 13.18 ($p<.004$) and F = 6.19 ($p<.03$) respectively.

These results indicate that the proposed methods can be used to provide parsimonious, "law-like" descriptions of the change in social networks. They are summarized in table 2 and figure 1.

TABLE 2 AND FIGURE 1 ABOUT HERE

While both functions account for a sizable proportion of the variance in the trace over time, they provide different information. The exponential decay function does not account for the sign reversal of at the last two points. It treats them as deviations from the asymptote, as it fits the entire data set. While the polynomial accounts for less variance, it does account for the reversal. The quadratic term is significant. Thus, while the overall pattern of connectedness in the air traffic network has increase exponentially, this pattern may be changing. Connectedness may be decreasing. Two points are two few to make any definitive statements about this trend. The curve may be oscillating about the eventual asymptote. It should be examined in the future to verify the trend. Additional insights may be gained by examining the magnitude of change between the sociomatrices.

THE CHANGE SCORES

The overall change scores (differences between time i and time i+1) from Galileo (tm) reveals a consistant pattern. 10 They are presented in table 3 and are graphically displayed in figure 2. The data suggest two distinct epoches, an early period, 1968-1974, characterized by a high rate of change and, a stable later period, 1974-1981. The magnitude of difference between averages of these sets of points indicates that the rate of change was 7.51 times greater for the early epoch. The airline network initially changed rapidly and then slowed to a stable pattern with a slow rate of change. As will be discussed later, this difference may be attributable to the opening of the Dallas-Ft. Worth airport which acts as a central hub or switching facility. These results suggest that the exponential decay may provide a better description of the pattern of change in connectedness than the polynomial because the rate of change at the end is so small.

TABLE 3 AND FIGURE 2 ABOUT HERE

THE CHANGE OF INDIVIDUAL NODES

Insights into the changing pattern of the nodes' relationships may be gained by examining specific nodes. In the early years, Ft. Lauderdale and San Diego changed more than twice the overall average for each of these seven years. During the later period, Tampa and Dallas changed more than the average. But, these changes were quite small when the overall magnitude of change during the early epoch is considered.

Specifically, how did the position of these nodes change over time? Ft. Lauderdale moved from the periphery toward the center of the network. Through hierarchical cluster analysis, it was determined that it was the least central node in 1968. 11 By 1976, it was the fourth least central. In 1981, it was the seventh least central node. San Diego, likewise moved from the periphery to the center of the network. In 1968, it was the second least central node. Within two years, it stabilized as the tenth least central node. These nodes were replaced at the periphery by the smaller cities in the midwest, Columbus, Cincinnati and Indianapolis.

Tampa clustered with Atlanta during the early years. Between 1975 and 1977, it moved from being a "branch" of Atlanta to become directly interconnected with the other nodes at a national level. Dallas continued to become more central in the network throughout the later period.

CHANGE IN THE AIR TRAFFIC NETWORK STRUCTURE

How did the overall network structure change over time? Groups within the network were identified by hierarchical cluster analysis. In the early years, there were two regional groups or clusters. One was centered about Chicago and New York and included all the eastern and midwestern cities from Miami to Minneapolis. The other cluster was centered on the west coast around Los Angeles and San Francisco. It included another cluster which contained New Orleans, Dallas and Houston. Hierarchical clustering combines all nodes into a single cluster. These two regional groups were combined at iteration 22, in 1968, and iteration 17, in 1969. Worth noting were the positions of Kansas City and St. Louis. While the later was part of the eastern cluster, the former was grouped with the west. The break in the air traffic network in 1968-69 appeared to go through the middle of Missouri north to the west of Minneapolis and south to the east of New Orleans.

Analysis of the later years, fails to find as profound regional variation in the network structure. The cluster analysis shows that New York, Chicago, Los Angeles, San Francisco, Dallas and Houston are combined into a single cluster immediately, at iteration 4. The other individual nodes were then added to this hub with little prior regional clustering.

This conclusion was confirmed through regression analysis. 12 The coordinate values of an early year, 1969, and a later year, 1980, were regressed on laditude and longitude, the dimensions of physical space. In the early year, the first dimension accounted for 70.4% of the variance in longitude and the first four, 83.6%. The first dimension accounted for 25.1% of the variance in network structure and the four together, 34.3%. In the later year, the first dimension accounted for only 63.7% of the east-west variation. It took six dimensions to account for an equivalent 83.3%. The first dimension accounted for only 13.4% of the network structure and the six, 41.1%. The variation attributable to longitude is more homogeneously distributed during the later years, indicating a breakdown of the regional grouping.

The regression analysis also revealed a change in north-south variation. In 1969, there was no clear relation between laditude and the network dimensions. The largest proportion of variance in laditude accounted for by a single dimension was 25.5% and it accounted for only 0.5% of network structure. The second largest was 18.8%. It accounted for only 0.6%. It took 11 dimensions of the network to account for 86.7% of the variance in laditude. The variation attributable to laditude was homogeneously distributed at that point in time. In 1980, it took only six dimensions to account for 88.5% of the variance in laditude. The first two accounted for 19.9% of the network. This indicates there was greater north-south differentiation during the later year than in the early one. Thus, while the network in the early years was characterized by east-west differentiation, the later years seem to be characterized by north-south differentiation. This suggests that the fundamental change in the network differentation occurred from coast-to-coast to frost belt-sunbelt.

NETWORK DENSITY

The air traffic network's density increased. As in the case of connectedness, the distance at which the least central node was clustered to the air traffic network decreased over time with a slight reversal in 1980 and 1981. The correlation between the trace (connectedness) and the distance at which the least central node was clustered was .952. This fact coupled with the breakdown of the regional clusters suggests that the distribution of air traffic in the United States has become more homogeneous.

This may be supported by examining the distribution of variance among the dimensions (eigenvalues) and the warp in the network's spaces over time. If the network became homogeneous, that is, the links became equally strong, then the space would become Euclidean (warp = 1.0) and the variance explained by the largest single dimension would decrease over time. In 1968, it was 330.8%. Percentages greater than 100 are due to the warp. In 1969, it was 43.0%, and by 1981, 14.0%. If all dimensions were equivalent, then each would account for 3.2% of the variance. The variance in the size of the eigenvalues also decreased. In 1968, the standard deviation was 77.73, in 1969, 9.37. After 1971, it stabilizes between 2.89 and 4.52.

The warp, likewise, suggests that the links become more homogeneous. In 1968, it was 3.97, in 1969, 1.25, and by 1978, 1.04. Warp, however, has a distinct reversal during the last two years. It rises to 1.09 (1980) and 1.10 (1981). This suggests that the distribution of air traffic is becoming less homogeneous. This is consistant with the findings that the network is less connected and dense. The value of the distance required to add the least central node to the network, the per cent variance of the first dimension, standard deviation of the eigenroots' per cent variance and the warp for each year are presented in table 4.

TABLE 4 ABOUT HERE

GRAPHIC RESRESENTATIONS OF NETWORK STRUCTURE

One of the advantages of metric multidimensional scaling is its ability to graphically represent the relationships among the nodes. Plots have not yet been presented because of the low rcentage of variance attributable to any two dimensions. However, since one goal of this paper is to demonstrate the utility of this method, three plots will be presented. They are 1970 (figure 3), 1975 (figure 4) and 1980 (figure 5). The two plotted dimensions (the two largest real dimensions) account for 37.7% of the variance in 1970, 26.5% in 1975 and 23.0% in 1980. The later percentages are smaller due to the increased density of the network. The 1970 plot is presented with longitude regressed on the first dimension. The 1980 plot includes the cluster analysis. There is considerable distortion in all three cases due to low percentages of explained variance.

These three plots were chosen to demonstrate the change in the network over time. The 1970 plot shows a midwest-eastern cluster with Ft. Lauderdale and Portland at the periphery. Also, there is a prominent east-west dimension. By examining the scale across all three spaces, it is clear that density and connectedness became greater. By 1975, the regional clustering is less prominent. Also, the distribution of the nodes within the space became more homogeneous. The 1980 plot shows further breakdown of the regional clustering an increase in homogeneity, density and interconnectedness. If these two dimensions accounted for all the variance in the network, centrality could be represented as a node's distance from the origin. In all three plots, Chicago, is closest to the origin. It is the most central node. This conclusion is consistant with the results of the cluster analysis in which Chicago has the shortest distance for inclusion.

FIGURES 3 TO 5 ABOUT HERE

STABILITY WITHIN THE NETWORK

Up to this point, this paper has concentrated only on changes in the network. The issue of stability has not been addressed. Stability may be inferred through an examination of the correlations of the nodes' locations on the dimensions at adjacent points in time. The mean correlation for the first dimension was .981. It was .986 for the second. This indicates that the network is highly stable.

Early in this paper an argument was made supporting those use of the dimensions with negative eigenroots in the analysis of social networks. The mean correlation among the largest (absolute value) of these dimensions across adjacent points in time was .67. For the last ten points it was .82 and .99 for the final four. This indicates that the variance on the imaginary eigenvector is not random error and that change in the size of this dimension and the arrangement of the nodes on it should be examined.

One reason for stability within the network is the fixed physical distances among the nodes. Physical proximity is one determinant of network structure (Olsson, 1965; Rogers & Kincaid, 1980). To determine how physical structure impacts on network structure, two multiple regressions were performed with the 14 sets of network coordinates as the independent variables and the 31 cities' laditude and longitude as the dependent variables.

The zero-order correlations with laditude were: .45 for dimension 3, .25 for dimension 4, .22 for dimension 31 and .18 for dimension 2, $R^2 = .35$ for those dimensions accounting for 2% or more of the variance in laditude. The correlations with longitude were .83 for dimension 1, and .38 for dimension 2, $R^2 = .84$. The multiple correlations were multiplied by the mean proportion of variance accounted for by the respective dimensions across the 14 data sets. 13 Since laditude and longitude are orthogonal, these two values were summed. The results indicate that approximently 18.3% of the variance in network structure may be accounted for by the physical relations among the nodes. Thus, one source of network stability may be attributed to physical proximity.

Another factor contributing to the stability within the network is the populations of the nodes. Population is a major determinant of the frequency of interaction among cities (Olsson, 1965; Hamblin, 1977). The correlation between the cities' populations in 1970 and 1980 is .99. Those nodes which moved greater than the average (Ft. Lauderdale, San Diego, Dallas and Tampa) all grew at least 24.7% between 1970 and 1980. Ft. Lauderdale, the node whose position changed the greatest, grew 68.2%. These nodes along others with comparable growth rates (Houston, Denver and Phoenix) all moved from the periphery to the center of the network. This suggests that population stability may contribute to the overall network stability and that change in the network may be due to population dynamics.

DETERMINANTS OF CHANGE IN NETWORK STRUCTURE

The network structure appears to change in an orderly manner which can be described by simple mathematical functions. However, those variables which facilitate or inhibit this change must be identified before an explanatory theory about social networks can be developed. A number of variables may be suggested to account for the change in the air traffic network. Among them are economic factors (GNP, GNP service, inflation as represented by producer and consumer prices, personal income, unemployment, automobile sales and fuel prices), the diffusion of new communications technologies, population growth and mobility, and changes within the airline industry (deregulation) and the network itself (the opening of the Dallas- Ft. Worth and Atlanta airports and the shifting operations to these nodes). In order to determine the impact of these factors annual data on these variables must be available on a national level or for each individual node. It was not available for this time period for the population or communication technologies. It was available for the economic factors and those internal to the industry.

The 14 annual values for the variables were correlated with the trace and 13 difference scores (time i+1 - time i) with the change in the trace and the overall change in the network between adjacent points in time. Due to the limited number of points in time, only bivariate linear relations were examined. They are presented in table 5. Worth noting is the consistantly high relation between the trace and all the variables with the exception of annual automobile sales. This is due to the variables' autocorrelation (Box & Jenkins, 1976). As a result, the linear trend was removed by taking first-order differences. Thus, the change in the variables were correlated with the change in the trace and the overall mean change between adjacent points in time.

The opening of the Dallas-Ft. Worth airport correlates -.89 with the mean change. Prior to its opening, there is a consistant high rate of change (See figure 2.). Afterward, the rate of change is lower. The network becomes stable. Both change in personal income and GNP correlate significantly with the overall rate of change. While none of these three variables have a significant relation with the change in the trace, they do have among the highest correlations. Although not significant, change in fuel prices has the highest correlation with the change in trace (.53) and a strong relation with the overall change (r = .31).

TABLE 5 ABOUT HERE

Descriptively, how do these variables relate to the critical points in time that have been identified through the analysis of the network structure? In 1974, there was a slight reversal in the trend towards greater connectedness. In 1974, Fuel prices had their first large increase due to the Arab embargo. The later may account for the trend reversal. Jet fuel prices caused an increase in ticket prices which may have resulted in fewer trips among the nodes and thus lower connectedness in this network. Between 1974 and 1975 the network stabilized. In 1974, the Dallas-Ft. Worth airport opened. Its use as a central hub seems to have stabilized air traffic.

1980 began a trend towards lower connectedness. That same year Atlanta's Hartsfield airport opened. One interpretation may be that there was no longer a need to travel through a more central node since Atlanta's traffic expanded as it became a regional hub. Thus, the network decentralized. 1980 also began an increase in unemployment and a smaller increase in personal income than in previous years. Thus, the change in the trend may be due to economic factors, the current recession.

The airline industry was deregulated in 1979. After that many flights and routes were abandoned because they were not profitable. The change in trend may be due to dereguation of the industry. It may have taken a year for its impact of show up in the state of the network. Determining the precise lag between deregulation and network characteristics would require more than the 13 changes scores available. Thus, this interpretation may only be suggested.

In summary, change in the air traffic network appears to be related to the changes in economic conditions and the changes within the airline industry.

DISCUSSION AND SUMMARY

This paper has demonstrated the utility of a variant of metric multidimensional scaling to describe changes in social networks over time. It uses a Riemannian manifold, rather than an Euclidean space, to represent the relative positions of the nodes. The results suggest, that change in America's air traffic network has been orderly and that it can be described precisely by simple mathematical functions that can be readily be interpreted when exogenous factors are examined. The trace of the spatial coordinates matrix, a negative indicator of network connectedness, decreased rapidly between 1968 and 1974, remained stable until 1980, when a reversal began. The only exception to this trend was 1974, when fuel prices rose greatly for the first time and the Dallas-Ft. Worth airport opened. Change in the trace may be described by two functions, an exponential decay, $R^2=.864$, and a polynomial to the second degree, $R^2=.752$. In the later case, the quadratic term was significant, indicating true change in the direction of connectedness during the last two years. This reversal may be attributed to a number of factors. Among them, the economic recession during that period which was characterized by high unemployment and slow growth in the GNP and the opening of Atlanta's airport. Consistant with this pattern was one independently obtained through a cluster analysis of the frequencies of interaction among the nodes (r=.95).

The rate of change in network structure may be described by two epoches. The first epoch (1968-1974) was characterized by a high rate of change, while the second (1974-1981) was relatively stable. During the first epoch, the network was differentiated by an east-west dimension. The second was differentiated by a north-south dimension. These changes may be attributable to the opening of the Dallas-Ft. Worth airport and the increased use of its facilities as a central hub for air traffic.

The network also exhibited a high degree of stability. The mean correlation between the first two dimensions at adjacent points in time were .981 and .986. For the largest imaginary dimension, it was .67 and .82 for the last ten points in time. This demonstrates the necessity of using a Riemannian manifold to describe social networks.

Future research with this data set is planned using alternative rotational algorithms rather than the ordinary least squares procedures reported here. During the period examined in this paper the critical event in the air traffic network's history appeared to be the opening of the Dallas-Ft. Worth airport. To examine its impact on the network and the node's changing position within the network, a rotational scheme which hold the other cities in the network constant relative to a free moving Dallas should be applied. This analysis has not been performed due to the high cost of computing and restrictions on computer limitations at SUNY-Buffalo. Plans have been made to perform these analyses at SUNY-Albany with the assistance of Joseph Woelfel and Richard Holmes.

There is a family of models developed by geographers to describe the frequency of interaction among collectives such as cities. They are known as Gravity Models (Hamblin, 1977; Olsson, 1965). Originally proposed by Zipf (1949), they predict the frequency of interaction as a function of the product of two nodes' population divided by the distance between them raised to some power. It is presented a equation 7.

$$i = c(p_1 p_2)/d^{exp} \quad 7$$

Test of this model have resulted in explained variances in the range, .592-.774 (Howrey, 1969; Long, 1970). The exponent ranges from .14 to less than 3.0 (Olsson, 1965), depending on the type of network examined. Tests with the 1980 data have resulted in explained variances between .46 and .58. 14 The exponent ranged from .11 to .40, depending on the restrictions placed on the model. These results suggest another analysis, a rotation in which the nodes are weighted by their population. While Woelfel, et al.(1975, 1979) describe the algorithm for a weighted rotational procedure, the software necessary to perform this analysis is not operational.

Rogers and Kincaid (1980) propose two determinants of network structure, physical proximity and homophily or similarity among the nodes. This paper discusses only the former. To evaluate the latter causal mechanism, data on the similarity among the nodes must be collected. This paper has indicated that economic variables may predict change in the network. Thus, a logical starting point would be to gather economic data on the cities. Other variables such as ethnic makeup, mobility patterns, cultural, educational and political factors could be examined. From these data, an index of similarity among the nodes may be developed. This would allow the construction of a sociomatrix based on the nodes' structural equivalence. Structural equivalence occures when two nodes occupy equivalent positions in a network due to the pattern of relations (Burt, 1982). This assumes that if two nodes are similar, their position in the network should be equivalent despite not necessarily being in direct communication contact. For example, both the Florida nodes and Phoenix and San Diego have equivalent positions based upon tourism. However, they have little direct interaction. Thus, a structural equivalence approach may reveal many insights into the changing nature of the air traffic network. A matrix of structural equivalence could be directly compared to the sociomatrices of air traffic in much the same manner as any two sociomatrices. In this way, the extent of influence of homophily on the structure of the network can be determined.

This paper has focused upon the network among American cities which resulted from the frequencies of their interaction via airplane. These results could be applied to those innovations which are diffused primarily by this network and to the communication-transportation tradeoff issue (Barnett, 1979). The reversal in trend in connectedness could be the beginning of the discontinuance of the use of air travel which coincides with the development of alternative communications technologies which may make travel unnecessary.

Future research should apply the methods described here to other communication networks. One such application would be to the communication networks of formal organizations as suggested by Goldhaber, et al. (1983). In that case, the unit of analysis (nodes) were branches or functions of an organization. At that level, the system was highly interconnected and the information on the strength of link among the nodes could have been treated as distances in matrix S. Such research on a formal organization is currently underway. These procedures are directly applicable to Computer-Mediated Communication Systems (Danowski, 1982; Rice, 1982). In those cases, the data are error-free, time sensitive and may include a quantitative measure of interaction which could be converted to communication distances.

There are a number of drawbacks with this method that should be discussed. One is that it can only be applied to those systems that are completely interconnected. The reason is that in those cases where the frequency of interaction between two nodes is zero, the reciprocal becomes infinity. To apply the proposed procedures to those situations, rules must be established to deal with links with a value of zero. One simple solution is to assign an arbitrarily large value. In that case, the value of the trace, the indicator of connectedness discussed in this paper, would also be arbitrary and no inferences about it could be drawn.

Another problem with these procedures is that their application is to relatively small networks. The Galileo (tm) software is limited to 40 nodes. Also, it is limited to 40 points in time. This limits the potential application of classical time-series analysis (Box & Jenkins, 1976; Jenkins & Watts, 1968) for analyzing the periodicity of changing network parameters as proposed by Barnett and Woelfel (1979). Although there are procedures to work around the points in time limitation. It should be noted, however, that the software may not be the ultimate limitation. As pointed out above, the time required to perform these operations may exceed the limitations of university computers because the algorithms are iterative solutions.

A final complication concerns those cases where the research is interested in directional or nonreciprocated links. In that case, $s_{ij} \neq s_{ji}$. Currently, Galileo (tm) has no procedures to directly analyze assymetrical matrices. Although plans to calculate both the left and right handed eigenroots have been discussed (Woelfel, personal correspondence), a simpler method, currently available, would be to create two matrices S(send) and S(receive), and then compare as if they were separate points in time. In this way, one can determine the differences between the incoming and outgoing links.

In summary, this paper has proposed that a variant of metric multidimensional scaling, the Galileo System (tm) be used to analyze over-time changes in social networks. The paper discussed the theoretical necessity of using these procedures and certain methodological problems associated with this approach. Next, it examined the air traffic network among 31 major American cities over the 14 year period, 1968-1981. It demonstrated how the proposed method can provide insights into the activity within the network and the impact of exogenous factors upon the structure of the network.

NOTES

1. The authors would like to thank the following people for their input into this research. They include, Bruce Alesse, Mary Russell, Edward L. Fink, Shirley Masse, Joseph Woelfel, James A. Danowski, Gary Siegel, Gerald Goldhaber, Charles Petrie, Kenneth Day, Richard Holmes, James Gillham, D. Lawrence Kincaid, Diana Miller, Nicholas Illich, and Mark Feldman. This research was made possible in part by a grant from the Faculty of the Social Sciences at the State University of New York at Buffalo.

2. This assumes that the relationship among the nodes is relatively stable over time. In the case where the relations among the nodes are dynamic, this process of change can be precisely described by gathering matrix S at a number of points in time and calculating the changes between S_{t_1} and S_{t_2}.

3. This matrix is typically "double centered". That is the grand mean of the distance matrix is subtracted from all values, giving the matrix a grand mean of zero. Thus, the matrix is centered about zero. As a result, the centrality of any individual node, i, may be found on the diagonal of the scalar products matrix (S S). the value on the diagonal, sii, represents the distance of node i from the center of the network, such that the greater the value of sii, the less central the node is to the network.

4. Matrix S in the example would produce a two-dimensional space because any matrix of N points may be described without the loss of any information by a manifold of N-1 dimensions. For example, any two points may be precisely described by a line. Three points may be described by a plane (two dimensions) and four points by a cube (three dimensions). N points by a space of N-1 dimensions.

5. In this example, one assumed that there was some communication between b and c. If all communiction between b and c occurred through a, one must identify a maximum value for the frequency (discrepancy) of the bc link. Without such a value, be would be equal and the calculations could not be performed.

6. These data may be obtained from,
The Economics and Finance Council
Air Transport Association of America
1709 New York Avenue, N.W.
Washington, D.C. 20006

7. Not included were San Antonio and Sacramento. Oakland, San Jose and San Francisco were combined into a single node. SCSA Ft. Lauderdale was treated separately from Miami because of the great frequency of air traffic into its airport.

8. The data were extracted from table 11 "Domestic Origin-Destination Survey of Airline Passenger Traffic."

9. When equation 1 was applied to the San Diego-Ft. Lauderdale-New York triad, cos θ = 16,011.24, indicating a violation of the rule of triangular inequalities. However, when the same equation was applied to the triad after the inverse transformation was performed, cos θ = .56, an Euclidean solution.

10. Root mean squares ($\sqrt{\Sigma x^2/n}$) of the changes were examined rather than simple means because many of the distances moved were negative. That is the differences occurred on those dimensions with negative eigenroots.

11. BMDP's P2M hierarchical cluster analysis program was used to identify subgroups with the network (Dixon, 1981). It forms clusters of cases based on a measure of association or similarity between the cases. Here, the distances separating the nodes (cases) were the measure of association. Initially, each mode is considered a separate group or cluster. Cases and/or clusters of cases are joined in a stepwise process until all cases are combined into one cluster. Hence, the label, Hierarchical Cluster Analysis. The algorithm begins by computing a matrix of distances between each pair of cases (nodes). In this case, a distance matrix, S, was provided. Then, disks are placed about each point and their radii expanded until the intersection of two disks or until one covers another point. Their distance is the length of the radius. A matrix of these pseudo-distances is then stored. The two cases with the smallest distances are joined first. The process is repeated. During the amalgamating process, (a case with another case, a case with a cluster or two clusters), distances are read from the initial distance matrix. The results include a distance or density measure indicating the distance at which the n-th case was clustered and a tree diagram which reveals the sequence of cluster formation. The distance measure may be used as a measure of centrality. The more central nodes are clustered first and have a lower distance value. The more peripherial ones are added later and have a greater distance value.

12. To check the validity of these procedures, the physical distances among the cities were transformed into spatial coordinates and then the coordinate values regressed on laditude and longitude.

Dimension one's correlation with longitude was .993 (r^2=.988). Dimension two's correlation with laditude was .982 (r^2=.964). These dimensions accounted for 81% and 19% of the variance in the distances among the cities respectively. Together, they account for 98.7% of the variance in the distances among the cities. This was determined by summing the proportions of explained variance multiplied by the correlation squared. The remaining 1.3% may be attributable to measurement and rounding error and the curvature of the earth. Thus, regressing matrix S 's coordinate values upon laditude and longitude can be used to determine the impact of physical location on network structure.

13. The decision to combine the 14 sets of coordinates was based upon the high correlations among the respective dimensions at adjacent points in time. Combining the dimensions results in a conservative estimate of the variance attributed to physical proximity. If two dimensions are not identical random error is entered into the analysis and the estimate of goodness-of-fit are lowered. The reason for this is that dimension n at time k may not be dimension n at time k+1 due to change in the network which changes the order in which the dimensions are extracted.

14. Only 26 nodes (325 pairs) were included in testing the gravity model. Excluded were, Columbus, Ft. Lauderdale, Milwaukee, San Diego and Tampa. The specific form of the model tested was,

$$\ln(i) = \ln(c) + m \ln(p\ ) + n \ln(p\ ) - r \ln(d).$$

The coefficients were, $c = 316$, $m=.88$, $n=.83$, $r=-.11$, $R^2=.463$. Taking the antilogs, the predictive model becomes,

$$\hat{y} = 316\ p_1^{.88}\, p_2^{.83} / d^{.11}.$$

Predicting interaction from the nodes' population alone, produced an R^2=.458. This indicates that population alone is the best predictor of interaction by air. An examination of of the residuals revealed that the greatest deviations (greater than 3.0 standard deviations) occurred between nodes less than 125 miles apart. Travel between them is most efficient using other modes of transportation (automobile, bus or train). For example, 4 of the 5 pairs with the greatest residuals are on Amtrack's New York to Washington corridor (New York, Philadelphia, Baltimore and Washington). Here travel is most efficient via rail. As a result of this analysis, the model was tested with only those links whose distance was greater than 125 miles. Seven of 325 pairs were eliminated. The coefficients were, $c = 242$, $m=.93$, $n=.85$, $r=-.397$, $R^2=.582$. R^2 with the nodes' population only was .522. For the reasons why c is only an approximate value see Hamblin (1974).

REFERENCES

Alba, R.D., "A graph-theoretic definition of a sociometric clique." Journal of Mathematical Sociology, 1973,3,1-13.

Barnett, G.A., "Spatial modelling of social networks with applications to the diffusion process." Paper presented to the International Communication Association, Philadelphia, April, 1979.

Barnett, G.A., "Inertial mass as a function of frequency of occurrence." Paper presented to the International Communication Association, Acapulco, May, 1980.

Barnett, G.A., "Seasonality in television viewing: a mathematical model." Paper presented to the International Communication Association, Boston, May, 1982a.

Barnett, G.A., "A modest proposal: standards for the measurement of time and distance in communication research." Paper presented to the International Communication Association, Boston, May, 1982b.

Barnett, G.A. and D.L. Kincaid, "A mathematical theory of cultural convergence." In W. Gudykunst (ed.) International-Intercultural Communication Annual VII, Beverly Hills, CA: Sage, 1983.

Barnett, G.A. and J. Woelfel, "On the dimensionality of psychological processes." Quality and Quantity, 1979,13,215-232.

Bernard, H.R. and P.D. Killworth, "On the social structure of an ocean-going research vessel and other important things." Social Science Research, 1973,2,145-184.

Bernard, H.R. and P.D. Killworth, "Informant accuracy in social networks." Human Communication Research, 1977,4,3-18.

Breiger, R.L., S.A. Boorman and P. Arabie, "An algorithm for clustering relational data with applications to social networks and comparison with multidimensional scaling." Journal of Mathematical Sociology, 1975,12,328-383.

Box, G.E.P. and G.M. Jenkins, Time Series Analysis: Forcasting and Control. San Francisco: Holden-Day, 1976.

Burt, R.S., "Stratification and prestige among experts in methodological and mathematical sociology circa 1975." Social Networks, 1978,1,105-158.

Burt, R.S., "Cohesion versus structural equivalence as a basis for network subgroups." In R.S. Burt and M.J. Minor (Eds.) Applied Network Analysis: A Methodological Introduction. Beverly Hills, CA Sage, 1982.

Burt, R.S. and N. Lin, "Network time series from archival records." In D. Heise (Ed.) Sociological Methodology-1977. San Francisco: Jossey-Bass, 1977.

Danowski, J.A., "Uniformity of group attitude-belief and connectivity of organizational communication networks for production, innovation and maintainance." Human Communication Research, 1980,6,299-308.

Danowski, J.A., "Computer-mediated communication: a network-based content analysis using a CBBS conference." In M. Burgoon (Ed.) Communication Yearbook-6. Beverly Hills,CA: Sage, 1982.

Dixon, W.J., BMDP Statistical Software 1981. Berkeley University of California Press, 1981.

Festinger, L. "The analysis of sociograms using matrix algerba." Human Relations, 1949,2,153-158.

Forsyth, E. and L. Katz, "A matrix approach to the analysis of sociometric data." Sociometry, 1946,9,340-347.

Freeman, S.C. and L.C. Freeman, "The networkers network: a study of the impact of a new communications medium on sociometric structure." Paper presented to Seminar on Communication Network Analysis, East-West Communication Institute, Honolulu, January, 1979.

Gillham, J.R. and J. Woelfel, "The Galileo system of measurement: preliminary evidence for precision, stability, and equivalence to traditional measures." Human Communication Research, 1977,3,222-234.

Goldhaber, G.M., G.A. Barnett, J.J. Mitchell and R.W. Bales, "Network analysis from a functional perspective." Paper presented to the International Communication Association, Dallas, May, 1983.

Goldstein, K.M., S. Blackman and D.J. Collins, "Relationship between sociometric and multidimensional scaling measures." Perceptual and Motor Skills, 1966,23,639-643.

Hamblin, R.L., "Social attitudes: magnitude measurement and theory." In H.M. Blalock (Ed.), Measurement in the Social Sciences Theories and Strategies, Chicago: Aldine, 1974.

Hamblin, R.L., "Behavior and reinforcement: a generalization of the matching law." In R.L. Hamblin and J.H. Kunkel (Eds.) Behavioral Theory in Sociology. New Brunswick, NJ: Transaction Books, 1977.

Howrey, E.P., "On the choice of forecasting models for air travel." Journal of Regional Science, 1969,9,215-224.

Jenkins, G.M. and D.G. Watts, Spectral Analysis and its Application. San Francisco: Holden-Day, 1968.

Jones, L.E. and F.W. Young, "Structure of a social environment: longitudinal individual differences of an intact group." Journal of Personality and Social Psychology, 1972,24,108-121.

Krumhansl, C.L., "Concerning the application of geometric models to similarity data: the interrelationship between similarity and spatial density." Psychological Review, 1978,85,445-463.

Kruskal, J.B., "Multidimensional scaling by optimizing goodness of fit to a non-metric hypothesis." Psychometrika, 1964a,29,1-28.

Kruskal, J.B., "Non-metric multidimensional scaling: a numerical method." Psychometrika, 1964b,29,115-159.

Lankford, P.M., "Comparative analysis of clique identification methods." Sociometry, 1974,37,287-305.

Lindman, H. and T. Caell, "Constant curvative Riemannian scaling." Journal of Mathematical Psychology, 1978,17,89-109.

Lloyd-Kolkin, D., "Communication network analysis of an educational dissemination system: the research and development exchange." Ph.D. Thesis, Stanford, CA: Stanford University, 1979.

Long, W.H., "The economics of air travel gravity models." Journal of Regional Science, 1979,10,353-363.

Messick, S. and R. Abelson, "The additive constant problem in multidimensional scaling." Psychometrika, 1956,21,1-15.

Morett-Lopez, F.J., "Communication networks among marginals in a Mexican city." Ph.D. Thesis, Stanford, CA: Stanford University, 1979.

Olsson, G., Distance and Human Interaction. Philadelphia Regional Science Research Institute, 1965.

Pieszko, H., "Multidimensional scaling in Riemannian space." Journal of Mathematical Psychology, 1975,12,449-477.

Pitts, F.R., "Recent trends in social network analysis: a bibliography." Paper presented at Seminar on Communication Network Analysis, East-West Communication Institute, Honlulu, January, 1979.

Rice, R., "Communication networking in computer-conferencing systems: a longitudinal study of groups, roles and system structure." In M. Burgoon (Ed.) Communication Yearkbook-6. Beverly Hills, CA: Sage, 1982.

Richards, W.D. "Network analysis in large complex systems." Paper presented to the International Communiction Association, New Orleans, April, 1974.

Roberts, K.H. and C.A. O'Reilly III, "Organizations as communication structures: an empirical approach." Human

Communication Research, 1978,4,283-293.

Rogers, E.M. and D.L. Kincaid, Communication Networks: Towards a New Paradigm for Research. New York: Free Press, 1980.

Rogers, E.M. and R. Agarwala-Rogers, Communication in Organizations. New York: Free Press, 1976.

Romney, A.K. and K. Faust, "Predicting the structure of a communications network from recallled data." Social Networks, 1982,4,285-304.

Schwartz, M., Computer-communication Network Design and Analysis. Englewood Cliffs, NJ: Prentice-Hall, 1977.

Shepard, R.N., "The analysis of proximities: multidimensional scaling with an unknown distance function I. Psychometrika, 1962a,27,125-140.

Shepard, R.N., "The analysis of proximities: multidimensional scaling with an unknown distance function II. Psychometrika, 1962b,27,219-246.

Stern, R.N., "The development of an interorganizational control network: the case of intercollegiate atheltics." Administrative Science Quarterly, 1979,5,845-867.

Stevens, S.S., "Mathematics, measurement and psychophysics." In S.S. Stevens (Ed.) Handbook of Experimental Psychology. New York Wiley, 1951.

Togerson, W.S., Theory and Methods of Scaling. New York Wiley, 1958.

Tversky, A., "Features of similarity." Psychological Review, 1977,84,327-352.

Van de Geer, J.P., Introduction to Multivariate Analysis for the Social Sciences. San Francisco: Freeman, 1971.

Woelfel, J. and G.A. Barnett, "Multidimensional scaling in Riemann space." Quality and Quantity, 1982,16,469-491.

Woelfel, J., M.J. Cody, J.R. Gillham and R. Holmes, "Basic premises of multidimensional attitude change theory." Human Communication Research, 1980,6,153-168.

Woelfel, J and E.L. Fink, The Galileo System: A Theory of Social Measurement and Its Application. New York: Academic, 1980.

Woelfel, J., E.L. Fink, K.B. Serota, G.A. Barnett, R. Holmes, M. Cody, J. Saltiel, J. Marlier and J.R. Gillham, Galileo-A Program for Metric Multidimensional Scaling. Honolulu: East-West Communication Institute, 1976.

Woelfel, J., R. Holmes and D.L. Kincaid, "Rotation to congruence for general Riemann surfaces under theoretical constraints." Paper presented to the International Communication Association, Philadelphia, April, 1979.

Woelfel, J., J. Saltiel, R. McPhee, J.E. Danes, M.J. Cody, G.A. Barnett and K.B. Serota, " Orthogonal rotation to theoretical criteria: comparison of multidimensional spaces." Paper presented to the Mathematical Psychology Association, West Lafayette, IN, August, 1975.

Zipf, G.K., Human Behavior and the Principle of Least Effort. New York: Hafner, 1949.

TABLE 1
SELECTED CITIES (SMSA) AND POPULATIONS (1980)

1	Atlanta	2,010,000
2	Baltimore	2,166,000
3	Boston	3,443,000
4	Buffalo	1,241,000
5	Chicago	7,697,000
6	Cincinnati	1,651,000
7	Cleveland	2,830,000
8	Columbus	1,089,000
9	Dallas-Fort Worth	2,964,000
10	Denver	1,615,000
11	Detroit	4,606,000
12	Fort Lauderdale-Hollywood (SCSA)	1,006,000
13	Houston	3,086,000
14	Indianapolis	1,162,000
15	Kansas City	1,322,000
16	Los Angeles	11,439,000
17	Miami (without Fort Lauderdale)	1,573,000
18	Milwaukee	1,566,000
19	Minneapolis-St. Paul	2,109,000
20	New Orleans	1,184,000
21	New York City	16,065,000
22	Philadelphia	5,530,000
23	Phoenix	1,612,000
24	Pittsburg	2,261,000
25	Portland	1,236,000
26	San Diego	1,860,000
27	San Francisco-Oakland-San Jose	4,845,000
28	Seattle	2,084,000
29	St. Louis	2,345,000
30	Tampa-St. Petersburg	1,550,000
31	Washington	3,045,000
	TOTAL POPULATION OF CITIES	94,092,000
	TOTAL POPULATION OF U.S. 1980	225,479,000

Sample contains 43.5% of Total

TABLE 2

RESULTANT COEFFICIENTS, OBSERVED TRACES, PREDICTED VALUES AND RESIDUALS

T	OBSERVED TRACE*	EXPONENTIAL DECAY PREDICATED	RESIDUAL	POLYNOMIAL PREDICTED	RESIDUAL
1	67,730	66,729	1001	59,404	8325
2	50,699	53,931	-3232	56,105	-5407
3	48,657	48,288	369	53,145	-4489
4	48,331	45,800	2531	50,524	-2193
5	46,718	44,703	2015	48,241	1523
6	45,817	44,220	1597	46,296	-479
7	47,987	44,007	3980	44,690	3297
8	45,157	43,913	1244	43,422	1735
9	44,323	43,871	452	42,492	1831
10	43,421	43,853	-432	41,901	1520
11	41,587	43,845	-2258	41,648	-61
12	40,507	43,841	-3334	41,733	-1226
13	41,567	43,840	-2273	42,156	-590
14	42,178	43,839	-1661	42,919	-741

EXPONENTIAL DECAY: Y = a + b (exp * kt)

a = 43,838
b = 51,919
k = -.82

R square = .864

POLYNOMIAL: Y = a + b(1)t + b(2)t **2

a = 63,041
b(1) = -3,806 F = 13.18 p = .004
b(2) = 169 F = 6.19 p = .03

R square =.752

* = thousands

TABLE 3

OVERALL CHANGE BETWEEN ADJACENT POINTS IN TIME

YEARS	DIFFERENCE
1968-1969	585.8
1969-1970	2,278.5
1970-1971	2,316.8
1971-1972	2,362.9
1972-1973	2,370.2
1973-1974	2,350.6
1974-1975	383.2
1975-1976	259.8
1976-1977	216.1
1977-1978	253.5
1978-1979	302.6
1979-1980	248.0
1980-1981	241.9

TABLE 4

HOMOGENEITY OF THE NETWORK OVER TIME

YEAR	MAXIMUM DISTANCE FOR CLUSTER	PER CENT VARIANCE FIRST DIMENSION	STANDARD DEVIATION OF EIGENROOTS	WARP
1968	20,407	330.8	77.73	3.97
1969	6,529	43.0	9.37	1.25
1970	4,380	28.4	5.22	1.13
1971	3,888	25.5	4.64	1.11
1972	3,244	21.9	3.98	1.07
1973	3,045	20.7	3.97	1.07
1974	4,472	23.8	4.52	1.09
1975	2,782	17.8	3.33	1.06
1976	2,768	16.6	3.23	1.06
1977	2,733	16.3	3.10	1.04
1978	2,723	15.0	2.97	1.05
1979	2,665	14.6	2.89	1.05
1980	2,942	14.6	3.22	1.09
1981	3,024	14.4	3.30	1.11

TABLE 5

CORRELATION OF EXTERNAL VARIABLES WITH CHANGE IN NETWORK STRUCTURE

	TRACE	CHANGE IN TRACE	MEAN OVERALL CHANGE
DEREGULATION	-.51	.26	-.46
ATLANTA'S OPENING	-.31	.26	-.36
PERSONAL INCOME	-.69*	.38	-.56*
DALLAS'S OPENING	-.60*	.27	-.89*
GNP	-.68*	.41	-.67*
GNP SERVICE	-.61*	.25	-.38
PRODUCER PRICES	.75*	-.24	-.29
CONSUMER PRICES	.78*	-.02	.12
UNEMPLOYMENT	-.65*	.10	.03
AUTOMOBILE SALES	.17	.00	.00
FUEL PRICES	-.80*	.53	.31

* SIGNIFICANT AT .05 LEVEL

Figure 1

Trace Over Time

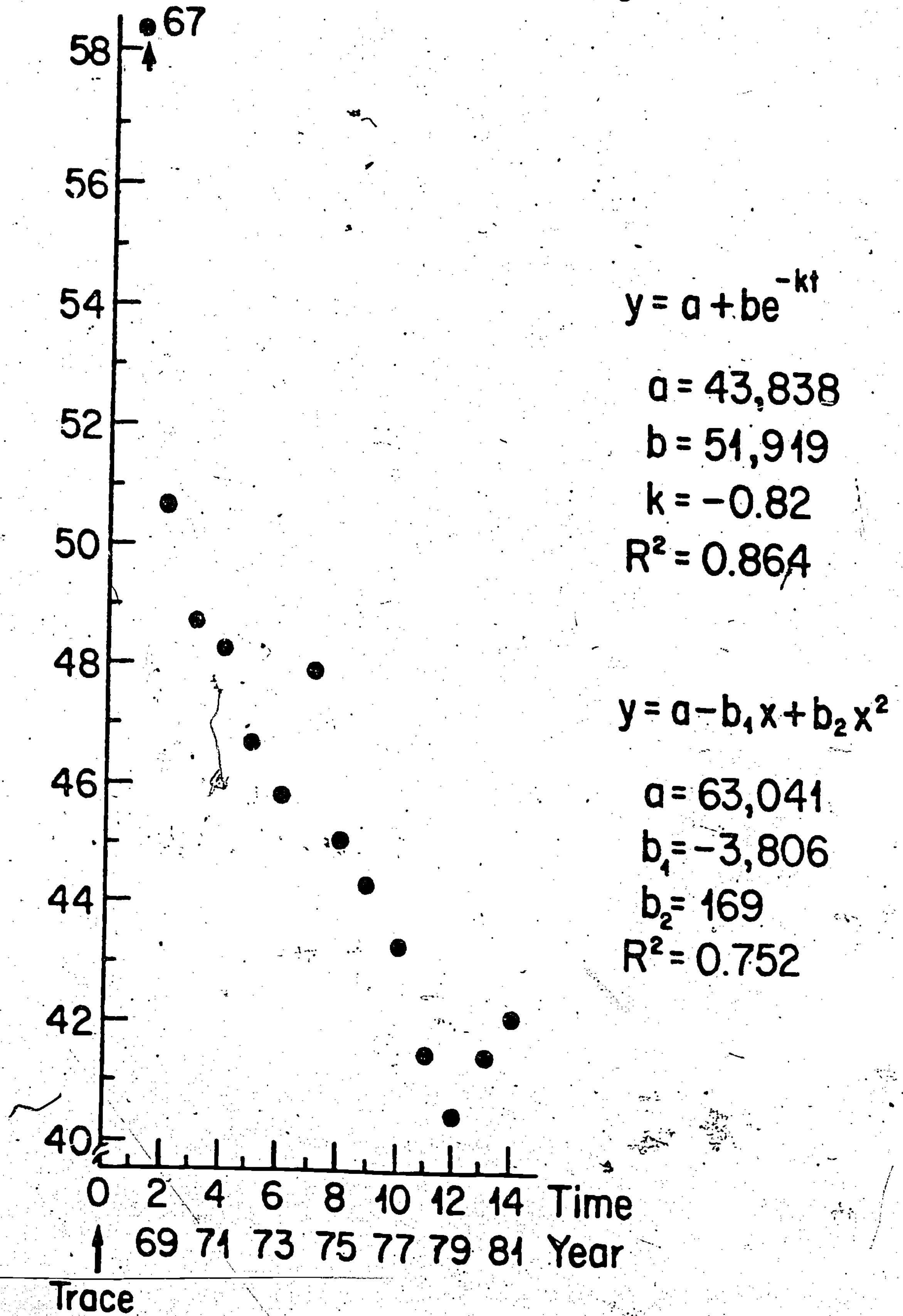


Figure 2

Overall Change Over Time

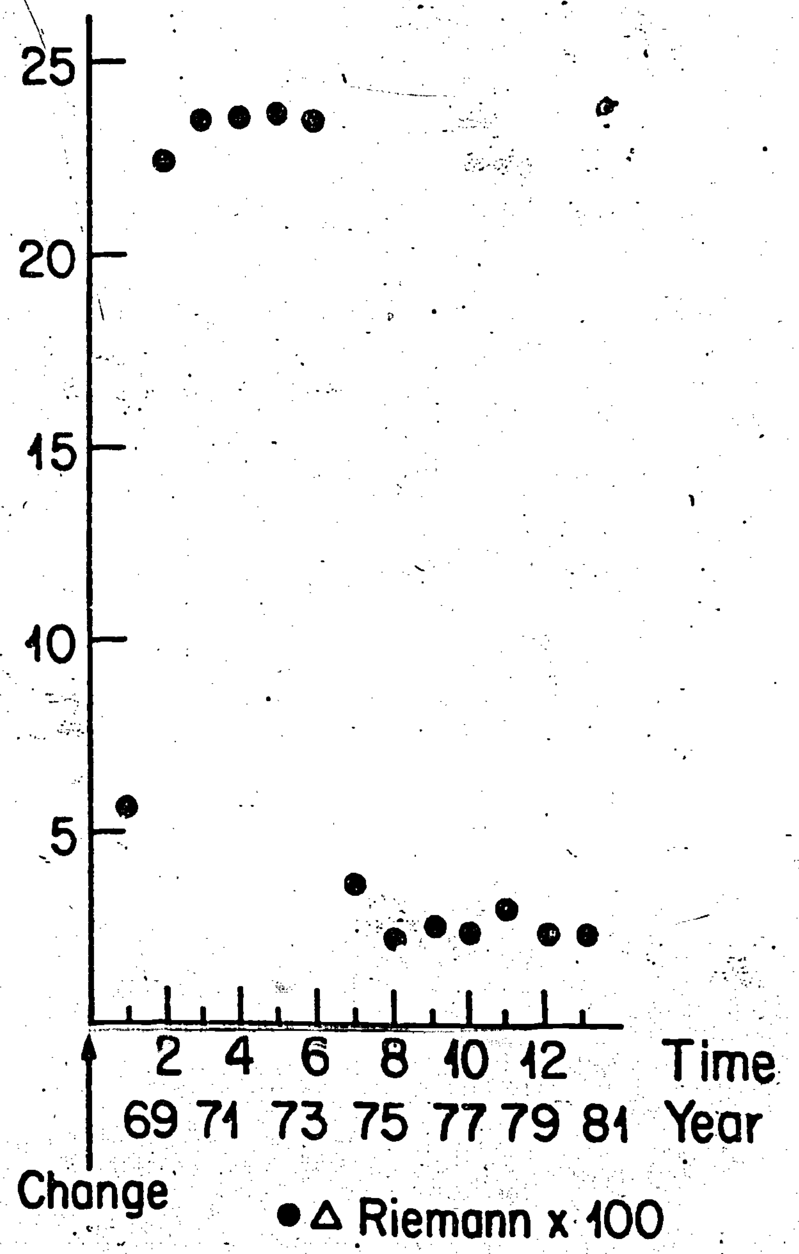
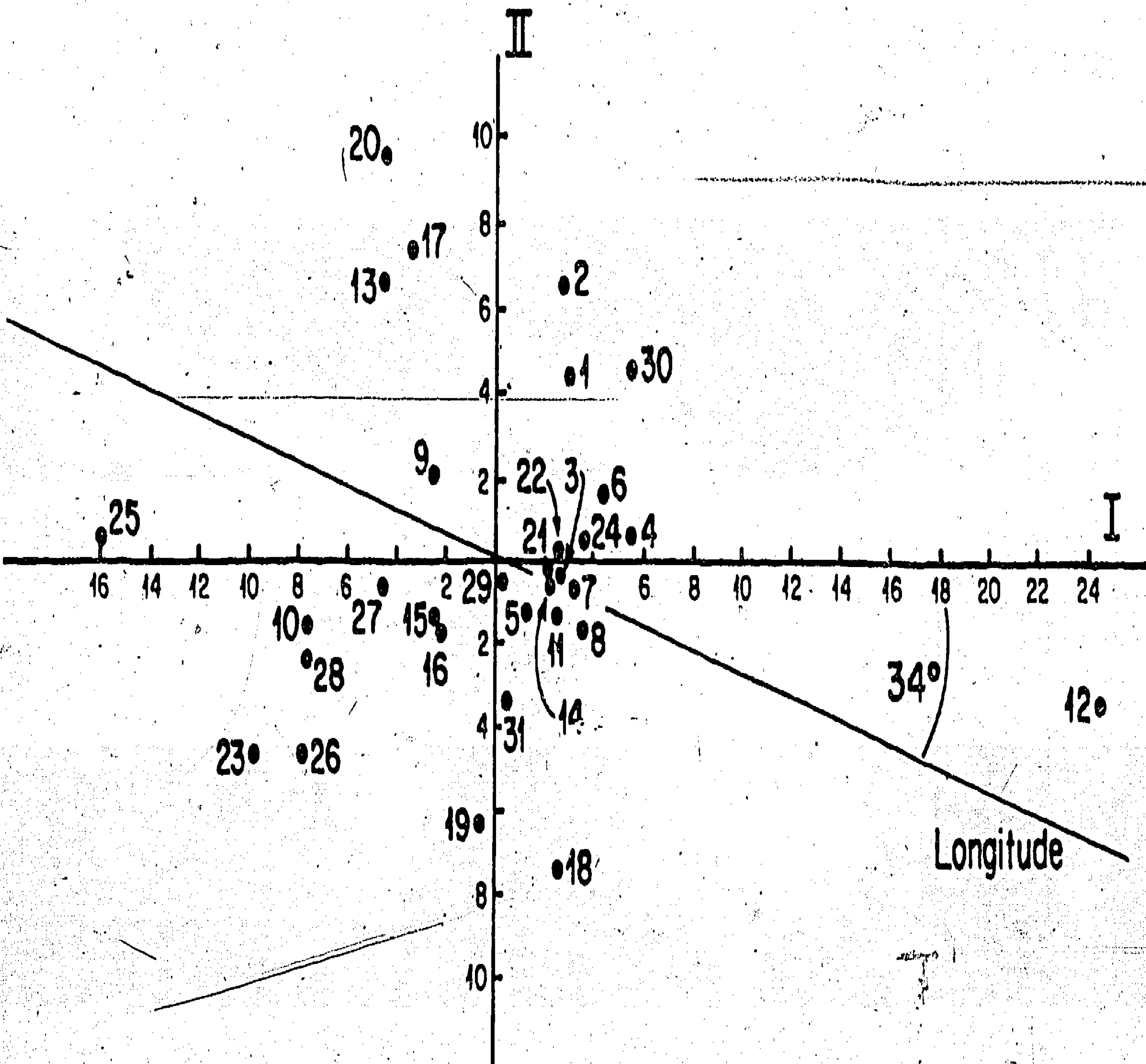


Figure 3

1970 Dimensions 1 and 2 with Regression of Longitude



41

Figure 4

1975 Dimensions 1 and 2

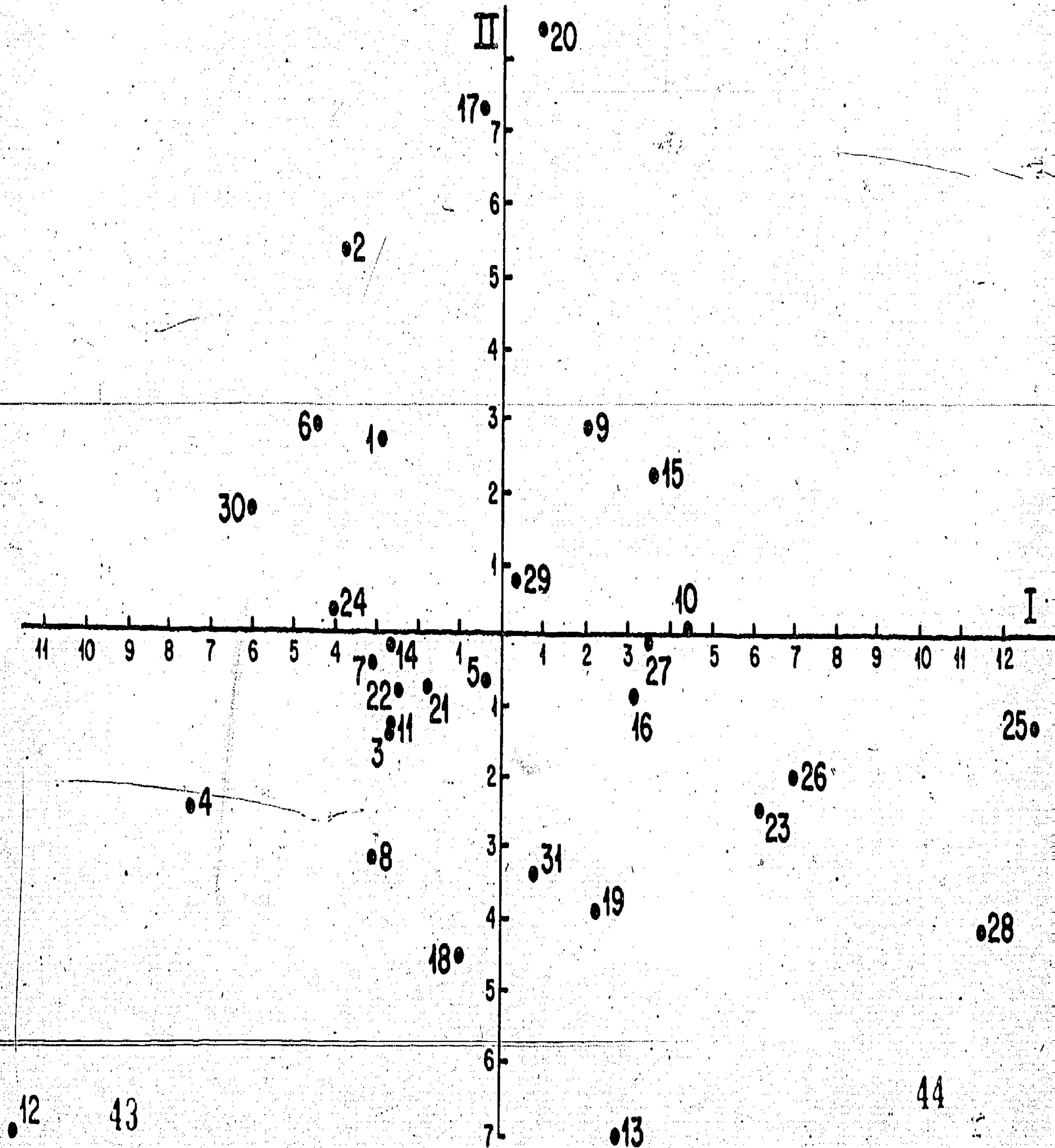
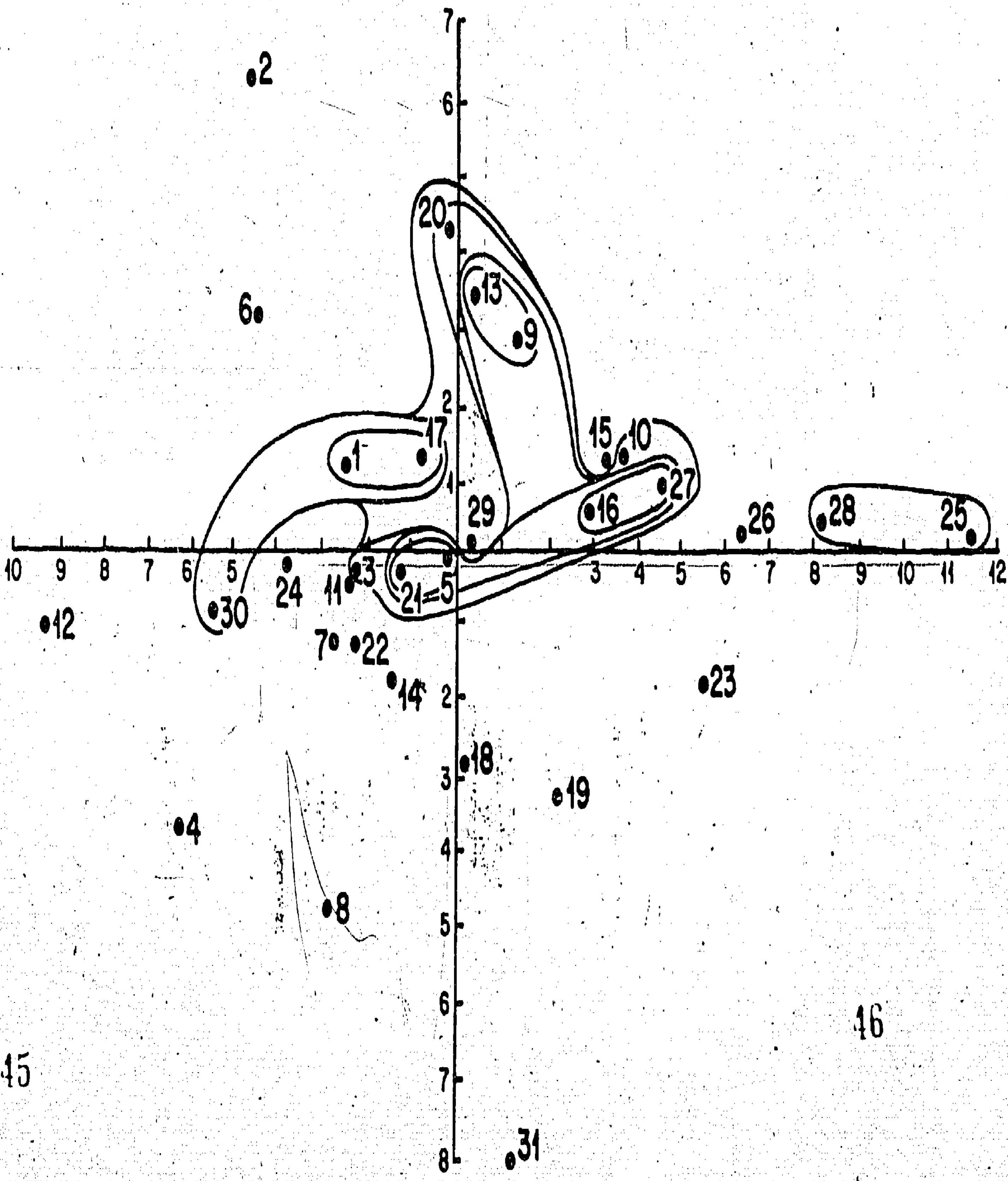


Figure 5

1980 Dimensions 1 and 2 with Cluster Analysis



45

46