Social Networks 7 (1985) 287-322
North-Holland

LONGITUDINAL NON-EUCLIDEAN NETWORKS: APPLYING GALILEO

George A. BARNETT *
State University of New York at Buffalo

Ronald E. RICE **
University of Southern California

This article discusses the theoretical utility of using a non-Euclidean spatial manifold when describing social networks. It proposes that a variant of metric MDS, the Galileo System, can be particularly useful in analyzing social networks and their changes over time, partially because it does not impose Euclidean assumptions on the data. Two sets of longitudinal network data are examined with Galileo. One is the American air traffic network from 1968-81. The other is ten groups engaged in a computer conference over a 24 month period. In both cases, the results indicate that a Riemannian spatial manifold is required to describe the network structure. Consistent theoretically valid results based upon the non-Euclidean components of spatial manifold are obtained. Further, they could be readily explained by exogenous factors. The implications of these results for network analysis are discussed.

1. Introduction

Multidimensional scaling (MDS) is frequently used to describe social networks (Goldstein *et al.* 1966; Jones and Young 1972; Lankford 1974; Breiger *et al.* 1975; Gillham and Woelfel 1977; Freeman and Freeman 1979; Romney and Faust 1982; Barnett 1979, 1984; Rice and Barnett 1985). It is only one of many network analysis methods currently used (Burt and Minor 1983; Knoke and Kuklinski 1982; Moreno 1960; Rice and Richard 1985; Rogers and Kincaid 1981).

* Department of Communication, State University of New York at Buffalo, Buffalo, NY 14260, U.S.A.
** Annenberg School of Communications, University of Southern California, Los Angeles, CA 90089, U.S.A.

0378-8733/85/$3.30 © 1985, Elsevier Science Publishers B.V. (North-Holland)

Despite its great use, less than satisfactory results have been reported (Lankford 1974; Breiger *et al.* 1975). One reason for this may be the assumption that networks are best explained by an Euclidean spatial manifold. This article will argue that social network data need not be Euclidean, but may be Riemannian. It proposes the use of a variant of metric MDS, the Galileo System (Woelfel and Fink 1980) which does not make Euclidean assumptions. Finally, it demonstrates Galileo's utility for describing longitudinal social networks, using two sets of data: Barnett's (1984) data on the frequency of air traffic among 31 American cities and Rice's (1982) data on the communication among groups engaged in a computer conference.

2. Theory

Social networks may be conceptualized and protrayed in a variety of ways. One way is to portray a network as a $N \times N$ matrix S, where N equals the number of nodes or interacting units in the network. The value in each cell (s_{ij}) is some measured attribute of the relationship or link between nodes i and j. This value may be the frequency of communication, often weighted by perceived importance. Or, the value may indicate distance between nodes. Distance may be a direct (perhaps perceived) measure, or the result of computations.

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

Consider the matrix S above, where s_{ij} is a measure of network distance. The diagonal contains zeros because the distance between any node and itself is zero by definition. If matrix S were converted to Cartesian coordinates (through MDS) by finding the eigenvector of its scalar products matrix ($S^T S = B$), one would find that the eigenroots or characteristics roots of B would include one negative root. The reason is that the triangle formed from the links of the *abc* triad cannot exist in a two-dimensional Euclidean space. [1] The *abc* triangle has two short legs (ab and ac) and one long one (bc). As a result, the sum of the triangle's angles exceeds 180°. Any set of N nodes will represent an Euclidean configuration if and only if the triangular inequalities rule is not violated for any triple of points. Thus, the triad cannot be described without a complex dimension (one with a negative root) to foreshorten the bc leg. If the values of s_{ij} were converted to binary values measuring absence or presence of links (0 or 1), which foreshortens the bc leg, the triangle becomes Euclidean – but this throws away information.

Therefore, network data need not be Euclidean, i.e., at least one of the characteristic roots of B may be imaginary. One solution is a Riemann manifold represented by a coordinate system in which some of the dimensions are imaginary. The locations of the non-Euclidean relations among the nodes may be determined by Equation 1.

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

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 the negative eigenroots result (Woelfel and Barnett 1982).

While multidimensional scaling has frequently been applied to analyze social networks, less than satisfactory results have been reported. One reason for this may be the failure to take imaginary dimensions into account. Historically, psychometricians have treated the variance on these dimensions as error to be removed through the addition of an additive constant (Messick and Abelson 1956) or adjusted away by a non-metric algorithm (Kruskal 1964a,b; Shepard 1962a,b). They assumed that social and psychological structures were Euclidean and that any departure from a positive semi-definite scalar products matrix (one with only positive values in its eigenvector) was caused exclusively by measurement error. Thus, imaginary dimensions were ignored or transformed and inadequate descriptions of sociometric data resulted. Additionally, the stated purpose of MDS was to identify an underlying structure, such as the dimensions by which a network

[1] In the example, matrix S results in a two-dimensional (one real and one imaginary) space because any matrix of N points may be described without the loss of information by a space of N-1 dimensions. For example, any two points may be precisely described by a line.

was differentiated. This resulted in focusing upon only the dimensions which accounted for the greatest proportion of the variance. Other dimensions with less variance were removed (see, for example, Levine 1972, on interpreting corporate interlock factors) and imaginary dimensions were ignored. However, since the underlying dimensions are only orthonormal reference vectors upon which no meaning may be directly attributed, all dimensions should be retained for analysis, including those with negative eigenroots (Barnett and Woelfel 1979). Attribution of meaning to the dimensions may be made by regressing an attribute vector through the space.

Recently, psychometricians have become interested in MDS in Riemann space (Lindman and Caell 1978; Pieszko 1975). Griffith (1984) has modelled transportation networks using non-Euclidean geometries. However, his models are constrained by physical space which limits the dimensionality to two or three axes. The structure of social space may be more complex.

3. Galileo – a MDS algorithm

3.1. Scaling and longitudinal analysis

One algorithm exists which allows for the analysis of all the dimensions in a multidimensional manifold including those with negative roots. It is known as Galileo (Woelfel *et al.* 1977; Woelfel and Fink 1980). The computer software takes ratio level measurements of discrepancies (distances or dissimilarities), such as matrix S, and converts them to an adjusted scalar products matrix following Torgerson (1958). It then determines the eigenroots and Cartesian coordinates for all dimensions, real and imaginary through Jacobi's method (Van de Geer 1971).

Previous research with Galileo has shown that the loadings on the imaginary dimensions are reliable both across groups and over time (Woelfel and 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 are made with stimuli from two or more semantic domains or when the stimuli are incongruent or produce a psychological state of imbalance. Krumhansl (1978) examined psychological non-Euclideanisms in geometric models and found that violations of triangular inequalities resulted when 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. Similar results occur with social network data.

One reason for performing network analysis has been clique or group identification. Two procedures may perform this function in conjunction with scaling algorithms, cluster analysis or multiple discriminant analysis (MDA). Once the Riemann space has been obtained, the researcher may perform a cluster analysis to identify groupings within the space, or when group identification is known or hypothesized, MDA may be used (Jones and Young 1972). In the latter 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, the spaces are translated to a common origin and rotated to a least squares fit which minimizes the departure from congruence among the spaces. The individual spaces are not standardized. This allows one to examine the dilation or contraction of a social network. 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, hypotheses about future network structures may be tested (Barnett 1984; Rice and Barnett 1985).

The ability to compare manifolds in this manner represents an advance over standard practices in psychometrics. A typical solution is to use a Procrustean rotation which first dilates or contracts the spaces to a common size and performs a least squares rotation to minimize departure from congruence (Schonemann and Carroll 1970). This procedure is unacceptable because change in the size of the space may represent true change. The network may change in density or connectedness and this information would be lost by standardizing the volume of the network space.

When no additional information about the relative stability of the nodes exists, the ordinary least squares procedure may be applied.

However, when the nodes' stability or the extent to which the position of certain nodes has changed is known, alternative rotational algorithms exist (Woelfel *et al.*, 1980). The least squares procedure has the effect of overestimating some changes while underestimating others. This may lead to erroneous conclusions. 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 theoretically 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 positions 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 node 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 (Barnett 1980; Woelfel *et al.* 1980).

3.2. Network indices from Galileo

Galileo provides the dimensionality (eigenroots) for each sociomatrix, the locus for each node on all dimensions (real and imaginary) of the network's space, and the changes in the network at adjacent points in time. Since it is a metric algorithm, Galileo also produces the following statistics which may be used to describe the networks.

Centrality may be defined as the average distance of a node to all others in the network (Freeman 1979). Because MDS places the centroid of the nodes at the origin, the centrality of any node may be determined from the diagonal of the adjusted scalar products matrix. The value on the diagonal, b_{ii}, represents the squared distance of node i from the center of the network. The greater the square root of the absolute value of b_{ii}, the less central the node is in the network. [2]

System connectedness has been defined by Rogers and Kincaid (1981 : 346) as "the degree to which members of a system are connected to others in the system." The trace of sum of the eigenroots of the

[2] The absolute value of b_{ii} is taken because b_{ii} is negative if the dimension is imaginary.

coordinates provides an indicator of connectedness. The smaller the distance, or the greater the frequency of interaction among the system's nodes, the smaller this coefficient.

Network density is defined as a ratio of the number of nodes to the volume of the space produced by scaling the social distance matrix. The trace may be considered a measure of the volume of the coordinate space. It is the sum of the squared lengths of the dimensions or coordinate axes (equal to the sum of the eigenroots). Density is a useful indicator when comparing different systems. However, because the number of nodes generally is constant when comparing the same system over time, the measures of system connectedness and network density for such networks produce equivalent results.

Homogeneity of linkage among the nodes is indicated by the warp of the coordinate space. *Warp* is defined as the ratio of the real variance (from all dimensions with positive eigenroots) to the total variance (from positive and negative roots) in the space. It is a convenient measure of the degree to which a space is non-Euclidean. It will be greater than or equal to 1.0. An Euclidean space will have a warp of 1.0. The greater the warp, the less Euclidean and more heterogeneous the pattern of interaction among the nodes and the more frequently contacts go through a limited number of more central nodes. Thus, warp is an indicator of (though not equivalent to) network integration. *Integration* is the extent to which the nodes which are linked to a focal node are linked together (Rogers and Kincaid 1981). A well integrated network will have a warp approaching 1.0.

To demonstrate these concepts we will return to the three-node example presented at the beginning of this article. When the three nodes were entered into Galileo, the following coordinates and eigenroots resulted.

	Dimension		
Node	1	2	3
a	0.00	0.000	−2.93
b	4.50	0.000	1.46
c	−4.50	0.000	1.46
Eigenroot	40.50	0.000	−12.83
Sum of roots =	27.67		
Warp =	1.46		

The nodes are arrayed in two-dimensional space. One of these dimensions (1) is positive and one (3) is negative. Warp is greater than 1.0, indicative of a non-Euclidean spatial manifold. The square root of the diagonal of the scalar products matrix provided measures of the nodes' relative centrality, or distance from the origin of the network space. They were 2.93, 4.26 and 4.26. As suggested earlier, node *a* is the most central. It is at the origin on dimensions 1 and 2 and 2.93 units from the origin on 3. Since the indicators of connectedness and density have meaning only relative to each other over time or when comparing networks they will not be discussed.

Up to this point, the discussion has focused on the theoretical desirability of using a non-Euclidean MDS algorithm to describe social or communication networks. Any new methodology's utility should be evaluated against theoretical as well as technical criteria. The following sections empirically demonstrate these procedures with two sets of data.

4. Empirical examples

The two following sections are included primarily as examples of applications of Galileo. Theoretical foundations and implications are discussed more thoroughly in the original papers.

4.1. American air traffic

Data from the annual "Domestic Origin-Destination Survey of Airline Passenger Traffic" conductd by the U.S. Civil Aeronautics Board (CAB), were analyzed with Galileo. [3] Flight coupons surrendered by passengers upon boarding were the source of the survey data. The universe consisted of all coupons lifted by participating air carriers. Coupons with ticket serial numbers ending in zero were selected, resulting in a 10 percent systematic sample.

[3] These data may be obtained from:
The Economics and Finance Council, Air Transport Association of America, 1709 New York Avenue, N.W., Washington, DC 20006, U.S.A.
The CAB edited the data, removing inconsistencies 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 were records which fail computer editing tests. Less than 2 percent of the total reported number of flights were dropped from the survey.

Table 1
Population of 31 cities selected as nodes in air traffic analysis. [a]

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	Pittsburgh	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 populations of U.S. 1980	225,479,000

[a] Sample contains 43.5% of total. See footnotes 3 and 4.

Thirty-one cities (SMSA) with a population greater than one million were selected as nodes. [4] They are listed in Table 1. In 1980, these 31 cities had a cumulative population over 94 million or 43.5 percent of the U.S. population. The links in the symmetric sociomatrix S were the

[4] San Antonio and Sacramento were not included. 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 at its airport.

number of passengers outbound plus inbound (nondirectional) between the cities. The diagonal contained zeros. Fourteen separate sociomatrices were created, one for each year 1963 to 1981.

These data are not subject to the criticism of self-report network data (Berger and Roloff 1980; Bernard and Killworth 1977; Bernard *et al.* 1980, 1982; Nisbett and Wilson 1977). Rather than being reports of travel by individuals, they are objective, coming from used airline tickets. The nodes are cities, not individuals. Thus, the interaction among aggregates was examined. Danowski (1982) and Barnett (1982) have shown that the process of aggregating interactions to the group level filters out significant measurement error because individual variation is averaged. This results in stable estimates which improve the ability to describe the mathematical relations among the variables of interest. In this case, a 10 percent sample of air traffic is sufficiently large to assume that random perturbations contribute little to the description of the network.

These data were transformed from matrices of frequencies of interaction to social or communication distance so that the greater the interaction between two nodes, the closer they are in network space. The transformation function was

$$S' = \log S(K).$$

K was set equal to 14,638, the value required for equivalent traces between the spaces produced by the physical distance between the cities and the air traffic.

4.1.1. Results

Change in connectedness. The 14 sociomatrices were transformed into multidimensional spaces and comparisons made using a rotation to a least-squares best fit which minimizes overall departure from congruence. 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 is the trace of the coordinates matrix, an indicator of network connectedness. The smaller the value, the greater the connectedness. The traces for the 14 years are presented in Table 2, and are plotted in Figure 1.

To describe the change in connectedness, these 14 values were plotted against time. Connectedness increased continuously until the last two years, when it decreased.

Table 2
Resultant coefficients, observed traces, predicted values and residuals using air traffic data. [a]

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	50,526	−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

[a] Values in thousands.

The data were fit to both an exponential decay and a polynomial with a negative linear term and a positive quadratic term. For the exponential decay function, $R^2 = 0.864$. For the polynomial, $R^2 = 0.752$. Both the linear and quadratic terms were significant. That is, the overall process of change in network connectedness is increased connectivity according to a decay function, but the reversal in the last two time periods is captured by the polynomial.

The change scores. The overall change scores (the sum of the differences between the spaces at t_{i+n} and t_i reveal a consistent pattern. They are presented in Table 3 and Figure 2. The data suggest two distinct epochs: an early period, 1968–74, characterized by a high rate of change, and a stable period, 1974–81. The airline network initially changed from a rapid but nearly constant rate of change to a slow and nearly constant rate of change.

The change in individual nodes. Insights into the changing pattern of the nodes' relationships may be gained by examining specific nodes. For each of the first seven years, Ft. Lauderdale and San Diego changed more than twice the overall average. During the later period, Tampa and Dallas changed more than the average, but these were small

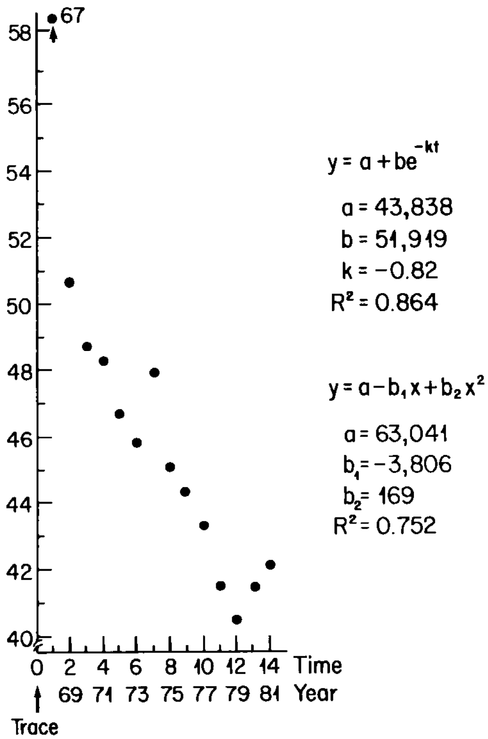

Figure 1. Trace of air traffic network over time.
Note: Significance of polynomial coefficients: b_1 ($p < 0.005$); b_2($p < 0.03$).

when the overall magnitude of change is considered. Specifically, Ft. Lauderdale and San Diego moved from the periphery toward the center of network. These nodes were replaced at the periphery by smaller midwestern cities – Columbus, Cincinnati and Indianapolis.

Change in the air traffic network structure. The network structure changed over time. Groups within the network were identified by hierarchical cluster analysis. During the first epoch, there were two regional groups. One was centered around Chicago and New York and included all the eastern and midwestern nodes from Miami to Minneapolis. The other cluster was centered on the west coast around Los Angeles and San Francisco. It included another regional cluster containing New Orleans, Dallas and Houston. Worth noting were the positions of Kansas City and St. Louis. While the latter was part of the

Table 3
Overall change between adjacent points in time using air traffic data

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

eastern cluster, the former was grouped with the west. The break in the air traffic network in 1969 appeared to go through 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 clusters.

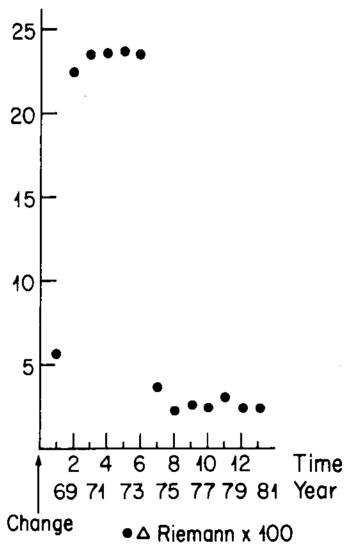

Figure 2. Overall change in coordinate value of air traffice network over time.

New York, Chicago, Los Angeles, San Francisco, Dallas and Houston combine into a single cluster. Other individual nodes are then added to this hub with little prior regional clustering.

This pattern was confirmed through regression analysis. The coordinate values of 1969 and 1980 were regressed on latitude and longitude. [5] In 1969, the first dimension accounted for 70.4 percent of the variance in longitude and the first four, 83.6 percent. The first dimension accounted for 25.1 percent of the variance in network structure and the four together, 34.3 percent. In 1980, the first dimension accounted for only 63.7 percent of the east-west variation. It took six dimensions to account for 83.3 percent. The first six accounted for only 13.4 percent of the network structure and the six 41.1 percent. The variation attributable to longitude is more homogeneously distributed in 1980, indicating a breakdown of the regional groupings.

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

Network homogeneity of linkage. The distribution of air traffic in the U.S. became more homogeneous. This may be supported by examining

[5] 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 latitude and longitude. Dimension 1's correlation with longitude was 0.993. Dimension 2's correlation with latitude was 0.982. These dimensions accounted for approximately 81 and 19 percent of the variance in the distances among the cities. Together, they account for 98.7 percent of the variance in the distance among the cities. The remaining 1.3 percent may be attributable to measurement and rounding error and the curvature of the earth. Thus, regressing matrix S's coordinate values upon the latitude and longitude can be used to determine the impact of physical location on network structure.

changes in the warp and the distribution of variance among the dimensions (eigenvalues). If the network became more homogeneous, that is, the links became equally strong, then the space would become more Euclidean (warp = 1.0) and the variance explained by the single largest dimension would decrease. In 1968, it was 330.8 percent (Percentages greater than 100 are due to the warp.) In 1969, it was 43.0 percent, and by 1981, 14.0 percent. If all dimensions were equivalent, each would account for only 3.2 percent 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 stabilized between 2.89 and 4.52.

The warp indicates that the links became more homogeneous. In 1968, it was 3.97, in 1969, 1.25, and by 1978, 1.04. Warp, however, showed a very slight increase during the last two years of the analysis. It rose to 1.09 (1980) and 1.10 (1981). This suggests that the distribution of air traffic was becoming slightly less homogeneous. This is consistent with the finding that the network became somewhat less interconnected. The values of the percent variance of the first dimension, standard deviation of the eigenroots' percent variance and the warp for each year are presented in Table 4.

Graphic representations of network structure. Scaling programs produce coordinate values which can be used to graphically represent the

Table 4
Homogeneity of the air traffic network over time

Year	Percent variance first dimension	Standard deviation of eigenroots	Warp
1968	330.8	77.73	3.97
1969	43.0	9.37	1.25
1970	28.4	5.22	1.13
1971	25.5	4.64	1.11
1972	21.9	3.98	1.07
1973	20.7	3.97	1.07
1974	23.8	4.52	1.09
1975	17.8	3.33	1.06
1976	16.6	3.23	1.06
1977	16.3	3.10	1.04
1978	15.0	2.97	1.05
1979	14.6	2.89	1.05
1980	14.6	3.22	1.09
1981	14.4	3.30	1.11

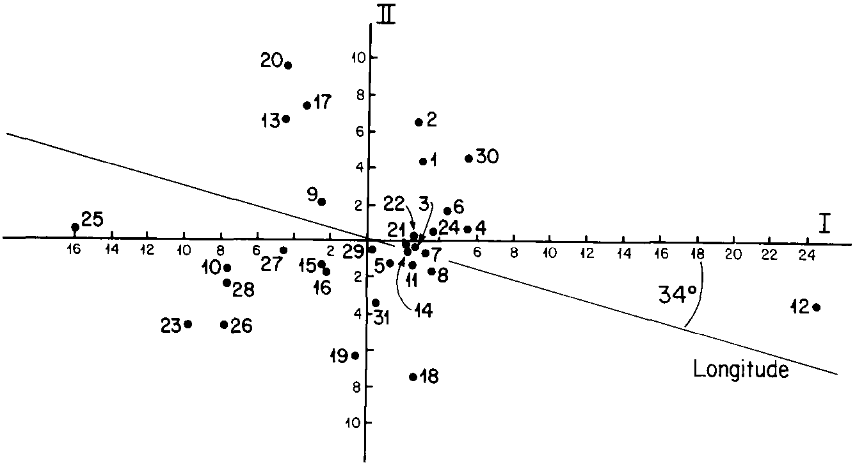

Figure 3. Two-dimensional portrayal of 1970 air traffic network with regression of longitude of airports.

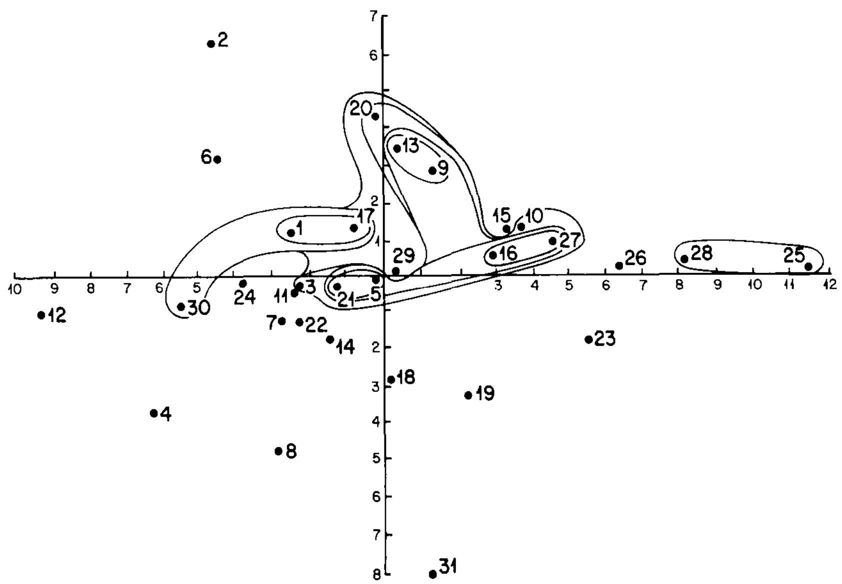

Figure 4. Two-dimensional portrayal of 1980 air traffic network with hierarchical clustering solution.

relationships among the nodes. Plots have not been presented because of the low percentage of variance attributable to any two dimensions. However, since the goal of this article is to demonstrate the utility of this method, two plots are presented. They are 1970 (Figure 4) and 1980 (Figure 5). The two plotted dimensions account for 37.7 percent of the variance in 1970 and 23.0 percent in 1980. The later percentage is smaller due to the increased connectedness of the network. The 1970 plot has longitude regressed on the first dimension. The 1980 plot includes its cluster analysis. There is considerable distortion in both plots due to the low percentages of explained variance.

These plots were chosen to demonstrate network changes. 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 it is clear that density and connectedness became greater. The 1980 plot shows a further breakdown of the regional clusters, an increase in homogeneity and interconnectedness. If these two dimensions accounted for all the variance in the network, centrality could be visually represented by the node's distance to the origin. In the plots Chicago is closest to the origin. The issue of centrality will be discussed at greater length in the second example.

Stability within the network. Up to this point only changes in the network have been discussed. Stability has not been addressed. Stability may be inferred through an examination of the correlations of the node's locations on the dimensions at adjacent points in time. The mean correlation for the first dimension was 0.981. It was 0.986 for the second, indicating high stability.

A critical argument in this article concerns the use of the dimensions with negative eigenroots. The mean correlation on the largest (absolute value) of these dimensions across adjacent points in time was 0.67. For the last ten points it was 0.82 and 0.99 for the final four. This indicates that the variance on the imaginary eigenvectors is not random. Change in the size of these dimensions, as reflected in the warp, and the arrangement of the nodes on them, should be examined.

One reason for stability within the network is the physical distance among the nodes. Physical proximity is one determinant of network structure (Olsson 1965; Rogers and Kincaid 1981). To determine how physical structure influences network structure, two multiple regressions were performed with the 14 sets of network coordinates as the independent variables and the 31 cities' latitude and longitude as the dependent variables. $R^2 = 0.35$ for latitude and 0.84 for longitude. These coefficients were multiplied by the mean proportion of variance accounted for by the dimensions across 14 data sets. [6] Since latitude and longitude are orthogonal, these values were summed. The results indicate that 18.3 percent of the variance in network structure was accounted for by the physical relations among the nodes. That is, a fifth of the structure in air travel in the U.S. can be explained simply by the distance between and the location of cities.

Another factor contributing to stability within the network is the population of the nodes. Population is a major determinant of the frequency of interaction among cities (Hamblin 1977; Olsson 1965). The correlation between the cities' populations in 1970 and 1980 was 0.99. Those nodes which moved greater than the average (Ft. Lauderdale, San Diego, Dallas and Tampa) all grew at least 24.7 percent. Ft. Lauderdale, the node whose position changed the greatest, grew 68.2 percent. These nodes along with others with comparable growth rates (Houston, Denver and Phoenix) all moved from the periphery to the center of the network, suggesting that population stability may contribute to the overall network stability and that change in the air traffic network is, in part, 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. Further, those variables which facilitate or inhibit this change can account for the change in air traffic. Among them are economic factors (GNP, GNP service, personal income, unemployment, automobile sales, and fuel prices), changes within the airline industry (deregulation) and the network itself (the opening of Dallas' and Atlanta's airports). To determine their impact, annual data on the variables were correlated with the trace and 13 change scores. To control for autocorrelation, difference scores were used. Thus, the change in the variables were correlated with the change in the trace and the overall change (including that occurring on the imaginary dimensions) between adjacent points in time.

[6] 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, error is entered into the analysis and the estimates of goodness-to-fit are lowered. Dimension k at time i may not be dimension k at time $i+1$ due to change in the network which changes the order in which the dimensions are extracted.

Table 5
Correlation of external variables with change in air traffic network structure

Correlates	Trace	Change in trace	Mean overall change
Deregulation	−0.51	0.26	−0.46
Atlanta's opening	−0.31	0.26	−0.36
Personal income	−0.69 *	0.38	−0.56 *
Dallas's opening	−0.60 *	0.27	−0.89 *
GNP	−0.68 *	0.41	−0.67 *
GNP service	−0.61 *	0.25	−0.38
Producer prices	0.75 *	−0.24	−0.29
Consumer prices	0.78 *	−0.02	0.12
Unemployment	−0.65 *	0.10	0.03
Automobile sales	0.17	0.00	0.00
Fuel prices	−0.80 *	0.53	0.31

* Significant at 0.05 level

The opening of the Dallas airport correlates − 0.89 with the overall change. Prior to its opening, there is a consistent high rate of change (see Figure 2). Afterwards 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 variables has 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 the trace (0.53) and a strong relation with the overall change. These results are presented in Table 5.

Descriptively, these variables relate to the critical points in time that have been identified through the analysis of the network. In 1974, there was a slight reversal in the trend toward greater connectedness. That year, fuel prices had their first large increase. Jet fuel prices caused an increase in ticket prices which may have resulted in fewer trips and thus lower connectedness. Between 1974 and 1975 the network stabilized. In 1974, the Dallas airport opened. Its use as a central hub seems to have stabilized air traffice.

1980 began a trend toward lower connectedness. Atlanta's airport opened. There was no longer a need to travel through a central node since Atlanta expanded into a regional hub. Thus, the network decentralized. 1980 also saw an increase in unemployment and a smaller increase in personal income. Thus, the change in the trend may be due to economic factors. 1981 was also marked by the air traffic controllers' strike which forced a cutback in air traffic. The continuation of the trend may have been due to this event.

4.1.2. Summary

This example has demonstrated the utility of Galileo in describing changes in the U.S. air traffic network, 1968–81. The results suggest that Galileo can be used to describe the orderly change in network data and that changes can be described by simple mathematical functions that can readily be interpreted when exogenous factors are examined. Much of the analysis (such as the regressions of physical distances) reported here would not have been possible if a non-metric algorithm had been employed because such algorithms alter the relations among the scaled nodes. Futher, important insights were made because the data were analyzed with a non-Euclidean algorithm. Warp, the measure of the extent to which the space departs from being Euclidean, provided important insights about the homogeneity of the linkage structure. Similar results were obtained with a more laborious examination of the distribution of variance of the eigenroots. Further, loadings on those dimensions with negative eigenroots were highly correlated with one another at adjacent points in time. This indicates that these dimensions are not composed of random error but rather true variance which facilitates the understanding of social networks.

4.2. Computer conferencing groups [7]

The second example is the "private messaging" behavior of the users of the EIES computer conferencing system. The study reporter here is part of a larger analysis of new communication media (Hiltz 1983; Hiltz and Turoff 1978; Kerr and Hiltz 1982; Rice 1980a, b; Rice *et al.* Associates 1984; Rice and Case 1983; Rice and Paisley 1982). Other network analyses of portions of these data are reporter by Freeman (1980, 1984). These messaging data constituted 70 percent of all items on the system. A total of of 10 groups made up the EIES system for its first months. Group 0 ($n = 17$) included user consultant and system personnel, whose role was to maintain the system and help others users. They communi-

[7] Partial support for preparation of the network data was provided by National Science Foundation Grant NSF-MCS-77-27813 to Dr. Roxanne Hiltz. Much of this section is a revised version of portions of Rice and Barnett (1985).

cated more with the other groups than any other specific group. Groups 1 ($n = 45$), 5 ($n = 56$), 6 ($n = 25$), 7 ($n = 30$) and 8 ($n = 76$) were task-oriented groups, while groups 2 ($n = 32$), 3 ($n = 46$) and 4 ($n = 67$) were not formally mandated to accomplish a task. Group members were typically researchers, university faculty or government-sponsored agency personnel. Group 9 included all those users not otherwise members of a specific group. They did not perceive themselves as members of any group, so were free to roam throughout the conferencing system.

Some of the groups were part of EIES system staff, some were researchers invited to join the system as part of the evaluation, others were ongoing research groups, and not all groups were on the system for all 25 months. Groups 0 and 9 were, as were 2 through 5. But, 6, 7 and 8 entered during month 12. Group 1 left the system at this time, but entered again in month 21 with a different identity.

One important attribute of the data is that they were collected by computer. Computer-monitored data for network research have a number of advantages (Danowski 1982; Rice and Borgman 1983; Rice and Rogers 1984). One, the data represent acutal communication behavior as noted earlier. Two, a full census of system users can be obtained, which better represents the interactive nature of communication data and which allows network analysis for which samples of the entire population are typically needed. Three, extensive longitudinal data are accessible so that limitations of cross-sectional network research may be overcome (Monge 1982; Rice 1981).

Over the 25 months, more than 700 unique users participated. The first month was start-up time, so this analysis begins with month 2. Although data were collected continuously, they were aggregated into monthly intervals to facilitate handling and analysis. The monthly aggregation resulted in nearly 87,000 data points each a link identified by individual and group, sender and receiver, and month. At the system level, the data indicate how many messages were sent within each group and to each of the other groups in each month.

Prior anlaysis of these data used both the intragroup and intergroup data to describe and test models of network development. Specifically, the system as a whole was very well described by a log-linear model which posited reciprocal information flows between groups but similar flows within groups. Groups were well-categorized into network roles based upon estimated parameters of these information flows (Rice

1982). For example, a group which sent and received more messages than the average group (in that month) was classified as a "carrier". "Receiver" groups received more messages but sent fewer messages than did the average group. "Transmitters" did the opposite. "Isolate" groups sent and received less. Each of these roles can also be differentiated according to their having higher or lower than average levels of within-group messaging. Using the four-fold role typology, however, analysis showed that groups 0 and 9 were consistently information carriers. Task groups generally remained isolates after a few months as carriers or receivers, while nontask groups shifted from isolate roles through transmitters to carriers or receivers (Rice 1982: 939). In general, an electronic environment is entropic and it is difficult for a group to sustain a carrier role.

Previous analysis of these data found the relations among the groups to be best modelled as reciprocal across groups (Rice 1982). Thus, matrix S was treated as symmetrical. Therefore, the nondirectional interaction frequencies (the average of s_{ij} and s_{ji}) were entered into both the upper and lower triangles of the 24 matrices, one for each month.

Two scaling approaches were used. In the first, matrix S was composed of communication frequencies instead of distances. This operation has a number of advantages. One, connectedness is scaled positively. Two, it is simpler than using distance matrices and does not alter the dimensionality of the network space. Three, when three is no theoretical criterion for selecting a function for transforming similarities into dissimilarities, the direct scaling of the frequencies allows the direct comparison of changes in the network (Barnett 1984). Scaling frequencies is the same as using similarity data, rather than dissimilarity data, to scale psychological stimuli (Shepard et al., 1972). Its major disadvantage is that the greater the frequency of interaction among two nodes, the farther apart, rather than closer, the nodes are in space. Graphic interpretation is difficult.

The second approach transformed the frequencies of interaction to social distances. All off-diagonal elements were subtracted from 655, the largest intragroup communication frequency. However, because the largest frequency of intragroup communication was 405, this transformation added a constant to all values in the sociomatrices, altering the dimensionality of the coordinates and the value of a number of descriptive indicators including warp. These matrices, once transformed into social distances, were entered into Galileo. Results from both approach will be discussed.

4.2.1. Results

Centrality. Table 6 presents the centralities of the groups for each month. [8] Groups 0 and 9 are the most central. They are closest to the origin. At nearly every month, group 0 is the most central node, as befits its service role. Nearly all system-wide "broadcast" messages emanated from this group. All users could send comments or help messages to this group. It was an information "carrier." The group of unaffiliated users, 9, was the most central in the remaining periods. In general, nontask groups were more central than the task groups, supporting their prior role-categorization as information "carriers." Group 8 was the exception, becoming more central. The other task groups are less central and were categorized as information isolates by prior research (Rice 1982).

Connectedness. Table 7 presents the two indicators of system connectedness and homogeneity of linkage. The first indicator, the trace of the communication distance matrix, is inversely related to connectedness. The second, the trace of the frequency matrix, is a direct measure of connectedness. Since the participants knew that the conferences were ending at the end of month 25, use of the system decreased before then. As a result, data from month 25 were dropped from further analysis.

These indicators were plotted against time, revealing that connectedness increased over time. There was no evidence of non-linear trends. A linear regression using successive time periods was performed with both indicators. For communication distance, $R^2 = 0.80$, $b = -3{,}150$ and for the frequencies, $R^2 = 0.78$, $b = 429$. An examination of the residuals failed to identify any additional pattern.

Homogeneity of linkage. The network's linkages were not distributed homogeneously. Warps ranged from 3.32 to 4.97. Homogeneity of linkage remained stable, with only 10 percent variation between time periods. This stability is partially due to the fact that increases in linkage remained proportional across the groups. Groups 0 and 9 were the most central over time and their frequency of contacts remained

[8] Because the matrix cell values have been transformed by the greatest intragroup frequency, 655, rather than one greater than the largest intergroup frequency, the centrality values have meaning only relative to one another.

Table 6
Centrality of groups at 24 monthly intervals

Month	Groups									
	0	1	2	3	4	5	6	7	8	9
2	421.7	440.6	441.4	440.5	440.9	440.2	441.4	441.4	441.4	423.6
3	421.1	441.8	437.5	438.9	433.4	442.1	443.0	443.0	443.0	421.9
4	407.5	442.8	437.4	440.2	431.2	441.7	443.7	443.7	443.7	417.2
5	397.4	444.9	434.4	430.2	431.9	438.7	445.8	445.8	445.8	410.4
6	399.2	444.0	433.2	423.3	434.4	439.0	447.0	447.0	447.0	399.5
7	399.4	443.7	434.2	432.3	429.4	441.5	445.9	445.9	445.9	407.3
8	396.9	445.0	436.0	430.9	427.0	441.3	446.1	446.1	446.1	407.6
9	399.3	445.8	433.5	436.6	424.0	436.3	446.1	446.1	446.1	408.6
10	403.2	443.3	435.1	437.6	433.6	438.9	444.7	444.7	444.7	411.7
11	401.4	444.8	433.3	434.2	433.7	437.4	445.7	445.7	445.7	405.7
12	397.3	445.7	437.5	440.0	437.7	442.0	440.3	444.0	439.7	402.6
13	392.7	447.1	439.1	438.3	435.7	440.7	440.7	442.9	438.4	396.3
14	377.2	449.1	437.0	440.5	436.7	439.3	443.5	441.7	436.2	486.5
15	388.1	447.6	437.0	440.7	436.5	441.0	444.3	442.1	436.5	391.1
16	389.2	450.8	435.1	436.7	430.7	440.2	445.1	442.5	429.1	369.9
17	386.0	448.7	439.5	430.3	437.7	440.8	441.8	442.9	434.4	391.1
18	385.9	449.3	439.4	429.8	436.5	442.0	444.6	442.5	429.5	388.0
19	386.0	449.9	438.4	438.1	436.8	440.9	443.0	442.7	424.0	379.5
20	384.8	449.3	434.4	440.5	434.8	444.4	444.7	443.7	429.3	379.6
21	391.4	443.4	435.6	437.7	434.1	444.9	444.9	441.1	436.1	391.6
22	379.1	443.2	433.4	441.4	431.6	442.3	445.9	444.9	433.4	385.8
23	369.8	442.4	433.2	442.5	433.8	444.1	448.6	444.5	432.8	374.7
24	370.9	437.6	432.6	442.1	435.8	444.3	449.3	445.8	426.4	373.7
25	412.5	439.7	438.7	441.1	438.4	441.0	443.5	440.8	436.6	418.9

[a] Underlined values indicate isolates, or groups not on the system at that time period. Lower values are more central. Values are relative to highest intergroup communication frequency plus one. Because the matrix cell values have been transformed by a factor of 655, the table values have meaning only relative to one another.

disproportionally greater than the other nodes. This network may be simplistically described as a hub-spoke network with two groups at the hub. This structure did not change much over time.

The warps in this example are much larger than in the first (see Table 4 and 7). In the prior example, the network was highly interconnected. It may be described as having a common-communication or all-links structure. None of the nodes was significantly more central. As a result, a multidimensional Euclidean space may have been satisfac-

Table 7
System connectedness and warp at 24 monthly intervals

Month	Inverse connectedness: Trace of distance matrix	Connectedness: Trace of frequencies matrix	Warp
2	1,912,942.8	1546.8	4.97
3	1,899,942.5	1646.9	3.97
4	1,892,836.8	2521.3	4.01
5	1,873.427.2	3406.4	3.32
6	1.863,521.5	4960.6	3.78
7	1,873,365.0	3996.4	3.91
8	1,871,554.2	4477.0	3.85
9	1,870,964.2	4215.0	3.92
10	1,883,379.7	3237.4	4.20
11	1,875,145.2	4496.3	4.34
12	1,874,990.5	6145.2	4.74
13	1,862,680.9	5888.3	4.39
14	1,844,001.1	9999.5	4.35
15	1,857,397.2	7483.0	4.37
16	1.828,973.0	8857.8	3.76
17	1,847,636.0	7538.3	4.19
18	1,842,895.6	7783.9	4.06
19	1.837,735.7	8695.1	3.85
20	1,842,339.6	9518.4	4.12
21	1,856,089.0	6693.4	4.32
22	1,838,332.3	7805.0	3.85
23	1,827,824.1	14393.5	4.33
24	1,821,040.8	12130.9	3.90
25	1,894,242.7	1421.3	4.08

tory to describe the network. In this example, however, the nodes are not highly linked and two nodes are more central than the rest. As a result, the warps are much higher and a Riemann manifold of limited dimensionality is required to describe the network.

Dimensionality. Scaling the frequency matrices resulted in a two-dimensional solution for all time points. These were the largest real and imaginary dimensions (the overall solution reported four real and six imaginary dimensions). They accounted for virtually all (99%) the variance in the coordinate space. The mean correlation among the dimensions at adjacent points in time were 0.999 (real) and 0.740 (imaginary). The amount of variance, the high correlation among the loadings on the imaginary dimensions, and the values of the warp indicate that this network must be described in a Riemann, rather than Euclidean space, and that the network's overall structure was relatively stable over time.

The nodes' loading on these two dimensions is instructive. For month 2, all nodes except 0 and 9 were approximately at the origin, on the dimension with a positive eigenroot. Groups 0 and 9 had loadings at 62.02 and -61.98 respectively. On the one with a negative root all the nodes except 0 and 9 were at about 12.35. Groups 0 and 9 were -49.55 and -49.58. Thus, the coordinates which result from MDS provide information which can be used to interpret network structure and determine where violations of the triangular inequalities occur.

Scaling the distance matrices did not result in as simple a solution. The two largest dimensions accounted for an average of only 15.2 and 11.7 percent of the variance (these comprised nine real dimensions and one imaginary dimension). (An equal distribution of variance among the dimensions would have resulted in 10 percent.) The mean correlations among these dimensions between adjacent points in time were only 0.69 and 0.58 ($F = 18.5$, 11.3; $p < 0.005$). Because of the lower percentage of variance explained and the lower correlations among the dimensions, all further analysis compared the network's coordinate spaces from the frequencies rather than from the distances.

However, for graphic purposes, months 2, 14 and 24 of the distances have been plotted in Figure 5. [9] The two dimensions in Figure 5 account for only 25–28 percent of the variance in network structure. Therefore, conclusions about how the network changed should not be based upon this representation. Figure 5 indicates that the network became more connected. Again, groups 0, 8 and 9 appear to move toward the origin, while the task groups other than 8, identified as increasingly more information isolates by Rice (1982), drift away from the origin. [10] Nontask groups 2, 3 and 4 cut across the central area of the space, remaining information carriers and relatively central to the system.

[9] This plot was drawn from the coordinates after they were rotated, to minimize departure from congruence from the previous point in time.

[10] Rice and Barnett (1985) explain group 8's apparently anomalous categorization as a result of the differences in MDS *vs.* log-linear analyses. The log-linear analysis took into consideration intragroup communication, while MDS does not. Group 8's heavy intragroup interaction reduces its relative centrality in the three-way log-linear analysis.

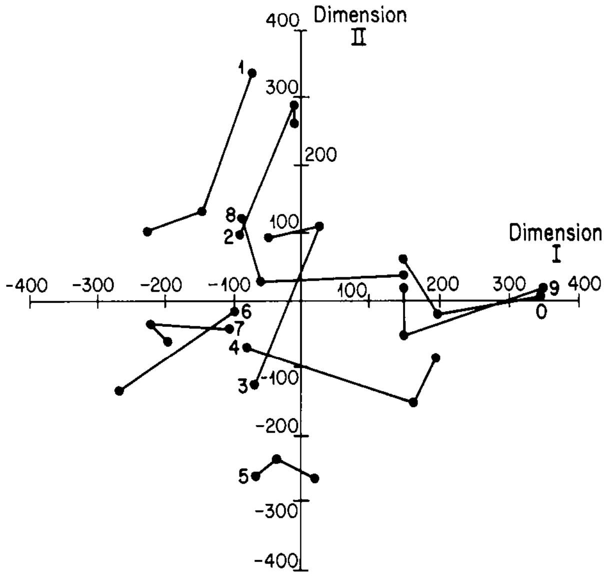

Figure 5. Change in spatial coordinates of ten computer conferencing groups, at three time periods, based upon distance matrix.
Note: Position at Month 2 is closest to group number. 25.4 percent of variance explained at Month 2. Position at Month 13 is middle point of line segment. 26.8 percent of variance explained at Month 13. Position at Month 24 is end point of line segment. 28.2 percent of variance explained at Month 24.

Change in the structure of the network coordinates. The coordinates which resulted from the MDS were next compared using an ordinary least-squares rotation. The overall differences among the nodes on all dimensions are presented in Table 8 and Figure 6.

Table 8 and Figure 6 show an increase in the network's rate of change. A linear regression with monthly intervals as the independent variable suggested that the network did not stabilize ($R^2 = 0.8$, $a = -0.24$, $b = -12.6$, $F = 88$, $p < 0.001$) in spite of the high correlations between dimensions across time. However, a closer examination of the changes reveals an initial slow rate of change for months 2 to 5, with an acceleration between months 5 and 9. The differences became smaller for months 9 to 11, suggesting that the network structure was beginning to stabilize. Groups 6, 7 and 8 entered the network at month 12. As a result, the rate of change increased between months 11 and 12. The reciprocity model did not fit the data well at this time either (Rice

Table 8

Months	Dimensions		
	Overall	Real	Imaginary
2–3	−1.49	1.46	2.86
3–4	−4.97	1.98	4.18
4–5	−10.80	2.48	6.41
5–6	−55.47	3.37	11.98
6–7	−56.84	2.75	8.40
7–8	−88.91	3.01	9.72
8–9	−75.03	3.13	10.27
9–10	−38.77	2.65	7.47
10–11	−86.33	3.22	11.50
11–12	−169.94	4.11	18.72
12–13	−156.98	3.78	15.88
13–14	−244.87	4.87	26.42
14–15	−194.80	4.27	20.34
15–16	−180.80	4.20	19.84
16–17	−186.78	4.21	19.69
17–18	−183.72	4.27	19.49
18–19	−188.33	4.40	21.19
19–20	−219.21	4.57	23.68
20–21	−162.20	4.01	17.90
21–22	−169.41	4.01	18.06
22–23	−316.29	5.65	35.52
23–24	−253.32	4.94	27.71
24–25	230.13	2.05	4.54

1982). This high rate of change continued throughout the next year with the greatest change between months 23 and 24. The rate of change, however, stabilized between months 14 and 19 before oscillating and rising at the end suggesting that the network may have again begun to stabilize at this level of interaction. These rates of change indicate shifting of *levels* of communication rather than of the *pattern* of communication, so they do not reflect the stability of the fit of the data to the reciprocity model discussed earlier (Rice 1982).

The overall motion was negative because there was greater change on the imaginary dimension than on the real. As a result, motion on both dimensions was analyzed separately. There is considerably more motion on the imaginary dimension. The mean changes were 15.58 (imaginary) and 3.62 (real), a ratio of 4.3 to 1. The variances were also considerably different, 8.66 for the imaginary and only 1.06 for the real. Clearly, the positions of the groups on the imaginary dimensions were volatile. Had

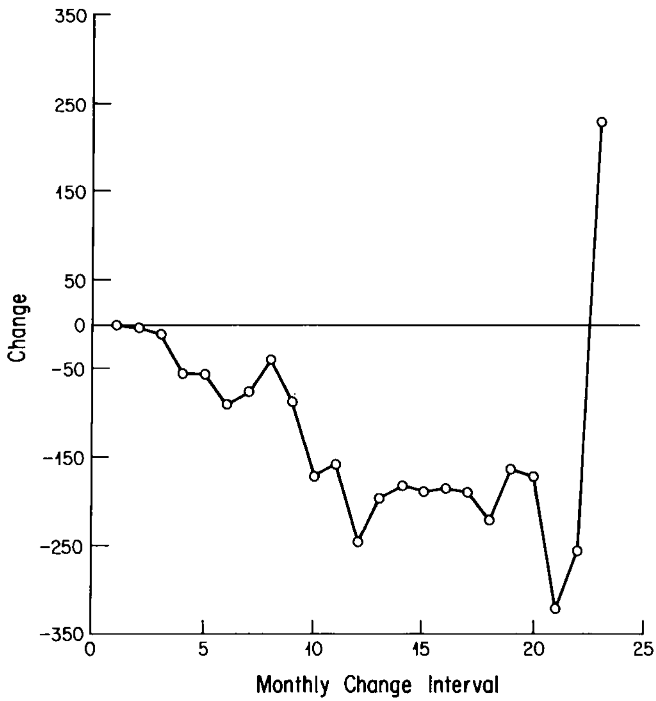

Figure 6. Overall change in coordinate values of computer conferencing network over time.
Note: See Table 8.

an Euclidean model been used to describe the changes in the network, this information would have been lost. A linear regression was also applied to each dimension separately. For the real dimension, $b = 0.138$, $a = 2.11$ and $R^2 = 0.77$. For the imaginary dimension, $b = 1.14$, $a = 2.94$ and $R^2 = 0.76$ ($F = 73$, $p < 0.001$ for both). These results suggest that the rate of change in the communication patterns in the network increased. There was an initial slow rate of change which then accelerated, oscillated, accelerated and showed a period of stability before accelerating at the end. The implications of different conclusions about network stability from different measures are discussed in Rice and Barnett (1985).

4.2.2. Summary

In this example, Galileo was applied to describe two years of computer conferencing network data. Derived measures of centrality, system connectedness and homogeneity of linkage were examined. How these indicators and the overall network structure changed were discussed. This network required a non-Euclidean multidimensional space because the groups varied in centrality. Two groups were near the origin of the space while the other 8 resided at the periphery. Thus, triads composed of either group and any other pair of nodes violated Euclidean properties. The spatial manifold which resulted from scaling the nodes' frequency of interaction was two-dimensional, one real and one imaginary. The imaginary dimension correlated 0.74 with itself at adjacent points in time despite changes in the network which occurred predominantly on this dimension. The linear component accounted for 76 percent of the variance in the change scores. Further, deviations from the linear were interpretable. They resulted from the entrance of groups into the system or the length of residence of the nodes. This example has provided further evidence of the utility of a Riemannian spatial manifold for describing social networks.

5. Discussion

This article has suggested theoretical and empirical rationales for using a non-Euclidean manifold when using a spatial model, such as MDS, to describe social networks. It advocated the use of a variant of metric MDS, Galileo, whose algorithm allows for the analysis of network data in Riemannian space. It then applied Galileo to two different sets of network data. In both cases the data were behavioral, rather than self-reported, and based upon a large number of interactions. Thus, the data are of sufficient accuracy and reliability to obviate discussions of measurement or instrument error. Further, these data may be easily obtained if one wishes to replicate the research or compare the results from Galileo with more traditional MDS algorithms.

The Galileo results described an orderly pattern of change in both networks. In the air traffic network, connectedness (as measured by the trace of the social distance matrix) increased over time. The change in connectedness was described by a function which accounted for 87 percent of the variance in connectedness. Further, the residuals could be accounted for by exogenous factors. A second function, a polynomial, significantly describes the reversal in the trend in connectedness. As the network became more connected, warp, the degree to which the space is non-Euclidean, became smaller. These results are readily interpretable by physical, historical and economic factors.

The results from the EIES network indicate that two groups are much more central than the other eight. Thus, the network structure remained highly non-Euclidean. Further, change in the system's connectedness was orderly. The rate of change was probably a function of whether or not new nodes entered the network and the nodes' length of residence. Again, these results were readily interpretable.

Internally consistent, theoretically valid predictions from exogenous variables were obtained with Galileo. For both data sets, loadings on the imaginary dimensions at adjacent points in time were highly correlated despite true change in the network. As predicted in the theoretical discussion, the levels of warp were much higher in the network in which a subset of nodes were more central than the rest.They ranged from 3.32 to 4.97 in the EIES network. For the highly interconnected airline network, the warps were much lower, 1.04 to 3.97. The highest value occurred when two of the nodes resided at the periphery.

The results from Galileo are consistent with other methods of network analysis. The data reported in the second example are consistent with those reported by Rice (1982), who used a log-linear analysis to describe transitions in the network. Barnett, *et al.* (1985) found that Galileo produced comparable results with NEGOPY (Richards and Rice 1981; Rice and Richards 1985) when used to identify cliques.

There are a number of relative advantages of Galileo over other methods of MDS for network analysis. One, its algorithm allows a non-Euclidean spatial manifold. This is a theoretical advantage when dealing with networks which vary greatly in centrality. Two, it provides a convenient measure, warp, which describes the structure's departure from being Euclidean. This measure along with the imaginary eigenroots and the nodes' loadings on these dimensions provides a detailed description of the network. Three, Galileo is appropriate for longitudinal analysis. It automatically provides detailed information on the changes in the node's relative positions over time on both the real and imaginary dimensions. This is done without standardizing the data. In this way, it is possible to describe the dilation and contraction of the network space, and thus make statements about its changing density. Four, Galileo is a metric scaling program. This is required for regressing exogenous factors through single spaces and fitting curves to the changes in longitudinal data. Fifth, Galileo provides a measure of centrality. It may be determined by taking the square root of the diagonal of the scalar products matrix.

There are some problems with Galileo. Its application is limited to relatively small networks. Currently, the software is limited to 40 notes and 40 points in time, although one could change the program's dimension statements for larger networks. Galileo is designed for symmetrical matrices ($s_{ij} = s_{ji}$). In those cases where the researcher is interested in directional or non-reciprocal networks, special procedures are required. One would have to create two matrices S (send) and S (receive), and then compare as if they were separate points in time. Galileo requires variance in strength of link (frequency of interaction). It may not converge to a solution if the cells in the matrix are too

[11] One interpretation of the Riemannian distances in social networks involves the concept of transitivity. Transitivity posits the presence of links among a triad of nodes, and thus some small social distance among them. That is, if *a* talks with *b* and *b* talks with *c*, then *c* talks with *a*. The direction of these linkages is an important aspect of transitivity which is not discussed here. An entire methodology of analyzing these "local structures," or triad census, is heavily concerned with the amount and distribution of transitivity (Holland and Leinhardt 1976). Most studies of closed networks do, indeed, find triads moving away from intrasitivity over time, although the majority of structures remain intransitive (Killworth 1974).

It may be argued that Euclidean space allows for a continuum of transitiveness, from completely connected triads in a network, to completely unconnected isolates. A network in the aggregate is more or less transitive. In that sense, Euclidean space allows for the full range of network structure. However, two aspects of transitivity are confounded in this portrayal. This notion of transitivity allows only for the presence or absence of links. That is, a triad is transitive or it is not, though there are many varieties of intransitivity (see Holland and Leinhart 1976; Killworth 1974), but that is determined solely by where links exist or do not exist.

But, what is to be done about a relation among, i, j and k where $ij = 5$, $jk = 3$ and $ik = 17$? That is transitive and clearly non-Euclidean. Nodes k and i have frequent or strong or important messages, and no Euclidean representation can reconcile this with the other two relationships, except by transforming scalar values to binary, by ignoring complex roots in the mathematical solutions, or by forcing all values to conform to Euclidean calculations (as in Burt's 1980 method for computing social distances).

Further, the assumption of transitivity in social relations as being predominant seems unnecessary, and sometimes contradictory to the realities and economics of social structure. Some kinds of nonreciprocity and structural differentiation enhance a group's stability, and intransitivity and liaisons promote individual survivability by preventing direct competition for resources (Amir 1979; Killworth 1974). Further, groups or individuals with partially differentiated resources can operate more efficiently by depending on liaisons (which may imply intransitivity) to manage some resource acquisitions (such as new information - Granovetter 1973). Interest in an analogous form of intransitivity - "Q-holes" in Q-analysis or the relations of relations (Freeman 1980) - may be specified by similar arguments. Insofar as intransitivity in social relations (the absence of some linkages or large local distances) implies non-Euclidean space, analyses which assume Euclidean space may be insufficient.

similar. Thus, one must use more than simply binary data (link–no link) or Likert type items to describe relationships among nodes.

The intent of this article has been to discuss some of the utilities in using a multidimensional scaling algorithm which provides the capability of analysing non-Euclidean networks over time. Galileo is one such algorithm, and, as such, should be considered as one more available tool on the network analyst's workbench. Its advantages and disadvantages should be considered when choosing how best to analyze one's data, and in thinking about the assumptions underlying the measurement and meaning of network data. [11]

References

Amir, S.
1979 "Economic interpretations of equilibrium concepts in ecological systems." *Journal of Social and Biological Structure*, 2:293–314.

Barnett, G.A.
1979 "Spatial modelling of social networks with applications to the diffusion process." Paper presented to the International Communication Association, Philadelphia, May.
1980 "Inertial mass as a function of frequency of occurrence." Paper presented to the International Communication Association, Acapulco, May.
1982 "Seasonality in television viewing: A mathematical model." Paper presented to the International Communication Association, Boston, May.
1984 "Precise procedures for longitudinal network analysis." Paper presented to the Sunbelt Social Networks Conference, Phoenix, February.

Barnett, G.A., R.W. Bales and K.D. Day
1985 "A general systems approach to network analysis." Paper presented to the Sunbelt Social Networks Conference, Palm Beach, Florida, February.

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

Berger, C. and M. Roloff
1980 "Social cognition, self-awareness and interpersonal communication." In Dervin, B. and Voigt, M. (eds.) *Progress in Communication Sciences, Vol. II.* Norwood, NJ: Ablex.

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

Bernard, H.R., P. Killworth and L. Sailer
1980 "Informant accuracy in social network data IV." *Social Networks* 2: 191–218.
1982 "Informant accuracy in social network data V." *Social Science Research* 11: 30–66.

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

Burt, R.
1980 "Models of network structure." *Annual Review of Sociology* 6: 79–141.

Burt, R. and M. Minor

1983 *Applied Network Analysis: A Methodological Introduction.* Beverly Hills, CA: Sage.

Danowski, J.A.

1980 "Uniformity of group attitude-belief and connectivity of organizational communication networks for production, innovation and maintainance." *Human Communication Research* 6: 229–308.

1982 "Computer-mediated communication A network-based content analysis using a CBBS conference." In Burgoon, M. (ed.) *Communication Yearbook 6.* Beverly Hills, CA: Sage, pp. 905–924.

Freeman, L.C.

1979 "Centrality in social networks." *Social Networks* 1: 215–239.

1980 "Q-analysis and the structure of friendship networks." *International Journal of Man-Machine Studies* 12 (3): 367–378.

1984 "The impact of computer based communication on the social structure of an emerging scientific specialty." *Social Networks* 6: 201–221.

Freeman, S.C. and L.C. Freeman

1979 "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.

Gillham, J.R. and J. Woelfel

1977 "The Galileo system of measurement: preliminary evidence for precision, stability, and equivalence to traditional measures." *Human Communication Research* 3: 222–234.

Goldstein, K.M., S. Blackman and D.J. Collins

1966 "Relationship between sociometric and multidimensional scaling measures." *Perceptual and Motor Skills* 23: 639–643.

Granovetter, M.

1973 "The strength of weak ties." *American Journal of Sociology.* 78, 1360–1380.

Griffith, D.A.

1984 "Identifying metric spaces associated with transport networks." Proposal to the National Science Foundation. Department of Geography, State University of New York, Buffalo.

Hamblin, R.L.

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

Hiltz, S.R.

1983 *Online Communities: A Case Study of the Office of the Future.* Norwood, NJ: Ablex.

Hiltz, S.R. and M. Turoff

1978 *The Network Nation: Human Communication via Computer.* Reading, MA: Addison-Wesley Advanced Book Program.

Holland, P. and S. Leinhardt

1976 "The statistical analysis of local structure in social networks." In Heise, D. (ed.) *Sociological Methodology, 1975.* San Francisco, CA: Jossey-Bass pp. 1–45.

Jones, L.E. and F.W. Young

1972 "Structure of a social environment: Longitudinal individual differences of an intact group." *Journal of Personality and Social Psychology* 24: 108–121.

Kerr, E. and Hiltz, S.R.

1982 *Computer-Mediated Communication Systems.* NY: Academic Press.

Killworth, P.

1974 "Intransitivity in structure of small closed groups." *Social Science Research* 3: 1–23.

Knoke, D. and J. Kuklinski
1982 *Network Analysis. Quantitative Applications in the Social Sciences, Vol. 28.* Beverly Hills, CA: Sage.
Krumhansl, C.L.
1978 "Concerning the application of geometric models to similarity data: The interrelationship between similarity and spatial density." *Psychological Review* 85: 445–463.
Kruskal, J.B.
1964a "Multidimensional scaling by optimizing goodness of fit to a non-metric hypothesis." *Psychometrika* 29: 1–28.
1964b "Non-metric multidimensional scaling: a numerical method." *Psychometrika* 29: 115–159.
Lankford, P.M.
1974 "Comparative analysis of clique identification methods." *Sociometry* 37: 287–305.
Levine, J.
1972 "The sphere of influence." *American Sociological Review* 37: 14–27.
Lindman, H. and T. Caell
1978 "Constant curvative Riemannian scaling." *Journal of Mathematical Psychology* 17: 89–109.
Messick, S. and R. Abelson
1956 "The additive constant problem in multidimensional scaling." *Psychometrika* 21: 1–15.
Monge, P.
1982 "Systems theory and research in the study of organizational communication: The correspondence problem." *Human Communication Research* 3: 245–261.
Moreno, J.
1960 *The Sociometry Reader.* Glencoe, IL: Free Press.
Nisbett, R. and T. Wilson
1977 "Telling more than we can know: Verbal reports on mental processes." *Psychological Review* 84: 231–259.
Olsson, G.
1965 *Distance and Human Interaction.* Philadelphia, PA: Philadelphia Regional Science Research Institute.
Pieszko, H.
1975 "Multidimensional scaling in Riemannian space." *Journal of Mathematical Psychology* 12: 449–477.
Rice, R.E.
1980a "The impacts of computer-mediated organizational and interpersonal communication." In M. Williams (ed.) *Annual Review of Information Science and Technology, Vol. 15.* White Plains, NY: Knowledge Industry Publications pp. 221–249.
1980b "Computer conferencing." In B. Dervin and M. Voigt (eds.) *Progress in Communication Sciences, Vol. II.* Norwood, NJ: Ablex, pp. 215–240.
1981 "Resources for longitudinal network analysis." *Connections* 4: 10–21.
1982 "Communication networking in computer-conferencing systems: A longitudinal study of groups, roles and system structure." In M. Burgoon (ed.) *Communication Yearbook 6.* Beverly Hills, CA: Sage pp. 925–944.
Rice, R.E. *et al.*
1984 *The New Media: Communication, Research and Technology.* Beverly Hills, CA: Sage.
Rice, R.E. and G.A. Barnett
1985 "Group communication networking in an information environment: Applying metric multidimensional scaling." In M. McLaughlin (ed.) *Communication Yearbook 9.* Beverly Hills, CA: Sage.

Rice, R.E. and C. Borgman
1983 "The use of computer-monitored data in information science and communication research." *Journal of the American Society for Information Science* 34 (4): 247–256.

Rice, R.E. and D. Case
1983 "Electronic message systems in the university: A description of use and utility." *Journal of Communication* 33 (1): 131–152.

Rice, R.E. and W. Paisley
1982 "The green thumb videotex experiment: Evaluation and policy implications." *Telecommunications Policy* 6 (3): 223–235.

Rice, R.E. and W. Richards Jr.
1985 "An overview of network analysis methods." In B. Dervin and M. Voigt (eds.) *Progress in Communication Sciences, Vol. VI.* Norwood, NJ: Ablex.

Rice, R.E. and E.M. Rogers
1984 "New methods and data for the study of new media." In Rice, R.E. and Associates. *The New Media: Communication, Research and Technology.* Beverly Hills, CA: Sage pp. 81–100.

Richards, W. Jr. and R.E. Rice
1981 "Negopy network analysis program." *Social Networks* 3 (3): 215–223.

Rogers, E.M. and D.L. Kincaid
1981 *Communication Networks: Towards a New Paradigm for Research.* NY: Free Press.

Romney, A.K. and K. Faust
1982 "Predicting the structure of a communications network from recalled data." *Social Networks* 4: 285–304.

Schonemann, P.H. and R.M. Carroll
1970 "Fitting one matrix to another under choice of a central dilation and a rigid motion." *Psychometrika* 35: 245–256.

Shepard, R.N.
1962a "The analysis of proximities: multidimensional scaling with an unknown distance function I." *Psychometrika* 27: 125–140.
1962b "The analysis of proximities: multidimensional scaling with an unknown distance function II." *Psychometrika* 27: 219–246.

Shepard, R.N., A.K. Romney and S. Nerlove (eds.)
1972 *Multidimensional Scaling: Theory and Applications in the Behavioral Sciences.* NY: Seminar Press.

Torgerson, W.S.
1958 *Theory and Methods of Scaling.* NY: Wiley.

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

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

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

Woelfel, J. and E.L. Fink
1980 *The Measurement of Communication Processes: Galileo Theory and Method.* NY: Academic.

Woelfel, J., E.L. Fink, K.B. Serota, G.A. Barnett, R. Holmes, M. Cody, J. Saltiel, J. Marlier and J.R. Gillham
1977 *Galileo-A Program for Metric Multidimensional Scaling.* Honolulu, HI: East-West Communication Institute.